Modeling the Evolution of Age and Cohort Effects in Social Research*

Sam Schulhofer-Wohl† and Yang Yang‡

March 3, 2009

DRAFT: FOR PRESENTATION AT 2009 PAA MEETING

Abstract

It is well known that the conventional linear age-period-cohort model suffers from an identification problem: If we assume that an outcome of interest depends on the sum of an age effect, a period effect and a cohort effect, then it is impossible to distinguish the separate effects of age, period and cohort because, for any individual, *birth year* = *current year* − *age*. Less well appreciated is that the model also suffers from a conceptual problem: It assumes that the influence of age is the same in all time periods, the influence of present conditions is the same for people of all ages, and cohorts do not change over time. We argue that in many substantive applications of APC analysis, these assumptions fail. We propose a more general model of age, period and cohort effects in which age profiles can change over time; period effects can have different influences on people of different ages; and cohorts can evolve from one period to the next. Our model operationalizes Ryder's (1965) concept of cohort effects as an accumulation of age-by-period interactions. We show that the additive model is a special case of our model and that, except in special cases, the parameters of the more general model are identified. We apply our model to analyze changes in age-specific mortality rates in Sweden over the past two centuries.

*Revised version of a paper presented at the 2008 annual meeting of the American Sociological Association. We are grateful to audiences there and at the University of Cape Town and the Federal Reserve Bank of Minneapolis, as well as to Stephen Raudenbush, for helpful comments. We thank Kristin Plys and Thu Vu for excellent research assistance. Research supported by a pilot grant from the Princeton Center for the Demography of Aging, NIH P30 AG024361.

†Department of Economics, Woodrow Wilson School of Public and International Affairs, and Office of Population Research, Princeton University, and National Bureau of Economic Research. Address: 363 Wallace Hall, Princeton, NJ 08544. Phone: (609) 258-7392. E-mail: sschulho@princeton.edu.

‡Department of Sociology, Population Research Center, and Center on Demography and Economics of Aging at NORC, The University of Chicago. Address: 1126 E. 59th St., Chicago, IL 60637. Phone: (773) 834-1113. E-mail: yangy@uchicago.edu.

1 Introduction

Social scientists conceive of many phenomena as depending on age, period, and cohort (APC) effects. For example:

- In demography, vital rates may depend on a person's age, on environmental conditions in the current year (period), and on conditions early in life that created scarring or selection effects (cohort).
- In sociology, behaviors such as going to college or forming a family may depend on individual physiological and social development (age), on major historical events and social structural changes that individuals encounter in the current year (period), and on formative experiences of groups of individuals coming of age in different historical and social contexts (cohort).
- In economics, consumption inequality among a group of people born in the same year may depend on stages of the life cycle (age), on economic conditions in the current year (period), and on the group's initial level of inequality (cohort).

Despite the analytic importance of age, period and cohort effects, how to empirically distinguish them is among the best-known and longest-standing methodological problems in the social sciences. Age, period, and cohort are linearly dependent; for any person, *birth year* = *current year* − *age*. Therefore, a linear regression model can include at most two of these three variables as regressors. The problem persists even if one specifies age, period, and cohort effects non-parametrically with dummy variables for each possible value, as in the additive model

$$y_{at} = \alpha_a + \beta_t + \gamma_j, \quad j = t - a, \tag{1}$$

where y_{at} is some outcome for people of age a in year t (who are therefore members of birth cohort $j = t - a$); α_a is the effect of being age a; β_t is the effect of living in year t; and γ_j is the effect of being a member of cohort j. The separate effects of age, period, and cohort cannot be distinguished in the additive model (1) because, if (1) is true, then for any constant δ, we also have

$$y_{at} = (\alpha_a + \delta a) + (\beta_t - \delta t) + (\gamma_j + \delta j). \tag{2}$$

That is, age, period, and cohort effects are identified only up to an unknown trend δ.

Even though the additive model (1) is not identified, it has been widely adopted to study age, period, and cohort effects. (Examples date to Greenberg et al., 1950; for reviews, see, e.g., Hobcraft et al., 1982 and Robertson et al., 1999.) Researchers typically solve the identification problem by imposing one or more constraints on the parameters (e.g., Deaton and Paxson, 1994; Mason et al., 1973; Mason and Smith, 1985). But such constraints are often unsatisfying because they must depend on potentially unavailable outside information, on the researcher's subjective preferences, or on purely mathematical (as opposed to substantive) considerations.

We approach the APC identification problem by noticing and then resolving a conceptual problem. The additive model (1) is a quite simple approximation to the process of social change and does not adequately describe most of the phenomena where age, period, and cohort effects are of interest:

- *The additive model specifies that the influence of age is the same in all time periods and for all cohorts.* In fact, however, the influence of age changes over time and across cohorts; consider, for instance, the dramatic declines in infant mortality over the past century (United Nations Demographic Yearbook, 1997).

- *The additive model specifies that the influence of conditions in the present period is the same for people of all ages.* In reality, period effects are often age-specific; for example, the influenza epidemic of 1918 caused especially high mortality among people in their teens and twenties (Noymer and Garenne, 2000).

- *The additive model specifies that cohorts do not change over time.* But cohorts must change, not least because – most obviously in the context of studies of mortality – some members of the cohort die each year, and they are not necessarily identical to those who remain alive (Vaupel et al., 1979).

Cohorts can also change over time for reasons other than composition effects. As Ryder (1965) explained in his seminal article:

> The case for the cohort as a temporal unit in the analysis of social change rests on a set of primitive notions: persons of age a in time t are those who were age $a - 1$ in time $t - 1$; transformations of the social world modify people of different ages in different ways; the effects of these transformations are persistent.

In other words, cohort effects arise because different cohorts live through different social events, or live through the same events at different ages. But because cohort effects result from living through social events, a model with unchanging cohort effects is appropriate only if the relevant events occur before the initial observation and only if these events' impact stays fixed as the cohort ages (Hobcraft et al., 1982). One can model the effect of events experienced at earlier ages by including lagged period effects if these events and conditions affect all age groups similarly. However, if, as Ryder argues, cohorts are continuously exposed to events that affect people of different ages in different ways, one needs a more general model – a framework that (Hobcraft et al., 1982) labeled "continuously accumulating cohort effects." Despite the widespread theoretical influence of Ryder's paper, the concept of continuous cohort change appears never to have been mathematically formalized or taken to data.

