"The interplay of lifecycle, history and generation – or age, period and cohort, in humbler terms – is one of the great puzzles in the study of individual and social change. This landmark collection is a major contribution on a topic that reaches from philosophy to social statistics. Don't be overtaken by time: start reading now!"
– Professor David Voas, University College London, UK

"Understanding age, period, and cohort effects is a mainstream concern in all the social sciences. But researchers have struggled with how exactly to approach 'APC' issues. This book provides invaluable, up-to-date methodological guidance."
– Professor Malcolm Fairbrother, Umeå University, Sweden

AGE, PERIOD AND COHORT EFFECTS

Statistical Analysis and the Identification Problem

Edited by Andrew Bell

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

6

LEARNING FROM AGE–PERIOD–COHORT DATA

Bounds, mechanisms, and 2D-APC graphs

Ethan Fosse, Christopher Winship and Adel Daoud

"The formulation of the problem is often more essential than its solution."
Albert Einstein

"A problem well put is half solved."
John Dewey

Introduction[1]

Social scientists have long sought to explain social change by using age–period–cohort (APC) models, attempting to estimate the unique contributions of age, period and cohort effects on a particular outcome. A major obstacle, however, has been the issue of model identification that arises from the linear dependence among age, period and cohort. Many solutions have been proposed to the identification problem, but none have gained wide acceptance. The lack of consensus, along with limited understanding of seemingly complicated models, has greatly hindered both methodological development and the application of APC methods to substantive analyses of social change.

As the quotes above suggest, solving a problem may require redefining it. In this chapter we present a set of related strategies in which the goal is to determine the unique contributions of age, period and cohort on an outcome. In doing so, we argue that the challenge is to not to achieve point identification but rather to determine what and how much can be learned from a particular APC dataset. In some instances, it may be possible to identify or nearly identify point estimates for APC causal effects using weak theoretical assumptions, but in others not. However, even if point identification is not possible, a great deal can often be learned from the data.

As we show below, the great advantage of redefining the goal of an APC analysis in this way is that it is possible to carry out a series of analyses using techniques

that have a solid methodological foundation, something that has eluded the great majority of past efforts and arguably is the much more fundamental problem. Our claim is not that we have 'solved' the APC identification problem, but rather that, by changing the definition of the problem, we have developed ways to rigorously analyze APC data involving transparent assumptions that can be clearly defended using substantive theory.

Central to our approach is the belief that there is an inherent trade-off between theoretical assumptions and what can be learned from data. The fewer the assumptions, the less is revealed. More assumptions mean that more can be learned, but what is learned will depend directly on the assumptions made. Given this trade-off, researchers need to demonstrate what can be learned with different assumptions. This is ultimately no different from any causal inference problem, which always requires information external to the data to justify a causal story.

Our suggested strategy for analyzing APC data consists of three general stages, each containing more detailed steps. The first stage (Analysis without assumptions) consists of two parts: an analysis of the nonlinear effects and different combinations of the linear effects.[2] As we discuss below, the nonlinear effects are identified under the standard assumption that there are no interactions between the APC variables. The nonlinearities may be of considerable interest in their own right and worthy of detailed analysis. Furthermore, in many cases it may be reasonable to assume that the absence of nonlinear effects for one or more of the three APC variables implies the likely absence of their related linear effects. We call this the Coupled Temporal Effects assumption (CTE).[3] If there are no nonlinear effects for one of the APC variables and the CTE assumption seems reasonable, identification then can be achieved by dropping that variable from the model and standard methods can be used.

The second step in the first stage is to carry out an analysis of the linear effects that are consistent with the data. This is done using what is called a 2D-APC graph that indicates that set of parameter values for the three APC variables that are consistent with the data. These values always fall on a one-dimensional line, what we term the canonical solution line.

What has not been previously recognized is that the location of the solution as determined by the data potentially rules out various combinations of linear effects. For example, there might be no point on the solution line where the linear age effect was positive and both the linear period and cohort effects are negative. Thus even in the total absence of assumptions, the data may rule out certain combinations of effects.

In the second stage (Partial Identification Using Bounds) we propose carrying out a bounds analysis. In many cases there may good theoretical reasons for assuming that the overall effect (i.e., the combined linear and nonlinear effects) of an APC variable is monotonically increasing or decreasing over some range of the variable. By monotonically increasing we mean that as the APC variable of interest increases, the outcome increases or at least remains constant. By monotonically decreasing we mean that as the APC variable of interest increases, the outcome decreases or at

least remains constant. For example, there is established biological theory as to why immunity to many diseases increases from infancy to childhood to adulthood, but then decreases at older ages. Similarly, theories of the human brain with regards to the development of the prefrontal cortex imply that an individual's attraction to risk will decrease as they transition from adolescence to adulthood.

Assumptions about whether or not ordered outcome values are monotonically increasing or decreasing are important in that they have the potential to bound the effects of the APC variables. As discussed below, fixing the value of any one APC value determines the values of the other two. Analogously, it is the case that bounding one variable potentially bounds the values of the other two variables. For instance, depending on the data, a monotonically increasing age effect might imply a monotonically decreasing period effect or a monotonically increasing cohort effect. The resulting bounds one obtains for each of the three APC variables may be wide or narrow. In our experience, it is often possible to achieve quite narrow bounds resulting in approximate point estimates for the effects of each of the three APC variables.

In the third and last stage (Mechanism-based models of APC effects), we recommend incorporating into one's analysis, when available, measured mechanisms (or mediators) of the bundles of causal effects thought to be related to the APC variables. As explained below, key here will be the assumption that the mechanism is affected by at most two of the APC variables. Mechanisms have the potential to explain the effects of the APC variables on an outcome. For example, below we examine the potential for educational attainment to explain age-related differences in religious disaffiliation in the United States. Importantly, whether one or more mediating variables explains the nonlinear effects of an APC variable is testable. If the nonlinear effects of an APC variable are explained and it is reasonable to assume that this implies its linear effect is also explained (as discussed in the CTE assumption above), then the APC model is fully identified and a full set of point estimates for each of the three APC variables can be obtained.

We suggest that researchers learn from APC data by carrying out their analysis following the three stages above. Each successive stage represents an analysis built on successively stronger assumptions. The first stage involves no assumptions. However, a researcher in finding the absence of nonlinear effects for one of the three APC variables may drop that variable invoking the CTE assumption, namely, that the absence of nonlinear effects of an APC variable implies the same for its linear effect. In this case, analyses can proceed using conventional methods. The second stage involves carrying out a bounds analysis in which assumptions about whether or not an APC variable is monotonically increasing or decreasing over some range of an APC variable is used to bound the effects of other APC variables. The third stage involves the inclusion of mediators and the assumption that they are affected at most by two of the APC variables. This last stage is often the most challenging, in that the researcher needs data on mechanisms to estimate the causal pathways.

The remainder of the chapter is organized as follows. In the next section we provide a very brief history of APC analysis before providing the logic of APC analyses.

In the following section we formally discuss how to organize temporal data, the relationship between models of temporal effects, and the nature of the identification problem. In the next sections we discuss the three stages of APC analysis.[4] We focus in particular on describing the usefulness of 2D-APC graphs and the importance of theoretical arguments in developing APC models. In each section we illustrate how these methods can be used by analyzing changes in religious disaffiliation. Specifically, we examine changing patterns of religious intensity from 1977 to 2018 as measured in the General Social Survey (GSS).

Background of APC analysis

Brief history

The history of APC analysis goes back to at least the 1860s, pre-dating Wilheim Lexis's introduction of the Lexis diagram in his book *Introduction to the Theory of Population Statistics* in 1875 (for a review, see Keiding 2011). In his well-regarded essay, written in the 1920s, the German sociologist Karl Mannheim demonstrated the importance of generations indicating how a population could change without individuals changing as a result of cohort replacement (Pilcher 1994). Wade Hampton Frost's analysis in the 1930s changes in tuberculosis rates marks the modern period of APC analysis in epidemiology (Frost 1995). Norman Ryder's classic (1965) paper on cohorts similarly delineates this for demography and sociology.

Over the decades social and behavioral scientists have researched a wide variety of topics using APC methods. Two general types of outcomes have been studied. Both epidemiologists and demographers have examined changes in rates of disease and other health-related behaviors, (Chen et al. 2003; Kerr et al. 2004; O'Malley, Bachman, and Johnston 1984; Vedøy 2014), obesity (Diouf et al. 2010; Fu and Land 2015), cancer (Clayton and Schifflers 1987; Liu et al. 2001) and mental health (Lavori et al. 1987; Yang 2008). Sociologists, demographers and others have examined a variety of outcomes related to social change, including verbal ability (Alwin 1991; Hauser and Huang 1997; Wilson and Gove 1999), social trust (Clark and Eisenstein 2013; Putnam 1995; Robinson and Jackson 2001; Schwadel and Stout 2012), party identification (Hout and Knoke 1975; Tilley and Evans 2014) and religious affiliation (Chaves 1989; Firebaugh and Harley 1991).

APC effects as bundles of unmeasured causes

In the APC literature, researchers have typically viewed age, period and cohort as indicators of distinct sets of unmeasured causes (e.g., see Mason and Fienberg 1985; Rodgers 1990). The notion that the APC variables are causal variables themselves is hard to support since it is not possible through manipulation to change an individual's age, the year they were born, or the current year. To quote Clifford Clogg (1982): 'age, period, and cohort are merely indicators of other variables which actually "cause" the observed variation in the dependent variable under

study. The age–period–cohort framework is properly interpreted as an accounting scheme, not a "causal model"' (460). Mason and Fienberg (1985) make the same point: 'these models do not explain so much as they provide categories with which to seek explanation. For accounting models to have value, the parameterizations of the general framework must be linked to phenomena presumed to underlie the accounting categories' (46–47).

If age, period and cohort can be understood as indices, what are they indices of? Generally, the APC literature describes the 'effects' of the three APC variables as consisting of a set of underlying causal processes each associated with one or more of the three APC variables.[5] For example, consider a study examining temporal effects in happiness. Age might be an indicator for the stress associated with having children of different ages. Period might reflect changes in the employment and political environment. Cohort might reflect differences in opportunities over the entire life course.

Figure 6.1 represents the general logic of the APC model with each of the three APC variables being associated with some underlying causal mechanism (*MA*, *MP* and *MC*), which are unobserved. The double-headed lines denote associational linkages between the three time scales and unobserved causal mechanisms. The directed arrows indicate causal relations between the unobserved mechanisms and the outcome. The logic represented by Figure 6.1 now clarifies what one might mean when one says that age, period or cohort have an 'effect': such a statement is just shorthand for the longer statement that the causal variables associated with each of the three APC have a causal effect on the outcome.

If the APC variables are simply indices, what then are the regression parameters associated with the three APC variables? The most basic way to understand a regression model is as a conditional expectation or conditional mean operator

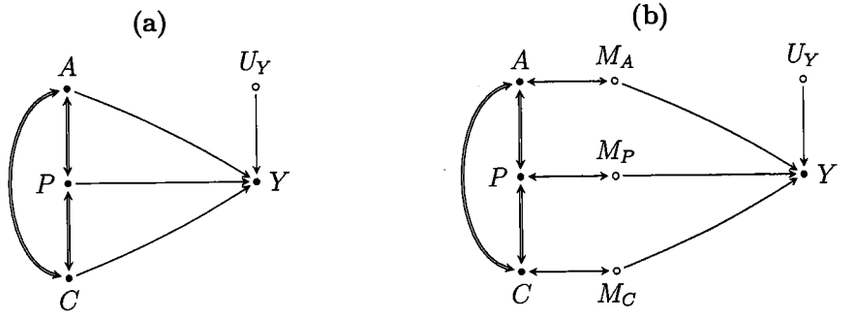

FIGURE 6.1 Graphical models of temporal variables
Notes: Panel (a) shows the simplified APC graphical model for age (A), period (P) and cohort (C). The double-line arrows denote the linear dependence among the three time scales. Filled circles denote observed variables, while hollow circles denote unobserved variables. Panel (b) shows the graphical model with a full set of mechanism variables (MA, MP and MC), which are unobserved. The double-headed lines denote associational linkages between the three time scales and unobserved causal mechanisms.

