Published in final edited form as:

Soc Forces. 2013 ; 92(1): 221–248. doi:10.1093/sf/sot066.

Assessing the Significance of Cohort and Period Effects in Hierarchical Age-Period-Cohort Models: Applications to Verbal Test Scores and Voter Turnout in U.S. Presidential Elections

Steven M. Frenk,
University of North Carolina–Chapel Hill

Yang Claire Yang, and
University of North Carolina–Chapel Hill

Kenneth C. Land
Duke University

Abstract

In recently developed hierarchical age-period-cohort (HAPC) models, inferential questions arise: How can one assess or judge the significance of estimates of individual cohort and period effects in such models? And how does one assess the overall statistical significance of the cohort and/or the period effects? Beyond statistical significance is the question of substantive significance. This paper addresses these questions. In the context of empirical applications of linear and generalized linear mixed-model specifications of HAPC models using data on verbal test scores and voter turnout in U.S. presidential elections, respectively, we describe a two-step approach and a set of guidelines for assessing statistical significance. The guidelines include assessments of patterns of effects and statistical tests both for the effects of individual cohorts and time periods as well as for entire sets of cohorts and periods. The empirical applications show strong evidence that trends in verbal test scores are primarily cohort driven, while voter turnout is primarily a period phenomenon.

Social scientists often study time-specific phenomena for which there may be age, period, and/or cohort effects. Because age-period-cohort (APC) analysis has the capacity to depict the entire complex of social, historical, and environmental factors that shape individual life courses parsimoniously, it is important for constructing and refining theories of social change. One common goal of such analysis is to distinguish the unique effects associated with age, period, and cohort (Hobcraft, Menken, and Preston 1982). Attempts to estimate these effects, however, must address the model identification problem that occurs due to the exact linear dependency between the three variables: cohort = period – age. Thus, one cannot separately estimate them using conventional linear regression models without adding

Please direct correspondence to Steven M. Frenk, CB#8120, University Square, 123 West Franklin Street, Chapel Hill, NC 27516-2524; phone: (919) 843-5548; frenk@live.unc.edu.

An earlier version of this manuscript was presented at the Measurement and Models Session at the Annual Meeting of the American Sociological Association, San Francisco, CA (August 8–11, 2009) and at Demography Daze, a symposium of demographic research organized by the Carolina Population Center and the Duke University Population Research Institute, Durham, NC (July 14, 2012).

constraints to identify the model. A vast literature in demography, biostatistics, epidemiology, and sociology has used this general model with various sorts of constraints (Mason and Fienberg 1985; Mason and Wolfinger 2001; Mason et al. 1973; Kupper et al. 1985), yet differences among the solutions often produce ambiguous and inconsistent results.

Beginning in 2000, new interest in APC models and methods has emerged in the social sciences to address this question (see special issue of Sociological Methods and Research [2008]; Yang et al. 2008). Statistics has continued to develop as a discipline since the Mason-Feinberg synthesis of 1985, and new statistical models and new computationally intensive estimation methods have been developed (i.e., mixed [fixed and random] effects models, Markov chain Monte Carlo methods). In addition, data sets with research designs that invite or even require the analysis of separate age, period, and cohort components of change are increasingly available.

One type is repeated cross-sectional surveys. It is composed of multiple waves of cross-sectional surveys, which over time lead to the accumulation of data spanning several decades. Although the same respondents are not surveyed in each wave, by combining each survey wave together, one can produce synthetic cohorts, which trace the same groups of people from the same birth cohorts over a large segment of the life span (Mason and Fienberg 1985; Preston, Heuveline, and Guillot 2001). Recent methodological research has highlighted the opportunities they provide for examining cohort and period effects (Yang 2006; Yang and Land 2006, 2008). One of their crucial features is that they provide a multilevel data structure (individuals are nested in a cross-classification of survey periods and birth cohorts) that allows researchers to specify a different family of models rather than conventional linear models composed of fixed and additive age, period, and cohort effects. These latter models are not only certain to produce the identification problem, but are a poor approximation to the process of social change.

Recently, a mixed effects modeling framework was developed for APC analysis that utilizes data from repeat cross-sectional surveys—the class of *Hierarchical APC (HAPC) models* (Yang 2006; Yang and Land 2006, 2008). The HAPC approach conceptualizes time periods and cohort memberships as social historical contexts within which individuals are embedded and ordered by age and models them as random as opposed to fixed effects additive to that of age. This contextual approach broadens the theoretical foundation of APC analysis, helps deal with (by completely avoiding) the identification problem, and also accounts for potentially correlated errors. In addition, mixed-model specifications are more statistically efficient, which is useful because repeated cross-sectional surveys produce highly unbalanced data (Yang and Land 2008, 321).[1] HAPC mixed models also have parameterizations that produce estimates of period and cohort effects that are interpretable as effects that apply across all cohorts in the former case and across all periods in the latter; by comparison, the estimated partial regression coefficients of a HAPC fixed model must be

[1]Sample respondents in a repeated survey design are cross-classified by cohort (arrayed in rows) and time period of observation (arrayed in columns). The number of observations in cells above the diagonal is not symmetric with those below the diagonal.

transformed algebraically to produce estimates of such effects (Yang and Land 2008, 317–18).[2]

The simplest form of model specification of a HAPC mixed model is that of a *linear mixed effects model (LMM)* (Yang and Land 2006). It consists of a two-level model: the level-1 component is a regression of an individual-level outcome variable on a set of individual-level explanatory variables (or covariates) with regression slope coefficients (fixed effects), a constant or intercept term, and an individual-level random error term. Level-2 models use level-1 regression coefficients as outcomes and contain intercepts and specifications of random effects of each cohort and time period distinguished in the model. The level-2 model may also contain cohort or time-period explanatory variables with fixed effect coefficients that are hypothesized to explain, at least in part, the cohort or period effects (Yang 2006). As formulated by Yang and Land (2006), a key feature of the HAPC approach to modeling age, period, and cohort effects is the recognition that *an individual's age is intrinsically an individual-level characteristic* that each individual in a sample survey carries with him or her. On the other hand, *an individual's presence in a birth cohort or time period* and the potential effects that such a presence can have on their outcomes *may be best conceived*, as emphasized by Ryder (1965), *as group memberships and thus modeled as contextual effects*. A second key feature is the recognition that time periods in a repeated cross-sectional survey design are not fully nested within cohorts and cohorts are not fully nested within time periods. Rather, one obtains a cross-classified structure with individuals nested within cells defined by birth cohorts and time periods.

The development of HAPC models provides a useful apparatus for modeling and estimating distinct age, period, and cohort effects in repeated cross-sectional survey. Several studies have already utilized these methods on topics ranging from the obesity epidemic (Reither, Hauser, and Yang 2009) to attendance at religious services (Schwadel 2010). In this context, however, the question arises: How can one assess or judge the significance of estimates of cohort and period effects in such models? This question may be addressed by examining the statistical significance of the estimated random effect for each individual cohort and time period in a study. But it may be the case that some cohorts have statistically significant effect coefficients and some do not, and the same may be true for the estimated period effects. In such a case, how does one assess the overall statistical significance of the cohort and/or the period effects? Beyond statistical significance is the question of substantive significance. It could be the case, for example, that most of the individual estimates of cohort or period effects are not statistically significant at a conventional level of significance but exhibit a substantively interesting trend or pattern that merits substantive interpretation.

This paper addresses these questions. In the context of empirical applications of linear and generalized linear mixed-model specifications of HAPC models, we describe a two-step

[2]For brevity and simplicity we will use the terms “estimated period and cohort effects” or “estimated effect coefficients.” Other terms used in the general mixed-models statistical literature are “random effect estimates,” “residual estimates,” or “empirical Bayes estimates.” They are posterior estimates in the sense that they are not parameters estimated along with the covariate coefficients and random effect variance. Statistically, they are “shrunk” by a factor that depends on the number of observations in a particular cohort/period and the size of the between-cell variances relative to the within-cell variance (see Demidenko 2004; Raudenbush and Bryck 2002; or Snijders and Bosker 1999).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

approach and set of guidelines for assessing statistical significance. The guidelines include assessments of patterns of effects and statistical tests both for the effects of individual cohorts and time periods as well as for entire sets of cohorts and periods. This procedure and set of guidelines build upon a large body of literature on methods for hypothesis testing in mixed (fixed and random effects) models in statistics. We claim no originality for these general statistical methods. Rather, the contribution of this paper is to organize them into a set of methods specifically adapted to the features of HAPC models and to illustrate their application in the context of empirical analyses of two data sets widely used by social scientists—the General Social Survey (trends in verbal ability) and the American National Election Studies (trends in voter turnout in U.S. presidential elections). In addition to our presentation of a set of methods for hypothesis testing in HAPC models, substantive findings from the empirical applications clearly demonstrate the dominance of cohort effects in the former case and period effects in the latter and thus help resolve long-standing empirical questions and disputes in each case. The procedure and guidelines that are articulated and illustrated in these two empirical analyses can be readily adapted and applied more generally to other empirical APC analyses focused on outcomes of interest to social scientists (behaviors, beliefs, health outcomes, etc.).

