2

A METHODOLOGICAL COMPARISON OF AGE-PERIOD-COHORT MODELS: THE INTRINSIC ESTIMATOR AND CONVENTIONAL GENERALIZED LINEAR MODELS

*Yang Yang**
Wenjiang J. Fu†
*Kenneth C. Land**

Age-period-cohort (APC) accounting models have long been objects of attention in statistical studies of human populations. It is well known that the identification problem created by the linear dependency of age, period, and cohort (Period = Age + Cohort or P = A + C) presents a major methodological challenge to APC analysis, a problem that has been widely addressed in demography, epidemiology, and statistics. This paper compares

Revision of a paper presented at the annual meeting of the Population Association of America in May 2003. We thank Robert Schoen and Shiro Horiuchi and the anonymous reviewers for useful comments and discussions that have helped to improve the paper. The research reported herein was supported in part by NIH/NIA Grant Numbers R01AG07198, P30AG12852, and K07AG00892, and in part by NIH/NCI Grant Number R03CA83010 and R25CA90301 and the Michigan State University IRGP program. Direct correspondence to Yang Yang, Department of Sociology and Center for Demographic Studies, P. O. Box 90088, Duke University, Durham, NC 27708-0088;
E-mail: yangy@soc.duke.edu.

*Duke University
†Texas A&M University and Michigan State University

parameter estimates and model fit statistics produced by two solutions to the identification problem in age-period-cohort models—namely, the conventional demographic approach of constrained generalized linear models (Fienberg and Mason 1978, 1985; Mason and Smith 1985) and the intrinsic estimator method recently developed by Fu (2000; Knight and Fu 2000; Fu, Hall, and Rohan 2004). We report empirical analyses of applications of these two methods to population data on U.S. female mortality rates. Comparisons of parameter estimates suggest that both constrained generalized linear models and the intrinsic estimator method can yield similar estimates of age, period, and cohort effects, but estimates obtained by the intrinsic estimator are more direct and do not require prior information to select appropriate model identifying constraints. We also describe three statistical properties of the estimators: (1) finite-time-period bias, (2) relative statistical efficiency, and (3) consistency as the number of periods of observed data increases. These empirical analyses and theoretical results suggest that the intrinsic estimator may well provide a useful alternative to conventional methods for the APC analysis of demographic rates.

1. INTRODUCTION

The *age-period-cohort (APC) accounting model* has long been an object of attention in statistical studies of human populations. It serves as a general methodology for cohort analysis when all three factors—age, period, and cohort—are potentially of interest. It has been a popular tool in demography and epidemiology to identify age, period, and cohort trends in disease incidence and/or mortality rates (Hobcraft, Menken, and Preston 1982; Robertson, Gandini, and Boyle 1999). One common aim of fitting APC models is to assess the effects of the three factors on demographic or disease rates. The *age effects* represent differing risks associated with different age groups. The *period effects* represent variation in vital rates over time that is associated with all age groups simultaneously. The *cohort effects* are associated with changes in rates across groups of individuals with the same birth years—that is, for successive age groups in successive time periods. To date, scientists have reached a consensus that methodological guidance is needed to address the fundamental

question in cohort analysis: to determine whether the phenomenon of interest is cohort-based or some other factors such as age or calendar year are more relevant. This points to the necessity of statistically estimating the age, period, and cohort effects.

A major methodological challenge arises, however, because of the *identification problem* induced by the exact linear dependency between age, period, and cohort: Period = Age + Cohort, which can be viewed as a special case of collinear regressors that yields, in this case, a singular design matrix of one less than full rank. Since a singular design matrix produces multiple estimators of the three effects, it is difficult to estimate the true separate effects. This problem has been widely addressed in demography and epidemiology. The pioneering work of Mason et al. (1973) laid out a general framework for cohort analysis—namely, the multiple classification model. It was followed by a number of methodological discussions on the specifications and estimations of the APC model in social and demographic research (e.g., Glenn 1976; Fienberg and Mason 1978, 1985; Hobcraft, Menken, and Preston 1982; Wilmoth 1990; O'Brien 2000). In biostatistics and epidemiology, a number of solutions to the identification problems have also been proposed over the past two decades (e.g., Osmond and Gardner 1982; Clayton and Schifflers 1987; Holford 1992; Tarone and Chu 1992; Robertson and Boyle 1998; Fu 2000; Knight and Fu 2000). Generally speaking, the limitations of existing approaches have been widely acknowledged by statisticians and statistical methodologists in all disciplines. As a result, the identification problem still remains largely unsolved.

Recent developments in APC methodology in biostatistics have emphasized the utility of estimable functions that are invariant to the selection of constraints on the parameters (Clayton and Schifflers 1987; Holford 1983, 1991, 1992; Robertson, Gandini, and Boyle 1999; Tarone and Chu 1992, 2000). This is the approach applied by Fu (2000; Knight and Fu 2000; Fu, Hall, and Rohan 2004) in the derivation of a new APC estimator—which he calls the *intrinsic estimator*—based on estimable functions and the singular value decomposition of matrices.

In this paper, we present a methodological comparison of two approaches to the identification problem in APC models—namely, the intrinsic estimator method and the constrained generalized linear models estimator that has become conventional among demographers

and other social scientists (Fienberg and Mason 1978, 1985; Mason and Smith 1985). We demonstrate, through data analyses of population mortality rates, similarities and differences of these two methods in parameter estimates and model fit. We also discuss some key statistical properties, such as finite-time-period bias, the relative efficiency of the intrinsic estimator compared to the constrained generalized linear models estimator, and the consistency of the estimators. We conclude from these results that the intrinsic estimator appears to offer a useful alternative to conventional methods for the APC analysis of demographic rates.

2. THE IDENTIFICATION PROBLEM IN APC MODELS

The *APC accounting/multiple classification model* was articulated for demographic and social research some 30 years ago by Mason and colleagues (1973). For mortality rates, 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 ith age group for $i = 1, \ldots, a$ age groups at the jth time period for $j = 1, \ldots, p$ time periods of observed data; D_{ij} denotes the number of deaths in the ijth group; P_{ij} denotes the size of the estimated population in the ijth group, the population at risk of death; μ denotes the intercept or adjusted mean death rate; α_i denotes the ith row age effect or the coefficient for the ith age group; β_j denotes the jth column period effect or the coefficient for the jth time period; γ_k denotes the kth diagonal cohort effect or the coefficient for the kth cohort for $k = 1, \ldots, (a+p-1)$ cohorts, with $k = a - i + j$; and ϵ_{ij} denotes the random errors with expectation $E(\epsilon_{ij}) = 0$.

Conventional APC models as represented in (1) fall into the class of generalized linear models (GLIM) that can take various alternative forms such as the following. First, model (1) can take a *log-linear regression form* via a *log link* as

$$\log(E_{ij}) = \log(P_{ij}) + \mu + \alpha_i + \beta_j + \gamma_k, \tag{2}$$

where E_{ij} denotes the expected number of deaths in cell (i, j) that is assumed to be distributed as a Poisson variate, and $\log(P_{ij})$ is the log of the exposure P_{ij} in (1) and is called the "offset" or adjustment for the log-linear contingency table model. Models of this type are most popular in epidemiology where the disease counts generally follow Poisson distributions and the disease rates are estimated through log-linear models (for a review of log-linear models, see, e.g., Agresti [1996]). The second alternative formulation of the model is to treat the underlying mortality as dichotomous, distributed as a binomial variate. The canonical link changes from a log link to a *logit link*, which yields a *logit model*:

$$\theta_{ij} = \log\left(\frac{m_{ij}}{1 - m_{ij}}\right) = \mu + \alpha_i + \beta_j + \gamma_k, \tag{3}$$

where θ_{ij} is the log-odds of death and m_{ij} is the probability of death in cell (i, j). This model has been implemented more widely in demographic research (e.g., Mason and Smith 1985). Regression models (1), (2), and (3) can be treated as *fixed-effect generalized linear models* after a reparameterization to center the parameters:

$$\Sigma_i \alpha_i = \Sigma_j \beta_j = \Sigma_k \gamma_k = 0. \tag{4}$$

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

After reparameterization (4), model (1) can be written in the conventional matrix form of a least-squares regression:

$$Y = Xb + \varepsilon. \tag{5}$$

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 dimension $m = 1 + (a - 1) + (p - 1) + (a + p - 2)$) of model parameters b

$$b = (\mu, \alpha_1, \ldots \alpha_{a-1}, \beta_1, \ldots, \beta_{p-1}, \gamma_1, \ldots, \gamma_{a+p-2})^T, \tag{6}$$

with the T superscript denoting vector transposition, and where ϵ in (5) is a vector of random errors with mean 0 and constant diagonal variance matrix $\sigma^2 I$, with I denoting an identity matrix. Note that the parameters α_a, β_p, and γ_{a+p-1} are not included in the parameter vector b because of the constraints (4) and can be uniquely determined by use of (4) in conjunction with each estimator of b. An alternative parameterization often used in APC models consists of setting one of each of the α, β, and γ coefficients—say, either the first or the last—equal to zero and then estimating the remaining coefficients relative to these "reference" age, period, and cohort categories. We later illustrate here, by an empirical example, that the use of reference categories is equivalent to the translation by a constant of the parameter estimates produced by the constraints (4) and thus of no substantive importance.