We fill this gap by developing a new model of age, period, and cohort effects that can accommodate the various processes of change described above. In our model, age profiles can change over time, period effects can have different influences on people of different ages, and cohorts can evolve from one period to the next. Our model operationalizes Ryder's concept of continuously evolving cohort effects and specifies both age profiles and cohort effects as accumulations of age-by-period interactions. We show that our model nests the additive model as a special case. Apart from a set of measure zero of special cases, however, the parameters of our model are identified, unlike those of the additive model.

Previous researchers, of course, also extended the APC accounting model (1) to include interactions (Fienberg and Mason, 1985; James and Segal, 1982; Moolgavkar et al., 1979). Our model differs from previous models of interactions both substantively and mathematically. Our model allows outcomes to depend on the accumulation of all the events a group of people experiences over the life course, whereas previous models have assumed that only events in the birth year and in the present year are relevant and that the influence of the birth year never changes. Previous models, further, remain unidentified because the additive part can never be identified without additional constraints.

The paper proceeds as follows. In section 2, we describe the model and discuss how to interpret its parameters. In section 3, we analyze conditions under which the parameters are identified when outcomes are measured without error, while section 4 extends the analysis to allow measurement error. Section 5 applies the model to analyze the evolution of human mortality – a fundamentally important phenomenon in demography, and section 6 concludes. Proofs appear in the appendix.

2 Model

We model an outcome y_{at} as an accumulation of age-by-period interactions. Specifically, there are $K \geq 1$ sequences of time effects $\mathbf{e}_1, \ldots, \mathbf{e}_K$, where K is assumed to be known *a priori*. Each sequence $\mathbf{e}_k$ is a list of time effects in various years s: $\mathbf{e}_k = \{e_{k,s}\}_{s=-\infty}^{\infty}$. Time effects that occur in year s affect every cohort alive in that year. However, the impact may depend on the cohort's age and on which sequence contains the time effect: $w_{k,a}e_{k,s}$ is the contribution of time effects from sequence k in year s to the outcomes of people who are age a in year s. We refer to $w_{k,a}$ as the age weight for sequence k at age a. Each sequence of time effects should be thought of as representing a different factor that contributes to the outcome of interest. For example, if the outcome is mortality, one sequence of time effects might represent environmental conditions that affect infant mortality and another might represent medical technologies that affect the mortality of older people.

Past time effects' influence may increase or fall off over time. We let $r_k \geq 0$ be the rate of increase or decay, so $r_k^{t-s} w_{k,a} e_{k,s}$ is the impact in year t of time effects from sequence k occurring in year s for people who were age a in year s. (We leave for future research the problem of modeling time effects whose influence alternates between positive and negative effects or decays in a non-exponential pattern.) We add an intercept and sum up the entire history of time effects to obtain our model for the outcomes for a particular cohort in a particular year:

$$y_{at} = \mu + \sum_{k=1}^{K} \sum_{a'=0}^{a} r_k^{a-a'} w_{k,a'} e_{k,t-a+a'}. \tag{3}$$

We now consider how to interpret the parameters of the model shown in equation (3). For some parameter values, age and cohort effects in our model evolve over time. For other parameter values, our model generates time-invariant age effects, time-invariant cohort effects, and period effects that have the same influence on people of all ages. We first discuss the parameter values that generate these pure effects before showing how other parameter values can produce effects that evolve over time.

- *Pure age effects:* Suppose that, for the kth sequence of time effects, the same time effects occur every year: $e_{k,s} = \bar{e}_k$ for all years s. Then the contribution of this sequence of time effects to outcomes for people of age a in year t is $\sum_{a'=0}^{a} r_k^{a-a'} w_{k,a'} \bar{e}_k$, which depends only on age a, not on the period t or the cohort $j = t - a$.

- *Pure period effects:* Suppose that, for the kth sequence of time effects, $r_k = 0$ and $w_a^k = 1$ for all a. We adopt the convention that $0^0 = 1$. Then the contribution of the kth sequence to outcomes for age a in year t is simply e_t^k, which depends only on the current year and not on age or birth year.

- *Pure cohort effects:* Suppose that, for the kth sequence of time effects, $r_k = 1$, $w_0^k = 1$ and $w_a^k = 0$ for $a > 0$. Then the contribution of the kth sequence to outcomes for age a in year t is simply e_{t-a}^k, which depends only on the birth year $j = t - a$ and not separately on age or the current year.

Because our model can generate pure age, period and cohort effects, it nests the additive model (1). Specifically, suppose that $K = 3$, $e_{1,s} = \bar{e}_1$ for all s, $r_2 = 0$, $w_{2,a} = 1$ for all a, $r_3 = 1$, $w_{3,0} = 1$, and $w_{3,a} = 0$ for $a > 0$. Then (3) reduces to

$$y_{at} = \mu + \sum_{a'=0}^{a} r_1^{a-a'} w_{1,a'} \bar{e}_1 + e_{2,t} + e_{3,j}, \quad j = t - a, \tag{4}$$

which is equivalent to (1) with $\alpha_a = \mu + \sum_{a'=0}^{a} r_1^{a-a'} w_{1,a'} \bar{e}_1$, $\beta_t = e_{2,t}$ and $\gamma_j = e_{3,j}$.

Researchers are often interested in estimating an age profile of outcomes. For example, how does the mortality rate depend on age? Or, how does within-cohort consumption inequality change as the cohort ages? A major goal of our model is to make it possible to estimate such age profiles while allowing the possibility that the age profile is not the same for all cohorts. In our model, we conceive of changes in the age profile over historical time as changes in the time effects that accumulate for different cohorts. Permanent changes in time effects lead to permanent changes in the age profile. Define $m_k(a) = \sum_{a'=0}^{a} r_k^{a-a'} w_{k,a'}$. Then a hypothetical cohort that experienced the same time effects $(\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_K)$ in every year of its life would, at age a, have outcomes

$$y_a(\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_K) = \mu + \sum_{k=1}^{K} \bar{e}_k m_k(a), \tag{5}$$

which depends only on the cohort's age a. Now consider a different hypothetical cohort that experienced a different set of constant time effects $(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_K)$ in every year of its life. The second cohort would have outcomes

$$y_a(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_K) = \mu + \sum_{k=1}^{K} \tilde{e}_k m_k(a). \tag{6}$$

