9

VARIATION IN VITAL RATES BY AGE, PERIOD, AND COHORT

*John R. Wilmoth**

The analysis of age-specific vital rates is studied, and special attention is given to the problem of decomposing an array of rates into factors related to age, period, and cohort. A complete, symmetric decomposition of the data array into age, period, and cohort components is not attempted. Instead, the paper focuses on the age and period dimensions and derives an initial description of the matrix's structure with regard to changes only in those two directions. This two-dimensional description is then augmented by a consideration of residual patterns that seem clearly linked to cohorts. The use of a model that is asymmetric in age, period, and cohort is justified by a detailed discussion of the problems of identification in models involving perfectly collinear independent variables. An important conclusion is that traditional modeling approaches that treat age, period, and cohort in a symmetric fashion are fundamentally flawed.

For their stimulating comments on this article, I thank Andrew Foster, David Lam, Germán Rodríguez, William Mason, and several anonymous reviewers. For her diligence in preparing a difficult manuscript, I also express my sincere gratitude to Judy Mullin. A previous version of this paper was presented at the 1989 Annual Meetings of the Population Association of America in Baltimore. This paper was written with support from an Andrew W. Mellon Postdoctoral Fellowship at the Population Studies Center of the University of Michigan. Most of the research was supported by an NIH traineeship for graduate study at Princeton University, Office of Population Research and Department of Statistics, and by a Fulbright-IIE grant for study at the Università di Roma "La Sapienza," Dipartimento di Scienze Demografiche.

*University of California, Berkeley

Vital rates trace the evolution over age and time of major life events such as birth, death, marriage, divorce, and migration. Although aggregate rates are a less than perfect source of information for studying life's critical transitions, they are nevertheless one of the main windows through which demographers and sociologists view social change. Methods of summarizing changes in aggregate vital rates become, then, a topic of considerable interest.

The research presented here grew out of an analysis of the components of mortality change and, more specifically, out of a desire to answer the question, Do cohorts remember their past mortality experience? In other words, Is current mortality a function of past mortality? These questions necessitated a very careful consideration of the means available for decomposing large arrays of age-specific mortality rates into separate age, period, and cohort components. This paper, while offering a qualified "yes" to the above questions, concentrates on the more general methodological problem of summarizing variation in vital rates by age, period, and cohort.

The message is sobering and yet optimistic. Using models of the form

$$f_{ij} = \alpha_i + \beta_j + \sum_{m=1}^{\rho} \phi_m \gamma_{im} \delta_{jm} + \theta_k + \epsilon_{ij}, \qquad (1)$$

where $k = j - i$, I develop two separate but related arguments. First, I argue that in all legitimate analyses of vital rates, the data-description process must focus on two of the three age, period, or cohort "dimensions." Second, and equally important, I argue that all is not lost as concerns the third and final dimension. I demonstrate, taking age and period to be the primary dimensions in a study of French male mortality for the years 1946–1981, that there are evident sources of variability that can plausibly be associated uniquely with cohorts.

The paper is divided into three major sections. I begin by motivating the choice to use the above model through an exploratory analysis of a large array of French mortality data. In the second and shortest section, I review the literature on the influence of the cohort life history on subsequent mortality and discuss the significance of the current substantive findings. Finally, I examine the issue of age-period-cohort modeling from a theoretical perspective to ground the proposed method in a general and coherent framework.

Since this paper concentrates on the general problems of model development and interpretation, it is imperative to give some consideration to the constraints imposed on the parameters of the model. A brief outline of the fitting algorithm is also given in the appendix. All other details of the fitting procedures, however, are omitted for brevity and are available in a separate technical report (Wilmoth 1989).

1. EXPLORATORY DERIVATION OF THE MODEL

Flexible empirical modeling of arrays of vital rates has several precedents in the work of McNeil and Tukey (1975), Hobcraft and Gilks (1984), and particularly Breckenridge (1983, 1989). The methods used in these cases bear a notable resemblance to the association models (ANOAS) of Goodman (1986), Clogg (1982*b*), and Becker and Clogg (1989). These studies have all made important contributions to the problem of adapting models to empirical reality. Flexible empirical modeling of vital rates has not yet, to my knowledge, been successfully combined with an approach that accounts simultaneously for variation by age, period, and cohort. Most notably, Breckenridge (1983) adopted an exploratory approach toward modeling an age × cohort array of Swedish fertility data and discovered clear period patterns in the residuals, but she did not include a term in the full model to account for this particular form of period-specific variation. In the present study, I sacrifice a degree of flexibility, since I do not consider the rich diversity of robust fitting mechanisms advocated in some of these previous works. However, by restricting my fitting methods to ordinary least squares (OLS), I succeed in combining the most important elements of the flexible modeling approach (i.e., those related to model choice) with a simultaneous consideration of the age, period, and cohort dimensions of variation.

The data examined throughout this paper consist of a rectangular array of age-specific probabilities of death, q_{ij}, for French males aged 0–89 over the years 1946–1981 (Vallin 1973, 1984). Thus, $i = 0, \ldots, 89$ and $j = 1946, \ldots, 1981$, implying that there are $I = 90$ rows (ages) and $J = 36$ columns (periods) in the matrix. The cohorts concerned are those whose life experience intersects this age × period array. Thus, $k = 1857,\ldots,1981$, indicating that $I + J - 1 = 125$ cohorts are included in the study for at least one year. At most, a cohort can be observed for 36 years, as is the case for cohorts 1892–1946. Some cohorts are observed for very few years, such as those just after 1857 or just before 1981, so we can expect to learn very little about their particular mortality experience. An age × period array is chosen mainly for convenience. In addition, we will see that the rows and columns of such an array demonstrate a very high degree of regularity and that we can exploit this regularity of structure by age and period to learn something of the nature of cohort mortality.

It is generally necessary to transform the raw data before proceeding with any form of model fitting. In the present case an appropriate transformation[1] is

$$f_{ij} = \log\left(\frac{q_{ij}}{1-\frac{1}{2}q_{ij}}\right). \tag{2}$$

This transformation has the useful property that

$$f_{ij} \approx \log(\mu_{i+\frac{1}{2},j}), \tag{3}$$

where $\mu_{i+\frac{1}{2},j}$ is the force of mortality (or hazard rate) at the midpoint of the age interval. Thus, the chosen transformation has a theoretical interpretation, since the pieces of an additive model of f_{ij} can be expressed as proportional adjustments to an underlying hazard. As pointed out by Emerson and Stoto (1983), a transformation chosen for one reason often brings with it several serendipitous effects. In this case, beyond the helpful theoretical interpretation, we can cite variance stabilization and increased additivity of the data matrix as two additional arguments in favor of the above transformation (Wilmoth 1989, pp. 2–3).

The exploratory phase of analysis proceeds by fitting a series of simple models to the transformed rates or to the residuals of a previous model. It is sensible to begin by fitting an additive model,

$$f_{ij} = \alpha_i + \beta_j + \epsilon_{ij}, \tag{4}$$

[1]If the available data consist of rates, m_{ij}, instead of probabilities, q_{ij}, the transformation analogous to (2) is $f_{ij} = \log(m_{ij})$.

where the parameters are constrained such that $\Sigma_j \beta_j = 0$. Figure 1 shows the row effects, $\hat{\alpha}_i$, which represent the average age pattern of mortality (in logarithmic scale) over the period of study, 1946–1981. Figure 2 shows the column effects, $\hat{\beta}_j$, which represent the average evolution of mortality over time. The first of these two curves depicts the well-known pattern of mortality change over the age range: high at birth, falling to a minimum around 11, rising to a local peak in the late teens and early twenties, leveling for about 10 years, and then increasing exponentially after around age 30.[2] The second curve documents the very rapid decline in mortality in the period immediately following World War II, a period of relative stability in the 1960s, and then a resumption of the downward trend (albeit at a slower pace) in the 1970s.

This simple model can explain approximately 99.5 percent of the total variance in the matrix. My interest in this analysis, however, is in examining the ways in which an additive model fails to reflect completely the structure of the matrix. Although by this

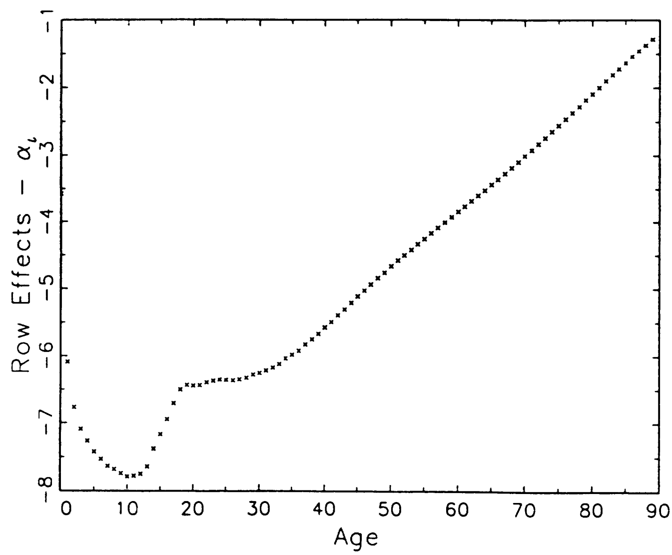


FIGURE 1 Row effects for the additive model. French male mortality, ages 0–89, years 1946–1981.

[2]Obviously, the increase after age 30 in Figure 1 is linear, since the rates have been transformed.

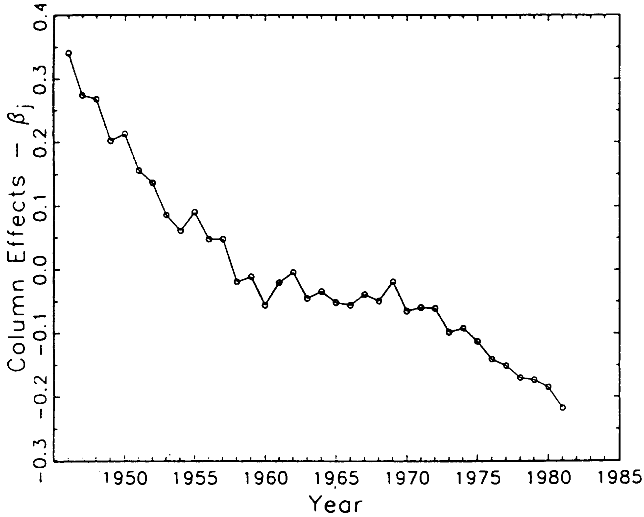


FIGURE 2 Column effects for the additive model. French male mortality, ages 0–89, years 1946–1981.

simple measure one could argue that the matrix demonstrates near perfect additivity and that further analysis is therefore unwarranted, the patterns of nonadditivity in the matrix present considerable substantive interest and should not be ignored. The high percentage of variance explained by a strictly additive model reflects primarily the stark contrasts in mortality across the age range, a pattern whose general contours have changed only slightly since World War II. The nature of those changes, although small relative to the variation in mortality over the age range, may nevertheless have implications for an understanding of the components of mortality change and thus should not be ignored.[3]

Residual analysis for the additive model is quite revealing. Figure 3 shows the average residuals by period for three age groups: 0–4, 15–24, and 50–89. The fit seems particularly deficient for ages 0–4 and 15–24, although the two patterns of deviation are different.