Two Applications

Verbal Test Scores

The first application involves trends in verbal ability among American adults. A series of articles published in the *American Sociological Review* in 1999 center upon the existence of an inter-cohort decline in verbal ability in the General Social Survey. The debate was initiated by Alwin's (1991) and Glenn's (1994) finding of a long-term inter-cohort 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 confused cohort effects with aging effects and ignored time-period effects. In response, Glenn (1999) and Alwin and McCammon (1999) disagreed that the decline in GSS vocabulary scores resulted solely from period influences and argued against the Wilson and Gove claim that cohort differences actually reflected only age effects—to which Wilson and Gove (1999b) responded by continuing to insist on the strength of the period effects. Yang and Land (2006) used hierarchical age-period-cohort models to provide better estimates of the effect of period and cohort on verbal test scores. Their findings led to the conclusion that there were significant cohort effects and only modest time-period effects. The cohort effects were bimodal, with an increase in verbal knowledge from the early 1900s to the 1940s and then declining until increasing again in the 1980s.

To further test whether the birth-cohort and time-period effects make statistically significant contributions to explained variance in an outcome variable—taken in their entirety as sets of effects, as opposed to individual period and cohort effects—a general linear hypothesis may be applied. Specifically, one can use an *F*-test to test the hypothesis of the presence of random effects. The sampling distribution of the *F*-statistic is exact when the random effects are independently distributed as normal random variables. This test statistic is preferred over the *z*-score when the sample sizes for random effects are small (Littell et al. 2006). The statistical theory for such tests has been developed in a very general LMM context (Demidenko 2004).

How can one assess or judge the significance of estimates of cohort and period effects in such models? This question may be addressed by examining the statistical significance of the estimated effect coefficients for each individual cohort and time period in a study. But it may be the case that, say, some cohorts have statistically significant effect coefficients and some do not, and the same may be true for the estimated period coefficients. In such a case, how does one assess the overall statistical significance of the cohort and/or the period effects? Beyond statistical significance is the question of substantive significance. It could be the case, for example, that most of the individual estimates of cohort or period effects are not statistically significant at a conventional level of significance but exhibit an interesting trend or pattern that merits substantive interpretation.

Data

Data for this analysis come from 17 waves of the General Social Survey (1974–2006). The GSS is a nationally representative survey of non-institutionalized adults age 18 and older in the United States (Davis, Smith, and Marsden 2007). The GSS collects data on a wide variety of demographic characteristics, attitudes, and behaviors as well as a measure of respondents' verbal vocabulary knowledge operationalized as a ten-item vocabulary scale. The scale serves as the outcome variable in our analysis. Age is transformed by centering it around the grand mean.[3] Three other independent variables are included in the analysis: sex, race, and years of education (centered around cohort means).[4] Birth cohort is measured in five-year intervals, and period is measured by year of the GSS wave. Table 1 reports the mean, central tendency, and range for all variables used in the analysis and gives a brief description of each measure. All models were estimated using SAS PROC MIXED.[5]

Analysis

Because the WORDSUM outcome variable has a relatively bell-shaped sample frequency distribution, it is reasonable to use an HAPC mixed-model specification that has a conventional normal-errors level-1 regression model. In the absence of evidence to the contrary, this level-1 model can be combined with a conventional normal period and cohort

[3]An important decision in hierarchical modeling pertains to "centering" or choosing the location of the individual level (Raudenbush and Bryk 2002; Yang and Land 2006, 88). The choices include (a) using the natural metric of the variables [NM], (b) grand-mean centering by subtracting the complete sample or grand mean from the observed values [GMC], and (c) centering within subgroups or contexts [CWC]. When the minimum value of an explanatory variable does not include zero, as is the case of age (since the GSS sampling frame is for age 18 and over) in the model of equation (1), conventional methodological guidelines (Raudenbush and Bryk 2002, 32) indicate that one of the latter two options should be used. Furthermore, the literature on the effects of age on vocabulary knowledge cites a pure physiological age effect that does not vary by cohort context (Wilson and Gove 1999a, 257–58). Thus, we apply centering around the grand mean to the individual-level age variable. 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 equations (1)–(3), 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 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.

[4]Wilson and Gove (1999a, 255–56) argue that changing average levels of school years completed varies substantially across the cohorts survey in the GSS. Thus, education is centered around the cohort means.

[5]Cross-classified mixed models can be estimated by using HML 6 (Raudenbush et al. 2004) or other statistical software packages.

residual effects specification at level 2.[6] Specifically, after testing for the presence of cross-level interactions of time periods and birth cohorts with individual-level covariates on the WORDSUM scores following standard multilevel modeling methods (see, e.g., Raudenbush and Bryk 2002; Snijders and Bosker 1999) and finding them insignificant, Yang and Land (2006, 87) specified the following *Hierarchical Age-Period-Cohort-Cross-Classified Random Effects Model (HAPC-CCREM)* with random effects of periods and cohorts on the intercept of the individual-level regression model:

Level-1 or "Within-Cell" Model:

$$\begin{aligned} \mathrm{WORDSUM}_{ijk} = & \beta_{0jk} + \beta_1 \mathrm{AGE}_{ijk} + \beta_2 \mathrm{AGE}^2_{ijk} + \beta_3 \mathrm{EDUCATION}_{ijk} \\ & + \beta_4 \mathrm{FEMALE}_{ijk} \\ & + \beta_5 \mathrm{BLACK}_{ijk} + e_{ijk}, e_{ijk} \sim N(0, \sigma^2) \end{aligned} \quad (1)$$

Level-2 or "Between-Cell" Model:[7]

$$\beta_{0jk} = \gamma_0 + u_{0j} + v_{0k}, u_{0j} \sim N(0, \tau_u), v_{0k} \sim N(0, \tau_v) \quad (2)$$

Combined Model:

$$\mathrm{WORDSUM}_{ijk} = \gamma_0 + \beta_1 \mathrm{AGE}_{ijk} + \beta_2 \mathrm{AGE}^2_{ijk} + \beta_3 \mathrm{EDUCATION}_{ijk} + \beta_4 \mathrm{FEMALE}_{\mathrm{ijk}} + \beta_5 \mathrm{BLACK}_{ijk} + u_{0j}, v_{0k} + e_{ijk} \quad (3)$$

for $i = 1, 2, \ldots, n_{jk}$ individuals within cohort j and period k;

$j = 1, \ldots, 20$ birth cohorts;

$k = 1, \ldots, 17$ 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 her/his age, age squared, educational attainment, and two covariates, gender and race, that have been found in previous research to be related to verbal ability. Since, in preliminary analyses, Yang and Land (2006) found that none of the level-1 slope

[6]With the random errors at level 1 and the random effects at level 2 both assumed to be normally distributed, this model is a member of the class of *Gaussian linear mixed models*. This is the most widely used specification of LMM and is the standard specification applied in hierarchical linear model applications in the social sciences. However, statistical methods also have been developed for alternative models that assume that the random effects and errors are independent, or simply uncorrelated, but not normal (see Jiang [2007] for a review of these alternative models and methods). After estimation of a Gaussian LMM version of a HAPC model, at the "model diagnostics" phase of analysis, an analyst can apply various diagnostic plots and goodness-of-fit tests (Jiang 2007, 88–92) to assess the normality assumptions. If these diagnostics indicate substantial departures from normality of the errors or random effects and/or nonzero correlations thereof, then a non-Gaussian LMM should be specified and estimated in order to assess the robustness of the estimates from the Gaussian specification and possibly to replace them. For the empirical application to the WORDSUM data described here, the model diagnostics indicate that the Gaussian assumptions are acceptable.

[7]In mixed (fixed and random) effects linear or generalized linear regression models, the usual specification on the random effects—the time periods and birth cohorts in the analyses in the paper—is that they are normally distributed with an expectation of zero. We use the term "usual" here because there are versions of mixed effects models that specify frequency distributions other than the normal for the random effects (e.g., Lee and Nelder 2006). However, it is the normal distribution assumption that is most common and that is programmed into the standard software for estimation of mixed effects models. And for most empirical applications it is adequate. Accordingly, it is the specification used in the analyses reported in the paper. Other than this distributional assumption, the estimated time-period and cohort effects are not constrained. Thus, for example, if the data indicated such, the estimates of the period effects could, perhaps with some stochasticity, line up pretty much in a downward-sloping linear trend. These estimated effects then would add a positive adjustment to the estimated intercept of the level-1 regression model in the early periods of a study, decreasing with each period until they hit zero or very close to zero, and then further decreasing and becoming negative such that they would adjust the estimated level-1 intercept down. Estimates of the cohort effects similarly could show a linear trend.

coefficients exhibit significant random variation across cohorts and periods in the GSS verbal test score data, 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 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 are surveyed in year *k* at zero values of age and the covariates; $\beta_1 \ldots \beta_5$ are the level-1 fixed effects; e_{ijk} is the random individual effect, that is, the deviation of individual *ijk*'s score from the mean for cohort period *jk*; the e_{ijk} 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 sampled individuals at zero values of age and the covariates; u_{0j} is the 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 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, controlling in both cases for the effects of age and the covariates. In addition, at zero values of age and the covariates, $\beta_{0j} = \gamma_0 + u_{0j}$ is the cohort verbal test score mean, and $\beta_{0k} = \gamma_0 + v_{0k}$ is the period verbal test score mean.