The outcomes in (6) again depend only on the cohort's age, but they differ from the outcomes of the first hypothetical cohort in (5). Thus, given any set of time effects, we can calculate the hypothetical age profile that would result if those time effects continued for the entire life of a cohort. So, for example, we can calculate different age profiles corresponding to the time effects of 1900 and the time effects of 2000:

$$\begin{aligned} y_a(e_{1,1900}, e_{2,1900}, \ldots, e_{K,1900}) &= \mu + \sum_{k=1}^{K} e_{k,1900} m_k(a), \\ y_a(e_{1,2000}, e_{2,2000}, \ldots, e_{K,2000}) &= \mu + \sum_{k=1}^{K} e_{k,2000} m_k(a). \end{aligned} \tag{7}$$

The profile $y_a(e_{1,1900}, e_{2,1900}, \ldots, e_{K,1900})$ tells us the effect of age on outcome y in 1900. We can interpret $y_a(e_{1,1900}, e_{2,1900}, \ldots, e_{K,1900})$ as a prediction for the outcomes of the 1900 birth cohort if conditions never changed after the cohort's birth. In other words, $y_a(e_{1,1900}, e_{2,1900}, \ldots, e_{K,1900})$ describes the effect of age on outcomes y, holding time effects constant. Similarly, the profile $y_a(e_{1,2000}, e_{2,2000}, \ldots, e_{K,2000})$ tells us the effect of age on outcome y in 2000. By comparing the profiles, we can see how the effect of age on y changed over the 20th century.

3 Identification

We have claimed that one advantage of our model over the additive model (1) is that the parameters of our model are identified. We now make this claim precise. Because the additive model is unidentified even when (1) does not contain an error term, we assume for now that the outcomes y_{at} are measured without error; in section 4, we show how to handle measurement error. We consider identification when the data consist of an $(A + 1) \times T$ matrix of outcomes for ages $a = 0, \ldots, A$ and dates $t = 1, \ldots, T$, such as a table of age-specific mortality rates in various years. We leave extensions to other data structures for future research.

We say the parameters of our model are identified if there exists a unique set of parameters that can generate any given matrix of outcomes y_{at}. That is, the parameters are identified if there is a unique vector $\boldsymbol{\theta} = \left[\mu, \left\{\{e_{k,t}\}_{t=1-A}^{T}, \{w_{k,a}\}_{a=1}^{A}, r_k\right\}_{k=1}^{K}\right]$ such that (3) holds for all a and t. It turns out that our model is identified for some values of the true parameters and not for other values. The following definition is therefore helpful:

Definition. *The parameter vector $\boldsymbol{\theta}$, an element of a parameter space $\boldsymbol{\Theta}$, is identified with respect to $\boldsymbol{\Theta}$ if there does not exist any vector $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ distinct from $\boldsymbol{\theta}$ such that,* *given* $\left\{\{y_{at}\}_{a=0}^{A}\right\}_{t=1}^{T}$, (3) *holds for both* $\boldsymbol{\theta}$ *and* $\tilde{\boldsymbol{\theta}}$ *for all* $a = 0, \ldots, A$ *and* $t = 1, \ldots, T$.

Under normalizations on the parameter space $\boldsymbol{\Theta}$ that do not affect the interpretation of the model, the set of parameter vectors that are not identified with respect to $\boldsymbol{\Theta}$ is of measure zero. The normalizations are:

Normalization 1. $r_k \leq r_{k'}$ *for all* $k < k'$.

Normalization 2. $w_{k,0} = 1$ *for all* k.

Normalization 3. *If* $K > 1$, *then* $e_{k,s} = 0$ *for* $s < k - A$.

Normalization 1 puts the time effect types in order, which is necessary because switching k with k' would not change the model. (We show below that the unidentified set of measure zero includes the case $r_k = r_{k'}$, so the ordering is strict.) Normalization 2 fixes the sign and scale of the age weights $w_{k,a}$ and the time effects $e_{k,s}$; for any $c_k \neq 0$, replacing $w_{k,a}$ by $c_k w_{k,a}$ for all a and $e_{k,s}$ by $e_{k,s}/c_k$ for all s would not change the model. The normalization does not affect the interpretation of results since only the product $w_{k,a}e_{k,s}$ enters the age profiles (5). Finally, we need normalization 3 because the data do not contain adequate information about time effects in the distant past. The normalization is equivalent to dropping all data on the K oldest cohorts. To see why, notice that time effects $e_{k,s}$ at any date $s \leq K - A$ influence only the K oldest cohorts; that there are K^2 such time effects e_s^k in the model; and that we have $K(K+1)/2 \leq K^2$ observations (with strict inequality for $K > 1$) on the K oldest cohorts. We therefore have no hope of identifying all the time effects at dates $s \leq K - A$. In addition, by appropriately choosing $\{e_{k,s}\}_{s \leq k-A}$, we can perfectly fit the data on the K oldest cohorts regardless of how we choose $\mathbf{r}$, $\mathbf{w}$, μ and $\{e_{k,s}\}_{s > K-A}$. Since the K oldest cohorts are uninformative, we could drop them and avoid estimating $\{e_{k,s}\}_{s \leq k-A}$. Equivalently, we can normalize some elements of

$\{e_{k,s}\}_{s\leq k-A}$ to zero. Since the normalization does not affect $\mathbf{r}, \mathbf{w}, \mu, \{e_{k,s}\}_{s>K-A}$, it does not affect the substantive results.

Proposition 1. *Let $K \in \{1,2,3\}$ be known, and let the parameter space $\mathbf{\Theta}$ consist of all vectors $\left[\mu, \left\{\{e_{k,t}\}_{t=1-A}^{T}, \{w_{k,a}\}_{a=0}^{A}, r_k\right\}_{k=1}^{K}\right]$ that satisfy normalizations 1 to 3. Suppose further that $A \geq K$, that $T \geq A+K$, and that if $K=1$, then $T \geq 4$; if $K=2$, then $T \geq 12$; and if $K=3$, then $T \geq 32$. Then there exists a set $X_K \subset \mathbf{\Theta}$ such that X_K is of measure zero and all $\boldsymbol{\theta} \in \mathbf{\Theta} \setminus X_K$ are identified.*

Proposition 1 says there may be parameter vectors $\boldsymbol{\theta}$ that are not identified: For each of these $\boldsymbol{\theta}$, there exists some $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ that would generate the same data as $\boldsymbol{\theta}$. However, the set X of unidentified parameter vectors is of measure zero. For almost all $\boldsymbol{\theta}$, therefore, there does not exist any $\boldsymbol{\theta}' \neq \boldsymbol{\theta}$ that would generate the same data, and by observing y_{at}, we can uniquely determine the true parameter vector $\boldsymbol{\theta}$. We have not proved versions of proposition 1 for $K > 3$, but we conjecture that it holds; a proof would require tedious algebra.