[3]It is in fact the row effects, α_i, which alone account for around 99 percent of the total variance. From this perspective, the column effects, β_j, account for only about 50 percent of the remaining variance and hence of changing mortality patterns over time. The remaining 50 percent needs to be examined subsequently.

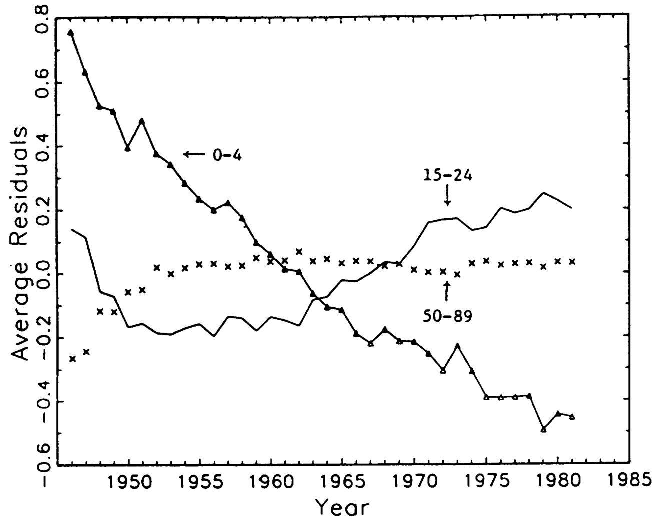


FIGURE 3 Average residuals from the additive model for age groups 0–4, 15–24, and 50–89. French male mortality, ages 0–89, years 1946–1981.

At the youngest ages, the decline in mortality has been much faster than average, resulting in strongly positive residuals for the early years and negative ones for the later years. This tendency is balanced by the evolution of mortality at older ages (50–89 in this case): Across a very broad age range, mortality has declined at a rate that is somewhat less than average. In the case of the 15–24 age group, the decline was somewhat faster than average in the first half-decade, then followed the average pattern of decrease for around 13 years, and finally began a period of *relative* increase compared with the average rate of decline in the period since 1963. In examining the original data, we can see that mortality at these ages actually increased in *absolute* terms during the 1960s and 1970s: The minima over the period 1946–1981 for ${}_5q_{15}$ and ${}_5q_{20}$ were reached in 1963 and 1959, respectively.

To account for these kinds of deviations from additivity, the additive model can be extended by including one or more rank-one multiplicative terms. Defining residuals from the model given in (4) as

$$r_{ij} = f_{ij} - \hat{\alpha}_i - \hat{\beta}_j, \qquad (5)$$

we then fit another model,

$$r_{ij} = \sum_{m=1}^{\rho} \phi_m \, \gamma_{im} \, \delta_{jm} + \epsilon_{ij}. \tag{6}$$

Equivalently, we calculate the first ρ terms of the singular value decomposition (SVD). (See Good [1969] and Gabriel [1971] for useful discussions of the SVD.)

At this stage of analysis, we are faced with the problem of choosing ρ, the proper number of multiplicative terms for the model. The first piece of evidence to be considered is the relative magnitude of the singular values, $\hat{\phi}_m$, since they are a measure of the variance explained by each term of the SVD. From Table 1 it is apparent that the first two terms are dominant, thus suggesting the choice $\rho = 2$.

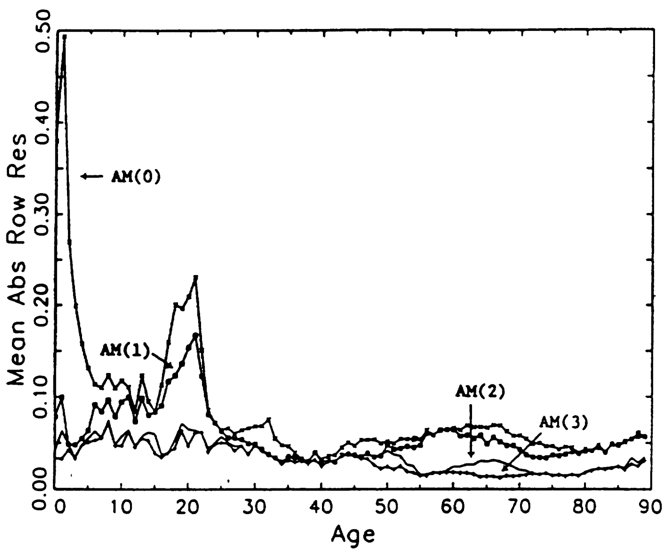
It is possible to confirm that the structure of the matrix by row and column is well represented by an additive and two multiplicative terms by examining successive sets of residuals, formed by removing an additive term and zero, one, two, or three multiplicative terms. (We call such models AM(0), AM(1), etc.) Figures 4a and 4b show the mean absolute residuals by row and by column in these four cases. These graphs demonstrate the successive improvements in fit when one, two, and three multiplicative terms

TABLE 1
Singular Values, $\hat{\phi}_m$, from a Singular Value Decomposition of the Residual Matrix $(f_{ij} - \hat{\alpha}_i - \hat{\beta}_j)$

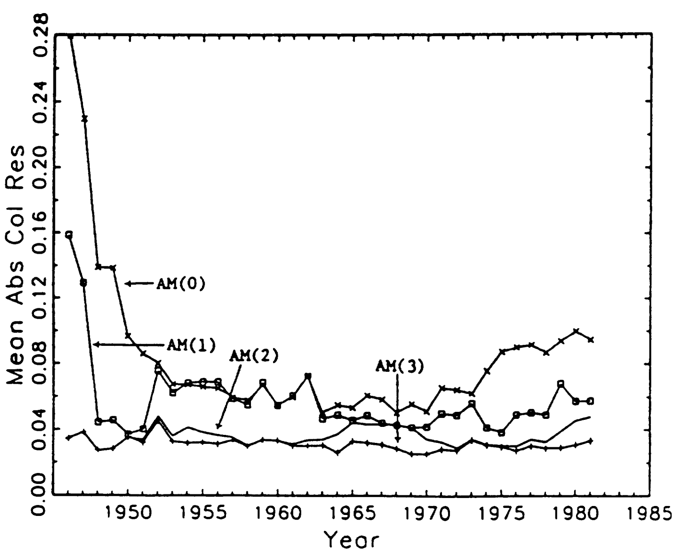
m	$\hat{\phi}_m$
1	6.3
2	3.6
3	1.4
4	1.0
5	0.9
6	0.8
7	0.64
8	0.62
9	0.56
10	0.53

FIGURE 4 Residuals from models AM(0) through AM(3). French male mortality, ages 0–89, years 1946–1981.

a. Mean absolute row residuals by age.



b. Mean absolute column residuals by period.



are added to the additive model. When two such terms are added, the description offered by the model appears to be roughly uniform across all rows and columns of the matrix. The addition of a third multiplicative term, however, results in relatively minor improvements and is thus recommended neither by the previous criterion (see Table 1) nor by the current one.

Accepting the choice $\rho = 2$, then, we may begin to examine the patterns of mortality change that are brought about by these multiplicative terms. Figure 5 shows a four-level contour map representation (Vaupel, Gambill, and Yashin 1985) of the first two multiplicative terms and thus depicts the predominant residual patterns of the simple additive model. Positive areas (shaded dark) mark regions of mostly positive residuals, indicating that the additive model has consistently underestimated the true level of mortality for certain ages and periods. Negative areas (shaded light) correspond to areas of negative residuals and hence to systematic overestimates of mortality.

It is useful to view the joint description of the additive and multiplicative terms in a comparison of the shape of the age curve of mortality for, say, the first and last years of our study, 1946 and 1981. The changing shape of the mortality curve over this period is quite apparent in Figure 6, which combines the additive pattern from Figure 1 with the multiplicative adjustments to this pattern from Figure 5. The two multiplicative terms thus reflect the slow transformation in the shape of the age curve of mortality over the years of the study. In particular, they depict the relatively faster pace of decline at ages under 40 and the slower improvement at higher ages. In addition, the increasing prominence of an "accident hump" around ages 15–25 is quite evident in Figure 6.

The third and fourth terms of the SVD are of much smaller magnitude and seem to depict a different kind of nonadditive structure. As can be seen in Figure 7, these two terms combine to form a series of alternating positive and negative diagonal strips across the matrix, undoubtedly reflecting the influence of cohort factors. These contours can be represented more simply by calculating the average residuals along diagonals after fitting an AM(2) model. Formally, we define

$$s_{ij} = f_{ij} - \hat{\alpha}_i - \hat{\beta}_j - \sum_{m=1}^{2} \hat{\phi}_m \hat{\gamma}_{im} \hat{\delta}_{jm} \qquad (7)$$

