March 18, 2004

AGE-PERIOD-COHORT ANALYSIS OF REPEATED CROSS-SECTION SURVEYS: FIXED OR RANDOM EFFECTS? *

Yang Yang and Kenneth C. Land

Department of Sociology and Center for Demographic Studies

Duke University

*Paper prepared for presentation at the Annual Meeting of the Population Association of America, Boston, MA, April 1-3, 2004. We thank Guang Guo for comments on another paper that led to the research reported in this paper. The research reported herein was supported in part by NIH/NIA Grant Numbers R01AG07198, P30AG12852, and K07AG00892. Address all correspondence to Yang Yang, Department of Sociology and Center for Demographic Studies, PO Box 90088, Duke University, Durham, NC 27708-0088; e-mail: **yangy@soc.duke.edu**

Abstract

Yang and Land (2003) and Yang (2004) developed a mixed (fixed-random) effects model for the age-period-cohort (APC) analysis of micro datasets in the form of a series of repeated cross-section sample surveys that are increasingly available to demographers. To estimate the mixed effects APC models, Yang and Land applied the statistical methodology of hierarchical regression models by treating cohort and period effects as random. An alternative approach to model specification could be based on a purely fixed- cohort-and-period-effects regression model, which does not require large numbers of cohorts and/or periods for reliable statistical estimation and the assumption that the period and cohort effects are independent of individual-level regressors. We examine these assumptions and compare the fixed- versus random-effects model specifications for APC analysis. We continue to use the data on verbal test scores from 15 cross-sections of the General Social Survey, 1974-2000, for substantive illustrations. Strengths and weaknesses are identified for both the random- and fixed-effects formulations. However, under each of the two data conditions studied – one with a moderate number of cohorts and time periods and another with a small number of cohorts and periods – the random effects hierarchical age-period-cohort model is the most appropriate specification.

AGE-PERIOD-COHORT ANALYSIS OF REPEATED CROSS-SECTION SURVEYS: FIXED OR RANDOM EFFECTS?

INTRODUCTION

For the past 80 years or so, demographers, epidemiologists, and social scientists 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, 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 some other factors such as age or calendar year are more relevant.

The Age-Period-Cohort (APC) accounting/multiple classification model developed by Mason, Mason, Winsborough, and Poole (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, crimes, etc. 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 age-period-cohort models have occurred in recent decades in a wide variety of disciplines, including social and demographic research (e.g., Fienberg and Mason 1978, 1985; Glenn 1976; Hobcraft, Menken, and Preston 1982; O'Brien 2000; Wilmoth 1990), biostatistics and epidemiology (e.g., Clayton and Shifflers 1987; Fu 2000; Holford 1992; Knight and Fu 2000; Kupper, Janis, Salama, Yoshizawa, and Greenberg 1983; Osmond and Gardner 1982; Roberton and Boyle 1998). Most of these studies focus on aggregate population data where researchers have few choices of time-interval widths for the age and period groups.

Increasingly, however, micro datasets in the form of series of repeated cross-section sample surveys are available to demographers, epidemiologists, and social scientists. These datasets create both new opportunities and challenges to APC analysis. The opportunities lie in the fact that these repeated cross-section survey data can not only be aggregated into population-level contingency tables for conventional multiple classification models, but 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 (Yang and Land 2003). The challenge for APC analysis of repeated cross-section surveys then becomes how social scientists can take advantage of the individual-level data in these datasets as opposed to grouping the data.

While straightforward regression analyses on the micro sample data can be conducted, Yang and Land (2003) noted that this may violate the independence-of-errors assumption on which conventional fixed-effects regression models (e.g., ordinary least squares or logit regression) are based. They developed a *hierarchical age-period-cohort models* (HAPC models) approach to address this problem. Specifically, they applied two-level and three-level models to APC data to ascertain whether or not there are any clustering effects in survey responses by higher level units, survey year or birth cohort. In recognition of a more complex data structure in which lower-level units (individuals) are cross-classified by two higher-level units (cohorts and periods), Yang (2004) refined this HAPC approach to micro APC data by introducing a cross-classified random-effects model (CCREM).

Both Yang and Land's (2003) and Yang (2004)'s approaches prove to be promising strategies for APC analyses. In this paper, we compare the random-effects model specification to an alternative specification, namely, a fixed-effects hierarchical APC model – with a focus on primarily two considerations. First, the assumption of the random-effects model that the level-2 effects are independent of the level-1 regressors needs to be examined. Second, in this HAPC application of hierarchical linear models, the sample sizes at the level-2 or contextual effects-level, that is, the numbers of birth cohorts and periods, are small and therefore can be viewed as specific groups. Cohort and period effects, therefore, may just as easily be viewed as fixed rather than random. In this paper, we lay out and compare these two approaches side-by-side, and we exploit the properties of each model to assess the empirical applicability of the independence assumption. As a substantive illustration, we continue to analyze 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. While the primary focus of this paper is methodological, we note how our approach can be used to address these debates by identifying and estimating the separate age, period, and cohort components of change.

The paper is organized as follows. In the next section, we briefly review the APC identification problem and recent methodological developments in solving this problem. We then summarize conventional methodological guidelines for and against the treatment of certain effect parameters in hierarchical regression models as fixed or random and guidelines for model specifications. This is followed by sections that describe the data to be analyzed and the models to be compared. Results then are reported. A conclusion section discusses the findings and reports conclusions from this methodological study.