The conditions in proposition 1 are sufficient but not necessary for identification. In particular, the parameters may be identified for T smaller than the values stated, so long as A is sufficiently large. We have not completely characterized the sets X_K of unidentified parameter vectors. In one sense, this is unimportant since almost all parameter vectors lie outside X_K. However, to understand the source of identification, it is helpful to partially characterize X_K. The next proposition gives some necessary conditions for a parameter vector to be identified.

Proposition 2. *Under the hypotheses of proposition 1, any parameter vector $\boldsymbol{\theta} \in \mathbf{\Theta}$ is not identified if either:*

(a) $e_{k,t} = \bar{e}_k$ *for some* k *and all* $t = 1-A, \ldots, T$, *or*

(b) $K > 1$ *and* $r_k = r_{k'}$ *for some* $k \neq k'$.

Further, $\boldsymbol{\theta}$ remains unidentified in each of these cases even if μ is known.

Condition (a) in proposition 2 is the case where the model contains pure age effects. Therefore, although the additive model (1) is a special case of our model, it is an unidentified special case. We emphasize that the potential need to identify the intercept μ has nothing to do with this failure of identification. It is clear that pure age, period or cohort effects will be unidentified in our model without some normalization on μ for the usual reason that – even without the APC identification problem – one dummy variable in any given category must be omitted in any linear model that contains an intercept. But proposition 2 shows that pure age effects will remain unidentified even with a normalization on μ. The intuition is as follows. Suppose the same time effect happens over and over, i.e., $e_{k,t} = \bar{e}_k$. Then it will be impossible to distinguish whether this time effect has a transitory impact that directly affects people of all ages (a period effect) or a persistent impact that directly affects only the young (so that the effect on the old is indirect, a cohort effect). Pure age effects, in other words, make it impossible to distinguish period from cohort.

4 Identification With Measurement Error in y

Suppose that, instead of observing y_{at}, we have data only on a noisy measurement $\bar{y}_{at}$, where

$$\bar{y}_{at} = y_{at} + \epsilon_{at}. \tag{8}$$

For example, y_{at} could be the probability of death for individuals age a in year t, and $\bar{y}_{at}$ could be the observed mortality rate, which is a random variable with mean y_{at} when the population is finite. Alternatively, y_{at} could be a measure of consumption inequality among all people age a in year t, and $\bar{y}_{at}$ could be an estimate of inequality calculated from a random sample of the population. We now show conditions on the measurement error ϵ_{at} under which our model remains identified.

Assumption 1. *$\mathrm{E}[\epsilon_{at}|y_{at}] = 0$ and $\mathrm{E}[\epsilon_{a,t}^2|y_{at}] = \sigma^2$ for all a, t, and $\mathrm{E}[\epsilon_{a,t}\epsilon_{a',t'}|y_{at}, y_{a't'}] = 0$ whenever $a' \neq a$ or $t' \neq t$.*

Assumption 1 restricts the variance-covariance matrix of the measurement error. We must impose such a restriction because age-period-cohort analysis is, in essence, a decomposition of variance. In section 5, we will consider an application in which assumption 1 is plausible.

Proposition 3. *Suppose assumption 1 and the hypotheses of proposition 1 hold. Let*

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \sum_{a=0}^{A} \sum_{t=1}^{T} \left(\bar{y}_{at} - \tilde{\mu} - \sum_{k=1}^{K} \sum_{a'=0}^{a} \tilde{r}_k^{a-a'} \tilde{w}_{k,a'} \tilde{e}_{k,t-a+a'} \right)^2 . \tag{9}$$

Then, subject to regularity conditions on ϵ_{at}:

(a) $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$ in the limit as $\sigma^2 \to 0$ with A and T fixed, and

(b) If $e_{k,t}$ is a stationary and ergodic process, then $\left(\left\{\{\hat{w}_{k,a}\}_{a=1}^{A}, \hat{r}_k\right\}_{k=1}^{K}, \hat{\mu}\right) \xrightarrow{p} \left(\left\{\{w_{k,a}\}_{a=1}^{A}, r_k\right\}_{k=1}^{K}, \mu\right)$ in the limit as $T \to \infty$ with A fixed.

Proposition 3 says certain parameters can be consistently estimated by nonlinear least squares when outcomes are measured with uncorrelated, homoskedastic, mean-zero error. In the limit as the variance of the measurement error goes to zero, all of the parameters can be consistently estimated; this limit applies when $\bar{y}_{at}$ is computed from large populations in each (a, t) cell, as in the case of mortality rates calculated from vital records. In the limit as T goes to infinity with A fixed – as when small samples are collected in each of many years – all parameters except the time effects $e_{k,t}$ can be consistently estimated; parameters indexed by t cannot be consistently estimated because adding data on new time periods does not add information about parameters relevant only to earlier time periods. We do not consider limits as A goes to infinity because the human life span is finite. One can test whether the homoskedasticity requirement ($\mathrm{E}[\epsilon_{a,t}^2|y_{at}] = \sigma^2$) in assumption 1 holds by examining whether the squared residuals $\left(\bar{y}_{at} - \hat{\mu} - \sum_{k=1}^{K} \sum_{a'=0}^{a} \hat{r}_k^{a-a'} \hat{w}_{k,a'} \hat{e}_{k,t-a+a'}\right)^2$ are systematically related to the predicted values $\hat{y}_{at} = \hat{\mu} + \sum_{k=1}^{K} \sum_{a'=0}^{a} \hat{r}_k^{a-a'} \hat{w}_{k,a'} \hat{e}_{k,t-a+a'}$.

