Bounding Analyses

of Age-Period-Cohort Effects

Abstract

For over a century researchers from a wide range of disciplines have sought to estimate the unique contributions of age, period, cohort (APC) effects on a variety of outcomes. A key obstacle to these efforts is the perfect linear dependence among the three time scales. Various methods have been proposed to deal with this issue, but they have suffered, arguably universally, from either ad hoc assumptions or extreme sensitivity to small differences in model specification. After briefly reviewing past work, we outline a new approach for identifying temporal trends in population-level data. Fundamental to our framework is the recognition that it is only the linear trends of an APC model that are unidentified, not the nonlinearities or particular combinations of the linear components. One can thus use constraints implied by the data along with explicit theoretical claims to bound one or more of the APC effects. Bounds on these parameters may be nearly as informative as point estimates, even with relatively weak assumptions. To demonstrate the usefulness of our approach we examine temporal trends in prostate cancer incidence and homicide rates. We conclude with a discussion of guidelines for further research on APC effects.

1 Introduction

Researchers in a wide range of fields have long sought to understand social and cultural change by identifying the separate contributions of age, period, and cohort (APC) effects[1] on various outcomes (Glenn 1976; O'Brien 2015; Ryder 1965). The core idea is that any temporal change can be attributed to three kinds of processes (Glenn 2005: 11; Yang and Land 2013b: 1-2): (1) changes over the life course of individuals, or *age effects*; (2) changes due to the events in particular years, or *period effects*; (3) changes arising from the replacement of older cohorts of individuals with younger ones with different characteristics, or *cohort effects*. However, in what has been called the *APC identification problem* (Mason and Fienberg 1985: 68), the slopes[2] of an APC model cannot be uniquely estimated due to the perfect linear dependency among the age, period, and cohort variables. A variety of approaches have been attempted to overcome the APC identification problem, with recent attempts based on constraining the temporal trends through hierarchical (or multilevel) models or variations of the Moore-Penrose generalized inverse.

In this paper we propose an alternative approach based on placing bounds on one or more of the temporal trends from an APC model. Our approach has important similarities to the work by Charles Manski on partial identification (1990; 1993; 2003). In contrast to problems of statistical inference, which involve understanding how sampling variability can affect conclusions based on samples of limited size, problems of identification entail understanding what conclusions can be drawn even with a sample of infinite size. The lack of a unique solution for the linear APC trends is a classic identification problem since it cannot be resolved by gathering larger samples (Manski 2003: 12). Our approach begins by examining what the data alone, with as few constraints as possible, can tell us. Then, since full identification of the parameters from an APC model is not possible, we proceed with the goal of partially identifying the slopes using explicit theoretical considerations that are based on the expected size, direction, or overall shape of one or more of the

[1] We follow the convention in the APC literature and use the shorthand of "effects" when referring to age, period, and cohort processes (e.g., Glenn 1981: 249; Mason, Mason, et al. 1973: 243; Fienberg and Mason 1979: 133; O'Brien 2015: 1; Yang and Land 2013b: 1-2). These "effects" need not refer to causal effects in the sense of parameters associated with well-defined potential outcomes (Morgan and Winship 2014: 37-76).

[2] Researchers have used various terms in the literature to refer to the linear and nonlinear components of an APC model. In this paper we refer to the linear components as "slopes," "linear trends," or simply "linear components"; conversely, we refer to the nonlinear components as "nonlinearities", "deviations", or just "nonlinear components." By "trends" or "overall trends" we refer to the combination of the linear and nonlinear components.

temporal trends.

The rest of this paper is organized as follows. First, we offer a brief history of APC analysis, providing a conceptual overview of the identification problem and summarizing recent methods to address the problem in sociology, epidemiology, and demography. Second, we discuss in detail the formal model of APC analysis, showing how the identification problem is restricted to the linear components. Third, we demonstrate how, despite the nonidentifiability of the individual slopes, the data provide important information that allow us to derive formal bounds on the linear components of APC models. Fourth, based on a geometric derivation of the identification problem, we outline a novel graphical tool that summarizes what can be known about the linear trends from an APC model. Fifth, we discuss various strategies for bounding APC effects by specifying the size and direction of one or more slopes, fixing only the sign of one or more slopes, or applying a shape constraint using data on the nonlinear components. Sixth, we provide two examples of our approach, examining temporal trends in prostate cancer incidence and homicide rates. Finally, we conclude with a summary of our framework and an outline for further research on APC trends.

2 Background

The APC effects has interdisciplinary origins, with seminal works arising in epidemiology and medicine in the 1930s (e.g., Frost 1939) and later in sociology and demography in the 1950s and 1960s (Mannheim 1952; Ryder 1965).[3] In the 1930s the epidemiologist W.H. Frost sought to identify the cause of a tuberculosis outbreak that had ravaged Massachusetts in the 19th century. In a classic work, Frost (1939) noted that changes in observed tuberculosis mortality rates could be attributed not only to aging effects but also period and cohort effects (O'Brien 2015: 11:13; see also Mason and Smith 1985).[4] Similarly, in a posthumous theoretical essay, the sociologist and philosopher Karl Mannheim (1952)[5] emphasized the importance of generations in understanding social change (see also Demartini 1985; Pilcher 1994; Simirenko 1966). Among Mannheim's insights

[3]However, APC analysis arguably dates back to at least the 1860s, predating the eponymous diagrams of Wilhelm Lexis (see Keiding 2011)

[4]See Case (1956) for an overview of other early studies in epidemiology and medicine on cohort effects and mortality rates (e.g., Andvord 1930; Derrick 1927; Kermack et al. 1934).

[5]Mannheim's essay was first published in German in 1927, but it was not published in English until 1952, five years after his death in 1947.

was that major shifts in the population can occur not only because individuals change their beliefs and values, but because each cohort is imprinted with different values and beliefs that its members carry with them until their death (1952: 292-302). Likewise, in an influential article linking population studies with cultural analysis, the demographer Norman Ryder (1965) observed that, through the biological processes of birth and death, different cohorts are continually entering and leaving society, potentially transforming society. By recognizing that social change can occur even though no particular individual has changed, the arguments by Ryder appealed to a broad swath of quantitatively-oriented researchers seeking to understand the sources of social change (for example, see Firebaugh and Harley 1991; Firebaugh 1989; Hobcraft et al. 1982; Inglehart 1971).

2.1 The APC Identification Problem

The insights of Frost, Mannheim, and Ryder spawned generations of researchers to estimate cohort trends for a wide range of social, biological, and cultural outcomes.[6] Unfortunately, however, because of the exact linear relationship between age, period, and cohort, one cannot use conventional statistical methods to estimate the separate contributions of age, period, and cohort on a particular outcome (Mason, Mason, et al. 1973). The APC identification problem is simply the fact that if we know the age of a person in years and the year in which their outcome was measured, then we know their birth year. That is, we have the exact linear dependency:

$$\text{cohort} = \text{period} - \text{age} \tag{1}$$

To illustrate how the perfect linear identity in Eq. 1 results in an identification problem, 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. A direct way to understand this problem is to use age, period, and cohort variables as inputs in a multiple linear regression model:

[6] For example, researchers have examined verbal ability (Alwin 1991; Hauser and Huang 1997; Wilson and Gove 1999; Yang and Land 2006), social trust (Schwadel and Stout 2012; Clark and Eisenstein 2013; Putnam 1995; Robinson and Jackson 2001), party identification (Bartels and Jackman 2014; Ghitza and Gelman 2014; Hout and Knoke 1975; Tilley and Evans 2014), religious affiliation (Chaves 1989; Firebaugh and Harley 1991), drug use (Chen et al. 2003; Kerr et al. 2004; O'Malley et al. 1984; Vedøy 2014), obesity (Diouf et al. 2010; Fu and Land 2015; Reither et al. 2009), cancer (Clayton and Schifflers 1987; Liu et al. 2001), and mental health (Lavori et al. 1987; Lewinsohn et al. 1993; Yang 2008).

$$Y = \mu + \alpha(\text{age}) + \pi(\text{period}) + \gamma(\text{cohort}) + \epsilon \tag{2}$$

where Y is the outcome variable to be explained; μ is the intercept; age, period, and cohort are age, period, and cohort 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 data set.

In Eq. 2 we are attempting to estimate the effect of each variable holding the other variables fixed at some particular value. What provides information for estimating the effect of the variable of interest is the extent to which Y varies with that variable holding constant the other variables. Without loss of generality assume that we want to estimate the slope of cohort (γ in Eq. 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.

However, the APC identification problem is limited to the linear trends only, a point under-appreciated in the literature.[7] We provide some intuition here. Let $f(.)$, $g(.)$, and $h(.)$ denote functions relating age, period, and cohort, respectively. There is no general equation[8] of the form $g(\text{period}) = f(\text{age}) + h(\text{cohort})$ except in the case where these functions are the identity functions.[9] For example, $\text{age}^2 + \text{cohort}^2 \neq \text{period}^2$. What this means in practice is that the identification problem is confined to the linear trends of the model; the nonlinear components are thus unproblematic in terms of identifiability.[10] This point becomes important when constructing bounds, since we can use the nonlinearities to dramatically narrow the range of the overall APC

[7] As Fienberg (2013) has stated: "The APC problem is a linear effects problem (1982)."

[8] This claim no longer holds if we expand the APC model to allow for cross-product terms. Obviously $\text{period}^2 = (\text{age} + \text{cohort})^2 = \text{age}^2 + 2(\text{age} \times \text{cohort}) + \text{cohort}^2$. The question is whether we think that $(\text{age} + \text{cohort})^2$ represents a different set of substantive processes than period^2. An additional issue is that allowing for cross-product terms results in a far greater degree of underidentification. For a technical discussion of interaction terms in APC models, see Mason and Fienberg (1985: 71-80).

[9] That is, we have the equation $\text{period} = \text{age} + \text{cohort}$, such that $f(\text{age}) = \text{age}$, $g(\text{period}) = \text{period}$, and $h(\text{cohort}) = \text{cohort}$.

[10] The nonlinearities may be require some smoothing, especially in the extreme levels of age and cohort, but this is a relatively tractable problem.

trends.

2.2 Approaches to Identifiability

Since the seminal works by Frost, Mannheim, and Ryder many approaches to the APC identification problem have proposed by sociologists, epidemiologists, demographers, and other social scientists (O'Brien 2015; Yang, Fu, et al. 2004; Yang, Schulhofer-Wohl, et al. 2008; Firebaugh 1989; Hobcraft et al. 1982; Mason and Fienberg 1985; Fu 2008; Fu, Land, and Yang 2011; Schaie 1986; Lewinsohn et al. 1993; Costa and McCrae 1982). These proposed solutions can be divided into two generations (see also Glenn 2005; O'Brien 2015; Yang and Land 2013b).

2.2.1 1st Generation

The first generation of solutions to the identification problem, developed in the 1970s and 1980s (Mason and Fienberg 1985), consist of omitting one of the linear components of an APC model, using a "proxy" variable in place of age, period, or cohort, or constraining some set of the parameters to be equal (Yang and Land 2013b: 63-67). Among the most widely-used approach from this first generation is the equality constraints model, in which the researcher specifies each age, period, and cohort group as a dummy variable in a regression model, but collapses two groups into a single group (Mason, Mason, et al. 1973; Mason and Fienberg 1985; Fienberg and Mason 1979). For example, rather than modeling each birth cohort as a distinct group, an analyst might assume that the parmeters for cohorts born in 1980-1984 and 1985-1989 are equal. As noted by some researchers (Rodgers 1982a; Rodgers 1982b; Palmore 1978; Glenn 1976), models using these kinds of equality constraints are problematic in that they are very sensitive to minor differences in model specification and impose implicit assumptions about the size and direction of the underlying age, period, cohort slopes (O'Brien 2015: 41-43).[11] Moreover, despite the caution by Fienberg and colleagues that equality constraints should be based on overt theoretical assumptions (Mason and Fienberg 1985; Smith et al. 1982), in practice researchers have used such constraints in arbitrary

[11] As Rodgers (1982) has cautioned: "Although a constraint of the type described by Mason et al. [1973; 1979] seems trivial, in fact it is exquisitely precise and has effects that are multiplied so that even a slight inconsistency between the constraint and reality, or small measurement errors, can have very large effects on estimates (785)." However, see also the reply by Smith et al. (1982) as well as the rejoinder by Rodgers (1982).

and often atheoretical ways.[12] Furthermore, researchers may be misled by failing to recognize that models with differing equality constraints have identical fit statistics (O'Brien 2015: 32; Yang and Land 2013b: 65-66).

2.2.2 2nd Generation

Over the past decade, a second generation of models for disentangling the unique contributions of age, period, and cohort effects has emerged (Fu, Land, and Yang 2011; Yang and Land 2006; Fu 2000). This second generation has led to a wave of prominent studies on various topics, including verbal ability (Yang and Land 2006), infant mortality (Powers 2013), heart disease (Lee and Park 2012), obesity (Reither et al. 2009), and perceived happiness (Yang 2008). To achieve identification, these models in general use either shrinkage (or regularization) of the parameters or some variant of the Moore-Penrose generalized inverse for singular matrices. The two most widely-used techniques are the intrinsic estimator (IE) and the hierarchical age-period-cohort (HAPC), but similar results can be obtained from ridge regression, partial least squares regression, and principal components regression (O'Brien 2011; O'Brien 2015).

These techniques have highlighted the importance of addressing the APC identification problem when modeling temporal trends, but they share many of the same shortcomings as the first generation of approaches, as demonstrated by a spate of recent studies (for example, see Bell and Jones 2015a; Bell and Jones 2015b; Bell and Jones 2014b; Bell and Jones 2014a; Pelzer et al. 2014; Fienberg, Hodges, et al. 2015; Luo et al. 2016). Like the first generation of techniques, these methods require researchers to impose constraints, typically implicit, on the model parameters, relying on potentially strong assumptions that may fail to consistently estimate the true underlying age, period, and cohort parameters.[13] Moreover, recent studies have shown that these new methods can be very sensitive to model parameterization in ways that are not likely to be obvious to applied researchers (Bell and Jones 2015b; Pelzer et al. 2014; Luo et al. 2016; Luo 2013). The result is that a number of important findings in the social and biomedical sciences are now coming under height-

[12] For notable examples of researchers identifying APC effects using explicit, theoretically-motivated equality constraints, see Mason and Smith (1985: 175-178) and O'Brien (2015: 183-190).

[13] For example, with reference to the IE, Yang and Land (2013) note: "[T]he objective of the IE is not to estimate the unidentifiable regression coefficient vector (119)." That is, the IE finds the point on the solution line closest to the origin in terms of Euclidean distance, but it does not necessarily recover the actual data generating parameters of the age, period, and cohort effects.

ened scrutiny. For example, a widely-cited study using the HAPC model concluded that the obesity epidemic is primarily the result of period effects (Reither et al. 2009). However, simulation studies have shown that it is just as plausible that the epidemic is the result of cohort effects, consistent with previous theories on the topic (Bell and Jones 2014b).

In short, there has been some progress since the foundational works by Frost, Mannheim, and Ryder, but problems remain. Either the assumptions used to identify an APC model are ad hoc or the estimates are highly sensitive to the exact model specification. Moreover, many researchers appear to be unaware of the identification problem altogether, dropping one or more of the temporal variables without an explicit proxy (O'Brien 2015: 5). As we show in the next section, this entails an unnecessarily strong assumption about one or more of the APC effects, since the nonlinear components are fully identifiable.

