

Federal Reserve Bank of Minneapolis  
Research Department Staff Report 461

September 2011

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

Sam Schulhofer-Wohl

Federal Reserve Bank of Minneapolis

Yang Yang

University of North Carolina at Chapel Hill

\*Revised version of a paper presented at the 2008 annual meeting of the American Sociological Association and the 2009 annual meeting of the Population Association of America. We are grateful to audiences there and at the University of Cape Town, the Federal Reserve Bank of Minneapolis, Princeton University, and the Massachusetts Institute of Technology, as well as to Stephen Raudenbush, for helpful comments. We thank Kristin Plys and Thu Vu for excellent research assistance. This research was supported by a pilot grant from the Princeton Center for the Demography of Aging, NIH P30 AG024361. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

## ABSTRACT

---

This paper proposes a new way of modeling age, period, and cohort effects that improves substantively and methodologically on the conventional linear model. The linear model suffers from a well-known identification problem: If we assume 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 these three separate effects 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 applications, these assumptions fail. We propose a more general model in which age profiles can change over time and period effects can have different influences on people of different ages. Our model defines cohort effects as an accumulation of age-by-period interactions. Although a long-standing literature on theories of social change conceptualizes cohort effects in exactly this way, we are the first to show how to statistically model this more complex form of cohort effects. 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 150 years. Our model fits the data dramatically better than the additive model. The estimates show that the rate of increase of mortality with age among adults became more steep from 1881 to 1941, but since then the rate of increase has been roughly constant. The estimates also allow us to test whether early-life conditions have lasting impacts on mortality, as under the cohort morbidity phenotype hypothesis. The results give limited support to this hypothesis: The impact of early-life conditions lasts for several years but is unlikely to reach all the way to old age.

Keywords: Cohort effects; Age-period-cohort identification problem; Mortality; Cohort morbidity phenotype hypothesis; Sweden  
JEL: C23, I15, J11, N33

---

# 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. Researchers commonly analyze data consisting of an  $(A + 1) \times T$  matrix of outcomes for ages  $a = 0, 1, \dots, A$  and dates  $t = 1, 2, \dots, T$ , such as a table of age-specific mortality rates in various years. The diagonals of this matrix correspond to birth cohorts: people born in the same year and aging together. Figure 1 illustrates a typical dataset on mortality in Sweden; we will analyze these data later in the paper. Most attempts to date to distinguish age, period, and cohort effects in such data have used linear models. However,

it is impossible to distinguish the separate effects of age, period, and birth cohort in a linear regression model because age, period, and cohort are linearly dependent; for any person, *birth year* = *current year* - *age*. The problem persists even if one specifies age, period, and cohort effects nonparametrically with dummy variables for each possible value, as in the additive model

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

where  $y_{at}$  is some outcome for people of age  $a$  in year  $t$  (who are therefore members of birth cohort  $j = t - a$ );  $\alpha_a$  is the effect of being age  $a$ ;  $\beta_t$  is the effect of living in year  $t$ ; and  $\gamma_j$  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  $\delta$ , we also have

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

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

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.

Besides being unidentified, the conventional additive model (1) has serious substantive limitations. The additive model is a quite simple approximation to the process of social change and does not adequately describe many 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, 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, and these events change the cohort in long-lasting ways. But because new events constantly occur, a model with unchanging cohort effects is appropriate only if all relevant events occur before the initial observation and only if the impact of these events 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. Our model improves on the additive model in two ways. First, in our model, age profiles can change over time and period effects can have different influences on people of different ages. Second, our model operationalizes Ryder’s concept of continuously evolving cohort effects by specifying cohort effects as accumulations of age-by-period interactions. These substantive contributions lead to a methodological contribution. 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. In other words, by broadening the concept of cohort effects, our model avoids the identification problem that has bedeviled the previous literature on 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, with some additional details in a separate technical 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  $\mathbf{e}_1, \dots, \mathbf{e}_K$ , 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.

The influence of past time effects may increase or fall off over time. Suppose we let  $r(k, a, a')$  represent the increase or decay between ages  $a'$  and  $a$ , where  $r(k, a, a) \equiv 1$ . Then the impact in year  $t$  of time effects from sequence  $k$  occurring in year  $s \leq t$  for people who were age  $a'$  in year  $s$  would be

$$r(k, a, a')w_{k,a'}e_{k,s}, \quad a = a' + t - s. \quad (3)$$

For example,  $r(k, a, a')$  could be a step function if time effects at a young age have no further impact until old age, as in the case of increases in old-age lung cancer mortality that result from increasing popularity of cigarettes when people are young. The general form of increase or decay in (3) may, however, be difficult to analyze. In particular, modeling  $r(k, a, a')$  as a step function with steps at unknown ages would lead to nonsmooth likelihood functions, and allowing  $r(k, a, a')$  to depend nonparametrically on  $a$  and  $a'$  would add  $A(A + 1)K/2$  parameters to the model. Therefore, in this paper we restrict attention to the simplest possible smooth approximation — exponential increase or decay:

$$r(k, a, a') = r_k^{a-a'}, \quad r_k \geq 0. \quad (4)$$



(We adopt the convention that  $0^0 = 1$ .) Although the exponential form does not encompass all possible forms of increase or decay, its simplicity makes it easy to analyze, and we will show that it generalizes the additive model. We recognize that other choices of  $r(k, a, a')$  may be valuable in particular applications and leave the analysis of such models for future research.

Having defined our building blocks  $e_{k,t}$ ,  $w_{k,a}$ , and  $r(k, a, a')$ , 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:

$$y_{at} = \mu + \sum_{k=1}^K \sum_{a'=0}^a r_k^{a-a'} w_{k,a'} e_{k,t-a+a'}. \quad (5)$$

We now consider how to interpret the parameters of the model. 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  $k$ th 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  $k$ th sequence of time effects,  $r_k = 0$  and  $w_a^k = 1$  for all  $a$ . Then the contribution of the  $k$ th 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  $k$ th sequence of time effects,  $r_k = 1$ ,  $w_0^k = 1$ , and  $w_a^k = 0$  for  $a > 0$ . Then the contribution of the  $k$ th 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 (5) reduces to

$$y_{at} = \mu + \sum_{a'=0}^a r_1^{a-a'} w_{1,a'} \bar{e}_1 + e_{2,t} + e_{3,j}, \quad j = t - a, \quad (6)$$

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  $\gamma_j = e_{3,j}$ .