5 Example: Mortality Rates in Sweden

In Western developed countries, the demographic transition in the past two hundred years featured gradual mortality declines in response to improvements in features of the environment including water quality, sanitation, nutrition, prevalence of infectious diseases, and medical technology (Elo and Preston, 1992; Omran, 1982). How did these changes impact mortality for various birth cohorts? And how did they differentially affect people of different ages? We answer these questions by applying our model to estimate the age profiles of mortality under present and past conditions using the long time series of high quality mortality data from Sweden in the Human Mortality Database (2007). Our approach is to compare the observed and model-predicted age-specific mortality rates across cohorts. Because our model yields age profiles under the hypothetical scenario in which conditions at birth remain constant and operate throughout (and hence accumulate over) a cohort's life, discrepancies between the predicted and actual age profiles indicate changes in conditions in historical time.

We analyze the 5-by-5 table of mortality rates for ages 0-4, 10-14, ..., 75-79 and years 1800-1804, 1805-1809, ..., 2000-2004, dropping older ages and earlier years due to data quality concerns discussed in the data documentation. (We do not use annual data for the same reason.) Figure 1 displays the data. Infant mortality has decreased proportionately much more than adult mortality over the past two centuries – exactly the kind of shift in an age profile of outcomes that our model aims to capture.

The dependent variable we analyze is the natural logarithm of the realized mortality rate among people who are age a in year t. We treat (3) as a model of the underlying log probability of death and (8) as a model of log realized mortality, which randomly differs from the log probability of death in a finite population. We estimate the model by nonlinear least squares as in (9), weighting each age-year cell by a consistent estimate of the inverse of the variance of observed log mortality in that cell. Appendix B shows that this procedure is equivalent to maximum likelihood.

We estimate four models: the additive model (1) as well as the continuously accumulating model (3) for $K = 1$, $K = 2$ and $K = 3$. (We did not attempt models with $K > 3$ due to the large number of parameters involved.) Our purpose in estimating the additive model is not to interpret its parameters but only to test it against the more general $K = 3$ model in which it is nested. For this purpose, the failure of identification in the additive model does not cause problems: We need to obtain only the log likelihood of the additive model, which does not depend on which single identifying constraint we impose on the parameters.

We will present results at the PAA meeting.

6 Conclusion

The conventional linear model of additive age, period, and cohort effects has been widely used to analyze tabular population level data. The literature, however, often concludes that it is impossible to obtain meaningful estimates of the distinct contributions to social change of age, time period, and cohort. The methodological problem underlying this conclusion is well recognized: In the additive model, one must resolve the identification problem induced by the exact linear dependency between age, period, and cohort indicators by imposing some identifying constraint, and there is no consensus as to what constitutes a satisfactory constraint.

In this paper, we emphasize that the APC identification problem is inevitable only under the conventional specification of fixed, additive age, period, and cohort effects. But additive effects are merely one approximation to the process of social change. A prominent example of an alternative process is that of continuously accumulating or evolving cohort effects, described decades ago by social demographers who also noted the absence of procedures for empirically investigating such a process (Hobcraft et al., 1982; Ryder, 1965). It is this process that we attempt to model in this paper.

The new model relaxes the assumption of the conventional additive model that the influence of age is the same in all time periods, the influence of present conditions is the same for people of all ages, and cohorts do not change over time. We show that the failure of identification in the conventional model stems precisely from the strong assumptions it makes. When we generalize the model to allow age profiles to change over time, period effects to have different influences on people of different ages, and cohorts to evolve from one period to the next, we obtain a model that *is* identified. More important, we can better capture the essence of social change by taking into account the fact that cohorts are continuously exposed to influences that cumulatively alter their trajectories. As an example, our data analysis illustrates the utility of this model in studying the evolution of human mortality. Estimates of age profiles of mortality allow us to discern the effects of various historical conditions in shaping cohort mortality at specific ages. We believe that, beyond demography, this model can find wide application in economics, sociology, and political science and can potentially provide new stylized facts that are fundamental to construction of basic explanations and evaluation of theories of social change and structure.

A Proofs

A.1 Proposition 1

We prove the result separately for $K = 1$, $K = 2$ and $K = 3$. In each case, the strategy will be to construct a set $X_K \subset \mathbf{\Theta}$ such that X_K is of measure zero and such that, unless $\boldsymbol{\theta} = \left[\mu, \left\{\{e_{k,t}\}_{t=1-A}^{T}, \{w_{k,a}\}_{a=1}^{A}, r_k\right\}_{k=1}^{K}\right]$ is in X_K, the equality

$$\mu+\sum_{k=1}^{K}\sum_{a'=0}^{a} r_k^{a-a'} w_{k,a'} e_{k,t-a+a'} = \tilde{\mu}+\sum_{k=1}^{K}\sum_{a'=0}^{a} \tilde{r}_k^{a-a'} \tilde{w}_{k,a'} \tilde{e}_{k,t-a+a'},\ a = 0, \ldots, A,\ t = 1, \ldots, T, \tag{A.1}$$

implies, under the hypotheses of the proposition, that $\left[\mu, \left\{\{e_{k,t}\}_{t=1-A}^{T}, \{w_{k,a}\}_{a=1}^{A}, r_k\right\}_{k=1}^{K}\right] = \left[\tilde{\mu}, \left\{\{\tilde{e}_{k,t}\}_{t=1-A}^{T}, \{\tilde{w}_{k,a}\}_{a=1}^{A}, \tilde{r}_k\right\}_{k=1}^{K}\right] \equiv \tilde{\boldsymbol{\theta}}$.

Case 1: $K = 1$. Let X_1 be the set of $\boldsymbol{\theta} \in \mathbf{\Theta}$ such that either $r_1 + w_{1,1} = 1$ or the vectors $(e_{1,1}, \ldots, e_{1,T-1})$ and $(e_{1,2}, \ldots, e_{1,T})$ are collinear with a constant. X_1 is a set of measure zero. Assume $\boldsymbol{\theta} \in \mathbf{\Theta} \setminus X_1$. Specializing (A.1) to $K = 1$, $a = 0$ and $a = 1$ (by hypothesis, $A \geq 1$) and using normalization 2, we have

$$\mu + e_{1,t} = \tilde{\mu} + \tilde{e}_{1,t}, \quad t = 1, \ldots, T, \tag{A.2a}$$

$$\mu + r_1 e_{1,t-1} + w_{1,1} e_{1,t} = \tilde{\mu} + \tilde{r}_1 \tilde{e}_{1,t-1} + \tilde{w}_{1,1} \tilde{e}_{1,t}, \quad t = 2, \ldots, T. \tag{A.2b}$$