THE APC ACCOUNTING MODEL AND THE IDENTIFICATION PROBLEM

In order to specify the models to be estimated, it will be useful to review briefly the *Age-Period-Cohort accounting/multiple classification model* that was articulated for demographic and social research some 30 years ago by Mason, Mason, Winsborough, and Poole (1973). For mortality rates tabulated in standard arrays with age groupings defining the rows, periods of data defining the columns, and cohorts defining the diagonals, this model can be written in *linear regression form* as:

$$M_{ij} = D_{ij} / P_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij} \tag{1}$$

where M_{ij} denotes the observed occurrence/exposure rate of deaths for the *i-th* age group for i = 1, …, a age groups at the j-th time period for j = 1, …, J, D_{ij} denotes the number of deaths in the ij-th group, P_{ij} denotes the size of the estimated population in the ij-th group, the population at risk of death, μ denotes the intercept or adjusted mean death rate, α_i denotes the *i-th* row age effect or the coefficient for the *i-th* age group, β_j denotes the *j-th* column period effect or the coefficient for the *j-th* time period, γ_k denotes the *k-th* diagonal cohort effect or the coefficient for the *k-th* cohort for $k = 1,\ldots,(a+p-1)$, with $k=a-i+j$, and ε_{ij} denotes the random errors with expectation $E(\varepsilon_{ij}) = 0$.

Conventional age-period-cohort models as represented in Eq. (1) fall into the class of generalized linear models (GLIM) that can take various alternative forms such as the *log-linear regressions* and *logit regressions*. They can be treated as *fixed effect generalized linear models* after a reparametrization to center the parameters:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0 \tag{2}$$

Alternatively, constraints may be set by identifying one of each of the age, period, and cohort categories as the *reference category*.

After the re-parameterization of Eq. (2), model (1) can be written in the conventional matrix form of a least-squares regression:

$$Y = X\beta + \varepsilon \tag{3}$$

where Y is a vector of mortality rates or log-transformed rates, X is the regression design matrix consisting of "dummy variable" column vectors for the vector of model parameters β:

$$\beta = (\mu, \alpha_1, \ldots \alpha_{a-1}, \beta_1, \ldots, \beta_{p-1}, \gamma_1, \ldots, \gamma_{a+p-2})^T \tag{4}$$

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. (Note: In order to keep to standard linear regression model notation, we use β to denote the entire vector of model coefficients in equation (3).) The ordinary least squares (OLS) estimator of the matrix regression model (3) is the solution b of the normal equations:

$$b = (X^T X)^{-1} X^T Y \tag{5}$$

But this estimator *does not exist* (i.e., is not a unique vector of coefficient estimates) since the design matrix X is singular with one less than full rank and $(X^T X)$ is not invertible, which is due to a perfect linear relationship between the age, period and cohort effects:

Period – Age = Cohort.

This is the *model identification problem* of APC analysis. It implies that there are an infinite number of possible solutions of the matrix equation (5) (i.e., OLS estimators of model (3)), one for each possible linear combination of column vectors that results in a vector identical to one of the columns of X. Therefore, it is not possible to separately estimate the effects of cohort, age, and period without assigning certain constraints to the coefficients in addition to the reparameterization (2). Since Fienberg and Mason (1978), the conventional approach to identification of the APC accounting model in demography has been to impose equality constraints on two or more coefficients of the parameter vector (4). Yang, Fu, and Land (2004) compare this standard approach with the *Intrinsic Estimator* (IE) method, an approach to the identification problem that yields a unique solution to (5) determined by the Moore-Penrose generalized inverse and removes the arbitrariness brought about by constraints on parameters. This comparison shows that the IE possesses the desirable statistical properties of unbiasedness and relative statistical efficiency in APC analyses of tabulated rates with a finite number of time periods of data.

SOLVING THE IDENTIFICATION PROBLEM IN REPEATED CROSS-SECTION DATA DESIGNS – HAPC MODELS

In the context of micro-level data from repeated cross-section sample surveys, Yang and Land (2003) noted that the existence of individual-level data on the age, period, and cohort membership of each respondent in the surveys opens up new opportunities for solving the APC identification problem. Specifically, access to the individual-level observations allows the analyst to group the age, period, and/or cohort properties of sample members into time intervals of different lengths. This breaks the underidentification problem of Eq. (3) and allows finite-valued numerical solutions of Eq. (5) to exist.

To illustrate this, it will be helpful to consider the application of the classical APC accounting model of Eq. (3) to, say, the following five sample members, ages 60, 61, 62, 63, and 64, each of whom is a member of the same five-year birth cohort, the 1930-34 birth cohort, and each of whom is a respondent in a sample survey conducted in 1990:

$$Y_{60,1990,1930\text{-}34} = \mu + \alpha_{60} + \beta_{1990} + \gamma_{1930\text{-}34} + \varepsilon_{60,1990,1930\text{-}34} \quad (6.1)$$

$$Y_{61,1990,1930\text{-}34} = \mu + \alpha_{61} + \beta_{1990} + \gamma_{1930\text{-}34} + \varepsilon_{61,1990,1930\text{-}34} \quad (6.2)$$

$$Y_{62,1990,1930\text{-}34} = \mu + \alpha_{62} + \beta_{1990} + \gamma_{1930\text{-}34} + \varepsilon_{62,1990,1930\text{-}34} \quad (6.3)$$

$$Y_{63,1990,1930\text{-}34} = \mu + \alpha_{63} + \beta_{1990} + \gamma_{1930\text{-}34} + \varepsilon_{63,1990,1930\text{-}34} \quad (6.4)$$

$$Y_{64,1990,1930\text{-}34} = \mu + \alpha_{64} + \beta_{1990} + \gamma_{1930\text{-}34} + \varepsilon_{64,1990,1930\text{-}34} \quad (6.5)$$

It can be seen from this representation of the classical APC model (1) for these sample members that the exact linear dependence of the A, P, and C categories that occurs in tabulated data with age, period, and cohort time intervals of equal length is broken.[1] That is, from knowledge of the period of this survey (1990) and the birth cohort (1930-34) in which each of these samples respondents is a member, it is not possible to determine the exact age of each respondent. In fact, the general five-year age category (60-64) of which each of these respondents is a member can be determined. But, at the level of the individual respondent, this is not an exact linear dependence.