Three questions are of substantial interest in analyzing age- and period-specific data. First, how do outcomes vary with age, and how has the effect of age changed over time? For example, how does the mortality rate depend on age? Or, how does within-cohort consumption inequality change as the cohort ages? Second, how do outcomes depend on conditions in the current period? And third, does history influence current outcomes in a way that age effects do not fully capture — in other words, is a cohort theory appropriate? The parameters of our model help answer all three of these questions.

We can address the first question by estimating age profiles of outcomes. 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  $(\bar{e}_1, \bar{e}_2, \dots, \bar{e}_K)$  in every year of its life would, at age  $a$ , have outcomes

$$y_a(\bar{e}_1, \bar{e}_2, \dots, \bar{e}_K) = \mu + \sum_{k=1}^K \bar{e}_k m_k(a), \quad (7)$$

which depends only on the cohort's age  $a$ . Two kinds of comparisons are in order. First, by comparing the model profile with the observed outcomes  $y_{a,j+a}$  for a particular cohort  $j$ , we can see how the cohort's outcomes differ from what we would have predicted if conditions had never changed throughout its life. This comparison applies to actual people — it can help us see whether, and in what way, conditions changed over a particular cohort's life course. Second, consider a different hypothetical cohort that experienced a different set of constant time effects  $(\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_K)$  in every year of its life. The second cohort would have outcomes

$$y_a(\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_K) = \mu + \sum_{k=1}^K \tilde{e}_k m_k(a). \quad (8)$$

The outcomes in (8) again depend only on the cohort's age, but they differ from the outcomes of the first hypothetical cohort in (7). 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:

$$\begin{aligned} y_a(e_{1,1900}, e_{2,1900}, \dots, e_{K,1900}) &= \mu + \sum_{k=1}^K e_{k,1900} m_k(a), \\ y_a(e_{1,2000}, e_{2,2000}, \dots, e_{K,2000}) &= \mu + \sum_{k=1}^K e_{k,2000} m_k(a). \end{aligned} \quad (9)$$

The profile  $y_a(e_{1,1900}, e_{2,1900}, \dots, e_{K,1900})$  tells us the effect of age on outcome  $y$  in 1900. We can interpret  $y_a(e_{1,1900}, e_{2,1900}, \dots, e_{K,1900})$  as a prediction for the outcomes of the 1900 birth cohort if conditions never changed after the cohort's birth. In other words,  $y_a(e_{1,1900}, e_{2,1900}, \dots, e_{K,1900})$  describes the effect of age on outcomes  $y$ , holding time effects constant. Similarly, the profile  $y_a(e_{1,2000}, e_{2,2000}, \dots, e_{K,2000})$  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. This comparison applies not to particular cohorts but to history — it can help us see whether, and in what way, conditions changed between two perhaps widely separated eras.

We can address the second question — how do outcomes depend on current conditions? — by examining both the period effects  $e_{k,t}$  and the age weights  $w_{k,a}$ . Suppose we are examining mortality, and suppose that for some age  $a$ ,  $w_{k,a}$  is positive. Also suppose that for two years  $s$  and  $t$ ,  $e_{k,s} < e_{k,t}$ . The immediate impact of conditions in year  $t$  on people of age  $a$  is  $w_{k,a}e_{k,t}$ . Thus, all else equal, mortality for people of age  $a$  is predicted to be lower in year  $s$  than in year  $t$ .

We can address the third question — is a cohort theory appropriate? — by examining the rates of decay  $r_k$ . A cohort theory predicts that past events continue to affect a cohort's outcomes at much later ages. Recall that the impact at age  $a$  of conditions at age  $a' < a$  is  $r_k^{a-a'} w_{k,a'} e_{k,t-a+a'}$ . A cohort theory says that even if  $a'$  is much less than  $a$ , this impact is large. But in that case,  $r_k$  must be close to one. Therefore, we can tell whether a cohort theory describes the data by examining whether any  $r_k$  is close to one.

By combining all of the parameters, we can use our model to see how a particular cohort evolves over the life course. For simplicity, suppose  $K = 1$ , so there is only one type of time effect, and suppose we are studying mortality. Consider the mortality rate of cohort

$j$ . At age  $a$ , its mortality is

$$y_{a,j+a} = \mu + \sum_{a'=0}^a r_1^{a-a'} w_{1,a'} e_{1,j+a'}. \quad (10)$$

The next year, at age  $a + 1$ , the cohort's mortality is

$$y_{a+1,j+a+1} = \mu + \sum_{a'=0}^{a+1} r_1^{a+1-a'} w_{1,a'} e_{1,j+a'}. \quad (11)$$

The change in mortality from age  $a$  to age  $a + 1$  is thus

$$y_{a+1,j+a+1} - y_{a,j+a} = w_{1,a+1} e_{1,j+a+1} + (r_1 - 1) \sum_{a'=0}^a r_1^{a-a'} w_{1,a'} e_{1,j+a'}. \quad (12)$$

In words, the change between age  $a$  and age  $a + 1$  is a combination of what happens to the cohort when it reaches age  $a + 1$  (the term  $w_{1,a+1} e_{1,j+a+1}$ ) and a decrease in the influence of its past experiences (the term consisting of  $(r_1 - 1)$  multiplied by an accumulation of time effects at ages  $a' < a + 1$ ). The cohort evolves because it has new experiences and because the influence of the past diminishes.