Substituting (A.2a) into (A.2b) and collecting terms gives

$$0 = (\mu - \tilde{\mu})(1 - \tilde{r}_1 - \tilde{w}_{1,1}) - (\tilde{r}_1 - r_1) e_{1,t-1} - (\tilde{w}_{1,1} - w_{1,1}) e_{1,t}, \quad t = 2, \ldots, T. \tag{A.3}$$

By hypothesis, $T \geq 4$, so (A.3) contains at least three equations. Since (given $\boldsymbol{\theta} \notin X_1$) $e_{1,t-1}$ and $e_{1,t}$ are not collinear with a constant, (A.3) can hold only if $(\mu - \tilde{\mu})(1 - \tilde{r}_1 - \tilde{w}_{1,1}) = 0$ and the coefficients on $e_{1,t-1}$ and $e_{1,t}$ are both zero. Hence $\tilde{r}_1 = r_1$, $\tilde{w}_{1,1} = w_{1,1}$, and, since $1 - r_1 - w_{1,1} \neq 0$ for $\boldsymbol{\theta} \notin X_1$, we must have $\tilde{\mu} = \mu$. It follows from (A.2a) that $\tilde{e}_{1,t} = e_{1,t}$ for $t = 1, \ldots, T$. Finally, substituting the foregoing results into (A.1) for $a \geq 2$ shows that $\tilde{e}_{1,t} = e_{1,t}$ for $t \leq 0$ and $\tilde{w}_{1,a} = w_{1,a}$ for $a \geq 2$.

Case 2: $K = 2$. Define the following sets:

$$
\begin{aligned}
&X_{2,1} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \colon \exists k \text{ s.t. } r_k = 0\}, \quad X_{2,2} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \colon r_1 = r_2\}, \\
&X_{2,3} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \colon w_{1,1} = w_{2,1}\}, \\
&X_{2,4} = \left\{ \boldsymbol{\theta} \in \boldsymbol{\Theta} \colon \operatorname{rank} \begin{bmatrix} 1 & \begin{pmatrix} e_{k,j} \\ \vdots \\ e_{k,T-4+j} \end{pmatrix}_{\substack{k \in \{1,2\}, \\ j \in \{1,2,3,4\}}} \\ \vdots & \\ 1 & \end{bmatrix} < 9 \right\}, \\
&X_{2,5} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \colon (w_{1,1} - w_{2,1})[-r_2 w_{1,1} + r_1 w_{2,1}] + (w_{1,2} - w_{2,2})(r_2 - r_1) = 0\}, \\
&X_{2,6} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} \colon w_{1,1} - w_{2,1} + r_1 - r_2 + (r_2^2 + r_2 w_{2,1} + w_{2,2})(1 - r_1 - w_{1,1}) \\
&\qquad\qquad - (w_{1,2} + r_1 w_{1,1} + r_1^2)(1 - r_2 - w_{2,1}) = 0\}.
\end{aligned}
\tag{A.4}
$$

Let $X_2 = \cup_{j=1}^{6} X_{2,j}$. X_2 has measure zero. Our web appendix shows that under normalizations 1 to 3 and the hypotheses of the proposition, if $\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus X_2$, then the unique solution to (A.1) is $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$. The algebra proceeds by using (A.1) at $a = 0$ and $a = 1$ to eliminate $\tilde{e}_{2,t}$ and obtain a first-order difference equation in $\tilde{e}_{1,t}$; substituting the difference equation into (A.1) at $a = 2$ to eliminate $\tilde{e}_{1,t}$; and observing that coefficients in a linear combination of a constant with $\{e_{k,t-3}, \ldots, e_{k,t}\}_{k=1}^{2}$ must be zero given $\boldsymbol{\theta} \notin X_{2,4}$. Setting the coefficients to zero yields quadratic equations with two solutions, $(\tilde{r}_1, \tilde{r}_2) = (r_1, r_2)$ and $(\tilde{r}_1, \tilde{r}_2) = (r_2, r_1)$; normalization 1 rules out the latter to give uniqueness.

Case 3: $K = 3$. The approach parallels the $K = 2$ case; see the web appendix. $\square$

A.2 Proposition 2

We must show that under each of conditions (a) and (b), (A.1) has multiple solutions for $(\tilde{\mu}, \tilde{\mathbf{r}}, \tilde{\mathbf{e}}, \tilde{\mathbf{w}})$ in terms of $(\mu, \mathbf{r}, \mathbf{e}, \mathbf{w})$, and that this is so even if $\tilde{\mu} = \mu$.

Condition (a): Without loss of generality, suppose $e_{1,t} = \bar{e}_1$. Choose any $r^* \in [0, 1]$. Let $\{w_a^*\}_{a=1}^A$ be the unique solution to the following nonsingular triangular system of linear equations given r^*, r_1 and $\{w_{1,a}\}_{a=1}^A$:

$$\sum_{a'=1}^{a} (r^*)^{a-a'} w_{a'}^* = -(r^*)^a + \sum_{a'=0}^{a} r_1^{a-a'} w_{1,a'}, \quad a = 1, \ldots, A. \tag{A.5}$$

Given $e_{1,t} = \bar{e}_1$, the following solves (A.1): $\tilde{\mu} = \mu$; $\tilde{e}_{j,t} = e_{j,t}\ \forall\ j, t$; $\tilde{r}_1 = r^*$; $\tilde{r}_j = r_j\ \forall\ j > 1$; $\tilde{w}_{1,a} = w_a^*\ \forall\ a$; $\tilde{w}_{j,a} = w_{j,a}\ \forall\ j > 1, a$. Therefore, (A.1) has a continuum of solutions indexed by $r^* \in [0, 1]$.