The ordinary least squares (OLS) estimator of the matrix regression model (5) is the solution b of the normal equations

$$\hat{b} = (X^T X)^{-1} X^T Y. \tag{7}$$

But this estimator *does not exist* (i.e., it is not a unique vector of coefficient estimates). This is due to the fact that the design matrix X is singular with one less than full column rank (Kupper et al. 1985). Therefore, $X^T X$ is not invertible, due to the perfect linear relationship between the age, period, and cohort effects:

$$\text{Period} - \text{Age} = \text{Cohort}.$$

This means that some of the columns of X can be linearly combined or summed up to produce a column identical to other columns in X. This is the *model identification problem* of APC analysis. It implies that there are an infinite number of possible solutions of the matrix equation (7)—that is, OLS estimators of model (5)—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 (4).

3. TWO APPROACHES TO SOLVING THE MODEL IDENTIFICATION PROBLEM

3.1. *The Conventional Approach in Demography: Identifying Constraints*

Since the work of Fienberg and Mason (1978, 1985) more than two decades ago on the development of a log-linear contingency tables representation of the APC analysis problem, the following has become the conventional approach to APC analysis in demography: Place (at least) one additional identifying constraint on the parameter vector (6)—for example, constrain the effect coefficients of the first two periods to be equal, $\beta_1 = \beta_2$. With this one additional constraint, the matrix $(X^T X)$ becomes nonsingular and the least squares estimator (7) exists, as do related maximum-likelihood estimators for models (1), (2), and (3). An example of this approach is the APC analysis of the tuberculosis mortality of U.S. males by Mason and Smith (1985) that imposed equality constraints on neighboring age and period groups to identify the model. We will refer to this approach to estimation as the *constrained generalized linear models estimator* (*CGLIM*).

The main problem of this approach is that its methodological usefulness depends on strong prior information for identifying these restrictions. As has been known at least since the work of Mason and Smith (1985), and as will be demonstrated with U. S. female mortality rates herein below, estimates of model effect coefficients are sensitive to the arbitrary choice of the identifying constraint. This has led to a large methodological literature in demography, epidemiology, and statistics (e.g., see the review of cohort analysis by Mason and Wolfinger [2002]) advising APC analysts of this sensitivity to model specification and that the choice of the model identifying constraint must be based on prior theoretical or empirical information that, unfortunately, rarely exists.

3.2. *The Intrinsic Estimator*

Another line of research using APC models has evolved in epidemiological studies of human disease and cause-specific mortality rates. For the past 20 years or so, biostatisticians have addressed the

identification problem with similar approaches. They have also demonstrated the effects of using different arbitrary constraints on the parameters to ensure identifiability (Clayton and Schifflers 1987; Holford 1992). More recent methodological research in biostatistics has proposed some solutions other than those based on arbitrary linear constraints. A complete review and comparison of these methodologies is available by Roberton, Gandini, and Boyle (1999).

Recent developments in APC methodology in biostatistics have emphasized the utility of estimable functions that are invariant to the selection of constraint on the parameters (Clayton and Schifflers 1987; Holford 1983, 1991, 1992; Kupper et al. 1983, 1985; Robertson, Gandini, and Boyle 1999; Tarone and Chu 1992, 2000). This is the approach taken by Fu (2000; Knight and Fu 2000; Fu, Hall, and Rohan 2004) in the derivation of the intrinsic estimator. Based on estimable functions and the singular value decomposition of matrices, the intrinsic estimator yields robust estimates of disease trends by age, period, and cohort and uniquely determines the coefficient estimates. In the following, we briefly summarize the structure and statistical properties of the intrinsic estimator and its computational algorithms.

First, note that the linear relationship between the age, period, and cohort effects in the APC regression model (5) can be expressed mathematically in matrix form for some nonzero vector B_0 as

$$XB_0 = 0. \tag{8}$$

Equation (8) is an expression of the property that X is singular—i.e., that there is some linear combination of the columns of the design matrix X that equals a zero vector. Since the design matrix X is one-less-than-full column rank (Kupper et al. 1985), the parameter space P can be decomposed as $\mathrm{P} = \mathrm{N} \oplus \Theta$, where $\oplus$ is the direct sum of two linear spaces that are perpendicular to each other; N is the one-dimensional null space of X spanned by the vector $\{sB_0\}$ with real number s; and Θ is the complement subspace orthogonal to N.

Second, note that the matrix X^TX has a unique eigenvalue 0 (Christensen 2002: 381; Fu, Hall, and Rohan 2004). Denote the corresponding eigenvector by B_0; then B_0 represents a special direction in the parameter space with a Euclidean norm equal to 1. This special direction in the parameter space is defined by the difference

$\hat{b}_1 - \hat{b}_2$ between two arbitrary OLS estimators (7) of model (5)—that is, any two estimators obtained by imposing arbitrary identifying constraints on model (5)—and that

$$X(\hat{b}_1 - \hat{b}_2) = X(tB_0) = 0, \tag{9}$$

where t is an arbitrary real number and B_0 is the eigenvector corresponding to the zero eigenvalue of X. Thus, the difference of any two arbitrary estimators of model (5)—each obtained by the imposition of one equality constraint on the regression coefficients—must be in the null space of X—that is, the space defined by the eigenvector of the eigenvalue zero.

Third, due to the orthogonal decomposition of the parameter space noted above, each of the infinite number of estimators of model (5) can be written as

$$\hat{b} = B + tB_0, \tag{10}$$

where tB_0 is as defined in (9). The eigenvector B_0 depends only on the design matrix X, not on the observed rates Y, and thus it is completely determined by the numbers of age groups and period groups regardless of the event rates. In other words, B_0 has a specific form that is a function of the design matrix; for a specific example, see Appendix A. B_0 yields a linear trend among the age, period, and cohort coefficients; and *the arbitrary term tB_0 represents arbitrary linear trends*. This is a key point, as intuition suggests that the eigenvector corresponding to the zero eigenvalue should be an arbitrary vector. And, indeed, tB_0 is arbitrary. But B_0 is not arbitrary; it is fixed by the design matrix. Furthermore, equation (10) establishes that any APC estimator obtained by placing any identifying constraint(s) on the design matrix can be written as a linear combination $B + tB_0$, where B *is termed the intrinsic estimator* (IE) that lies in the space Θ that is orthogonal to N, the null space, and is uniquely determined by the rates Y and matrix X.[1] The special estimator B in equation (10) is determined by the

[1]Fu, Hall, and Rohan (2004) show that $E(B)$ is the only estimable function that determines the coefficient estimates of APC trends, and we will note later, in a section on properties of the APC estimators, that B satisfies a certain condition for estimability.

Moore-Penrose generalized inverse.[2] This is due to the structure of the estimators (10) and the parameter space. As noted earlier, the parameter space can be decomposed into a null space of dimension one and a non-null space. The estimator B is in the non-null space and thus corresponds to the Moore-Penrose generalized inverse matrix.

In brief, corresponding to the decomposition of the estimators in equation (10), we have the corresponding decomposition of the parameter vector (6) as

$$b = b_0 + tB_0, \tag{11}$$

where $b_0 = P_{proj}\, b$ is a special parameter vector that is a linear function of b corresponding to the projection of the parameter vector (6) to the non-null space of X. Specifically, the special parameter vector b_0 corresponding to $t = 0$ satisfies the projection

$$b_0 = (I - B_0 B_0^T)b, \tag{12}$$

where b is the parameter vector (6). As illustrated in Figure 1, this shows the projection of two arbitrary parameterizations, b_1 and b_2, onto the non-null parameter space (the vertical axis in Figure 1), which is independent of the arbitrary real number t. The geometric representation in this figure can be thought of either as a simple parameter space of dimension two or as multidimensional with the vertical axis representing a direction in a multidimensional non-null space. In either case, since the projection of any arbitrary parameterization b in Figure 1 yields the same parameter vector b_0, the latter is estimable.[3]

Figure 1 also helps to illustrate geometrically that the IE may in fact also be viewed as a constrained estimator. But, in contrast to the equality constraints on two or more coefficients of the parameter vector b that are imposed by the CGLIM models, the constraint imposed by the IE to identify model (5) is a constraint on the

[2]For example, see Searle (1971: 16–19) for a definition of the Moore-Penrose generalized inverse and its properties.

[3]Figure 1 provides only a geometric illustration. Full algebraic details of the proof that b_0 is estimable and the only estimable function that determines both the linear and nonlinear trends in the age, period, and cohort coefficients, see Fu, Hall, and Rohan (2004).

