Population Center Working Papers (PSC/PARC)
Population Studies Center

5-30-2020

Age-Period-Cohort Analysis: What Is It Good For?

Herbert L. Smith
University of Pennsylvania, hsmith@pop.upenn.edu

Recommended Citation
Smith, Herbert 2020. "Age-Period-Cohort Analysis: What Is It Good For?." *University of Pennsylvania Population Center Working Paper (PSC/PARC)*, 2020-46. https://repository.upenn.edu/psc_publications/46.

Paper prepared for the volume *Age, Period and Cohort Effects: The Identification Problem, and What to Do about It*, edited by Andrew Bell. Routledge. Version of 30 May 2020. Thanks to Theodore Holford for a helpful reading, and to Tomáš Sobotka for his generosity in providing the data and graphics that appear here as Figures 4 and 5.

Age-Period-Cohort Analysis: What Is It Good For?

Abstract

If you know when someone was born, and you know what time it is, you know how old they are. If you know how old someone is and when they were born, you know the date on which they are being observed. If you know someone's age as of a given time, you know when they were born. These are ineluctable features of algebra (age ≡ period – cohort) and geometry, as reflected in the Lexis diagram (Chauvel 2014, 384-389). There are many ways that one can turn the problem (e.g., cohort ≡ period – age) and thus many alternative forms of observation, classification, and depiction. However, there is a strong statistical sense in which there are only two pieces of information, not three.

Keywords

age, cohort analysis, age-period-cohort model

Disciplines

Demography, Population, and Ecology | Family, Life Course, and Society | Social and Behavioral Sciences | Sociology

Comments

Paper prepared for the volume *Age, Period and Cohort Effects: The Identification Problem, and What to Do about It*, edited by Andrew Bell. Routledge. Version of 30 May 2020. Thanks to Theodore Holford for a helpful reading, and to Tomáš Sobotka for his generosity in providing the data and graphics that appear here as Figures 4 and 5.

Age-Period-Cohort Analysis: What Is It Good for?

Herbert L. Smith (https://orcid.org/0000-0001-8974-2788)
Department of Sociology and Population Studies Center
University of Pennsylvania
3718 Locust Walk
Philadelphia, Pennsylvania 19104-6298 USA
hsmith@pop.upenn.edu

Paper prepared for the volume *Age, Period and Cohort Effects: The Identification Problem, and What to Do about It*, edited by Andrew Bell. Routledge. Version of 30 May 2020. Thanks to Theodore Holford for a helpful reading, and to Tomáš Sobotka for his generosity in providing the data and graphics that appear here as Figures 4 and 5.

If you know when someone was born, and you know what time it is, you know how old they are. If you know how old someone is and when they were born, you know the date on which they are being observed. If you know someone's age as of a given time, you know when they were born. These are ineluctable features of algebra (age ≡ period – cohort) and geometry, as reflected in the Lexis diagram (Chauvel 2014, 384-389). There are many ways that one can turn the problem (*e.g.*, cohort ≡ period – age) and thus many alternative forms of observation, classification, and depiction. However, there is a strong statistical sense in which there are only two pieces of information, not three.

This chafes, because in sociology and other social and population sciences, it is not hard to conjure for each temporal dimension "a distinct causal interpretation" (Mason *et al*. 1973, p. 243). A primordial example is political party identification (Oppenheim 1970):

> … [T]he political environment in which a particular birth cohort first enters the electorate may help determine the extent to which individuals in that cohort identify with a political party for the remainder of their lives. As that party experiences normal fluctuations in political fortunes, however, some members of the cohort may temporarily shift their loyalties. Both cohort and short-term period effects can thus contribute to party identification. Since the aging process might also independently affect party identification (as persons become more "conservative" with age, for example, they may find the Republicans increasingly attractive), we have here another example in which age, period and cohort conceptually have distinct causal impacts on the dependent variable. (Mason *et al*. 1973, 244-245).

Persons reaching their majority in the New Deal era may have been more likely to be Democrats than those in cohorts before or after. Net of this, aging into mid-life may have pushed them more toward the Republicans. With a bit more precision, you can know that we are looking at party identification circa 1964, when the Republicans nominated a candidate viewed as too conservative to win, and the Democrats benefited from the holdover popularity of a recently assassinated young president. The period under observation is knowable from the age and the birth cohort of a set of respondents. But there is nothing about the events of the period shaping party identification that follow necessarily from the experience of this cohort in early adult life or the leavening effects of prosperity and responsibility on youthful idealism. The temporal markers are linear functions of one another. Their manifestations are not.

On Cohort Analysis

"Any quantitative *cohort* analysis is a form of time-series analysis" (Fienberg and Mason 1985, 85 [emphasis added]). A corollary is that the concept of a cohort is what is distinct about age-graded data analyzed over historical time. Keeping this in mind helps in specifying models and analyses that convey meaning in the presence of a linear identity that is algebraically tautological and sociologically meaningless.

In sociology, the foundational paper on cohort analysis is Ryder (1965). At the time of the publication of Mason et al. (1973), it hit a steady pace of citations (**Figure 1**) that persisted through its re-publication in the volume on *Cohort Analysis in Social Research: Beyond the Identification Problem* (Mason and Fienberg 1985) to 2005, when it was recognized as being among the "greatest hits" of the *American Sociological Review* (Jacobs 2005). This was also the time when it became clear that the field had not yet moved beyond the identification problem. Citations to Ryder (1965) accelerated coincident with a new wave of papers on issues of identification in age-period-cohort analysis—in particular, the Intrinsic Estimator (Yang *et al.* 2004). This is notwithstanding that Ryder (1965) is not a technical paper and does not mention the identification problem. The paper concludes with a summary statement that does sound like a call for the estimation of something like cohort effects: "The purpose of this essay is to direct the attention of sociologists toward the study of time series of parameters for successive cohorts of various types, in contradistinction to conventional period-by-period analyses" (861).

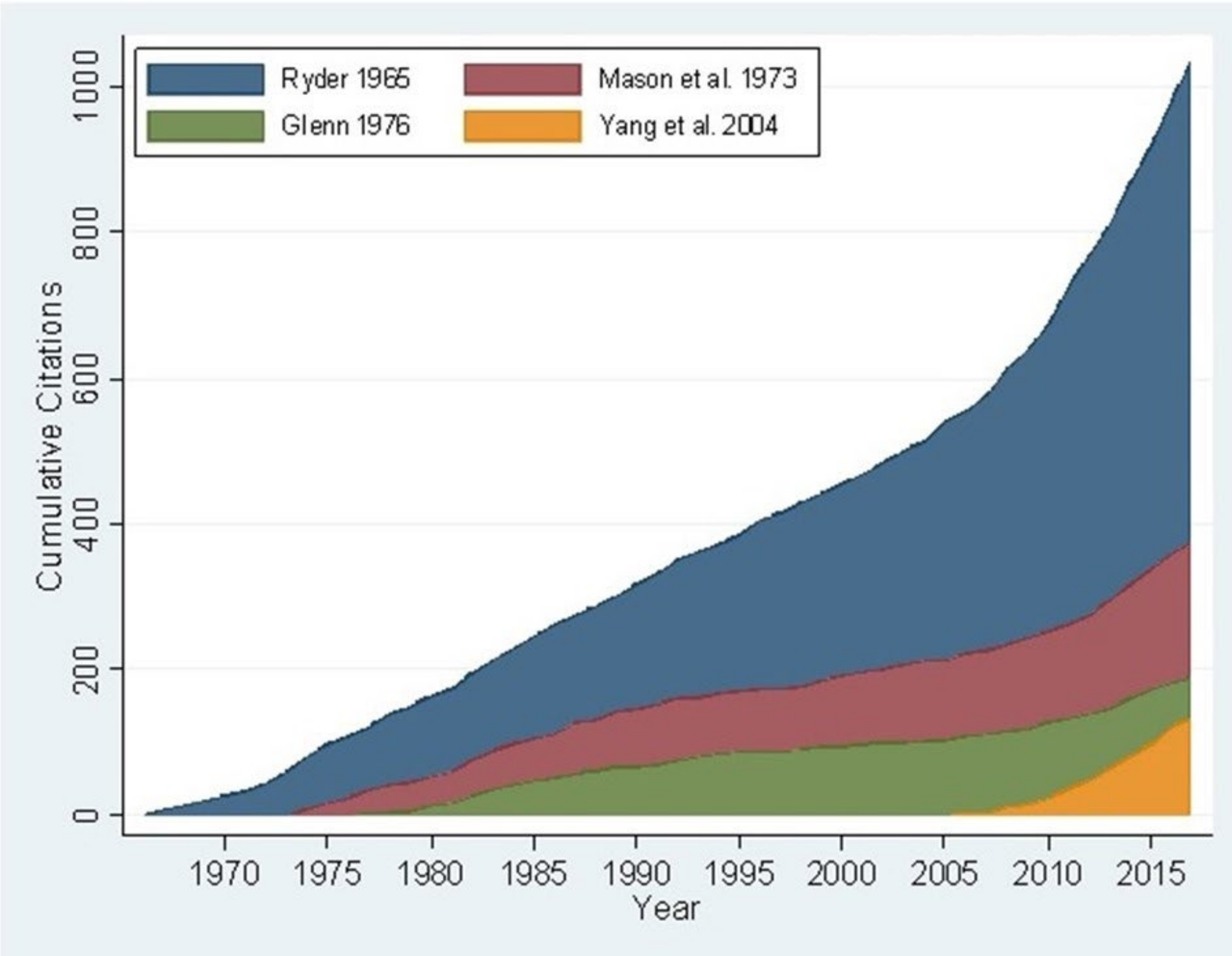


Figure 1. *Cumulative Web of Science Citations to Ryder (1965) and to Selected Subsequent Papers in Sociology on Cohort Analysis.*

But the methodological concerns that are touched on subsequently concern the onerous demands for data collection and the difficulty of making comparisons across cohorts under conditions of differential selectivity, and are, in the event, essentially second-order: "Yet such difficulties are not so much those of the method itself as meaningful reflections of the research investment necessary to study a long-lived species experiencing structural transformation" (861). People live a long time, but they don't live forever. Experiences at early ages can have profound effects. Events at a given time can have differential effects on individuals at different ages. Social change via the succession of differentiated cohorts variously implicates the distinction between changes in an individual and changes in a population; socialization and social control across the life course; the adaptiveness of a society versus the "limited intellectual flexibility" (844) of an individual; and the role of a cohort as the embodiment and historical representation of an admix of social events and individual experiences. Ryder's (1965) synthesis of ideas from the sociology of generations (Mannheim 1952), demography, history, developmental psychology, and what is now known as life course analysis has become such a touchstone for the concept of cohort analysis that is easy to miss something that he does not talk about: *linear trends*. Nor are they present in analyses that are faithful to this conceptual perspective.

Consider Yang's (2016) cohort biography of the Red Guard generation, the Chinese who were teens and young adults at the time of the 1966-68 Cultural Revolution. They were the first cohort raised under Communism. They were mobilized to fight against functionaries in the Cultural Revolution, were subsequently exiled to the countryside, and were variously rehabilitated, de-radicalized, and re-integrated in a fraught journey that shaped and re-shaped Chinese history—and their cohort.