Condition (b): Without loss of generality, suppose $r_1 = r_2$. Choose any $x \in (1/2, 1]$. Given $r_1 = r_2$, the following solves (A.1): $\tilde{\mu} = \mu$; $\tilde{r}_j = r_j\ \forall\ j$; $\tilde{w}_{1,a} = x w_{1,a} + (1 - x) w_{2,a}\ \forall\ a$; $\tilde{w}_{2,a} = (1 - x) w_{1,a} + x w_{2,a}\ \forall\ a$; $\tilde{e}_{1,t} = \frac{(1-x)e_{2,t} - x e_{1,t}}{1-2x}\ \forall\ t$; $\tilde{e}_{2,t} = \frac{(1-x)e_{1,t} - x e_{2,t}}{1-2x}\ \forall\ t$; $\tilde{e}_{j,t} = e_{j,t}\ \forall\ j > 2, t$; $\tilde{w}_{j,a} = w_{j,a}\ \forall j > 2, a$. Therefore, (A.1) has a continuum of solutions indexed by $x \in (1/2, 1]$. $\square$

A.3 Proposition 3

We assume the distribution of ϵ_{at} satisfies regularity conditions such that a uniform law of large numbers (ULLN) holds. Case (a): If $\sigma^2 = 0$, (9) becomes (A.1); hence the true parameters uniquely solve (9) when $\sigma^2 = 0$. Since the objective function in (9) is continuous, a ULLN applies, and solutions for $\sigma^2 > 0$ converge to the solution for $\sigma^2 = 0$. Case (b): Note that the predicted values can be written as $\hat{\mathbf{y}}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu}, \tilde{\mathbf{e}}) =$

$\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\tilde{\mathbf{e}}$ and hence, if we solve (9) for $\hat{\mathbf{e}}$ as a function of the remaining parameters, we obtain $\hat{\mathbf{e}} = [\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})]^{-1}\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'(\mathbf{y} + \boldsymbol{\epsilon})$. (The inverse should be interpreted as a generalized inverse when $\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})$ is not of full rank. Substituting this solution into (9), we obtain

$$\begin{aligned}
&\left(\left\{\{\hat{w}_{k,a}\}_{a=1}^{A}, \hat{r}_k\right\}_{k=1}^{K}, \hat{\mu}\right) \\
&\quad = \arg\min_{\tilde{\mathbf{w}},\tilde{\mathbf{r}},\tilde{\mu}} \Big\{ \big[\mathbf{y} - \mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})[\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})]^{-1}\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'(\mathbf{y} + \boldsymbol{\epsilon}) + \boldsymbol{\epsilon}\big] \Big\}' \\
&\qquad\qquad \big[\mathbf{y} - \mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})[\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})]^{-1}\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'(\mathbf{y} + \boldsymbol{\epsilon}) + \boldsymbol{\epsilon}\big] \\
&\quad = \arg\min_{\tilde{\mathbf{w}},\tilde{\mathbf{r}},\tilde{\mu}} \Big\{ (\mathbf{y} + \boldsymbol{\epsilon})'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})(\mathbf{y} + \boldsymbol{\epsilon}) \Big\} \\
&\quad = \arg\min_{\tilde{\mathbf{w}},\tilde{\mathbf{r}},\tilde{\mu}} \Big\{ \mathbf{y}'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\mathbf{y} + 2\mathbf{y}'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\boldsymbol{\epsilon} \Big\}
\end{aligned} \tag{A.6}$$

where $\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu}) = \mathbf{I} - \mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})$ is a symmetric and idempotent matrix. Since $e_{k,t}$ is stationary and ergodic, so is y_{at}, and so the ergodic theorem and ULLN apply to the new objective function. Hence as $T \to \infty$, the second term in the objective function converges uniformly in probability to zero. Further, since ϵ_{at} is serially uncorrelated and homoskedastic by assumption 1, the third term converges uniformly in probability to $\sigma^2 \text{tr}[\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})]$. Since $\mathbf{M}$ is idempotent, its trace equals its rank, which is no smaller than its rank when $\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})$ has full rank. At the true parameters, $\mathbf{X}$ has full rank. Hence, in the limit as $T \to \infty$, the true parameters minimize the third term. Further, in the limit as $T \to \infty$, the first term converges uniformly in probability to a function that is zero at the true parameters and, by proposition 1, strictly positive otherwise. Thus the objective function converges uniformly in probability to a function minimized by the true parameters. It follows that $\left(\left\{\{\hat{w}_{k,a}\}_{a=1}^{A}, \hat{r}_k\right\}_{k=1}^{K}, \hat{\mu}\right)$ converges in probability to the true parameters. □

B Equivalence of weighted NLS and MLE

Suppose each individual who is age a at time t has a probability of death p_{at}, and let N_{at} be the population at risk in cell (a, t). If $\bar{p}_{at}$ is the realized mortality rate in the cell, then by the central limit theorem, $\sqrt{N_{at}}(\bar{p}_{at} - p_{at}) \xrightarrow{d} \mathcal{N}[0, p_{at}(1 - p_{at})]$ as $N_{at} \to \infty$. (The smallest cell in our data has $N_{at} = 97{,}551$, and the median cell has $N_{at} = 1{,}570{,}678$, so approximating the distribution by the limit as $N_{at} \to \infty$ seems reasonable.) By the delta method, $\sqrt{N_{at}}(\ln \bar{p}_{at} - \ln p_{at}) \xrightarrow{d} \mathcal{N}[0, (1 - p_{at})/p_{at}]$. We observe realized log mortality $\bar{y}_{at} \equiv \ln \bar{p}_{at}$ and population N_{at} but not true log mortality $y_{at} \equiv \ln p_{at}$; indeed, the goal is to estimate parameters determining y_{at}. But $\bar{p}_{at} \xrightarrow{p} p_{at}$, so by the continuous mapping theorem, $\sqrt{N_{at}\bar{p}_{at}/(1 - \bar{p}_{at})}(\bar{y}_{at} - y_{at}) \xrightarrow{d} \mathcal{N}(0, 1)$. If p_{at} depends on parameters $\boldsymbol{\theta}$, the log likelihood for data on ages $a = 0, \ldots, A$ and years $t = 1, \ldots, T$ is $\ln L = -\frac{(A+1)T}{2} \ln(2\pi) - \frac{1}{2} \sum_{a=0}^{A} \sum_{t=1}^{T} \frac{N_{at}\bar{p}_{at}}{1-\bar{p}_{at}} [\bar{y}_{at} - y_{at}(\boldsymbol{\theta})]^2$. Maximizing the likelihood is thus equivalent to minimizing the weighted nonlinear least squares objective function for the model $\bar{y}_{at} = \ln(p_{at}(\boldsymbol{\theta})) + \epsilon_{at}$ with weights $\hat{\sigma}^2_{at} = N_{at}\bar{p}_{at}/(1 - \bar{p}_{at})$. The minimized WNLS objective function, divided by the residual degrees of freedom, is an estimate of dispersion; the dispersion should be 1 if the model fully accounts for variation in mortality. In practice, since we estimate dispersion greater than 1, we compute the log likelihood and standard errors without assuming the dispersion equals 1.