Under Eq. (6), the specification of the error terms indicates the possibility that sample respondents in the same cohort group and/or survey year may be similar in their responses to the verbal test questions due to the fact that they share random error components (i.e., through random cohort and/or period components of $e_{i,1990,1930\text{-}34}$) unique to their cohorts or periods of the survey. 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 standard regression models like Eq. (3) may be underestimated, leading to inflated t-ratios and actual alpha levels that are larger than the nominal .05 or .01 levels[2]. This implies that multilevel or hierarchical regression models should be employed for adequate estimates of the error variances (Goldstein 2003; Raudenbush and Bryk 2002; Snijders and Bosker 1999). Yang and Land (2003) proposed both two-level and three-level APC model specifications and labeled this approach as the Hierarchical APC (HAPC) model.

---

[1] In their empirical application of the APC model to the verbal test score data described later in the text, Yang and Land (2003) noted that the linear dependence/identification problem could be further broken by using the results of their application of the Intrinsic Estimator to the verbal test score data aggregated and grouped into standard APC table format. This application of the IE found that the age curve of the data was well described as a quadratic function of age. By using this information to specify the relationship between an individual's verbal test score and her/his age as quadratic, Yang and Land further broke the underidentification problem of the APC accounting model. This quadratic specification of the individual-level age curve also is used in the models analyzed in the present paper.

[2] 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.

The two-level HAPC is useful if we separately consider the periods or the cohorts as level-2 units. The three-level HAPC is useful when we are interested in birth cohorts at level-2 who are nested within periods at level-3. But it requires a strict nested 3-level design where level-1 units are nested within level-2 units, which are then completely nested within level-3 units. So it is assumed that each individual belongs to one and only one cohort, and each cohort belongs to one and only one period. This is a simplification of the APC data. In a recent paper that further extends the HAPC models, Yang (2004) pointed out that in typical cross-sectional surveys, 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. Yang (2004) therefore formulated a cross-classified random-effects APC model (CCREM) to account for this particular APC data structure.

The HAPC modeling framework developed by Yang and Land (2003; Yang 2004) has enhanced our ability to estimate separate age, period, and cohort effects through the estimation of variance components. It also enriches the families of analytical models in demography that can be applied to study APC trends by incorporating explanatory variables in hierarchical regression models. In this paper, we consider two additional conditions that may affect applications of the HAPC models in demographic studies. First, as is the case for all mixed effects regression models, desirable statistical properties of HAPC models rest on the assumption that the level-2 or contextual effects – the cohort and period effects – are independent of the level-1 or individual-level regressors. An alternative approach for HAPC models could be based on a fully fixed-effects HLM regression formulation for which this independence assumption is not necessary. Second, in finite time period demographic data, the number of periods or birth cohorts are usually smaller than 10. As commented below, this suggests a possible advantage in modeling these effects as fixed. In the following section, we discuss the rationale of the model specification in more details.

WHEN TO USE FIXED OR RANDOM COEFFICIENT MODELS? THE CONVENTIONAL WISDOM

The literature on hierarchical/multilevel regression models contains some general guidelines on when certain effect coefficients should be treated as fixed or random. To articulate these guidelines, consider the simplest possible hierarchical model that has a level-1 or individual-level model:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij} \qquad (7)$$

where Y_{ij} is a response or dependent variable for individual i in group j, x_{ij} is an explanatory variable or regressor for individual i in group j, β_0 is the intercept parameter for the regression model, β_1 is the slope parameter of the regression, and e_{ij} is a random error term. Suppose that the intercept β_0 is group-dependent and that it varies randomly among J observed groups. To model this random variation, we specify the level-2 or group-level model:

$$\beta_0 = \gamma_{00} + r_{0j}\,. \qquad (8)$$

This level-2 model separates the group-dependent intercept into an average intercept among the groups, γ_{00}, plus a group-level deviation or error, r_{0j}. Substitution of Eq. (8) into Eq. (7) then yields the combined model:

$$Y_{ij} = \gamma_{00} + \beta_1 x_{ij} + r_{0j} + e_{ij}\,. \qquad (9)$$

The values of the r_{0j} are the main effects of the groups: conditional on having a specific X-value and being in group j, the expected Y-value for individual i deviates by r_{0j} from the average expected value for all individuals over all groups. Note again that this is the simplest possible formulation of a hierarchical model; a more general formulation would allow for the possibility that the slope coefficient in Eq. (7) could vary among the groups, and there could be more than one explanatory variable in Eq. (7).

As a statistical model, Eq. (9) can be treated in two ways:

(1) As a *fixed-effects model*, wherein the r_{0j} are treated as *fixed* parameters, J in number. This approach leads to specific instance of a fixed-effects regression model, namely, the conventional analysis of covariance model, in which the grouping variable is a covariate.

(2) As a *random coefficients* or *random intercepts model*, wherein the r_{0j} are assumed to be independent identically distributed *random* variables. These errors now are assumed to be randomly drawn from a population with zero mean and an *a priori* unknown variance. This assumption is equivalent to the specification that the group effects are governed by mechanisms or processes that are roughly similar from group to group and operate independently among the groups. This is termed the *exchangeability* assumption. The random coefficients model also requires the assumption that the random level-2 or contextual effects, i.e., the r_{0j} coefficients, are distributed *independently* of the level-1 regressors.

These two approaches to the model of Eqs. (7)-(9) imply that hierarchical data generally can be analyzed in two different ways, using models with fixed or random group-level coefficients. Which of these two specifications is the most appropriate in a given situation depends on a number of considerations.

Goldstein (2003, pp. 3-4) and Snijders and Bosker (1999, pp. 43-44) provide summaries of conventional statistical wisdom and methodological guidelines for choosing between the fixed or random specifications. They point out that:

- If the groups are regarded as *unique entities* and the objective of the analysis is primarily to draw conclusions pertaining to each of the *J* groups, then it is appropriate to use the conventional analysis of covariance model.
- If the groups are regarded as a *sample* from a (real or hypothetical) population and the objective of the analysis is to make inferences about this population, then the random coefficients model is appropriate.
- The *fixed-effects model explains all differences among the groups* by the fixed-effect adjustments (through the use of indicator or dummy variables to represent the group-level adjustments) to the intercept coefficient of Eq. (7). This implies that there is no between-group variability left that could be explained by group-level variables. Therefore, if the objective of the analysis is to test effects of group-level variables, the random coefficient model should be used. The exception to this guideline pertains to the case wherein the analyst introduces explicitly measured group-level variables that are hypothesized to account for the group-level effects.[3] In this case, however, the model cannot at the same time incorporate indicator variables for the group-level fixed-effect adjustments. Rather, the analyst must assume that the group-level fixed-effect adjustments are completely accounted for by the explicitly measured group-level variables.