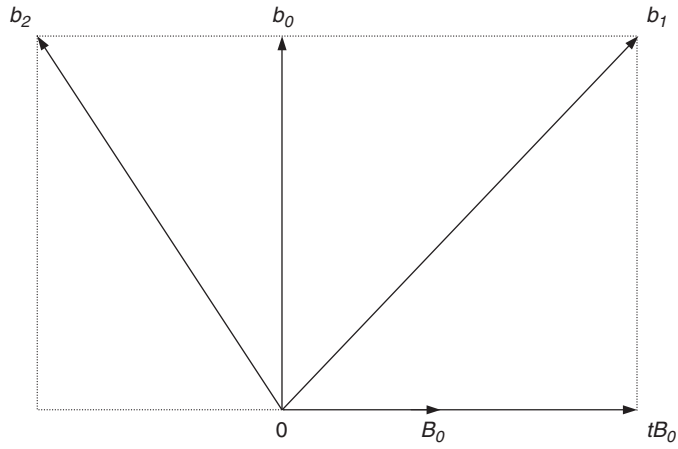

FIGURE 1. Geometric projection of parameter vectors b_1 and b_2 to the vertical axis to yield the estimable function b_0.

Note: b_1 and b_2 are two arbitrary sets of parameter vectors that result from the imposition of equality constraints on elements of the parameter vector b of the APC regression model; tB_0 represents the arbitrary term in these parameter vectors; projection of the parameter vectors to the vertical axis yields the estimable function b_0. The vertical vector b_0 is orthogonal to and therefore independent of the arbitrary component tB_0. Although b_0 is estimable, none of the others, $b_0 + tB_0$ with $t \neq 0$ such as b_1 and b_2, are.

geometric orientation of the parameter vector b in parameter space. Specifically, the IE imposes the constraint that the direction in parameter space defined by the eigenvector B_0 in the null space of the design matrix X has zero influence on the parameter vector b_0 (i.e., on the specific parameterization of the vector b that is estimated by the IE). Since B_0 is a fixed vector that is a function solely of the design matrix (e.g., the number of time periods of data in an analysis) and does not depend on the observed event rates or frequencies being analyzed, this would seem to be a desirable constraint. And, indeed, later herein we show that the IE has some nice statistical properties.

Corresponding to the projection of the parameter vector b onto b_0, we have the following projection of the estimators of equation (10) onto the intrinsic estimator B

$$B = (I - B_0 B_0^T)\hat{b}. \tag{13}$$

This equation provides one algorithm for computing the IE—namely, compute an arbitrary estimator $\hat{b}$ of model (5), say by an equality constraint on two of the age, period, or cohort parameters, and then geometrically project $\hat{b}$ to the intrinsic estimator B by removing the component in the B_0 direction. A second computational algorithm for the IE is a principal components regression method, whereby we (1) compute the eigenvalues and eigenvectors (principal components) of the matrix X^TX; (2) normalize them to have unit length; (3) identifiy the eigenvector B_0 corresponding to the unique eigenvalue 0; (4) estimate a regression model with response vector Y as in equation (5) and design matrix U whose column vectors are the principal components determined by the eigenvectors of nonzero eigenvalues, (i.e., we estimate a principal components regression model); and (5) use the orthonormal matrix of all eigenvectors to transform the coefficients of the principal components regression model to the regression coefficients of the intrinsic estimator B.

4. DATA ANALYSIS: COMPARISONS OF MODEL ESTIMATES

4.1. *Data*

We now develop an empirical application and systematic comparison of the conventional CGLIM and IE estimators by an analysis of population data. The data we used are from the Berkeley Human Mortality Database. For the present methodological discussion, we focus on female mortality rates in the United States from 1960 to 1999. The mortality data are shown in Age $\times$ Period rectangular arrays in Tables 1 and 2. Both age and period are of five-year interval lengths, so the diagonal elements of the matrices correspond to cohorts.

The observed age-specific death rates, ${}_nM_x$, are plotted by period and cohort from age 0 to ages 110+ in Figure 2. Panel (a) shows the death rates by age for the past 40 years. Mortality rates show slight decreases in more recent years as compared to earlier years for ages less than 95. For females ages 95 and over, however, the age-specific death rates appear to have increased in recent periods. It is possible that age misreporting could be responsible, at least in

TABLE 1
Deaths by Age and Period for U.S. Females: 1960–1999

Age/Year	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99
0	260,722	204,152	155,978	120,659	108,658	100,491	90,870	74,991
5	18,778	18,286	15,525	11,542	9,143	8,651	8,059	7,732
10	13,631	14,736	15,070	11,749	9,496	8,201	8,314	8,478
15	19,550	26,070	29,909	28,350	23,736	21,678	19,322	19,897
20	21,079	27,062	32,464	32,502	30,822	26,953	23,818	21,113
25	24,381	26,524	31,161	32,240	33,087	34,666	32,126	27,975
30	36,181	34,724	35,449	34,855	37,659	43,692	47,488	42,973
35	58,069	54,865	49,078	43,017	45,283	53,505	62,876	67,249
40	84,879	88,611	79,611	63,376	59,764	67,023	79,507	91,724
45	118,793	128,535	126,307	102,225	86,975	87,857	101,409	119,147
50	162,651	172,901	177,271	162,403	142,094	128,388	131,767	152,750
55	205,521	226,392	233,485	226,245	224,558	203,293	185,474	195,702
60	280,559	289,307	308,622	302,720	317,144	322,048	290,982	273,073
65	378,465	390,665	397,338	397,335	421,324	442,461	438,269	408,104
70	483,099	504,954	509,123	493,894	539,572	570,458	585,914	601,328
75	537,764	597,621	619,849	595,027	640,503	705,885	728,370	782,586
80	510,834	587,648	650,397	659,074	707,360	786,926	848,119	934,583
85	366,273	429,646	500,196	559,903	650,291	732,089	815,174	955,085
90	167,327	205,192	248,234	305,174	404,208	500,808	563,213	694,251
95	43,082	55,762	71,744	94,421	138,464	197,922	240,546	294,860
100	6,464	7,860	11,231	15,621	24,588	37,262	52,017	65,560
105	743	869	1,196	1,660	2,372	3,550	5,109	6,833
110+	224	178	231	274	339	369	396	425

part, for this pattern. It is also possible that recent rapid increases of the numbers of females in these oldest age categories may have contributed to this crossover in the pattern of period-specific changes in mortality rates.

Panels (b) and (c) of Figure 2 present the mortality data by age and cohort. Since there is a large array of age and cohort groups, we separately show the cohort mortality rates for the younger and older age groups. In Panel (b), for ages 5 to 64 it is evident that there are successive declines in mortality across the cohorts from the earliest to the most recent. The decreases of mortality rates for younger cohorts continue to the age of 95, as shown in panel (c), after which the pattern reverses. Again, these period-specific mortality crossovers could be

TABLE 2
Population Exposure by Age and Period for U.S. Females: 1960–1999

Age/Year	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99
0	50,310,260	45,991,285	42,023,196	39,081,739	41,932,221	44,241,299	47,304,665	46,927,652
5	48,272,344	50,375,584	46,204,563	42,905,645	39,579,319	42,404,636	44,771,611	48,022,494
10	43,830,192	48,793,799	51,257,058	47,565,854	43,963,283	40,600,404	43,991,851	46,839,889
15	36,747,979	44,564,478	49,600,051	52,694,796	49,040,322	45,350,815	42,465,690	46,380,435
20	29,984,490	37,224,831	45,340,756	50,625,585	53,696,982	49,913,107	46,429,447	43,711,718
25	27,706,134	30,805,800	37,955,164	46,321,783	51,698,474	54,165,472	50,197,433	47,095,284
30	29,415,506	28,230,509	31,481,816	38,813,980	47,310,125	52,577,177	55,541,038	52,342,130
35	31,645,368	29,721,112	28,583,726	31,935,110	39,206,345	47,182,440	52,970,578	56,606,591
40	30,820,270	31,444,607	29,805,756	28,728,957	32,148,626	39,356,065	47,752,030	53,860,328
45	28,351,470	30,494,593	31,202,356	29,707,802	28,647,933	31,751,777	38,794,611	47,205,699
50	25,784,806	27,646,443	30,010,301	30,986,955	29,449,496	28,279,290	31,440,374	38,710,059
55	22,674,388	25,026,733	26,892,978	29,367,252	30,321,055	28,608,590	27,633,141	31,039,156
60	19,835,899	21,635,314	24,077,211	26,175,548	28,155,503	28,982,029	27,644,166	26,927,628
65	16,985,992	18,431,049	20,196,085	22,832,713	24,985,091	26,712,185	27,555,437	26,453,875
70	13,388,911	14,853,761	16,360,007	18,309,306	20,704,192	22,253,102	23,965,739	24,768,431
75	9,135,706	10,685,549	12,052,751	13,584,477	15,672,095	17,665,239	19,186,308	20,712,650
80	5,278,197	6,414,038	7,746,974	8,983,373	10,183,117	11,849,161	13,457,082	14,715,755
85	2,338,233	2,865,646	3,635,884	4,659,723	5,661,301	6,523,176	7,804,965	8,735,081
90	707,960	898,443	1,167,544	1,606,965	2,214,476	2,727,596	3,237,862	3,929,784
95	130,920	174,268	235,314	337,808	506,426	704,332	883,461	1,056,582
100	16,905	21,431	31,006	42,965	66,100	96,589	133,789	162,818
105	2,284	2,568	3,597	4,617	5,887	7,790	10,696	13,633
110+	773	768	865	920	915	832	841	951