3 Modeling Temporal Effects

3.1 The Classical APC Model

Temporal effects in an age-period array of categorically-coded data can be represented using the *classical APC (C-APC) model*, also referred to as the *multiple classification model* (Mason, Mason, et al. 1973: 243) or *accounting model* (Mason and Fienberg 1985: 46-47, 67). Following the convention in the APC literature, we 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 $k = j - i + I$ and $K = I + J - 1$.[14] With this notation, we can express the C-APC as follows (Mason and Fienberg 1985: 67-68; O'Brien 2015: 24-29; Yang and Land 2013b: 61):

$$Y_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k + \epsilon_{ijk} \tag{3}$$

where Y_{ijk} is the outcome variable to be explained, μ is the intercept, α_i represents the $i\text{th}$ age effect, π_j represents the $j\text{th}$ period effect, γ_k represents the $k\text{th}$ cohort effect, and ϵ_{ijk} is the error term. To avoid overparameterization, we apply the so-called usual constraints that the parameters

[14] For simplicity of exposition we assume that age and period are aggregated into intervals of equal width. 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).

sum to zero, such that $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \pi_j = \sum_{k=1}^{K} \gamma_k = 0$.[15]

The parameterization shown in Eq. 3 is very flexible, allowing the age, period, and cohort effects to be highly nonlinear since there is one parameter for each age, period, and cohort category (Mason, Mason, et al. 1973: 246). In an age-period array, each cell is represented by a unique set of parameters (O'Brien 2011: 1430-1431). However, like the simple model in Eq. 2, the C-APC model suffers from a fundamental identification problem due to perfect linear dependence in the columns of the design matrix (O'Brien 2015: 25-26; Yang and Land 2013b: 63). As a result, the model in Eq. 3 cannot be estimated using conventional statistical techniques.

To clarify the nature of this problem, we can write the C-APC model in matrix notation as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\epsilon} \tag{4}$$

where $\mathbf{y}$ is an $(I \times J) \times 1$ vector; $\mathbf{X}$ is an $(I \times J) \times 2(I + J) - 3$ design matrix; $\mathbf{b}$ is a vector of coefficients with $2(I + J) - 3$ rows; and $\boldsymbol{\epsilon}$ is an $(I \times J) \times 1$ vector representing random error. The least-squares estimate is:

$$\widehat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{5}$$

where the superscripted T denotes the transpose and the superscripted -1 indicates the inverse. Since at least one of the columns of $\mathbf{X}$ can be rewritten as a function of the other columns, the design matrix $\mathbf{X}$ is rank deficient one (i.e., singular). Consequently, a regular inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ does not exist and Eq. 5 cannot be estimated without an additional constraint.[16]

3.2 The General APC Model

An alternative representation of the C-APC orthogonally decomposes the linear from the nonlinear components (see Holford 2006). We can accordingly specify a *general APC (G-APC) model* with the form:

[15] Alternatively, one could fix the parameters at one of the levels to zero. By convention researchers typically fix to zero 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 other sets could be used.

[16] The perfect linear dependence is reflected in the null vector of the design matrix, which has non-zero elements (see Kupper et al. 1985: 829; O'Brien 2015: 56-57).

$$Y_{ijk} = \mu + \alpha(i - i^*) + \pi(j - j^*) + \gamma(k - k^*) + \widetilde{\alpha}_i + \widetilde{\pi}_j + \widetilde{\gamma}_k + \epsilon_{ijk} \tag{6}$$

where the asterisks denote midpoint or referent indices $i^* = \left(\frac{I+1}{2}\right)$, $j^* = \left(\frac{J+1}{2}\right)$, and $k^* = \left(\frac{K+1}{2}\right)$. As before, we refer to the linear trends 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 C-APC and G-APC are equivalent representations of temporal data grouped by age, period, and cohort.[17] As with the C-APC, each cell in an age-period array is modeled by a unique combination of parameters under the usual constraints $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \pi_j = \sum_{k=1}^{K} \gamma_k = 0$. For example, the $i\text{th}$ age effect in the C-APC is represented in the G-APC by the overall linear age trend along with a unique parameter for the $i\text{th}$ age nonlinearity: $\alpha_i = (i - i^*)\alpha + \widetilde{\alpha}_i$. In other words, each age effect α_i is decomposed into the sum of a common parameter α representing the 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 $\widetilde{\alpha}_i$, which is a nonlinearity specific to each row (or age category). We can similarly decompose each of the period and cohort effects into linear and nonlinear components.

The importance of the G-APC model is that, by explicitly separating the slopes from their deviations, it clearly shows that the identification problem is confined to the linear trends. This is reflected in the null vector of the G-APC, which consists of $\{1, -1, 1\}$ for the age, period, and cohort linear trends, respectively, and a set of zeros for the nonlinearities.[18] In fact, the basic linear model of Eq. 2 is just a special case of the G-APC model. When the nonlinear terms are zero in the population, the G-APC is equivalent to the basic linear model:

$$\begin{aligned} Y_{ijk} &= \mu + \alpha(i - i^*) + \pi(j - j^*) + \gamma(k - k^*) + (0) + (0) + (0) + \epsilon_{ijk} \\ &= \mu + \alpha(\text{age}_i) + \pi(\text{period}_j) + \gamma(\text{cohort}_k) + \epsilon_{ijk} \end{aligned} \tag{7}$$

where age_i, period_j, and cohort_k are the midpoint values for each of the categories, which are simply the indices recentered and rescaled, and the nonlinearities are zeroed out so that $\tilde{\alpha}_i = \tilde{\pi}_j = \tilde{\gamma}_k = 0$. [19]

[17] They are equivalent in the sense that as basis vectors they span the same space.

[18] Note that the null vector is unique up to multiplication by a scalar.

[19] A simple linear transformation can be used to convert age_i to $i - i^*$, since $i - i^* = (\text{age}_i - \text{age}^*)/(\Delta\text{age})$, where age^* is the midpoint for all age groups and Δage is the fixed difference between the midpoints. For example, suppose we have $\text{age}_1 = 32$, $\text{age}_2 = 37$, $\text{age}_3 = 42$, $\text{age}_4 = 47$, and $\text{age}_5 = 52$. The midpoint across all age groups is 42 and

4 APC Bounding Formulas

As discussed in the preceding section, it is only the linear components of an APC model that are unidentified. As such, what we need to understand is how the values of the three linear parameters, one each for age, period, and cohort, are related to each other. For ease of exposition, we assume that we have an APC model in which there are only linear components; that is, the parameters for the nonlinear terms are assumed to equal zero in the population (e.g., see Eq. 7). In subsequent sections, we drop this assumption and clarify how our approach applies to the more general case when nonlinearities are present. In this section we first show how the APC slopes can take on any real number between negative and positive infinity, producing the same set of values of the outcome regardless of the constraint. Although this suggests that the data are entirely uninformative when there are no nonlinearities, we then show how the linear trends are systematically related to each other. It is from these interdependencies that we derive a set of formal bounds on the slopes.

4.1 The Nonidentifiability of the Temporal Trends

Since APC models are not fully identified, we cannot obtain point estimates for each of the age, period, and cohort slopes without an additional constraint. A convenient way to express the non-identifiability problem is to note that for any particular APC model we can specify the linear trends as (O'Brien 2015: 70; Rodgers 1982a: 782):

$$\begin{aligned} \alpha^* &= \alpha + \nu \\ \pi^* &= \pi - \nu \\ \gamma^* &= \gamma + \nu \end{aligned} \tag{8}$$

where the asterisk ($*$) indicates an arbitrary set of estimated slopes from an APC model under particular constraints and ν is a scalar fixed to some value. As Eq. 8 indicates, the true unobserved slopes α, π, and γ are simple additive transformations from the estimated slopes α^*, π^*, and γ^*, shifted by a single arbitrary scalar, ν. Eq. 8 also shows that the true slopes are unbounded on the real number line. For example, suppose we fix $\nu = +\infty$. Then it must be the case that

the fixed difference the groups is 5. Thus, for example, we can calculate that $\text{age}_1 = 32$ equals $(32 - 42)/5 = -2$ which is equivalent to $i - i^* = 1 - 3 = -2$.

$$
\begin{aligned}
&\alpha^* < \alpha + \infty \Rightarrow -\infty < \alpha \\
&\pi^* > \pi - \infty \Rightarrow +\infty > \pi \\
&\gamma^* < \gamma + \infty \Rightarrow -\infty < \gamma
\end{aligned} \tag{9}
$$

where we assume that α^*, π^*, and γ^* are finite. Likewise, by fixing $v = -\infty$, we obtain

$$
\begin{aligned}
&\alpha^* > \alpha - \infty \Rightarrow +\infty > \alpha \\
&\pi^* < \pi + \infty \Rightarrow -\infty < \pi \\
&\gamma^* > \gamma - \infty \Rightarrow +\infty > \gamma
\end{aligned} \tag{10}
$$

where again we assume that α^*, π^*, and γ^* are finite. Combining the inequalities from Eqs. 9 and 10, we obtain the following bounds:

$$
\begin{aligned}
&-\infty < \alpha < +\infty \\
&-\infty < \pi < +\infty \\
&-\infty < \gamma < +\infty.
\end{aligned} \tag{11}
$$

The conclusion from Eq. 11 is seemingly quite dire for researchers modeling temporal trends, since the slopes can take on any value between $-\infty$ to $+\infty$. As a consequence, it would appear that the data on age, period, and cohort are totally uninformative with respect to the parameter estimates of interest; we have no way of evaluating which set of estimated slopes α^*, π^*, and γ^* are preferable over another set.

4.2 Identical Values for the Outcome

A related problem is that, regardless the value of the slopes under a given constraint, we will obtain the same values for the outcome (O'Brien 2015: 104). In other words, not only are the slopes entangled but it would seem that the data (i.e., the values of Y) provide no guidance for constructing bounds. To see this, consider the following. Without loss of generality, we can express the APC model in terms of the linear trends as:

$$
Y^* = \mu + (\text{age})(\alpha^*) + (\text{period})(\pi^*) + (\text{cohort})(\gamma^*) + \epsilon \tag{12}
$$

where the subscripts have been dropped for simplicity. We can rewrite this equation by substituting the values from Eq. 8:

$$Y^* = \mu + (\text{age})(\alpha + \nu) + (\text{period})(\pi - \nu) + (\text{cohort})(\gamma + \nu) + \epsilon. \tag{13}$$

Rearranging terms we obtain:

$$Y^* = \mu + (\text{age})(\alpha) + (\text{period})(\pi) + (\text{cohort})(\gamma) + (\text{age} - \text{period} + \text{cohort})(\nu) + \epsilon. \tag{14}$$

However, $(\text{age} - \text{period} + \text{cohort}) = 0$ so we have

$$\begin{aligned} Y^* &= \mu + (\text{age})(\alpha) + (\text{period})(\pi) + (\text{cohort})(\gamma) + \epsilon \\ &= \mu + (\text{age})(\alpha^*) + (\text{period})(\pi^*) + (\text{cohort})(\gamma^*) + \epsilon. \end{aligned} \tag{15}$$

Thus the APC model with the original parameters (α, π, γ) and the shifted parameters (α^*, π^*, γ^*) give us exactly the same values of the outcome Y (as well as predicted values of the outcome). As such, the data (i.e., the outcome Y) is uninformative as to whether the original or shifted parameters should be our estimates.

4.3 Interdependencies Among the Slopes

The fact that the true slopes are indistinguishable from any set of estimated slopes under an arbitrary constraint would seem highly problematic for our goal of constructing meaningful bounds. It would also appear that the outcome Y provides no information on how to restrict the range of the slopes, since regardless of our estimates we obtain the same set of values of Y. However, this depiction is incomplete: although the slopes can take on any value, they are not independent of each other. Certain linear combinations of the slopes are identifiable, so if we set any one slope to a particular value then this constraint determines the values of the remaining two slopes.

As we discuss in detail below, if one is willing to make an informed statement about the qualitative nature of any one of the parameters associated with the linear components of an APC model, then the data may be quite informative about the size and direction of the remaining parameters. The relationships among the slopes is the key to understanding why placing bounds on the parameters associated with the linear components of an APC model can be informative. Although the value of ν can equal any number between minus to plus infinity, the data constrains the values for the age, period, and cohort slopes to be consistent with relations in Eq. 8. That is, the scalar ν creates the new parameters on the left of each equation (denoted by an asterisk) by shifting the linear age parameter α and cohort parameter γ by the same amount in the same direction, and the period parameter π by an equal amount in the opposite direction. These interdependencies result in identifiable linear combinations of the slope parameters, which are based entirely on the data. That is, using Eq. 8, we can prove that although the individual slopes are unidentifiable particular combinations are known since the arbitrary constant ν cancels out:

$$
\begin{aligned}
&(\alpha+\nu)+(\pi-\nu) \Rightarrow \alpha+\pi=\theta_1 \\
&(\pi-\nu)+(\gamma+\nu) \Rightarrow \pi+\gamma=\theta_2 \\
&(\alpha+\nu)-(\gamma+\nu) \Rightarrow \alpha-\gamma=\theta_1-\theta_2 \\
&(\gamma+\nu)-(\alpha+\nu) \Rightarrow \gamma-\alpha=\theta_2-\theta_1 .
\end{aligned} \tag{16}
$$

These combinations of slopes are invariant to any given constraint defined by ν. For example, regardless of the constraint imposed by the researcher the sum of the age and period slopes will always equal the sum of the true age and period slopes: $\alpha^* + \pi^* = \alpha + \pi + (\nu - \nu) = \alpha + \pi$. Moreover, we can further simplify the equations by noting that although there are four different combinations of slopes in Eq. 16, there are actually only two unique quantities generated by the data, which we have labeled θ_1 and θ_2. Once we know θ_1 and θ_2, the right side of all four equations above are determined.

Since θ's in Eq. 16 are identifiable, they can be estimated from the data. To obtain values for these parameters, suppose again that we want to estimate following basic linear model:

$$Y = \mu + (\text{age})\alpha + (\text{period})\pi + (\text{cohort})\gamma + \epsilon \tag{17}$$

where as before we have assumed that the nonlinearities are zero in the population. As discussed previously, since $\text{period} = \text{age} + \text{cohort}$, Eq. 17 is not identified and as such the slope parameters cannot be estimated. However, we can re-express Eq. 17 as (cf. Eq. 13):

$$Y = \mu + (\text{age})(\alpha + \nu) + (\text{period})(\pi - \nu) + (\text{cohort})(\gamma + \nu) + \epsilon. \tag{18}$$

There are three constrained models that allow us to estimate the θ's, which are generated from the combinations of the true, unknown slopes in Eq. 16. Each of these models is based on setting either age, period, or cohort to zero. For example, to estimate the period and cohort slopes under the constraint that the age slope is zero, we substitute $\text{period} - \text{cohort}$ for age in Eq. 18:

$$\begin{aligned} Y &= \mu + (\text{period} - \text{cohort})(\alpha + \nu) + (\text{period})(\pi - \nu) + (\text{cohort})(\gamma + \nu) + \epsilon \\ &= \mu + (\text{period})(\theta_1) + (\text{cohort})(\theta_2 - \theta_1) + \epsilon \end{aligned} \tag{19}$$

where $\theta_1 = \alpha + \pi$ and $\theta_2 - \theta_1 = \gamma - \alpha$. Likewise, we can estimate the age and cohort slopes under the constraint that the period slope is zero by substituting $\text{age} + \text{cohort}$ for period:

$$\begin{aligned} Y &= \mu + (\text{age})(\alpha + \nu) + (\text{age} + \text{cohort})(\pi - \nu) + (\text{cohort})(\gamma + \nu) + \epsilon \\ &= \mu + (\text{age})(\theta_1) + (\text{cohort})(\theta_2) + \epsilon \end{aligned} \tag{20}$$

where $\theta_1 = \alpha + \pi$ and $\theta_2 = \pi + \gamma$. Finally, we can estimate the age and period slopes under the constraint that the cohort slope is zero by substituting $\text{period} - \text{age}$ for cohort:

$$\begin{aligned} Y &= \mu + (\text{age})(\alpha + \nu) + (\text{period})(\pi - \nu) + (\text{period} - \text{age})(\gamma + \nu) + \epsilon \\ &= \mu + (\text{age})(\theta_1 - \theta_2) + (\text{period})(\theta_2) + \epsilon \end{aligned} \tag{21}$$

where $\theta_1 - \theta_2 = \alpha - \gamma$ and $\theta_2 = \pi + \gamma$.