[3] In the context of the age-period-cohort models that are the subject of this paper, such an incorporation of explicitly measured group-level variables modifies the basic APC accounting model described later in the text towards that of the age-period-cohort-characteristics model as described by O'Brien (2000) and variations thereon.

- The random coefficient model typically is used with some additional assumptions. Most importantly, as noted above, it requires that the *random residuals by groups are orthogonal to independent variables*, which implies that $corr(r_{0j}, X) = 0$. In addition, in conventional normal errors HLM models, it is assumed that the *random coefficients r_{0j} and e_{ij} are normally distributed*. If these assumptions are poor approximations to the characteristics of a specific set of empirical data (e.g., the regressors are not independent of the random coefficients, or there is high density in the tails of the distributions of the errors), then they assumptions should be modified.
- The choice between fixed- and random-effects formulations can be related to sample sizes. Snijders and Bosker (1999, p. 44) state that the following *rule of thumb* often works in educational and social research: when J, the number of groups is small, say $J < 10$, use the analysis of covariance approach, because the small number of groups does not contain sufficient information about the population of groups from which the J groups are sampled to make reliable inferences; if J is not small, say $J \geq 10$, but n_j, the number of observations in group j is small or of moderate size, say $n_j < 100$, then use the random coefficients model, as 10 or more groups is too large to regard each group as a unique entity; if the group sizes are large, say $n_j \geq 100$, then it does not matter which view we take.

These, then, are several of the main considerations that conventional statistical wisdom indicates should be taken into account in deciding of fixed versus random-effects formulations of hierarchical statistical models. After describing the data and the specific model specifications to be studied herein, we assess how well these guidelines hold up in the context of the HAPC and CCREM models developed by Yang and Land (2003) and Yang (2004).

THE VERBAL TEST SCORES CONTROVERSY AND DATA

Questions Regarding Trends in Verbal Ability

A series of articles published in the *American Sociological Review* in 1999 center upon the existence of an intercohort decline in verbal ability in the GSS 1974 to 1996. The debate was initiated by Alwin (1991) and Glenn's (1994) finding of a long-term intercohort decline in verbal ability beginning in the early part of the 20th century. Wilson and Gove (1999) 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 (1999: 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.

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 some aspects of these studies invite further examination before definitive conclusions can be drawn. First, although the graphs presented in Wilson and Gove (1999) 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, Fu, and Land 2003). 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, but need to be made through statistical modeling.[4]

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 (1999) estimated age-period regression models for four age groups; in reply to Wilson and Gove, Glenn 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.

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

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

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 are, therefore, presented for the purpose of illustration. A full substantive analysis will follow in another paper.

The General Social Survey (GSS) Data and Variables

We analyze verbal test score data from 15 cross-sections of the General Social Survey (GSS), 1974-2000. This is an extension of the 1974-1996 data on which the controversy is based to the most recently available wave of the GSS. In these surveys, a survey respondent's verbal ability 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 zero to 10. WORDSUM is reported by previous studies to have an internal reliability of .71 (Wilson and Gove 1999). The data include 19,500 respondents who had measures on WORDSUM and other covariates across all survey years. Variable descriptions and summary univariate statistics for the data used herein are shown in Table 1.

[Table 1 about here]

The variable WORDSUM is available for 15 survey years of the GSS from 1974 to 2000. It is approximately normally distributed with a mean around 6. The respondents age from 18 to 89. The oldest cohort member was born in 1890 and the youngest born in 1982. We grouped individuals into 19 five-year birth cohorts for our analyses. In the pooled cross-section data, 57% are female and 15% are Black. The average years of education completed is around 12.7 years.[5]

Like typical cross-sectional surveys, the data on verbal ability include individuals who are nested within cells created by the cross-classification of two types of social context: birth cohorts and survey years. This data structure is displayed in Table 2. Each row is a cohort and each column is a year. The numbers in this matrix are the counts – the numbers of individuals who belonged to a given birth cohort and were surveyed in a given year.

[Table 2 about here]

We next describe the format of the cross-classified random effects model (CCREM) that Yang (2004) applied to the GSS verbal test score data. We then describe the format of the corresponding fixed effects hierarchical model for analysis of these data with which comparisons will be reported below.

HIERARCHICAL APC MODELS OF THE GSS DATA: CCREM vs. CCFEM

Yang (2004) specified a *cross-classified random-effects model* (CCREM) for the APC analysis to assess the relative importance of the two contexts, cohort and period, in understanding individual differences in verbal test outcome. Two prominent examples of

---

[5] In addition to the individual-level explanatory variables, Yang and Land (2003) and Yang (2004) used contextual variables in their analysis that represent cohort characteristics. As noted earlier in the text, however, it is not possible to use explicitly measured group-level variables in fixed-effects multilevel models without making the assumption that the measured group-level variables account for all of the group-level effects – cohort and period effects in the present context. For this reason, and in order to keep the comparison of the random and fixed effects specifications as direct and simple as possible, the analyses reported herein will not incorporate group-level measured variables.

applications of such models to social data can be found in Raudenbush's (1993, 2002) 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 such a model applied to the verbal test data, one specifies variability in WORDSUM associated with individuals, cohorts, and periods.[6]

Level-1 or "Within-Cell" Model:

$$WORDSUM_{ijk} = \beta_{0jk} + \beta_1 AGE_{ijk} + \beta_2 AGE^2_{ijk} + \beta_3 EDUCATION_{ijk} + \beta_4 FEMALE_{ijk} + \beta_5 BLACK_{ijk} + e_{ijk}$$
$$e_{ijk} \sim N(0, \sigma^2) \quad (10)$$

Level-2 or "Between-Cell" Model:

$$\beta_{0jk} = \gamma_0 + u_{0j} + v_{0k}\,,\ u_{0j} \sim N(0, \tau_u),\ v_{0k} \sim N(0, \tau_v) \quad (11)$$

Combined Model:

$$WORDSUM_{ijk} = \gamma_0 + \beta_1 AGE_{ijk} + \beta_2 AGE^2_{ijk} + \beta_3 EDUCATION_{ijk}$$
$$+ \beta_4 FEMALE_{ijk} + \beta_5 BLACK_{ijk} + u_{0j} + v_{0k} + e_{ijk} \quad (12)$$

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, the respondent i's verbal score is modeled as a function of age, age squared, education, gender and race, the intercept then varies by birth cohort and time period, and all continuous covariates are centered around their means.