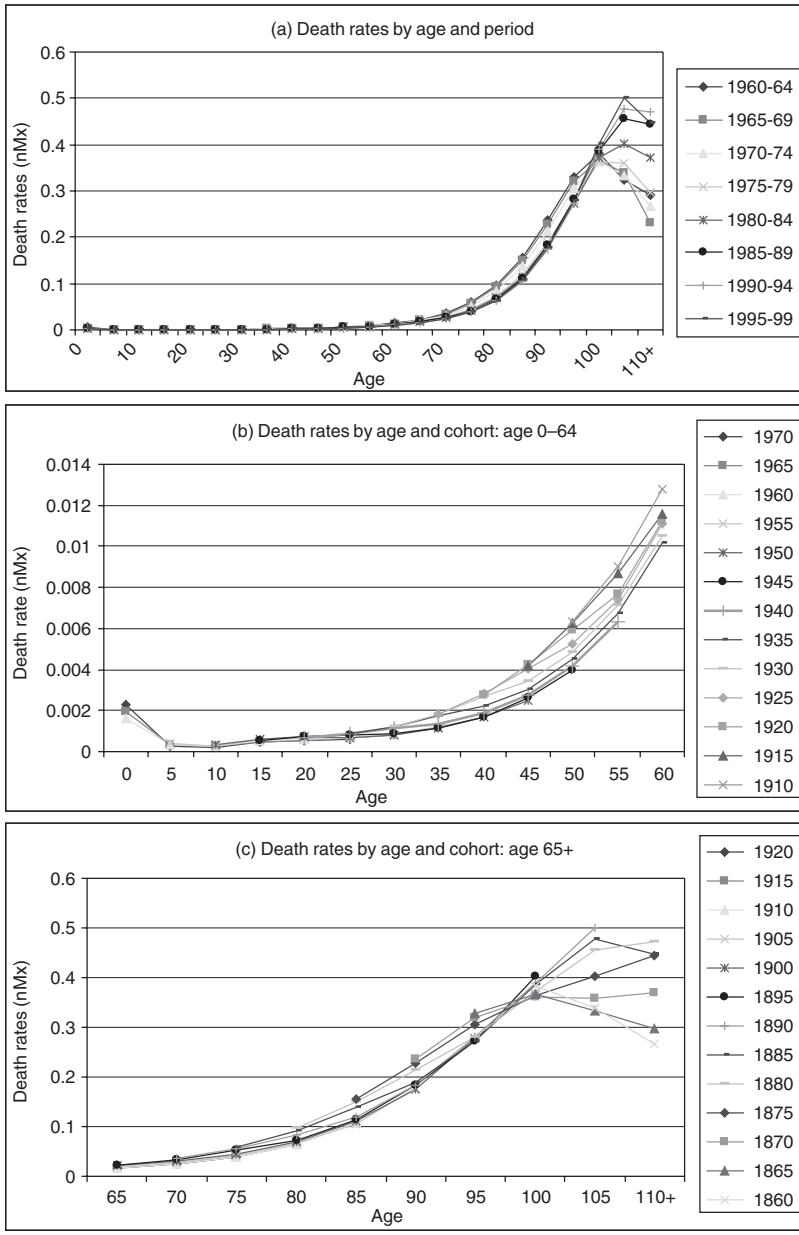

FIGURE 2. Observed death rates for U.S. females: 1960–1999.

due either to age misreporting or to an increase in the proportions of cohort members who survive to the very old ages, thus leading to increasing proportions of old frail individuals in more recent cohort.

Since it is not clear as to what extent the crossover in the mortality patterns are due to age misreporting, and in order not to complicate the APC analyses with interaction effects for the ages 95+ groups, we focus the regression analyses and comparisons reported below on the ages 0 to 94. In this subset of data, there are 19 five-year age groups, 8 time periods with five-year intervals, and 26 cohorts. This yields 152 degrees of freedom.

4.2. *Results: GLIM with Identifying Constraints*

We first estimated generalized log linear models in equation (2) for the female mortality data. Table 3 shows the coefficient estimates and model fit statistics. We commenced by estimating six reduced models: one for each of the three dimensions without controlling for the other two (i.e., the *marginal* or *gross effects models* with estimates given in columns I through III in Table 3), and one for each of the three possible pairs of dimensions (i.e., the two-way models with estimates given in columns IV through VI in Table 3). No linear constraints on coefficients other than the usual normalization are necessary to identify these models.

We then estimated the full three-way APC model where all three dimensions are simultaneously controlled. In order to fit an identified APC model, we considered three models that place an equality restriction on adjacent age, period, and cohort contrasts. In model VII, we equate the contrasts for the first two age groups—namely, 5–9 and 10–14. Demographically, mortality rates are relatively close and low in these early ages. In model VIII, we set the first two period effects to be equal because the data show no substantial period variability before 1970. In model IX, we impose an equality restriction on the coefficients for the two most recent cohorts, as the observed death rates show similar age patterns for these two cohorts. Overdispersion coefficients are estimated for the log-linear models

TABLE 3
Coefficients from Generalized Log-Linear Models of U.S. Female Mortality

	I [A]	II [P]	III [C]	IV [AP]	V [AC]	VI [PC]	VII [A**PC]	VIII [AP**C]	IX [APC**]
Constant	−5.770	−4.838	−1.442	−5.540	−4.315	−1.442	−7.758	−4.515	−8.742
Age									
0–4	*0.000**			*0.000**	*0.000**		*0.000**	*0.000**	*0.000**
5–9	−2.449			−2.453	−2.577		*−2.387***	−2.567	−2.332
10–14	−2.547			−2.548	−2.773		*−2.387***	−2.747	−2.277
15–19	−1.804			−1.794	−2.111		−1.527	−2.068	−1.363
20–24	−1.641			−1.618	−2.023		−1.243	−1.964	−1.024
25–29	−1.495			−1.459	−1.948		−0.974	−1.875	−0.700
30–34	−1.208			−1.166	−1.732		−0.570	−1.651	−0.242
35–39	−0.827			−0.789	−1.420		−0.074	−1.336	0.309
40–44	−0.401			−0.371	−1.056		0.477	−0.965	0.914
45–49	0.048			0.069	−0.684		1.040	−0.582	1.532
50–54	0.487			0.503	−0.324		1.591	−0.212	2.137
55–59	0.900			0.917	0.018		2.124	0.142	2.725
60–64	1.323			1.345	0.383		2.677	0.515	3.334
65–69	1.740			1.769	0.751		3.233	0.891	3.944
70–74	2.185			2.220	1.147		3.819	1.296	4.584
75–79	2.643			2.687	1.557		4.421	1.719	5.242
80–84	3.143			3.195	2.006		5.064	2.181	5.939
85–89	3.638			3.705	2.456		5.707	2.644	6.637
90–94	4.095			4.178	2.872		6.316	3.072	7.300

continued

TABLE 3
Continued

	I [A]	II [P]	III [C]	IV [AP]	V [AC]	VI [PC]	VII [A**PC]	VIII [AP**C]	IX [APC**]
Constant	−5.770	−4.838	−1.442	−5.540	−4.315	−1.442	−7.758	−4.515	−8.742
Period									
1960–64		*0.000**		*0.000**		*0.000**	*0.000**	*0.000**	*0.000**
1965–69		0.005		−0.050		0.383	−0.180	*0.000**	−0.235
1970–74		−0.007		−0.130		0.737	−0.388	−0.028	−0.498
1975–79		−0.068		−0.270		1.028	−0.658	−0.118	−0.822
1980–84		−0.047		−0.329		1.401	−0.844	−0.123	−1.062
1985–89		−0.015		−0.357		1.796	−1.000	−0.099	−1.273
1990–94		−0.021		−0.413		2.156	−1.183	−0.102	−1.511
1995–99		0.010		−0.416		2.548	−1.316	−0.054	−1.698
Cohort									
1870			*0.000**		*0.000**	*0.000**	*0.000**	*0.000**	*0.000**
1875			−0.292		−0.009	−0.413	0.178	−0.002	0.233
1880			−0.616		−0.048	−0.863	0.333	−0.027	0.443
1885			−0.934		−0.114	−1.322	0.475	−0.066	0.639
1890			−1.222		−0.191	−1.775	0.607	−0.113	0.826
1895			−1.475		−0.258	−2.215	0.740	−0.161	1.013
1900			−1.708		−0.330	−2.656	0.862	−0.219	1.190
1905			−1.898		−0.391	−3.086	0.983	−0.278	1.366
1910			−2.200		−0.435	−3.524	1.127	−0.315	1.564
1915			−2.558		−0.486	−3.979	1.264	−0.357	1.756
1920			−2.943		−0.531	−4.422	1.409	−0.393	1.955