### 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 say the parameters of our model are identified if there exists a unique set of parameters that can generate any given matrix of outcomes  $y_{at}$  for  $a = 0, \dots, A$  and  $t = 1, \dots, T$ . 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 (5) holds for all  $a = 0, \dots, A$  and  $t = 1, \dots, T$ . 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  $\Theta$ , is identified with respect to  $\Theta$  if there does not exist any vector  $\tilde{\boldsymbol{\theta}} \in \Theta$  distinct from  $\boldsymbol{\theta}$  such that, given  $\{\{y_{at}\}_{a=0}^A\}_{t=1}^T$ , (5) holds for both  $\boldsymbol{\theta}$  and  $\tilde{\boldsymbol{\theta}}$  for all  $a = 0, \dots, A$  and  $t = 1, \dots, T$ .*

Under normalizations on the parameter space  $\Theta$  that do not affect the interpretation of the model, the set of parameter vectors that are not identified with respect to  $\Theta$  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 (7). 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  $\mathbf{r}$ ,  $\mathbf{w}$ ,  $\mu$ , 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  $\Theta$  consist of all vectors  $[\mu, \{\{e_{k,t}\}_{t=1-A}^T, \{w_{k,a}\}_{a=0}^A, r_k\}_{k=1}^K]$  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  $X_K \subset \Theta$  such that  $X_K$  is of measure zero and all  $\theta \in \Theta \setminus X_K$  are identified with respect to  $\Theta$ .*

Proposition 1 says there may be parameter vectors  $\theta$  that are not identified: For each of these  $\theta$ , there exists some  $\theta' \neq \theta$  that would generate the same data as  $\theta$ . However, the set  $X$  of unidentified parameter vectors is of measure zero. For almost all  $\theta$ , therefore, there

does not exist any  $\theta' \neq \theta$  that would generate the same data, and by observing  $y_{at}$ , we can uniquely determine the true parameter vector  $\theta$ . 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  $\theta \in \Theta$  is not identified if either:*

- (a)  $e_{k,t} = \bar{e}_k$  for some  $k$  and all  $t = 1 - A, \dots, T$ , or
- (b)  $K > 1$  and  $r_k = r_{k'}$  for some  $k \neq k'$ .

*Further,  $\theta$  remains unidentified in each of these cases even if  $\mu$  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  $\mu$  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  $\mu$  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  $\mu$ . 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

$$\bar{y}_{at} = y_{at} + \epsilon_{at}. \tag{13}$$

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  $\epsilon_{at}$  under which our model remains identified.

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

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{a=0}^A \sum_{t=1}^T \left( \bar{y}_{at} - \tilde{\mu} - \sum_{k=1}^K \sum_{a'=0}^a \tilde{r}_k^{a-a'} \tilde{w}_{k,a'} \tilde{e}_{k,t-a+a'} \right)^2. \quad (14)$$

Then, subject to regularity conditions on  $\epsilon_{at}$ :

- (a)  $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$  in the limit as  $\sigma^2 \rightarrow 0$  with  $A$  and  $T$  fixed, and
- (b) If  $e_{k,t}$  is a stationary and ergodic process, then  $(\{\{\hat{w}_{k,a}\}_{a=1}^A, \hat{r}_k\}_{k=1}^K, \hat{\mu}) \xrightarrow{p} (\{\{w_{k,a}\}_{a=1}^A, r_k\}_{k=1}^K, \mu)$  in the limit as  $T \rightarrow \infty$  with  $A$  fixed.

Proposition 3 says certain parameters can be consistently estimated by nonlinear least squares when outcomes are measured with serially 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 ( $E[\epsilon_{a,t}^2 | y_{at}] = \sigma^2$ ) 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

The demographic transition in Western developed countries over the past 200 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). When did these changes occur? How did they differentially affect people of different ages? And did they have lasting impacts on particular birth cohorts? We answer these questions by estimating our model on the long time series of high-quality mortality data from Sweden in the Human Mortality Database (2007).

We analyze data from 1861 to 2005 on ages 64 and younger. (We drop earlier years and older ages due to data quality concerns described in the Human Mortality Database documentation.) To keep the sample size and number of parameters manageable, we use only data on ages that are multiples of two  $(0, 2, \dots, 64)$  and at two-year intervals  $(1861, 1863, \dots, 2005)$ . 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 (5) as a model of the underlying log probability of death and (13) 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 (14), weighting each age-year cell by a consistent estimate of the inverse of the variance of observed log mortality in that cell. Appendix A4 shows this procedure is equivalent to maximum likelihood estimation and generates residuals that satisfy assumption 1.

We estimate the additive model (1) as well as the continuously accumulating model (5) for  $K = 1$ ,  $K = 2$  and  $K = 3$ . (We did not attempt models with  $K > 3$  because of 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.

Table 1 reports goodness-of-fit statistics for the additive and continuously accumulating models. The continuously accumulating model with  $K = 3$  fits best by any criterion: log likelihood, Akaike information criterion (AIC), or Bayesian information criterion (BIC). A likelihood ratio test of the  $K = 3$  continuously accumulating model against the additive model nested within it rejects the additive model with a  $p$ -value of zero. A possible concern is that our model uses so many parameters that it may overfit the data. However, even when we penalize our model for using more parameters by examining the BIC, we still find that our model is preferred to the additive model.

Figure 2 plots the residuals for each model. We observe a great deal of heteroskedasticity in the additive and  $K = 1$  models, but relatively little heteroskedasticity in the  $K = 3$  model, further evidence that the  $K = 3$  model accounts better for the data. (The unusually small residuals at the highest predicted values for  $K = 3$  correspond to data points for early years and young ages; one can show that our weights are noisily estimated for these data points and thus that the points may be overweighted in a finite sample, leading to incorrectly small residuals.) None of the models fits perfectly: Even with  $K = 3$ , the estimated dispersion is 1.89 times what it would be if our model accounted perfectly for the individual

probability of death and the residuals were due only to randomness in realized death rates. Still, the fit of the  $K = 3$  model is quite good. It accounts for 99.84 percent of the variation in the data and has one-eighth the overdispersion of the additive model. Also, as long as the residuals are homoskedastic, the overdispersion does not make our parameter estimates inconsistent.

We investigate the overdispersion by estimating two alternative models besides the additive model. The first alternative is the Lee and Carter (1992) model, which takes the form  $y_{at} = w_{1,a} + w_{2,a}e_{2,t}$ . The Lee-Carter model is a widely adopted statistical method for forecasting age-specific mortality rates (United Nations, 2003); although our model is explicitly *not* intended for forecasting, it is still interesting to compare our model fit with that of another model commonly estimated on mortality data. The second alternative is a three-factor model without lagged effects,  $y_{at} = \sum_{k=1}^3 w_{k,a}e_{k,t}$ . This model is simply our model (5) with  $K = 3$  under the restriction that  $r_1 = r_2 = r_3 = 0$  — that is, under the restriction that lagged effects do not accumulate over time — so estimating the three-factor nonaccumulating model is a way to tell whether the process of accumulation that we model is important. The  $K = 3$  continuously accumulating model fits better than either of these two alternatives by all criteria considered. The Lee-Carter model has four times as much overdispersion as the continuously accumulating model, demonstrating that the overdispersion problem is not limited to our model; mortality clearly is a complex phenomenon that any simple statistical model can only imperfectly describe. The rejection of the nonaccumulating model provides evidence that the accumulation of lagged effects is important in mortality data.