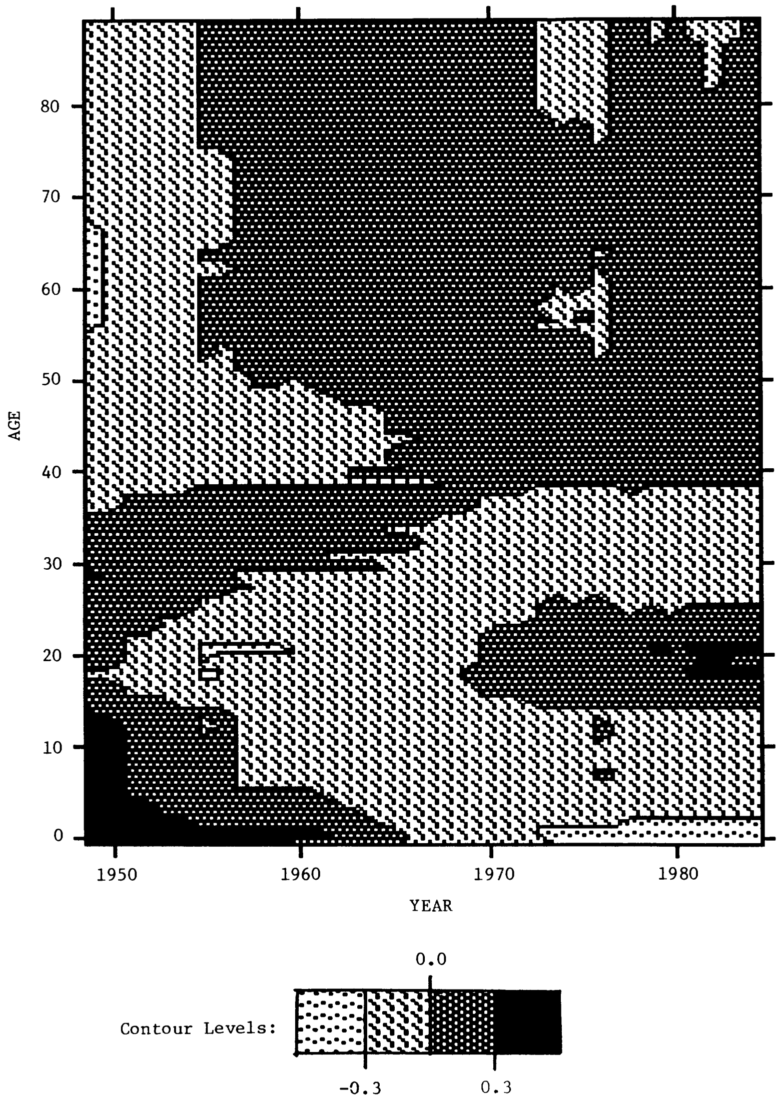


FIGURE 5 Four-level contour map representation of the first two terms of the singular value decomposition, applied to the residuals of an additive model. French male mortality, ages 0–89, years 1946–1981.

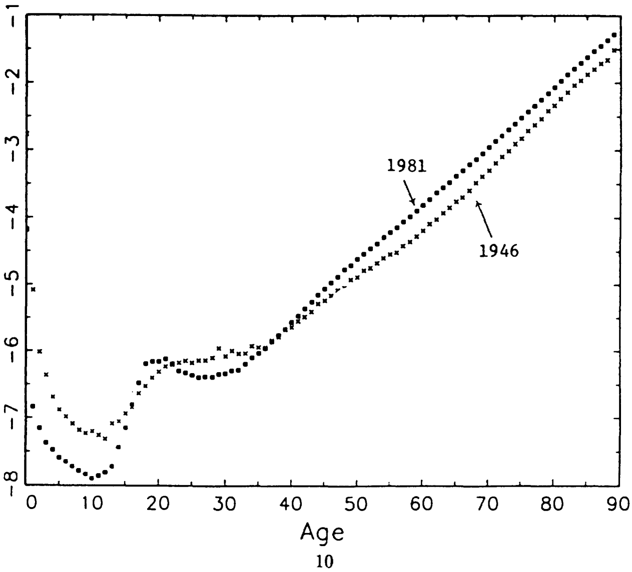


FIGURE 6 Comparison of the shape of predicted age curves for 1946 and 1981, combining additive row effects and multiplicative terms. French male mortality, ages 0–89, years 1946–1981.

and fit a model of the form

$$s_{ij} = \theta_k + \epsilon_{ij}, \tag{8}$$

where $k = j - i$, as before. The resulting estimates, $\hat{\theta}_k$, are called residual diagonal effects and are graphed in Figure 8. The description they offer of the unusual mortality experience of certain cohorts is simpler than, but analogous to, the patterns shown in Figure 7.

The exploratory analysis thus suggests the model referred to in the introduction:

$$f_{ij} = \alpha_i + \beta_j + \sum_{m=1}^{\rho} \phi_m \gamma_{im} \delta_{jm} + \theta_k + \epsilon_{ij}, \tag{1}$$

where i, j, and $k = j - i$ index rows (ages), columns (periods), and diagonals (cohorts), respectively. This is called the AMD model or, when ρ is specified, the AM(ρ)D model. The α_i's and β_j's are the row and column effects, respectively. The α_i's give the shape of the underlying age curve of mortality over the entire period, while the

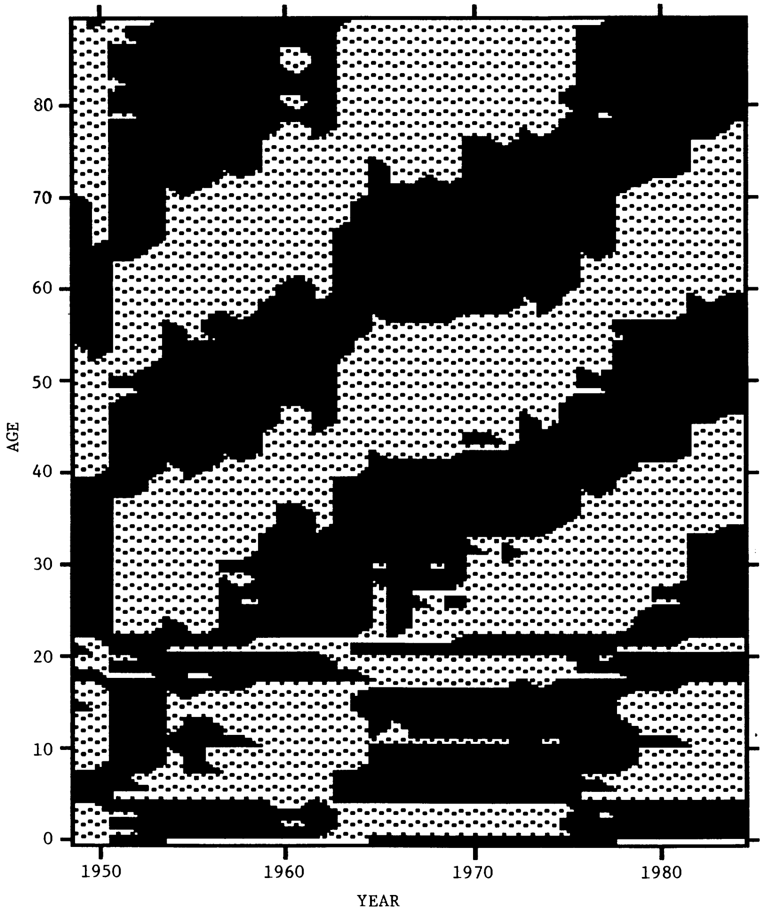


FIGURE 7 Two-level contour map representation of the third and fourth terms of the singular value decomposition, applied to the residuals of a simple additive model. French male mortality, ages 0–89, years 1946–1981 (dark = positive, light = negative).

β_j's indicate the level of the curve for year j. The multiplicative part shows the slow evolution in the shape of the age curve over time. Finally, the diagonal effects, θ_k, depict the average amount of "excess mortality" for cohort k over the period of study.

Since the terms of this model are not independent, it is not optimal to fit successive portions of the model to the residuals of a previous fit, as was done in the exploratory analysis. Rather, it is

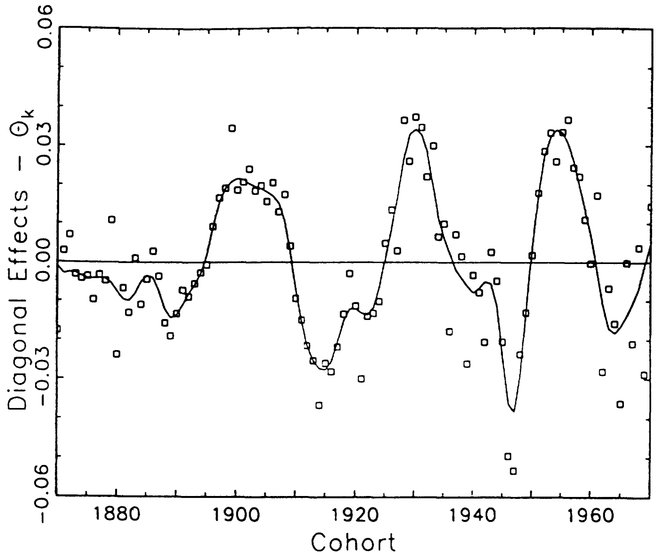


FIGURE 8 Residual diagonal effects (original and smoothed), calculated directly after removal of an additive term and two multiplicative terms. French male mortality, ages 0–89, years 1946–1981.

preferable to fit the model globally, for example, by minimizing the sum of squared residuals for the entire model. It is important for now to limit the fitting procedure to the simplest technique available—i.e., OLS—both to facilitate computation and to center the discussion on more crucial topics, such as the choice of constraints, parameter interpretation, and, eventually, the arrangement of the matrix being analyzed (age × period, etc.). In any case, previous work has indicated that alternative fitting routines, such as weighted least squares or robust methods, are unlikely to substantially alter the results (Wilmoth 1988, pp. 61–64).

The difference between fitting the complete AMD model and the step-by-step exploratory procedure demonstrated above is most evident in the magnitude of the estimates for θ_k. Thus, Figure 9 shows that the residual (i.e., exploratory) diagonal contours recovered in the current example are only about one half the size of those found by an iterative (i.e., complete) solution. This difference is due to the additive part and the first two multiplicative terms, which together consume a portion of the diagonal patterns in the initial steps of the exploratory procedure. Through an iterative

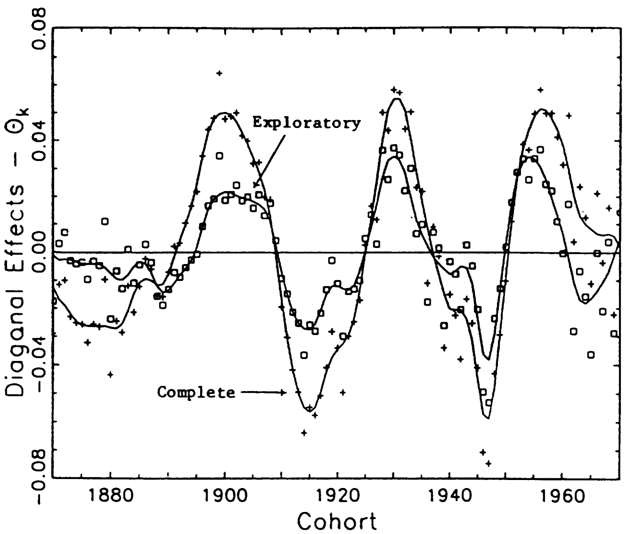


FIGURE 9 Estimated diagonal effects (original and smoothed) from exploratory and complete fits. French male mortality, ages 0–89, years 1946–1981.

procedure that minimizes the sum of squared residuals of (1), the full magnitude of the underlying diagonal contours is recovered. (See section 3 below).