1925			−3.321		−0.566	−4.840	1.563	−0.419	2.165
1930			−3.724		−0.629	−5.273	1.690	−0.473	2.346
1935			−4.140		−0.708	−5.709	1.799	−0.544	2.510
1940			−4.590		−0.831	−6.173	1.865	−0.658	2.630
1945			−5.020		−0.947	−6.610	1.937	−0.766	2.757
1950			−5.401		−1.010	−6.998	2.065	−0.818	2.940
1955			−5.687		−1.016	−7.294	2.250	−0.813	3.180
1960			−5.297		−0.999	−6.907	2.453	−0.790	3.437
1965			−5.420		−1.114	−7.136	2.521	−0.902	3.560
1970			−5.532		−1.265	−7.358	2.572	−1.032	3.665
1975			−5.606		−1.426	−7.554	2.655	−1.130	3.802
1980			−5.648		−1.593	−7.728	2.678	−1.287	3.880
1985			−5.615		−1.754	−7.835	2.685	−1.460	3.942
1990			−5.432		−1.933	−7.804	2.690	−1.635	*4.002***
1995			−4.997		−2.124	−7.545	2.635	−1.870	*4.002***
Deviance	877108	95739867	20225299	96431	64385	4186420	17531	17531	17531
DF	133	144	126	108	119	0102	102	102	
Overdispersion	6813.0	1605358.4	200370.2	749.7	598.9	102417.3	171.7	171.7	171.7

*Omitted categories from the models.
**Coefficients constrained to be equal.

using the deviance divided by degrees of freedom through the quasi-likelihood method (McCullagh and Nelder 1989).[4]

The likelihood ratio test statistics reported in Table 4 for each of these nine models show that the three full APC models with the identifying constraints (each of which yields the same deviance) fit the data significantly better than any of the restricted models. Therefore, it can be concluded that none of the three elements of the APC model should be eliminated from the model specification.

Graphs of the estimated coefficients that aid the visual inspection of the temporal trends of the female mortality are shown in Figure 3. These represent the comparisons between the coefficients of successive categories within classifications. Since they indicate changes in mortality from one age group to the next, from one time period to the next, and from one cohort to the next, they actually

TABLE 4
Likelihood Ratio Test of Model Fit Using the CGLIM Estimator: U.S. Female Mortality 1960–1999

Model	Deviance	DF	LR test	(d.f.)	p-value
A	877108	133	859576.7	(31)	<.0000
P	95739867	144	95723085.8	(42)	<.0000
C	20225299	126	20208518.2	(24)	<.0000
AP	96431	126	79650.3	(24)	<.0000
AC	64385	108	47603.5	(6)	<.0000
PC	4186420	119	4169638.6	(17)	<.0000
APC	17531	102			

[4]The estimate of dispersion after fitting, as measured by the deviance or Pearson's chi-square, divided by the degrees of freedom, is greater than 1, indicating the data may be overdispersed. A simple way to model this situation is to allow the variance function of the distribution of the data—namely, Poisson—to have a multiplicative overdispersion factor ϕ. The parameter estimates—say, μ—are not affected by ϕ, but the covariance matrix is multiplied by ϕ—i.e., $V(\mu) = \phi\mu$—and the scaled deviance and log likelihoods used in likelihood ratio tests are divided by ϕ. In SAS GENMOD, the SCALE = option in the MODEL statement specifies a value of $\sigma^2 = \phi$ for the Poisson distributions. We use deviance divided by d.f. as an estimate of ϕ by specifying SCALE = DEVIANCE. An alternative is to use the Pearson's chi-squares. The results are the same for the mortality data. The function obtained by dividing a log-likelihood function for the Poisson distribution by a dispersion parameter is an example of a quasi-likelihood function. Refer to McCullagh and Nelder (1989, chap. 9) for details.

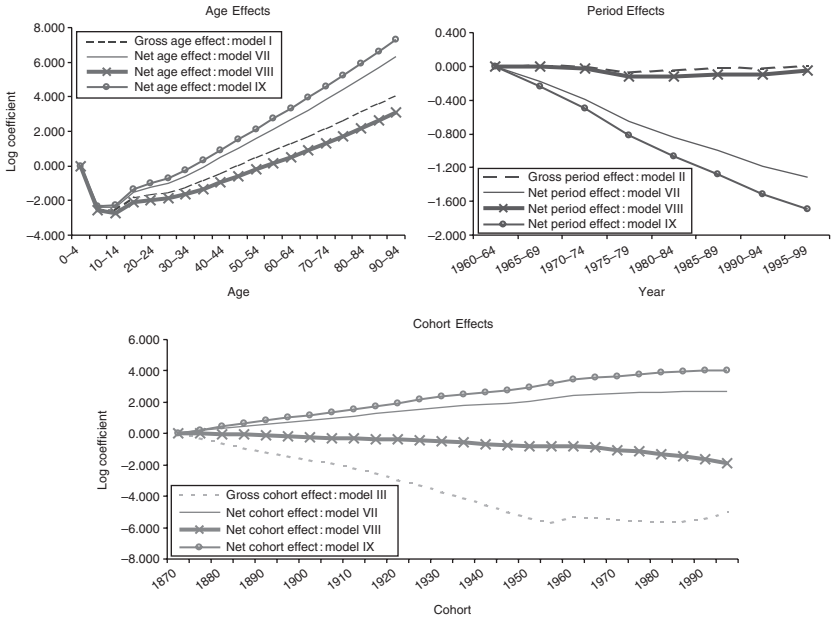

FIGURE 3. Age, period, and cohort effects estimated from the CGLIM with three identifying constraints.

show the trends of mortality along the above dimensions. For each effect, we emphasize comparisons of the coefficient estimates from the gross effect models and the full APC net effect models. The estimated net age effects from models VII, VIII, and IX with different constraints largely follow the same pattern, but indicate different slopes and magnitudes of increases in mortality. Specifically, the two age-effect curves from models VII and IX with age and cohort constraints show steeper and larger increases in mortality than shown by model I for gross age effects, whereas the curve with the period constraint (model VIII) shows less increase.

The deviations of the estimated net period and cohort effects from the gross effects are much more evident for models with different identifying constraints. Here we witness the well-known sensitivity referenced above of the coefficient estimates to the choice of linear restrictions on coefficients. For the period effects, the estimates from the model equating the first two period coefficients (model VII) show a consistent trend with the gross period effects model II, but the estimates from the other two APC model with age and cohort

constraints indicate largely monotone declines in mortality over time. For estimates of cohort effects, we also see drastically different trends in estimated mortality change in successive cohorts. The net cohort effects estimated from model VIII with period constraints show a gradual decline in mortality that is less steep than the gross cohort effect. The imposition of identifying constraints on age (model VII) and cohort coefficients (model IX) produces the reverse trends in the net cohort effects, i.e., the estimates from these two models show increases in mortality across cohorts.

The above results illustrate two essential points in the CGLIM-APC analysis of U.S. female mortality rates. First, it is necessary to simultaneously include all three effects in the model to estimate the net effects, which, more often than not, are different from the gross effects. Second, estimates from the full APC models are sensitive to the choice of equality constraints on the parameters of the model. Specifically, different restrictions that equate coefficients of different subsets of adjacent categories can lead to widely different trend estimates of age, period, and cohort effects, all of which fit the data equally well (see again Table 4).

A natural question arises as to where such a restriction should be placed. The practice of using prior knowledge or external information that was recommended by Mason and Smith (1985) provides a general guideline for model specification and selection. It is this approach that we adopt here. We appeal both to external information and to comparisons of the net and gross effect models to evaluate the three identifying constraints in models VII to IX. We first rule out an equality constraint on the age groups. Age is closely related to physiological state (cf. Strehler and Mildvan 1960) and is the most important source of variation in vital rates. The model fit statistics in Table 4 show that the models including age variables show a more substantial reduction in the deviance than otherwise. Cohort mortality also shows substantial variability, as exemplified in the observed death rates and the gross cohort effect. This is consistent with general conclusions from demographic studies of mortality (Hobcraft, Menken, and Preston 1982), and we therefore infer that we should not place an equality restriction on the cohort effect coefficients. An inspection of the variation of the estimated gross period effect coefficients in Figure 2, however, suggests no large changes before 1970.

Since the mortality rates are close for these two periods within ages and cohorts, there is no strong reason for letting them vary. We therefore prefer an equality restriction on the two initial periods that would result in little loss of information. In other words, we choose model VIII as our final specification.