One possible reason for the overdispersion may be that we use a simple functional form for  $r(k, a, a')$ , the impact on outcomes at age  $a$  of time effects at age  $a' < a$ . If the

true relationship between time effects in youth and subsequent mortality is not exponential — for example, if early life conditions can cause mortality in early life and old age but not in between (e.g., Horiuchi, 1983) — then the form we use for  $r(k, a, a')$  will not adequately fit the data. As discussed in section 2, more complicated forms for  $r(k, a, a')$  would be difficult to implement, and we leave them for future research.

Because the  $K = 3$  continuously accumulating model fits the data best, we present parameter estimates only from that model. Table 2 reports the estimated rates of decay, and figure 3 shows the estimated age weights and sequences of time effects.

The first type of time effect,  $k = 1$ , has short-lasting effects, with an estimated half-life of about four months. The estimated age weights  $w_{1,a}$  show that these time effects impact mainly the young, not the middle-aged or the old. The estimated time effects  $e_{1,t}$  show that mortality related to these time effects first rose and then fell over the years we study. Examining the estimated age weights and time effects together, we conclude that there was a sharp spike in mortality related to these time effects in 1919, when there was a global influenza epidemic.

The second type of time effect,  $k = 2$ , displays an interesting pattern: It has opposite impacts on the very young as on all others. A time effect of type  $k = 2$  that lowers the mortality of 2-year-olds by one log point is predicted to *raise* the mortality of 18-year-olds by about the same amount; impacts on older people are smaller but still raise mortality. In other words, time effects of type  $k = 2$  describe a process in which falling infant mortality and rising adult mortality, especially young-adult mortality, are two sides of the same coin. At least two explanations are possible. First is a selection pattern: Reductions in infant mortality may mainly save the lives of unhealthy people who will soon die anyway, so that saving in-

fants inevitably raises adult mortality. Second is a historical explanation: Technological and social changes that reduced infant mortality, such as better sanitation, may have happened in Sweden around the same time as other technological changes, such as the introduction of machinery and motor vehicles, that raise the “accident hump” frequently observed among young adults (Heligman and Pollard, 1980). Our model is not designed to discriminate between these or other possible explanations for the estimated pattern, but further investigation via other methods would be worthwhile. There is a downward spike in these time effects in 1919, which raises mortality of young adults and lowers mortality of infants. Combined with the impact of time effects of type 1, events in 1919 then have little impact on infants but increase the mortality of young adults, consistent with the findings of Noymer and Garenne (2000).

The third type of time effect,  $k = 3$ , has the longest-lasting effects, with a half-life of more than four years. It impacts mainly infants, with smaller impacts on the young and virtually no impact on the old. These findings are somewhat consistent with a theory of cohort effects in mortality, since they demonstrate that conditions early in life have lasting consequences. However, given the estimated half-life, the consequences do not necessarily last into old age. Thus, our findings fall in the middle of the debate between Finch and Crimmins (2004), who proposed the cohort morbidity phenotype hypothesis that suggests reductions in early-life mortality due to infections are associated with reductions in mortality at all subsequent ages for the same cohort, and Barbi and Vaupel (2005), who contend that cohort effects are unimportant.

The estimated sequences of time effects  $e_{k,t}$  show that mortality related to time effects of type  $k = 1$  and  $k = 3$  largely fell over the 20th century. For time effects of type  $k = 2$ ,

infant mortality largely fell, but the decrease was disrupted around the time of the Great Depression and World War II.

Figure 4 shows the observed mortality rates of several birth cohorts and the predicted age profile of mortality based on conditions in each cohort's birth year. Forecasting is not our goal, so differences between the predicted and observed outcomes do not reflect a failure of our model; rather, these differences are interesting because they show us how conditions changed over a cohort's life course. For the 1881 birth cohort, the predicted age profile closely matches the observed mortality rates, showing that conditions changed little over the cohort's life course. The findings for the 1881 cohort are in sharp contrast with those for later cohorts, where observed mortality at most ages is strikingly lower than the predicted age profile based on conditions in the cohort's birth year. The gap between predicted and observed mortality grows larger with each successive cohort, suggesting not only that conditions improved during each cohort's life course but also that the rate of improvement grew over time.

Figure 5 shows the predicted age profiles of mortality based on conditions in several birth years. The graph allows us to examine historical declines in mortality through the lens of age profiles — that is, we can analyze how the effect of age on mortality has changed over time. Except for 1881 conditions, we see continuous improvements over time in mortality at all ages. However, the improvements are larger at young adult ages, and improvements in mortality among people in their 50s and 60s are quite small until recent years. Thus, the rate of increase of mortality with age among adults became more steep from 1881 to 1941, but since then the rate of increase of mortality with age among adults has been roughly constant. In the future, it would be worthwhile to extend this analysis to ages beyond 64 when data quality permits.



## 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 in many applications, 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 directly modeling the process that generates cohort effects: As they age, cohorts are continuously exposed to influences that cumulatively alter their trajectories. Our substantive model of cohort effects is what allows identification in our model, because the model restricts the possible forms that cohort effects can take. As an example, our data analysis illustrates the utility of our model in studying the evolution of human mortality. We believe that, beyond demography, the model can find application in economics, sociology, and other social sciences and can potentially provide new stylized facts that are useful for explaining and evaluating theories of social change and structure.