Assessing the significance of estimated cohort and period effects

Table 2 reports the parameter estimates and model fit statistics for the HAPC-CCREM of equations 1–3 estimated on the 17 GSS repeated cross-section surveys.[8] These results were obtained using the restricted maximum-likelihood-empirical Bayes (REML-EB) estimation method (Raudenbush and Bryk 2002: Chapters 3 and 12).[9] Yang and Land (2006) gave detailed substantive interpretations of the coefficients and statistics reported in table 2. For present purposes, the key focus is on how one assesses and tests the statistical significance of the cohort and time-period effects reported in the table. We suggest a two-step approach.

Step 1: Study the Patterns and Statistical Significance of the Individual Estimated Coefficients for Time Periods and Birth Cohorts

As an initial step, the individual estimated period and cohort effects should be studied for both substantively meaningful patterns and statistical significance. This can be done in two parts.

Step 1.1: Graphically Plot the Estimated Cohort and Period Effect Coefficients

While the numerical values of the estimated cohort and period effects in table 2 contain the same information, as a first step in the analysis of their substantive and statistical significance, we recommend that analysts graphically plot the estimates. This facilitates a quick visual check of the extent to which the estimated effects exhibit patterns that are of

---

[8]Model diagnostics (see endnote 3) indicated that the assumptions of a Gaussian linear mixed-model specification (normally distributed errors at level 1 and independently and normally distributed random effects at level 2) were acceptable for the WORDSUM data.

[9]The REML-EB estimation algorithm uses a restricted maximum likelihood estimator to estimate the fixed effects of a mixed effects, or multilevel, model and an empirical Bayes estimator to estimate the random effects.

substantive significance. Particular periods or birth cohorts that stand out also may be identified.

Figure 1 contains graphs of the estimated cohort effects (i.e., the estimated $\beta_{0j} = \gamma_0 + u_{0j}$ cohort verbal test score effects averaged over all time periods for each cohort *j*) and the time-period effects (i.e., the estimated $\beta_{0k} = \gamma_0 + v_{0k}$ period verbal test score effects averaged over all cohorts for each time period *k*) with their 95-percent confidence bounds. Each graph also has a horizontal line at 6.175, the numerical value of the estimated γ_0 or intercept coefficient reported in table 2. This line facilitates a visual inspection of those cohorts and time periods, if any, that have substantial deviations from the overall intercept.

With 20 and 17 data points, respectively, the 95-percent confidence bounds for the birth-cohort and time-period effects are relatively broad. The pattern of the estimated time-period effects does not show much variability, with only the 1988 data deviating substantially from the overall average. By comparison, it can be seen that the estimated pattern of birth-cohort effects contains some fluctuations that are quite pronounced and relevant to substantive debates concerning historical trends in verbal ability (see Glenn 1999; Alwin and McCammon 1999; Wilson and Gove 1999a, 1999b). Specifically, there are peaks in the cohort effects for the 1940–1944, 1945–1949, and 1950–1954 War Babies and Early Baby Boomer birth cohorts followed by declining effects in subsequent cohorts. In addition, however, there is evidence of an early twentieth-century peak in cohort effects, specifically of the 1910–1914 and 1915–1919 birth cohorts. There also is a dip in the estimated effect for the 1905–1909 cohort.

Step 1.2: Examine the Statistical Significance of Individual Cohort and Period Effect Coefficients

Turning from visual and substantive assessments of estimated cohort and period effects, the next step is to examine the statistical significance of the individual effect coefficients for the birth cohorts and time periods—the estimates of the u_{0j} and the v_{0k} random effects with the null hypothesis in each case being that the respective coefficient is zero, that is:

$$H_0: u_{0j} = 0 \, versus \, H_a: u_{0j} \neq 0, \text{and} \, H_0: v_{0k} = 0 \, versus \, H_a: v_{0k} \neq 0.$$

If these null hypotheses are not rejected, this implies that the mean of the WORDSUM outcome variable for the *j*th time period or the *k*th cohort is no different than the overall average. The coefficients and their standard errors for the birth cohorts and time periods in table 2 were estimated by the REML-EB method, and their ratios can be interpreted as asymptotic/large sample *t*-ratios in the conventional way (Raudenbush and Bryk 2002, 57–58).[10]

[10]"Asymptotic" is a term used to describe statistical estimators and hypothesis tests that are derived under assumptions of large samples as opposed to estimators and tests that are "exact" or apply for any sample size, including small samples of, say, 30 or fewer observations. Because some of the statistical estimators and tests described in this paper are exact and some hold only in large samples, we use the asymptotic term to identify the latter.

As will often be the case with HAPC models, some cohorts and time periods may have statistically significant effects, as measured by *t*-ratios, while others do not. For instance, in table 2, the effect coefficients for the 1905–1909, 1915–1919, 1940–1945, 1945–1949, 1950–1954, 1965–1969, and 1970–1974 cohorts are statistically significant at conventional levels, whereas those for the other cohorts are not. For time periods, only the effect coefficient for 1988 attains conventional levels of statistical significance.

These assessments of statistical significance of the individual cohort and period estimated coefficients are consistent with the graphical representations in figure 1. Given that the estimated cohort and period coefficients are displayed in the figure with their 95-percent confidence bounds, there is a correspondence between those cohort and period coefficients with asymptotic *t*-ratios that are statistically significant at the 0.05 level in table 2 and those for which the 95-percent confidence bounds do not cross the 6.175 horizontal line in figure 1, that is, the 1905–1909, 1915–1919, 1940–1945, 1945–1949, 1950–1954, 1965–1969, and 1970–1974 cohorts and the 1988 period. In general, however, even when none of the individual birth-cohort and time-period coefficients are statistically significant, it often is useful to examine graphically the patterns of each set of coefficients for trends that could be of substantive interest.

Step 2: Test for the Statistical Significance of the Period and Cohort Effects Taken as a Group

When some cohort and period effect coefficients attain statistical significance, but some or most do not, the analyst next may address the question of whether the birth-cohort and/or time-period effects, taken as a set, are statistically significant. This question is one of whether these effects, taken altogether, contribute to explained variance in the model.

Step 2.1: Deviance and Variance Components Analysis

First, study the model deviance statistic and variance components.[11] In table 2, the bottom rows show that the deviance statistic is very large compared to the degrees of freedom of the model, thus indicating a highly significant association of the explanatory variables with the WORDSUM response variable. In addition, the variance components show that most of the variance in WORDSUM is the individual-level regressors at level 1. Level-2 variance components results indicate that variation by cohorts is statistically significant, whereas that for time periods is not. This variance component analysis based on *z*-scores is consistent with the results noted above for the individual cohort and period coefficients. That is, a sufficient number of estimated cohort effects are statistically significant for the cohort variance component to attain statistical significance. But, since only one of the time-period effects is statistically significant, the overall contribution of the random effects for time periods is not sufficiently large for its variance component to attain statistical significance.[12]

[11]The deviance is defined as minus twice the natural logarithm of the likelihood of an estimated model and can be regarded as a measure of lack of fit between model and data (see Raudenbush and Bryk 2002, 63–65; Snijders and Bosker 1999, 88–109).
[12]These conclusions are not changed if the level-1 model is specified with only the linear and quadratic age effects.

Step 2.2: An F-Test for the Presence of Random Effects

The above results have been obtained using REML, which rests on the assumption that the error terms are asymptotic normally distributed, and yields random effect estimators with good large sample properties. When the number of level-2 units, in this case cohorts and periods, is not large, this assumption may not be appropriate. And the *z*-scores for the REML estimates of the variance components are only proximate.

To further test whether the birth-cohort and time-period effects make statistically significant contributions to explained variance in an outcome variable, a general linear hypothesis may be applied. Specifically, one can use an *F*-test to test the hypothesis of the presence of random effects.[13] The sampling distribution of an *F*-statistic is exact when the random effects are independently distributed as normal random variables. This test statistic is preferred over the *z*-score when the sample sizes for random effects are small (Littell et al. 2006).[14] The statistical theory for such tests has been developed in a very general LMM context (Demidenko 2004).[15]

In the present case, for the CCREM model of equations (1)–(3), there are only two sets of random effect coefficients that are estimated, namely, the set of residual random effects of cohort *j*, u_{0j}, and the set of residual random effects of period *k*, v_{0k}. Each of these sets of random coefficients is assumed to be independently, normally distributed with mean 0 and variances τ_u and τ_v, respectively. Thus, for a CCREM model with random intercepts of the form of equations (1)–(3), the exact *F*-test amounts to testing null hypotheses for the relevance of either of the birth-cohort random effects:

$$H_0{:}\tau_u{=}0 vs. H_a{:}\tau_u{>}0 \quad (4)$$

or the time-period effects:

$$H_0{:}\tau_v{=}0 vs. H_a{:}\tau_v{>}0. \quad (5)$$

Alternatively, one can test for the joint relevance of both the cohort and period effects:

$$H_0{:}\tau_u{=}\tau_v{=}0 vs. H_a{:}\tau_u{>}0 or \tau_v{>}0 \quad (6)$$

[13]The exposition presented here focuses on the *F*-test as an exact test when the numbers of estimated random effects are small and independent. An alternative is the likelihood ratio test statistic (Wasserman 2004, 299), the chi-squared asymptotic distribution of which has long been studied. Recently, however, Crainiceanu and Ruppert (2004) derived the finite sample distribution of likelihood ratio statistics in linear mixed models with one variance component, and Lu and Zhang (2010) proved an equivalence between the likelihood ratio and *F*-test for testing variance components in a balanced one-way random effects model. Accordingly, it is likely that likelihood ratio test equivalents to the *F*-test described here will be developed for all forms of linear and generalized linear mixed models in the near future.