4.3. *Results: IE Estimates and Comparisons with CGLIM*

We next applied the IE method to analyze the female mortality data. We fit the log-linear APC model (2) to the data. Using a standard statistical software package, *S-Plus* (Venables and Ripley 2000), the IE of model regression coefficients and its standard errors were computed using the algorithms described above. The results for female mortality rates are displayed in Table 5. This model yields a deviance of 17530.5 on 102 degrees of freedom. Overdispersion is estimated for the APC log-linear model with a dispersion coefficient of 171.7. It can be seen that the model fit statistics are exactly the same with those reported for the CGLIM models in Table 4.

How do the coefficient estimates from the IE compare to those from the chosen CGLIM-APC model with identifying constraints on the first two periods, as described above? First, in order to compare the coefficients, note that the IE estimator does not use reference categories for the age, period, and cohort coefficients, whereas the CGLIM estimator does. This means that there are two different parameter centralization methods employed in the two sets of coefficient estimates. The difference between them is a translation by a constant. For example, subtracting the first age parameter estimate from the IE estimates in Table 5 for age effects leads to a reference category on the first age level, which facilitates comparisons of the coefficient estimates. With translations of this type, the two different parameter centralization methods yield similar values of estimated coefficients and trends across the age, period, and cohort dimensions. However, due to the use of reference categories in the CGLIM models, no standard errors are available for those reference categories since they are treated as constants in the CGLIM models. The identifiability problem remains the same for both methods—i.e., both have a singular design matrix though they may differ slightly.

To illustrate further the comparison of the coefficients and the trends across the age, period, and cohort dimensions, Figure 4

TABLE 5
Intrinsic Estimator—Estimated Regression Coefficients and Standard Errors for U.S. Female Mortality Rates[a, b]

Intercept	Age	Effect (s.e.)	Period	Effect (s.e.)	Cohort	Effect (s.e.)
−5.400(.006)	0–4	0.453(.016)	1960–64	−0.039(.008)	1870	1.008(.031)
	5–9	−2.144(.039)	1965–69	−0.009(.007)	1875	0.977(.019)
	10–14	−2.354(.041)	1970–74	−0.007(.006)	1880	0.922(.014)
	15–19	−1.704(.029)	1975–79	−0.067(.006)	1885	0.853(.011)
	20–24	−1.630(.028)	1980–84	−0.043(.006)	1890	0.776(.010)
	25–29	−1.571(.026)	1985–89	0.011(.006)	1895	0.698(.009)
	30–34	−1.377(.023)	1990–94	0.038(.007)	1900	0.611(.008)
	35–39	−1.091(.020)	1995–99	0.116(.007)	1905	0.522(.008)
	40–44	−0.751(.018)			1910	0.455(.008)
	45–49	−0.398(.016)			1915	0.383(.009)
	50–54	−0.057(.014)			1920	0.317(.011)
	55–59	0.266(.012)			1925	0.262(.012)
	60–64	0.610(.010)			1930	0.178(.015)
	65–69	0.956(.009)			1935	0.077(.017)
	70–74	1.331(.008)			1940	−0.067(.020)
	75–79	1.724(.008)			1945	−0.204(.021)
	80–84	2.157(.008)			1950	−0.287(.023)
	85–89	2.590(.009)			1955	−0.312(.025)
	90–94	2.989(.010)			1960	−0.319(.020)
					1965	−0.620(.026)
					1975	−0.748(.030)
					1980	−0.934(.033)
					1985	−1.137(.036)
					1990	−1.342(.039)
					1995	−1.607(.048)

[a] Deviance (d.f.) = 17530.5 (102).
[b] Dispersion coefficient = 171.7.

displays the point estimates of regression coefficients and 95 percent Wald confidence interval estimates from the two models. The estimated trends across ages, periods, and cohorts are strikingly similar for the two models, especially for the age and period effects. The overall trend in the cohort effect is also remarkably similar, but the slopes and ranges of the coefficient estimates differ. The IE estimates show a steeper decline in mortality from slightly above 1 to about −2.0, whereas the CGLIM estimates show less rapid decline and range from around 0 to −2.5.

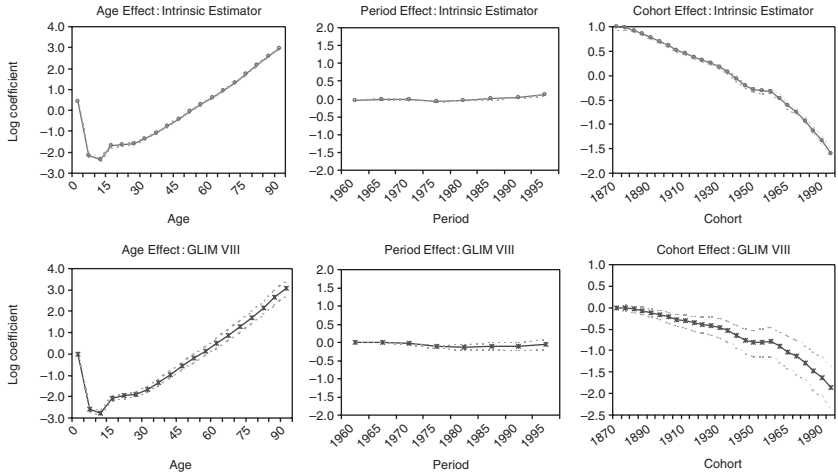

FIGURE 4. Coefficient estimates and 95 percent confidence intervals compared: IE and GLIM VIII.

Two additional points are noteworthy concerning the estimated 95 percent confidence intervals shown in Figure 4. First, it can be seen that the confidence intervals for the CGLIM estimates grow in width across the age, period, and cohort categories, whereas this is not the case for the confidence intervals for the IE estimates. The growth in the confidence intervals for the CGLIM is purely a function of the design matrix and, in particular, the choice of the reference categories. For instance, if instead of choosing the first age, period, and cohort categories as the reference categories for the CGLIM estimates we had chosen the last age, period, and cohort categories, then the confidence intervals would decrease across the age, period, and cohort dimensions—i.e., the confidence intervals would be largest for the first age, period, and cohort categories and decrease in magnitude toward the last categories. Second, it also can be seen that the 95 percent confidence intervals estimated from the CGLIM generally are much wider than those from the IE for most of the age, period, and cohort coefficients, which indicates more variance in the CGLIM estimates. This difference suggests that there may be a difference between the two estimators with respect to relative statistical efficiency, which we explore analytically in the next section.

Overall, these results from both the IE and CGLIM estimators show that the mortality variations for U.S. females are due mainly to

age and cohort effects. The age effects as estimated by both models depict the well-known pattern of mortality change over the life course: high at birth, falling to a minimum around 12, rising to a local peak in the late teens and early twenties, leveling off for the next 10 years or so, and then increasing exponentially after the age of 30 (cf. Wilmoth 1990). While this estimated trend of the mortality age effect coefficients is hardly surprising, the estimated trend for the cohort effect coefficients is revealing. The association between cohort and mortality displayed in Figure 4 is characterized by a gradual and persistent decline in mortality rates for early cohorts until the 1955–1959 birth cohort, a slight increase afterward until the 1965–1969 cohort, and a continuous and more rapid decline in the most recent birth cohorts. This cohort trend is suggestive of a progressive decline in mortality rates in successive cohorts. The estimated rapid declines of mortality in the youngest cohorts may partly reflect progressive improvement in nutrition, health behaviors, and the availability of medical care to the younger generations due to the development of medical research and the increase in women's educational levels that enables them to access healthy ways of living. All these may translate into lower prevalence of disease risk factors and subsequently lower death rates. By comparison, the curves for the period effects indicate relatively less net variation in mortality associated with periods, except for a slight upward turn in the most recent years. The reason for the slow or no improvement of female mortality across the most recent four decades is not clear, but it may partly reflect the influence of increasing cigarette smoking in females (Pampel 2002).

5. BIAS, RELATIVE EFFICIENCY, AND CONSISTENCY OF THE ESTIMATORS

We now discuss some key statistical properties of the CGLIM and IE methods of APC analysis. In general, it is desirable that estimators of coefficients in statistical models be unbiased in finite samples, relatively efficient in finite-time-period analyses in the sense of having a sample variance, or mean squared error (MSE) in the case of biased estimators, that is smaller than that of other estimators, and consistent or asymptotically unbiased in large samples (Casella and Berger 2002).