In this way I have motivated by mostly empirical arguments the choice to analyze the array of French male mortality rates using the particular model given in equation (1). I have referred to equation (1) as the AMD model because it contains additive, multiplicative, and diagonal components. This particular analysis was motivated not simply by the regularity of empirical structure thus documented but also by specific substantive issues concerning the factors affecting the long-term mortality experience of cohorts and by general methodological questions regarding the difficulties of analyzing vital rates within an age-period-cohort framework. The next two sections of this paper deal in turn with each of these topics.

2. THE PECULIAR MORTALITY EXPERIENCE OF CERTAIN COHORTS

The effects of the life history on the subsequent mortality experience of cohorts has been a topic of great interest to demographers. The theoretical debate centers around the probable net effects of two countervailing processes: debilitation and selection. The surviving members of a cohort may be debilitated by adverse environmental influences that shift their distribution of frailty, resulting in a higher level of subsequent mortality than would otherwise have been the case. Conversely, these same adverse environmental conditions may select out the relatively weaker members of the cohort through premature death without altering the future susceptibility of the survivors, resulting later on in a lower level of observed aggregate mortality than otherwise expected.

Numerous demographic works have emphasized the debilitative effects of adverse environmental circumstances, which may act to increase mortality in the long term (Kermack, McKendrick, and McKinley 1934*a*,*b*; Coale and Kisker 1986; Preston and van de Walle 1978; Horiuchi 1983; Vallin 1973; Okubo 1981). Others have postulated the potential selective effects of high mortality early in life on the subsequent mortality of a population that is heterogeneous in its susceptibility to death and disease (Vaupel, Manton, and Stallard 1979; Bourbeau and Legare 1981; Levinson 1959; Beard 1961). The particular influence of war on subsequent morbidity and mortality has also been a frequent topic of inquiry in psychological and medical journals (Archibald and Tuddenham 1965; Murphy 1975; Hocking 1970; Maskin 1941; Déthienne and Donnay 1976; Hearst, Newman, and Hulley 1986), where the general consensus has been that war has an unambiguously debilitative effect (both physiologically and psychologically) on survivors.

Some of the demographic studies listed above, particularly those by Horiuchi, Vallin, and Okubo, have also been concerned with the long-term effects of war on the aggregate mortality experience of certain cohorts. These authors documented patterns of unusually high or low levels of mortality in the postwar period for various European countries and Japan. Their methods were generally less sophisticated, less powerful, and in some instances, more difficult to interpret, but they succeeded in documenting many of the same unusual patterns of cohort mortality that are evident in Figure 9 of this paper (see also Wilmoth 1988, pp. 98–113). Still, their descriptions were less complete and, apparently, less accurate. In particular, I have shown elsewhere (Wilmoth 1988; Wilmoth, Vallin, and Caselli 1989) that many of these patterns are present for females as well as males, whereas Horiuchi claimed that they were a peculiarly male phenomenon. Furthermore, the excess mortality of cohorts born in the late 1950s and early 1960s had not been noted before application of the current methods.

The current technique also allows us to quantify more precisely the size of the deviation in the level of cohort mortality. Since, as noted above,

$$f_{ij} \approx \log(\mu_{i+\frac{1}{2},j}), \tag{3}$$

we can interpret the diagonal term of the AMD model as a proportional adjustment to the underlying hazard of death. The estimated adjustment to the cohort force of mortality, μ, at the midpoint of the age interval is

$$\exp(\hat{\theta}_k) \approx 1 + \hat{\theta}_k. \tag{9}$$

Referring back to Figure 9, then, we see that the mortality of the most affected cohorts deviates by approximately 5 or 6 percent from the underlying levels. The peak-to-trough difference is thus on the order of 10 percent or slightly higher.

Unfortunately, a fully adequate interpretation of these patterns in light of processes of debilitation and selection has yet to be found. It had been suggested that excess mortality for groups of male cohorts born around 1900 and 1930 may be related to early combat experiences near the end of the World Wars or to nutritional deprivation in adolescence resulting from the social and economic dislocations at these times (see especially Horiuchi 1983; Wilmoth et al. 1989). These explanations fail to account for the presence of similar patterns of excess mortality among Japanese cohorts, both male and female, born around 1900, since Japan was only nominally involved in World War I (see Wilmoth 1988, pp. 107–13). Furthermore, the presence of similar patterns for females suggests that combat experience in the wars may not be the crucial explanatory variable. Finally, the apparent recurrence of a similar pattern for cohorts born after World War II suggests that the phenomenon may operate through mechanisms at least partially unrelated to the two wars.

Resolution of these problems of interpretation will probably come only through extensive comparative study, through consideration of differences in the form and magnitude of these peculiar patterns of cohort mortality across a variety of countries and causes of death. I do not intend in this paper to resolve this issue. Rather, I intend to outline a method that seems to be a powerful means of documenting the existence of these unusual and strangely persistent patterns. It is worth noting that when this method is applied separately to French male mortality data from two disjoint time periods (1946–1963 and 1964–1981), it recovers a similar description of cohort mortality in both cases, indicating that the peculiarity is indeed persistent across the entire postwar period up to 1981. Furthermore, any concerns that the accuracy of the general description may be due to artifacts of the method or of the data being analyzed may safely be put aside. The method has been thoroughly tested using simulated data to verify its ability to accurately recover patterns of this kind (see section 3), and this accuracy is further corroborated by the findings of authors who used different methods and yet found similar results. That the findings are not a simple artifact of the data is supported by the fact that strikingly similar results have been found for a variety of countries with different systems of vital registration (France, Japan, Austria, the Netherlands, Italy, and Sweden).

The task remains, then, to understand these patterns in light of theories concerning the relations of cohort mortality at different points in the life cycle. This topic will be taken up in future work. The next section considers the relationship between these patterns of cohort mortality and the general subject of age-period-cohort analysis.

3. DATA ANALYSIS WITHIN THE AGE-PERIOD-COHORT FRAMEWORK

It is well known that the analysis of vital rates within an age-period-cohort (APC) framework presents substantial methodological problems relating to the identity

$$\text{cohort} + \text{age} = \text{period}. \qquad (10)$$

The book *Cohort Analysis in Social Research*, edited by Mason and Fienberg (1985), brings together a seminal article by Ryder (1965), a review of the literature by Hobcraft, Menken, and Preston (1982), and several other articles by Mason, Fienberg, and others that provide considerable insight into the general problem of APC analysis. The purpose of this paper is not to provide a comprehensive review of the topic but rather to justify a particular perspective on data analysis within the APC framework and to explicate the relationship between the general problem and the approach offered in the previous sections of this paper.

First, it is notable that my approach to analyzing the given array of French mortality data is strictly descriptive. It is not, however, a blind foray into empirical modeling of data structures; as outlined in the preceding section, it is a description informed by a demand for substantive understanding. The notion that the APC framework can best be thought of as a descriptive accounting scheme was quite effectively argued by Clogg (1982*a*, p. 461), who noted that "because of the difficulties in specifying what it is that the age, period, and cohort variables actually represent in terms of proximate causal agents, an analysis that makes as few assumptions as possible about the causal agents is certainly legitimate." Clogg's particular application dealt with rates of women's labor force participation. His argument that the APC accounting framework is an appropriate first step toward an understanding of the causal processes that have generated an array of vital rates is valid in the present case as well, and it probably holds true in general: One chooses to work within the APC framework precisely because one lacks sufficient knowledge of the relevant explanatory variables that might substitute for the age, period, and cohort proxies.

Yet the notion that the APC framework is merely an accounting scheme through which one may derive an informative description of an array of numbers seems antithetical to the recurrent argument that the identification problem in these cases should be resolved through reference to prior information about the phenomenon being studied. Choosing on the basis of prior information to constrain two age, period, or cohort parameters to be equal in order to achieve model identification, as popularized by Fienberg and Mason (1978) and replicated in one form or another by numerous researchers, implies that the said parameters represent some sort of causal factor about which some *a priori* knowledge is available. If one possesses sufficient information to legitimately make such restrictions, then it must surely be possible to go one step further and, abandoning the accounting framework altogether, define and measure the underlying variables for which age, period, and cohort are mere proxies. I am strictly opposed, therefore, to a hybrid approach that advocates, on the one hand, an accounting procedure (implying a lack of prior information) and, on the other, the imposition of identifying restrictions supposedly based on a refined knowledge of the substantive phenomenon (implying possession of reliable prior information).

There are several pertinent characteristics of descriptive modeling within an APC framework that bear consideration. In brief, the description should be accurate, informative, and parsimonious. Accuracy implies that the residuals of a fitted model should contain no systematic patterns. An informative model is one that lends itself to interpretations related to the causal processes that have in fact yielded the data. Finally, a parsimonious model is one that minimizes the number of fitted parameters without unduly sacrificing accuracy and information.

Notably absent from this list is a concern for parameter bias (cf. Clogg 1982*a*, pp. 465–66). In an accounting scheme, parameter estimates are not compared with an underlying probabilistic model that is assumed to have some real-world validity. The parameter estimates derived in this paper have no direct interpretation other than in relation to the data structures they describe. Alternative choices for the identifying restrictions used in fitting the AMD model in section 1 would not be justified on the basis of a lesser bias or a truer adherence to underlying age, period, and cohort effects; rather, alternative descriptions could be justified, but only on the grounds that they would be more informative, i.e. that they would focus our attention more keenly on data patterns of demographic interest.

To fit any model that contains age, period, and cohort elements to an array of vital rates, one must choose a set of identifying restrictions. In a descriptive analysis of the kind advocated here, one chooses identifying constraints that facilitate the process of informative data description. The constraints used in the previous sections of this paper were a natural by-product of the stepwise fitting procedure used in choosing the model: Each identifying restriction was the result of fitting an additional model to the residuals of a previous model (as described in detail below). They are by no means the only correct constraints. Their justification rests on their utility in highlighting certain interesting features of the data: the predominant additivity of the age and period patterns, the evolutionary change in the shape of the age curve, and the diagonal overlay corresponding to cohorts.

Most of the controversy in APC analysis has revolved around identification of the linear trend in the simple three-way main-effects model. The empirical or theoretical inadequacy of a model that allows for no interaction terms should be obvious (as noted by Glenn 1976, among others); but the inclusion of interaction terms has been thought to so complicate the identification problem that they have generally been avoided (see Fienberg and Mason 1985, p. 71; Clogg 1982*a*, p. 464). Nevertheless, inattention to the relationship between interaction terms and the identification of higher-order patterns (quadratic, cubic, etc.) results in an over-simplification of the problem, as can be readily demonstrated (cf. Fienberg and Mason 1985, pp. 71–74).