> … I use terms such as *trajectories*, *journey*, and *life course* to talk about the history of the Red Guard generation. These words may convey a sense of linear progress, as if, from the time of birth, members of the Red Guard generation were destined to march toward a clear, fixed, and grand goal…. By analyzing the longer history of the Red Guard generation, which will highlight the many ups and downs of the generation, I … will show the futility of grand teleological perspectives for understanding history. There is neither linearity nor teleology to the trajectory of the Red Guard generation, or perhaps other political generations in other times and places…. For the protagonists of my story, the history of a generation was nothing less than a history of perpetual disruption of personal lives. (Yang 2016, 5)

Yang's (2016) study, which pertains to a single cohort and is non-quantitative at that, is an example of the kind of "composite cohort biograph[y]" whose comparison with other cohorts "would yield the most direct and efficient measurement of the consequences of social change" (Ryder 1965, 847). Which can be done quantitatively. The radicalizing political and social events in China circa 1968 had contemporaneous parallels globally, including in France. Thus it is not surprising that *les soixante-huitards*—the generation that came of political age, at least emblematically, among the civil unrest of May 1968, figures often, via comparison of cohort "destinies," in Chauvel's (2014) analysis of changes in French social structure across the 20th Century (and into the current one). This was a generation that turned from street protest to electoral participation, displacing in the legislative elections of 1981 prior generations of politicians whose youth in the aftermath of the Second World War had provided *them* with opportunities given the number of *their* elders tarnished by the war and the Occupation. The *soixante-huitards* thereupon encrusted themselves in the French body politic in a manner that left scant room for subsequent generations (Chauvel 2014, 36-43).

This is a form of "generational domination" (Chauvel 2014, 10) that highlights not just the social, historical, and political biographies of generations, but their interplay. In France, a cohort's experience at a given point in time reflects not just its past and its development, but the opportunity structure of a period as conditioned by the experience of *other* cohorts at other times. Policy decisions made at one time for one reason—on the expansion of education, for example—can be a boon for one generation, to the detriment of those that follow (Chauvel 2014, 232-239).

Generational dominance, or the hoarding of opportunities, is only one way in which cohorts are useful for understanding how historical events re-shape social structure. "To some extent all cohorts respond to any given period- specific stimulus. Rarely are changes so localized in either age or time that their burden falls exclusively on the shoulders of one cohort" (Ryder 1965, 847).

So which cohorts shoulder the burden—or reap the rewards? The Nineteenth Amendment to the United States Constitution, enacted in 1920, extended to women the right to vote. The degree to which women took up the franchise depended on how long they had been living in a polity in which voting was forbidden to them. The longer they had lived under the old regime, the more they seem to have been inculcated with the idea that voting is not for women, and the less likely they were to vote during the remainder of their lives. We know this because voting among women, *relative to voting among same-generation men*, picked up across cohorts in function of the age of a cohort in 1920, to the point where women who were either too young in 1920 to have been aware of their disenfranchisement or who were born after 1920 could no longer be distinguished from men in their propensity to vote (Firebaugh and Chen 1995). The power of the cohort as a tool for understanding social change is reinforced by another feature of this study: The data on voting that are analyzed are for national elections between 1952 and 1988, long after the historical event whose effects are being inferred. The concept of a cohort thus provides both a window on the past… and a glimpse into the future.

I once wrote that sociologists and demographers are "mad for cohorts" (Smith 2008, 289), which is infelicitous insofar as it implies that there is something either frivolous or romantic in the frequency with which the concept figures in our research. There can be a real power to cohort analysis. At the same time, we should not reify the construct unduly. Ryder (1965) was simultaneously of two minds about this:

> As a minimum, the cohort is a structural category with the same kind of analytic utility as a variable like social class. Such structural categories have explanatory power because they are surrogate indices for the common experiences of many persons in each category. Conceptually the cohort resembles most closely the ethnic group: membership is determined at birth, and often has considerable capacity to explain variance, but need not imply that the category is an organized group (847).

So is a cohort like a social class, or isn't it? An occupation indexes a class in the same sense that a birthdate indexes a cohort, but in the former instance the analytic category entails interests as well, if not self-consciousness (Wright 1997). There may indeed be historical cohorts that are like social classes in that sense—the Red Guard generation being an obvious example, perhaps the *soixante-huitards* as well—but the women who were sentient at the time of the Nineteenth Amendment probably did not constitute cohorts in the same fashion. Their birthdates are just a way of keeping track of them.

> From this view comes an operating conception of a cohort, conceived not as a concrete group, but as a possible key to understanding social change. As a result it is very different from the concept of social class: It is based above all on its technical construction and constitutes an instrument of objectification. The cohort is thus a tool and not necessarily a strong element of the theoretical apparatus of sociology: It's characterization as a relevant group can only be a result—true, false, or somewhere in between—according to the purpose [of the exercise], and not an a priori hypothesis. (Chauvel 2014, 383 [my translation])

"'Cohort analysis,' after all, is a means and not an end in itself" (Duncan 1985, 300); the proof of the pudding is inevitably in the tasting (Smith, Mason, and Fienberg 1982, 792). Many of the perceived analytic problems associated with so-called age-period-cohort analysis recede if we keep the cohorts in the fore, not for any special statistical reason, but because of the conceptual utility for reading the data. In the sections that follow, I revisit from this perspective several of the key issues in the identification of these models, and conclude with some orienting suggestions. There is no panacea, but this has less to do with the remorselessness of algebra and geometry than it does with the absence of an ailment. Once one stops thinking about events (or is it their causes?) flowing simultaneously (and, implicitly, infinitely) through historical time in two non-concomitant directions, things improve considerably.

Identification Entails a Constraint on Linear Terms

The mésalliance between an algebraic identity and theoretical aspirations put a bad curse on *statistical* treatments of "age-period-cohort analysis." This is unfortunate. From the beginning, it was recognized that the algebraic identity which reduces three dimensions to two obtains only with respect to the *linear* terms in age, period, and cohort. Thus whereas

$$Y = \alpha + \beta_1 A + \beta_2 P + \beta_3 C + \varepsilon \tag{1}$$

is not estimable,

$$Y = \alpha + \beta_1 A^2 + \beta_2 P + \beta_3 C + \varepsilon \quad (2)$$

Is estimable, and under certain circumstances it might be appropriate to think about models in which at least one of the three temporal dimensions is not linearly related to the response:

> For example, if we were studying age-period-specific fertility rates as a cohort problem, we clearly would not want to specify a linear relationship between fertility rates and age.... Women do not, we know, have a higher probability of a birth with each additional year that they age. (Mason *et al*. 1972, 7)

This led to a consideration of "relatively functional free" (Mason *et al*. 1973, 246) specifications linking age, period, and cohort to outcomes of interest based on multiple classification analysis (sets of categorical variables representing each temporal dimension). Parsing the temporal dimensions into categorical indicators does not in and of itself "solve" the estimation issue posed by the linear identity $A \equiv P - C$. Rather, the linear identity means that not even differences between category coefficients are uniquely estimable (Mason *et al.* 1973, 246-247). The algebra and geometry of this point have been made many times since, and there is nothing that I will or can say that will change them. Nonetheless, since a theme of this chapter is that it is rarely useful to *think* of the *linear* effects of period and cohort as coexisting independent of one another, it is useful to have a perspective and notation that explicitly distinguishes linear from non-linear effects. These I borrow liberally from Holford (1983), and adapt slightly.

Consider the classic data array in which there are I age-groups observed across J periods.[1] The general representation of the multiple classification analysis framework is

---

[1] I assume that age-groups conform to intervals between periods—*i.e.*, if periods are observed every five years, then age categories span five years—although all the points here generalize to more complicated data arrays and observation schemes (Fienberg and Mason 1978, 37-42).

(3) $$g(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k$$

where y_{ijk} is some measure (count, rate, *etc*.) at age i in period j hence cohort $k = j - i + J$.[2] A standard method for identifying the sets of effects $\{\alpha_i, \beta_j, \gamma_k\}$ is via effect coding:

(4) $$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{k=1}^{K} \gamma_k = 0$$

This is, however, insufficient for resolving the identification issue in the situation in which we are seeking to maintain representations of all three temporal dimensions (Fienberg and Mason 1985, 67-68). To see this, create linear re-scalings of the dimensional indexes

(5a) $$A_i = i - \frac{I+1}{2} \qquad \left(\Rightarrow \sum_{i=1}^{I} A_i = 0 \right)$$

(5b) $$P_j = j - \frac{J+1}{2} \qquad \left(\Rightarrow \sum_{j=1}^{J} P_j = 0 \right)$$

(5c) $$C_k = k - \frac{K+1}{2} \qquad \left(\Rightarrow \sum_{k=1}^{K} C_k = 0 \right)$$

and use them to redefine the $\{\alpha_i, \beta_j, \gamma_k\}$ terms as

(6a) $$\alpha_i = \alpha^L A_i + \alpha_i^d$$

(6b) $$\beta_j = \beta^L P_j + \beta_j^d$$

(6c) $$\gamma_k = \gamma^L C_k + \gamma_k^d \qquad .$$

---

[2] The function $g(y_{ijk})$ implies the generalized linear model (McCullagh and Nelder 1989). The indexing of cohort (k) assumes that ages are indexed from youngest to oldest, and periods from most distant to most recent.

Each term $\{\alpha_i,\ \beta_j, \gamma_k\}$ is thus a function of a constant rate of change in, respectively, age (α^L), period (β^L), and cohort (γ^L); plus a corresponding deviation $\{\alpha_i^d, \beta_j^d, \gamma_k^d\}$ from the linear trend. These new terms are what would be estimated were we somehow to have known (or estimated) the sets of category-specific coefficients $\{\alpha_i,\ \beta_j, \gamma_k\}$ and regressed them on their respective linear locations $\{A_i, P_j, C_k\}$.

A corresponding implication is that

$$(7) \qquad \sum_{i=1}^{I} A_i \alpha_i^d = \sum_{j=1}^{J} P_j \beta_j^d = \sum_{k=1}^{K} C_k \gamma_k^d = 0$$

i.e., no covariance (correlation) between deviations (residuals) and trend (predictors).[3]

"[S]pecifying a model with linear and nonlinear effects does not solve the identification problem" (Fienberg and Mason 1978, 15), but it can keep clear what is at issue. Substitute equations $(6a-c)$ into equation (3), to yield

$$(8) \qquad g(y_{ijk}) = \mu + \alpha^L A_i + \beta^L P_j + \gamma^L C_k + \alpha_i^d + \beta_j^d + \gamma_k^d$$

There are three new terms in this functionally equivalent expression, but also, as per equation (7), three new restrictions—restrictions that render the category effects independent of their corresponding linear indexes. The model is still not identified, but any further single restriction on these parameters will make the model just-identified, with equivalent predicted values $\hat{y}_{ijk}$. Estimates of the category-specific (de-trended) deviations $\{\alpha_i^d,\ \beta_j^d,\ \gamma_k^d\}$ are unique and invariant regardless of the (single) identifying restriction that is chosen (Holford 1991, 432-433).

[3] Estimation of linear effects $\alpha^L, \beta^L, \gamma^L$ via regression is as per, *e.g.*, Holford (1991, 433). The normalization to 0 of both categorical effects (equation $[4]$) and linear trends (equations $[5a-c]$) eliminates the intercept term. Thus in the regression interpretation of equations $(6a-c)$, the deviations $\{\alpha_i^d, \beta_j^d, \gamma_k^d\}$ are akin to the respective error terms, not to intercepts.