## Appendix

### A1. Proof of 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  $X_K \subset \Theta$  such that  $X_K$  is of measure zero and such that, unless  $\theta = [\mu, \{\{e_{k,t}\}_{t=1-A}^T, \{w_{k,a}\}_{a=1}^A, r_k\}_{k=1}^K]$  is in  $X_K$ , the equality

$$\mu + \sum_{k=1}^K \sum_{a'=0}^a r_k^{a-a'} w_{k,a'} e_{k,t-a+a'} = \tilde{\mu} + \sum_{k=1}^K \sum_{a'=0}^a \tilde{r}_k^{a-a'} \tilde{w}_{k,a'} \tilde{e}_{k,t-a+a'}, \quad a = 0, \dots, A, \quad t = 1, \dots, T, \quad (\text{A1})$$

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

*Case 1:  $K = 1$ .* Let  $X_1$  be the set of  $\theta \in \Theta$  such that either  $r_1 + w_{1,1} = 1$  or the vectors  $(e_{1,1}, \dots, e_{1,T-1})$  and  $(e_{1,2}, \dots, e_{1,T})$  are collinear with a constant.  $X_1$  is a set of measure zero. Assume  $\theta \in \Theta \setminus X_1$ . Specializing (A1) to  $K = 1$ ,  $a = 0$ , and  $a = 1$  (by hypothesis,  $A \geq 1$ ) and using normalization 2, we have

$$\mu + e_{1,t} = \tilde{\mu} + \tilde{e}_{1,t}, \quad t = 1, \dots, T, \quad (\text{A2a})$$

$$\mu + r_1 e_{1,t-1} + w_{1,1} e_{1,t} = \tilde{\mu} + \tilde{r}_1 \tilde{e}_{1,t-1} + \tilde{w}_{1,1} \tilde{e}_{1,t}, \quad t = 2, \dots, T. \quad (\text{A2b})$$

Substituting (A2a) into (A2b) and collecting terms gives

$$0 = (\mu - \tilde{\mu})(1 - \tilde{r}_1 - \tilde{w}_{1,1}) - (\tilde{r}_1 - r_1)e_{1,t-1} - (\tilde{w}_{1,1} - w_{1,1})e_{1,t}, \quad t = 2, \dots, T. \quad (\text{A3})$$

By hypothesis,  $T \geq 4$ , so (A3) 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, (A3) 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 (A2a) that  $\tilde{e}_{1,t} = e_{1,t}$  for  $t = 1, \dots, T$ . Finally, substituting the foregoing results into (A1) 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:

$$\begin{aligned}
X_{2,1} &= \{\boldsymbol{\theta} \in \Theta : \exists k \text{ s.t. } r_k = 0\}, & X_{2,2} &= \{\boldsymbol{\theta} \in \Theta : r_1 = r_2\}, \\
X_{2,3} &= \{\boldsymbol{\theta} \in \Theta : w_{1,1} = w_{2,1}\}, \\
X_{2,4} &= \left\{ \boldsymbol{\theta} \in \Theta : \text{rank} \begin{bmatrix} 1 & \begin{pmatrix} e_{k,j} \\ \vdots \\ e_{k,T-4+j} \end{pmatrix}_{\substack{k \in \{1,2\}, \\ j \in \{1,2,3,4\}}} \\ 1 \end{bmatrix} < 9 \right\}, & & \text{(A4)} \\
X_{2,5} &= \{\boldsymbol{\theta} \in \Theta : (w_{1,1} - w_{2,1})[-r_2 w_{1,1} + r_1 w_{2,1}] + (w_{1,2} - w_{2,2})(r_2 - r_1) = 0\}, \\
X_{2,6} &= \{\boldsymbol{\theta} \in \Theta : w_{1,1} - w_{2,1} + r_1 - r_2 + (r_2^2 + r_2 w_{2,1} + w_{2,2})(1 - r_1 - w_{1,1}) \\
&\quad - (w_{1,2} + r_1 w_{1,1} + r_1^2)(1 - r_2 - w_{2,1}) = 0\}.
\end{aligned}$$

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  $\boldsymbol{\theta} \in \Theta \setminus X_2$ , then the unique solution to (A1) is  $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$ . The algebra proceeds by using (A1) 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 (A1) at  $a = 2$  to eliminate  $\tilde{e}_{1,t}$ ; and observing that coefficients in a linear combination of a constant with  $\{e_{k,t-3}, \dots, e_{k,t}\}_{k=1}^2$  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.  $\square$

## A2. Proof of Proposition 2

We must show that under each of conditions (a) and (b), (A1) 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  $\{w_{1,a}\}_{a=1}^A$ :

$$\sum_{a'=1}^a (r^*)^{a-a'} w_{a'}^* = -(r^*)^a + \sum_{a'=0}^a r_1^{a-a'} w_{1,a'}, \quad a = 1, \dots, A. \quad (\text{A5})$$

Given  $e_{1,t} = \bar{e}_1$ , the following solves (A1):  $\tilde{\mu} = \mu$ ;  $\tilde{e}_{j,t} = e_{j,t} \forall j, t$ ;  $\tilde{r}_1 = r^*$ ;  $\tilde{r}_j = r_j \forall j > 1$ ;  $\tilde{w}_{1,a} = w_a^* \forall a$ ;  $\tilde{w}_{j,a} = w_{j,a} \forall j > 1, a$ . Therefore, (A1) 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 (A1):  $\tilde{\mu} = \mu$ ;  $\tilde{r}_j = r_j \forall j$ ;  $\tilde{w}_{1,a} = xw_{1,a} + (1-x)w_{2,a} \forall a$ ;  $\tilde{w}_{2,a} = (1-x)w_{1,a} + xw_{2,a} \forall a$ ;  $\tilde{e}_{1,t} = \frac{(1-x)e_{2,t} - xe_{1,t}}{1-2x} \forall t$ ;  $\tilde{e}_{2,t} = \frac{(1-x)e_{1,t} - xe_{2,t}}{1-2x} \forall t$ ;  $\tilde{e}_{j,t} = e_{j,t} \forall j > 2, t$ ;  $\tilde{w}_{j,a} = w_{j,a} \forall j > 2, a$ . Therefore, (A1) has a continuum of solutions indexed by  $x \in (1/2, 1]$ .  $\square$