The APC identity, equation (10), is of course only a special instance of a more general problem of perfect collinearity. Therefore, I develop the argument for any set of perfectly collinear variables, X, Y, and Z. Suppose that data are collected in the form (w_n, x_n, y_n, z_n), for $n = 1, \ldots, N$. The observations w_n, x_n, y_n, and z_n are realizations of random variables W, X, Y, and Z. Suppose furthermore that a given statistical model assumes that W is the appropriate dependent variable and that X, Y, and Z are explanatory (proxy) variables affecting W. We are interested in a particular situation in which the explanatory variables defy traditional assumptions of uncorrelatedness. Quite to the contrary, in this case they are related by the simple identity

$$X + Y = Z. \qquad (11)$$

This instance of perfect collinearity is clearly analogous to the relationship between age, period, and cohort expressed in equation (10).

Define the main-effects model for W in terms of X, Y, and Z as follows:

$$E[W] = f(X) + g(Y) + h(Z), \qquad (12)$$

where $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$ are functions of variables X, Y, and Z, respectively, and where the expectation operator, $E[\cdot]$, is used merely to avoid the necessity of writing an error term in the equations.[4] In all practical appplications, we can write f, g, and h as polynomial functions. Hence,

$$E[W] = \begin{aligned} & a_0 + a_1X + a_2X^2 + \ldots \\ & b_0 + b_1Y + b_2Y^2 + \ldots \\ & c_0 + c_1Z + c_2Z^2 + \ldots. \end{aligned} \tag{13}$$

It is thus clear that the main-effects model in (12) is not identifiable, since the coefficients pertaining to both the constant and the linear terms of the polynomial expansions of functions f, g, and h cannot be uniquely determined. That the constant terms of f, g, and h are not estimable comes as no surprise and is not unique to the situation in which the explanatory variables are perfectly collinear (being shared, for example, by traditional two-factor ANOVA). The inability to derive unique estimates for the coefficients of the linear term can be demonstrated simply by noting that

$$\begin{aligned} a_1X + b_1Y + c_1Z &= (a_1 + \lambda)\, X + (b_1 + \lambda)\, Y + (c_1 - \lambda)Z \\ &= a_1^*X + b_1^*Y + c_1^*Z \end{aligned} \tag{14}$$

for any real λ. Since either a_1, b_1, and c_1 or a_1^*, b_1^*, and c_1^* reproduce the same probabilistic model, the model implicit in equation (12) is not identified. To estimate the main-effects model, it will be necessary to constrain not only a_0, b_0, and c_0, but also a_1, b_1, and c_1 in some convenient manner.

Quadratic and higher terms in (13) do not present the same problems of estimation: Though it is true that $X + Y = Z$, similar identities do not hold at higher orders. If $X + Y = Z$, it is trivial to show that $X^2 + Y^2 \neq Z^2$, $X^3 + Y^3 \neq Z^3$, etc.; therefore, only the constant and linear terms require identifying constraints. If we consider the main-effects model in which the functions f, g, and h are assumed to have quadratic forms, the relationship between interaction terms and the nonlinear part of the main effects becomes apparent. Since $X + Y = Z$, the quadratic main-effects model,

[4] In the discussion that follows, I assume that X, Y, and Z are fixed, i.e., that they are measured with precision. This assumption is conventional in much of statistical reasoning, and it is certainly not too far from accurate in this case, where X, Y, and Z correspond to ages, periods, and cohorts.

$$E[W] = a_0 + a_1X + a_2X^2 + b_0 + b_1Y + b_2Y^2 + c_0 + c_1Z + c_2Z^2, \tag{15}$$

can be rewritten as

$$\begin{aligned} E[W] = {} & a_0 + a_1X + a_2X^2 + b_0 + b_1Y + b_2Y^2 \\ & + c_0 + c_1(X+Y) + c_2(X+Y)^2 \\ = {} & a_0+b_0+c_0 \\ & + (a_1+c_1)X + (b_1+c_1)Y \\ & + (a_2+c_2)X^2 + (b_2+c_2)Y^2 + 2c_2XY. \end{aligned} \tag{16}$$

Thus, a second-order main-effects model in X, Y, and Z can be rewritten as a second-order model in X and Y alone if we include an XY interaction term. Similarly, an nth-order main-effects model in X, Y, and Z can be rewritten as an nth-order model in X and Y alone if we include all appropriate interaction terms.

Hence, the more general model,

$$E[W] = f(X) + g(Y) + h(Z) + \text{interaction terms in } X, Y, \text{and } Z, \tag{17}$$

requires identifying constraints not only for the constant and linear parts of f, g, and h but also for the quadratic and higher-order terms. It is seen in this manner that there exists a separate identification problem for each order of the polynomial expansions of f, g, and h. I thus define nth-order identification to be the process of constraining polynomial terms of order n, along with the corresponding interaction terms, to produce an estimable model. Note that the main-effects model, equation (12), resolves the second- and higher-order identification problems by setting all interaction terms equal to zero; it retains only the zeroth- and first-order problems.

I have thus demonstrated that the difficulties of modeling posed by perfect collinearity among the explanatory variables can have no purely statistical solution. Furthermore, the problem is not limited to the linear term but affects constant, quadratic, and all higher-order terms as well. It might seem that the only sensible solution to the problem is to abandon one of the three explanatory variables, although the findings of earlier sections of this paper suggest that such a conclusion would be premature. On the one hand, the low-order identification problems may have no reasonable solution other than combining the constant terms and dropping one of the three explanatory variables from the linear, quadratic, and other low-order terms (necessitating, of course, the inclusion of interaction terms). On the other hand, there may be sound theoretical reasons for preferring a description of high-order data fluctuations in terms of simple main effects ascribed to a single dimension, say Z, rather than as complicated combinations of interaction terms in the other dimensions, in this case X and Y. Stated otherwise, as long as the aliasing that occurs between explanatory variables, X, Y, and Z can be expressed in a relatively simple algebraic form (such as occurs at low orders), there may be no substantive grounds for distinguishing between main effects in Z or combinations of main effects and interaction terms in X and Y. As the complexity of the aliasing increases (that is, as we consider high-order terms of polynomial functions f, g, and h), we can argue, on grounds of substantive plausibility, that a representation in terms of main effects in Z is more informative than one involving high-order interaction terms in X and Y. Distinctions between high and low orders in this sense can never be based on conventional statistical criteria. In all instances, we must have recourse to arguments of theoretical plausibility.

This discussion has outlined the critical mathematical properties of models containing perfectly collinear explanatory variables. These arguments lead naturally to a rather pertinent conclusion: Given the inevitable confounding between the main effects and interaction terms in models of this kind, there is little utility in attempting to build descriptions of the data, W, that are *symmetric* in X, Y, and Z. On the contrary, the simplest and most informative descriptions will result from building models that treat X, Y, and Z in an *asymmetric* fashion. More precisely, it is eminently reasonable to choose two primary dimensions of analysis, say X and Y, and to relegate Z to a secondary status. One can then build a model in X and Y, including both main effects and interaction terms, until the resulting description of the dependent variable, W, is complete from the standpoint of those patterns that may at least plausibly be associated with the two primary explanatory variables. At this point it is appropriate to search for residual variation that may be associated with variable Z. If any such residual variation is found, one can then go back and include a term for Z in the original model; at this point it will be necessary to choose appropriate constraints so as not to change the character of description, which should still be focused primarily on X and Y and only secondarily on Z.

Because in this particular case complete arrays of mortality data are most readily available in the age × period format, it was natural first to formulate the description in terms of age and period and then to search for residual variation relating to cohorts. Thus, the main and interaction effects corresponding to age and period capture by far the largest part of the variation in the observed rates. The multiplicative terms are a particular form of age-period interaction that quite successfully depicts the changing shape of the age curve of mortality (see Figure 6). A third set of main effects was added to describe patterns that could be modeled as an age-period interaction (see Figure 7) but are more appropriately analyzed in terms of cohorts (see Figures 8 and 9).

The constraints used in fitting the full model were derived during the exploratory analysis. This initial description of the data array began by extracting a set of row effects, $\hat{\alpha}_i$, which were indexed by age. This meant that in calculating column effects, $\hat{\beta}_j$, from this first set of residuals, the following property held true:[5]

$$\sum_j \hat{\beta}_j = 0. \tag{18}$$

Modeling the residuals of the additive model using the first two terms of the SVD yielded estimates of the multiplicative parameters, $\hat{\phi}_m$, $\hat{\gamma}_{im}$, and $\hat{\delta}_{jm}$, with the following properties:

$$\hat{\phi}_1 > \hat{\phi}_2 > 0, \tag{19}$$

$$\sum_i \hat{\gamma}_{im} = \sum_j \hat{\delta}_{jm} = 0, \tag{20}$$

$$\sum_i \hat{\gamma}_{i1}\hat{\gamma}_{i2} = \sum_j \hat{\delta}_{j1}\hat{\delta}_{j2} = 0, \tag{21}$$

$$\sum_i \hat{\gamma}^2_{im} = \sum_j \hat{\delta}^2_{jm} = 1, \tag{22}$$

[5]These two steps were combined into one in fitting the additive model. Disaggregating this step tells us how the constraints can be derived in the stepwise process of model building. Of course, all the constraints depend on the choice to fit successive portions of the model by the OLS method.

where $m = 1$ and 2. These properties imply that the multiplicative terms are arranged in descending order of importance (equation (19)); that the left and right singular vectors, corresponding to $\hat{\gamma}_{im}$ and $\hat{\delta}_{jm}$, respectively, are centered around zero (equation (20)); and that these latter form orthonormal sets of vectors (equations (21) and (22)). In the final step of the exploratory analysis, a set of diagonal averages was calculated from the residuals remaining after the extraction of an additive and two multiplicative terms. The resulting parameter estimates, $\hat{\theta}_k$, have the following properties:

$$\sum_k w_k \hat{\theta}_k = 0, \tag{23}$$

$$\sum_k k w_k \hat{\theta}_k = 0, \tag{24}$$

where w_k is the number of observations in the kth diagonal. These last two equations indicate that the weighted sum and "slope" of the estimated diagonal effects is zero. More precisely, a weighted least squares regression line (with weights w_k) fit to the estimated values would be identical to the horizontal axis (see Wilmoth 1989, pp. 5–7).