The values of θ_1 and θ_2 as well as their differences $\theta_2 - \theta_1$ and $\theta_1 - \theta_2$ are crucial to understanding the interdependencies among the temporal trends. Although each slope can take on any value between negative and positive infinity, their interdependencies result in identifiable linear combinations, which we have denoted θ_1 and θ_2. These quantities are generated by the unknown values of the slopes, not by an ad hoc constraint imposed by the researcher, and can be easily estimated from the data. As we show in the next section, we can accordingly use the estimates from Eqs. 19, 20, and 21 to derive a set of bounding formulas for the slopes.

4.4 Bounding the Slopes

As we have established, the quantities θ_1 and θ_2 and their differences define the interdependencies among the values of the true unknown slopes α, π, and γ. To construct bounding formulas for the slopes, we can use the constrained formulas in Eqs. 19, 20, and 21. For example, when we set the period slope to zero we obtain the constrained estimates $\alpha^* = \theta_1$, $\pi^* = 0$, and $\gamma^* = \theta_2$ (see Eq. 20), which have the following relationship to the true age, period, and cohort slopes:

$$\begin{aligned} \theta_1 &= \alpha + \nu \\ 0 &= \pi - \nu \\ \theta_2 &= \gamma + \nu \end{aligned} \tag{22}$$

Since we have constrained the period slope to be zero in the estimated model, we can, in a meaningful way, interpret the arbitrary constant ν in terms of upper and lower bounds on the unknown period slope. That is, if we let $\nu = \pi_{\max}$ then we have the following inequalities:

$$\begin{aligned} \theta_1 - \pi_{\max} &\le \alpha \\ \pi_{\max} &\ge \pi \\ \theta_2 - \pi_{\max} &\le \gamma. \end{aligned} \tag{23}$$

Likewise, if we let $\nu = \pi_{\min}$, then we have:

$$\begin{aligned} \theta_1 - \pi_{\min} &\ge \alpha \\ \pi_{\min} &\le \pi \\ \theta_2 - \pi_{\min} &\ge \gamma. \end{aligned} \tag{24}$$

Given we have set bounds on the age slope, Eqs. 23 and 24 together provide the formulas for placing bounds on the period and cohort slopes. In a similar way, we can construct bounding formulas using the constrained estimates when the age slope is zero (see Eq. 19) and when the cohort slope is zero (see Eq. 21).[20] We have summarized these bounding formulas in Table 1. These inequalities indicate that, once we set the bounds for any particular slope, all we need are the values of θ_1 and θ_2 to calculate the bounds on the remaining two slopes. Since θ_1 and θ_2 are invariant to

[20]The estimates when the age slope is constrained to equal zero are $\alpha^* = 0$, $\pi^* = \theta_1$, and $\gamma^* = \theta_2 - \theta_1$ and the corresponding estimates when the cohort slope is fixed to zero are $\alpha^* = \theta_1 - \theta_2$, $\pi^* = \theta_2$, and $\gamma^* = 0$.

the constraint imposed on the APC model, they are identifiable and can be estimated from the data.

Table 1 about here

For example, suppose we estimate an APC model and find that $\theta_1 = 3$ and $\theta_2 = -2$, such that $\theta_2 - \theta_1 = -5$ and $\theta_1 - \theta_2 = 5$. If we have reason to assume that the age slope has a lower bound of $\alpha_{\text{min}} = 0$ and an upper bound of $\alpha_{\text{max}} = 2$, then we can use the θ_1 and θ_2 to bound the period and cohort slopes as well. Specifically, by setting $\alpha_{\text{max}} = 2$, we know that the lower bound for the period slope is $\theta_1 - \alpha_{\text{max}} = \alpha + \pi - 2 = 3 - 2 = 1$. Likewise, by setting $\alpha_{\text{min}} = 0$, we know that the upper bound for the period slope must be $\theta_1 - \alpha_{\text{min}} = \alpha + \pi - \alpha_{\text{min}} = 3 - 0 = 3$. In a similar way we can find the bounds for the cohort slopes. By setting $\alpha_{\text{min}} = 0$ and $\alpha_{\text{max}} = 2$, we know that the lower bound of the cohort slope must be $\theta_2 - \theta_1 + \alpha_{\text{min}} = \gamma - \alpha + \alpha_{\text{min}} = -5 + 0 = -5$ and the upper bound must be $\theta_2 - \theta_1 + \alpha_{\text{max}} = \gamma - \alpha + 2 = -5 + 2 = -3$. Thus, by restricting the age slope to lie within zero and two we can conclude from the bounding formulas that the period slope must range from one to three and the cohort slope from negative three to negative five.

The bounding formulas also reveal that by setting the minimum age slope we are setting the minimum cohort slope but the *maximum* period slope; conversely, by setting the maximum age slope we set the maximum cohort slope but the *minimum* period slope (see Table 1). This distinct pattern can also be seen by examining the signs of ν in Eq. 8. The age and cohort slopes have a perfectly direct relationship with each other but a perfectly inverse relationship with the period slope. Relative to an arbitrary set of constrained estimates α^*, π^*, and γ^*, fixing ν at its maximum will result in the maximum value for the age and cohort slopes but the minimum value for the period slope. Likewise, setting ν at its minimum will result in the minimum value for the age and cohort slopes but the maximum value for the cohort slope. These interdependencies become important when considering various bounding strategies, because they suggest how under certain conditions we only need to specify the sign (and not the magnitude) of two slopes to obtain finite limits on the solution line.

5 Graphical Tools for APC Analysis

Although data visualization has a long history in APC modeling (Keiding 2011), there are no commonly-used graphical tools for evaluating and understanding the linear components of APC models. Previously we revealed how, although the individual slopes can take on any value, the data inform us of the interdependencies among the slopes by determining the values of θ_1 and θ_2. From these identifiable quantities we derived bounding formulas which indicate how setting the lower and upper limits on any particular slope restricts the range of the remaining two slopes. Now we show how the bounding formulas in Table 1 can be aided by a visualization of the interdependencies among the slopes, which we derive from a geometric interpretation of the APC identification problem (see O'Brien 2015: 59-89).

5.1 The Solution Line

In the absence of data, the age, period, and cohort slopes may take on any combination of values in a three-dimensional space. In this space, the equations $\theta_1 = \alpha + \pi$ and $\theta_2 = \pi + \gamma$ each define a two-dimensional plane. Again, suppose we have APC data with $\theta_1 = 3$ and $\theta_2 = -2$. In Fig. 1 (a) we depict the age-period plane defined by $\theta_1 = 3$ while in Fig. 1(b) we depict the period-cohort plane defined by $\theta_2 = -2$. The intersection of these two planes then defines a line, as illustrated in Fig. 1 (c) and Fig. 1 (d). This line is known as the *solution line*, since any point on the line is a potentially valid set of estimates for the unknown parameters α, π, and γ (O'Brien 2015: 72). That is, all points on the solution line give exactly the same values for Y and the slopes may take on any real number between negative and positive infinity. As a result, without additional assumptions the data are uninformative as to which combination of parameter values on this line should be preferred. Note that Fig. 1 is a graphical representation of the APC identification problem, since if there were no exact linear dependency then we would have three planes intersecting at a single point in the parameter space (O'Brien 2015: 67-68).

Fig. 1 about here

What is important to understand from Fig 1 is what, precisely, is accomplished when data are applied to an APC model. In our simple case where there are only linear components, the data has taken us from a three-dimensional space where all parameter values are possible to a one dimensional space, where only certain combinations of estimates lying on a line are consistent with the data. This same extent of reduction holds even if our data has nonlinearities, since they are fully identified. As such, the data has gone a long way towards constraining the possible estimates for the age, period, and cohort slopes, restricting an initial set of parameters that could be anywhere in a three-dimensional space to a set of values lying on a line. The data have been quite informative about our parameter values, just not quite as informative as we might like in the sense of providing unique point estimates for the linear trends.

5.2 2D-APC Graph of the Solution Line

Since the solution line is a geometric representation of the APC identification problem, we can use Fig. 1(d) to evaluate the consequences of setting bounds on one or more of the slopes. Although possible, this is difficult given that the solution line runs through three dimensions. Fortunately, since the solution line is determined by only two quantities, $\theta_1 = \alpha + \pi$ and $\theta_2 = \pi + \gamma$, we can flatten our three-dimensional representation of the solution line to two dimensions without loss of information. One 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. We call this a *2D-APC graph*.

Fig. 2 about here

In Fig. 2 we depict a 2D-APC graph based on APC data with values of $\theta_1 = 3$ and $\theta_2 = -2$. The solution line shown in Fig. 2 is identical to that in Fig. 1(d). The left vertical axis refers to the values of the age slope, the top and bottom horizontal axes to period, and right vertical axis to cohort. We will refer to each point in the coordinate space in terms of age, period, and cohort so that, for example, the point $(1, 3, -5)$ refers to $\alpha = 1$, $\pi = 3$, and $\gamma - 5$. The solution line runs from the upper-left to the bottom-right of the 2D-APC graph and the three dotted lines indicate when each of the age, period, and cohort slopes are set to zero, respectively. The point at which each dotted line intersects with the solution line gives the estimates for the "drop-one variable" models in Eqs. 19, 20, and 21. For example, the bottom horizontal dotted line, indicating when $\alpha = 0$, intersects the solution at $(0, 3, -5)$, which give the estimates from Eq. 19. If the true slope for age were zero, then we would know that the period slope is three and the cohort slope is negative five. The three dotted lines intersect each other at the origin for the age-period dimension and that for the period-cohort dimension.[21] In Fig. 2 we denote the age-period origin and the period-cohort origin with hollow circles at $(0, 0-5)$ and $(5, 0, 0)$, respectively.

Importantly, Fig. 2 gives us information on three important features of the data: (1) the slope of the solution line, which is always negative one; (2) the direction and scale of the axes; and (3) the values of θ_1, θ_2, and their differences. For all APC data (1) and (2) will be the same: the slope of the solution line relating period to age and cohort will always be negative one and, accordingly, the age and cohort axes will always run in the same direction while the period axis will run in the opposite direction.[22] However, depending on the unknown parameters α, π, and γ, the data will generate different values of θ_1, θ_2, and their differences. These values in turn affect the bounding formulas as well as the location of the solution line in the 2D-APC graph relative to the age-period origin and period-cohort origin.

To clarify the connection between the θ's and the 2D-APC graph in Fig. 2, it is useful to think of the relations among the APC slopes in terms of simple linear equations.[23] Since $\theta_1 = \alpha + \pi = 3$, we can write $\alpha = 3 - \pi$. We can think of $\alpha = 3 - \pi$ as a linear equation relating the scale for age (α) to the scale for period (π) with an intercept of three and a slope of negative one. Because the slope is negative one, the age and period scales have flipped signs and since the intercept is three, they are offset by three units. The vertical (and horizontal) distance from the age-period origin $(0, 0-5)$ to the solution line is given by $\theta_1 = 3$, as shown in Fig. 2. Thus, if the age (or period) slope is set to zero, then the value of the period (or age) slope must be $\theta_1 = 3$.

Similarly, since $\theta_2 = \pi + \gamma = -2$, we can write $\gamma = -2 - \pi$. This is a linear equation relating the scale for cohort (γ) to the scale for period (π) with an intercept negative two and a slope of negative one. Again, because the slope is negative the cohort and period scales have flipped signs and since the intercept is negative two, they are offset by two units. The vertical (and horizontal)

[21]The age-period-cohort origin is $(0, 0, 0)$, but this is not directly visible on the 2D-APC graph unless $\theta_2 - \theta_1 = 0$.

[22]This is also reflected in the opposing sign of ν in Eq. 8 for the period slope compared to that for the age and cohort slopes.

[23]We can likewise think of each plane in Fig. 1 as functions of linear equations based on θ_1 and θ_2. For example, the age-period plane is defined by $\theta_1 = \alpha + \pi$, so it is equivalent to the linear equation $\pi = \theta_1 - \alpha$. Likewise, we can think of the period-cohort plane as a function of $\theta_2 = \pi + \gamma$, with $\pi = \theta_2 - \gamma$.

distance from the period-cohort origin $(5, 0, 0)$ to the solution line is given by $\theta_2 = -2$, as shown in Fig. 2. This is again due to the fact that if the cohort (or period) slope is fixed to zero, then the period (or cohort) slope must be $\theta_2 = -2$. Finally, the offset between the age and cohort scales is given by $\theta_2 - \theta_1 = \gamma - \alpha = -5$. For example, when $\alpha = 0$ (as indicated by the bottom dotted horizontal line in Fig. 2), then we know $\gamma = -5$. Because the scales for age and cohort are offset by a known quantity, we can use the vertical axis to represent the linear trends for both age and cohort.

Fig. 3 about here

5.3 Bounds with the 2D-APC Graph

The 2D-APC graph is a visual representation of the solution line and, by extension, the APC identification problem. As a result, the bounding formulas in Table 1 have a one-to-one correspondence with any set of bounds placed over a corresponding 2D-APC graph. Continuing with our example in which $\theta_1 = 3$ and $\theta_2 = -2$, suppose we have solid theoretical reasons to restrict the range of the cohort slope to $-4 \leq \gamma \leq -2$. Then, using the bounding formulas in Table 1, we know that $(\theta_1 - \theta_2) - 4 \leq \alpha \leq (\theta_1 - \theta_2) - 2$ and $\theta_2 + 2 < \pi \leq \theta_2 + 4$. Since we know from the data that $\theta_1 = 3$ and $\theta_2 = -2$, then we can conclude that if the cohort slope ranges from negative four to positive two, then it must also be the case that $1 \leq \alpha \leq 3$ and $0 \leq \pi \leq 2$. Since the APC identification problem can be represented both mathematically and geometrically, we can visualize the bounds in a 2D-APC graph.

In Fig. 3(a) we visualize the bounds for the cohort slope ($-4 \leq \gamma \leq -2$) in three-dimensional space. The blue box denotes the region of the parameter space corresponding to range of negative four to negative two. The solution line cuts through this region. Since we have ruled out cohort slopes greater than negative two and less than negative four, we have restricted the set of feasible APC trends to those lying on the line in the feasible region (i.e., to those points lying on the solid line rather than the dotted line in Fig. 3(a)).

However, the consequences of assuming these particular bounds for the cohort slope are not clear in Fig. 3 (a) due to the inherent difficulties of interpreting a three-dimensional visualization on a two-dimensional surface. In Fig. 3 (b) we depict the same information in a 2D-APC graph, which clarifies the nature of our assumptions. Restricting the cohort slope to range from negative four to negative two reduces the parameter space to the blue rectangle. Since each of the three APC slopes must lie on the solution line, the data informs us that with this assumption the two extreme points on the line are $(1, 2, -4)$ and $(3, 0, -2)$. Thus, we can conclude that the age slope must range from one to three and that the period slope must range from zero to two. We can arrive at the same conclusion using the bounding formulas in Table 1, but our 2D-APC graph provides a visual analogue that can help researchers understand the implications of setting one restriction rather than another on the slopes for age, period, and cohort. Using the bounding formulas and our 2D-APC graph, in the next section we provide a more detailed overview of various strategies that can be used to produce informative bounds from an APC model.