In this CCREM, 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 leve-1 fixed

[6] Note that this CCREM is a random-intercepts model, which is based on previous works by Yang (2004) and Yang and Land (2003). They suggest that only the intercepts, but not level-1 slopes, exhibit significant random variation across cohorts and periods in the GSS verbal test score data.

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 averaged over all periods; and $\beta_{0k} = \gamma_0 + v_{0k}$ is the period verbal test score averaged over all cohorts.

We seek to compare parameter estimates of the CCREM of Eqs. (10)-(12) with those obtained from a corresponding *cross-classified fixed-effects model* (CCFEM) where the effects of the cohorts u_{oj}, $j = 1,\ldots,J$ and the effects of the time periods (years) of the surveys v_{ok}, $k = 1,\ldots,K$ are assumed fixed and unique to each of the respective cohorts and period rather than variable and random. In practice, the fixed effects of the cohorts and periods are estimated by the incorporation of two sets of indicator/dummy variables for $J-1$ cohorts and $K-1$ periods. Therefore, Eq (11) changes to

$$\beta_{0jk} = \gamma_0 + \gamma_{1j}\sum_{j=2}^{19} Cohort_j + \gamma_{2k}\sum_{k=2}^{15} Period_k\ , \tag{13}$$

where the variance in the intercept, β_{0jk}, is assumed to be completely captured by the indicator variables for cohorts and periods. Substituting this expression into Eq (10) yields the combined CCFEM:

$$WORDSUM_{ijk} = \gamma_0 + \beta_1 AGE_{ijk} + \beta_2 AGE_{ijk}^2 + \beta_3 EDUCATION_{ijk}$$

$$+\beta_4 FEMALE_{ijk} + \beta_5 BLACK_{ijk} + \gamma_{1j}\sum_{j=2}^{19} Cohort_j + \gamma_{2k}\sum_{k=2}^{15} Period_k + e_{ijk} \qquad (14)$$

We noted earlier that there are two primary considerations on which we focus our methodological assessment of the comparative performance of the CCREM and CCFEM models. One of these is the assumption of the random-effects model that the level-2 effects are independent of the level-1 regressors. Note that most conventional empirical applications of hierarchical linear models proceed without a careful examination of the empirical veracity of this assumption. By contrast, the comparative performance of the fixed and random-effects model specifications is a standard part of model criticism and assessment in longitudinal panel models (often referred to as pooled time-series cross-section models) in econometrics (see, e.g., Greene 2000, pp. 837-841). This is due to the general results in statistical theory for mixed fixed-random effects models to the extent that, under the null hypothesis of zero correlation between the individual-level regressors and the contextual effects coefficients, both the ordinary least squares estimator (OLS) of the individual-level coefficients in the fixed-effects model and the restricted maximum likelihood (REML) estimator of those coefficients in the random-effects model are consistent, but the OLS estimator is inefficient. Therefore, under the null hypothesis, the two estimators should produce estimates of the individual-level coefficients that do not differ systematically.

The present application of the CCREM and CCFEM models to the repeated cross-section data on verbal ability in the GSS differs from standard longitudinal panel designs in that the same individuals are not repeatedly surveyed in consecutive waves of the GSS. However, given the temporal dimensions embedded in the cohort and time period contextual variables as we have defined them, it is important to explicitly address the independence assumption. We therefore examine this assumption of the independence of the random effects and the individual-level regressors in two ways. First, we estimate both the CCREM and the CCFEM models and qualitatively assess the resulting model fits and the parameter estimates and performance of each with respect to the data. Second, we obtain statistical tests by applying a form of what is known in the econometric analysis of pooled time series cross section regression models as a Hausman specification test (see Hausman and Taylor 1981; Baltagi 1995). The Hausman test is a Wald chi-squared test of the form:

$$W = \chi^2[K] = [b - \hat{\beta}]^T \hat{\Sigma}^{-1}[b - \hat{\beta}]$$

where, in the present case, b denotes the vector of individual-level regression coefficients estimated from the CCFEM model, $\hat{\beta}$ denotes the corresponding vector of regression coefficients estimates from the CCREM model, and $\hat{\Sigma} = Var[\hat{b}] - Var[\hat{\beta}]$ is the difference of the variance-covariance matrices of the two estimators (the constant term is excluded from all vectors and matrices). Under the null hypothesis that the cohort and period random effects in the CCREM model are independent of the individual-level regressors, W is distributed as chi-squared with K degrees of freedom, where K is the dimension of the b and β vectors.

A second focus of our assessment of the comparative performance of the CCREM and CCFEM models pertains to their ability to handle the relatively small sample sizes at the level-2 or contextual effects-level, that is, the relatively small numbers of birth cohorts and periods. But even with 19 cohorts and 15 time periods, the GSS data analyzed herein have larger numbers of cohorts and periods than would be the case with many repeated cross-section demographic surveys (e.g., the Current Population Survey Supplements on fertility histories or National Health Interview Surveys). Therefore, in order to make our assessment of the performance of these two models more like that typically encountered in demographic surveys, we "thin" the GSS data further by selecting five recent cohorts from the last five of the 15 GSS surveys and reestimating the CCREM and CCFEM models on this reduced set of data. A critical question pertains to whether the performance of the two models is comparable on this reduced dataset, as the conventional methodological wisdom cited earlier would suggest that fixed-effects model specifications would be better when there is such a small number of contextual units in the analysis.