In fitting the entire AMD model, equation (1), these relationships were retained as identifying constraints to avoid altering the particular asymmetric quality of the description; age and period were treated as primary dimensions and cohort as secondary. The iterative routine that was used to minimize the sum of squared residuals consisted of repeated calculations of additive, multiplicative, and diagonal terms, always in that order. At each step of the iteration, the original data were adjusted by removing the previously calculated diagonal term (see the appendix). Still, the constraints given above do not fully identify the model: From the various combinations of parameters that satisfy the OLS criterion, we must choose that solution that also minimizes the equation

$$\sum_k w_k \theta_k^2. \tag{25}$$

Thus, one additional constraint is needed to insure that the diagonal term continues to assume a subordinate position in the fully specified model. The minimization of (25) is achieved by maintaining the original order of fitting (additive, multiplicative, then diagonal) at each step of the iterative algorithm. (See the appendix or Wilmoth

[1989, pp. 7–12] for a more detailed discussion.)

Within the framework of this particular model, we may justifiably inquire about the types of cohort patterns that will be recovered separately in the diagonal term. Recoverability, in this sense, will be distinguished from the more familiar notion of translation. Noting that observed period and cohort patterns are often statistically indistinguishable, Ryder (1964, 1980) defined the process of demographic translation to be the establishment of relationships between these two dimensions of temporal change. Foster (1986), following up on Ryder's work, suggested that period and cohort patterns of an event are likely to be distinguishable only in the presence of high-frequency changes in the period or cohort processes contributing to the observed rates. Thus, in fitting the AMD model, equation (1), to age × period arrays of mortality rates, we can expect to recover discernible diagonal patterns only in the presence of rapid changes in the cohort factors generating those rates. Slow, evolutionary changes in cohort processes will, I predict, be easily translated by the model into combinations of age and period parameters. This translation process is used not to deny the existence or importance of long-term transformations of cohort variables but to show that these changes are, as noted by Ryder, observationally indistinguishable from comparable period movements and, as such, cannot be separated within the context of an APC accounting model.

High-frequency changes in the cohort dimension, on the other hand, can often be recovered separately. It is impossible to define *mathematically* where to draw the line between high-frequency recoverable patterns and low-frequency translatable ones. In *practical* terms, however, a cohort pattern is recoverable (or high frequency) when its translation into age and period patterns lacks a plausible interpretation in terms of age- and period-related variables. Thus, as Figure 7 demonstrates, it is certainly *possible* to translate the diagonal patterns recovered in the study of postwar French male mortality into an age-period interaction (in the form of two additional multiplicative terms). It is scarcely *plausible*, however, to suggest that these patterns are the result of changes in period factors (even in interaction with age) affecting mortality during this period. Instead, it is suggested that the most informative description of these patterns is obtained by relating them to cohorts, as achieved by the addition of a diagonal term to the model.

I have argued that period and cohort influences on age-specific vital rates are observationally indistinguishable except in the presence of rapid changes in one of these two sets of factors. This assertion can be stated in the form of a hypothesis, which we can test by simulating a matrix of size 90×36, consisting of an additive term, a rank-two multiplicative structure, and a diagonal term. (The error term is omitted so that we can concentrate on the underlying structure.) Holding constant the additive and multiplicative inputs to the simulated matrix, we can investigate the success with which a variety of diagonal inputs are recovered. (See the appendix for parameter values for the simulated matrix.) We can make two predictions concerning the recoverability of diagonal patterns in applying the AM(2)D model to the simulated matrices:

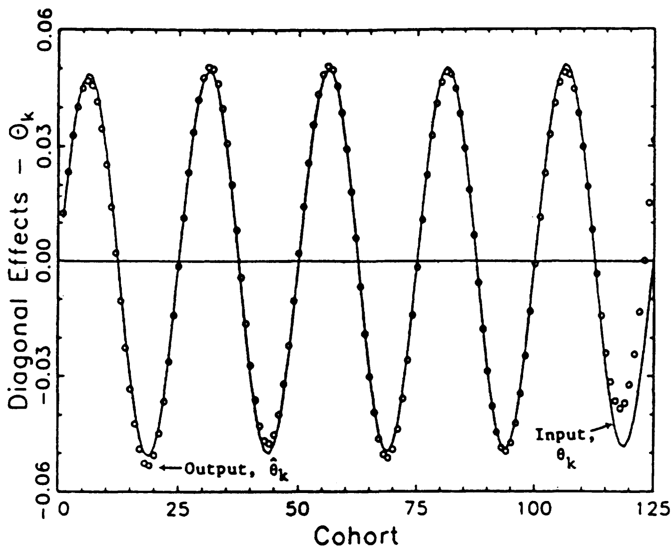
1. If we compare simple diagonal inputs consisting of single-frequency oscillations, relatively high-frequency inputs will be more accurately recovered than relatively low-frequency ones.
2. If an input diagonal pattern is composite, consisting of both low- and high-frequency components, then only the high-frequency part will be well recovered.

Figures 10a, 10b, and 10c present results that are consistent with these predictions. When we compare Figures 10a and 10b, the superior recoverability of high-frequency diagonal patterns is evident. Here, diagonal inputs taking the form of simple sine waves are well recovered when the period of oscillation is relatively short (25 years) and poorly recovered when the wavelength is considerably greater (42 years). That 25-year diagonal cycles are well recovered in simulations is reassuring, since the diagonal patterns discovered in the French example are of approximately that frequency. In this particular case, it appears that the accuracy with which cyclical diagonal inputs are recovered is a function of the length of the cycle relative to the total number of diagonals in the matrix. Thus, cyclical patterns of length equal to around one fifth (or less) of the total number of diagonals are well recovered; those around one third (or greater) are poorly recovered; those around one fourth are recovered well except near the ends (i.e., for the shortest diagonals).

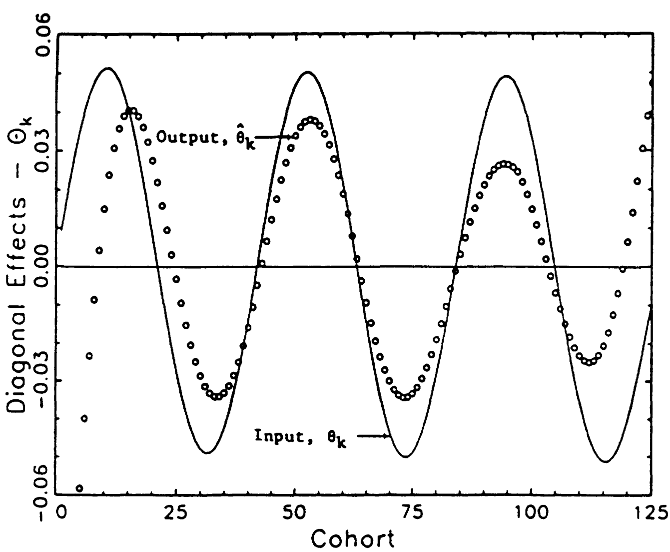
Figure 10c demonstrates the model's ability to separate high- and low-frequency diagonal patterns. From the input diagonal effects, only the high-frequency component was well recovered in the estimated values, $\hat{\theta}_k$. All that was retained in this case were the

FIGURE 10 Examples of the recoverability of diagonal patterns from simulated data.

a. High-frequency input diagonal effects (period = 25).

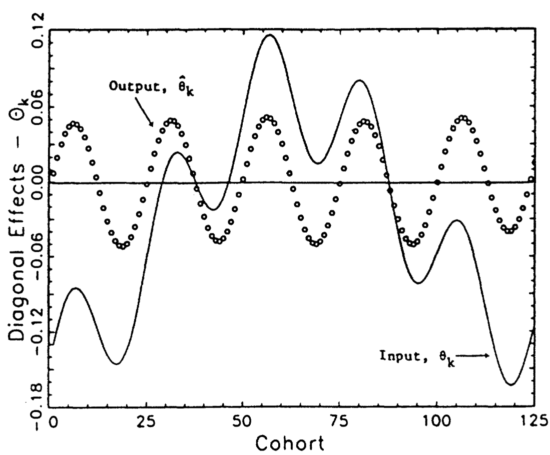


b. Low-frequency input diagonal effects (period = 42).



Continued

c. Composite input diagonal effects containing high- and low-frequency components.



rapid oscillations around a longer trend that first rises and then falls. This example seems to confirm that long swings in period or cohort influences are statistically indistinguishable, especially when the fitted model contains interaction terms. Nevertheless, it does appear possible to separate out the rapid changes in one of these two factors using a model of the kind proposed here.

Further study is needed to determine exactly what kinds of diagonal patterns are recoverable or, conversely, translatable in different circumstances. In particular, it would be useful to investigate the behavior of noncyclical patterns. The proposed model appears capable of serving as a means of investigating these questions in greater depth.

4. DISCUSSION

The methods of analysis advocated in this paper could be questioned or criticized for a number of different reasons. In particular, there may be a few words of dissent related to the difficulties of making causal interpretations of the estimated parameters, the somewhat nebulous distinction between high- and low-frequency patterns, the seemingly arbitrary choice to study an age × period array, the general applicability of the AMD model to events other than mortality, the reliance on traditional and potentially troublesome fitting routines (i.e., OLS), and the perhaps excessive number of parameters used in fitting the model. I will comment on each of these points in order.

First, it is clear that the parameters must be interpreted in light of the model's asymmetry. In the example studied in this paper, age and period are indices for the primary dimensions, whereas cohort is treated as a secondary explanatory variable. Thus, the parameter estimates related to age and period are detailed descriptions of mortality patterns from the perspective of the age and year of death. It can be misleading to refer to the $\hat{\alpha}_i$'s and $\hat{\beta}_j$'s as age and period effects, since they do not isolate causal processes related uniquely to age and period. As descriptions of age and period patterns of mortality, they may contain clues regarding causal influences by age and period. In particular, rapid oscillatory changes in age and period *patterns* undoubtedly point to the causal importance of age- or period-specific *factors*. Ages and periods are, in this sense, no different from cohorts. The age and period parameters in this model, however, contain both high- and low-frequency patterns of mortality change and hence probably reflect the causal influence of factors related to age, period, *and* cohort. The cohort parameters, since they are limited to depicting high-frequency patterns, reflect changes that are more certainly related to cohort membership alone.