6 Bounding Strategies

The bounding formulas and 2D-APC graph provide the basis for evaluating the consequences of setting one or more bounds on the magnitude and direction of one or more of the slopes from an APC model. In the foregoing section we outlined an example in which we specified lower and upper limits on the sign and size of a single slope. However, other strategies are possible, some of which may entail considerably weaker assumptions or facilitate the systematic evaluation of competing theoretical claims. In general, there are three main bounding strategies for identifying APC effects. Researchers can place bounds reflecting assumptions regarding: (1) the size and sign of one or more linear trends; (2) just the sign of one or more linear trends; and (3) the shape of one or more overall trends. We review each of these in turn.

6.1 Constraints on the Sign and Size

The most straightforward strategy entails restricting the direction as well as the magnitude on the linear trends. For example, we can set lower and upper limits on a single slope, reflecting our assumptions about its size and sign. The disadvantage with this strategy is that it may require relatively strong claims about a single linear trend to obtain meaningful lower and upper limits for all temporal trends. Narrower bounds can typically be achieved by setting bounds on more than one slope. Importantly, however, in some instances we can obtain narrower bounds by specifying *wider* bounds on two or more of the slopes. For example, suppose again that our data generates values of $\theta_1 = 3$ and $\theta_2 = -2$. In Fig. 2 we assumed that $-4 \leq \gamma \leq -2$, resulting in $1 \leq \alpha \leq 3$ and $0 \leq \pi \leq 2$. Yet we can place a wider bound on the cohort slope with an equally wide bound on the period slope to derive narrower bounds on all three linear trends. If we set $-4 \leq \gamma \leq -1$, then from the bounding formulas we know that $1 \leq \alpha \leq 4$ and $-1 \leq \pi \leq 2$. Similarly, if we set $1 \leq \pi \leq 4$, then this implies that $-1 \leq \alpha \leq 2$ and $-6 \leq \gamma \leq -3$. Combining these two sets of inequalities results in overall bounds of $1 \leq \alpha \leq 2$, $1 \leq \pi \leq 2$, and $-4 \leq \gamma \leq -3$. Consequently, we have been able to obtain narrower bounds on the trends by placing wider bounds than previously, but on multiple slopes.

Another advantage of placing multiple bounds is that we can rule out particular theoretical assertions, since certain combinations of bounds may lie off the solution line. For example, if we assume that $-3 \leq \gamma \leq 0$ and $2 \leq \pi \leq 5$, then the final set of bounds excludes the solution line. If there is a theory that the period and cohort slopes are within these ranges, then we can conclude that these claims are not supported by the data; one or both assumptions must be wrong.

6.2 Constraints on the Sign Only

In many empirical examples, theorizing about the magnitude of an effect may be difficult or even impossible. Another strategy is to constrain only the direction of one or more slopes. These bounds are necessarily one-sided, with one end of the interval at zero and the other to either positive or negative infinity. Remarkably, 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. The top part of Table 2 shows the simplified form of the bounding formulas when considering only the sign of a single linear trend.

To illustrate the application of bounds based only specifying the sign of a slope, consider APC data with values of $\theta_1 = 1$ and $\theta_2 = -1$. Suppose we have theoretical reasons to claim that the age slope is non-positive. Using the one-sided bounding formulas in Table 2, we know that $\theta_1 \leq \pi < +\infty$ and $-\infty < \gamma \leq \theta_2 - \theta_1$. Since $\theta_2 - \theta_1 = -2$, then we can conclude that the period slope is $1 \leq \pi < +\infty$ and the cohort slope is $-2 \leq \gamma < -\infty$. In other words, by positing the direction of the age linear trend, we can conclude that the cohort slope must be negative. Moreover, even though we only constrained *sign* of a single slope, we are able to make conclusions about the *magnitude* of the cohort slope; specifically, we know that the cohort slope must be less than or equal to negative two if the age slope is non-positive. Likewise, we can conclude that the period slope must be greater than or equal to positive one.

Table 2 about here

One drawback of setting the sign of a single slope is that the final set of bounds are not finite. However, we only obtain finite bounds for the slopes when one-sided intervals operate in opposite directions. Since the age and cohort slopes have a perfectly direct relationship with each other but a perfectly inverse relationship with the period slope, some sets of one-sided intervals will never result in finite bounds. For example, suppose we assume that α and γ are both non-negative and that we know from our data that $\theta_1 = 1$ and $\theta_2 = -1$. By assuming the age slope is non-negative we can conclude that $0 \leq \alpha < +\infty$, $-\infty < \pi \leq 1$, and $-2 \leq \gamma < +\infty$. By assuming the cohort slope is non-negative, then we know that $-2 \leq \alpha < +\infty$, $-\infty \leq \pi < -1$, and $0 \leq \gamma < +\infty$. Combining these inequalities gives us final bounds of $0 \leq \alpha < +\infty$, $-\infty < \pi \leq -1$, and $0 \leq \gamma < +\infty$. Unfortunately, because the one-sided bounds are infinite on the same side (i.e., the bounds are in the same direction), our final bounds are not finite.

In the lower part of Table 2 we show the combinations of one-sided bounds that can result in final bounds that are finite. Bounds based on these assumptions will result in finite bounds or bounds that can be ruled out as inconsistent with the data. Returning to our example in which we have $\theta_1 = 1$ and $\theta_2 = -1$, suppose we assume that age and period are both non-negative. The formulas in Table 2 tell us that $0 \leq \alpha \leq 1$, $0 \leq \pi \leq 1$, and $-2 \leq \gamma - 1$. Consequently, we can conclude that if the age and period slopes are both non-negative, then the age slope must lie within zero and one, the period slope must lie within zero and one, and the cohort slope must lie within negative two and negative one. Conversely, if we assume that the age and cohort slopes are both non-positive, then we know that one or both of these assumptions must be incorrect since we would obtain $\theta_1 \leq \alpha \leq 0$ is impossible if $\theta_1 = 1$.

The key to our approach entails using the bounding formulas and a 2D-APC graph to analyze the consequences of assumptions on the size and sign of the slopes from an APC model. However, so far we have only considered the case where age, period, and cohort only consist of linear trends.

This is a worst-case scenario, since in reality temporal trends are composed of both linear and nonlinear components, the latter which are fully-identified and will help us narrower the bounds even further.

6.3 Shape Constraints

We have only considered bounding the linear trends so far. However, as shown in Eq. 6, a more general APC model partitions the overall temporal trends into linear and nonlinear components. Because the nonlinearities are identified, these can help us further reduce the overall bounds through the specification of shape constraints. We introduce some additional notation to clarify how we can incorporate nonlinearities into our bounding analyses. Let $\alpha_{m.i.}$ denote a monotone increasing ($m.i.$) age slope and $\alpha_{m.d.}$ a monotone decreasing ($m.d.$) age slope. If we have an APC model with only linear components the situation is simple: a monotonically increasing age trend is equivalent to a positive slope and a monotonically decreasing trend is equivalent to a negative slope. However, with nonlinear components we can distinguish between shape (e.g., monotone decreasing or increasing) and direction (e.g., non-positive or non-negative), resulting in a considerably expanded set of possible bounding strategies.

One possibility is to specify that the overall age trend is monotonically increasing over some range of values corresponding to an age slope of $\alpha_{m.i.} \leq \alpha < +\infty$. For example, it is reasonable to assume that the probability of developing certain kinds of cancer will either increase or stay the same as one ages (e.g., Ames et al. 1993; Harman 1956; Pryor 1982). A second possibility is to specify an overall trend that is decreasing monotonically, with an age slope of $-\infty < \alpha \leq \alpha_{m.d.}$. For instance, crime rates among men are widely believed to decrease or stay the same after young adulthood (e.g., Farrington 1986; Hirschi and Gottfredson 1983). In some cases it might be reasonable to assume that a trend is monotonically increasing over one range but monotonically decreasing over another. For example, crime rates among men might be assumed to increase monotonically in adolescence, but decrease monotonically in later adulthood, corresponding to what has been called the "age-crime curve" (Loeber and Farrington 2014).

Non-monotonicity constraints are also possible, such as the assumption that an age trend is *not* monotone decreasing ($-\infty < \alpha \leq \alpha_{m.i.}$) or *not* monotone decreasing ($\alpha_{m.d.} < \alpha \leq +\infty$).

For example, one might assume that the probability of developing a particular cancer increases with age, but allow for declines so long as they are not monotonically decreasing across all age groups. Furthermore, in some instances it might make sense to assume a temporal trend is *neither* monotone increasing nor decreasing, such that $\alpha_{m.d.} \leq \alpha \leq \alpha_{m.i.}$. For example, when examining unemployment rates it might make sense to rule out any monotonic trend across periods, as this may violate theoretical assumptions regarding a business cycle (e.g., Burns and Mitchell 1946).

Fig. 4 about here

To illustrate how we can set bounds using shape constraints, let us consider APC data with $\theta_1 = 2$ and $\theta_2 = 1$ as well nonlinearities for age and period. The nonlinearities for the age groups are $\widetilde{\alpha} = \{-5, 0, 2, 3, 2, 0, -1.5, -3\}$ and are displayed as a dotted line in Fig. 4 (a). Assume that we believe that the overall trend in age is monotonically increasing. As can be seen in Fig. 4 (a), the nonlinearities are increasing at some intervals and decreasing at other intervals, relative to an unknown linear age trend. We want to specify a value for the linear age trend that ensures that between any two adjacent age categories that at the minimum the trend is flat. To do so we need only find that pair of neighboring age categories in which the downward trend is the most negative. This can be computed using the following equation:

$$\alpha_{m.i.} = \begin{cases} (-1) \times min(\Delta\widetilde{\alpha}_{I-1}) & \text{if} \quad min(\Delta\widetilde{\alpha}_{I-1}) < 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{25}$$

where the Δ notation indicates forward differences (i.e., $\widetilde{\alpha}_{i+1} - \widetilde{\alpha}_i$) and the subscript $m.i.$ indicates the resulting linear trend is monotone increasing. We can use Eq. 25 to calculate the minimum age slope required for a monotonic increasing overall trend. The forward differences of the age nonlinearities in Fig. 4 (a) are $\Delta\widetilde{\alpha}_{I-1} = \{5, 2, 1, -1, -2, -1.5, -1.5\}$. The minimum of these differences is negative two, which is the difference between the 35 and 40 age groups. To counter this downward deviation, the parameter value for the linear age term must be greater to or equal positive two. In Fig. 4 (b) we illustrate what happens to the overall age trend when the slope is set to positive two, which is shown by the solid line. As can now be seen the overall trend in age is monotonically increasing. Any slope less than positive two will result in an overall trend that is either not monotonic or monotonic decreasing over the full set of age groups. Alternatively, any slope greater than two will be monotonically increasing; this is the *minimum* monotonically increasing age slope consistent with our data.

We can determine the implications of a monotone increasing age trend by using our 2D-APC graph. The results for age are shown in Fig. 5 (c). By assuming the age trend increases monotonically, we can conclude that the period slope must be negative and the cohort slope must be positive. Although we have determined the direction of the period and cohort slopes, we may want finite bounds for our slopes. One possibility is to assume the overall period trend is *not* monotone decreasing. This is a weaker assumption than assuming the trend is monotonically increasing, since we allow downward deviations as long as they are not *always* decreasing or staying the same across adjacent age groups.

The nonlinearities for the period groups are $\widetilde{\pi} = \{0, -1, 0, 1, 2, -2\}$ and are displayed as a dotted line in Fig. 4 (c). To find the period slope consistent with a monotonically increasing trend, we need only find that pair of adjacent period groups in which the downward trend is most positive. This can be computed using the following equation:

$$\pi_{m.d.} = \begin{cases} (-1) \times max(\Delta\widetilde{\pi}_{J-1}) & \text{if} \quad max(\Delta\widetilde{\pi}_{J-1}) > 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{26}$$

where the Δ notation again indicates forward differences (i.e., $\widetilde{\pi}_{j+1} - \widetilde{\pi}_j$), and the subscript $m.d.$ indicates the linear trend is monotone decreasing. Using Eq. 26 we can calculate the period slope corresponding to an overall monotone decreasing trend. The forward differences for the period nonlinearities are $\Delta\widetilde{\pi}_{J-1} = \{-1, 1, 1, 1, -4\}$ and the maximum of these differences is positive one. To counter this upward deviation, the parameter value for the period slope must be less than or equal to negative one. When we set the slope to negative one we obtain the solid line in Fig. 4 (d), which corresponds to the maximum period slope required for a monotonically decreasing overall period trend. Any slope less than negative one will result in a steeper downward period trend.

Fig. 5 about here

However, in our case we would like to assume that the overall period trend is *not* monotone decreasing. Thus, the solid line in Fig. 4 (c) is the most downward trend we are willing to consider; that is, we are assuming the period slope is negative one or greater. The consequences of this assumption are shown in Fig. 5 (c), which restricts the age slope to be less than three and the cohort slope less than two. Reflecting the weak assumptions of our constraint, we cannot make any decisive conclusions about the direction of any of the temporal slopes.

Notwithstanding, we can combine our assumptions to obtain finite bounds for all three slopes. This is shown in Fig. 4 (d). By assuming the overall age trend is monotone increasing and the period trend is not monotone decreasing, we can conclude that the age slope is between two and three, the period slope is between negative one and zero, and the cohort slope is between one and two.

The bounding strategies outlined here are flexible and can be combined in a multitude of ways. In a worst-case scenario in which we only have the linear components (that is, the nonlinearities are zero for all three time scales), we can still set up two-sided bounds reflecting assumptions about the size and direction of one or more of the slopes. We can also obtain finite bounds under certain conditions by specifying only the direction of two or more slopes. With the nonlinear components we can include assumptions about the shape, size, and direction of the overall trends for specified ranges of each variable, greatly expanding the variety of bounding strategies available.

7 Empirical Examples

To demonstrate the applicability of our approach, we provide two empirical analyses. In the first example we bound the temporal trends of the incidence of prostate cancer while in the second we examine over-time trends in homicide rates. Both examples have received considerable attention in the APC literature (for example, see Holford 1983; O'Brien 2015).

7.1 Prostate Cancer

A large body of research shows that the risk of prostate cancer increases with age and is diagnosed in very few people aged 50 years or younger (Holford 1983; James and Segal 1982). After this age, incidence rates increase greatly. The reason for this increase is not clear, but in general it is thought to be due to an accumulation of exposure risks combined with a loss of effectiveness in cellular repair mechanisms. Untangling temporal trends in prostate cancer is crucial for understanding likely causes, particularly societal or environmental, that may be increasing or decreasing the incidence. In light of this background knowledge, a reasonable assumption for APC modeling is that the risk of prostate cancer among men increases monotonically over at least some range of the age groups.

To examine prostate cancer trends, we collected data from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. The SEER program is the primary source of over-time population-based cancer incidence in the United States and is considered the "gold standard" for cancer registration data worldwide (Siegel et al. 2017: 2). We used data on white men and black men with prostate cancer who were diagnosed from 1973 to 2013 among residents of nine geographic areas: Connecticut, Hawaii, Iowa, New Mexico, Utah, Atlanta, Detroit, Seattle-Puget Sound, and San Francisco-Oakland. Given the rarity of prostate cancer among young people, following previous research we restricted our analysis to men aged 40-44 to 85-89. We grouped the years into periods from 1970-1974 to 2010-2014.[24] The outcome variable is the number of malignant prostate cancer diagnoses per $100,000$ men for each age-period cell. We logged the outcome since it is highly skewed.