Since $P_j \equiv A_i + C_k$, equation (8) is equivalently

$$(9) \qquad g(y_{ijk}) = \mu + (\alpha^L + \beta^L)A_i + (\gamma^L + \beta^L)C_k + \alpha_i^d + \beta_j^d + \gamma_k^d$$

Although all just-identified models will yield unique and invariant estimates of these additive combinations of coefficients on linear trend $\alpha^L + \beta^L$ and $\gamma^L + \beta^L$, the particular values of this set of coefficients $\{\alpha^L,\ \beta^L,\ \gamma^L\}$ will vary according to the identifying restriction that is employed.

Holford (1983) suggested identification by a constraint on a linear parameter itself, in particular period, as per

$$(10) \qquad \beta^L = 0\ ,$$

or zero linear trend (ZLT) in the period coefficients (O'Brien 2015, 51). This reduces equations (8) and (9) to

$$(11) \qquad g(y_{ijk}) = \mu + \alpha^L A_i + \gamma^L C_k + \alpha_i^d + \beta_j^d + \gamma_k^d$$

The assumption underlying this identifying restriction may or may not be accurate. If it is at all inaccurate, we can see from equation (9) that estimates for α^L and γ^L will be too high (low) to the extent that the assumption of no linear trend in period (equation $[10]$) is too low (high) relative to "whatever the true value of [β^L] happens to be" (Holford 1983, 316). Even though the deviations $\{\alpha_i^d, \beta_j^d, \gamma_k^d\}$ are not affected by this assumption, the category effects $\{\alpha_i,\ \beta_j, \gamma_k\}$ from the multiple classification schema (equations$[3]$ and$[4]$) will be, via equations $(6a - c)$.[4]

[4] The exceptions, as can be derived from equations $(6a - c)$ and equations $(5a - c)$, are the terms $\alpha_{\frac{I+1}{2}},\ \beta_{\frac{J+1}{2}}, \gamma_{\frac{K+1}{2}}$, which will exist when $I, J,$ and/or $K$are odd, and which will not vary under alternative identification assumptions. In one sense this is a trivial artifact of normalization assumptions (equations $[4]$ and $[5a - c]$). On the other hand, It does explain why plots of effects for given dimensions under alternative identifying restrictions rotate around a central value (*e.g.*, Holford 1991, Figure 1; O'Brien 2019, Figures 5-7).

All assumptions underlying just-identified age-period-cohort models are constraints on the linear coefficients $\{\alpha^L,\ \beta^L,\ \gamma^L\}$, if only implicitly. Thus identification by equating two adjacent terms in the multiple classification specification (Mason *et al*. 1973)—the constrained generalized linear model (CGLIM) (Yang and Land 2013, 66)—is equivalent to a constraint on the corresponding linear term as a function of the deviation terms. In general, for adjacent categories w and $w+1$, $\theta \in \{\alpha, \beta, \gamma\}$, and $V_w \in \{A_i, P_j, C_k\}$, from equations $(6a-c)$ we see that

$$(12) \qquad \theta_w = \theta_{w+1} \quad \Rightarrow \quad \theta^L V_w + \theta_w^d = \theta^L V_{w+1} + \theta_{w+1}^d \quad \Rightarrow \quad \theta^L = \theta_{w+1}^d - \theta_w^d \ .$$

This is *not* what any researcher to date has been arguing *explicitly* in estimating a CGLIM model, and yet there it is. The corollary is that any model identified by a constraint on one of the linear coefficients implies a CGLIM model. O'Brien (2019, 218) distinguishes identification via constraints on a pair of effect coefficients ("simple constraints") from "more complicated just identifying constraints such as setting the linear trend of the period effects to some value such as zero…". Equation (12) shows that these are two sides of the same coin.

Recall from equation (9) that two additive combinations of linear terms $\{\alpha^L,\ \beta^L,\ \gamma^L\}$ are estimable (*e.g.*, $\alpha^L + \beta^L$ and $\gamma^L + \beta^L$), so that "[i]f any one of the slopes is fixed at a particular value, then the other two are immediately determined, as well" (Holford 1991, p. 434). Any slope determined in one dimension must find some "compensation" in the other dimensions, subject to the two estimable (hence observable) sums of linear coefficients. This is why the literature is replete with demonstrations that different just-identified CGLIM models yield wildly varying estimates of $\{\alpha_i,\ \beta_j, \gamma_k\}$ (*e.g.*, Mason *et al.* 1973, Table 4; Kupper *et al*. 1985, Tables 3a-c; Holford 1991, Table 2; Yang, Fu, and Land 2004, Table 3).[5]

[5] But not estimates of $\{\alpha_i^d, \beta_j^d, \gamma_k^d\}$; these are unique (*e.g.*, Holford 1991, Table 2).

When William Mason and I (1985) reconstructed and extended Frost's (1939, Table 1) data on tuberculosis (TB) mortality in Massachusetts, we noted that classification of the non-pulmonary forms of TB had been changing over time (pp. 169-170), and that these non-respiratory tuberculoses were concentrated prior to adulthood (p. 177). In focusing on pulmonary TB alone—a disease classically associated with young adulthood—we were struck that mortality in both of the first two decades of life was comparatively low and virtually indistinguishable in level, whether viewed from a period or a cohort perspective (Mason and Smith 1985, 175-178; *cf*. Smith 2004, 114). In adopting the identifying restriction

$$(13) \qquad \alpha_{0-9} = \alpha_{10-19}$$

in our age-period-cohort analysis of pulmonary tuberculosis mortality, we ignored the implications of this restriction for the *linear* effects of age (hence period and cohort). Implicitly, however, we were fixing the linear effect of age as (from equation [12])

$$(14) \qquad \alpha^L = \alpha^d_{0-9} - \alpha^d_{10-19} \quad .$$

It is hard to think about what (if anything) the perspective afforded by equation (14) means *a priori* since, by construction (equation [7]), these category-specific deviation coefficients (α^d_{0-9}, α^d_{10-19}, …) are uncorrelated with the linear term for age (α^L). As it turns out, this was one instance where the implicit constraint on the linear term does not seem to have done any violence to the analysis, at least in terms of estimated parameters (*pace* Rodgers 1982, 785). Estimation of the parameters in equation (8), as a log-rate model fitted to the Massachusetts TB mortality data (Mason and Smith 1985, Tables A1 and A2)[6], with identifying restrictions as per equations (4), (7), and (14), yields $\widehat{\alpha^L} = -0.0148\ (\widehat{SE_{\alpha^L}} = 0.0051)$, $\widehat{\beta^L} = 0.0022\ (\widehat{SE_{\beta^L}} = 0.0053)$, and $\widehat{\gamma^L} = -0.0545\ (\widehat{SE_{\gamma^L}} =$

[6] *I.e.*, $g(y_{ijk}) = \ln \frac{y_{ijk}}{N_{ijk}}$, with N_{ijk} as corresponding population counts (as proxies for person-years of exposure at age[s] i in year j), fixed by design; and $y_{ijk} \sim Poisson$ a random variate.

0.0053). Thus we had concomitantly and implicitly set mortality as declining by age at approximately 1.5% a year, a result that, in and of itself, is neither here nor there: Mortality rates for respiratory tuberculosis should decline with age after early adulthood. But before adulthood, rates of pulmonary TB are low, too; so that there is no substantive sense to a linear trend in age, even if it must exist algebraically. The more important consequence of the identification of the linear term for age (α^L) is that it simultaneously identifies linear terms for period (β^L) and cohort (γ^L). In the case of the latter, the decline is precipitous, at over 5% per annum, or a >40% decline in mortality at all ages across every ten years of births (*i.e.*., cohorts). This is a reasonable parameterization of something long known: that TB moves in waves, and that when mortality drops, it drops in a wave across cohorts (Andvord [1930] 2002; Frost 1939).

In contrast, under this model, identified even if only implicitly by $\widehat{\alpha^L} = -0.0148$ (equation [14] as implied by equation [13]), there is essentially no *linear* trend in period, as the estimated 0.2% increase per annum ($\widehat{\beta^L} = 0.0022$ or +2.2% per decade) is less than half of its estimated standard error ($\widehat{SE_{\beta^L}} = 0.0053$). The implication is that had the model been identified under the assumption of zero linear trend in period (equation [10] in lieu of equation[14]), the estimated parameters would have been the same. This is confirmed in **Figure 2** , where the coefficients for these two models—respectively, the blue and red dots—are indistinguishable. The coefficients plotted in this figure sum the estimated linear trends and deviations as per equations ($6a - c$) and thus a linear regression of time on the coefficients in any dimension has as slope the estimated effects discussed above, at least for the identifying assumption that the terms for the first two age categories are equivalent (equation [13]).[7]

[7] The pattern of coefficients is thus that depicted as "net effects Model VII" in Mason and Smith (1985, Figures 4-6), with three minor differences. First, the earlier analysis used a logistic response (rather than log-rate) model, although at such low rates, the difference in specifications is minuscule (*e.g.*, Clogg and