RESULTS

Table 3 reports the parameter estimates and model fit statistics for the CCFEM (Eq. 14) and CCREM (Eq. 12) models estimated on the 15 GSS repeated cross-section surveys. Results from both models show that all individual covariates are significantly related to WORDSUM. The age effect is curvilinear and concave. Not surprisingly, education has a strong positive effect on one's verbal ability. Females and whites tend to score higher on verbal tests. Taken together, these regressors account for about 30% of the unconditional level-1 variance (not shown).

The parameter estimates for these individual effects are remarkably similar for the CCFEM and CCREM. The main difference is in the estimated intercept and its standard errors. The CCFEM reports a larger mean verbal score and a standard error (0.285) that is more than five times larger than that produced by the CCREM (0.059), indicating much more uncertainty in the mean verbal score estimate. This shows that the indicator variables representing the fixed cohort and period effects do not explain well the variance for the intercept.

[Table 3 about here]

The next section in Table 3 shows the estimated fixed effects and random effects for the 19 cohorts and 15 time periods. In the CCFEM, these correspond to the coefficient estimates for the 18 cohorts, $\hat{\gamma}_{1j}$ and 14 periods, $\hat{\gamma}_{2k}$, and their standard errors, with the last cohort and the period being the reference groups. In the CCREM, the residual random effects are represented by $\hat{u}_{0j}$ for all 19 birth cohorts and $\hat{v}_{0k}$ for 15 survey years. They seem to be different in magnitude and directions from those estimated by the fixed-effects model. This is because in the CCFEM formulation, the fixed cohort and period effects are estimated jointly so that the reference group is 1980 birth cohort in year 2000. If we calculate the predicted value for each cohort averaging across the 15 time periods and vice versa, we obtain essentially the same effects revealed in the CCREM where the random residual effects for cohorts and periods are obtained so that each represent the net effects averaged across the other. The results for the variance components analysis suggest that, controlling for all the individual covariates, the residual variation between cohorts is still significant and is estimated to be 0.039, whereas the residual period variation is close to zero. The inclusion of the age, education, gender and race effects reduces the cohort variance by about 70% and the period variance by more than 90% (not shown). The AIC statistics show that the CCREM has a better model fit to the verbal ability data.

As a formal test of the equality of the coefficients estimated by the fixed- and random-effects models, we next apply the Hausman specification test described above. The results are shown in Table 4. The left panel summarizes the coefficient estimates under the CCFEM and CCREM specifications based on the full sample where there are 19 birth cohorts and 15 periods. The variance-covariance matrices for the two sets of parameter estimates are then obtained. The Hausman test is:

H_0: *Differences in the estimated coefficient vectors are not systematic*

$$\chi^2(5) = (\hat{b} - \hat{\beta})^T \Sigma^{-1} (\hat{b} - \hat{\beta}),\ \Sigma = Var(\hat{b}) - Var(\hat{\beta})$$

$$= 11.371\ (p{=}0.045)$$

Since the sample size on which these coefficient estimates are based is very large (N = 19,500), the standard errors of these coefficients are very small. It, therefore, is appropriate to use a p-value of .001 for assessing this Wald-χ^2 statistic. Accordingly, we fail to reject the null hypothesis of no systematic differences in the coefficient vectors. Therefore, the test shows that the assumption that the random effect effects, u_{0j} and v_{0k}, are uncorrelated with the regressors is acceptable and the fixed-effects model does not outperform the random-effects model for this particular dataset. Instead, the CCREM has the advantages of smaller standard error estimates for the level-1 coefficients and a better model fit. Therefore, the random-effects model should be chosen for the analysis.

[Table 4 about here]

To determine the possible effect of smaller numbers of surveys and birth cohorts that are more typical of demographic data on the model specification for APC analysis, we next replicate the above analysis for the sub-sample where only the last five periods (1993 to 2000) and five recent cohorts (1950 – 1970) are included. The right panel of Table 4 shows the comparison of the effect estimates of CCFEM and CCREM. Since the Hausman test is not significant, the fixed-effects model is not necessarily better than the random-effects model even in such a small sample. This contradicts the conventional methodological wisdom reviewed earlier on choosing between a fixed- versus a random-effects model specification.

Finally, we summarize the modeling results graphically with regards to the age, period, and cohort effects on the verbal ability. Note that in the HAPC model specification each cohort has a distinct intercept, β_{0j}, that represents the *cohort effect* net of all the individual effects, and averaged across all time periods. The estimated cohort effect and the 95% confidence intervals based on the CCREM[7] are plotted in Figure 1. There indeed is evidence in this graph for a decline in verbal ability for recent cohorts born since 1945. However, instead of a linear decline for cohort verbal ability starting in early 20th century estimated by previous researchers, there is evidence of more variation for older cohorts born before 1945.

Figure 1 also shows the *period effect* (β_{0k}) net of all the individual effects and averaged across all birth cohorts. A slight V-shape curve occurs. There was a decrease in verbal ability from mid 1970s to late 1980s. Then it increased to the beginning level, followed by some small fluctuations into the late 1990's.

Finally, the *age effect* is plotted as the predicted WORDSUM by each age, net of all the other factors and averaged across all cohorts and periods. The age effect is curvilinear over the life course and indicates that individual's verbal ability increases from late teens to about 50s as a result of accumulation of vocabulary through education and other social experiences. After the age of 55, however, one's verbal skills gradually decline due to many reasons related with aging, such as loss of memory. This is also consistent with theory of cognitive growth.

[Figure 1 about here]

[7] Since all continuous variables are mean centered, the cohort effect is the cohort intercept at the mean age and mean education, and for the reference sex-race group, white males. The same model was estimated for other three combinations of sex and race (white females, black females, and black males) and the results are similar.