### A3. Proof of Proposition 3

We assume the distribution of  $\epsilon_{at}$  satisfies regularity conditions such that a uniform law of large numbers (ULLN) holds. Case (a): If  $\sigma^2 = 0$ , (14) becomes (A1); hence the true parameters uniquely solve (14) when  $\sigma^2 = 0$ . Since the objective function in (14) is continuous, a ULLN applies, and solutions for  $\sigma^2 > 0$  converge to the solution for  $\sigma^2 = 0$ . Case (b): The predicted values can be written as  $\hat{\mathbf{y}}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu}, \tilde{\mathbf{e}}) = \mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\tilde{\mathbf{e}}$ . Hence if we solve (14) for  $\hat{\mathbf{e}}$  as a function of the remaining parameters, we obtain  $\hat{\mathbf{e}} = [\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})]^{-1}\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'(\mathbf{y} + \boldsymbol{\epsilon})$ . (Interpret the inverse as a generalized inverse when  $\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})$  is not of full rank.) Substituting this solution into (14), we obtain

$$(\hat{\mathbf{w}}, \hat{\mathbf{r}}, \hat{\mu}) = \arg \min_{\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu}} \frac{1}{T} [\mathbf{y}'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\mathbf{y} + 2\mathbf{y}'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})\boldsymbol{\epsilon}], \quad (\text{A6})$$

where  $\mathbf{M}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu}) = \mathbf{I} - \mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})[\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})]^{-1}\mathbf{X}(\tilde{\mathbf{w}}, \tilde{\mathbf{r}}, \tilde{\mu})'$  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  $T \rightarrow \infty$ , the second term in the objective function converges uniformly in probability to zero. Further, since  $\epsilon_{at}$  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  $T \rightarrow \infty$ , the true parameters minimize the third term. Further, in the limit as  $T \rightarrow \infty$ , 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.

Thus  $(\hat{\mathbf{w}}, \hat{\mathbf{r}}, \hat{\mu})$  converges in probability to the true parameters.  $\square$

#### A4. 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  $N_{at} \rightarrow \infty$ . (The smallest cell in our data has  $N_{at} = 19,856$ , and the median cell has  $N_{at} = 92,794$ , so approximating the distribution by the limit as  $N_{at} \rightarrow \infty$  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  $a = 0, \dots, A$  and years  $t = 1, \dots, T$  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  $\hat{\sigma}_{at}^{-2} = N_{at}\bar{p}_{at}/(1 - \bar{p}_{at})$ . 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

- Barbi, Elisabetta, and James W. Vaupel, 2005, "Comment on 'Inflammatory Exposure and Historical Changes in Human Life-Spans,'" *Science* 308(5729), 1743a.
- Deaton, Angus, and Christina Paxson, 1994, "Intertemporal Choice and Inequality," *Journal of Political Economy* 102(3), 437–467.
- Elo, Irma T., 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.
- Finch, Caleb E., and Eileen M. Crimmins, 2004, "Inflammatory Exposure and Historical Changes in Human Life-Spans," *Science* 305(5691), 1736–1739.
- Greenberg, B.G., John J. Wright and Cecil G. Sheps, 1950, "A Technique for Analyzing Some Factors Affecting the Incidence of Syphilis," *Journal of the American Statistical Association* 45(251), 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(1), 4–43.
- Horiuchi, Shiro, 1983, "The Long-Term Impact of War on Mortality: Old-Age Mortality of the First World War Survivors in the Federal Republic of Germany," *Population Bulletin of the United Nations* 15, 80–92.
- Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at [www.mortality.org](http://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.
- Lee, Ronald D., and Lawrence R. Carter, 1992, "Modeling and Forecasting U.S. Mortality," *Journal of the American Statistical Association* 87(419), 659–671.
- 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(2), 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(3), 565–581.
- Omran, Abdel R., 1982, "Epidemiologic transition," in *International Encyclopedia of Population*, pp. 172–183, New York: Free Press.
- Robertson, Chris, Sara Gandini, and Peter Boyle, 1999, "Age-Period-Cohort Models: A Comparative Study of Available Methodologies," *Journal of Clinical Epidemiology* 52(6), 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, 1997, *United Nations Demographic Yearbook 1997*, historical supplement, table 11, New York: United Nations.
- United Nations, 2003, "Long-Range Population Projections: Proceedings of the United Nations Technical Working Group on Long-Range Population Projections," available at [http://www.un.org/esa/population/publications/longrange/long-range\\_working-paper\\_final.PDF](http://www.un.org/esa/population/publications/longrange/long-range_working-paper_final.PDF).
- Vaupel, James W., Kenneth G. Manton, and Eric Stallard, 1979, "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography* 16(3), 439–454.

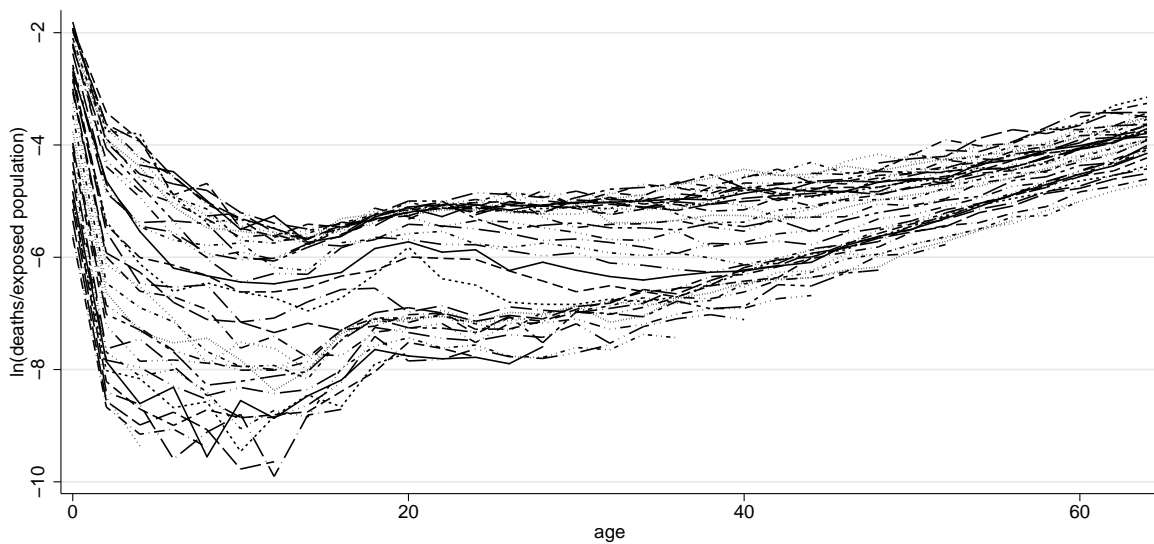


Figure 1: Mortality in Sweden, 1861–2005.

Each line shows the realized mortality of a particular birth cohort at various ages. Cohorts included are those born in 1781, 1783, ..., 2005. Lines for cohorts born before 1861 or after 1925 omit some ages because the dataset does not cover those ages for those cohorts. Data source: Human Mortality Database (2007).