[14]A reviewer suggested the use of an information criterion such as the Bayes Information Criterion (BIC) to ascertain which models among a set were best fitting. The BIC statistic is an approximation to the Bayes factor for assessing the posterior odds of one model as compared to another with an error of approximation that, for a reasonable choice of prior distributions, is on the order of the square root of the sample size (Raftery 1995). By comparison, the *F*-test is exact for the LMM when the random effects are Gaussian and thus applies to all sample sizes.

[15]Demidenko (2004, 138) noted that this *F*-test may be viewed as a generalization of *F*-tests previously developed for variance components models and for mixed effects models with one random effect.

These null hypotheses correspond to situations, respectively, in which the levels of variation in the cohort effects, the period effects, and the cohort and period effects taken together do not differ significantly from zero.

The results of the *F*-tests for hypotheses (4)–(6) for the GSS data are summarized in table 3. Consider the case of the null hypothesis (4), in which the birth-cohort effects are not relevant to explaining variation in the verbal test score outcome variable of equations (1)–(3). The idea of the *F*-test is that, when the variance of the random birth-cohort effects τ_u = 0, the difference between the minimum sum of squares (SS) of the model (1)–(3) with random cohort and period effects, $S_{\min}$, and the minimum sum of squares without the random effects, as estimated by an ordinary least squares (OLS) regression on the level-1 explanatory variables, S_{OLS}, should be close. Accordingly, we compute the residual SS:

$$S_{OLS}=\sum_{i=1}^{N}\|y_i - X_i\hat{\beta}_{OLS}\|^2 \quad (7)$$

for an ordinary fixed effects regression model that assumes no random effects of cohorts or time periods. Next, the minimum sum of squares in the presence of the random effects, i.e., the minimum:

$$S_{\min_\delta}=\min_\delta\|y - W\delta\|^2, \quad (8)$$

where the matrix *W* consists of the matrix *X* of data points on the individual-level explanatory variables adjoined with a design matrix *Z* for the random cohort effects, i.e., *W* = [*X*, *Z*], and $\delta = (\beta', u'_0)'$. In this example, S_{OLS} = 69,377 and $S_{\min}$ = 68,696. Under the null hypothesis (4), it can be shown that the ratio of two quadratic forms has an *F*-distribution, or more precisely:

$$\frac{(S_{OLS} - S_{\min})/(r - m)}{S_{\min}/(N_T - r)} \sim F(r - m, N_T - r), \quad (9)$$

where N_T denotes total sample size, *r* is the rank of the matrix *W*, and *m* is the number of explanatory variables in the OLS regression (Demidenko 2004, 137). When random cohort effects are present in an LMM model, such as equations (1)–(3), i.e., when τ_u is nonzero, $S_{\min}$ should be relatively small so that the ratio (9) becomes large. Thus, we reject the null hypothesis (4) if the left-hand side of (9) is large. More precisely, let $1 - \alpha$ be a chosen significance level, e.g., $\alpha = 0.05$, and $f_{0.95}$ be the quantile of the *F*-distribution with $r - m$ and $N_T - r$ degrees of freedom. Then the H_0 is rejected when the ratio in (9) exceeds $f_{0.95}$.

To apply the *F*-test (9) to models (1)–(3), note first that, under the assumption that the explanatory variables in the *X* matrix are linearly independent, the rank of *X* is *m*, the number of explanatory variables, five in this case. In addition, since individuals in the pooled GSS data may be members of different birth cohorts, the columns of the design matrix *Z* for the random effects will be linearly independent and thus have rank 20. Therefore, in the numerator of (9), $r = 25$ and $m = 5$, which gives $r - m = 20$ degrees of freedom. In the denominator, $N_T - r$ = 22,042 – 25 = 22,017. Under the null hypothesis (4) that the cohort effects have zero variance in the GSS verbal test score analysis, the *F*-ratio is 10.9. With 20 and 22,017 degrees of freedom, this far exceeds the critical value $f_{0.95}$ = 1.571. The corresponding *F*-ratio for hypothesis (5) of period effects is 2.03, which also exceeds the critical value of 1.623. Thus, we reject the null hypotheses that the variance of either the birth-cohort random effects or the period random effect is zero and conclude that inclusion of these sets of random effects is relevant to the explanation of variation of the GSS verbal test score data. In addition, the *F*-test can be applied to the sets of random effect coefficients taken as a whole, i.e., to test the null hypothesis (6). In this case, the *Z* matrix is expanded to include both u'_0 and v'_0, which changes the rank of *W* to 42 (= 5 + 20 + 17). The *F*-ratio is 7.096, which is significant at the 0.05 level.

The foregoing analyses indicate that there is evidence that the two sets of random effects taken together contribute significantly to the explained variance. Note that the *z*-score reported in table 2 for the variance component for period effects is 1.49 (p = 0.135), indicating the failure to reject the null hypothesis of zero variance for the period coefficients. The *F*-tests described here indicate the opposite, that is, that the period effects contribute significant variability that should not be ignored in the model. Because there are only a few levels of the period random effects, the *F* sum of squares method is a more statistically sensitive method for testing hypotheses about the variance components than the *z*-score method. In particular, the *F*-tests typically will indicate statistical significance of either the cohort or period effects, taken as a whole, when at least one of the members of these sets of effects is statistically significant, as is the case for the estimated period effects reported in table 2.

Voter Turnout in Presidential Elections

The second application involves trends in voter turnout in U.S. presidential elections. Previous studies have examined how age, period, and cohort affects voter turnout and have used these findings to explain the overall decline in voter turnout since the 1950s. Although many studies have established the importance of age as a major factor in voter turnout (see Wolfinger and Rosenstone 1980, 37), the effects of cohort and period are less settled. Using the Current Population Surveys (November edition) to gather data on voter turnout from the 1968–1980 elections, Land, Hough, and McMillen (1986) found that variation in turnout was driven by individual-level factors rather than time-related variables such as period and cohort. Thus, they attribute the decline in voter turnout to the population aging past the age at which turnout is most likely (i.e., mid-60s). Other studies, however, have found significant cohort effects. Using data from pooled cross-sectional surveys, Lyons and Alexander (2000) claim that generational effects can explain much of the decline in voter turnout. Using qualitative cutoff points to generate two cohorts (i.e., respondents born before and after 1932), they find that those born before 1932 are more likely to vote than people born after 1932. In addition, they find that survey year (period) did not affect turnout. Based on these findings, the authors posit that efforts to increase voter turnout should focus on combating the “radical individualism” of recent cohorts and less on structural voting reforms such as motor-voter laws. Despite the focus on only two cohorts, the authors are confident of the impact of generations on voter turnout and state—“While it is statistically impossible to isolate a pure generational effect, it is certainly significant and substantial” (Lyons and

Alexander 2000, 1031)—and thus argue that generational effects should be included in models examining voter turnout.

These analyses lacked the ability to measure age, period, and cohort effects without addressing the identification problem described in the introduction. We estimate an HAPC model that includes covariates found to be significant in prior voter-turnout studies in order to assess the effect of cohort and period on voter turnout. To further test whether the birth-cohort and time-period effects make statistically significant contributions to explained variance, an *F*-test will test the hypothesis of the presence of random effects.

Data

Data for this analysis come from the American National Election Studies (ANES) conducted by the Center for Political Studies at the University of Michigan (ANES 2012). The ANES surveys a nationally representative sample of the U.S. adult population and collects information on respondents' demographic characteristics, socioeconomic status, and voting behavior. These surveys have been conducted during presidential and midterm elections since 1948, but for our analysis only data from surveys conducted during the 1976 to 2008 presidential elections are used.[16]

Our outcome variable is a dichotomous measure of whether respondents reported that they voted in the U.S. presidential election (1 = yes). Age is centered around the grand mean, and five-year cohorts are used. Demographic characteristics and factors found in prior research to be associated with voter turnout in presidential elections are included in the model (Squire, Wolfinger, and Glass 1987; Wolfinger and Rosenstone 1980; Wolfinger and Wolfinger 2008). These variables include sex, race, education, marital status, and residential mobility (i.e., the length of time respondents have lived in their current dwelling). Table 1 reports the mean, central tendency, and range for all variables used in the analysis and gives a brief description of each measure. All models were estimated using SAS PROC GLIMMIX.