To identify these properties of the CGLIM and IE methods of APC analysis, consider first the analysis of an APC data set for a finite number of time periods, p. That is, suppose that an APC analysis is to be conducted for a fixed matrix of observed rates or event counts. This implies that the corresponding design matrix X is fixed (i.e., X has a fixed number of age groups and time periods). The randomness in the error term ϵ of model (5) then corresponds to measurement errors in the rates or in the event counts and/or to intrinsic randomness in the rates or counts. It is in this context that the property of estimability of functions of the parameter vector b of equation (6) has been discussed. Specifically, Kupper et al. (1985: 830) derived a condition for the estimability of linear functions of the parameter vector b, namely

$$l^T B_0 = 0, \tag{14}$$

where l^T is a constraint vector (of appropriate dimension) that defines a linear function $l^T b$ of b. Note that, since the IE imposes the constraint that $t = 0$—i.e., that the arbitrary vector B_0 has zero influence, $l^T = (I - B_0 B_0^T)$ by (13). Thus (14) holds for the IE. In other words, the Kupper et al. condition for estimability in the context of a finite-time-period analysis is satisfied. Since estimable functions are unbiased, it follows that the IE B is an unbiased estimator of the special parameterization b_0 of b defined in equation (11). The finite-time-period unbiasedness of the IE can also be proved directly.[5] Specifically, we have the following.

Theorem 1: The Intrinsic Estimator is an unbiased estimator of b_0 in finite-time-period APC analyses of any fixed number p of time periods.

The proof of this theorem is given in Appendix B. Note that this theorem implies that setting $t = 0$ in the representation of the infinite number of estimators of model (5) given above in Equation (10)—namely, $\hat{b} = B + tB_0$, is necessary to produce an estimator of the

[5]For a definition of estimability that includes nonlinear as well as linear functions of the parameter vector (6) and that is specific to APC analysis and a corresponding statement of theorems and proofs, see Fu, Hall, and Rohan (2004).

coefficients of the special parameterization b_0 of the regression vector b that is unbiased in APC analyses of a finite number of time periods. However, this estimator cannot be obtained by the CGLIM methods, as the imposition of identifying constraints in that method almost surely produces a nonzero value of t and thus induces some finite-time-period bias in the CGLIM-class estimators.

Being unbiased in finite-time-period analyses is desirable for a statistical estimator for APC models. In addition, it is preferable that the estimator has a sample variance that, for any finite number p of time periods of data, is small relative to other estimators. The following theorem addresses this property for the IE and CGLIM estimators.

Theorem 2: For any finite number p of time periods, the intrinsic estimator B has a variance smaller than that of any CGLIM estimator—i.e., $\text{var}(\hat{b}) - \text{var}(\text{B})$ is positive-definite for a nontrivial identifying constraint.

This theorem is also proved in Appendix B. Note that this theorem does not establish that the IE is the most statistically efficient possible estimator of the APC accounting model; it proves the more limited efficiency of the IE compared to any possible CGLIM-class estimator. Together, however, the finite-time-period unbiasedness and relative statistical efficiency properties of the IE are quite desirable.

In brief, the IE has nice finite-time-period properties. In addition to these properties, the asymptotic properties of APC estimators have been studied. Although a full discussion of these is beyond the scope of this paper, we cite here some key results that are proved in Fu, Hall, and Rohan (2004). First, the length of the arbitrary eigenvector B_0 decreases with increasing numbers of time periods of data, and in fact converges to zero.

Lemma 1: The eigenvector B_0 uniformly converges to 0 in coordinates as the number of periods increases—i.e., as $p \to \infty$.

Recall that equation (10) states that each of the infinite number of estimators of model (5) can be written as $\hat{b} = B + tB_0$, where the component tB_0 captures the effect of the design matrix on the estimator. For example, we could have two estimators—$\hat{b}_1 = B + t_1 B_0$ and $\hat{b}_2 = B + t_2 B_0$—where t_1 and t_2 correspond to different identifying

constraints placed on model (5). Lemma 1 implies that, as the number of time periods in an APC analysis increases, the difference between these two arbitrary estimators $\hat{b}_1$ and $\hat{b}_2$ decreases toward zero, and, in fact, that the estimators converge toward the intrinsic estimator B. Suffice it to say that the proof proceeds by demonstrating that the coordinates of B_0 are bounded by a quantity that is a function of the number of age groups and periods and the Euclidean norm of B_0 and this function converge to 0 as $p \to \infty$.

Fu, Hall, and Rohan (2004) also show that, under suitable regularity conditions on the error term process and a fixed set of age categories with effect coefficients, the IE will converge asymptotically to these "true" effect coefficients. In addition, Fu and Hall argue that, as the number of time periods of data increases, there are definite bounds on the differences between the effect coefficients estimated by the IE and the effect coefficients of the true period and cohort processes.

Theorem 3: The Intrinsic Estimator B is asymptotically consistent—i.e., it converges in probability to the true parameter vector $\theta(\infty)$ as $p \to \infty$.

Essentially, Theorem 3 is a consequence of Lemma 1. Theorems 1 and 3 and Lemma 1 also have important implications for the asymptotic behavior of CGLIM estimators. We summarize these in the following corollary.

Corollary 1: (a) Arbitrary estimators $\hat{b}$ in Equation (10) of the parameter vector in model (5) obtained by the imposition of arbitrary identifying constraints in CGLIM-class estimators are biased in finite-time-period APC analyses. (b) This bias decreases as the number of time periods in the analysis increases in which case the CGLIM estimators may converge to the IE estimator. (c) However, there are conditions under which the CGLIM estimator, $\hat{b} = B + tB_0$ with $t \neq 0$, may be inconsistent—namely, when t depends on p, the arbitrary term tB_0 may not vanish asymptotically as $p \to \infty$.

Property (a) of this corollary follows from noting that the unbiased property of the IE in finite-time-period analyses (Theorem 1) implies that any other estimator that does not set $t = 0$ must be biased in this context. Property (b) essentially is a consequence of the convergence toward zero of the arbitrary vector B_0 with an increasing number of time periods p of data (Lemma 1). But property (c) states that the

presence of the arbitrary constant t in a CGLIM estimator leads to unpredictable asymptotic behavior of the estimator in those cases wherein t depends on p. This results in the nonrandom term tB_0 not converging to 0 when t is a function of number of periods of data, p.[6]

6. DISCUSSION AND CONCLUSION

The APC multiple classification model is a widely used statistical procedure to estimate the trends in demographic rates along age, period, and cohort dimensions. The linear dependency between age, period, and cohort, however, poses a parameter identification problem. This paper revisited the identification problem by comparing two approaches—namely, the constrained generalized linear models method and the intrinsic estimator method.

The IE has several desirable statistical properties. First, the structure of the estimator guarantees its *estimability*. Through a decomposition of parameter space of the model, a special estimator B can be identified and estimated, either through the projection method or the principle components regression method. This special estimator lies in the subspace that is the orthogonal complement to the null space of matrix X spanned by the eigen-vector B_0 for the unique eigenvalue 0 of X^TX. This special estimator is the intrinsic estimator—obtained by averaging out the information presented by the arbitrary term tB_0 in the multiple estimators. It is a special estimator obtained by principal components regression with the principal components being the eigenvectors of all nonzero eigenvalues. The *asymptotic property* of the intrinsic estimator suggests that as the number of periods increases, the arbitrariness of the multiple possible CGLIM estimators is removed and these estimators converge to B.

We demonstrated through empirical analyses of population data on female mortality rates that the IE yields sensible coefficient estimates that are consistent with those estimated from CGLIM, and also in line with findings from previous epidemiological and demographic studies. We also showed that

[6]The proof of property (c) is given in Fu, Hall, and Rohan (2004).

the IE method produces a unique set of trend estimates that are independent of any arbitrary assignment of identifying constraints on age, period, or cohort coefficients that may not be verifiable in the data itself. And, perhaps most importantly for empirical applications of APC analysis, the IE produces estimated age, period, and cohort coefficients and their standard errors in a direct way, without the necessity of choosing among a large array of possible constraints on coefficients that may or may not be appropriate for a particular analysis.

Our empirical analyses suggested that the standard errors and confidence intervals produced by the IE are smaller than those of the CGLIM estimator. We followed this suggestion with an analysis of the relative statistical efficiency of the two methods and showed that, in addition to being an unbiased estimator for any finite number of time periods of data analyzed, the IE is efficient relative to that of the CGLIM estimator. In brief, the IE method provides a generic solution to APC model estimation without relying on strong priors or information external to the data. In practice, the presence of strong information external to a data set on which to base constraints on the design matrix is rare. In such instances, an informed analyst may well want to impose such constraints (e.g., see Mason and Smith 1985) and make comparisons with the coefficient estimates obtained from an application of the IE. In the more typical case in which strong external information is not available, the IE appears to provide a most useful approach to APC analysis. In either situation, our theoretical results imply that the impact of the identifying constraints on the estimated coefficients will be largest when the number of time periods of APC data is relatively small and that this impact will decrease as the number of time periods in the analysis increases.

Third, the intrinsic estimator method is easy to implement with a standard statistical software package, such as *Splus*. A Web site is being constructed to provide online access for the public to user-friendly computer programs that calculate the intrinsic estimator using APC data. It is currently going through the final stages of testing and security check and will soon be open. Upon completion, it should become an important tool for APC analysis in demography.