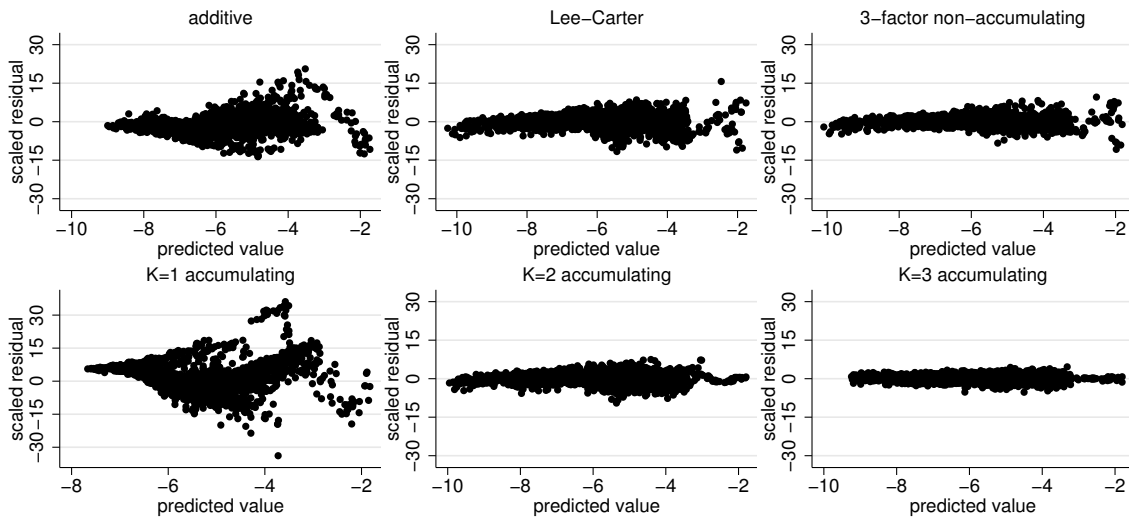


Figure 2: Scaled residuals vs. predicted values.

Graphs plot scaled residuals  $(\bar{y}_{at} - \hat{y}_{at})/\hat{\sigma}_{at}$  against predicted values  $\hat{y}_{at}$  for each model estimated.

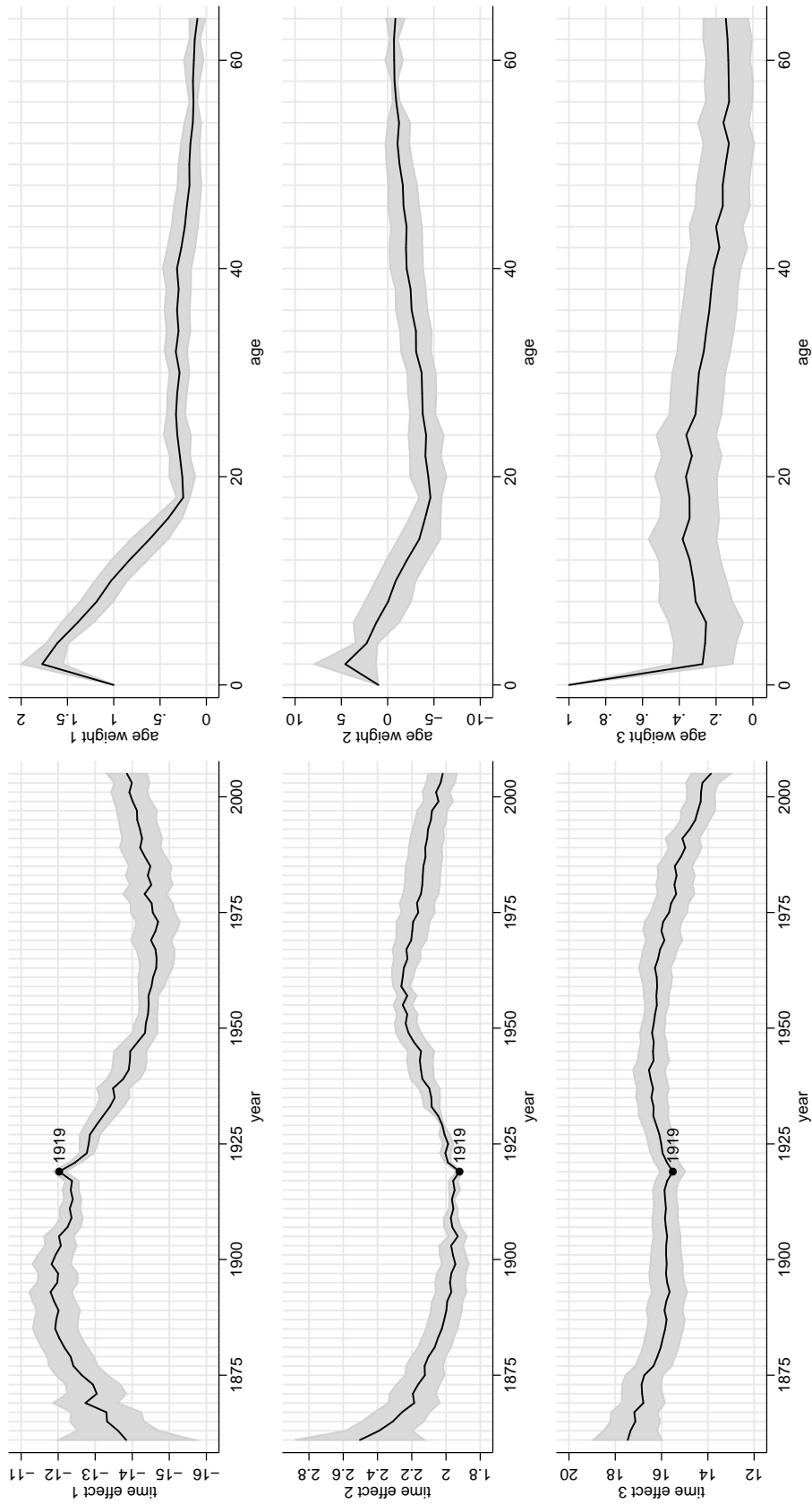


Figure 3: Estimated time effects and age weights.

Estimates from  $K = 3$  continuously accumulating model. Black lines are point estimates; gray lines are pointwise 95% confidence intervals.