Analysis

The HAPC approach to modeling age, birth-cohort, and time-period effects developed by Yang and Land (2006) is not restricted to applications to outcome variables that have relatively bell-shaped frequency distributions and can be modeled by a Gaussian LMM model specification. Rather, the HAPC approach can be applied to dichotomous and multiple categorical outcome variables. For such outcome variables, the HAPC framework takes the form of *generalized linear mixed model (GLMM)* specifications. Our suggested approach to testing for the statistical significance of the random effects again has two steps.

To model the likelihood of voter turnout, we specify the following HAPC-GLMM model:[17]

Level 1 or "Within-Cell" Model:

[16]Data on residential mobility, one of the covariates included in the analysis, were available only in the pre-election surveys covering the 1976–2008 presidential elections.

$$\text{LogitPr}(\text{VOTE}_{ijk}{=}1){=}\beta_{0jk}{+}\beta_1\text{AGE}_{ijk}{+}\beta_2\text{AGE}^2_{ijk}{+}\beta_3\text{MALE}_{ijk}{+}\beta_4\text{BLACK}_{ijk}{+}\beta_5\text{EDUCATION}_{ijk} {+}\beta_6\text{MARRIED}_{ijk}{+}\beta_7\text{MOBILITY}_{ijk} \quad (10)$$

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:

$$\begin{aligned}\text{LogitPr}(\text{VOTE}_{ijk}{=}1){=}&\beta_{0jk}{+}\beta_1\text{AGE}_{ijk}{+}\beta_2\text{AGE}^2_{ijk}{+}\beta_3\text{MALE}_{ijk}{+}\beta_4\text{BLACK}_{ijk}{+}\beta_5\text{EDUCATION}_{ijk}\\ &{+}\beta_6\text{MARRIED}_{ijk}\\ &{+}\beta_7\text{MOBILITY}_{ijk}\\ &{+}u_{0j}{+}v_{0k}\end{aligned} \quad (12)$$

for i = 1, 2, . . ., n_{jk} individual within cohort j and period k;

j = 1, . . ., 19 birth cohorts;

k = 1, . . ., 9 time periods (presidential elections).

Based on supplementary analyses, this HAPC-GLMM model again specifies that the random effects of birth cohorts and election time periods affect the overall intercept of the level-1 model but not the slope coefficients.

Model (10)–(12) differs from model (1)–(3), however, in that its level-1 model is specified in terms of the logit of the probability of voting in a presidential election (p) modeled as the log-odds of voting, logit(p) = log[p/(1-p)], i.e., as a logistic response function. This moves the HAPC-CCREM model from the LMM family of statistical models into the GLMM family.[18]

Table 5 reports parameter estimates and model fit statistics for model (10)–(12) estimated on the nine ANES repeated cross-sectional surveys (1976–2008). Many of the individual-level explanatory variables have highly statistically significant slope coefficients. Likelihood of voting increases with age, but the effect is curvilinear. Specifically, the effect of age reaches its peak during the respondents’ mid-60s, when people are the most likely to vote, and then declines afterward. As found in previous research, more highly educated respondents as well

[17]As indicated earlier in our LMM analysis of the GSS WORDSUM data, prior to assessing the substantive and statistical significance of estimated coefficients, preliminary model specification analyses should be performed. Additional analysis of the WORDSUM and TURNOUT data that examined whether statistically significant effects of time periods or birth cohorts on the individual-level slope coefficients of the level-1 regression model exist were performed. None were found for the WORDSUM data and for the TURNOUT data, none of the interactions with the level-1 slope coefficients affected the results, and thus only the estimates of effects of periods and cohorts on the level-1 intercepts are examined here.

[18]Estimation of a GLMM model is slightly more complicated than that of an LMM, but REML estimators still can be applied and were estimated via the GLIMMIX procedure available for SAS 9.1 (2005). Note that, as in the case of the Gaussian LMM version of the HAPC-CCREM of model (1)–(3), model (10)–(12) assumes normally distributed random effects. After model estimation, diagnostics also can be performed on GLMM versions of the HAPC-CCREM model to assess the applicability of the normal random effects specification. If this specification appears problematic, alternative model specifications, such as the hierarchical-likelihood-based generalized linear models of Lee and Nelder (1996), which permit alternative distributional assumptions on the random effects, can be used.

as those who are currently married are more likely to vote. Increased time of living in one's current dwelling (i.e., apartment, house) increases the likelihood of voting.

Beyond the effects of the individual-level covariates of this model, what is interesting about the estimates in table 5 is that, as contrasted to those for the GSS verbal test score data described above, of the two sets of effects, period effects are more relevant to the explanation of voter turnout than are birth-cohort effects, as measured by contributions to explained variance. Yet, as the *F*-test indicates, birth cohort cannot be discounted as an explanation of voter turnout.

Assessing significance of cohort and period effects in HAPC-GLMM models

To reach this conclusion, we apply the same sequence of graphical displays and statistical tests of significance as identified above for the LMM form of the HAPC-CCREM model.

Step 1: Study the Patterns and Statistical Significance of the Individual Estimated Coefficients for Time Periods and Birth Cohorts

As an initial step, the individual estimated period and cohort effects should be studied for both substantively meaningful patterns and statistical significance. This can be done in two parts.

Step 1.1: Graphically Plot the Estimated Cohort and Period Effect Coefficients

As was the case for the estimated random effects of time periods and cohorts in LMM HAPC models, a first step is to examine graphical displays of the patterns of the effects. These may reveal substantively interesting patterns and/or effects of particular periods or birth cohorts that stand out.

Figure 2 contains graphs of the estimated cohort effects (i.e., the estimated $\beta_{0j} = \gamma_0 + u_{0j}$ cohort voting effects averaged over all time periods for each cohort *j* and converted to probabilities of voting) and the time-period effects (i.e., the estimated $\beta_{0k} = \gamma_0 + v_{0k}$ election period voting effects averaged over all cohorts for each time period *k* and converted to probabilities of voting) with their 95-percent confidence bounds.[19] The graphs also have a horizontal line at 0.15, the transformed value of the intercept of the regression model given in table 5 and which, when multiplied by 100, is the grand-mean probability of voting of all sampled individuals at zero values of age and the covariates. With nineteen and nine observations, respectively, the 95-percent confidence bounds for the birth cohort and time period are relatively broad. In contrast to what was observed above for the WORDSUM cohort estimates in figure 1, the pattern for estimated birth-cohort effects in figure 2 is relatively constant except for a dip with the 1955–1959 cohort (Baby Boomer) followed by a modest increase with subsequent cohorts (1975–1979, 1980–1984, 1985–1990). The graph

[19]The general formula for conversion of logit regression coefficients to coefficients for the probability that the outcome of the regression $Y = 1$, voting in a presidential election in the present case, is *Probility*$(Y = 1) = (\exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K) / 1 + \exp(\alpha + \beta_1 + \beta_2 + \ldots + \beta_k X_K)$ where K is the number of regressors (see, e.g., Menard 2010, 15). This formula was applied to the estimated intercept term in table 5 to convert it to an estimated overall probability of voting in a presidential election and then adapted to convert the corresponding estimates of 95-percent confidence intervals, as shown in figure 2.

of the estimated election time-period effects shows one pronounced decline in likelihood of voter turnout with the 1988 election, followed by a small increase with the 2008 election.

Step 1.2: Examine the Statistical Significance of Individual Cohort and Period Effect Coefficients

Of the individual estimated random effect coefficients for birth cohorts and election periods given in table 5, it can be seen that the 1955–1959 birth cohort has a *t*-ratio that is statistically significant at the $\alpha = 0.10$ level. The time-period coefficient for the 1988 election is also significant at this level.

Step 2: Test for the Statistical Significance of the Period and Cohort Effects Taken as a Group

As in the analysis of the HAPC-LMM model for the verbal test score data described above, when some cohort and period effect coefficients in a HAPC-GLMM model attain statistical significance, but others do not, the analyst should address the question of whether the birth cohort and/or time period effects, taken as a set, are statistically significant.

Step 2.1: Deviance and Variance Components Analysis

As was the case for the GSS verbal test score data, the deviance statistic reported at the bottom of table 5 shows that the full GLMM model explains much variance in the voter-turnout outcome variable. In contrast to the verbal test score example, however, the variance component analysis indicates that the period random effect coefficients make a statistically significant contribution to explaining voter turnout.

Step 2.2: F-Test for the Presence of Random Effects

The model of equations (10)–(12) is that of a logistic regression model with a random intercept. Demidenko (2004, 374–75, 408–9) shows that the *F*-test (9) can be generalized to develop an asymptotic *F*-test of the null hypothesis that the intercepts are constant or homogeneous in a logistic regression model with random intercepts. We next describe and apply this test.