(Goldberger 1991). That is, a regression equation if properly specified indicates the mean of the outcome variable for individuals with fixed values for a specific set of X's. Thus, under this noncausal interpretation, a regression parameter indicates the difference in the conditional means for two groups with equivalent X's except that they differ by one unit with respect to the X whose parameter is of interest. For example, in a linear regression model where education is the outcome if the regression parameter for cohort is 0.25, this would indicate that a cohort born one year later, all the other X's being the same, would have 0.25 more years of education.

The logic of the APC identification problem

The APC identification problem, as it has become known, is simply the fact that if we know a person's age in years and the year in which their outcome was measured, then we know their birth year. That is, we know:

$$\text{cohort} = \text{period} - \text{age} \tag{6.1}$$

Suppose we have collected data on a set of individuals and have measured each person's birth year, age, year of measurement and their value on some outcome. An intuitive way to understand this problem is to use age, period and cohort variables as inputs in a multiple linear regression model:

$$Y = \mu + \alpha(\text{age}) + \pi(\text{period}) + \gamma(\text{cohort}) + \varepsilon \tag{6.2}$$

where Y is the outcome variable to be explained; μ is the intercept; age, period and cohort are measured in years; α, π and γ are the slopes for age, period and cohort, respectively; and ϵ is random error. For simplicity we have dropped the subscripts indexing each row (i.e., individual) of the dataset.

In Equation 6.2, we are attempting to estimate the effect of each variable holding the other variables constant. What provides information for estimating the effect of the variable of interest is the extent to which Y varies with that variable holding the control variables constant (i.e., varies within the levels of the control variables). Without loss of generality, assume that we want to estimate the linear effect of cohort (γ in Equation 6.2) holding age and period constant. Now consider only individuals of a certain age measured at a specific point of time: because of the perfect linear dependency among the three temporal variables, these individuals are not only the same age at the same period, but they also have the same birth year. There is no variance in cohort holding age and period constant, and as such it is impossible to estimate its linear effect.

For example, if we know a person was born in 1900 and that person's outcome was measured in 1950, then we know that person's age is 50. Then we are estimating the following equation for this person:

$$Y = \mu + \alpha(50) + \pi(1950) + \gamma(1950 - 50) + \varepsilon \tag{6.3}$$

Once we control for this person's age (50) and period (1950), then there is no information to estimate the linear cohort effect, γ, because the cohort is just the difference between period and age (1950 − 50).[6]

Modeling APC effects

Organizing temporal data

Previously in Equation 6.2 we outlined a simple linear regression model based on data in which each row is a different individual with columns consisting of the variables age, period, cohort and the outcome. More commonly, data on temporal effects is arranged in a rectangular age–period array of observations,[7] with age and period aggregated into intervals of equal width (Holford 1983: 311–318; Mason and Fienberg 1985: 67–68).[8] For example, Table 6.1 displays the data structure of an age–period array with five age groups and three period groups, both aggregated into five-year intervals. The number of cohort groups equals one less than the sum of the age and period groups, so there are 3 + 5 − 1 = 7 cohorts in Table 6.1. The number of observations on the outcome equals the number of cells in the array, which is simply the number of age groups multiplied by the number of period groups. Thus, in Table 6.1 there are $3 \times 5 = 15$ observations.

Cohort groups are labeled in the cells, calculated by cross-referencing corresponding age and period groups.[9] For instance, the cohort interval 1956–1964 is derived from the range of possible birth years for those observed during the period interval 2000–2004 and age interval 40–44. The history of each cohort can be traced along the diagonals of the table, moving forward in time from the upper-left to the lower-right. For example, the shaded diagonal set in Table 6.1 refers to the cohort born during the years 1956–1964. Following from the upper-left shaded cell to the lower-right shaded cell, we can track the cohort of people born in 1956–1964 as they advance in age from 40–44 to 50–54 and move through time from 2000–2004 to 2010–2014.

TABLE 6.1 Structure of a Lexis table with interval values of equal width for age and period groups

Age groups	*2000–2004*	*2005–2009*	*2010–2014*
30–34	1966–1974	1971–1979	1976–1984
35–39	1961–1969	1966–1974	1971–1979
40–44	1956–1964	1961–1969	1966–1974
45–49	1951–1959	1956–1964	1961–1969
50–54	1946–1954	1951–1959	1956–1964

Notes: Age and period aggregated into five-year intervals, with corresponding nine-year cohort intervals labeled in the cells. Shaded cells track the history of the cohort born in 1956–1964.

Classical APC model

It is common in the APC literature to use index notation to keep track of the dimensions of a temporal data structure such as that shown in Table 6.1 (Mason and Fienberg 1985: 67–71). We will let $i = 1,\ldots, I$ represent the age groups, $j = 1,\ldots, J$ the period groups, and $k = 1,\ldots, K$ the cohort groups with $k = j - i + 1$ and $K = I + J - 1$.[10] Using this index notation, temporal effects in an age–period array can be represented using the classical APC (C-APC) model, also known as the multiple classification model (Mason, Mason, et al. 1973: 243) or accounting model (Mason and Fienberg 1985: 46–47, 67), which has the following form (Mason and Fienberg 1985: 67–68; Yang and Land 2013: 61):

$$Y_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k + \varepsilon_{ijk} \tag{6.4}$$

where Y_{ijk} is the outcome variable to be explained, μ is the intercept, α_i represents the ith age effect, π_j represents the jth period effect, γ_k represents the kth cohort effect, and ε_{ijk} is the error term. To avoid overparameterization, we apply the so-called usual constraints that the parameters sum to zero, such that $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \pi_j = \sum_{k=1}^{K} \gamma_k = 0$.[11]

The parameterization shown in Equation 6.4 is very flexible, allowing the age, period and cohort effects to be highly nonlinear because there is one parameter for each age, period and cohort category (Mason, Mason, et al. 1973: 246). However, like Equation 6.2, the C-APC suffers from a fundamental identification problem due to perfect linear dependence in the columns (Yang and Land 2013: 63). The linear dependence can be difficult to spot visually, but what it means in practice is that at least one of the temporal variables must be dropped.

Linearized APC model

To clarify the nature of the identification problem with the C-APC, it is useful to provide an alternative representation of the C-APC that orthogonally decomposes the linear from the nonlinear components (Holford 1983, 2006). We can accordingly specify a linearized APC (L-APC) model with the form:

$$Y_{ijk} = \mu + \alpha\left(i - i^*\right) + \pi\left(j - j^*\right) + \gamma\left(k - k^*\right) + \tilde{\alpha}_i + \tilde{\pi}_j + \tilde{\gamma}_k + \varepsilon_{ijk} \tag{6.5}$$

where the asterisks denote midpoint or referent indices and are

$$i^* = \left(\frac{1+1}{2}\right), j^* = \left(\frac{J+1}{2}\right), \text{ and } k^* = \left(\frac{K+1}{2}\right)$$

As before, we refer to the linear effects as α, π and γ for age, period and cohort. However, we now introduce $\tilde{\alpha}$, $\tilde{\pi}$ and $\tilde{\gamma}$ to represent age, period and cohort nonlinearities, respectively. The L-APC model is based on setting up APC data so that the linear and nonlinear components are orthogonal to each other (see Fosse and Winship 2018).

The C-APC and L-APC are equivalent representations of the temporal data shown in Table 6.1. As with the C-APC, each cell in an age–period array is modeled by a unique combination of parameters under sum-to-zero constraints. For example, the ith age effect in the C-APC is represented in the L-APC by the overall linear age effect along with a unique parameter for the ith age nonlinearity: $\alpha_i = \left(i - i^*\right)\alpha + \tilde{\alpha}_i$. That is, each age effect α_i is decomposed into the sum of a common parameter α representing the (linear) age slope for the entire array, with a value shifting across rows (or age categories) as a function of the age index $_i$, and a unique parameter $\tilde{\alpha}$ which is a nonlinearity specific to each row (or age category) of the array. We can similarly decompose the period and cohort effects into linear and nonlinear components.

The importance of the L-APC model is that, by explicitly separating the slopes from their deviations, it clearly shows that the identification problem is limited to the linear effects.[12] When the nonlinear terms are zero in the population, the L-APC is equivalent to the basic linear model:

$$\begin{aligned} Y_{ijk} &= \mu + \alpha\left(i - i^*\right) + \pi\left(j - j^*\right) + \gamma\left(k - k^*\right) + (0) + (0) + (0) + \varepsilon_{ijk} \\ &= \mu + \alpha\left(\text{age}_i\right) + \pi\left(\text{period}_j\right) + \gamma\left(\text{cohort}^k\right) + \varepsilon_{ijk} \end{aligned} \tag{6.6}$$

where age_i, period_j and cohort_k are the midpoint values for each of the categories (see Table 6.2), which are simply the indices recentered and rescaled, and the nonlinearities are zeroed out.[13] A useful way to understand the nonidentifiability

TABLE 6.2 Structure of a Lexis table with midpoint values for age, period and cohort

	Period		
Age	*2002*	*2007*	*2012*
32	1970	1975	1980
37	1965	1970	1975
42	1960	1965	1970
47	1955	1960	1965
52	1950	1955	1960

Notes: Age and period values are midpoints of five-year intervals, while cohort values are midpoints of nine-year intervals calculated by cross-referencing age and period intervals. Shaded cells track the history of the cohort with a midpoint birth year of 1960.

problem is to note that for any particular APC model we can specify the linear effects as (Rodgers 1982: 782):

$$\begin{aligned} \alpha^* &= \alpha + v \\ \pi^* &= \pi - v \\ \gamma^* &= \gamma + v \end{aligned} \tag{6.7}$$

where the asterisk (*) indicates an arbitrary set of estimated slopes from an APC model and v is a scalar fixed to some value. As Equation 6.7 demonstrates these parameters are simple additive rescalings of the true unobserved slopes α, π and γ shifted by a single arbitrary scalar, v.

A three-stage approach to analyzing APC data

We are now in a position to discuss how to actually analyze APC data using our three-stage approach. We illustrate our approach by examining changes in religious disaffiliation from 1977 to 2018 using the GSS. The outcome is religious intensity, which measures the respondent's strength of religious affiliation on a scale from 1 to 4, where 1 = no religious affiliation, 2 = not very strong religious affiliation, 3 = somewhat strong religious affiliation, and 4 = strong religious affiliation.[14] After subsetting to respondents born in the United States, this gives us a sample of 48,598 respondents. Age and period are grouped into five-year intervals. To deal with the complex sampling of the GSS, estimates are adjusted using the appropriating sampling weights.

Because the predicted values of an APC model are identified under any particular constraint, we can estimate the expected average religious intensity for various age, period and cohort groups using the full linearized APC model.[15] These results are shown in Figure 6.2. Although it might be tempting to informally 'eyeball' the unique contributions of the temporal effects from this graph, it is important to note that each of the expected averages in Figure 6.2 are based on the combined effects of age, period and cohort. The patterns in this graph simply tell us that religious intensity is generally lowest among younger individuals measured during the most recent periods who are also members of the most recent birth cohorts, while religious intensity is highest among older individuals measured in earlier periods who are also members of earlier cohorts. To determine the unique effects of age, period and cohort, we need a more formal approach, which we outline below in three stages.