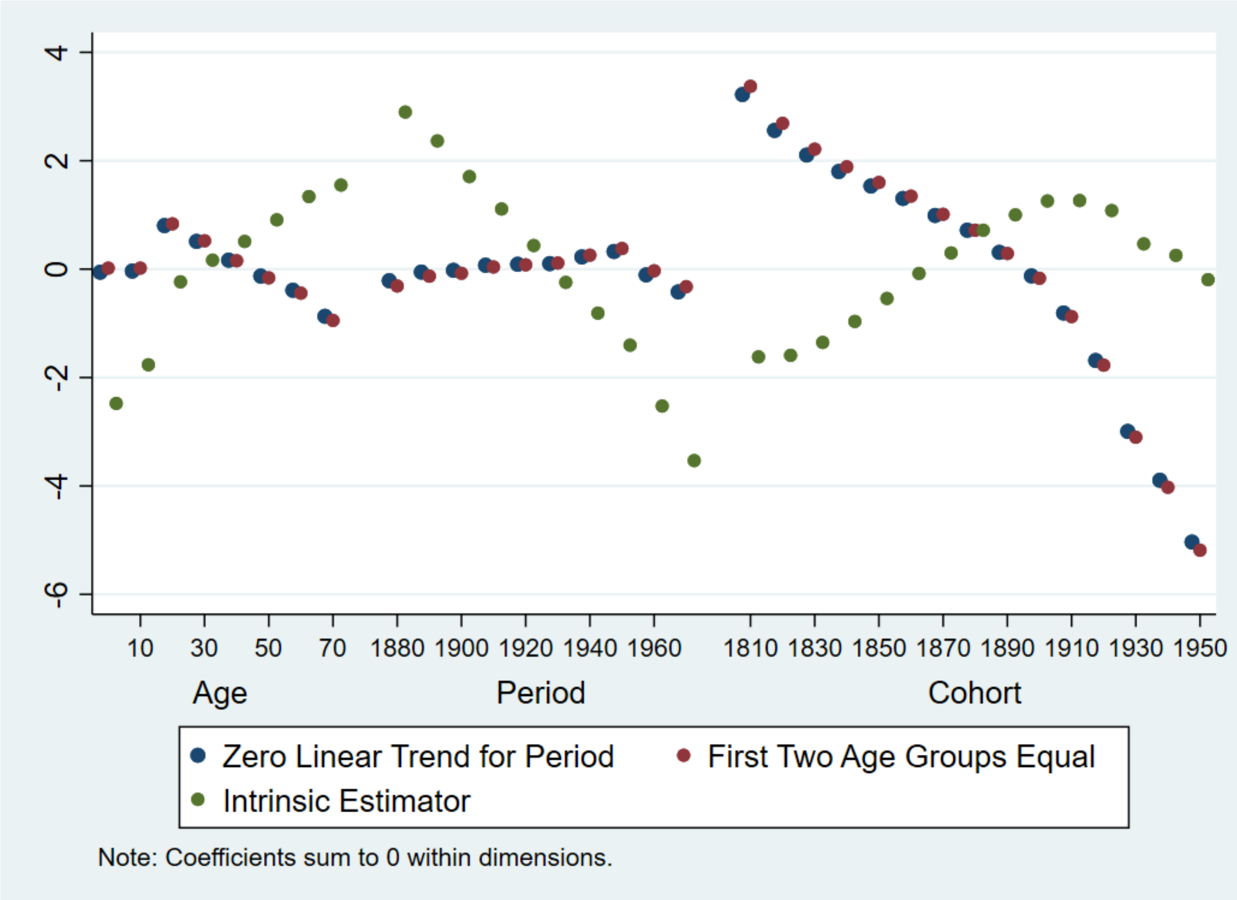


Figure 2*. Estimated Effect Coefficients (Including Linear Trend) for Three Just-Identified Age-Period-Cohort Models fit to the Massachusetts TB Data from Mason and Smith (1985).*

Eliason 1987, 28-29). Second, in the original analysis identification of the set of category coefficients was done by omitting dummy variables, *i.e.*, setting one or more to zero, rather than norming the set to sum to zero, as per equation (4). This affects the constant of the equation, hence the location of the coefficients—but not their difference from one another. Third, in Mason and Smith (1985), the absence of deaths in the three cells associated with the cohorts of 1960 and 1970 led to constraining the cohort effect for the cohorts of 1950, 1960, and 1970 to be identical. Here, no effects were estimated for these last two cohorts. Results are not sensitive to the treatment of these small cohorts and the corresponding cells.

It is by happenstance, then, that the CGLIM model proposed by Mason and Smith (1985) corresponds to a ZLT for *period* model: It *just happens* that $\alpha^L + \beta^L \approx \alpha^L$, where $\alpha^L + \beta^L$ is an estimable term that does not depend on the identifying restriction (as per equation $[9]$) and α^L is the constraint on the linear trend in *age* as per equation (14), which is implied by the CGLIM constraint in equation (13). Under this specification, we observe that $\beta^L \approx 0$ and infer that the ZLT for period specification would fit the data equivalently, as confirmed in **Figure 2**. Not all CGLIM models will imply such "clean" constraints (*e.g.*, a ZLT specification) on the linear terms $\{\alpha^L,\ \beta^L,\ \gamma^L\}$, but recognizing the correspondence between constraints on effect coefficients (such as CGLIM models) and constraints on linear terms is useful when thinking about identifying specifications (Fienberg and Mason 1985, 70). The plausibility of a set of estimated parameters is enhanced when they can be reproduced empirically from alternative constraints that do not necessarily imply one another.

In contrast, there is a large-class of just-identified specifications, including the Intrinsic Estimator (IE) (Yang, Fu, and Land 2004) that have some desirable *statistical* properties, but obscure constraints on the linear terms (O'Brien 2011). These constraints turn out to be non-intuitive functions of the *design* of the data array, *i.e.*, the values of I and J, hence K (Fosse and Winship 2018). **Figure 2** also shows (green dots) the estimated coefficients from the canonical IE model as applied to the Massachusetts respiratory TB data. It can be seen that linear trends contrary to those involved in either the Mason-Smith CGLIM constraint (equations $[13]$ or $[14]$) or period ZLT specification (equation $[10]$) give a different picture of the temporal structure of the disease as it disappeared in the commonwealth. The slopes under the IE constraint are, respectively, $\widehat{\alpha^L} = 0.0565$, $\widehat{\beta^L} = -0.0692$, and $\widehat{\gamma^L} = -0.0364$. The idea that TB mortality would rise monotonically with age, not to mention at such a high rate, is precisely what Frost (1939) was pushing back against 80 years ago.

From Bounding to Over-identification: Further Perspectives on Linear Trends

One could say that there is a lot of spurious precision in the various estimates presented above. Or if not in these estimates specifically—since we do generate the same estimates under two quite different identifying restrictions—in just-identified age-period-cohort models in general. How can we be so confident in a particular restriction on the linear parameters?

In brief: We can't. Confidence in the precise if occult IE identification for the Massachusetts TB data, for example, would have been misplaced. Cohort analysis thus shares with the more general social scientific literature (Manski 1995) an appreciation for the modesty (acknowledged ignorance) associated with identification through bounding the acceptable range of estimates of a parameter or set of parameters. The idea of identifying age-period-cohort models with reference to a restricted range of plausible values—*e.g.*, that linear trends in one or more dimensions are non-decreasing with time—dates to at least Wickramaratne *et al*. (1989); O'Brien (2019) proposes similar restrictions for a restricted range of one dimension, and Fosse and Winship (2019) give a generalized treatment of the underlying idea. The thread that I pull on in the remainder of this section has a particular orientation. Models with a linear bound in one dimension are equivalent to CGLIM models. Perhaps because no one seems to think about them as CGLIM models, no one seems to be as bothered by the identifying restriction as they are when an identifying restriction *starts* as a restriction on two adjacent terms in an ANOVA model. Nor am I bothered by this lack of visceral antipathy. It allows us to glide easily to interesting tests of models that are over-identified. At which point I draw the (precise algebraic) analogy with the problem of testing an age-period-cohort model against a model in which one of the three dimensions is excluded in its entirety… or in its linearity.

For any monotonic increasing constraint on effect coefficients $\{\theta_w\colon \theta_1 \leq \theta_2 \leq \cdots \leq \theta_{W-1} \leq \theta_W\}$, there is a corresponding minimal value to the linear trend θ^L. Call this θ_M^L. This value can be obtained with reference to a corresponding CGLIM model $\theta_w = \theta_{w+1}$ where w is determined by observing where $\theta_w^d - \theta_{w+1}^d$ is greatest across the $w = 1, \dots, W-1$ first differences among deviation coefficients $\{\theta_w^d\}$. The set of deviations $\{\theta_w^d\}$ for any temporal dimension θ is estimable and invariant across all just-identified models. Thus the initial model consulted (fitted) to establish the maximal $\theta_w^d - \theta_{w+1}^d$ does not affect the linear term θ_M^L obtained, or the other two linear terms, which are determined once θ_M^L is specified for one of the three dimensions. This does not mean that any CGLIM (or other just-identified) model will by itself generate sets of coefficients that maintain monotonicity in a given dimension. But the estimates from any such model, *combined with the identifying assumption regarding monotonicity*, will identify a CGLIM that defines a slope restriction (by reversing the derivation in equation $[12]$) that creates a boundary. All values $\theta^L > \theta_M^L$ are part of a region that maintains $\{\theta_w: \theta_1 \leq \theta_2 \leq \cdots \leq \theta_{W-1} \leq \theta_W\}$, albeit at the "cost" of inducing compensating change in the linear trends for the other two dimensions. With the boundary specifications of $\{\alpha^L, \beta^L, \gamma^L\}$ and the fitted $\{\alpha_i^d, \beta_j^d, \gamma_k^d\}$ in hand, equations $(6a-c)$ define the coefficients $\{\alpha_i, \beta_j, \gamma_k\}$ that maintain monotonicity in one dimension with the minimal linear slope.[8]

For example, if one wishes to stipulate that period coefficients $\{\beta_i\}$ be non-decreasing with time, then the minimal linear slope β_M^L that will sustain this stipulation is equal to the maximal first difference $\beta_w^d - \beta_{w+1}^d$, where the model is equivalently identified by setting $\beta_w = \beta_{w+1}$. In this sense,

---

[8] In the case where the identifying restriction is that coefficients in a given dimension be monotonically *decreasing*, then all of the above obtains with reversals of inequalities and the corresponding language, *i.e.*, "maximal" for "minimal," "least" or "lowest" for "greatest," and so on. In the case where monotonicity is only expected to obtain after a given time—*e.g.*, a peak in homicide rates at ages 20-24, with rates decreasing monotonically thereafter (O'Brien 2019)—then the definition of the lowest (most negative) value $\theta_w^d - \theta_{w+1}^d$ can be restricted to values of w that index categories over which monotonicity is assumed to obtain.

the bounding is weak. Consider, for example, a lower bound β_M^L conforming to the specification of monotonically increasing period effects. Then

$$(15a) \qquad (\alpha^L + \beta^L) - \beta_M^L \geq \alpha^L$$

$$(15b) \qquad \beta^L \geq \beta_M^L$$

$$(15c) \qquad (\beta^L + \gamma^L) - \beta_M^L \geq \gamma^L \,.$$

where both $\alpha^L + \beta^L$ and $\beta^L + \gamma^L$ are estimable and invariant (equation [9]) and thus can be determined from whatever just-identified model was estimated to determine β_M^L. Only for equation $(15b)$ does bounding identify the sign of a linear term (in this case, β^L)—by definition, since increasing monotonicity in period implies $\beta_M^L \geq 0$. Linear terms in age (α^L) and cohort (γ^L) can still take on both positive and negative values, at least until $\alpha^L + \beta^L$ and $\beta^L + \gamma^L$, respectively, are observed (estimated).[9]

We can advance matters by asking how putting further inequality restrictions on slope coefficients might plausibly tighten bounds. If in addition to monotonically increasing effects in period we further stipulate that the linear trend in cohort is also increasing—a less specific additional constraint, since it does *not* imply that cohort coefficients are necessarily monotonically increasing—then we are asserting that

$$(16) \qquad \gamma^L \geq 0$$

and thus defining tighter bounds:

$$(17a) \qquad (\alpha^L + \beta^L) - \beta_M^L \geq \alpha^L \geq (\alpha^L + \beta^L) - (\beta^L + \gamma^L)$$

$$(17b) \qquad \beta^L + \gamma^L \geq \beta^L \geq \beta_M^L$$

[9] Wikramaratne *et al*. (1989) posit $\beta_M^L = 0$, observe $\beta^L + \gamma^L > 0$, and note (338) that this implies (as per equation $[15c]$) that γ^L could be either positive or negative absent further assumptions. A corollary is that under the same maintained assumption, the counterfactual but a priori plausible observation that $\beta^L + \gamma^L < 0$ would guarantee $\gamma^L < 0$.

$$(17c) \qquad (\beta^L + \gamma^L) - \beta_M^L \geq \gamma^L \geq 0 \ .$$

These are elaborations of inequalities found in Holford (1991, 447), "for using knowledge about the underlying biology of a disease to understand something about the trends involv[ing] a restriction on one or more time factors." They differ from Holford's (1991) inequalities primarily in having sharpened the specification of the linear trend in period to include not just non-negativity in the linear trend, but also monotonicity in effect coefficients.