To generalize the *F*-test (9) and to specify the homogeneity test of Demidenko (2004, 374–75, 408) to GLMM formulations of HAPC-CCREM models, recall that the deviance statistic, twice the negative log-likelihood function, l, asymptotically behaves as the sum of squares (McCullagh and Nelder 1989). Given this, S in the test statistic (9) can be replaced by $-2l$ to obtain

$$\frac{(l_0 - l_{\max})/(r-m)}{l_{\max}/(N_T - r)} \cong F(r-m, N_T - r), \quad (13)$$

where l_0 is the maximum of the log-likelihood of the standard level-1 logistic regression model with no level-2 controls for cohort and period effects, $l_{\max}$ is the maximum of the log-likelihood treating the cohort and period effects as fixed parameters,[20] and, as in (9), N_T denotes total sample size, r is the rank of the matrix W, and m is the number of explanatory variables in the level-1 model.

To apply the asymptotic *F*-test (13) to the ANES voter turnout model (10)–(12), note that $m = 7$, the number of explanatory variables in the level-1 logistic regression model, and $r = 35 = m$ + the number of birth cohorts + the number of election time periods = 7 + 19 + 9. Then the value of the *F*-ratio for the statistical significance of the birth-cohort effects alone is

$$F=2.09, p<0.05,$$

that for the statistical significance of the period effects alone is

$$F=3.83, p<0.05,$$

and that for the statistical significance of cohorts and periods together is

$$F=2.51, p<0.05.$$

Numerical details for these *F*-tests for hypotheses (4)–(6) for the ANES data are given in table 6.

The foregoing analysis indicates that there is evidence that the two sets of random effects taken together contribute significantly to the explained variance. The *F*-scores reported in table 6 indicate the ability to reject the null hypothesis of zero variance for the cohort and period effects. The incorporation of election time-period and birth-cohort random effects into the model produces statistically significant variation in the level-1 model intercepts. In addition, the incorporation of both sets of effects produces a statistically significant *F*-ratio.

Discussion

We have described a two-step approach to testing the statistical significance of random birth-cohort and time-period effects in the class of Hierarchical Age-Period-Cohort models introduced by Yang and Land (2006). This approach focuses first on the examination of the statistical and substantive significance of estimates of individual time-period and cohort effect coefficients and second on assessing the contributions of the period and cohort effects as a set to explained variance in the outcome variable. It moves in four sub-steps from (1) graphical displays of the estimated random effects of individual cohort and period effects to (2) the examination of the statistical significance of the estimated individual cohort and period effects to (3) a variance components analysis of the statistical significance of each of the individual sets of cohort and period effects to (4) formal *F*-tests for null hypotheses of homogeneity of the sets of estimated cohort and period random effects individually and jointly. Empirical examples used for illustrating this strategy came from both the Linear Mixed Effects and Generalized Linear Mixed Models families of statistical models.

[20]Demidenko (2004, 54–55) shows that a fixed effects model that treats the random effects as fixed corresponds to a random effects model with infinite covariance matrix. Thus, l_{max} is an upper-bound estimate for the sum of squares of the mixed model.

Verbal Knowledge

When taking into account all of the foregoing assessments of individual coefficients and sets of coefficients, it must be concluded that, while there is evidence of statistical significance of one time-period effect, and while this is sufficient to conclude that the period effects make a statistically significant contribution to explained variance, the dominant explanation on trends in verbal ability, as measured by the GSS WORDSUM data, is a cohort as opposed to a period one. That is, net of the effects of individual-level covariates in accounting for trends in individuals' verbal test scores, cohort effects are much more prominent than period effects and researchers should indeed study cohort-based explanations for these trends (Alwin 2008, 2009; Alwin and McCammon 1999; Glenn 1999).[21]

Voter Turnout

This analysis confirms the importance of age as a factor in explaining voter turnout and also clarifies the effects of cohort and period on voter turnout. In contrast to Lyons and Alexander (2000), we find only modest evidence that birth cohorts affect voter turnout. Individual cohorts showed little variation, with only a single cohort (1955–1959) reaching statistical significance. Yet, the *F*-test did reveal that birth cohorts improve the fit of the model. The HAPC model as well as the *F*-test found that period had a significant effect on voter turnout. These findings suggest that attempts to explain voter turnout by focusing exclusively on birth cohorts may miss the contributions of period, and in turn, structural factors. Additional analyses can be utilized to test hypotheses about the underlying mechanisms behind the cohort and period findings. This could include random coefficient models to determine variation in how social factors (i.e., residential mobility, marital status) and contextual factors (i.e., unemployment rate, amount spent on political television ads) affect voter turnout across birth cohorts or periods. Results from these analyses could be used to cultivate a more nuanced explanation of why voter turnout has declined and thus develop more precise initiatives to increase voter turnout.

Conclusion

HAPC models have been identified as one way to circumnavigate the identification problem that arises when researchers attempt to assess age, period, and cohort effects using repeated cross-sectional surveys. This paper provides guidelines for assessing the significance of the effects generated by these models and also an *F*-test to further determine the overall contribution of cohort and period effects to the model. By applying these guidelines and the *F*-test to two dissimilar outcome variables (verbal knowledge and voter turnout), we have provided evidence of the wide-ranging potential use of these models to social scientists engaged in research on a variety of topics.

[21]Alwin (1991, 2008, 2009) has consistently argued that, because vocabulary knowledge is acquired first via schooling and then gradually during the life course, the action in the vocabulary test data is in the cohorts. Prior to the development of HAPC models and the careful and detailed applications of statistical tests for the significance of effect coefficients and sets of effect coefficients described in this paper, however, no prior studies have assessed this argument in the context of statistical models that fully incorporate all three sets of effects (age, periods, cohorts). The main substantive inferences from the results reported here provide strong statistical evidence for Alwin's position.

Acknowledgments

The research was supported by University Cancer Research Funds at the Lineberger Comprehensive Cancer Center, University of North Carolina–Chapel Hill, and by grant # K01AG036745 from the National Institute of Aging awarded to Dr. Yang. We are grateful to the Carolina Population Center (R24 HD050924) for general support. The American National Election Studies was supported by the National Science Foundation under grants SES-0535334, SES-0720428, SES-0840550, and SES-0651271; Stanford University; and the University of Michigan. Any opinions, findings, and conclusions or recommendations expressed in these materials are those of the authors and do not necessarily reflect the views of the funding organizations. The authors would like to thank the editor and three anonymous reviewers for comments and suggestions on an earlier version of the manuscript.

References

Alwin, Duane F. Family of Origin and Cohort Differences in Verbal Ability. American Sociological Review. 1991; 56:625–638.

Alwin, Duane F. Social Structure and Cognitive Aging. In: Hofer, SM.; Alwin, DF., editors. Handbook of Cognitive Aging: Interdisciplinary Perspectives. Thousand Oaks, CA: Sage Publications; 2008.

Alwin, Duane F. History, Cohorts, and Patterns of Cognitive Aging. In: Bosworth, HB.; Hertzog, C., editors. Aging and Cognition: Research Methodologies and Empirical Advances. Washington, D.C.: American Psychological Association; 2009. p. 9-38.

Alwin, Duane F.; McCammon, Ryan J. Aging versus Cohort Interpretations of Inter-Cohort Differences in GSS Vocabulary Scores. American Sociological Review. 1999; 64:272–286.

American National Election Studies (ANES; www.electionstudies.org). 2012. The ANES 2008 Time Series Study [data set]. Stanford University and the University of Michigan [producers].

Crainiceanu, Ciprian M.; Ruppert, David. Likelihood Ratio Tests in Linear Mixed Models with One Variance Component. Journal of the Royal Statistical Society: Series B. 2004; 66:165–185.

Davis, James A.; Smith, Tom W.; Marsden, Peter V. General Social Surveys, 1972–2006 [Cumulative File] [Computer file]. ICPSR04697-v2. Chicago: National Opinion Research Center [producer]. Storrs, CT: Roper Center for Public Opinion Research, University of Connecticut/Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributors], 2007-09-10. 2007. doi: 10.3886/ICPSR04697

Demidenko, Eugene. Mixed Models: Theory and Applications. Hoboken: Wiley; 2004.

Frenk, Steven M. Paper presented at the Annual Meeting of the American Sociological Association. Boston: 2008. Voting with My Generation: The Effects of Age, Period, and Birth Cohort on Voting Behavior in U.S. Presidential Elections, 1952–2004.

Glenn, Norval D. Television Watching, Newspaper Reading, and Cohort Differences in Verbal Ability. Sociology of Education. 1994; 67:216–230.

Glenn, Norval D. Further Discussion of the Evidence for an Inter-Cohort Decline in Education-Adjusted Vocabulary. American Sociological Review. 1999; 64:267–271.

Hobcraft, John; Menken, Jane; Preston, Samuel. Age, Period, and Cohort Effects in Demography: A Review. Population Index. 1982; 48:4–43. [PubMed: 12338741]

Jiang, Jiming. Linear and Generalized Linear Mixed Models and Their Applications. New York: Springer; 2007.

Kupper, Lawrence L.; Janis, Joseph M.; Karmous, Azza; Greenberg, Bernard G. Statistical Age-Period-Cohort Analysis: A Review and Critique. Journal of Chronic Disease. 1985; 38:811–830.

Land, Kenneth C.; Hough, George C., Jr; McMillen, Marilyn M. Voting Status Life Tables for the United States, 1968–1980. Demography. 1986; 23:381–402. [PubMed: 3758447]

Lee, Youngjo; Nelder, John A. Hierarchical Generalized Linear Models (with discussion). Journal of the Royal Statistical Society B. 1996; 58:619–678.