Stage 1: analysis without assumptions

The APC literature has generally failed to recognize that data are informative about the three APC effects, even in the absence of any assumptions.[16] The data are informative in two ways. First, as discussed previously, the nonlinear APC effects are identified. This means that they can be estimated and graphed. Furthermore, it is possible to test whether or not these effects are zero. If various tests or visual

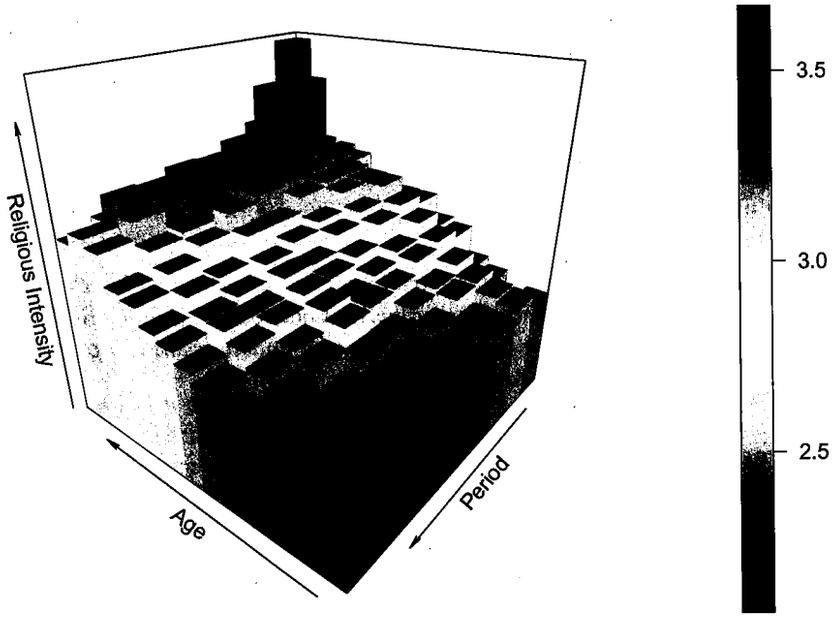

FIGURE 6.2 3D Lexis histogram of religious intensity
Notes: Results are based on a full linearized APC model with the period linear effect fixed to zero. Each cell gives the expected average religious intensity on a scale from 1 to 4, where 1 = no religious affiliation, 2 = not very strong religious affiliation, 3 = somewhat strong religious affiliation, and 4 = strong religious affiliation. Estimates are adjusted using appropriate sampling weights. Estimates based on 59,573 respondents in the United States GSS.

inspection suggest that it is reasonable to assume that one or more of the APC nonlinear effects are zero, then it may reasonable to assume that there is no linear effect as well (the CTE assumption) and drop these variables from the model, thus suggesting a fully identified analysis is appropriate.

Second, particular combinations of linear effects are identified, which define a set of solutions. Moreover, as discussed in detail below, because the set of solutions lie on a line, certain combinations of effects are ruled out. This means that particular theories may be shown to be inconsistent with the data. In general, the data are likely to be consistent with multiple patterns of effects, but not all combinations of APC effects. As such, the data, in the absence of any assumptions, may well provide evidence against certain theoretical claims.

Nonlinear effects: no assumptions needed

Because the nonlinear effects are identified, we can get point estimates of their values. The corresponding figure shows the nonlinear effects for age, period and cohort on religious disaffiliation.

As can be seen in Figure 6.3, there are clear nonlinear fluctuations in the data for all three temporal variables.[17] In Table 6.3 we present a set of tests for helping to ascertain whether or not the nonlinear effects for age, period or cohort should be dropped. For example, the AP Model in Table 6.3 refers to a model in which the nonlinear effects for cohort have been dropped. Both visual inspection of the graphs for the nonlinear effects as well as these results allow the researcher to determine whether or not – in terms similar to those outlined by Yang and Land – all three APC dimensions are operative. If, based on these tests, one is willing to assume that one or more effects are not operating, one can drop the corresponding nonlinear components. Accordingly, one can fit the sub-model rather than the full APC model.[18] Table 6.3 includes the model log-likelihoods, AIC and BIC fit statistics, as well as chi-square statistics from Wald tests comparing the candidate model with the full APC model. The AIC favors the model with period and cohort nonlinear effects, suggesting that it might be reasonable to favor a model in which the age linear effect is zero. In contrast, BIC, which prefers parsimonious models, suggests one should just fit a model with period nonlinear effects, implying that the age and cohort linear effects could be fixed to zero. In this case, we would want to be careful about making assumptions about any of the effects, because model results (not shown here) indicate that they are all statistically significant at conventional levels, notwithstanding the imprecision of the estimates for the cohort effects due to the unbalanced nature of the data.[19]

Constraints on linear effects absent assumptions

Without data, the age, period and cohort parameters can take on any combination of values in a three-dimensional space. The data, however, constrains all of the effects to lie on a line determined by the unidentified linear parameters. This implies that if we fix anyone of the three linear effects, the values of the other two are determined.

The canonical solution line

To appreciate how the data constrains the linear effects, one needs to understand the geometric representation of the linear dependence problem. Let $\theta_1 = \alpha + \pi$ and $\theta_2 = \gamma + \pi$. Figure 6.4(a) shows the age–period plane defined by the identified quantity for hypothetical data where $\theta_1 = 3$, while Figure 6.4(b) shows the period–cohort plane defined by the identified quantity $\theta_2 = -2$. Intersecting these two planes defines a line, as shown in Figure 6.4(c) and Figure 6.4(d). This is what is known as the *canonical solution line* as all points on it represent parameter estimates for α, π and γ that are equally consistent with the data.[20] As such, this visually represents the APC identification problem. If the age, period and cohort parameters were identified, the planes would intersect at a single point in the parameter space. Here the intersection consists of all points along a line. The scalar v in Equation 6.7 essentially moves the possible estimates up and down the canonical solution line.

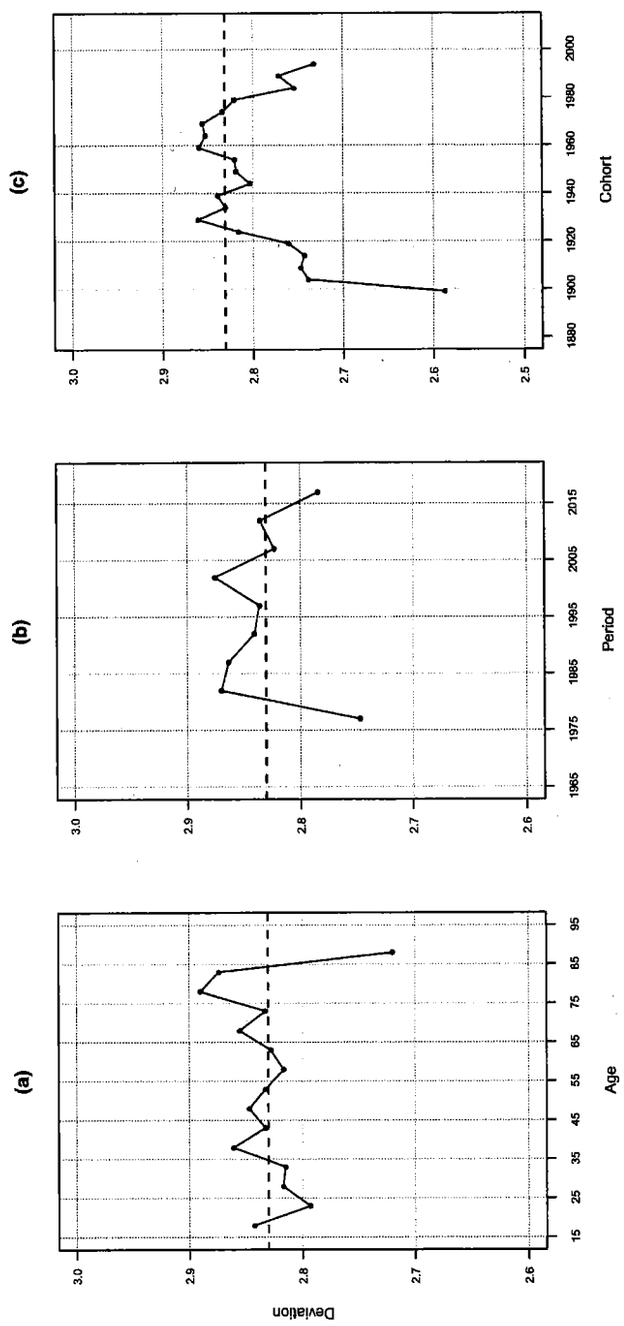

FIGURE 6.3 Nonlinear effects: religious intensity

TABLE 6.3 Fit statistics for various APC models: religious intensity

	Log-likelihood	*AIC*	*BIC*	*Chi-square*	*P-value*
Full APC Model	56,343.81	56,466,78	56,818,63		
PC Model	56,403.80	56,452,78	56,698,78	16.70	0.213
AC Model	56,403.80	56,505.65	56,785.81	39.86	<0.0001
AP Model	56,411.18	56,474.52	56,647.26	49.59	0.0004
A Model	56,482.29	56,524.13	56,622.24	97.24	< 0.0001
P Model	56,440.18	56,467.81	56,528.50	67.46	0.0005
C Model	56, 425.38	56,491.92	56,666.70	57.31	< 0:0001

Notes: All models include the same set of linear effects, which are estimated under the constraint that the period linear effect is zero. Chi-square statistics and p-values based on Wald tests comparing the full APC model with the specified candidate model.

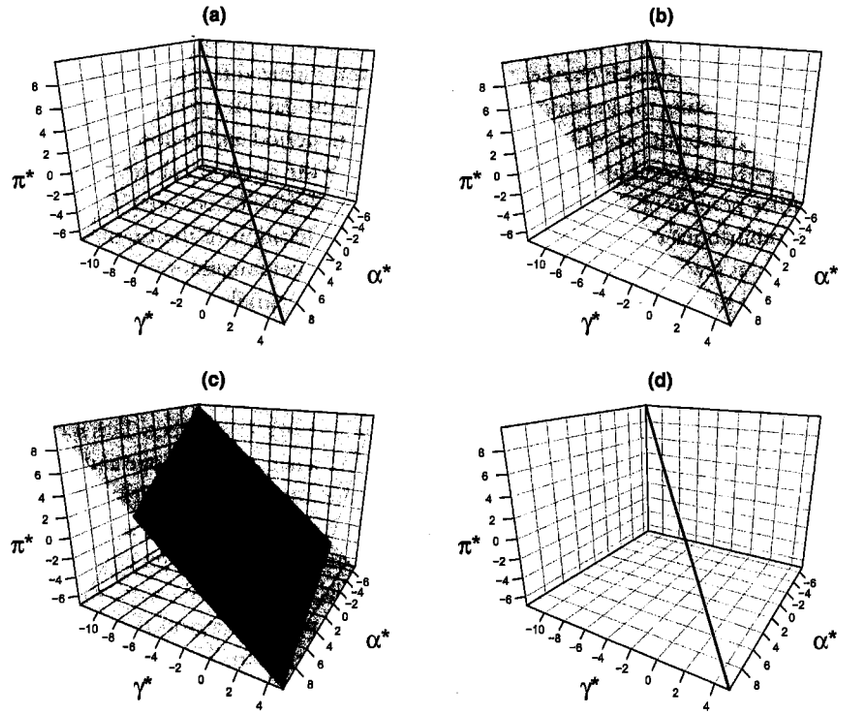

FIGURE 6.4 Geometric derivation of the canonical solution line

What is important to appreciate is what is accomplished when data are used in an APC model. If there are only linear effects, the data has taken us from a situation where all parameter values in a three-dimensional space are possible to one where only estimates lying on a one-dimensional line are consistent with the data. This same reduction also occurs if our model has nonlinear effects as they are fully identified.