APPENDIX A: EXPLICIT FORM OF THE B_0 VECTOR

Kupper et al. (1985) provided a closed-form representation for the eigenvector B_0 of the eigenvalue 0. When the responses are in the order by row,

$$\tilde{B}_0 = (0, A, P, C)^T, \tag{15}$$

where

$$A = \left(1 - \frac{a+1}{2}, \cdots, (a-1) - \frac{a+1}{2}\right)$$
$$P = \left(\frac{p+1}{2} - 1, \cdots, \frac{p+1}{2} - (p-1)\right)$$
$$C = \left(1 - \frac{a+p}{2}, \cdots, (a+p-2) - \frac{a+p}{2}\right)$$

and B_0 is a normalized vector of $\tilde{B}_0$, i.e., $B_0 = \frac{\tilde{B}_0}{\|\tilde{B}_0\|}$. Equation (15) clearly shows that B_0 is independent of the response Y and is completely determined by the number of age groups, a, and the number of periods, p.

This can be illustrated with a specific numerical example. Suppose $a = 3$ and $p = 3$. Then from equation (15) we have

$$A = (1 - \frac{3+1}{2}, 2 - \frac{3+1}{2}) = (-1, 0)$$
$$P = (\frac{3+1}{2} - 1, \frac{3+1}{2} - 2) = (1, 0)$$
$$C = (1 - \frac{3+3}{2}, 2 - \frac{3+3}{2}, 3 - \frac{3+3}{2}, 4 - \frac{3+3}{2}) = (-2, -1, 0, 1)$$

which yields

$$\tilde{B}_0 = (0, -1, 0, 1, 0, -2, -1, 0, 1)^T.$$

It is then easy to compute the vector B_0 as below:

$$B_0 = \frac{\tilde{B}_0}{\|\tilde{B}_0\|} = \frac{\tilde{B}_0}{(\tilde{B}_0^T \tilde{B}_0)^{1/2}}$$
$$= (0, -0.354, 0, 0.354, 0, -0.707, -0.354, 0, 0.354)^T$$

where $(\tilde{B}_0^T \tilde{B}_0)^{1/2} = 8^{1/2}$.

APPENDIX B: PROOFS OF THEOREMS 1 AND 2

Proof of Theorem 1. This proof is for the linear model form of the APC accounting model. The same results apply, however, for generalized linear models through the Fisher information matrix, as the linear part of the models remains the same. The proof begins by citing the principal components regression formulation of the IE noted earlier herein one method of estimation of the IE. Denote by m the number of eigenvalues (including the zero eigenvalue) of the matrix $X^T X$ in the solution to normal equations (7), and denote by r the rank of the $X^T X$ matrix. A general property of the principal components estimator (Sen and Srivastava 1990: 256) in the linear model is that the bias in the principal components estimator of the regression coefficient vector β induced by the deletion of $m-r$ variables in the regression model (corresponding to eigenvalues equal to zero or nearly equal to zero) is $Q_{(m-r)}\beta_{(m-r)}$. In this expression, $Q_{(m-r)}$ denotes a $(m-r)\times(m-r)$ matrix, the columns of which are the eigenvectors corresponding to the $m-r$ eigenvalues that are zero or near zero, and $\beta_{(m-r)}$ denotes the regression coefficients of these eigenvectors. In the present case of the APC regression model (5), $m-r=1$, $Q_{(m-r)} = B_0$, and $\beta_{(m-r)} = t$. Since $t=0$ in the IE, the unbiased property applies here. Furthermore, since an orthonormal linear transformation of an unbiased estimator (which, as described in the text, is used in the principal components approach to estimation to transform the principal components regression estimator to the special parameterization b_0 of the parameter vector b that is estimated by the IE) remains unbiased, it follows that the IE estimate obtained by this algorithm is unbiased.

Proof of Theorem 2. As for Theorem 1, we prove the theorem only for linear models, as the same results can be derived for generalized linear models through the Fisher information matrix. Recall that X denotes the design matrix of the APC multiple classification model (5). Let x denote the row vector with elements -1, 0, or 1 corresponding to the identifying constraint on the parameters in β necessary to achieve a unique estimator. Let W denote the augmented design matrix in which X has been expanded by the constraint vector x, so that we can write $W^T = [X^T | x^T]$. Thus, $W^T W$ is invertible, and, by linear model theory, the variance-covariance matrix for the constraint estimator is $\text{var}(\hat{b}) = \sigma^2 (W^T W)^{-1}$, where σ^2 denotes the variance of the model random error term.

Let Q be the orthonormal matrix such that the matrix $X^T X$ is diagonalized—i.e., $X^T X = Q \Lambda Q^T$, where $\Lambda = diag\ [\lambda_1, \ldots, \lambda_{m-1}, \lambda_m]$ with $\lambda_1, \geq \ldots \geq \lambda_{m-1} > 0$ and $\lambda_m = 0$ being the eigenvalues of matrix $X^T X$. Let $\Lambda_1 = diag\ [\lambda_1, \ldots, \lambda_{m-1}]$. Since the intrinsic estimator is a special principal component estimator with the principal components $\lambda_1 \geq \ldots \geq \lambda_{m-1} > 0$, by linear model theory

$$\text{var}(B) = \sigma^2 Q \Lambda^G Q^T = \sigma^2 Q \begin{pmatrix} \Lambda_1^{-1} & \\ & 0 \end{pmatrix} Q^T \tag{16}$$

where Λ^G denotes the generalized inverse. Note also that $W^T W = X^T X + x^T x = Q(\Lambda + z^T z) Q^T$, where $z = xQ$.

By the principal component decomposition of the CGLIM estimator $\hat{b} = B + B_1$, B and B_1 are orthogonal in the parameter space. Since Λ_1 corresponds to the variance of B, $z^T z$ corresponds to the variance of B_1. Therefore, letting $z^T z = diag[0, \ldots, 0, c]$ with a constant $c > 0$, we have

$$\begin{aligned} \text{var}(\hat{b}) - \text{var}(B) &= \sigma^2 \left[(W^T W)^{-1} - Q \begin{pmatrix} \Lambda^{-1} & \\ & 0 \end{pmatrix} Q^T \right] \\ &= \sigma^2 c^{-1} Q_m Q_m^T \end{aligned} \tag{17}$$

is positive-definite for nontrivial constraint z with $c > 0$, where Q_m is the mth column vector of matrix Q, which is the eigenvector of matrix $X^T X$ with eigenvalue 0.

REFERENCES

Agretsi, Alan. 1996. *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics. New York: Wiley.

Casella, George, and Roger L. Berger. 2002. *Statistical Inference*. 2d ed. Duxbury Advanced Series.

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

Christensen, Ronald. 2002. *Plane Answers to Complex Questions—The Theory of Linear Models*. 3d ed. New York: Springer-Verlag.

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

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

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

Fu, Wenjiang J., Peter Hall and Thomas E. Rohan. 2004. "Age-Period-Cohort Analysis: Structure of Estimators, Estimability, Sensitivity, and Asymptotics." *Journal of the American Statistical Association*, revised and resubmitted.

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

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

Holford, Theodore R. 1983. "The Estimation of Age, Period and Cohort Effects for Vital Rates." *Biometrics* 39:311–24.

———. 1991. "Understanding the Effects of Age, Period, and Cohort on Incidence and Mortality Rates." *Annual Review of Public Health* 12:425–57.

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

Knight, Keith, and Wenjiang J. Fu. 2000. "Asymptotics for Lasso-Type Estimations." *Annals of Statistics* 28:1356–78.

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

Mason, William M., and Herbert 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 William M. Mason and Stephen E. Fienberg. New York: Springer-Verlag.

Mason, W. M. and N. H. Wolfinger. 2002. "Cohort Analysis." Pp. 2189–94 in *International Encyclopedia of the Social and Behavioral Sciences*, edited by Neil J. Smelser and Paul B. Baltes. New York: Elsevier.

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

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2d ed. New York: Chapman and Hall.

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

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

Pampel, Fred C. 2002. "Cigarette Use and the Narrowing Sex Differential in Mortality." *Population and Development Review* 28:77–104.

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

Robertson, Chris, Sara Gandini, and Peter Boyle. 1999. "Age-Period-Cohort Models: A Comparative Study of Available Methdologies." *Journal of Clinical Epidemiology* 52:569–83.

Searle, Shayle R. 1971. *Linear Models*. New York: Wiley.

Sen, Ashish, and Muni Srivastava. 1990. *Regression Analysis: Theory, Methods, and Applications*. New York: Springer-Verlag.

Strehler, Bernard L., and Albert S. Mildvan. 1960. "General Theory of Mortality and Aging." *Science* 132:14–21.

Tarone, Robert, and Kenneth C. Chu. 1992. "Implications of Birth Cohort Patterns in Interpreting Trends in Breast Cancer Rates." *Journal of National Cancer Institute* 1992:1402–10.

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

Venables, W. N., and B.D. Ripley. 2000. *Modern Applied Statistics with S-PLUS*. 3d ed. New York: Springer-Verlag.

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