Lee, Youngjo; Nelder, John A. Double Hierarchical Generalized Linear Models. Journal of the Royal Statistical Society C (Applied Statistics). 2006; 55:139–185.

Littell, Ramon C.; Milliken, George A.; Stroup, Walter W.; Wolfinger, Russell D.; Schabenberger, Oliver. SAS for Mixed Models. 2nd ed. Cary, NC: SAS Institute; 2006.

Lu, Yan; Zhang, Guoyi. The Equivalence between Likelihood Ratio Test and *F*-test for Testing Variance Component in a Balanced One-Way Random Effects Model. Journal of Statistical Computation and Simulation. 2010; 80:443–450.

Lyons, William; Alexander, Robert. A Tale of Two Electorates: Generational Replacement and the Decline of Voting in Presidential Elections. Journal of Politics. 2000; 62:1014–1034.

Mason, Karen O.; Mason, William M.; Winsborough, HH.; Poole, W Kenneth. Some Methodological Issues in Cohort Analysis of Archival Data. American Sociological Review. 1973; 38:242–258.

Mason, William M.; Fienberg, Stephen E., editors. Cohort Analysis in Social Research: Beyond the Identification Problem. New York: Springer-Verlag; 1985.

Mason, William M.; Wolfinger, Nicholas H. Cohort Analysis. In: Smelser, Neil J.; Baltes, Paul B., editors. International Encyclopedia of the Social and Behavioral Sciences. Amsterdam: Elsevier Science; 2001. p. 2189-2194.

McCullagh, Peter; Nelder, John A. Generalized Linear Models. 2nd ed. New York: Chapman and Hall; 1989.

Menard, Scott. Logistic Regression: From Introductory to Advanced Concepts and Applications. Thousand Oaks, CA: Sage; 2010.

Preston, Samuel H.; Heuveline, Patrick; Guillot, Michel. Demography: Measuring and Modeling Population Processes. Malden, MA: Blackwell; 2001.

Raftery, Adrian E. Bayesian Model Selection in Social Research. Sociological Methodology. 1995; 25:111–164.

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

Raudenbush, Stephen W.; Bryk, Anthony S.; Cheong, Yuk Fai; Congdon, Richard T, Jr. HLM6: Hierarchical Linear and Nonlinear Modeling. Scientific Software International; 2004.

Reither, Eric N.; Hauser, Robert M.; Yang, Yang. Do Birth Cohorts Matter? Age-Period-Cohort Analyses of the Obesity Epidemic in the United States. Social Science & Medicine. 2009; 69:1439–1448. [PubMed: 19773107]

Ryder, Norman B. The Cohort as a Concept in the Study of Social Change. American Sociological Review. 1965; 30:843–861. [PubMed: 5846306]

Schwadel, Philip. Age, Period, and Cohort Effects on U.S. Religious Service Attendance: The Declining Impact of Sex, Southern Residence, and Catholic Affiliation. Sociology of Religion. 2010; 71:2–24.

Snijders, Tom AB.; Bosker, Roel. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, CA: Sage; 1999.

Sociological Methods and Research. Special Issue: Advances in Age-Period-Cohort Analysis. 2008; 36:287–428.

Squire, Peverill; Wolfinger, Raymond E.; Glass, David P. Residential Mobility and Voter Turnout. American Political Science Review. 1987; 81:45–66.

Wasserman, Larry. All of Statistics: A Concise Course in Statistical Inference. New York: Springer; 2004.

Wilson, James A.; Gove, Walter R. The Inter-Cohort Decline in Verbal Ability: Does It Exist? American Sociological Review. 1999a; 64:253–266.

Wilson, James A.; Gove, Walter R. Reply: The Age-Period-Cohort Conundrum and Verbal Ability: Empirical Relationships and Their Interpretation. American Sociological Review. 1999b; 64:287–302.

Wolfinger, Nicholas H.; Wolfinger, Raymond E. Family Structure and Voter Turnout. Social Forces. 2008; 86:1513–1528.

Wolfinger, Raymond E.; Rosenstone, Steven J. Who Votes?. New Haven, CT: Yale University Press; 1980.

Yang, Yang. Bayesian Inference for Hierarchical Age-Period-Cohort Models of Repeated Cross-Section Survey Data. Sociological Methodology. 2006; 36:39–74.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Yang, Yang; Land, Kenneth C. 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. Sociological Methodology. 2006; 36:75–97.

Yang, Yang; Land, Kenneth C. Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Fixed or Random Effects? Sociological Methods and Research. 2008; 36:297–326.

Yang, Yang; Schulhofer-Wohl, Sam; Fu, Wenjiang J.; Land, Kenneth C. The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It. American Journal of Sociology. 2008; 113:1697–1636.

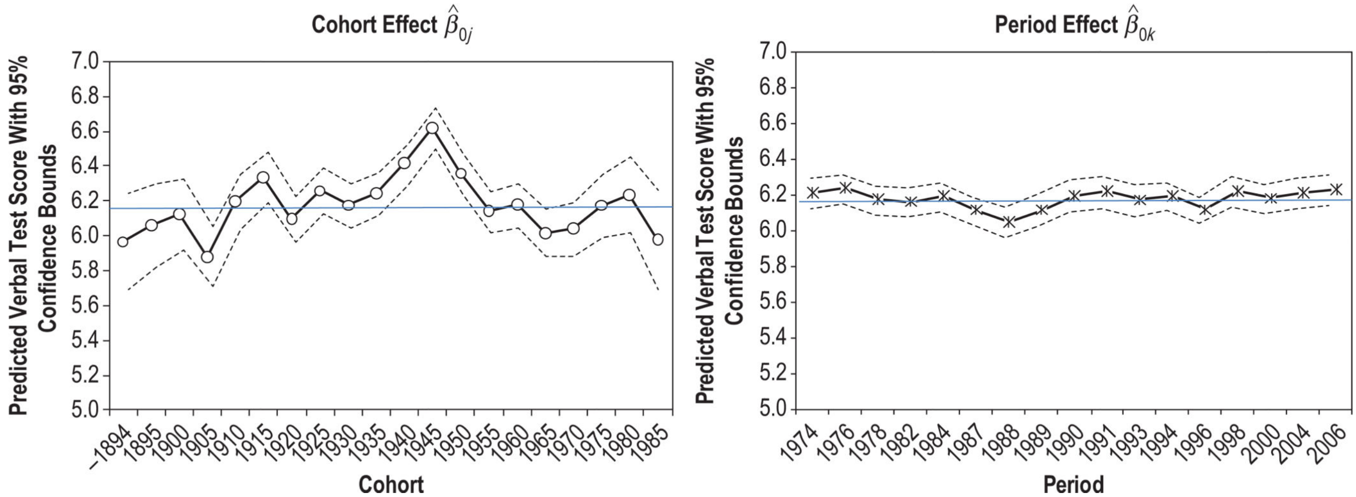

Figure 1.

Estimated Cohort and Period Effects and 95-Percent Confidence Bounds for GSS Verbal Ability Model

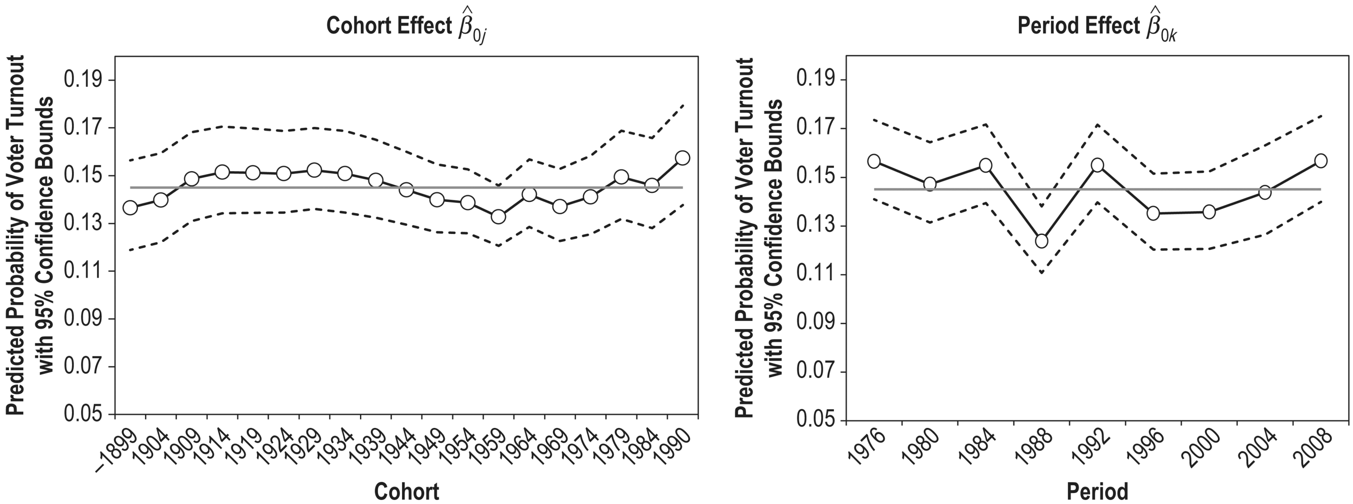

Figure 2.