Fig. 6 about here

To identify what can be known from the data without prior constraints, we fit a classical linear regression model predicting the logged prostate cancer rate using age, period, and cohort as inputs with a design matrix that partitions the linear and nonlinear components (cf. Eq. 6).[25] From this model we estimate that $\theta_1 = \alpha + \pi = 6.670$ and $\theta_2 = \pi + \gamma = 3.242$.[26] As discussed in the previous sections, the values of θ_1 and θ_2 define a solution line, which is visualized in Fig. 6 (a). The nonlinearities are visualized in Fig. 7 (a), (b), (c), with the horizontal dashed lines denoting the overall (or grand) mean in the data.[27] The nonlinearities can be difficult to interpret, since they are deviations from unknown linear trends. The easiest way to interpret these deviations is by recognizing that, if the slopes were all zero, then these are the temporal trends we would observe in the data.[28]

[24]The raw data are yearly, ranging from 1973-2013.

[25]Nearly identical results were obtained with Poisson regression, but for ease of exposition we present our findings using a classical linear regression with a logged rate outcome.

[26]Estimated values of both θ_1 and θ_2 are statistically significant at the conventional threshold of 0.05.

[27]F-tests indicate that all three time scales have statistically significant nonlinearities at the conventional threshold of 0.05.

[28]However, note that we know that these interpretations are not strictly correct, since we can estimate the values of θ_1 and θ_2. For example, if we assume that the period slope is zero, then it must be the case that the age slope is 6.670 rather than zero.

At this point we have exhausted what can be known from the data and we must use additional constraints to split the two estimates θ_1 and θ_2 into three separate age, period, and cohort slopes. An initial assumption is that prostate cancer incidence increases monotonically from ages 40 to 85, which results in a minimum age slope of 7.310. This weak constraint restricts the parameter space of the linear components to the shaded region shown in Fig. 6 (b). Under this assumption, which reflects what is widely believed about cancer etiology, we can conclude that the cohort linear trend is positive and that the period linear trend is either slightly positive or negative. Importantly, this initial bounding analysis informs us that, in general, more recent birth cohorts are at greater risk of being diagnosed with prostate cancer than earlier cohorts.

To obtain finite bounds on the solution line we need to apply an additional constraint. One plausible assumption is that the period slope is *not* monotone decreasing, since screening tests for prostate cancer have improved over calendar time. Most notably in the United States, the prostate-specific antigen (PSA) blood test introduced in the 1980s likely reduced prostate cancer rates. This assumption results in a minimum period slope of -3.453, with corresponding bounds visualized in Fig. 6 (c). Combining the assumption of a monotonic increasing age trend from 40 to 85 with a period slope that is *not* monotone decreasing, we obtain the bounds shown in Fig. 6 (d). With these two assumptions, we have restricted the temporal slopes to $7.400 \leq \alpha \leq 10.100$, $-3.430 \leq \pi \leq -0.730$, and $3.972 \leq \gamma \leq 6.672$.

Fig. 7 about here

When we incorporate the constrained linear trends with the identifiable nonlinearities, we obtain the overall trends shown in Fig. 7 (d), (e), (f). The shaded areas in Fig. 7 (d), (e), (f) indicate the range of estimates consistent with our assumptions, while the dashed lines denote the average of the upper and lower bounds for each of the age, period, and cohort trends. These dashed lines correspond to values of $\alpha = 8.750$, $\pi = -2.080$, and $\gamma = 5.322$, respectively. From these results, reflecting basic theories of the causes of cancer over the life course and through calendar time, we can make several conclusions. First, regarding the period trend, prostate cancer incidence remained relatively flat until the early 1990s, after which it clearly decreased. Second, the cohort trend shows a major increase across birth years, starting with those born in the 1920s. This may reflect the fact that fewer men are dying of preventable diseases and conditions, since prostate cancer occurs almost exclusively among older age groups. Further analyses, especially research that explicitly incorporates mechanisms for the temporal effects, can help explain these findings as well as assess the plausibility of these assumptions.

7.2 Homicide Rates

We now turn to our second empirical example, which deals with temporal trends in homicide arrest rates among men. For this analysis we obtained data on homicide arrest rates collected from 1980 to 2012 by the Federal Bureau of Investigation (FBI) through the Uniform Crime Reporting (UCR) Program. The UCR Program offers the highest-quality historical arrest data in the United States, based on contributions from over 18,000 agencies. Before analyzing the data, we collapsed the period groups into five-year intervals ranging from $1980 - 1984$ to $2010 - 2014$ and the age groups into five-year intervals ranging from $10 - 14$ to $60 - 64$.

The solution line for the linear components gives estimates of $\theta_1 = \alpha + \pi = -3.357$ and $\theta_2 = \pi + \gamma = -3.583$.[29] Both of these estimates suggest declines in the homicide rate across age and cohort groups, but these linear trends are aliased by the presence of the period slope. However, we can rule out from the data alone particular combinations of the slopes. Specifically, it is impossible that α and π are both positive or that π and γ are both positive. Any valid theory of changes in homicide arrest rates must assume that at least one of the temporal variables has a negative slope.

The nonlinearities are shown in Fig. 8 (a), (b), (c).[30] As with the previous example, these can be difficult to interpret since they are deviations from an unknown linear trend, but the shape of the age nonlinearities is consistent with an underlying age-crime curve. As visualized in Fig. 8 (a), if the age slope were zero then we would indeed observe a sharp increase in the homicide arrest followed by a steady decline.

As discussed previously, a reasonable assumption is that there in fact exists an age-crime curve for men. That is, we assume the crime rate increases monotonically to age 20-24, and then decreases monotonically afterwards. Assuming that there is an age-crime curve of this form, then

[29] Estimated values of both θ_1 and θ_2 are statistically significant at the conventional threshold of 0.05.

[30] F-tests show that, at the conventional threshold of 0.05, all three time scales exhibit statistically significant nonlinearities.

the age slopes lies between -3.026 to 1.247, inclusive.[31] If we assume that homicide arrest rates among men follows an age-crime curve, then we can rule out large sections of the parameter space. Specifically, we can conclude that the period slope must range from -0.331 to -4.604 while the cohort slope most range from -3.252 to 1.021. The corresponding overall trends are shown in Fig. 8 (d), (e), (f). The shaded regions indicate the upper and lower bounds for each of the temporal trends while the dashed lines indicate the average of the upper and lower bounds for each of the temporal trends. Specifically, the dashed lines correspond to values of $\alpha = -1.500$, $\pi = -1.857$, and $\gamma = -1.726$, respectively.

Several important conclusions follow from this initial bounding analysis of homicide arrest rates. As shown in Fig. 8 (e), we can conclude that the overall period trend did *not* increase from 1970 to 2010 and that crime rates decreased among cohorts born in the 1980s and later. Crucially, *we are able to make these conclusions even though the slopes for age and cohort might be positive, zero, or negative*. The substantial nonlinearities in the data, which are entirely identifiable, allow us to make substantively important conclusions even though our assumptions have only restricted the direction of one out of the three linear trends.

So far we have only examined bounding the overall age trend for homicide arrest rates. We can further narrow these bounds with one or more additional assumptions. One relatively weak assumption is that the homicide arrest rates from the 1980s to mid-1990s did not uniformly decrease. This reflects the assumption that it is unreasonable to conclude a steady decline occurred over this period since major social events, particularly the crack epidemic in the United States, plausibly increased violent crime rates. Note that we are not assuming that there is a monotonic increase; rather, we are positing the weaker claim that there is no monotonic decrease, allowing for the possibility that the crime rate increased in some periods but decreased in other periods. We further weaken this assumption by restricting this constraint to the periods 1980-1984 to 1990-1994. Thus, we are making no assumptions directly about time periods before or after the 1980-1994 epoch. With this assumption, we can conclude that the slopes are negative for all three temporal variables. Specifically, we know that $-3.026 \leq \alpha \leq -2.593$, $-0.764 \leq \pi \leq -0.331$, and $-3.252 \leq \gamma \leq -2.819$. When combined with the nonlinearities (see Fig. 8 (g), (h), (i)), we can

[31] The corresponding 2D-APC graphs are available in the online supplementary appendix.

conclude that age and cohort effects have been the primary sources of temporal variation in homicide arrest rates. The overall decline across cohorts is nearly monotonic, increasing very slightly only among cohorts born in the 1960s to early 1980s. This suggests that, although local variations likely exist, crime rates in the United States have plummeted mainly due to large differences between later and earlier birth cohorts.

8 Conclusion

Researchers in a variety of fields have sought to understand social change by identifying the effects of age, period, and cohort on a range of different outcomes. A variety of methods have been proposed to deal with the APC identification problem, most recently hierarchical (or multilevel) models and estimators based on the Moore-Penrose generalized inverse. In this paper we have outlined an alternative framework that entails placing bounds on the temporal trends. Our approach begins with identifying what can be known from the data alone with as few constraints as possible. Since full identification is impossible, we then illustrate how partial identification can be achieved based on constraining the expected size, sign, or overall shape of one or more of the temporal trends.

We make several important contributions in this paper. First, following previous work on APC models, we show that although nonlinear components are fully identified the linear components are not. Second, we demonstrate how the data provides us with identifiable linear combinations of the slopes even though each particular slope is unknown. These linear combinations in turn allow us to derive general bounding formulas on the slopes that in turn form the basis of our analyses. Third, we introduce a novel graphical display, the 2D-APC graph, which can be used to evaluate the consequences of various constraints on the linear trends. Finally, we present a variety of new bounding strategies, including those based on fixing the size and direction of one or more slopes, constraining only the sign of one or more slopes, or applying a shape constraint to the overall trends. These techniques offer a considerably wider range of identification approaches than commonly used in the literature.

The bounding analysis framework presented here is flexible and general. At least two major extensions are possible and warrant further research. First, our framework has a clear Bayesian interpretation. Specifically, rather than zeroing out regions of the parameter space, one could place a prior distribution over one or more parameters. For example, one might use a gamma distribution as a prior for the age slope, reflecting the assumption that the slope ranges from zero to positive infinity with some decreasing probability. Second, one can combine a bounds analysis with an APC model that includes mechanisms or proxy variables. While mechanisms and proxy variables introduce additional information into a traditional APC analysis, thereby allowing for identification, they are subject to potentially strong assumptions regarding model specification. If a researcher believes that important mechanisms or proxies are absent from the model, a bounds analysis such as that presented here can be used to evaluate the robustness of the estimated results.

Over the past several decades, a great deal of progress has been made on developing new methods for APC analysis. However, much work remains. Given the importance of estimating APC effects for understanding the basic contours of population-level change, it is vital that researchers ground their analyses in overt theoretical considerations and empirically-based constraints. Following earlier work by Holford (1985: 3), it is our belief that the bounds analysis presented here:

> …avoids the mystery of obtaining a completely different set of parameter estimates each time a different constraint is used. Instead, it points out exactly where the problem is and indicates which inferences can be made without ambiguity and which cannot. By presenting the data in this way, the interested reader can readily modify the results to see the effect of different assumptions, and translate the parameters accordingly.

Tables

Table 1: Bounding Formulas for APC Slopes

Age Bounds:	$\alpha_{\min} \leq \alpha \leq \alpha_{\max}$
	$\theta_1 - \alpha_{\max} \leq \pi \leq \theta_1 - \alpha_{\min}$
	$(\theta_2 - \theta_1) + \alpha_{\min} \leq \gamma \leq (\theta_2 - \theta_1) + \alpha_{\max}$
Period Bounds:	$\theta_1 - \pi_{\max} \leq \alpha \leq \theta_1 - \pi_{\min}$
	$\pi_{\min} \leq \pi \leq \pi_{\max}$
	$\theta_2 - \pi_{\max} \leq \gamma < \theta_2 - \pi_{\min}$
Cohort Bounds:	$(\theta_1 - \theta_2) + \gamma_{\min} \leq \alpha \leq (\theta_1 - \theta_2) + \gamma_{\max}$
	$\theta_2 - \gamma_{\min} \leq \pi \leq \theta_2 - \gamma_{\max}$
	$\gamma_{\min} \leq \gamma \leq \gamma_{\max}$

Notes: Age, period, and cohort slopes are α, π, and γ, respectively, with $(.)_{\min}$ and $(.)_{\max}$ denoting minimum and maximum values of the bounds. We denote $\theta_1 = \alpha + \pi$, $\theta_2 = \pi + \gamma$, $\theta_1 - \theta_2 = \alpha - \gamma$, and $\theta_2 - \theta_1 = \gamma - \alpha$.

Table 2: APC Bounds Given by Setting the Sign of One or Two Slopes

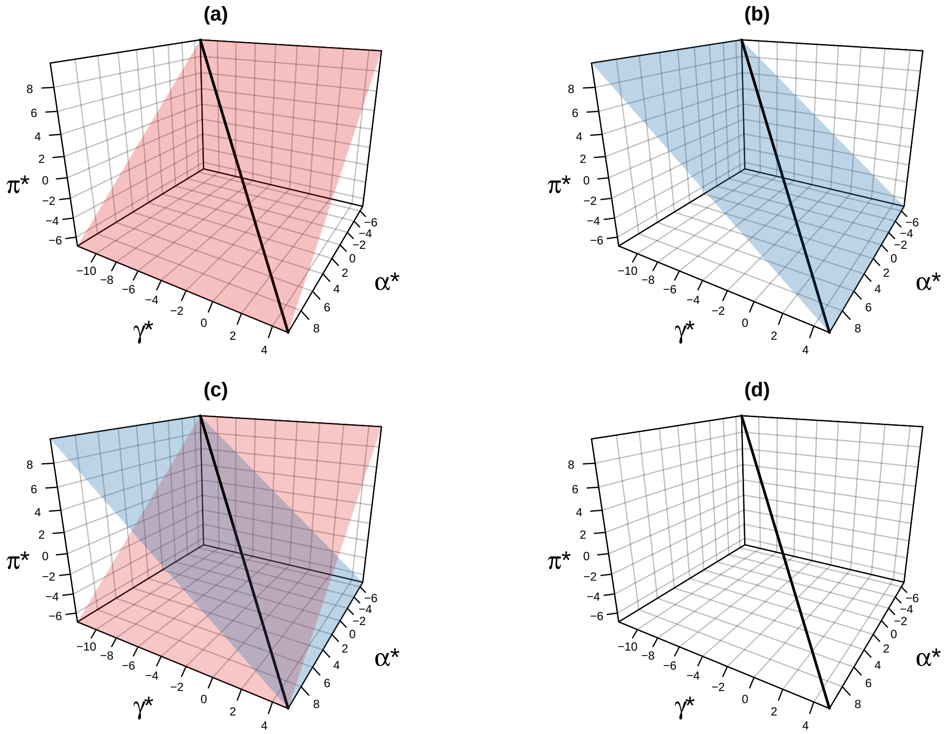
Sign of One Slope	Age	Period	Cohort
If $\alpha \geq 0$ then:	$0 \leq \alpha < +\infty$	$-\infty < \pi \leq \theta_1$	$(\theta_2 - \theta_1) \leq \gamma < +\infty$
If $\alpha \leq 0$ then:	$-\infty < \alpha \leq 0$	$\theta_1 \leq \pi < +\infty$	$-\infty < \gamma \leq (\theta_2 - \theta_1)$
If $\pi \geq 0$ then:	$-\infty < \alpha \leq \theta_1$	$0 \leq \pi < +\infty$	$-\infty < \gamma \leq \theta_2$
If $\pi \leq 0$ then:	$\theta_1 \leq \alpha < +\infty$	$-\infty < \pi \leq 0$	$\theta_2 \leq \gamma < +\infty$
If $\gamma \geq 0$ then:	$(\theta_1 - \theta_2) \leq \alpha < +\infty$	$-\infty < \pi \leq \theta_2$	$0 \leq \gamma < +\infty$
If $\gamma \leq 0$ then:	$-\infty < \alpha \leq (\theta_1 - \theta_2)$	$\theta_2 \leq \pi < +\infty$	$-\infty < \gamma \leq 0$

Sign of Two Slopes	Age	Period	Cohort
If $\alpha \geq 0$ and $\pi \geq 0$ then:	$0 \leq \alpha \leq \theta_1$	$0 \leq \pi \leq \theta_1$	$(\theta_2 - \theta_1) \leq \gamma \leq \theta_2$
If $\alpha \leq 0$ and $\pi \leq 0$ then:	$\theta_1 \leq \alpha \leq 0$	$\theta_2 \leq \pi \leq 0$	$\theta_2 \leq \gamma \leq (\theta_2 - \theta_1)$
If $\pi \geq 0$ and $\gamma \geq 0$ then:	$(\theta_1 - \theta_2) \leq \alpha \leq \theta_1$	$0 \leq \pi \leq \theta_2$	$0 \leq \gamma \leq \theta_2$
If $\pi \leq 0$ and $\gamma \leq 0$ then:	$\theta_1 \leq \alpha \leq (\theta_1 - \theta_2)$	$\theta_2 \leq \pi \leq 0$	$\theta_2 \leq \gamma \leq 0$
If $\alpha \geq 0$ and $\gamma \leq 0$ then:	$0 \leq \alpha \leq (\theta_1 - \theta_2)$	$\theta_2 \leq \pi \leq \theta_1$	$(\theta_2 - \theta_1) \leq \gamma \leq 0$
If $\alpha \leq 0$ and $\gamma \geq 0$ then:	$(\theta_1 - \theta_2) \leq \alpha \leq 0$	$\theta_1 \leq \pi \leq \theta_2$	$0 \leq \gamma \leq (\theta_2 - \theta_1)$

Notes: Age, period, and cohort slopes are α, π, and γ, respectively, with $(.)_{\min}$ and $(.)_{\max}$ denoting minimum and maximum values of the bounds. We denote $\theta_1 = \alpha + \pi$, $\theta_2 = \pi + \gamma$, $\theta_1 - \theta_2 = \alpha - \gamma$, and $\theta_2 - \theta_1 = \gamma - \alpha$.