Second, the method could be criticized for its implicit dependence on the distinction between high-frequency and low-frequency changes in vital rates (alternatively referred to as high-order versus low-order patterns, short-term versus long-term trends, rapid versus evolutionary changes, etc.). It bears repeating that this distinction, as I have operationalized it, is meaningful only within a substantive context. Low-frequency patterns are those for which a plausible explanation can be proposed relating to more than one of the three age, period, or cohort dimensions of causality. In the present case, the secular trend of mortality decline can reasonably be associated with either period or cohort forces of change, and we are powerless, at least in an examination of a single dataset, to discern what portion of that decline should be attributed to one or the other temporal dimension. That this decline is more rapid at younger than at older ages does not alter this fundamental conclusion, since this differential decrease may only implicate the importance of interaction effects. High-frequency patterns, on the other hand, are observable changes that must almost surely be attributed separately to one or another of the three causal domains. Thus, rapid oscillations in cohort patterns of mortality indicate deviations from an underlying, but probably unrecoverable, trend of cohort-related variables. Similar statements apply for ages or periods as well, but in all cases the distinction between high– and low–frequency patterns can be made only on the basis of substantive plausibility.

Third, since the model is asymmetrical in age, period, and cohort, it is appropriate to question the choice to study an age × period array. In particular, an age × cohort arrangement seems at least as valid on theoretical grounds. (Compare, for example, the contrasting depictions of period and cohort patterns of fertility in Ryder [1980] and Breckenridge [1983]). The comparison between the analysis of age × period arrays of mortality rates and the analysis of age × cohort arrays is in fact a topic of current research, although the problem is considerably more difficult than in the case of fertility, given that the length of the human life span typically exceeds the window of observation generally available in national mortality statistics. We can nevertheless expect numerous similarities: If we describe the structure of mortality first using age and cohort indices, there will still be residual elements of variation related to periods, which will fall along the reverse diagonals of an age × cohort array (see the discussion of Swedish fertility by Breckenridge [1983], in particular the graph on p. 85).

Several advantages to working with age × period arrays naturally present themselves. First, we avoid the problem of incomplete cohort data, although current work is aimed at discovering means of filling in missing data using the EM algorithm. Second, in working with age × period arrays, we can limit the analysis to high-quality postwar data, thus avoiding certain deficiencies in prewar vital registration systems (a problem for almost all countries, including France). Limiting the analysis to the postwar period also increases the number of countries available for detailed comparative study: While for France there are comparable mortality data dating from 1899 (although not always of a quality fully comparable to the postwar data), for other countries such as Japan and Austria, we possess data only for the postwar period. Beyond these practical considerations, there is an empirical argument in favor of the age

× period format: In these French data for the postwar era, there is a marked regularity across periods in the age pattern of mortality.[6] By exploiting this regularity of structure in an initial description by age and period, we succeed quite nicely in isolating patterns that point to the causal importance of cohorts. While at first glance it may appear that the study of cohorts is being slighted, in fact, the particular procedure developed here was motivated primarily by an interest in factors affecting the long-term mortality experience of cohorts, as should be clear from section 2.

Fourth, the applicability of the model to demographic events other than mortality is another subject of current interest. Certain adaptations will probably be necessary to account for the particularities of fertility and marriage patterns over age and time. Again, the work of Breckenridge (1983, 1989) is most relevant and suggests that similar, although probably not identical, models may indeed apply in a variety of circumstances. Even if adaptations of the current methods are called for, the basic principles of analysis will remain the same: We will be no more capable of distinguishing low-frequency cohort or period influences on fertility or marriage than in the case of mortality. Thus, while some details of model specification may change, the fundamentals of the approach will remain constant.

Fifth, it has proved difficult to develop the current methods of analysis outside an OLS framework. Although other researchers have succeeded in performing robust and resistant fits of similar models containing additive and multiplicative terms (McNeil 1974; McNeil and Tukey 1975; Breckenridge 1983), the addition of a diagonal term so complicates the fitting procedure that alternative fits seem to become a practical impossibility. The problem of missing data in analyzing age × cohort arrays of mortality rates, yet to be dealt with adequately in current applications, will undoubtedly complicate the fitting procedure still further and render alternative fits even less practical. Since it is possible to apply robust fitting procedures separately to the various parts of the model (fitting first additive and multiplicative models, then calculating diagonal effects

[6]Foster (1990) has noted a similar regularity in the period age pattern of fertility for several countries, a regularity that is less prominent in the corresponding cohort schedules. Pullum (1980) also pointed to more consistent patterns of period than cohort fertility in the U.S.

from the residuals), I did manage to compare the estimates from the initial exploratory analysis, based on OLS techniques, with analogous results derived using weighted least squares and robust alternative procedures (Wilmoth 1988). These comparisons suggest that there is little benefit to be gained, in this particular application, from a consideration of non-OLS methods. This result is perhaps not surprising given the high quality of the data, but in other applications the OLS technique could prove more troublesome.

Finally, the model suffers to some extent from an overabundance of parameters requiring estimation, and it could reasonably be suggested that the description of the structure of the matrix is exceedingly detailed. Thus, the method can serve at present to illustrate qualitative but not quantitative similarities and differences between comparable arrays. It might be useful, for example, to derive standard age schedules for the additive and multiplicative portions of the model (the α_i's and γ_{im}'s) to facilitate comparisons between sexes or among countries. While such a development might be desirable eventually, it is not essential at this time. I have thus far concentrated on several more immediate tasks (as summarized below) and have purposefully avoided an overly narrow specification of the model.

5. CONCLUSION

Although continued research on several related topics is both possible and desirable, a number of important accomplishments have already been achieved in this paper. First, a detailed description of the structure of one matrix of mortality rates has been given. The general nature of that structure is common to mortality data for both sexes and for several countries and causes of death. The three pieces of the description (corresponding to the additive, multiplicative, and diagonal parts of the model) have separate and informative interpretations. In particular, the diagonal portion of the model effectively isolates cohorts whose mortality experience seems either unusually high or low over an extended period of time. These patterns may help us gain an understanding of the dynamics of cohort mortality, especially as related to the effects of debilitation and selection.

Second, a flexible statistical model that serves as a tool for discovering the structure of an array of vital rates has been proposed. The model reflects an approach to the analysis of vital rates that attempts to separate only the high-frequency components of age, period, and cohort influences on the observed event. While two of these three dimensions are favored in the description of the underlying data structure, it is only in the remaining dimension that a complete isolation of high-frequency patterns is achieved.

Finally, I have discussed the AMD model and its results within the broader context of analysis within the APC framework. I have argued in favor of a purely descriptive approach to APC analysis, although it should be manifest that I intend for the process of description to be informed by a substantive understanding of the problem. The notion that analyses within the APC accounting framework are plagued by a single identification problem, relating to the linear trend of the three sets of effects, has been discarded in favor of a view that recognizes the presence of a range of identification problems. Stated otherwise, it has been shown that all parts, not just the linear one, of the main effects in any one of the three age, period, or cohort dimensions are confounded with the main effects and interaction terms in the other two dimensions. I have contrasted notions of the recoverability and translation of age, period, and cohort patterns. The manner in which high-frequency patterns can be recovered while low-frequency ones are quite simply translated has been illustrated, and the implications of this distinction for analysis within an APC accounting framework have been detailed. It is hoped that the utility of descriptive models isolating high-frequency patterns related to age, period, and cohort has been established and that the futility of similar analyses attempting a complete separation of both high- and low-frequency patterns has been amply demonstrated.

APPENDIX

This appendix contains a minimal description of the algorithm used in fitting the full AMD model, equation (1), and a more detailed description of the matrices used in the simulations exercises. A comprehensive discussion of the technical aspects of the procedure are available in a separate report (Wilmoth 1989).

Fitting Algorithm

The model is fit according to the following iterative procedure:

$$\alpha_i^{(n)} \leftarrow \frac{1}{J}\sum_j (f_{ij} - \theta_k^{(n-1)}), \tag{A-1}$$

$$\beta_j^{(n)} \leftarrow \frac{1}{I}\sum_i (f_{ij} - \alpha_i^{(n)} - \theta_k^{(n-1)}), \tag{A-2}$$

$$(\phi_m^{(n)}, \gamma_{im}^{(n)}, \delta_{jm}^{(n)}) \leftarrow \mathrm{SVD}_\rho(f_{ij} - \alpha_i^{(n)} - \beta_j^{(n)} - \theta_k^{(n-1)}), \tag{A-3}$$

$$\theta_k^{(n)} \leftarrow \frac{1}{w_k}\sum_{(k)} (f_{ij} - \alpha_i^{(n)} - \beta_j^{(n)} - \sum_{m=1}^{\rho} \phi_m^{(n)} \gamma_{im}^{(n)} \delta_{jm}^{(n)}), \tag{A-4}$$

where the third line indicates that the multiplicative terms are derived from the first ρ terms of the SVD of the matrix $(f_{ij} - \alpha_i^{(n)} - \beta_j^{(n)} - \theta_k^{(n-1)})$, w_k is the number of observations in the kth diagonal, and $\Sigma_{(k)}$ denotes summation over all elements within the kth diagonal of the array. For starting values, $\theta_k^{(0)} \equiv 0$ for all k.

As noted in the text, the order of fitting on each iteration is important and should not in general be altered. Iterations stop when differences between successive estimates of θ_k become sufficiently small. For example, one can calculate the weighted average absolute change in $\theta_k^{(n)}$,

$$\frac{\Sigma_k w_k \mid \theta_k^{(n)} - \theta_k^{(n-1)}\mid}{\Sigma_k w_k} = \frac{1}{IJ}\sum_k w_k \mid \theta_k^{(n)} - \theta_k^{(n-1)}\mid, \tag{A-5}$$

and assume convergence when this average becomes less than some predefined level (0.0005 for the matrices analyzed in this paper). It is also useful to establish a maximum number of iterations, such as $n = 25$.

Simulation Exercises

To study the recoverability of diagonal patterns as discussed in the text, I created three hypothetical matrices of size 90×36 having the form

$$\mathbf{F} = \mathbf{A} + \mathbf{M} + \boldsymbol{\theta}, \tag{A-6}$$

where $\mathbf{A} = (\alpha_i + \beta_j)$, $\mathbf{M} = (\Sigma_{m=1}^{2} \phi_m \gamma_{im} \delta_{jm})$, and $\boldsymbol{\theta} = (\theta_k)$. For $\mathbf{M}$, it is required that

$$\sum_i \gamma_{im} = \sum_j \delta_{jm} = 0. \tag{A-7}$$

Conversely, the orthogonality constraints on **M** can safely be ignored, as can the traditional restriction that the β_j's sum to zero, since these affect only the internal decompositions of **A** and **M** while our focus is on $\boldsymbol{\theta}$. We need to pay special attention, however, to the requirement that

$$\sum_k w_k \theta_k = \sum_k k w_k \theta_k = 0. \tag{A-8}$$

We note first of all that estimates for **M** and $\boldsymbol{\theta}$ do not in any way depend on the choice of **A** (see Wilmoth 1988, p. 139). Thus, it is sufficient to concentrate on the choices for **M** and $\boldsymbol{\theta}$ and on the proper manner of comparison between simulated and estimated values.