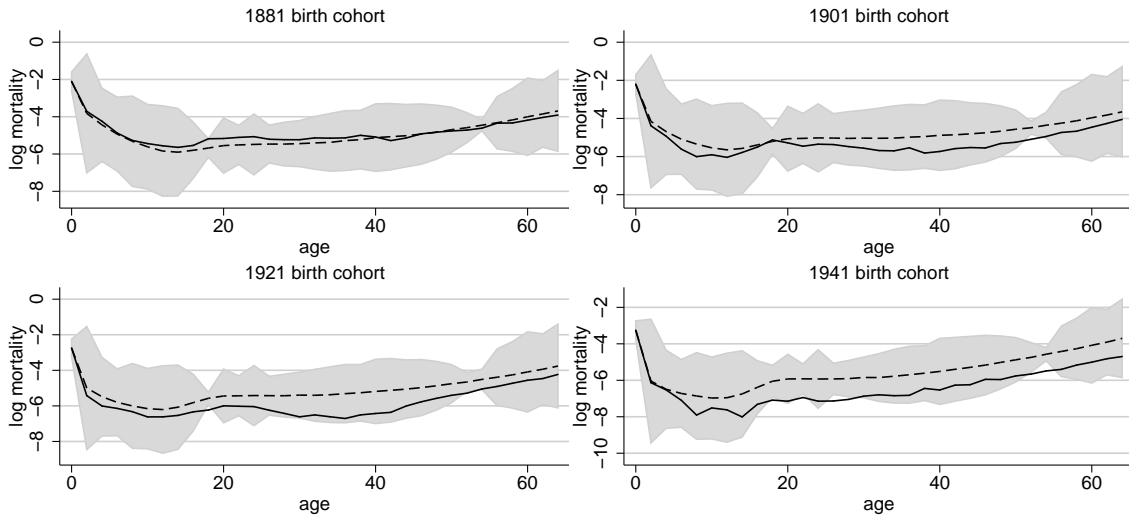


Figure 4: Predicted age profiles vs. actual mortality.

Each solid line shows the observed mortality of a particular birth cohort at various ages. The dashed line represents the mortality that the  $K = 3$  continuously accumulating model would predict for that cohort if conditions in its birth year continued throughout its life span.

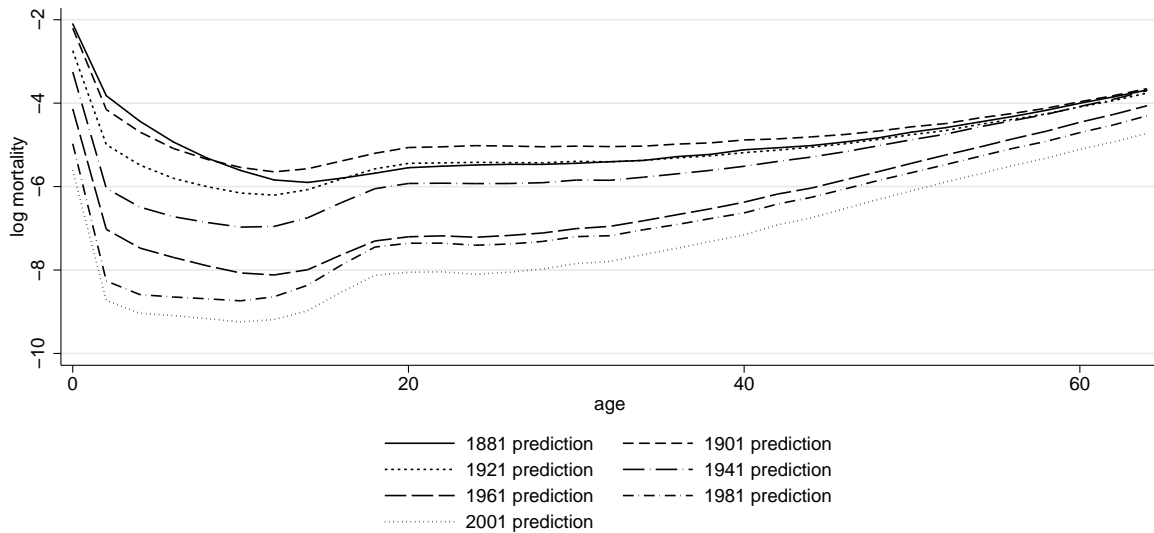


Figure 5: Comparing predicted age profiles.

Each line represents the mortality that the  $K = 3$  continuously accumulating model would predict for the cohort born in a given year if conditions in the birth year continued throughout the cohort's life span.

Table 1: Goodness-of-fit statistics for six models.

|                | continuously accumulating |        |        |        | Lee-Carter | 3-factor  |
|----------------|---------------------------|--------|--------|--------|------------|-----------|
|                | additive                  | K=1    | K=2    | K=3    |            | nonaccum. |
| log likelihood | -7728                     | -9321  | -6446  | -5163  | -7048      | -5969     |
| AIC            | 15872                     | 18920  | 13444  | 11150  | 14371      | 12567     |
| BIC            | 17075                     | 19724  | 15041  | 13535  | 15164      | 14390     |
| weighted $R^2$ | 0.9900                    | 0.9561 | 0.9965 | 0.9984 | 0.9943     | 0.9976    |
| dispersion     | 14.4                      | 52.5   | 5.13   | 1.89   | 7.94       | 3.52      |
| cells          | 2409                      | 2409   | 2409   | 2409   | 2409       | 2409      |
| parameters     | 208                       | 139    | 276    | 412    | 137        | 315       |
| residual d.f.  | 2201                      | 2270   | 2133   | 1997   | 2272       | 2094      |

$R^2$  calculated using the estimation weights shown in appendix A4. Data source: Human Mortality Database (2007).

Table 2: Estimated rates of decay for  $K = 3$  continuously accumulating model.

| parameter | estimate | standard error | half-life (years) |
|-----------|----------|----------------|-------------------|
| $r_1$     | 0.138    | 0.0000         | 0.35              |
| $r_2$     | 0.582    | 0.0005         | 1.28              |
| $r_3$     | 0.845    | 0.0005         | 4.11              |

Rates of decay are for one-year intervals. Half-life in years  $t_{1/2}$  solves  $r_k^{t_{1/2}} = 1/2$ . Data source: Human Mortality Database (2007).