DISCUSSION AND CONCLUSION

In this paper, we have described how to assess the adequacy of the hierarchical APC model's assumption in the context of repeated cross-section survey data that are increasingly available for demographic analysis. Specifically, we have compared the fixed- and random-effects HAPC models in applications to verbal ability data in the U.S. As noted earlier, both the fixed- and the random-effects models are superior to applications of conventional fixed-effects regression models estimated by ordinary least squares that do not take into account the effects of the contextual variables – the cohorts and time periods. In the presence of contextual effects, these conventional regression models tend to underestimate standard errors and overestimate t-tests, thus leading to incorrect inferences.

In addition, however, both the random-effects and the fixed-effects models have their strengths and weaknesses. Therefore, the choice between the two may be contingent upon a number of conditions such as correlations between the random components and the independent variables, sample sizes of cohorts and periods, whether contextual variables need to be incorporated, and properties of the specific demographic phenomena being modeled. Most generally, fixed-effects models require estimating unique effect coefficients for each higher-level unit: $(J - 1) + (K - 1)$ parameters in all. Random-effects models instead estimate one parameter that represents the distribution of the errors. With only small-to-moderate numbers of cohorts and time periods, conventional statistical methodology guidelines suggest that it might be more appropriate to treat the cohorts and time periods as unique entities and model them with a fixed-effects specification.

Contrary to this conventional wisdom, however, the results from the CCREM and CCFEM analyses reported above favor the random-effects model specification regardless of whether the numbers of birth cohorts and time periods are moderate (19 cohorts and 15 time periods) or small (5 cohorts and 5 time periods). A key problem with the fixed-effects specification appears to be the assumption that the indicator/dummy variables representing the fixed cohort and period effects fully explain all of the cohort and fixed effects. That is, the CCFEM model does not allow for the possibility of any additional random variance associated with the individual cohort and period effects. This implies that there is no unexplained between-cohort and/or between-period variability left beyond that captured by the fixed cohort and period effects. In the context of HAPC models, this appears not to be the best assumption. Rather, with relatively large numbers of sample respondents for each cohort and time period, the random-effects specification that allows for random variation in the cohort and/or period contexts appears to perform comparatively better.

While we have used the verbal test score data as a convenient laboratory for this evaluative experiment on the CCREM and CCFEM models, and while more typically demographic response variables from repeated cross-section surveys might display other types of behavior, these findings are encouraging with respect to the applicability of random-effects formulation of the HAPC modeling framework. In addition, we recommend that the evaluative strategy laid out here be used in assessments of the applicability of the fixed-effects and random-effects specifications of the HAPC modeling framework to repeated cross-section survey data in demography. Specifically, the version of the Hausman specification test described herein should be applied to assess the empirical veracity of the assumption that the cohort and time period random effects are distributed independently of the individual-level regressors. Assuming this specification test finds no evidence for systematic differences in the individual-level vectors of estimated coefficients from the fixed- and random-effects models, the analyst should proceed to use the random-effects model for substantive analyses.

References

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

_______ and R.J. McCammon. 1999. “Aging Versus Cohort Interpretations of Intercohort Differences in GSS Vocabulary Scores.” *American Sociological Review* 64:272-86.

Baltagi, B. H. 1995. *Econometric Analysis of Panel Data*. New York: Wiley.

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.” in *Sociological Methodology 1979* edited by K.F. Schuessler. San Francisco: Jossey-Bass, pp.1-67.

_______ and W.M. Mason. 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.

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* 29:263-278.

_______ and P. Hall. 2004. “Age-Period-Cohort Analysis: Structure of Estimators, Estimability, Sensitivity and Asymptotics.” *Journal of the American Statistical Association*, under review.

_______ and T. E. Rohan. 2004. “Age-Period-Cohort Analysis: Consistent Estimation by the Intrinsic Estimator.” *Journal of the American Statistical Association*, under review.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin. 2000. *Bayesian Data Analysis*. New York: Chapman & Hall/CRC.

Glenn, N.D. 2003. "Distinguishing Age, Period, and Cohort Effects." Pp. 465-76 in *Handbook of the Life Course*, edited by J. T. M. a. M. J. Shanahan. New York: Kluwer Academic/Plenum Publishers.

_______. 1999. “Further Discussion of the Evidence for An Intercohort Decline in Education-Adjusted Vocabulary.” *American Sociological Review* 64:267-71.

_______. 1994. “Television Watching, Newspaper Reading, and Cohort Differences in Verbal Ability.” *Sociology of Education* 67:216-30.

_______. 1976. “Cohort Analysts’ Futile Quest: Statistical Attempts to Separate Age, Period, and Cohort Effects.” *American Sociological Review* 41:900-904.

Goldstein, H. 2003. *Multilevel Statistical Models*. 3rd Edition. London: Oxford University Press.

Greene, W.H. 2000. *Econometric Analysis*. 4th edition. Upper Saddle River, NJ: Prentice Hall.

Hausman, J. A. and W. E. Taylor. 1981. “Panel Data and Unobservable Individual Effects.” *Econometrica* 49:1377-1398.

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* 22:283-299.

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, 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* 12:2779-2807.

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* 38:811-830.

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.

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.

O'Brien, R.M. 2000. "Age Period Cohort Characteristic Models." *Social Science Research* 29:123-139.

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.

_______ and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Edited by J. D. a. R. A. Berk. Thousand Oaks: Sage.

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 Methdologies." *Journal of Clinical Epidemiology* 52:569-583.

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

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

Wilmoth, J. R. 1990. "Variation in Vital Rates by Age, Period, and Cohort." *Sociological Methodology* 20: 295-335.

Wilson, J.A. and W.R. Gove. 1999. "The Intercohort Decline in Verbal Ability: Does it Exist?" and reply to Glenn and Alwin & McCammon. *American Sociological Review* 64:253-266, 287-302.

Yang, Y. 2004. A Bayesian Hierarchical Modeling Approach to Age-Period-Cohort Analyses of Repeated Cross-Section Survey Data. Master of Science Thesis. Institute of Statistics and Decision Sciences, Duke University. Paper winning the

2004 Student Paper Competition Award Sponsored by a Consortium of Sections of the American Statistical Association.

_______ and Kenneth C. Land. 2003. Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Towards an Integrated Methodology. Paper presented at the annual meeting of the American Sociological Association, Atlanta, Georgia.