We begin by defining the parameter sets γ'_{im} and δ'_{jm}, from which γ_{im} and δ_{jm} will be derived:

$$\gamma'_{i1} = i, \qquad i = 1, \ldots, 90 \tag{A-9}$$

$$\gamma'_{i2} = \begin{cases} 0, & i = 1, \ldots, 10 \\ i-10, & i = 11, \ldots, 20 \\ 30-i, & i = 21, \ldots, 30 \\ 0, & i = 31, \ldots, 90 \end{cases} \tag{A-10}$$

$$\delta'_{j1} = j, \qquad j = 91, \ldots, 126 \tag{A-11}$$

and

$$\delta'_{j2} = (j-108.5)^2, \quad j = 91, \ldots, 126. \tag{A-12}$$

All four vectors ($\boldsymbol{\gamma}'_1$, $\boldsymbol{\gamma}'_2$, $\boldsymbol{\delta}'_1$, and $\boldsymbol{\delta}'_2$) are centered (by subtracting out their means) and scaled (by dividing through by their L^2 norms) to obtain $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\delta}_1$, and $\boldsymbol{\delta}_2$. We take $\phi_1 = 6$ and $\phi_2 = 4$, so that they resemble $\hat{\phi}_1$ and $\hat{\phi}_2$ from Table 1. The first multiplicative term thus represents differential rates of mortality decline by age, while the second mirrors a pattern of initial decline and subsequent increase around age 20.

To obtain $\boldsymbol{\theta}$, we first define a diagonal matrix $\boldsymbol{\theta}'$, which does not necessarily conform to the constraints in (A-8). In the first two examples (Figures 10a and 10b), $\boldsymbol{\theta}'$ consists of a single sine curve that completes a cycle of amplitude a every p years:

$$\theta_k' = a \times \sin\left(\frac{2\pi k}{p}\right), \qquad k = 1, \ldots, 125. \qquad (A\text{-}13)$$

In the first example, $a = 0.05$ and $p = 25$; in the second, $a = 0.05$ and $p = 42$. In the final example (Figure 10c), $\boldsymbol{\theta}'$ is a composite of two sine curves:

$$\theta_k' = a_1 \times \sin\left(\frac{2\pi k}{p_1}\right) + a_2 \times \sin\left(\frac{2\pi(k-30)}{p_2}\right), \qquad k = 1, \ldots, 125, \qquad (\text{A-14})$$

where $a_1 = 0.05$, $a_2 = 0.10$, $p_1 = 25$, and $p_2 = 125$.

Since $\boldsymbol{\theta}'$ will not typically adhere to the constraints in (A-8), the θ_k' must be adjusted by fitting a weighted least squares regression line, R_k, and by taking θ_k to be the residuals, $\theta_k' - R_k$. We can write

$$\begin{aligned} R_k &= c + d \times k \\ &= c + d \times (j-i), \end{aligned} \qquad (A\text{-}15)$$

where c and d are some constants. Thus, the R_k can be expressed as a simple sum of functions of i and j. In formula (A-6), we then conveniently take $\mathbf{A} = \mathbf{R} = (R_k)$ and $\boldsymbol{\theta} = \boldsymbol{\theta}' - \mathbf{R}$. After estimation, we thus compare $\hat{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}$, and not with $\boldsymbol{\theta}'$.

REFERENCES

Archibald, H. C., and R. D. Tuddenham. 1965. "Persistent Stress Reactions After Combat." *Archives of General Psychiatry* 12: 475–81.

Beard, R. E. 1961. "A Theory of Mortality Based on Actuarial, Biological, and Medical Considerations." Pp. 611–26 in *International Population Conference of the IUSSP*. Vol. 1. New York: Wright.

Becker, M. P., and C. C. Clogg. 1989. "Analysis of Sets of Two-Way Contingency Tables Using Association Models." *Journal of the American Statistical Association* 84: 142–51.

Bourbeau, R., and J. Legare. 1981. "Introduction d'un processus de deterioration dans l'étude de la mortalité des adultes: Application d'un modèle de simulation aux générations norvégiennes 1866–1916." *Genus* 37: 1–40.

Breckenridge, M. B. 1983. *Age, Time, and Fertility: Applications of Exploratory Data Analysis*. New York: Academic Press.

———. 1989. "Age, Time, and Fertility: Model Building and Model Testing." Unpublished manuscript.

Clogg, C. C. 1982*a*. "Cohort Analysis of Recent Trends in Labor Force Participation." *Demography* 19: 459–80.

———. 1982*b*. "Some Models for the Analysis of Association in Multiway Cross-Classifications Having Ordered Categories." *Journal of the American Statistical Association* 77: 803–15.

Coale, A. J., and E. E. Kisker. 1986. "Mortality Crossovers: Reality or Bad Data?" *Population Studies* 40: 389–401.

Déthienne, F., and J. M. Donnay. 1976. "Variables environnementales et démographiques associées à la morbidité psychiatrique de l'ancien prisonnier de guerre." *Acta Psychiatrica Belgica* 76: 72–89.

Emerson, J. D., and M. A. Stoto. 1983. "Transforming Data." Pp. 97–128 in *Understanding Robust and Exploratory Data Analysis*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey. New York: Wiley.

Fienberg, S. E., and W. M. Mason. 1978. "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data." Pp. 1–65 in *Sociological Methodology 1979*, edited by K. F. Schuessler. San Francisco: Jossey-Bass.

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

Foster, A. 1990. "Cohort Analysis and Demographic Translation: A Comparative Study of Recent Trends in Age-Specific Fertility Rates from Europe and North America." *Population Studies* (forthcoming).

Gabriel, K. R. 1971. "The Biplot Graphic Display of Matrices with Application to Principal Components Analysis." *Biometrika* 58: 453–67.

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

Good, I. J. 1969. "Some Applications of the Singular Decomposition of a Matrix." *Technometrics* 11: 823–31.

Goodman, L. A. 1986. "Some Useful Extensions of the Usual Correspondence Analysis Approach and the Usual Log-Linear Models Approach in the Analysis of Contingency Tables." *International Statistical Review* 54: 243–309.

Hearst, N., T. Newman, and S. Hulley. 1986. "Delayed Effects of the Military Draft on Mortality." *New England Journal of Medicine* 314: 620–24.

Hobcraft, J., and W. Gilks. 1984. "Age, Period, and Cohort Analysis in Mortality Studies." Pp. 245–64 in *Methodologies for the Collection and Analysis of Mortality Data*, edited by J. Vallin, J. Pollard, and L. Heligman. Liège, Belgium: Ordina Editions.

Hobcraft, J., J. Menken, and S. H. Preston. 1982. "Age, Period, and Cohort Effects in Demography: A Review." *Population Index* 48: 4–43.
Hocking, F. 1970. "Extreme Environmental Stress and its Significance for Psychopathology." *American Journal of Psychotherapy* 24: 4–26.
Horiuchi, S. 1983. "The Long-Term Impact of War on Mortality: Old-Age Mortality of First World War Survivors in the Federal Republic of Germany." *United Nations Population Bulletin* 15: 80–92.
Kermack, W. O., A. G. McKendrick, and P. L. McKinlay. 1934*a*. "Death Rates in Great Britain and Sweden: Some General Regularities and their Significance." *Lancet* 226: 698–703.
————. 1934*b*. "Death Rates in Great Britain and Sweden: Expression of Specific Mortality Rates as Products of Two Factors, and Some Consequences Thereof." *Journal of Hygiene* 34: 433–57.
Levinson, L. 1959. "A Theory of Mortality Classes." *Transactions of the Society of Actuaries* 11: 36–96.
Maskin, M. 1941. "Psychodynamic Aspects of the War Neuroses." *Psychiatry* 4: 97–115.
Mason, W. M., and S. E. Fienberg, eds. 1985. *Cohort Analysis in Social Research: Beyond the Identification Problem.* New York: Springer-Verlag.
McNeil, D. R. 1974. "Fitting Models to Two-Way Tables." Technical Report 55, Series 2. Princeton: Princeton University, Department of Statistics.
McNeil, D. R., and J. W. Tukey. 1975. "Higher-Order Diagnosis of Two-Way Tables, Illustrated on Two Sets of Demographic Empirical Distributions." *Biometrics* 31: 487–510.
Murphy, J. M. 1975. "Psychological Responses to War Stress." *Acta Psychiatrica Scandinavica* (supp.) 263: 16–21.
Okubo, M. 1981. "Increase in Mortality of Middle-Aged Males in Japan." Research Paper, Series 3. Tokyo: Nihon University, Population Research Institute.
Preston, S. H., and E. van de Walle. 1978. "Urban French Mortality in the Nineteenth Century." *Population Studies* 32: 275–97.
Pullum, T. W. 1980. "Separating Age, Period, and Cohort Effects in White U.S. Fertility, 1920–1970." *Social Science Research* 9: 255–44.
Ryder, N. B. 1964. "The Process of Demographic Translation." *Demography* 1: 74–82.
————. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* 30: 843–61.
————. 1980. "Components of Temporal Variations in American Fertility." Pp. 15–54 in *Demographic Patterns in Developed Societies*, edited by R. Hiorns. London: Taylor and Francis.
Vallin, J. 1973. *La mortalité par génération en France, depuis 1899.* Travaux et Documents, Cahier No. 63. Paris: Presses Universitaires de France.
————. 1984. *Tables de mortalité du moment et par génération 1899–1981: Mise à jour provisoire des tables annexes du cahier 63.* Paris: INED.
Vaupel, J. W., B. T. Gambill, and A. I. Yashin. 1985. "Contour Maps of Demographic Surfaces." Working Paper 85–47. Laxenburg, Austria: Inter-

national Institute for Applied Systems Analysis.

Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography* 16: 439–54.

Wilmoth, J. R. 1988. "On the Statistical Analysis of Large Arrays of Demographic Rates." Ph.D. diss., Department of Statistics, Princeton University.

———. 1989. "Fitting Three-Way Models to Two-Way Arrays of Demographic Rates." Research Report 89–140. Ann Arbor: University of Michigan, Population Studies Center.

Wilmoth, J. R., J. Vallin, and G. Caselli. 1989. "Quand certaines générations ont une mortalité différente de celle que l'on pourrait attendre." *Population* 44: 335–76.