All of the bounds in equations $(17a - c)$ only hold conditional on

$$(18) \qquad \beta^L + \gamma^L \geq \beta_M^L$$

which can be viewed as an alternative hypothesis to the null hypothesis

$$(19) \qquad \beta^L + \gamma^L < \beta_M^L \ .$$

In this sense a test of equation (19) is a test of the over-identifying restriction provided by equation (16). (Compare equations $[15b]$ and $[17b]$.)

Why "[i]n this sense"? Because equations (17) through (19) lean on β_M^L, a quantity that is itself based on an assumption or identifying restriction. Were it to turn out that $\beta^L \ngeq \beta_M^L$—*i.e.*, that equation $(15b)$, the inequality conforming to the original identifying assumption, that period effects increase monotonically, were false—then the apparent tightness of the bounding on linear effects provided by equations $(17a - c)$ would be illusory. Estimation of linear trends continues to rely on an identifying restriction, even under an empirically maintained over-identifying restriction. This said, *conditional on* the stipulation provided by equation $(15b)$, the test of equation (19) and its potential sequelae—rejection in favor of equation (18) and adoption of the bounds in equations $(17a - c)$—could be quite informative. The fact that we do not *know* that period effects should increase monotonically does not mean that we cannot or should not *assume* (posit) that period effects increase monotonically, and to reason and observe from there.

Tests of over-identifying restrictions relative to the fit of a just-identified model have long been a feature of the cohort analysis literature (*e.g.*, Mason *et al*. 1973; Fienberg and Mason 1978, pp. 42-61).

> If the full logistic response model with age, period, and cohort effects, whether under-identified, just-identified, or over-identified, provides an acceptable fit to the data, then it will usually be of interest to explore whether we can set the effects of one or two dimensions to zero. Fitting age-period, age-cohort, cohort-period models, and even further reduced models, is a straightforward task with any computer program designed to fit standard log linear models to multidimensional arrays. Such reduced models pose no special identification problems because there is no way for the linear component of one dimension to become confounded with the linear components of the other two. (Fienberg and Mason 1978, 29).

Yang and Land (2013a, 109, Ch. 5) outline a method for age-period-cohort analysis that may culminate in the application of a just-identified model (including the IE), but only if more parsimonious two-dimensional models (*e.g.*, age-cohort) do not provide a satisfactory fit to the data. They attribute criticism of the IE by Luo (2013) to an instance where preference for a more restricted model should have ruled out the application of the IE (or any just-identified ["full-blown"] APC model) in the first place (Yang and Land 2013b).[10]

---

[10] Are the discrepant coefficients from the IE in **Figure 2** another instance in which this counsel was ignored? The estimates based on other constraints suggest very little variation in period, and one might well imagine that an age-cohort model would have / should have sufficed from the beginning. However, the improvement in fit observed by Mason and Smith (1985, Table 3)—a reduction in deviance of somewhat over 100 on 8 degrees of freedom—continues to obtain, even with the minor changes to the analysis detailed in footnote 7. A just-identified model would also be preferred if BIC is the model selection criterion ($BIC_{APC} = -73$ versus $BIC_{AC} = +12$).

Except that, as Holford (1991, 436-437) pointed out, two-factor models (over-identified models featuring only two of the three conceptual temporal dimensions) may not be what we think they are, at least with reference to a just-identified model.[11] Whereas in estimating a simplified multiple classification model of the form of (for example)

(20) $$g(y_{ijk}) = \mu + \alpha_i + \gamma_k$$

and comparing it with equation (3) we might imagine that we have tested as null the full set of omitted coefficients

(21) $$\beta_j = 0$$

for all j, hence linearity in period as well (equation $[10]$), we may have done no such thing. In particular, what we have tested is "only"

(22) $$\beta_j^d = 0$$

for all j, as in

(23) $$g(y_{ijk}) = \mu + (\alpha^L + \beta^L)A_i + (\gamma^L + \beta^L)C_k + \alpha_i + \gamma_k$$

versus equation (9), on $J - 2$, not $J - 1$, degrees of freedom (Fienberg and Mason 1985, 71-72). Uh-oh: There, in equation (23), is β^L again, free to be more or less anything, absent a further restriction on the linear terms!

This state of affairs—if state of affairs it is—discomfited Holford (1991, 437). Having noted that "[i]t is impossible to test the null hypothesis that the slope is zero [$\beta^L = 0$] when both age and cohort also are included in the model" he ventured, tentatively that "[i]t would be surprising if a causal agent that changed over time did so in a strictly linear fashion. Typically, you also would expect to see a certain amount of curvature…." Which is to say that if equation (22) holds, something—experience?

[11] Rodgers (1982, 780-782) was also getting at this point, albeit with an example that elided issues of measurement error with issues pertaining to identification.

theory? common sense?—would suggest that equation (21) should hold as well. This is notwithstanding that there is no algebraic or statistical reason to rule out the existence of a linear trend in period (rejection of equation [10]). Holford (1991, 437) thus yields to the dictates of algebra and statistics in concluding that "an analysis that is limited to the consideration of just two factors is not really a solution to the problem, because the possibility for bias has not been eliminated." This perspective is reprised by Fosse and Winship (2018, 322):

> Unfortunately, it is impossible to determine from the data alone whether or not all three temporal variable are operating. Believing otherwise can seriously mislead researchers…. [B]y fitting the two-factor model … one is imposing the identification assumption that the [omitted] linear effect is zero…. The zero-linear trend constraint on the cohort variable is external to the data, imposed by the researcher. Depending on the substantive application, it may or may not be reasonable to assume that because the nonlinear effects of [one temporal dimension] are observed to be zero, its linear effect is also zero…. [T]his is an assumption that can only be justified by appealing to theory or the inclusion of additional data.

O'Brien (2016, 366 [emphasis in original]) is even more admonitory:

> When the model contains just two factors, those two factors take credit for their own linear trend effects, their effects that involve deviations from their linear trends, *and they take credit for the linear trend effects of the third (left-out) factor*…. Leaving the third factor out of the model based on its incremental fit not being statistically significant will too often eliminate a substantively important factor. This elimination affects the coefficient estimates of the two factors in the model.

It is one thing to realize that all just-identified models featuring effects for all three temporal dimensions necessitate restrictions on the linear terms. It is quite another to see that even if one is willing to forego consideration of one of the temporal dimensions, one is (perhaps) still prey to the indeterminacy deriving from the fact that any of age, period, or cohort is a linear combination of the other two. What is going on here? In the case of two-factor models, "[t]he left-out factor's effects are constrained to be zero, both its linear trend effects and the effects of its deviations from the linear trend" (O'Brien 2016, 365). We can parse this further to specify that the constraint on the deviations (*e.g.*, equation $[22]$) is estimable hence testable, while the constraint on the linear term (*e.g.*, equation $[10]$) is an assumption. The situation is completely analogous to that discussed just above, in conjunction with an over-identifying restriction to tighten the bounds on effects (equations $[17-19]$). Holford (1991, 437) hazarded that if deviations from trend in one dimension are not statistically detectable, then one might plausibly infer that there is no linear trend in the same dimension, either. I would turn it around to ask, why not just acknowledge that the test of an over-identified two-factor model against a just-identified age-period-cohort model is, specifically, a test against the just-identified model in which the linear term in the soon-to-be-absent dimension is assumed to be zero? Conditional on this assumption, the test that there are also no non-linear effects in this dimension is a perfectly useful one.

How can we—or why might we—wish to entertain such an assumption? Theory or additional data specific to the analysis at hand (O'Brien 2016, 369; Fosse and Winship 2018, 322) are always welcome, but there may also be some general ways of thinking about the analysis of data in terms of age, period, and cohort that help to structure the models and methods in common use. The ones I discuss in the remainder of this chapter mostly abjure imagining that there *are* linear effects operating simultaneously in period and cohort. The problem, to my way of thinking, is less that age, period, and cohort are definitively bound (age ≡ period – cohort) than that we have come to substitute the algebra of the situation for the logic of social science. When O'Brien (2016, 322) writes that "[l]eaving the third factor out of the model based on its incremental fit not being statistically significant will too often eliminate a substantively important factor," I look for evidence of a social world, theoretical and/or empirical, in which it makes sense to think of all of these time dimensions as indexing factors that move both simultaneously and linearly in their effects on some phenomenon of interest. What I find instead are references to "the data-generating processes" (O'Brien 2016, 365) or "the true data generating parameters" (Fosse and Winship 2018, 324) or even "the true A, P, and C trends" (Luo and Hodges 2016, 710). Far be it from me to gainsay the utility of a received specification in adjudicating arguments in the statistical realm (Smith, Mason, and Fienberg 1982, Tables 1 and 2). But do such entities—data generating processes or parameters, "true" trends—really exist? I am mindful that all of our models are mental constructs; also, that it is hard to refute on the one hand or prove on the other a Platonic concept. Nonetheless, the indeterminacies intrinsic to the statistical perspective may be distracting us from some fundamentals that could make the ground feel less shaky underfoot.

Against Concomitant Linear Trends in Cohort and Period

Fosse and Winship (2018) conclude their survey of statistical approaches for identifying simultaneous age, period, and cohort effects with

> … a set of practical guidelines…. "[T]he full set of linear and nonlinear effects should be reported. This will allow the researcher to evaluate the legitimacy of the constraint imposed on the true, unknown linear APC effects…. Ultimately, any linear constraint should be grounded in an underlying social, cultural, or biological theory. (328)

If categorical terms for each dimension of time have been normalized as per equation (4), then equations (5) and (6) make it straightforward to deduce linear and nonlinear effects, in fitting a regression line to the observed coefficients in at least two of the dimensions (Holford 1983). It has long been known that all just-identified age-period-cohort models maintain and/or require a restriction on a linear term in time (Fienberg and Mason 1978, 6), even if the nature of those restrictions is sometimes obscure. Earlier in this chapter I treated some comparatively simple cases (ZLT, CGLIM, bounding) compared to those unearthed by Fosse and Winship (2018).

Given all the time that has passed since cohort analysis veered into the identification problem, and all of the work that has been done within the age-period-cohort frame as well as at the edges, it would be (a) hard to dispute the generic advice that any and all identifying restrictions be motivated with reference to the specific problem at hand, and the theory and state of knowledge that surround it; and (b) fatuous to suggest a general solution. But might not some re-orientation be in order, to keep us focused on the point of the enterprise, hence to avoid algebraic distractions? The association of the data structure with a set of challenges and strictures from statistics and mathematics has had the unfortunate consequences of (a) turning attention from substantive issues, first assuming that we are less knowledgeable than we are, then making it appear that this is true; and (b) causing us to overlook what the basic identity A≡P-C is telling us about how we should be thinking about age, period, and cohort as explanatory concepts. The nub of the problem is an epistemological one—what it is we think we are thinking about—and not a mathematical one. Age, period, and cohort might be exchangeable algebraically and geometrically, but conceptually they are distinct. For the two constructs associated with historical time, period and cohort, it is hard to think of when and why we would want to imagine *linear* trends in both.