_______, W.J. Fu, and K.C. Land. 2003. "A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models." *Sociological Methodology 2004*, forthcoming.

Table 1. Summary Statistics for Verbal Ability Data from GSS: 1974 – 2000

Outcome	***Description***	***N***	***Mean***	***SD***	***Min***	***Max***
WORDSUM	A composite vocabulary test score	19500	6.02	2.15	0	10
Independent Variables						
AGE	Respondent's age at survey year	19500	45.34	17.10	18	89
EDUCATION	Respondent's years of schooling	19500	12.72	3.02	0	20
FEMALE	Sex: = 1 if female; =0 if male	19500	0.57	0.50	0	1
BLACK	Race: = 1 if black; = 0 if white	19500	0.15	0.35	0	1
Group Variables						
Cohort	Five-year birth cohorts	19			1890	1980
Period	Survey year	15			1974	2000

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

Table 3 HAPC Models of the GSS Verbal Score Data: CCFEM vs. CCREM

	CCFEM			*CCREM*		
Individual Effects	***Coefficient***	***se***	***t Ratio***	***Coefficient***	***se***	***t Ratio***
INTERCEPT	6.456***	0.285	22.62	6.167***	0.059	103.73
AGE	0.266#	0.090	2.96	0.030#	0.017	1.75
AGE^2	-0.059***	0.006	-9.94	-0.065***	0.005	-11.83
EDUCATION	0.375***	0.005	82.95	0.374***	0.004	82.95
FEMALE	0.241***	0.026	9.38	0.242***	0.025	9.40
BLACK	-1.046***	0.037	-28.45	-1.051***	0.036	-28.74
	Fixed Effects			***Random Effects***		
Cohort	***Coefficient***	***se***	***t Ratio***	***Coefficient***	***se***	***t Ratio***
1890	-2.086	0.869	-2.40	-0.043	0.165	-0.26
1895	-2.035	0.797	-2.55	-0.123	0.140	-0.88
1900	-1.579	0.742	-2.13	0.069	0.113	0.61
1905	-2.022	0.695	-2.91	-0.403	0.099	-4.06
1910	-1.321	0.651	-2.03	0.079	0.088	0.89
1915	-1.070	0.607	-1.76	0.192	0.078	2.44
1920	-1.188	0.565	-2.10	-0.037	0.074	-0.50
1925	-1.016	0.524	-1.94	0.008	0.071	0.12
1930	-0.864	0.483	-1.79	0.030	0.071	0.46
1935	-0.769	0.441	-1.74	0.004	0.070	0.05
1940	-0.518	0.400	-1.29	0.126	0.068	1.85
1945	-0.162	0.360	-0.45	0.354	0.065	5.41
1950	-0.082	0.323	-0.25	0.326	0.065	4.99
1955	-0.279	0.287	-0.97	0.026	0.066	0.38
1960	-0.219	0.255	-0.86	-0.031	0.070	-0.44
1965	-0.160	0.228	-0.70	-0.079	0.076	-1.03
1970	-0.184	0.208	-0.88	-0.195	0.085	-2.29
1975	-0.076	0.203	-0.37	-0.178	0.102	-1.73
1980	0.000	.	.	-0.127	0.140	-0.91
Period						
1974	0.678	0.242	2.80	0.035	0.040	0.86
1976	0.676	0.225	3.00	0.063	0.040	1.58
1978	0.533	0.207	2.56	0.008	0.039	0.19
1982	0.418	0.173	2.41	-0.002	0.037	-0.06
1984	0.414	0.159	2.60	0.024	0.039	0.60
1987	0.234	0.134	1.74	-0.043	0.037	-1.15
1988	0.068	0.132	0.51	-0.103	0.042	-2.40
1989	0.159	0.124	1.28	-0.048	0.042	-1.13
1990	0.274	0.119	2.29	0.020	0.043	0.47
1991	0.288	0.111	2.59	0.041	0.042	0.95
1993	0.163	0.098	1.66	0.002	0.042	0.01
1994	0.174	0.084	2.05	0.022	0.037	0.60
1996	0.023	0.074	0.30	-0.048	0.037	-1.28
1998	0.117	0.073	1.59	0.037	0.040	0.92
2000	0.000	.	.	-0.005	0.041	-0.14
Variance Components	***Variance***	***se***	***p value***	***Variance***	***se***	***p value***
Cohort				0.039**	0.016	0.00
Period				0.003#	0.002	0.08
Individual	3.135***	0.032	0.00	3.136***	0.032	0.00
Model Fit						
Deviance (DF)	77732.9 (7)			77714.4 (9)		
AIC	77746.9			77732.4		

# indicates *p*<0.10* indicates *p*<.05; ** indicates *p*<.01; *** indicates *p*<.001

Table 4 Hausman Specification Tests Based on the Total Sample (N = 19,500; J = 19; K = 15) and the Subsample (N = 3,687; J = 5; K = 5)

	----------*Coefficients (N = 19,500)* ----------			----------*Coefficients (N = 3,687)* ----------		
WORDSUM	***Fixed Effects***	***Random Effects***	***Difference***	***Fixed Effects***	***Random Effects***	***Difference***
AGE	0.266	0.030	0.236	0.303	0.324	-0.021
AGE^2	-0.059	-0.065	0.006	-0.044	-0.020	-0.023
EDUCATION	0.375	0.374	0.001	0.352	0.352	0.001
FEMALE	0.241	0.242	-0.001	0.237	0.238	-0.001
BLACK	-1.046	-1.051	0.005	-1.049	-1.057	0.009
Hausman Test	χ^2	***df***	***p value***	χ^2	***df***	***p value***
	11.371	5	0.045	5.716	5	0.335

Figure 1 Estimated Cohort Effects, Period Effects, with 95% CIs, and Age Effects on GSS Verbal Test Scores