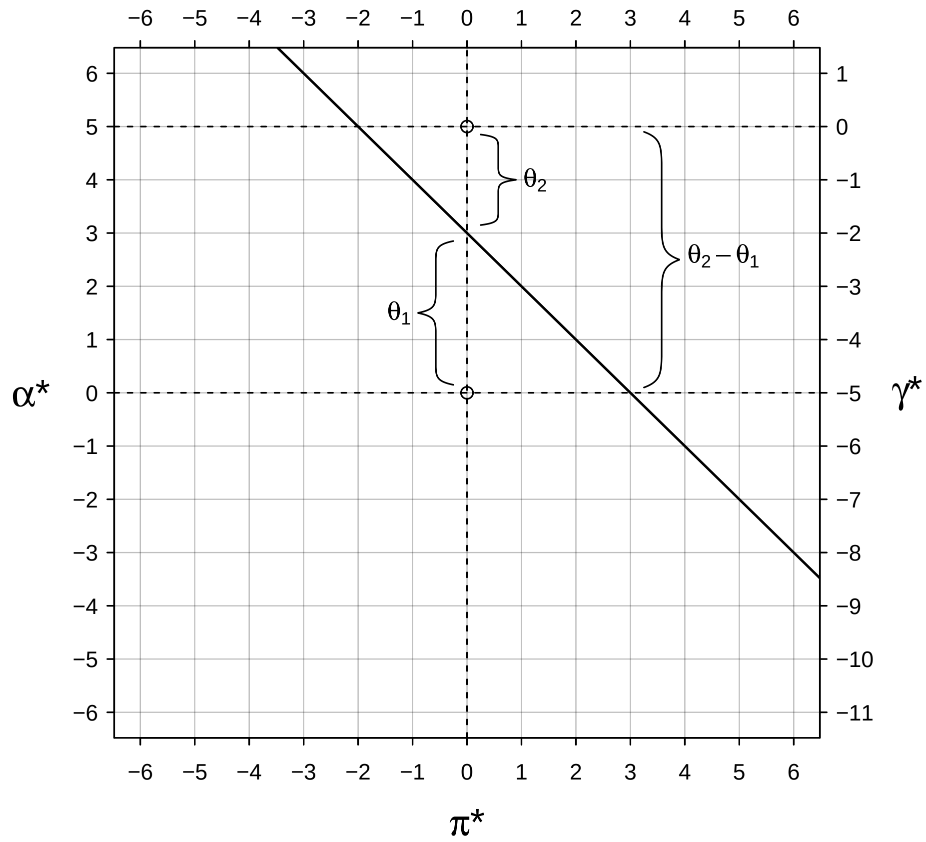
Figures

Fig. 1: Geometric Derivation of the Solution Line



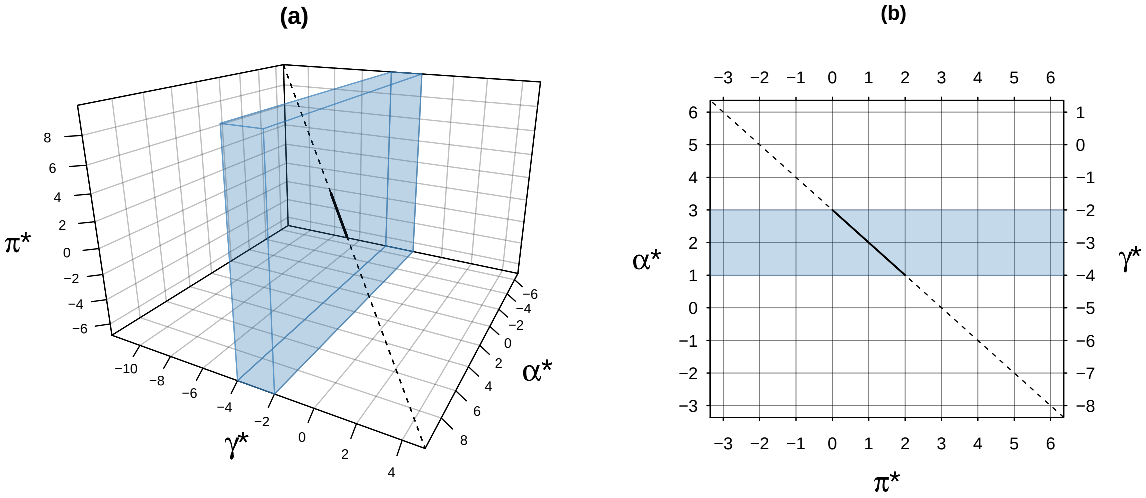
Notes: Panel **(a)** shows an age-period plane defined by $\pi = \theta_1 - \alpha$, where $\theta_1 = 1$. Panel **(b)** shows a period-cohort plane defined by $\pi = \theta_2 - \gamma$, where $\theta_2 = -1$. Panel **(c)** visualizes the intersection of the age-period and period-cohort planes, while **(d)** visualizes the solution line based on $\theta_1 = 1$ and $\theta_2 = -1$.

Fig. 2: 2D-APC Graph of the Solution Line



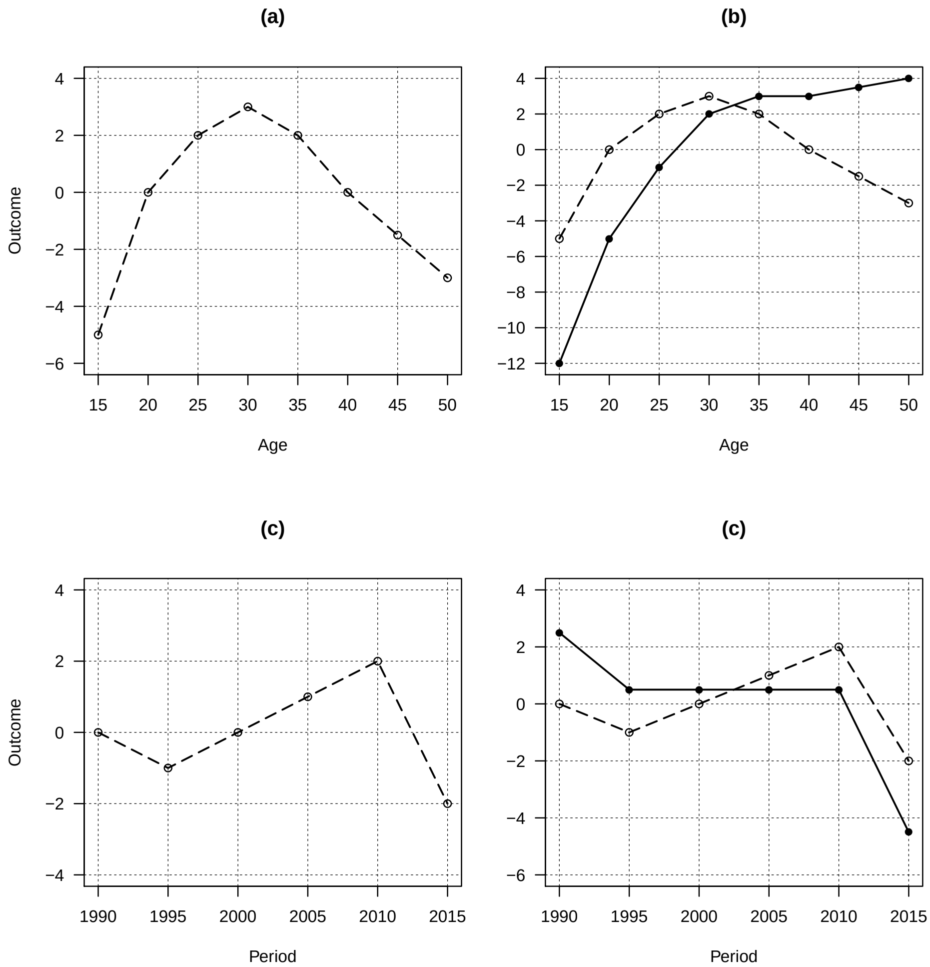
Notes: Solid black line denotes the solution line, with $\theta_1 = \alpha + \pi$, $\theta_2 = \pi + \gamma$, and $\theta_2 - \theta_1 = \alpha - \gamma$. Dotted lines refer to values in the parameter space when α^*, π^*, or γ^* are set to zero. Hollow circles denote the age-period and period-cohort origins. Simulated data based on values of $\theta_1 = 3$ and $\theta_2 = 2$.

Fig. 3: Solution Line with Bounds on Cohort Slope: 3D- vs. 2D-APC Graph



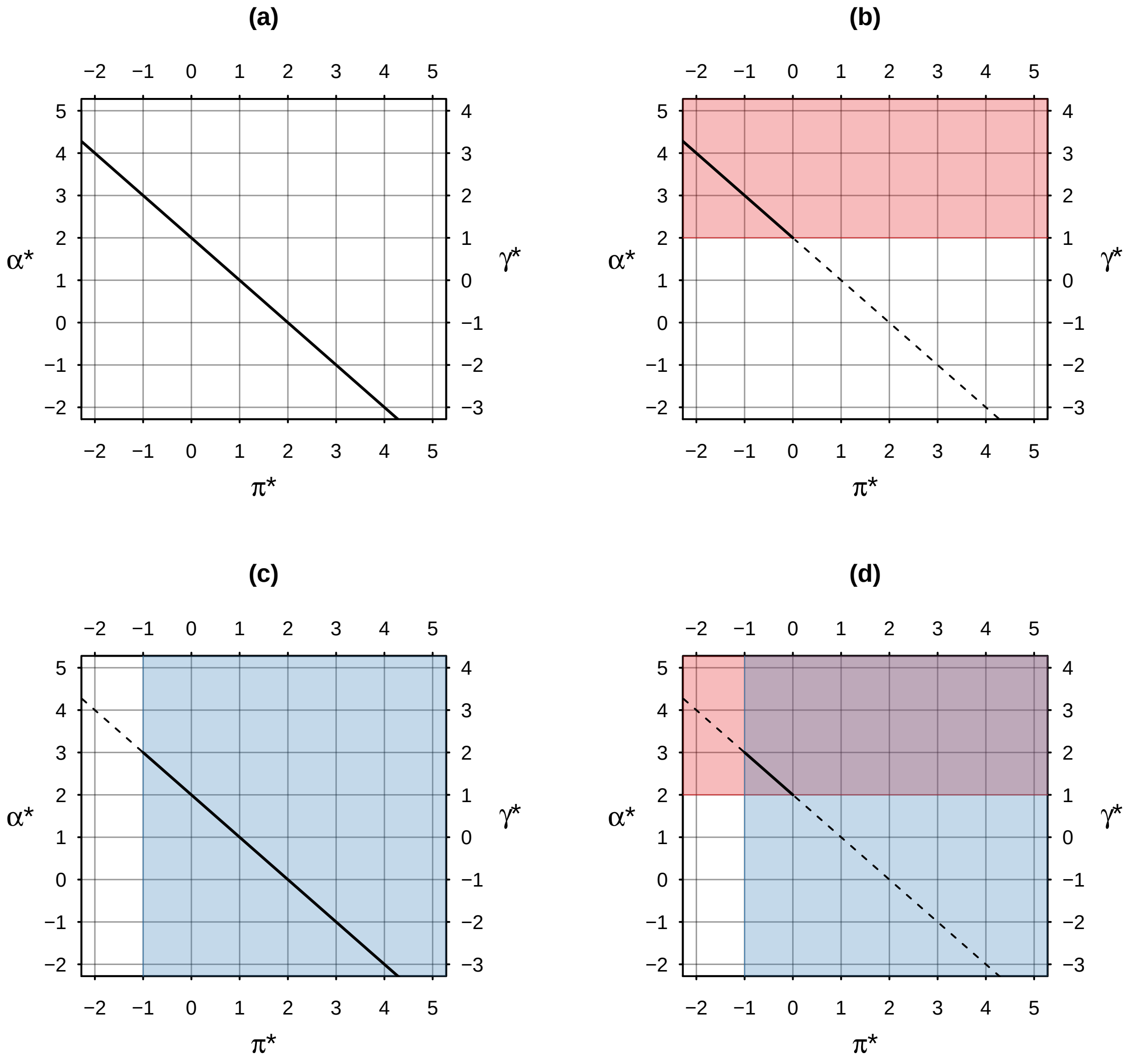
Notes: Panels **(a)** and **(b)** give 3D and 2D representations of specified bounds on the solution line. Shaded areas reflect the assumption that $1 \leq \alpha \leq 2$. Simulated data based on values of $\theta_1 = 3$ and $\theta_2 = 2$.