As Fosse and Winship (2018) demonstrate, the solution line always sits in a three-dimensional subspace of the overall space of possible parameter values. Thus, the data has substantially reduced the possible estimates for the linear effects of age, period and cohort to values that lie on a single one-dimensional line. The data are quite informative about parameter values, just not informative enough to give us point estimates for the linear effects. That said, the data, by constraining values to a single line, contains considerable information about the possible values of the APC parameters.

2D-APC graphs

There is a useful way to simplify the representation of the solution line in the previous section. Because of the linear relationships $\alpha + \pi = \theta_1$ and $\pi + \gamma = \theta_2$, our three-dimensional representation can be reduced to just two. A way of doing this is by having the horizontal axis represent the period slope, the left vertical axis the age slope and the right vertical axis the cohort slope. Fosse and Winship (2019) call this a 2D-APC graph.

The 2D-APC graph clarifies an important, unrecognized fact in the APC literature that, by fixing the location of the solution line, the data also determines which of the eight combinations of positive and negative age, period and cohort linear effects are empirically possible. Because the offset between the age and cohort slopes must be either positive or negative as determined by the difference between the θ's, only six combinations can exist. Then, depending on the location of the canonical solution line, as few as two and as many as four remaining combinations may be possible. Figure 6.5 shows an example of a 2D-APC graph using simulated data, where $\theta_1 = 3$ and $\theta_2 = -2$. There are six regions of the parameter space defined by the signed combinations of the slopes (α, π, γ). In the figure these are labeled as regions **I** (+,−,+), **II** (+,+,+), **III** (+,−,−), **IV** (+,+,−), **V** (−,−,−) and **VI** (−,+,−). Note first that there is no region representing either the (−,+,+) or (−,−,+) patterns of effects. This is due to the fact that $\theta_2 - \theta_1 < 0$. Thus, any theory that posits a negative age slope and positive slopes for period and cohort can be ruled out by the data alone. Similarly, we can rule out any theory that assumes that age and period have negative slopes while cohort has a positive slope. Furthermore, the canonical solution line runs through only four out of the six regions, so we can also rule out any social or biological theory that posits that the linear age, period and cohort effects are all positive (region **II**) or all negative (region **V**). Thus, despite the linear dependence problem, the data can eliminate a number of possibilities based on theorizing just about the direction of the slopes.

Turning to our empirical example, Figure 6.6 shows the bounding regions in the 2D-APC graph for religious intensity. The data, via the values of θ_1 and θ_2, determines the location of the solution line in the parameter space. The estimates of these parameters for religious intensity are displayed in Table 6.4. Also note that the data determines which of the signed plotting regions are even possible. This graph shows the various signed plotting regions: **I** (+,−,+), **II** (+,+,+), **III** (−,−,+), **IV** (−,+,+), **V** (−,−,−) and **VI** (−,+,−). Unlike the previous example with simulated data, with the religious disaffiliation data there is no region representing either the

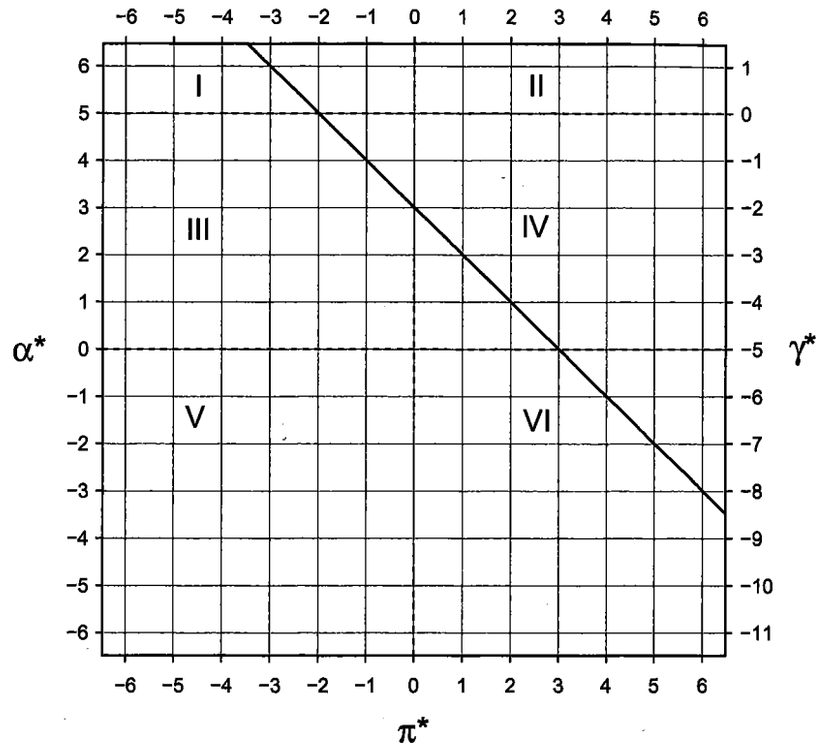

FIGURE 6.5 Example of a 2D-APC graph
Notes: This graph shows the various signed plotting regions: **I (+,−,+), II (+,+,+), III (+,−,−), IV (+,+,−), V (−,−,−)** and **VI (−,+,−)**. Note first that there is no region representing either the **(−,+,+)** or **(−,−,+)** patterns of linear effects.

TABLE 6.4 Intercept and combined linear effects: religious intensity

Parameter	*Est.*	*SE*	*t-ratio*	*P-value*
Intercept	2.8307	0.0129	219.49	< 0.0001
θ_1 or $(\alpha + \pi)$	0.1538	0.0351	4.38	< 0.0001
θ_2 or $(\gamma + \pi)$	−0.4760	0.0377	−12.6379	< 0.0001
$\theta_1 - \theta_2$ or $(\alpha + \gamma)$	−0.6298	0.0308	−20.46	< 0.0001
$\theta_2 - \theta_1$ or $(\gamma + \alpha)$	0.6298	0.0308	20.46	< 0:0001

Notes: Coefficients and standard errors based on the full linearized APC model.

(+,−,−) or (+,+,−) patterns of slopes. We can rule out any theory that makes the assumption that the age slope is positive while both the period and cohort slopes are negative. Likewise, we can rule out any theory that claims that the cohort slope is negative while both the age and period slopes are positive. Furthermore, note that

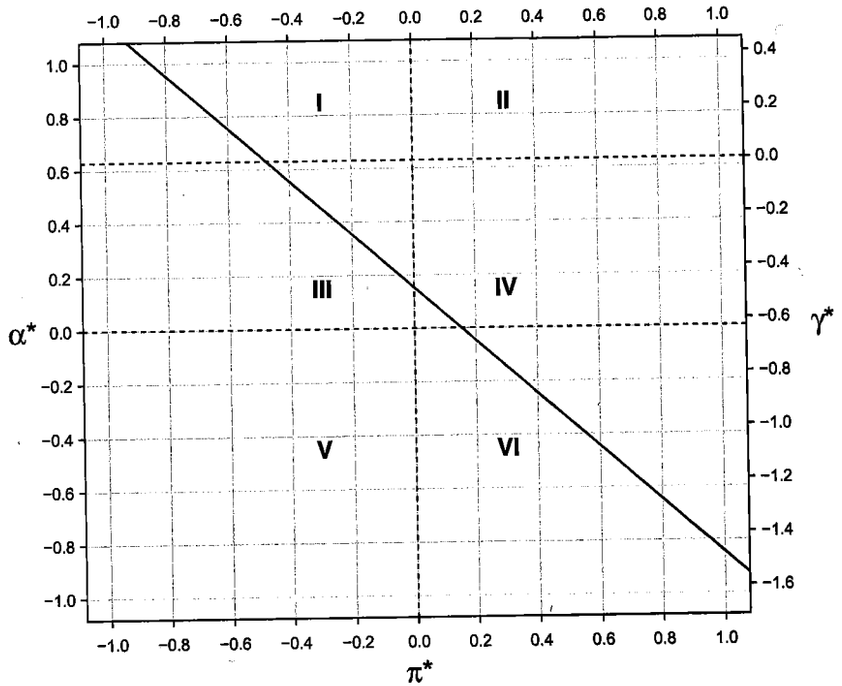

FIGURE 6.6 2D-APC graph: religious intensity
Notes: This graph shows the various signed plotting regions: **I (+,–,+), II (+,+,+), III (–,–,+), IV (–,+,+), V (–,–,–)** and **VI (–,+,–)**. Note first that there is no region representing either the **(+,–,–)** or **(+,+,–)** patterns of linear effects. Solid points denote zero-slope models, while hollow dots denote the age-period and cohort-period origins (i.e., where the age and period slopes are both zero as well as where the cohort and period slopes are both zero).

the canonical solution line does not go through regions **III** or **V**. In other words, we can reject any theory that assumes that the cohort slope is positive while both the age and period slopes are negative; additionally, we can reject any theory that assumes that all three slopes are negative. Another way of stating the above is that only the combinations of slopes in regions **I, II, IV** and **VI** are consistent with the data.

Combining the linear and nonlinear effects

If we had estimates of both the linear and nonlinear effects we could combine these to get the total effects of the three APC variables. Understanding how to do this is a useful preliminary to understanding how to combine point estimates of the nonlinear effects with bounds on the linear effects. We show how to do so using a hypothetical example.

Previously we have outlined how the C-APC model is related to the L-APC model. The advantage of using the L-APC is that we can focus on those parts of the

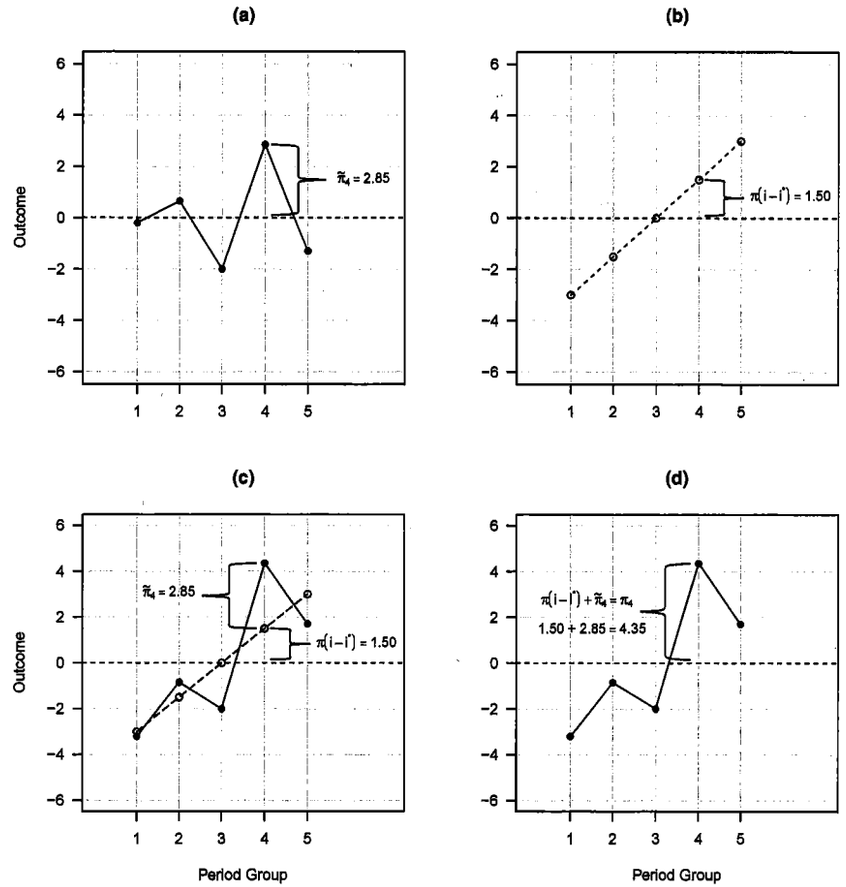

FIGURE 6.7 Combining linear and nonlinear APC effects
Notes: Panel (a) shows the nonlinear effects for a period variable (solid line). Panel (b) shows the assumed value of the period slope (dashed line). Panel (c) visualizes the combination of the linear and nonlinear effects as a solid line. Panel (d) shows the combined linear and nonlinear effects, resulting in the overall effects (solid line).