Estimated Cohort and Period Effects and 95-Percent Confidence Bounds for ANES Voter Turnout Model

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Descriptive Statistics for GSS Vocabulary Test Data, 1974–2006

Variables	Definition	*N*	Mean	SD	Min	Max
WORDSUM	A 10-item composite vocabulary scale score	22,042	6.05	2.13	0	10
AGE	Age at survey year	22,042	45.13	17.37	18	89
EDUCATION	Highest levels of education completed	22,042	12.84	3.02	0	20
FEMALE	Sex: 1 = female; 0 = male	22,042	0.57	.50	0	1
BLACK	Race: 1 = black; 0 = white	22,042	0.15	.36	0	1
COHORT	Five-year birth cohorts	20			−1894	1985–89
PERIOD	Survey years	17			1974	2006

Table 2

HAPC-CCREM Model of the GSS WORDSUM Data, 1974–2006

Fixed effects	**Coefficient**	**se**	***t*-ratio**	***p*-value**
INTERCEPT	6.175	0.055	112.50	< .001
AGE	0.026	0.015	1.71	.087
AGE^2	−0.057	0.005	−11.87	< .001
FEMALE	0.229	0.024	9.49	< .001
BLACK	−1.030	0.034	−30.07	< .001
EDUCATION	0.366	0.004	86.57	< .001
Random effects				
Cohort	**Coefficient**	**se**	***t*-ratio**	***p*-value**
1894	−0.210	0.142	−1.48	0.140
1895	−0.114	0.123	−0.93	0.353
1900	−0.051	0.104	−0.49	0.625
1905	−0.294	0.090	−3.27	0.001
1910	0.021	0.081	0.26	0.797
1915	0.163	0.073	2.22	0.027
1920	−0.079	0.068	−1.15	0.249
1925	0.083	0.068	1.23	0.220
1930	0.001	0.067	0.01	0.990
1935	0.068	0.064	1.06	0.289
1940	0.240	0.061	3.91	< .001
1945	0.447	0.060	7.50	< .001
1950	0.184	0.059	3.10	0.002
1955	−0.035	0.061	−0.57	0.568
1960	0.002	0.065	0.04	0.970
1965	−0.157	0.071	−2.20	0.028
1970	−0.135	0.080	−1.70	0.090
1975	−0.001	0.092	−0.01	0.990
1980	0.062	0.112	0.55	0.583
1985	−0.195	0.146	−1.34	0.180
Period				
1974	0.033	0.043	0.77	0.442
1976	0.060	0.043	1.41	0.158
1978	−0.002	0.042	−0.04	0.967
1982	−0.014	0.040	−0.36	0.718
1984	0.016	0.042	0.37	0.709
1987	−0.061	0.040	−1.52	0.129
1988	−0.128	0.046	−2.76	0.006

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Fixed effects	**Coefficient**	**se**	***t*-ratio**	***p*-value**
Fixed effects	**Coefficient**	**se**	***t*-ratio**	***p*-value**
1989	−0.061	0.046	−1.34	0.182
1990	0.020	0.047	0.43	0.670
1991	0.042	0.046	0.92	0.358
1993	−0.004	0.045	−0.09	0.926
1994	0.019	0.039	0.49	0.623
1996	−0.060	0.039	−1.52	0.128
1998	0.044	0.043	1.02	0.306
2000	0.005	0.043	0.11	0.915
2004	0.038	0.043	0.88	0.381
2006	0.052	0.045	1.16	0.247
Variance components	**Variance**	**se**	***z*-statistic**	***p*-value**
Cohort	0.034	0.013	2.56	.010
Period	0.005	0.003	1.49	.135
Individual	3.116	0.030	104.87	< .001
Deviance	87707.2	df = 21,999		

Table 3

F-Tests for the Presence of Random Effects, GSS WORDSUM Data

	Cohort effects $\tau_u = 0$ **vs.** $\tau_u > 0$	**Period effects** $\tau_v = 0$ **vs.** $\tau_v > 0$	**Cohort and period effects** $\tau_u = \tau_v = 0$ **vs.** τ_u **or** $\tau_v > 0$
S_{OLS}	69,377	69,377	69,377
S_{min}	68,696	69,268	68,558
R	25	22	42
M	5	5	5
N_T	22,042	22,042	22,042
$(S_{OLS} - S_{\min}) / (r - m)$	34.05	6.41	22.14
$S_{\min} / (N_T - r)$	3.12	3.15	3.12
F	10.9	2.03	7.096
$f_{0.95}(r - m, N_T - r)$	1.571	1.623	1.411

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Descriptive Statistics for ANES Voter Turnout Data, 1976–2008

Variable	Description	Mean	SD	Min	Max
VOTE	1 = voted in U.S. presidential election; 0 = did not vote	0.76	0.43	0	1
AGE	Age at survey year	46.63	17.71		
MALE	Sex: 1 = male; 0 = female	0.44	0.5	0	1
BLACK	Race: 1 = black; 0 = white	0.15	0.36	0	1
EDUCATION	1 = less than high school; 2 = high school; 3 = some college; 4 = bachelor or advanced degree	2.65	0.92	1	4
MARRIED	1 = married/cohabiting; 0 = not currently married/cohabiting	0.57	0.5	0	1
MOBILITY	Length of time in current dwelling: 1 = 0–4 years; 2 = 5–9 years; 3 = 10–19 years; 4 = 20–29 years; 5 = 30 or more years	2.23	1.34	1	5
COHORT	Five-year birth cohorts	19		−1899	1985–1990
PERIOD	Survey years	9		1976	2008
N = 13,508					

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

HAPC-GLMM Model of the ANES Voter Turnout Data, 1976–2008

Fixed effects	**Coefficient**	**se**	***t*-ratio**	***p*-value**
Intercept	−1.774	0.110	−16.18	< 0.001
AGE	0.023	0.002	14.05	< 0.002
AGE^2	−0.001	0.0001	−7.66	< 0.003
MALE	0.003	0.0001	0.06	0.951
BLACK	0.115	0.061	1.9	0.058
EDUCATION	0.942	0.03	31.46	< 0.001
MARRIED	0.47	0.046	10.11	< 0.001
MOBILITY	0.262	0.02	13.32	< 0.001
Random effects				
Cohort	**Coefficient**	**se**	***t*-ratio**	***p*-value**
1899	−0.07	0.081	−0.87	0.385
1904	−0.043	0.079	−0.55	0.583
1909	0.029	0.075	0.38	0.701
1914	0.051	0.072	0.71	0.478
1919	0.049	0.07	0.7	0.486
1924	0.046	0.068	0.68	0.494
1929	0.057	0.067	0.85	0.397
1934	0.046	0.068	0.67	0.5
1939	0.024	0.066	0.37	0.713
1944	−0.008	0.063	−0.13	0.9
1949	−0.042	0.06	−0.71	0.48
1954	−0.052	0.057	−0.91	0.364
1959	−0.103	0.056	−1.83	0.067
1964	−0.024	0.059	−0.41	0.683
1969	−0.066	0.065	−1.01	0.311
1974	−0.032	0.069	−0.47	0.642
1979	0.035	0.074	0.47	0.639
1984	0.007	0.077	0.09	0.93
1990	0.096	0.08	1.2	0.229
Period				
1976	0.09	0.063	1.42	0.156
1980	0.017	0.067	0.25	0.799
1984	0.077	0.063	1.22	0.224
1988	−0.183	0.064	−2.86	0.004
1992	0.078	0.062	1.25	0.212
1996	−0.082	0.068	−1.21	0.227

Fixed effects	**Coefficient**	**se**	***t*-ratio**	***p*-value**
Fixed effects	**Coefficient**	**se**	***t*-ratio**	***p*-value**
2000	−0.077	0.069	−1.11	0.266
2004	−0.01	0.076	−0.13	0.897
2008	0.091	0.068	1.33	0.182
Variance components	**Variance**	**se**	***z*−statistic**	***p*-value**
COHORT	0.008	0.008	0.88	0.188
PERIOD	0.013	0.008	1.51	0.066
Model fit				
-2 Res Log Pseudo-Likelihood	65678.03	df = 13474		

Table 6

F-Tests for the Presence of Random Effects, ANES Voter Turnout Data[a]

	Cohort effects $\tau_u = 0$ vs. $\tau_u > 0$	**Period effects** $\tau_v = 0$ vs. $\tau_v > 0$	**Cohort and period effects** $\tau_u = \tau_v = 0$ vs. τ_u or $\tau_v > 0$
l_0	12,934	12,934	12,934
l_{max}	12,896	12,901	12,867
R	26	16	35
M	7	7	7
N_T	13,508	13,508	13,508
$(l_0 - l_{max})/(r - m)$	2.00	3.67	2.39
$l_{max}/(N_T - r)$	0.96	0.96	0.96
F	2.09	3.83	2.51
$f_{0.95}\ (r - m, N_T - r)$	1.59	1.88	1.48