Thus given the following: social and/or demographic data arrayed over time by age, and a desire to know "what is going on" with respect to these age-graded rates and counts, including what has transpired in the past and what we have reason to think might transpire in the future. In such instances, which are ubiquitous, statistical accounts in terms of age (and aging) and cohorts (including trends) should come to the fore. What about period effects—don't *they* exist? Well, of course they exist; or might exist; or we might think that they exist; or thinking that they exist might help with our thoughts more generally. But, in the first instance, our preference should be for age-cohort models with specific interactions, and/or age-period-cohort models with identification residing in the presumption that the linear trend in period is zero. In many respects—as a *technical* treatment of the issue—this is not only not new, it is *old*: Adepts of the literature may recognize that these are the procession of models used by Holford (1983, 314-318) to illustrate estimation of age, period, and cohort effects. What I want to offer in addition, in the set of interconnected remarks that follow, is a *rationale* for thinking this way, as opposed to worrying, for example, that an implied cohort linear trend is somehow taking credit for something that in some sense belongs to period (*cf*. O'Brien 2016, 366). The orientation I proffer here is situational and not categorical. I am suggesting that cohort is in general the more logical "carrier" of historical or temporal trends.

One reason is that age effects usually are best interpreted with reference to cohorts. The theory of *age* effects is primarily developmental and longitudinal, hence within cohorts, not cross-sectional. This is true whether the primary emphasis is on understanding how some phenomenon unfolds across the life course (Baltes 1968), or whether the question is historical change across some phenomenon known to vary systematically with age (*e.g*., Dorius, Alwin, and Pacheco 2016). Sometimes the structure connecting rates with age—the linear increase in log-mortality with age over most of adulthood, for example (*e.g.*, Cohen, Bohk, and Rau 2018)—can have the same shape whether viewed from a cohort or period perspective (Lenert and Missov 2010). In other circumstances, the proponents of theories regarding age patterns of behavior and events have not noticed the distinction between age patterns hypothesized longitudinally but viewed cross-sectionally. In criminology, Hirschi and Gottfredson (1983) posited an age effect on crime that "is invariant across social and cultural conditions" (560), motivated it with respect to developmental factors, but illustrated it primarily with reference to cross-sectional patterns. This made matters ripe for misunderstanding as overall crime rates changed over time (Porter *et al*. 2016, 34-35).

What are the implications for identification? There are at least three. First, age-cohort models are in some sense primordial. This is unless we happen to be in a theoretical plane where age-period models are primordial. It is not impossible to give a scientifically coherent reading to an age pattern that applies in the cross-section: Social and economic punishments and rewards may be age-graded, and their relative application conditional on age might be a function of period-specific circumstances and resources alone. In this event, the cohort biographies found in the data could well be of historical interest, but the patterns within the cohorts of punishments and rewards would give an inaccurate rendering of how age figures in their allocation, absent adjustment for period. The distinction between the two is a theoretical one, as it has always been (*e.g.*, Fienberg and Mason 1978, 50, 58-59)—not a matter of comparative goodness-of-fit. The crucial point is that the mental baseline should not be an age-period-cohort model with no identification restriction in sight, because therein lies the madness of equation (23), a twisting whirligig we would not even have imagined unless our view of the world *began* with equations (8) and (9).

Second, to the extent that we have an understanding of what age patterns should look like, *e.g.*, within a cohort, then extended models—A-P-C models, with all three temporal dimensions represented—can be identified with reference to age effects. Fienberg and Mason (1978, 42-61) worked from the beginning with an over-identifying restriction on age effects: In an era with little adult education, the level of schooling in a cohort should be relatively fixed after young adulthood, at least through the middle ages, until mortality and poor recall kick in. Mason and Smith (1985) has provided a just-identified example. The fact that the restriction on age (equation $[13]$) proved isomorphic with a zero linear trend restriction on period (equation $[10]$) increases confidence that the estimated coefficients convey a reasonable partition of the various time dimensions, including the restriction of all trend in historical time to the coefficients indexing cohort.

The third implication is the converse of the second: If we are restricting trend in historical time to one dimension alone (period), one test of the plausibility of this restriction is whether or not it generates a pattern of age coefficients that comports with expectations regarding the age pattern of the phenomenon under study. **Figure 3** shows age, period, and cohort coefficients (blue points) under a zero linear trend (ZLT) constraint on period (equation $[10]$) for a just-identified model fit to data on correctional supervision and prison spells for North Carolina males, by age, between 1972 and 2016 (Shen *et al*. 2019). The coefficients for age show the anticipated pattern, with judicial sanction being highest in the mid- to late-20s, and dropping off at an accelerating pace thereafter. (The response variable is a *log*-rate.) Of course, such an "eye test" could (and should!) be made for *any* identifying restriction. But it may be more helpful and reassuring in the context of a decision to restrict the *trend* in effect coefficients to two dimensions alone: developmental and historical time. **Figure 3** also presents an alternative version of this partition, in which period is allowed to assume all historical trend (red points), and the general age pattern is not terribly different. One could not (nor should not) choose between the two specifications on this basis alone; but given the general reasonableness of the age pattern under two alternative apportionments of the linear component of historical trend, it is easier to turn to the substantive component of the exercise. This involved assessing the effects of a 1994 sentencing reform (clearly visible in the plot of period effect coefficients) against the backdrop of secular change in crime and punishment which, given the high propensity for crime in early adulthood and the subsequent effects of early incarceration on the chances of later involvement with the criminal justice system, can be best represented as so-called cohort effects (Porter *et al*. 2016).

Many other choices of identifying restriction, otherwise untethered and indifferently specified, could reorient the age pattern(s) in a manner that is un-credible. See again the IE estimates in **Figure 2**. One of the lessons of stipulating monotonicity in any set of effects $\{\theta_W\}$ is the "hands on" experience of finding that there is always an extreme value of θ_M^L that will functionally annihilate whatever information is (was) contained in the unique estimates of the deviation coefficients $\{\theta_w^d\}$. The arithmetic of the situation notwithstanding, this is not a desirable state of affairs. From an analytic standpoint, there really is a sense in which these coefficients need be privileged against the super-abundance (three versus two) of the linear terms, even if the elimination of a linear term in the potential presence of deviations from linearity in the same dimension does appear to stand on its head the common understanding of marginality in linear models (Nelder 1977, 49-50).

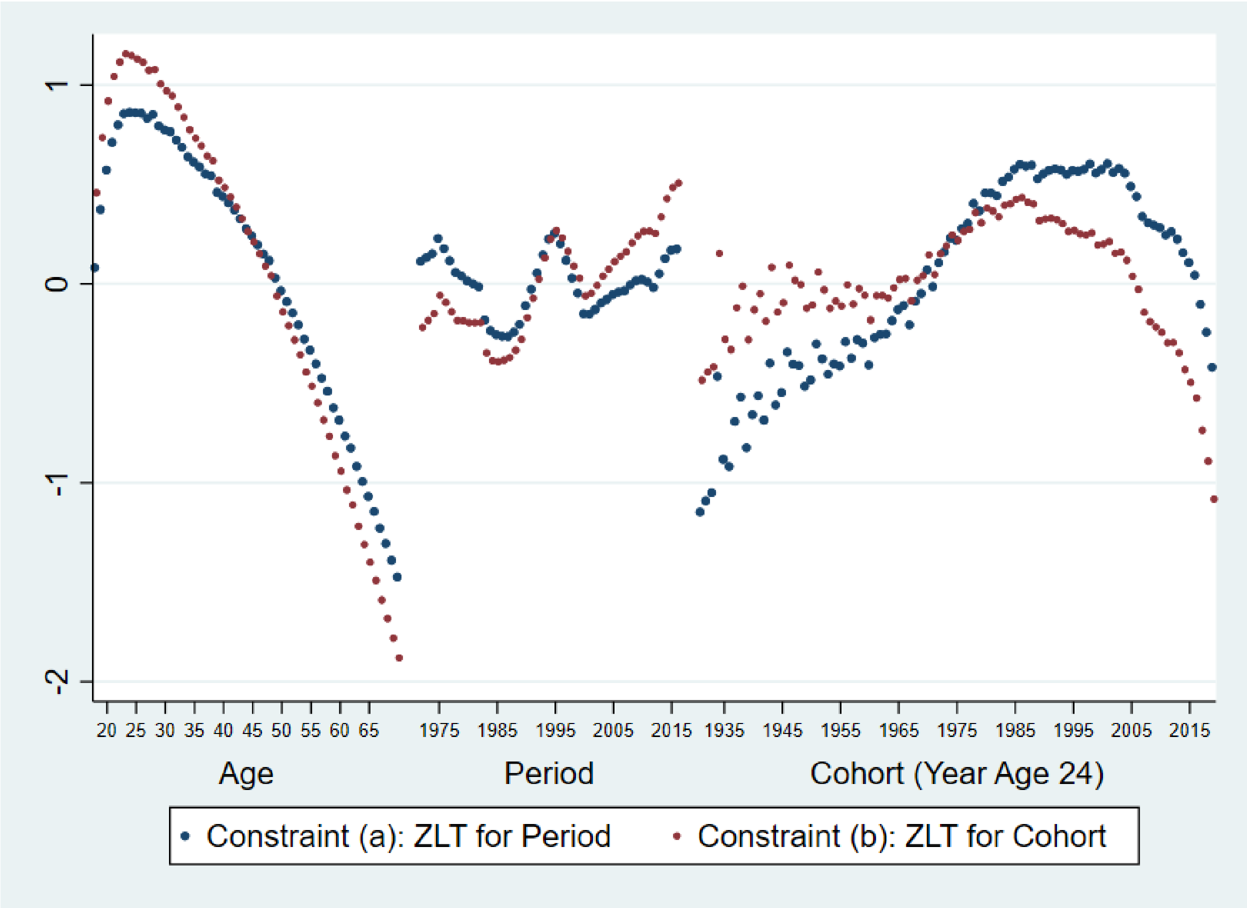


Figure 3. *Estimated Age, Period, and Cohort Coefficients for Two Zero Linear Trend (ZLT) Restrictions on a Model for Prison Sentence Spells in North Carolina, 1972-2016.*

Trends that are purportedly the combination of offsetting period and cohort effects are difficult to understand:

> For any dataset, there is a space of possible estimates, all equally consistent with the data, but with different linear trends in age, period, and cohort, and no way from the data alone to choose among these…. On the other hand, some of the estimates seem to make more sense than others. For example, consider a model which adds the following three trends: (i) since 1948, an increase of 1% per year in the overall probability of Democratic identification, (ii) starting at age 18, an increase of 1% per year in the probability of an individual being a Republican as he or she gets

> older, and (iii) for each cohort, an increase of 1% in the probability of being Republican, compared to the cohort that was born one year earlier. Add these three trends together and you get zero—the combination has no effect on any observable data—but they do not make much political sense. What does it really mean to talk about a linear time trend toward the Democrats if it is exactly canceled by each cohort being more Republican than the last? To put it another way, some methods of constraining the possible space of solutions seem more reasonable than others. (Gelman 2008, 2-3).