APC effects that are not identified in the standard model – that is, the linear effects. Once we use a 2D-APC graph to obtain a range of values for the linear effects, we can combine these effects with the nonlinear effects to obtain the overall effects. To illustrate how we can use the separate linear and nonlinear effects to obtain the overall effects, see Figure 6.7.

It is relatively simple to determine the total effects: we just need to add the nonlinear and linear effects together. In this hypothetical example, we have a set of nonlinear period effects that are identified: $\pi_1 = -0.20$, $\pi_2 = 0.65$, $\pi_3 = -2.00$, $\pi_4 = 2.85$, $\pi_5 = -1.30$. We also have a set of linear effects that are not identified, which can be represented generically as $\pi(j - j^*)$. With $J = 5$ period groups and $\pi = 1.5$, assume the following values for the period groups: $\pi(1 - 3) = -3$,

$\pi(2-3) = -1.5, \pi(3-3) = 0, \pi(4-3) = 1.5, \pi(5-3) = 3$. Adding the linear and nonlinear effects together gives us the overall effects: $\pi_1 = -3.20, \pi_2 = -0.85, \pi_3 = -2.00, \pi_4 = 4.35, \pi_5 = 1.70$. Figures 6.7(a) and (b) show the overall effects and how they are the simple sum of the linear and nonlinear effects. For example, the fourth period group has an overall effect of $\pi_4 = 4.35$, which is simply the sum of the period linear and nonlinear effect for that group, or 1.50 + 2.85.

The above illustration is vital for understanding our overall approach to APC models. By splitting the unidentified temporal effects of the C-APC into identified and unidentified linear effects, we can focus on that portion of these effects that can be estimated from the data. This idea of combining the linear and nonlinear effects extends easily to actual data. In Figure 6.8, we examine religious disaffiliation effects under the assumption that the age linear effect is zero. Specifically, these results are based on first estimating the nonlinear effects, assuming that the age linear effect is zero, and then using the estimated values of θ_1 and θ_2 to calculate the values of the period and cohort linear effects. Because we have obtained a single set of estimates, we have achieved what is called *point identification*. However, in practice one will want to specify a range of values of the effects, reflecting the inherent uncertainty of the linear effects, thereby achieving *partial identification*. We cover this topic in the next section on bounding analyses of APC effects.

Stage 2: partial identification using bounds

Bounding analyses entail restriction on the parameter space of the 2D-APC graph (or, equivalently, fixing a constraint). With bounding analyses, no direct measures of U_A, U_P or U_C are included in the model. Rather, constraints are added that entail assumptions about the effect of these underlying causes on the outcome of interest. One new approach is to specify bounds based on the sign, size or shape of one or more of the temporal effects. It is crucial to underscore that in many cases by constraining the direction of one temporal slope we can make conclusions about the direction and magnitude of at least one of the other slopes. To illustrate this, consider Figure 6.9(a)–(d). The graphs show various bounds on the canonical solution line using data on religious intensity.[21] Figure 6.9(a) shows that we can reject any theory that claims that the age and period slopes are both negative; this is easily falsified by the data. In other words, either age or period has a positive linear effect. Likewise, Figure 6.9(b) illustrates that it is impossible for there to be a positive age slope and a negative cohort slope. There is no point on the canonical solution line in which this combination of linear effects exists.

The above discussion mirrors our previous overview of the signed regions of the parameter space in a 2D-APC graph. However, we can go a step further and use a 2D-APC graph to obtain finite bounds (i.e., bounds that do not entail positive or negative infinity). In general, any particular APC dataset should have at least two combinations of slopes that, with a minimal assumption about the direction of the slopes, should result in finite bounds. Figure 6.9(c) shows that we can obtain narrow bounds by assuming that the age and period slopes are both positive. This assumption results in a relatively narrow bound on the cohort slope. Similarly,

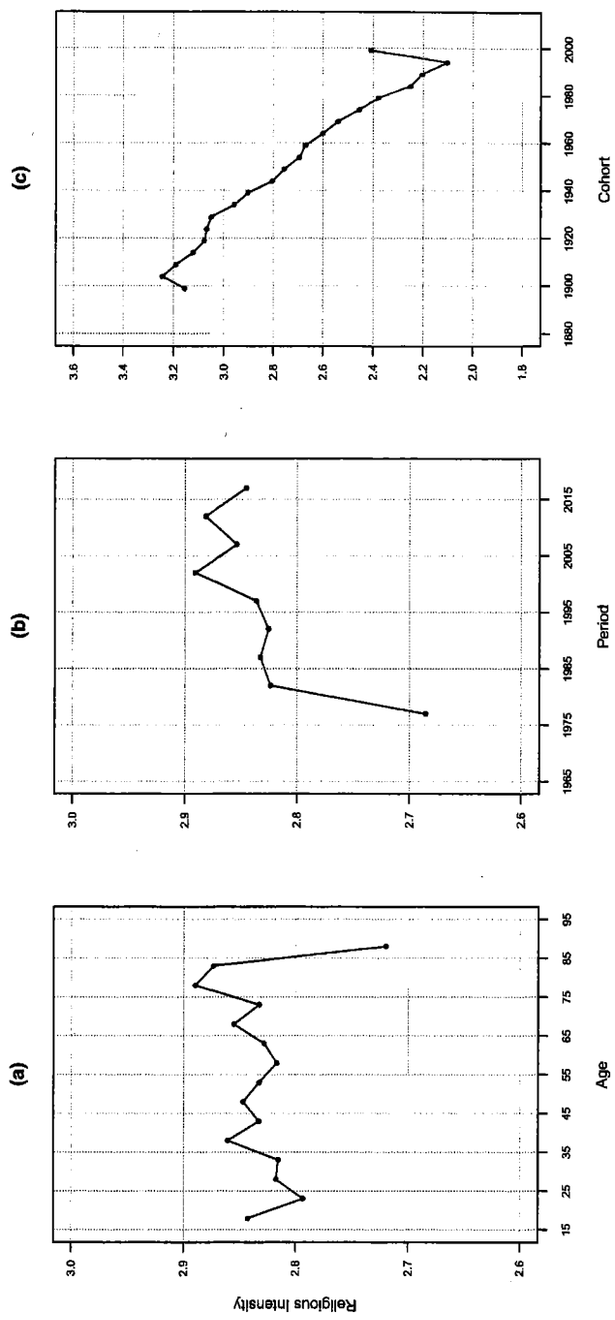

FIGURE 6.8 Combining linear and nonlinear APC effects: religious intensity

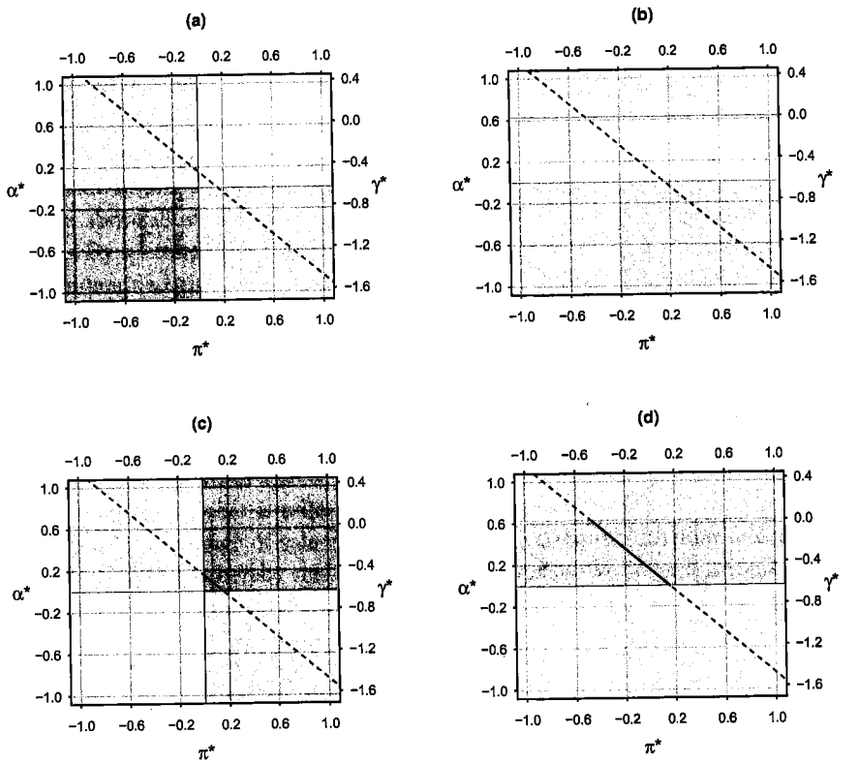

FIGURE 6.9 2D-APC graphs: examples of bounding strategies

Figure 6.9(d) illustrates the bounds, assuming that the age slope is negative and the cohort slope is positive. Given these assumptions, we can conclude that the period slope must be positive within a relatively narrow range of values.

The above bounds are based only on assumptions about the sign (or direction) of the linear effects. This is overly restrictive, because in practice we have data not only on the linear effects but the nonlinear effects as well. As a result, we can restrict not just the sign but the size of the slope by making assumptions about the shape of the temporal effects. For example, assume that we have strong theoretical reasons to believe that the overall set of effects in age is monotonically increasing. A set of simulated age nonlinear effects is displayed in Figure 6.10. As can be seen in Figure 6.10(a), the set of effects is at some groups increasing and at other groups decreasing. We want to specify a value for the linear age effect that ensures that, between any two adjacent age categories, the pair of effects is flat. To do so we need only find that pair of age adjacent categories in which the downward trend is most negative. This can be computed using the following equation: $\alpha_{m.i.} = (-1) \times min(\Delta\alpha_{I-1})$. The Δ notation indicates forward differences (i.e., $\alpha_{I+1} - \alpha_I$), and the subscript *m.i.* indicates the linear effect is monotonically increasing. We can use this equation to calculate the lower bound for a monotonically increasing set of effects.