References

Deaton, Angus and Christina Paxson, 1994, “Intertemporal Choice and Inequality,” *Journal of Political Economy* 102(3), 437–467.

Elo, Irma and Samuel H. Preston, 1992, “Effects of Early-life Conditions on Adult Mortality: A Review,” *Population Index* 58(2), 186–212.

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

Greenberg, B.G., J.J. Wright and C.G. Sheps, 1950, "A Technique for Analyzing Some Factors Affecting the Incidence of Syphilis," *Journal of the American Statistical Association* 45, 373-399.

Heligman, L. and J.H. Pollard, 1980, "The Age Pattern of Mortality," *Journal of the Institute of Actuaries* 107, 49-80.

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

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org (data downloaded November 13, 2007).

James, I.R. and M.R. Segal, 1982, "On a Method of Mortality Analysis Incorporating Age-Year Interaction, with Application to Prostate Cancer Mortality," *Biometrics* 38(2), 433-443.

Mason, Karen Oppenheim, William M. Mason, H.H. Winsborough and W. Kenneth Poole, 1973, "Some Methodological Issues in Cohort Analysis of Archival Data," *American Sociological Review* 38, 242-258.

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

Moolgavkar, Suresh H., Richard G. Stevens and John A.H. Lee, 1979, "Effect of Age on Incidence of Breast Cancer in Females," *Journal of the National Cancer Institute* 62(3), 493-501.

Noymer, Andrew and Michel Garenne, 2000, "The 1918 Influenza Epidemic's Effects on Sex Differentials in Mortality in the United States," *Population and Development Review* 26, 565-581.

Olshansky, S.J. and A.B. Ault, 1986, "The Fourth Stage of the Epidemiological Transition: The Age of Delayed Degenerative Diseases," *Milbank Memorial Fund Quarterly* 64, 355–391.

Omran, Abdel R., 1982, "Epidemiologic transition," in *International Encyclopedia of Population*, pp. 172–183, New York: Free Press.

Quigley, John Michael, 1972, "An Economic Model of Swedish Emigration," *Quarterly Journal of Economics* 86(1), 111-126.

Robertson, Chris, Sara Gandini and Peter Boyle, 1999, "Age-Period-Cohort Models: A Comparative Study of Available Methodologies," *Journal of Clinical Epidemiology* 52, 569-583.

Ryder, Norman B., 1965, "The Cohort as a Concept in the Study of Social Change," *American Sociological Review* 30(6), 843–861.

United Nations Demographic Yearbook 1997, historical supplement, table 11, New York: United Nations.

Vaupel, James W., Kenneth G. Manton and Eric Stallard, 1979, "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography* 16, 439-454.

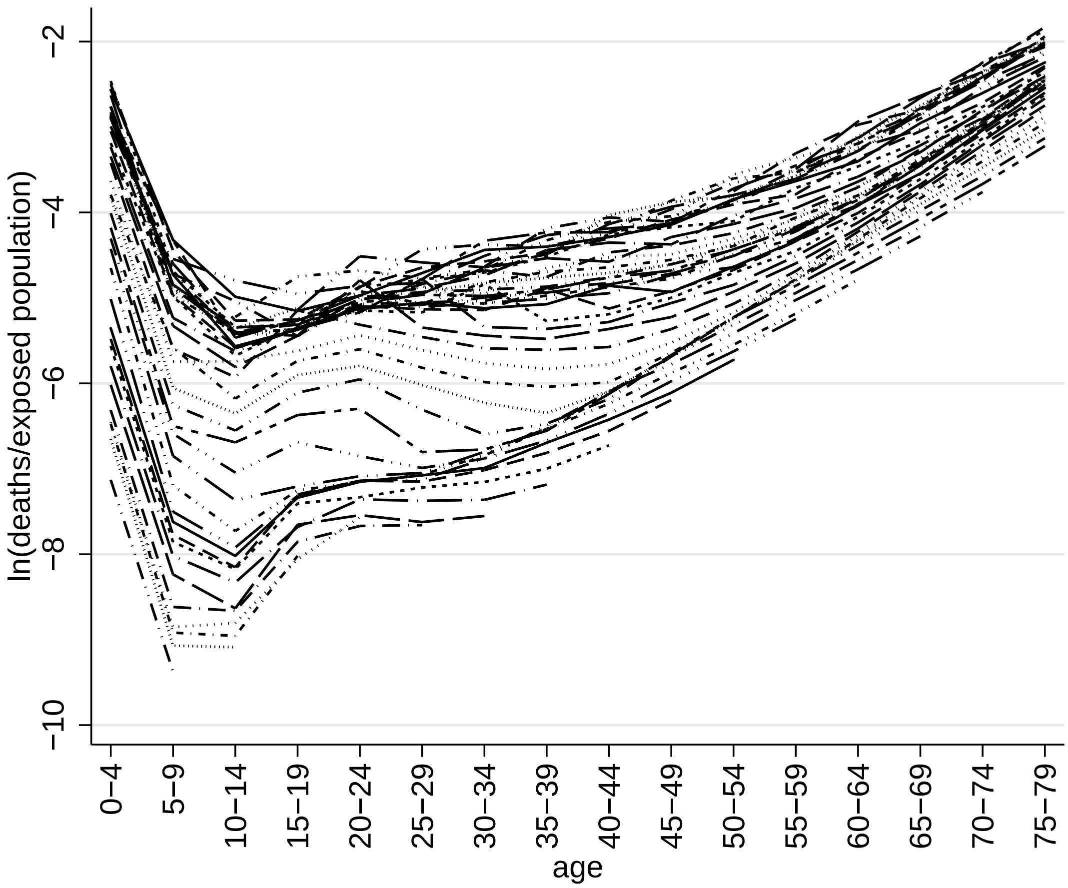

Figure 1: Mortality in Sweden at five-year intervals, 1800-2004.

Each line shows the realized mortality of a particular birth cohort at various ages. Cohorts included are those born in 1721 through 2004. Lines for cohorts born before 1800 or after 1929 omit some ages because the dataset does not cover those ages for those cohorts. Data source: Human Mortality Database (2007).