A similar argument can be made against taking too seriously the idea that there are conceptually separable period and cohort trends that are moving in the same direction. What exactly would that mean, given that we would be talking about *linear* trends? That the march of history is progressing or regressing in some continuous reapportionment of its effects, from universal with respect to age to transmitted via early socialization—or vice versa? The impossibility—and inutility—of attempting to distinguish these trends as between cohort and period is exemplified in the comparative study of income, where an historical trend is assumed, and the question of substantive interest is how different cohorts fare relative to this trend:

> As we explain how cohorts diverge from overall income trends, substantive reasons also exist for focusing on fluctuations around a linear trend, instead of focusing on the linear trend itself. Namely, the linear trend that one generation gets born into a society that is richer than the same society at an earlier point in time would generally not be considered unfair, but inevitable. Immanuel Kant (1784) most prominently argued that we are accustomed to one generation profiting from the efforts of the preceding one, so that overall, a long-run cohort- (or period-based; *one can never know*) progression of living standards is the baseline to expect. (Chauvel and Schröder 2014, 1285 [emphasis added]).

Into the Future

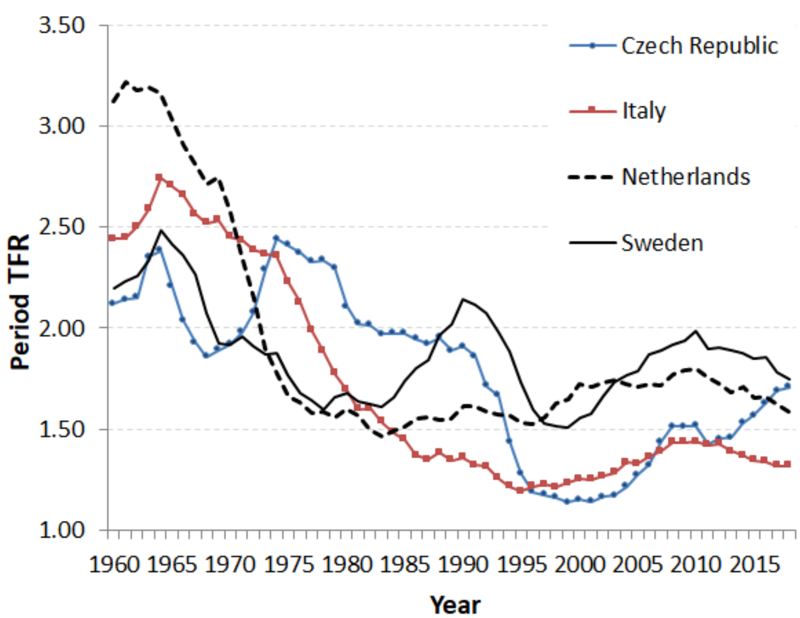
None of which is to imply that there is no reason to distinguish period from cohort effects, either conceptually or in age-period-cohort models. The enduring charm of these models is that there is age-graded variability in proportions, rates, counts and the like that derives from the irruption of historical events against the backdrop of the secular change embedded in cohorts. I close with a brief foray into how the allocation of historical linear trend to the dimension indexed by cohort may be of value when it comes to projecting or forecasting the future.

Demography is the only social science which routinely makes projections or forecasts over long horizons (Granger 2007, 6). The reason is the cohort conception that attaches to data arrayed by time and indexed by age (Smith 2009, 145-151). This, and the long-livedness of the species (Ryder 1965). Many of the men and women surveyed in the National Election Studies of the second half of the 20th Century had been alive, and at varying ages, when women were given the right to vote in 1920. This—coupled with observation of the age pattern of voting over the life course—was leveraged by Firebaugh and Chen (1995) to make inferences on the impact of this period-specific political reform on the political involvement of different generations. Conversely, the cohort differentiation that can be observed in the past opens up the possibility of using the characteristic age pattern to project phenomena forward, as when differences in the early life-course uptake of smoking foreshadow mortality many years into the future (Wang and Preston 2009).

Which is not to say that the variation that exists historically in rates will primarily be due to factors associated with cohort membership. To the contrary, there are many phenomena for which, from the standpoint of observed variation, "*short-term* period effects" (Mason *et al.* 1973, 245; emphasis added), such as those adduced with respect to political party identification, will dominate the partition of variance as between period and cohort. Consider **Figure 4** and **Figure 5**, which update figures that first appeared in Sobotka (2003, Figures 1a and 1b). **Figure 4** presents total fertility rates

(TFR), which are period-specific summations of age-specific fertility rates and, as such, are a strong first-approximation to the period effects in the age × period data array. **Figure 5** displays completed fertility rates (CFR) that are within cohort summations of age-specific rates for the corresponding age × cohort array of age-specific fertility rates. As such, they are a strong first-approximation to the cohort effects for this historical period. Period variation is paramount (Ní Bhrolcháin 1992). Moreover, this period variation is decidedly non-random. To focus, for example, on the dots and thin connecting line in **Figure 4**, for the Czech Republic: There was a substantial decline in fertility in the run-up to the Prague Spring (1968), with a substantial recovery thereafter, before fertility dropped precipitously again in the early 1990s, coincident with the end of the Soviet Union and its control over Czechoslovakia, and the subsequent separation of the Czech Republic from the union with Slovakia.

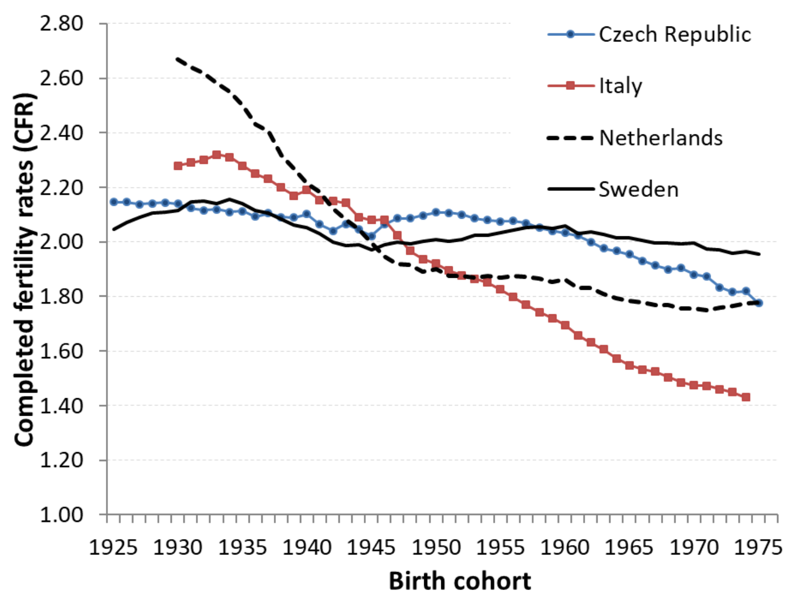
Figure 4 *Total Fertility Rates (TFR) for Four European Countries.*



The issue, however, is that these consequential period-specific shocks are, from a forecasting standpoint, all but unknowable. Things happen; we react. We can give knowledgeable accounts of the effects of past happenings, but we are not so good at predicting them, much less their timing (Smith 1987). The argument in favor of approaching fertility change from a cohort perspective (*e.g.*, Ryder 1960) always had two aspects to it: “first, the description or analysis of fertility patterns, and second, the formation of inferences about the course of fertility” (Hutchinson 1960, 117). It is with respect to the future course of events that a projection structure based on an underlying age-cohort model, with all historical trend allocated to cohort, has a crude but basic utility. It is possible to forecast from age-period-cohort models with linear trend in both cohort and period (Riebler, Held, and Rue 2012, 315-322; Yang and Land 2013a, 171-188); however, in all events, there is some added clarity to the exercise when the partition of trend is clear. In **Figure 2** and again in **Figure 3**, I have been at pains to present estimated coefficients on a common graph, hence on a common scale, which I claim immodestly only makes sense, given that age, period, and cohort have a common metric. I do not know why this practice is not more common. It has the great advantage of making clear not just the trend in estimated effects under alternative models, but the comparative variation in outcomes attributable to each temporal dimension conditional on contraints on linear trends.

The variation visible in these figures in, respectively, age, period, and cohort effects, and the dependence of this variation on our treatment of linear components, is evident in the following equations, which re-express the set of effect coefficients $\{\alpha_i,\, \beta_j, \gamma_k\}$ as variances, expanded with reference to equations $(6a-c)$:

Figure 5. *Completed Cohort Fertility (CFR) for Four European Countries.*



$$(24a) \qquad \sigma_{\alpha_i}^2 = (\alpha^L)^2\left(\frac{I^2-1}{12}\right) + \sigma_{\alpha_i^d}^2$$

$$(24b) \qquad \sigma_{\beta_j}^2 = (\beta^L)^2\left(\frac{J^2-1}{12}\right) + \sigma_{\beta_j^d}^2$$

$$(24c) \qquad \sigma^2_{\gamma_k} = (\gamma^L)^2\left(\frac{K^2-1}{12}\right) + \sigma^2_{\gamma^d_k}$$

As previously, the components corresponding to deviations from trend $\left\{\sigma^2_{\alpha^d_i}, \sigma^2_{\beta^d_j}, \sigma^2_{\gamma^d_k}\right\}$ are unique—they do not depend on the choice of identifying restriction. Nor do the dimension-specific scalars denominated by 12, which vary across age, period, and cohort in function of the span of the data array (*i.e.*, in a standard age × period design, I and J, hence K). Because $K > J$, *ceteris paribus* ($|\beta^L| = |\gamma^L|$ and $\sigma^2_{\beta^d_j} = \sigma^2_{\gamma^d_k}$)

$$(25) \qquad \sigma^2_{\gamma_k} > \sigma^2_{\beta_j}$$

Implying some normalization may be in order when comparing the extent of period and cohort effects (*cf*. Vaisey and Lizardo 2016, 9-13).

When there are linear trends in both period (β^L) and cohort (γ^L), the corresponding variances $\sigma^2_{\beta_j}$ and $\sigma^2_{\gamma_k}$ can become quite expansive. In contrast, under the assumption of zero linear trend in period (equation [10])

$$(26) \qquad \sigma^2_{\beta_j} = \sigma^2_{\beta^d_j}$$

and the proportion of historical variability attributable to differences between cohorts is

$$(27) \qquad \frac{\sigma^2_{\gamma_k}}{\sigma^2_{\beta_j}+\sigma^2_{\gamma_k}} = \frac{(\gamma^L)^2\left(\frac{K^2-1}{12}\right)+\sigma^2_{\gamma^d_k}}{\sigma^2_{\beta^d_j}+(\gamma^L)^2\left(\frac{K^2-1}{12}\right)+\sigma^2_{\gamma^d_k}} \quad .$$

This quantity can be either reassuring or sobering with respect to forecasts based on the set of age and cohort coefficients $\{\alpha_i, \gamma_k\}$, since under this baseline parameterization, $\sigma^2_{\beta^d_j}$ is a minimum estimate of the secular variability that is historical but not embedded in cohort effects. In this sense it is a non-stochastic measure of error—in the form of ignorance—in our quest to see into the future on the basis of what we have learned in the past. Further decomposition of $\sigma^2_{\gamma_k}$ in terms of trend and deviance from trend (via equation $[24c]$) can be similarly instructive.

What Is It Good for?