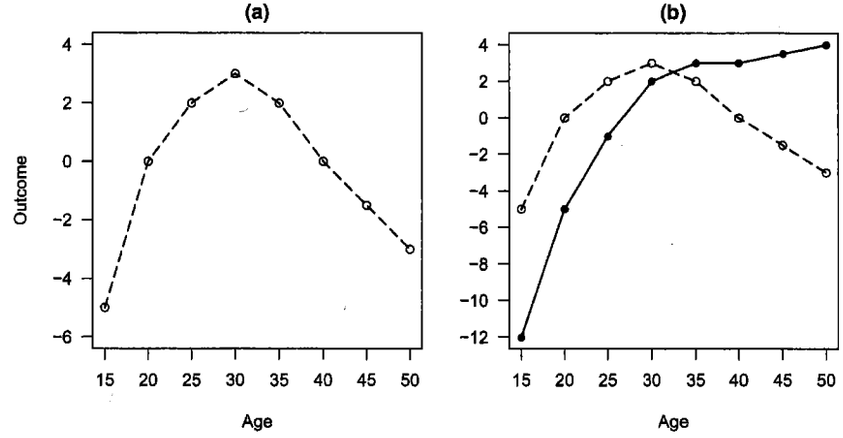

FIGURE 6.10 Specifying monotonicity constraints

The forward differences for the age nonlinearities are $\Delta\alpha_{I-1} = \{5,2,1,-1,-2,-1.5,-1.5\}$. The minimum of these differences is -2, which is between ages 35 and 40. To counter this downward deviation, the parameter value for the linear age term must be greater than or equal to +2. In Figure 6.10(b) we show what happens to the set of age effects when the slope is set to +2, which is illustrated by the solid line. As can now be seen the overall set of age effects is monotonically increasing. Any slope less than +2 will result in an overall effect that is neither monotonically increasing nor decreasing over the full set of age groups. Alternatively, any slope greater than +2 will be monotonically increasing; +2 is the minimum monotonically increasing age slope consistent with our data.

By incorporating nonlinearities, we are able to make weaker assumptions about the sign of the slopes. For example, for the age slope we can specify the range as $-\infty \leq \alpha_{m.d.} \leq 0 \leq \alpha_{m.i.} \leq +\infty$. The range $\alpha_{m.d.} \leq 0 \leq \alpha_{m.i.}$ is a weaker version of assuming a zero slope for age, since it takes into account some deviations above and below so long as the deviations are neither monotonically increasing nor decreasing. Similarly, instead of assuming the age slope is positive, we can say that $\alpha_{m.d.} \leq 0 \leq +\infty$. Finally, rather than assuming the age slope is negative, we can say that $-\infty \leq 0 \leq \alpha_{m.i.}$, which allows for any set of effects except those that are monotonically increasing.

Turning to our empirical example, we can consider placing bounds over one or more of the temporal effects using monotonicity constraints. Based on the fit statistics presented earlier, we might assume not that the age linear effect is zero, but rather that the age effects are neither monotonically increasing nor decreasing for some set of age groups. In Figure 6.11, we display the upper and lower bounds for the APC effects under the assumption that the age effects are neither monotonically increasing nor decreasing during middle age, defined here as the age groups 28–32 to 63–67. These constraints restrict the age linear effect within the range -0.4536 to 0.2772. We can conclude that, under this assumption, the great decline in religious intensity is mainly attributable to cohort effects.

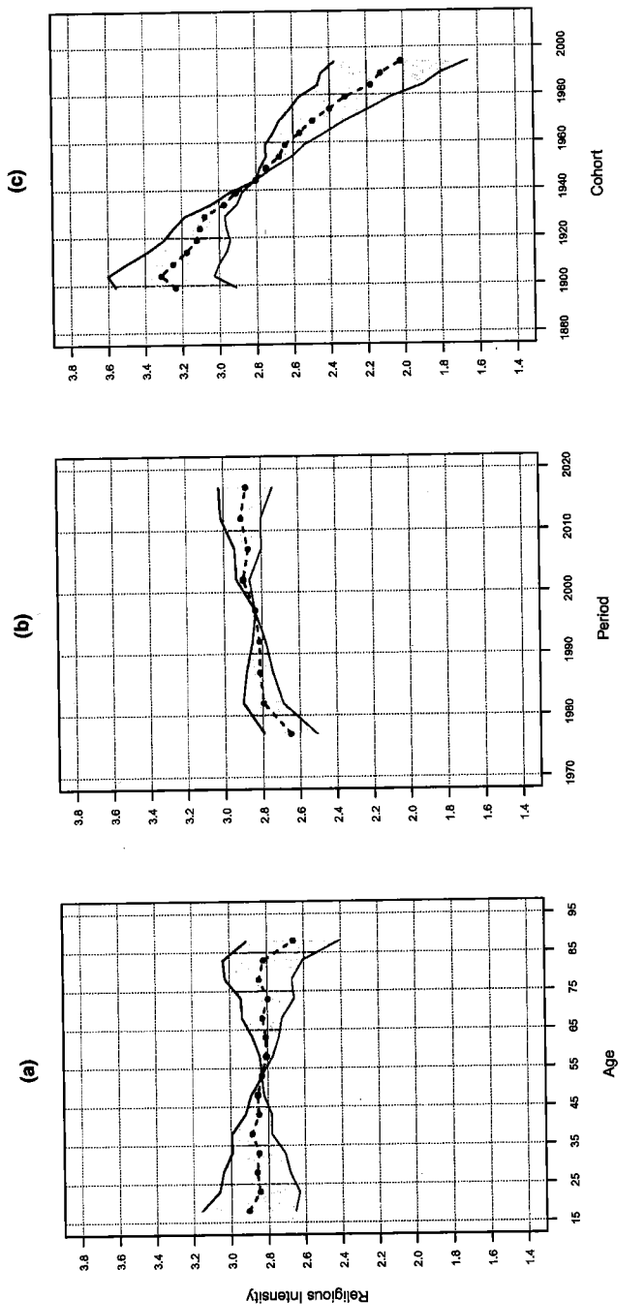

FIGURE 6.11 Specifying bounds with monotonicity constraints: religious intensity

To summarize, we can restrict the range of plausible estimates using a variety of bounding strategies, especially those about the sign, size and/or shape of the temporal effects. These should, at a minimum, reflect social or biological theories about the true causal effects thought to underlie the temporal variables. Although just-identified models fit the data equally well (and are thus observationally equivalent), some parameter values can be deemed as more plausible than others based on sociological or biological theory.

Stage 3: mechanism-based models of APC effects

Bounding analyses are based on the idea that we do not have measured mechanisms linking the APC variables to the outcome. However, in many instances we have causal mechanisms available. Such mechanisms can greatly aid in the identification of temporal effects. The mechanism-based approach identifies the causal effect of APC variables on an outcome by specifying at least one of the pathways between the temporal variables and the outcome (Winship and Harding 2008).

Incorporating observed mechanisms in an APC analysis

The first instance of the mechanism-based approach is Duncan's (1985) model shown in Figure 6.12, but it is relatively undeveloped and the causal assumptions are not fully specified. However, we present Duncan's example here because it helps to reveal the nature of the assumptions. In Duncan's basic model, M is a mechanism linking cohort, denoted by C, to the outcome, given by Y. In Duncan's particular instance, M is education and the outcome Y is a measure of generational attitudes.[22] This model identifies the putative effect of C on Y as long as M is the

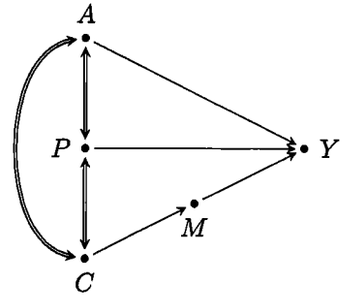

FIGURE 6.12 Duncan's basic model of APC effects
Notes: This figure displays Duncan's graphical model in which there is a mechanism between cohort and the outcome. Unobserved idiosyncratic causes of the mechanism and outcome are omitted for simplicity. The double-headed, double-lined arrows denote the inherently deterministic relationship among the three temporal variables. Filled circles indicate observed variables.

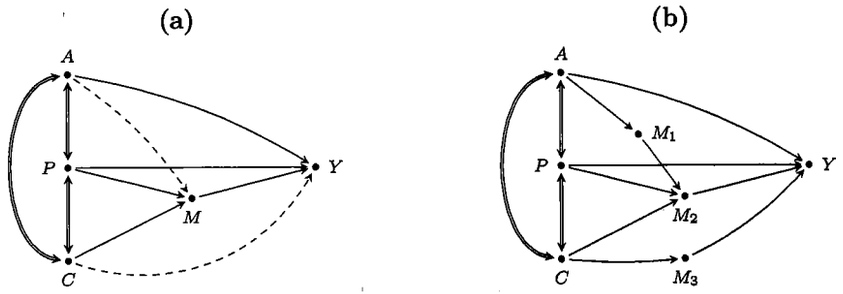

FIGURE 6.13 Omitted pathways in mechanism-based models
Notes: Panel (a) shows an extension of Duncan's APC model with omitted pathways $\mathbf{A} \rightarrow \mathbf{M}$ and $\mathbf{C} \rightarrow \mathbf{Y}$. Panel (b) shows the same model with the pathway $\mathbf{P} \rightarrow \mathbf{M}$ specified and mechanisms added for the other omitted pathways.

only mechanism associated with *C*. The model also assumes that neither *A* nor *P* are associated with *M*. These assumptions are visualized as dashed lines in Figure 6.13(a).

It is worth emphasizing that a richer, more detailed set of data can deal with expected violations of these assumptions. Figure 6.13(b) shows the same graphical model except *M* is now labeled M_2 and we have included two additional mechanisms, M_1 and M_3. As indicated in Figure 6.13(b), one can specify the pathway between *P* and *M* directly and then include an additional mechanism, M_1, between *A* and M_2. Likewise, if it is thought there are additional pathways not modeled by M_2 between *C* and *Y*, one can include them in the model. Here we weaken the assumption that the relationship between *C* and *Y* is entirely mediated by M_2 by including an additional mechanism, M_3. Even when these mechanisms are not available in a given dataset, theorizing about likely omitted pathways is helpful for orienting what new variables should be measured in future data collection efforts (Winship and Harding 2008).

Two points are worth emphasizing about the mechanism-based approach. First, the Duncan model is relatively simple in that it includes just one mechanism. However, as discussed by Fosse and Winship (2019) as well as Winship and Harding (2008), the mechanism-based approach is compatible with a very diverse set of models with multiple mechanisms. Second, so far we have said little about linear and nonlinear effects in a mechanism-based analysis. The nonlinear effects can be estimated directly, because they are identified. That is, one can specify the full set of nonlinear effects for the outcome as well as all mechanisms. In practice, we recommend controlling for all of the nonlinear effects when estimating the linear effects. Under the CTE assumption mentioned previously, the idea is that one can be relatively confident that the linear effect is zero if the corresponding nonlinear effects are zero after adjusting for one or more mechanisms. In our example, this would mean that, if the nonlinear effects of *C* on *Y* on *A* and *M* are zero, then we can be confident under the CTE that the linear effects of *C* on *Y* as well as *A* on *M* are zero. In those cases in which the CTE assumption is not met, then we recommend specifying bounds on the linear effects using the procedures outlined in the previous steps.

Mechanism-based analysis of religious intensity

To illustrate the mechanism-based approach, we use years of education (ranging from 0 to 20) as a mechanism between the age linear component and the outcome, religious intensity. This reflects the argument that attaining higher levels of education will cause, in general, individuals to turn away from organized religion. We also posit that education is associated with the cohort linear component. Because they are identified, we assume that the nonlinear components for all three variables are associated with years of education as well as religious intensity. The corresponding graphical model is visualized in Figure 6.14. Note that this graphical model represents our assumptions about the linear components only. Because the mechanism and outcome are both continuous, we can use the product rule to obtain the estimated APC effects (Winship and Mare 1983). The mechanism-based models produce point estimates of $\alpha = 0.010, \pi = 0.144$ and $\gamma = -0.620$.