Fig. 4: Monotone Constraints
on Nonlinear Components of Age and Period



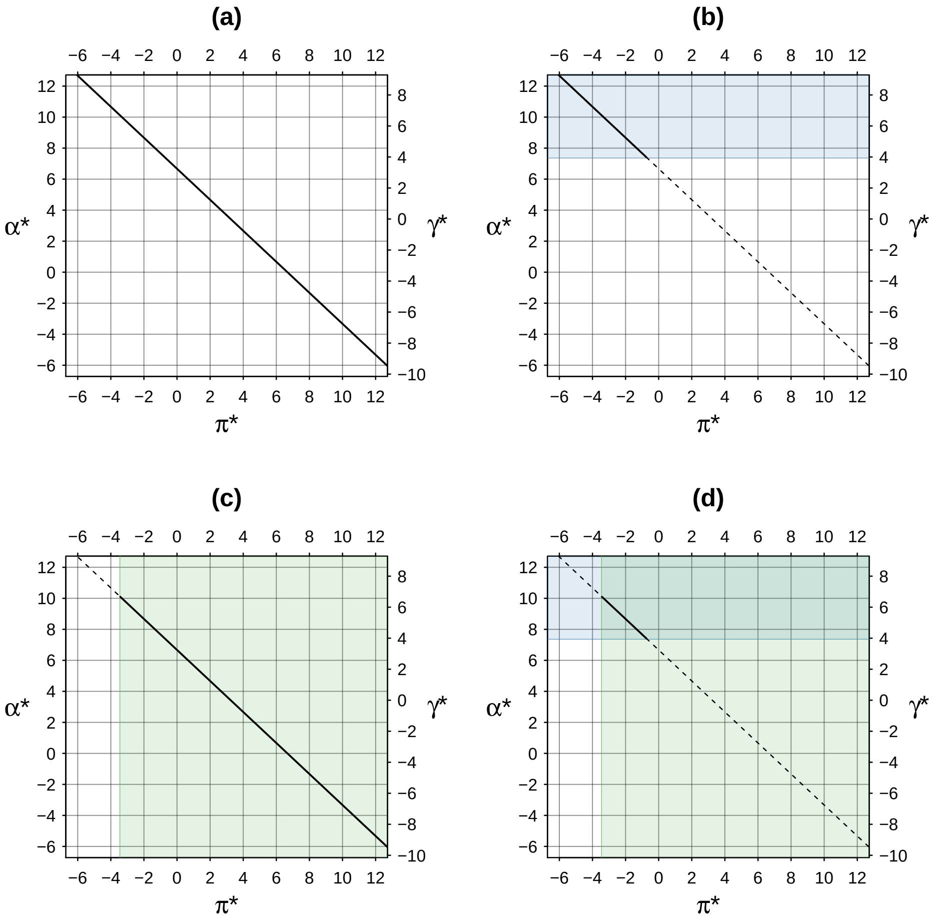
Notes: The dotted lines in panels **(a)** and **(c)** show simulated nonlinearities for age and period. The solid line in panel **(c)** reflects the assumption that $\alpha = 2$, which is the minimum age slope required for a monotonic increasing overall age trend. The solid line in panel **(d)** reflects the assumption that $\pi = -1$, which is the maximum period slope required for a monotonic decreasing overall period trend.

Fig. 5: 2D-APC Graphs with One-Sided Bounds on Age and Period Linear Components



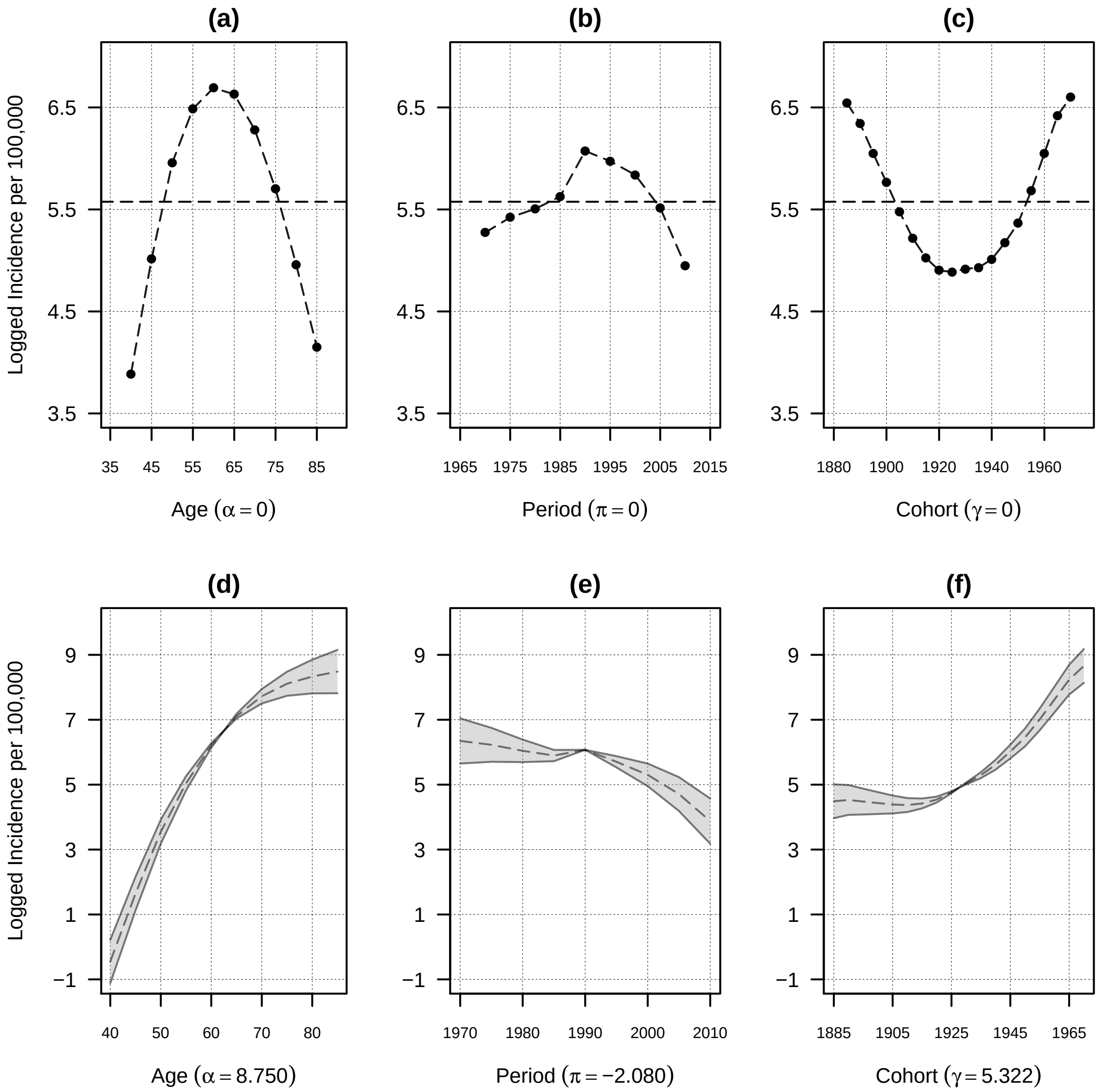
Notes: Panel **(a)** gives the 2D-APC graph of the solution line for simulated data with $\theta_1 = 2$ and $\theta_2 = 1$. Panel **(b)** shows the bounds for $2 \leq \alpha < +\infty$, reflecting the assumption that the overall age trend is monotonically increasing. Panel **(c)** shows bounds for $-1 < \pi < +\infty$, reflecting the assumption that the overall period trend is *not* monotonically decreasing. Panel **(d)** shows the bounds for $2 \leq \alpha < +\infty$ as well as $-1 < \pi < +\infty$.

Fig. 6: 2D-APC Graphs with Bounds
on Age and Period Linear Components: Prostate Cancer Incidence



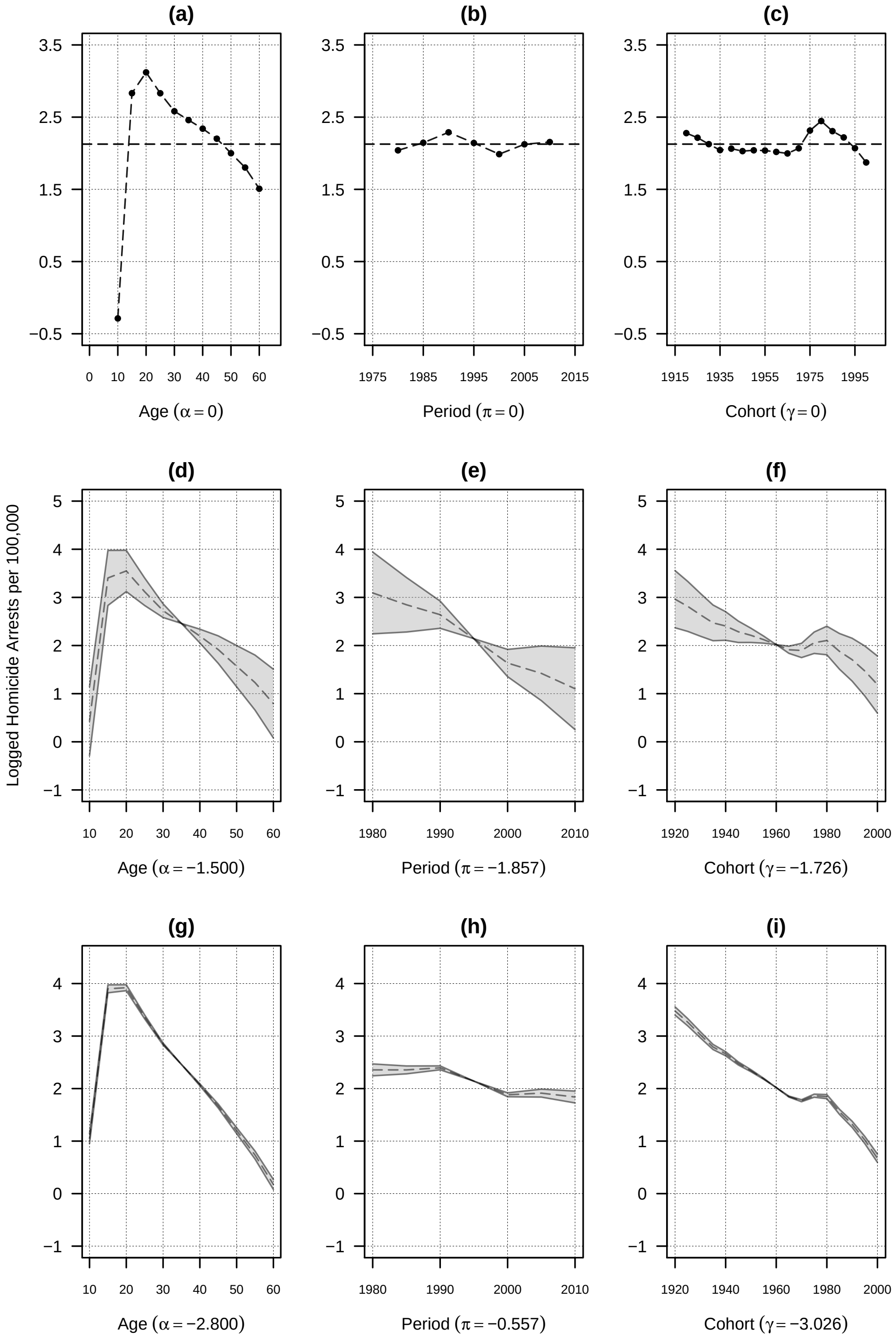
Notes: Panel **(a)** gives the 2D-APC graph of the solution line with $\theta_1 = 6.670$ and $\theta_2 = 3.242$. Panel **(b)** shows the bounds for $7.361 \leq \alpha < +\infty$, reflecting the assumption that the overall age trend is monotonically increasing from ages 40 to 85. Panel **(c)** shows bounds for $-3.453 < \pi < +\infty$, reflecting the assumption that the overall period trend is *not* monotonically decreasing. Panel **(d)** shows the bounds for $7.361 \leq \alpha < +\infty$ as well as $-3.453 < \pi < +\infty$.

Fig. 7: Nonlinear Components with and without Bounds: Prostate Cancer Incidence



Notes: Panels **(a)**, **(b)**, and **(c)** show the identifiable nonlinearities, with the horizontal dashed lines denoting the overall (or grand) mean in the data. Panels **(d)**, **(e)**, and **(f)** show the overall temporal trends under the assumption that the overall age trend is monotonically increasing from ages 40 to 85 and the overall period trend is *not* monotonically decreasing. Shaded areas indicate the range of estimates and the dashed lines denote the average of the upper and lower bounds for age, period, and cohort. Dashed lines correspond to $\alpha = 8.750$, $\pi = -2.080$, and $\gamma = 5.322$.

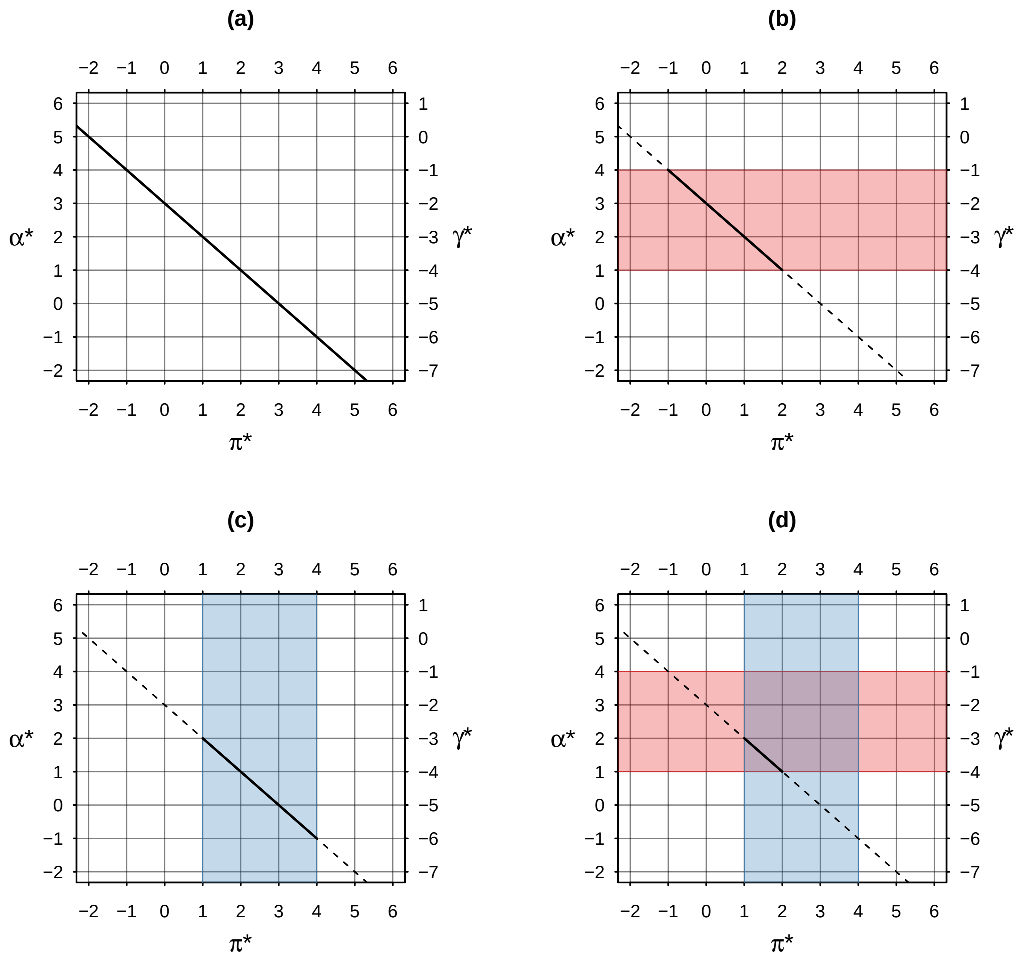
Fig. 8: Nonlinear Components with and without Bounds: Homicide Arrest Rate



Notes: Panels **(a)**, **(b)**, and **(c)** show the identifiable nonlinearities, with the horizontal dashed lines denoting the overall (or grand) mean in the data. Panels **(d)**, **(e)**, and **(f)** show the overall temporal trends under the assumption that the overall age trend conforms to an age-crime curve peaking at early adulthood. Shaded areas indicate the range of estimates and the dashed lines denote the average of the upper and lower bounds for age, period, and cohort. Dashed lines correspond to $\alpha = -1.500$, $\pi = -1.857$, and $\gamma = -1.726$. Panels **(g)**, **(h)**, and **(i)** show the overall temporal trends under the assumption of an age-crime curve as well as no monotonic decrease from the 1980-1984 to 1990-1994 periods. Dashed lines correspond to $\alpha = -2.800$, $\pi = -0.557$, and $\gamma = -3.026$.