Alas, there is not too much that is intrinsic in social life. Even if our range of behavior is circumscribed by our biology and our physical environment, what we make of it seems to be quite varied. Small wonder that our methods for making sense of the human world rarely have the fixity that our scientific minds crave. For roughly half a century, thinking about age-period-cohort models has tended to drift away from their specific utility for specific problems to absolutism, both for and against. This chapter represents an umpteenth effort—my fifth, personally (Fienberg 2013, 1982)—to explore how the age-period-cohort accounting model framework (Smith, Mason and Fienberg 1982; Mason and Smith 1985) can be used to illuminate the world around us, with reference to specific and general ideas in sociology, demography, and epidemiology. If I have perseverated a bit on the algebra of the situation, the homeliness of the effort is a reminder that this is not akin to splitting the atom on the one hand, or handling poisonous serpents on the other. It is just another way of getting temporary purchase on the social world around us. When it illuminates, it is good.

References

Andvord, Kristian F. [1930] 2002. "What Can We Learn by Following the Development of Tuberculosis from One Generation to Another?" *International Journal of Tuberculosis and Lung Disease* **6**(7):562–568.

Baltes, P.B. 1968. "Longitudinal and Cross-Sectional Sequences in the Study of Age and Generation Effects." *Human Development* **11**(3):145-171.

Chauvel, Louis. 2014. *Le destin des générations : Structure sociale et cohortes en France du XX^e siècle aux années 2010 [2^e édition]*. Paris : Presses Universitaires de France.

Chauvel, Louis, and Martin Schröder. 2014. "Generational Inequalities and Welfare Regimes." *Social Forces* **92**(4):1259-1283.

Clogg, Clifford C., and Scott R. Eliason. 1987. "Some Common Problems in Log-Linear Analysis." *Sociological Methods & Research* **16**(1):8-44.

Cohen, Joel E., Christina Bohk, and Roland Rau. 2018. "Gompertz, Makeham, and Siler Models Explain Taylor's Law in Human Mortality Data." *Demographic Research* **38**(29):773-842.

Dorius, Shawn F., Duane F. Alwin, and Julianna Pacheco. 2016. "Twentieth Century Intercohort Trends in Verbal Ability in the United States." *Sociological Science* **3**:383-412.

Duncan, Otis Dudley. 1985. "Generations, Cohorts, and Conformity." In *Cohort Analysis in Social Research: Beyond the Identification Problem*, edited by William M. Mason and Stephen E. Fienberg, 289-321. New York: Springer-Verlag.

Fienberg, Stephen E. 2013. "Cohort Analysis' Unholy Quest: A Discussion." *Demography* **50**(6):1981-1984.

Fienberg, Stephen E., and William M. Mason. 1978. "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data." In *Sociological Methodology 1979*, edited by Karl F. Schuessler, 1-67. San Francisco: Jossey-Bass.

-----. 1985. "Specification and Implementation of Age, Period and Cohort Models." In *Cohort Analysis in Social Research: Beyond the Identification Problem*, edited by William M. Mason and Stephen E. Fienberg, 45-88. New York: Springer-Verlag.

Firebaugh, Glenn, and Kevin Chen. 1995. "Vote Turnout of Nineteenth Amendment Women: The Enduring Effect of Disenfranchisement." *American Journal of Sociology* **100**(4):972-996.

Fosse, Ethan, and Christopher Winship. 2018. "Moore–Penrose Estimators of Age–Period–Cohort Effects: Their Interrelationship and Properties." *Sociological Science* **5**(14):304-334.

-----. 2019. "Bounding Analyses of Age-Period-Cohort Effects." *Demography* **56**(5):1975-2004.

Frost, Wade Hampton. 1939. "The Age Selection of Mortality from Tuberculosis in Successive Decades." *American Journal of Epidemiology* **30**(A3):91-96.

Gelman, Andrew. 2008. "Thoughts on New Statistical Procedures for Age-Period-Cohort Analyses." Unpublished paper last downloaded from http://www.stat.columbia.edu/~gelman/research/unpublished/apc.pdf on 4 April 2019.

Granger, Clive W. J. 2007. "Forecasting—Looking Back and Forward: Paper to celebrate the 50th Anniversary of the Econometrics Institute at the Erasmus University, Rotterdam." *Journal of Econometrics* **138**(1):3-13.

Hirschi, Travis, and Michael Gottfredson. 1983. "Age and the Explanation of Crime." *American Journal of Sociology* **89**:552–84.

Holford, Theodore R. 1983. "The Estimation of Age, Period and Cohort Effects for Vital Rates." *Biometrics* **39**(2):311-324.

-----. 1991. "Understanding the Effects of Age, Period, and Cohort on Incidence and Mortality Rates." *Annual Review of Public Health* **12**:425-457.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* **81**(396):945-960.

Hutchinson, Edward P. 1960. "Comment." Pp. 131-133 In Ryder (1960).

Jacobs, Jerry A. 2005. "*ASR*'s Greatest Hits." *American Sociological Review* **70**(1):1-3.

Kupper, Lawrence L., Joseph M. Janis, Azza Karmous, and Bernard G. Greenberg. 1985. "Statistical Age-Period-Cohort Analysis: A Review and Critique." *Journal of Chronic Disease* **38**(10):811-830.

Lenart, Adam, and Trifon I. Missov, 2010. "Linking Period and Cohort Life Expectancy In Gompertz Proportional Hazards Models." MPIDR Working Papers WP-2010-024, Max Planck Institute for Demographic Research, Rostock, Germany.

Luo, Liying. 2013. "Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem." *Demography* **50**(6):1945-1967.

Luo, Liying, and James S. Hodges. 2016. "Block Constraints in Age–Period–Cohort Models with Unequal-width Intervals." *Sociological Methods & Research* **45**(4):700-726.

Mannheim, Karl. 1952. "The Problem of Generations." In *Essays on the Sociology of Knowledge: Collected Works, Volume 5*, edited by Paul Kecskemeti, 276-322. New York: Routledge.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard.

Mason, Karen Oppenheim, William M. Mason, H. H. Winsborough, and W. Kenneth Poole. 1972. "Some Methodological Issues in Cohort Analysis of Archival Data." Working Paper 72-22. Center for Demography and Ecology, University of Wisconsin—Madison.

-----. 1973. ''Some Methodological Issues in Cohort Analysis of Archival Data.'' *American Sociological Review* **38**(2):242-258.

Mason, William M., and Stephen E. Fienberg (eds.) 1985. *Cohort Analysis in Social Research: Beyond the Identification Problem*. New York: Springer-Verlag.

Mason, William M., and Herbert L. Smith. 1985. "Age-Period-Cohort Analysis and the Study of Deaths from Pulmonary Tuberculosis." In *Cohort Analysis in Social Research: Beyond the Identification*

Problem, edited by William M. Mason and Stephen E. Fienberg, 151-227. New York: Springer-Verlag.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

Nelder, J. A. 1977. "A Reformulation of Linear Models." *Journal of the Royal Statistical Society. Series A (General)* **140**(1):48-77.

Ní Bhrolcháin, Máire. 1992. "Period Paramount? A Critique of the Cohort Approach to Fertility." *Population and Development Review* **18**(4):599-629.

O'Brien, Robert M. 2011. "Constrained Estimators and Age-Period-Cohort Models." *Sociological Methods & Research* **40**(3):419-452.

-----. 2015. *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data*. Boca Raton, FL: Chapman and Hall/CRC Press.

-----. 2016. "Model Misspecification When Eliminating a Factor in Age-period-cohort Multiple Classification Models." In *Sociological Methodology 2016*, edited by Duane F. Alwin, 358-372. Thousand Oaks, CA: Sage Publishing.

-----. 2019. "Homicide Arrest Rate Trends in the United States: The Contributions of Periods and Cohorts (1965–2015)." *Journal of Quantitative Criminology* **35**(2):211-236.

Oppenheim, Karen. 1970. "Voting in Recent American Presidential Elections." Ph.D. dissertation. University of Chicago.

Porter, Lauren C., Shawn D. Bushway, Hui-Shien Tsao, and Herbert L. Smith. 2016. "How the U.S. Prison Boom Has Changed the Age Distribution of the Prison Population." *Criminology* **54**(1):30-55.

Riebler, Andrea, Leonhard Held, and Håvard Rue. 2012. "Estimation and Extrapolation of Time Trends in Registry Data—Borrowing Strength from Related Populations." *The Annals of Applied Statistics* **6**(1):304-333.

Rodgers, Willard L. 1982. "Estimable Functions of Age, Period, and Cohort Effects." *American Sociological Review* 47(6):774-787.

Ryder, Norman B. 1960. "The Structure and Tempo of Current Fertility." In *Demographic and Economic Change in Developed Countries*, edited by the National Bureau of Economic Research, 117-136. New York: Columbia University Press.

-----. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* **30**(6):843-861.

Shen, Yinzhi, Shawn Bushway, Lucy Sorenson, and Herbert Smith. 2019. "Locking Up My Generation: Cohort Differences in Prison Sentence Spells over the Life Course." Poster presented at the Annual Meeting of the Population Association of America, April 2019, Austin, TX.

Smith, Herbert L. 1987. "The Social Forecasting Industry." *Climatic Change* **11**(1-2):35-60.

-----. 2004. "Response: Cohort Analysis Redux." In *Sociological Methodology 2004*, edited by Ross M. Stolzenberg, 111-119. Washington, D.C.: American Sociological Association.

-----. 2008. "Advances in Age-Period-Cohort Analysis." *Sociological Methods & Research* **36**(3):287-296.

-----. 2009. « Application de l'analyse des séries chronologiques à la projection d'effectifs de population scolaire par la méthode des composantes ». *Cahiers Québécois de Démographie* **38**(1):145-170.

Smith, Herbert L., William M. Mason, and Stephen E. Fienberg. 1982. "More Chimeras of the Age-Period-Cohort Accounting Framework: Comment on Rodgers." *American Sociological Review* **47**(6):787-793.

Sobotka, Tomáš. 2003. "Tempo-quantum and Period-cohort Interplay in Fertility Changes in Europe: Evidence from the Czech Republic, Italy, the Netherlands and Sweden." *Demographic Research* **8**(6):151-214.

Vaisey, Stephen, and Omar Lizardo. 2016. "Cultural Fragmentation or Acquired Dispositions? A New Approach to Accounting for Patterns of Cultural Change." *Socius: Sociological Research for a Dynamic World* **2**:1-15.

Wang, Haidong, and Samuel H. Preston. 2009. "Forecasting United States mortality using cohort smoking histories." *Proceedings of the National Academy of Sciences* **106**(2):393-398.

Wickramaratne, Priya J., Myrna M. Weissman, Philip J. Leaf, and Theodore R. Holford. 1989. "Age, Period and Cohort Effects on the Risk of Major Depression: Results from Five United States Communities." *Journal of Clinical Epidemiology* **42**(4):333-343.

Wright, Erik Olin. 1997. *Class Counts: Comparative Studies in Class Analysis*. Cambridge: Cambridge University Press.

Yang, Guobin. 2016. *The Red Guard Generation and Political Activism in China.* New York: Columbia University Press.

Yang, Yang, Wenjiang J. Fu, and Kenneth C. Land. 2004. "A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models." In *Sociological Methodology 2004*, edited by Ross M. Stolzenberg, 75-110. Washington, D.C.: American Sociological Association.

Yang, Yang, and Kenneth C. Land. 2013a. *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. Boca Raton, FL: Chapman and Hall/CRC Press.

-----. 2013b. "Misunderstandings, Mischaracterizations, and the Problematic Choice of a Specific Instance in Which the IE Should Never Be Applied." *Demography* **50**(6):1969-1971.