We could stop our analysis here. However, it is likely that several mechanisms are missing from our models (Chaves 1989; Firebaugh and Harley 1991). For example, we do not include variables that capture one's political identity, life course transitions or subjective health, all of which are likely associated with age and might cause religious disaffiliation. Under the CTE we would expect the nonlinear effects of age on the outcome to be zero after adjusting for education. However, statistical significance tests indicate that the CTE is not satisfied: specifically, after adjusting for education, there are still statistically significant nonlinear cohort effects on the

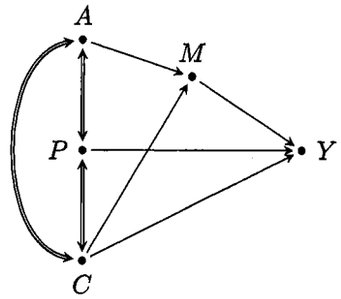

FIGURE 6.14 Mechanism-based model of religious intensity
Notes: This figure displays the graphical model used in the analysis of religious intensity. Unobserved idiosyncratic causes of the mechanism and outcome are omitted for simplicity. The double-headed, double-lined arrows denote the inherently deterministic relationship among the three temporal variables. Filled circles indicate observed variables. It should be noted that this graphical model applies to the linear components. All nonlinear components are included in the corresponding regression models.

outcome. Accordingly, we proceed to conduct a bounding (or sensitivity) analysis. To do so, we assume that the estimate of the mechanism-based age linear effect is an underestimate. Specifically, we assume that the influence of a set of unobserved mechanisms is no greater in absolute value than 20 times the size of the estimated age linear effect operating through education (denoted as the pathway from A to Y via M in Figure 6.14). This assumption is equivalent to claiming that the age linear effect is no less than -0.1857 and no greater than 0.2053. The bounds from this assumption are shown in Figure 6.15. These results reveal that cohort effects dominate in explaining changes in religious disaffiliation, with relatively minor effects of age and period.

Conclusion

Over a century ago, the sociologist Karl Mannheim attempted a reformulation of the problem of generations. In doing so, he outlined a distinctly sociological approach to the problem, rather than one based on what he termed 'positivist' or 'humanist' formulations, neither of which he thought were entirely satisfactory. In this chapter we have attempted a similar reformulation of the problem in its contemporary form, namely, APC analysis. Rather than using a just-identifying constraint to achieve a particular set of APC estimates or advocating for a particular statistical model, our goal has been to outline a general three-stage procedure for conducting APC analysis focused on examining what can be learned from the data without assumptions, using theory to specify bounds and, when variables are available, modeling mechanisms. Our approach is inherently sociological in that it requires thinking about theoretical assumptions at every stage of the analysis.

To summarize our approach, we recommend the following three stage process when conducting an APC analysis, each of which involves a set of smaller steps:

- **Stage 1** (Analysis without assumptions): Learn as much as possible from the data without making assumptions. Determine whether all three APC variables are operative.
  1. *Linearized APC model*: Separate the linear from the nonlinear components using the L-APC model. Using the L-APC, fit a model with the period linear effect fixed to zero.
  2. *Nonlinear effects*: Report the full set of nonlinear effects (e.g., α, π and γ). Because the nonlinear effects are point-identified, they can be visualized using traditional graphical techniques. Conventional significance tests and fit statistics can be applied. If it seems reasonable to conclude that the nonlinear effects for one of the three APC are zero and the CTE assumption seems reasonable, then that variable may be dropped from the model and analysis can proceed using conventional methods.
  3. *Linear effects*: Report the identifiable combinations of linear effects $\theta_1 = \alpha + \pi$ and $\theta_2 = \gamma + \pi$ that determine the location of the canonical solution line in the parameter space. Visualize the canonical solution line using a 2D-APC

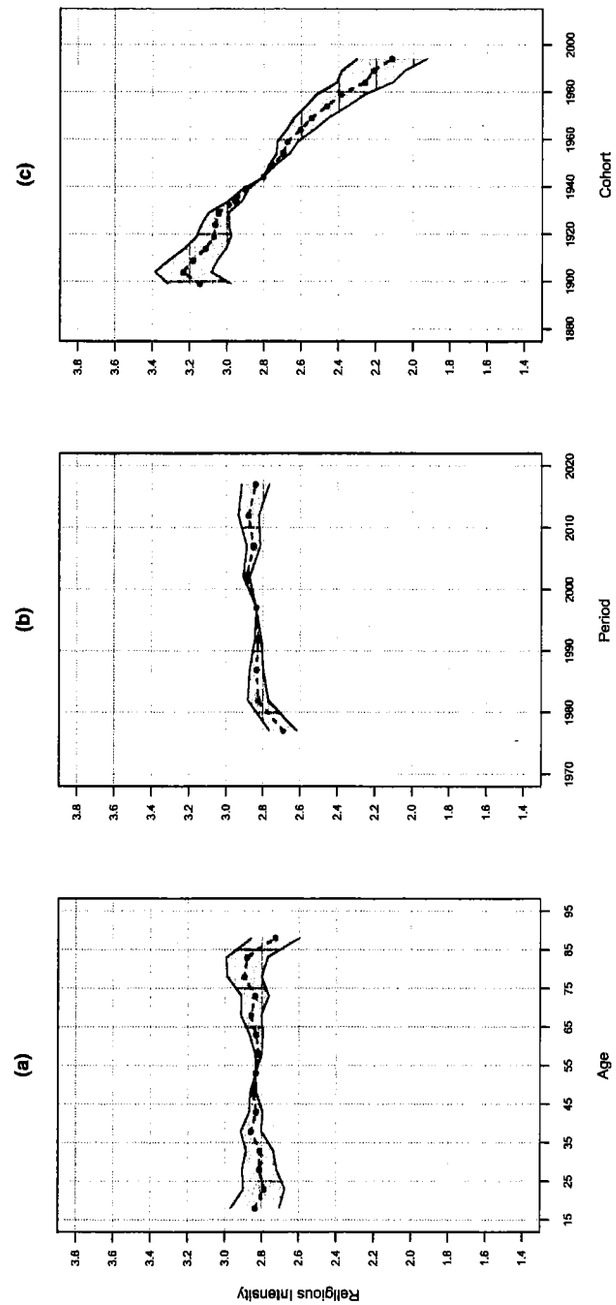

FIGURE 6.15 Religious intensity: upper and lower bounds on APC effects

graph. Consider what combinations of linear APC effects are ruled out by the 2D-APC graph and the location of the canonical solution line.

- **Stage 2** (Partial identification using bounds): Specify a series of bounds using explicit theoretical assumptions about the size, sign and/or shape of the temporal effects. In specifying bounds, one should ask whether there is value in making stronger assumptions to obtain more precise estimates. If not, one is done and should proceed to the next step.
  1. *Constrained 2D-APC graph*: Using the 2D-APC graph, analyze the effects of one's theoretical assumptions on the constraints imposed on the solution line. Are a broad set of parameter effects still plausible? If so, are there additional assumptions that can be defended that might be used to narrow the bounds? Note that it is possible for a set of assumptions to be inconsistent, leading to the whole solution line being excluded.
  2. *Bounds on overall effects*: Plot the bounded total effects for the three APC variables. Are the bounds sufficiently narrow to draw meaningful substantive conclusions? Are additional assumptions needed for this to be the case?
- **Stage 3** (Mechanism-based models of APC effects):
  1. *Mechanism-based models*: Using the observed measures of causes, fit one or more mechanism-based models to obtain point estimates of the temporal effects. Test whether any of the remaining nonlinear effects for the three APC variables are zero.
  2. *Sensitivity analyses*: After fitting mechanism-based models, one should consider conducting a sensitivity analysis with a 2D-APC graph to assess the robustness of findings in the presence of unobserved causal pathways.

The goal of our approach is to clarify what can be learned from the data itself and what can be concluded from data with the incorporation of various theoretical assumptions about the temporal effects. Mechanism-based models are promising in that they can be tested against the data and in many cases are overidentified. We view our approach as an attempt to redefine the problem of APC analysis away from a single identifying constraint to learning as much as possible using the weakest assumptions possible.

The approach outlined here is general, flexible and renders transparent many of the otherwise hidden assumptions of conventional APC models. As is generally the case with new methodologies, there is the question as to whether past substantive findings will be sustained or overturned. This is a critical task for future APC analysis given that many of the methods used have often rested on untested or untestable assumptions. We conjecture that the results will be mixed: when analyses have been driven by strong theory and/or conclusions are primarily based on nonlinear effects, new methods are likely to sustain old findings; when theory is lacking and/or conclusions are based on linear effects then it will be little more than luck if previous findings are not overturned. The important point, however, is that with the ongoing development of new methods, a broad set of new empirical analyses are needed.

Online supplementary material relating to this book can be found at www.routledge.com/9780367174439.

Notes

1 A Shiny program that allows replicating this work and analyzing new data can be found at https://github.com/adeldaoud/WhatIfAPC

2 A linear effect refers to a straight-line relationship, while a nonlinear effect refers to any deviation around this straight-line relationship. For example, a nonlinear effect might appear as an upside-down 'U' shape. The combination of linear and nonlinear effects we refer to as the total effect.

3 We have yet to identify a compelling case where CTE is unlikely to hold, that is, a case where there are linear effects, but no nonlinear effects. The converse situation – in which there is a nonlinear effect, but no linear effect, or what might be called 'trendless fluctuation' – does seem possible.

4 As noted above, numerous solutions to the APC identification problem have been offered over the years. These are not a concern of the present chapter. Detailed discussions can be found in Fosse and Winship (2018; 2019) as well as O'Brien (2015).

5 This conceptualization is close to that of Sen and Wasow's (2016) idea of race (and also gender) as a bundle of underlying causal processes.

6 Critical, although it has been under appreciated in the literature, the APC identification problem is restricted to the linear effects. As Fienberg (2013) has stated: 'The APC problem is a linear effects problem (1982).' We provide intuition for this point below. For a formal proof, see Fosse and Winship (2018).

7 For detailed reviews on the variety of temporal data structures and their associated problems, see the chapters by Yang and Land (2013: 15–53) as well as Mason and Fienberg (1985: 59–67).

8 Additional complications arise when the age and period intervals are not equally spaced, since this can generate artifactual cyclical patterns. For approaches to estimating temporal effects when age and period intervals are unequal, see Holford (2006).

9 Note, however, that calculating the cohorts from an age–period array introduces some ambiguity, since adjacent cohorts will overlap partially (Holford 1983: 311–312). This ambiguity is present even when the age and period groups are of equal width. For example, in Table 6.1 the 1956–1964 cohort overlaps partially with the cohort born during 1961–1969. Narrower age and period intervals will produce a finer grid of cells in the age–period array, but neighboring cohorts will nonetheless overlap to some extent. For instance, if age and period are tabulated into two-year intervals (Tarone and Chu 1996), then those in age 31–32 and period 2000–2001 will be in cohort 1968–1970, while those in age 30–31 and period 2000–2001 will be in the adjoining cohort 1969–1971. The inexactness can be mitigated when all three temporal variables are measured at the individual level, since this will allow for aggregation into nonoverlapping groups, but few datasets include all three variables.

10 Note that 1 is added to $j-i$ so that the cohort index begins at $k = 1$. This ensures that, for example, $i = j = k = 1$ refers to the first group for all three temporal measures. One could just as easily index the cohorts using $k = j - i$, but this identity would be lost.

11 Alternatively, one could fix the parameters at one of the levels to zero. By convention researchers typically fix the first set of levels (e.g., $\alpha_{i=1} = \pi_{j=1} = \gamma_{k=1} = 0$) or the last set (e.g., $\alpha_{i=I} = \pi_{j=J} = \gamma_{k=K} = 0$), although another set could be chosen to satisfy the constraints.

12 This is reflected in the null vector of the L-APC, which consists of [1, –1, 1] for the age, period and cohort slopes, respectively, and a set of zeros for the nonlinearities. In fact, the basic linear model of Equation 6.2 is just a special case of the L-APC model. Referring to our example, note that we can replace the intervals in Table 6.1 with single values, such as the midpoint of each category, to obtain the array shown in Table 6.2. For

example, we can replace the 30–34 age interval with the midpoint 32, the 2000–2004 period interval with the midpoint 2002, and the 1966–1974 cohort interval with the midpoint 1970.

13 A simple linear transformation can be used to convert age_i to $i - i*$, since $i - i* = (age_i - age*)/(\Delta age)$, where age* is the midpoint for all age groups and Δage is the fixed difference between the midpoints. For example, referring to Table 6.2 we have $age_1 = 32$, $age_2 = 37$, $age_3 = 42$, $age_4 = 47$ and $age_5 = 52$. The midpoint across all age groups is 42 and the fixed difference for the groups is 5. Thus, for example, we can calculate that for $age_1 = 32$, the transformed $age_1 = (32 - 42)/5 = -2$ which is equivalent to $i - i* = 1 - 3 = -2$.

14 For simplicity of exposition we treat this as a continuous variable and thus our estimates are based on classical linear regression models. We obtain similar results using ordinal or multinomial logistic regression models.

15 We fix the period linear effect to zero when estimating the predicted values of the outcome.

16 We note that there are, in fact, assumptions involved regarding sum-to-zero constraints and the additive nature of the classical APC model. However, these assumptions are relatively minor.

17 For visualization purposes we have dropped the first two cohort groups and last cohort group from the graphs. These groups are sparse, each consisting of 120 or fewer respondents.

18 Note that there is still an identification problem in the sense that one must assume, based on the principle of CTE, that the linear effect of the dropped variable is not operative. This assumption about the linear effects cannot be tested due to the identification problem.

19 Note that the cohort categories are unbalanced in two senses: first, there are fewer individual-level observations in the tails of the cohort categories; second, there are fewer age-period categories in the corners of the Lexis table. Both of these can result in greater imprecision of the cohort categories despite relatively large effect sizes of the nonlinearities.

20 The same dataset of APC variables can have any number of solution lines depending on how the variables are coded (Fosse and Winship 2018). However, in general these various solution lines can be expressed in their most basic form in terms of three dimensions, or what we call the canonical solution line.

21 Note that this discussion provides another way of thinking about the signed regions described in the previous section.

22 The question is as follows: "Which statement do you agree more with?" The answer consisted of two choices. The first choice is: "The younger generation should be taught by their elders to do what is right." The second choice is: "The younger generation should be taught to think for themselves even though they may do something their elders disapprove of."

References

Alwin, Duane F. (Oct. 1991). "Family of Origin and Cohort Differences in Verbal Ability". *American Sociological Review* 56 (5), pp. 625–638.

Chaves, Mark (1989). "Secularization and Religious Revival: Evidence from US Church Attendance Rates, 1972–1986". *Journal for the Scientific Study of Religion* 28 (4), pp. 464–477.

Chen, Xinguang et al. (2003). "Secular Trends in Adolescents Never Smoking from 1990 to 1999 in California: An Age-period-cohort Analysis". *American Journal of Public Health* 93 (12), pp. 2099–2104.

Clark, April K. and Marie A. Eisenstein (2013). "Interpersonal Trust: An Age–Period–Cohort Analysis Revisited". *Social Science Research* 42 (2), pp. 361–375.

Clayton, D. and E. Schifflers (1987). "Models for Temporal Variation in Cancer Rates. II: Age–period–cohort models". *Statistics in Medicine* 6 (4), pp. 469–481.

Clogg, Clifford C. (1982). "Cohort Analysis of Recent Trends in Labor Force Participation". *Demography* 19 (4), p. 459.

Diouf, Ibrahima et al. (2010). "Evolution of Obesity Prevalence in France: An Age-Period-Cohort Analysis". *Epidemiology* 21 (3): 360.

Duncan, Otis Dudley (1985). "Generations, Cohorts, and Conformity". In Mason, William M and Gienberg, Stephen E (eds), *Cohort Analysis in Social Research*. New York: Springer: 289–321.

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

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

Firebaugh, Glenn and Brian Harley (1991). "Trends in U.S. Church Attendance: Secularization and Revival, or Merely Lifecycle Effects?" *Journal for the Scientific Study of Religion* 30 (4), pp. 487–500.

Fosse, Ethan and Christopher Winship (2019). "Analyzing Age-Period-Cohort Data: A Review and Critique". *Annual Review of Sociology* (45), pp. 467–492.

Frost, Wade Hampton (1995). "The Age Selection of Mortality from Tuberculosis in Successive Decades". *American Journal of Epidemiology* 141 (1), pp. 4–9.

Fu, Qiang and Kenneth C. Land (2015). "The Increasing Prevalence of Overweight and Obesity of Children and Youth in China, 1989–2009: An Age–Period–Cohort Analysis". *Population Research and Policy Review* 34 (6), pp. 901–921.

Goldberger, Arthur Stanley (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Hauser, Robert M. and Min-Hsiung Huang (1997). "Verbal Ability and Socioeconomic Success: A Trend Analysis". *Social Science Research* 26 (3), pp. 331–376.

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

——— (2006). "Approaches to Fitting Age-Period-Cohort Models with Unequal Intervals". *Statistics in Medicine* 25 (6), pp. 977–993.

Hout, Michael and David Knoke (1975). "Change in Voting Turnout, 1952–1972". *Public Opinion Quarterly* 39 (1), pp. 52–68.

Keiding, Niels (2011). "Age-Period-Cohort Analysis in the 1870s: Diagrams, Stereograms, and the Basic Differential Equation". *Canadian Journal of Statistics*, 39 (3), pp. 405–420.

Kerr, William C. et al. (2004). "Age, Period and Cohort Influences on Beer, Wine and Spirits Consumption Trends in the US National Alcohol Surveys". *Addiction* 99 (9), pp. 1111–1120.

Lavori, Philip W. et al. (1987). "Age-Period-Cohort Analysis of Secular Trends in Onset of Major Depression: Findings in Siblings of Patients with Major Affective Disorder". *Journal of Psychiatric Research* 21 (1), pp. 23–35.

Liu, S. et al. (2001). "Increasing Thyroid Cancer Incidence in Canada, 1970–1996: Time Trends and Age-Period-Cohort Effects". *British Journal of Cancer* 85 (9), p. 1335.

Mason, Karen Oppenheim, William M. Mason, et al. (1973). "Some Methodological Issues in Cohort Analysis of Archival Data". *American Sociological Review* 38 (2), p. 242.

Mason, William M. and Stephen E. Fienberg, eds. (1985). *Cohort Analysis in Social Research*. New York, NY: Springer New York.
O'Brien, Robert (2015). *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data*. Boca Raton, FL: CRC Press.
O'Malley, Patrick M., Jerald G. Bachman, and Lloyd D. Johnston (1984). "Period, Age, and Cohort Effects on Substance Use among American Youth, 1976–82." *American Journal of Public Health* 74 (7), pp. 682–688.
Pilcher, Jane (1994). "Mannheim's Sociology of Generations: An Undervalued Legacy". *The British Journal of Sociology* 45 (3), pp. 481–495.
Putnam, Robert D. (1995). "Tuning in, Tuning out: The Strange Disappearance of Social Capital in America". *PS: Political Science & Politics* 28 (4), pp. 664–683.
Robinson, Robert V. and Elton F. Jackson (2001). "Is Trust in Others Declining in America? An Age–Period–Cohort Analysis". *Social Science Research* 30 (1), pp. 117–145.
Rodgers, Willard L. (1982). "Estimable Functions of Age, Period, and Cohort Effects". *American Sociological Review* 47 (6), pp. 774–787.
——— (1990). "Interpreting the Components of Time Trends". *Sociological Methodology* 20, p. 421.
Ryder, Norman B. (1965). "The Cohort as a Concept in the Study of Social Change". *American Sociological Review* 30 (6), pp. 843–861.
Schwadel, Philip and Michael Stout (2012). "Age, Period and Cohort Effects on Social Capital". *Social Forces* 91 (1), pp. 233–252.
Sen, Maya and Omar Wasow (2016). "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics". *Annual Review of Political Science* 19, pp. 499–522.
Tarone, R. E. and K. C. Chu (1996). "Evaluation of Birth Cohort Patterns in Population Disease Rates". *American Journal of Epidemiology* 143 (1), pp. 85–91.
Tilley, James and Geoffrey Evans (2014). "Ageing and Generational Effects on Vote Choice: Combining Cross-sectional and Panel Data to Estimate APC Effects". *Electoral Studies* 33, pp. 19–27.
Vedøy, Tord F. (2014). "Tracing the Cigarette Epidemic: An Age-Period-Cohort Study of Education, Gender and Smoking Using a Pseudo-panel Approach". *Social Science Research* 48, pp. 35–47.
Wilson, James A. and Walter R. Gove (1999). "The Age-Period-Cohort Conundrum and Verbal Ability: Empirical Relationships and Their Interpretation: Reply to Glenn and to Alwin and McCammon". *American Sociological Review* 64 (2), pp. 287–302.
Winship, Christopher and David J. Harding (2008). "A Mechanism-Based Approach to the Identification of Age–Period–Cohort Models". *Sociological Methods and Research* 36 (3), pp. 362–401.
Winship, Christopher and Robert D. Mare (1983). "Structural Equations and Path Analysis for Discrete Data". *The American Journal of Sociology* 89 (1), pp. 54–110.
Yang, Yang (2008). "Social Inequalities in Happiness in the United States, 1972 to 2004: An Age-Period-Cohort Analysis". *American Sociological Review* 73 (2), pp. 204–226.
Yang, Yang and Kenneth C. Land (2013). *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. Boca Raton, FL: Chapman and Hall/CRC.

7

MODELING FACTORS AFFECTING AGE, PERIOD AND COHORT TRENDS

The effect of cigarette smoking on lung cancer trends

Theodore R. Holford

Introduction

Temporal disease trends are driven by causal factors that are themselves changing over time. A time scale is measured in reference to a starting point, so the choice of scale should be based on the mechanism for the causal agent affecting outcome. Age, for example, measures time from birth so that changes in disease risk that result from the aging process can be captured to some extent by introducing age into a model used for analysis. If there are also exposure changes that affect all ages in much the same way, then we might expect these to be manifested as period effects on trend. Likewise, if the changes only affect individuals of an age group and not the entire population, then these effects can be represented as birth cohort effects.

Age–period–cohort models are useful at the early stages of trying to understand the causes of an outcome. If the trend is primarily due to a period effect, then we would primarily consider factors that affect all ages at the same time. Likewise, if trend is more closely related to birth cohort then better candidates for causal agents would be generational exposures that are often set early in life. These scenarios occur early in the research process when little is known about causes of the trends observed for an outcome. However, it is also useful to consider these trends after knowledge has advanced so that one can quantify the extent to which an identified factor can account for the observed trends. In this chapter I describe how age–period–cohort models can be extended to include the effect of exposure on the outcome that has been determined from analytical studies and assess the extent to which this knowledge accounts for the observed trends. This can not only suggest further areas of study to understand better what is not known, but it can also provide a framework that can be used to quantify programs intended to control the outcome.