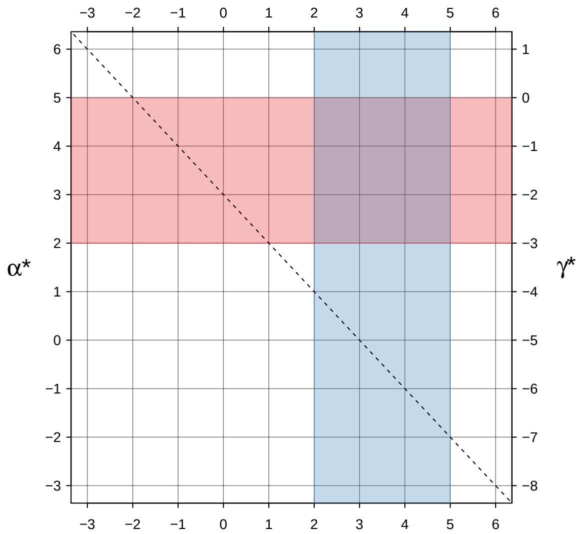
Online Supplementary Tables and Figures

Fig. 1: 2D-APC Graph:
Combining Multiple Two-Sided Bounds on APC Slopes



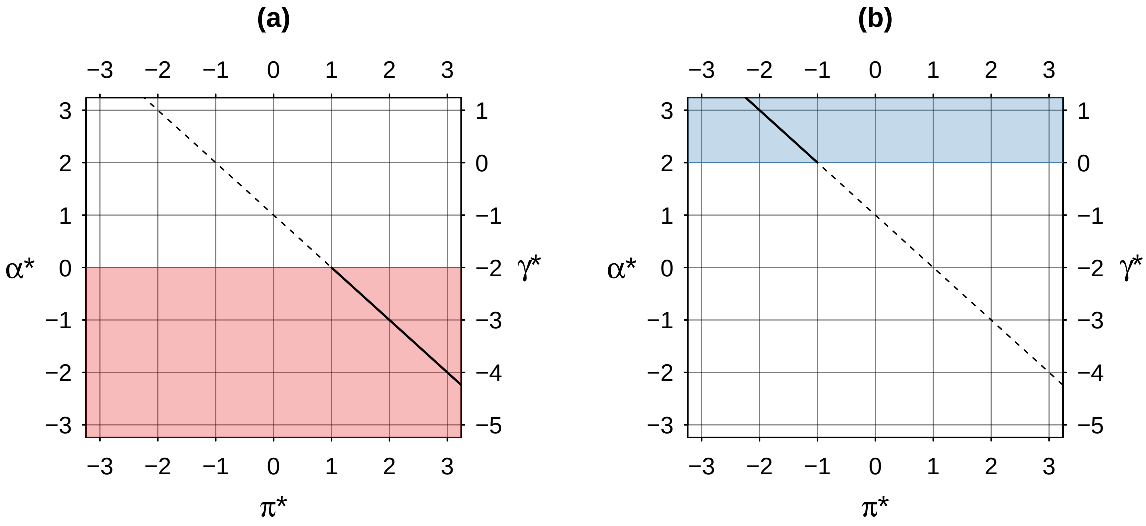
Notes: Illustration of how combining multiple two-sided bounds can result in narrower bounds for all three slopes. Panel **(a)** gives the 2D-APC graph of the solution line for simulated data with $\theta_1 = 3$ and $\theta_2 = -2$. Panel **(b)** shows the bounds for $-4 \leq \gamma \leq 1$ while panel **(c)** shows the bounds for $1 \leq \pi \leq 4$. The combination of these of these bounds is shown in panel **(d)**, which results in overall bounds of $1 \leq \alpha \leq 2$, and $1 \leq \pi \leq 2$, and $-4 \leq \gamma \leq -3$.

Fig. 2: 2D-APC Graph:
Ruling Out Claims with Multiple Two-Sided Bounds



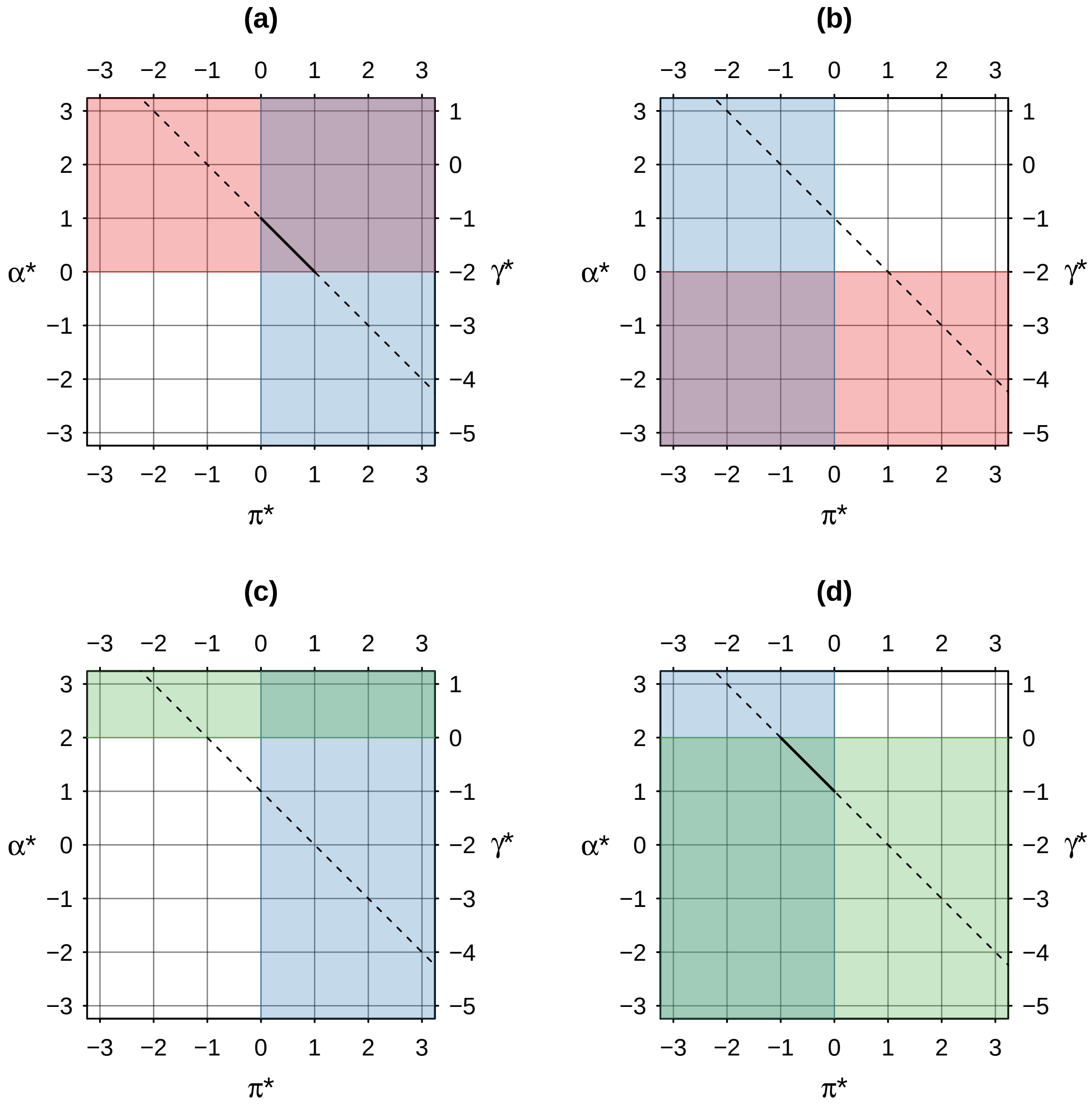
Notes: Illustration of how combining multiple two-sided bounds can rule out particular theoretical claims. Posited bounds of $2 \leq \alpha \leq 5$ and $2 \leq \pi \leq 5$ do not include the solution line and thus can be ruled as invalid. Based on simulated data with $\theta_1 = 3$ and $\theta_2 = -2$.

Fig. 3: 2D-APC Graph:
One-Sided Bounds on APC Slopes



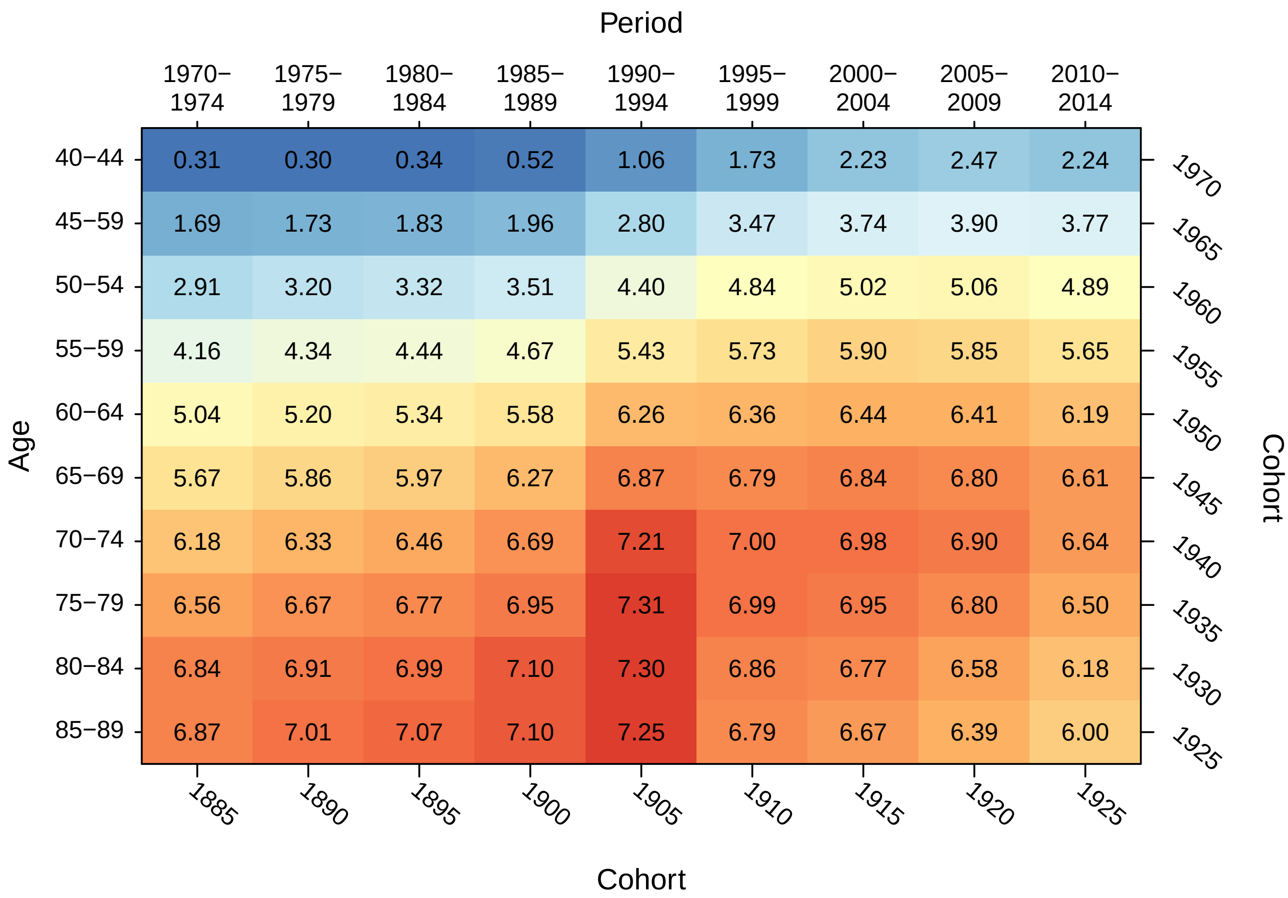
Notes: Examples of one-sided bounds. Panel **(a)** visualizes the assumption that age is non-positive (i.e., $-\infty < \alpha \leq 0$), while panel **(b)** visualizes the assumption that cohort is non-negative (i.e., $0 \leq \gamma < +\infty$). Based on simulated data with $\theta_1 = 1$ and $\theta_2 = -1$.

Fig. 4: 2D-APC Graph:
Multiple One-Sided Bounds on APC Slopes



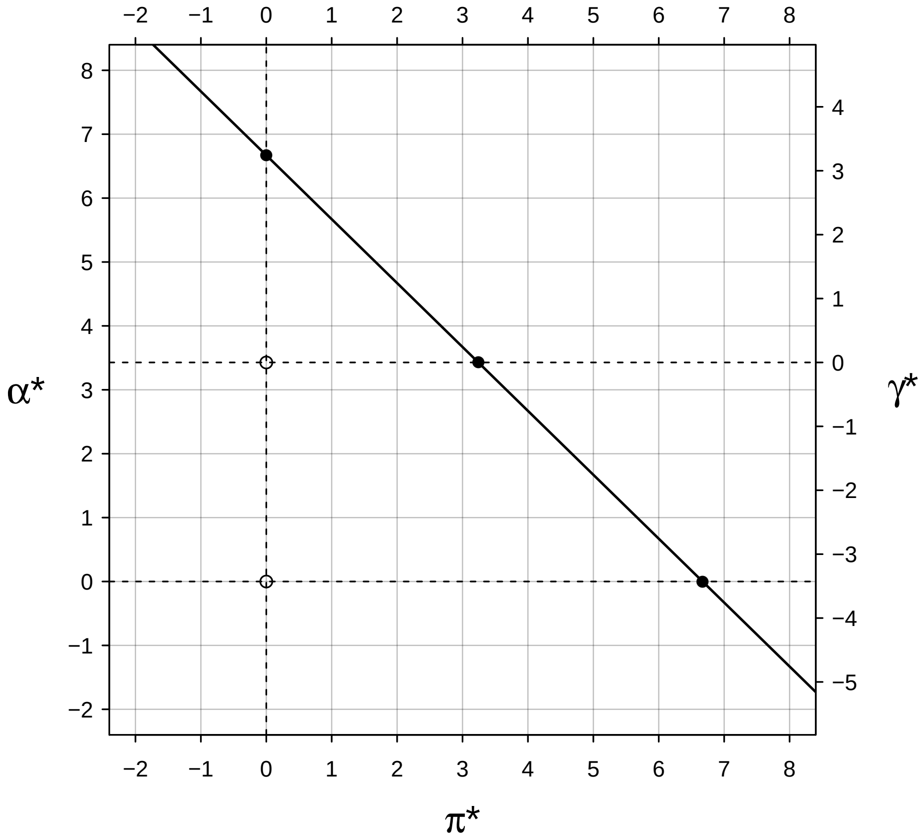
Notes: Illustration of obtaining finite overall bounds as well as ruling out particular theoretical claims by combining multiple one-sided bounds. Panel **(a)** shows finite bounds are obtained by assuming $0 \leq \alpha < +\infty$ and $0 \leq \pi < +\infty$. Panel **(b)** demonstrates that assuming $-\infty < \alpha \leq 0$ and $-\infty < \pi \leq 0$ is inconsistent with the data, and thus can be ruled out. Similarly, panel **(c)** shows that assuming $0 \leq \pi < +\infty$ as well as $0 \leq \pi < +\infty$ is inconsistent with the data, and accordingly can also be ruled out. Panel **(d)** shows finite bounds are obtained by assuming $-\infty < \pi \leq 0$ and $-\infty < \gamma \leq 0$. Based on simulated data with $\theta_1 = 1$ and $\theta_2 = -1$.

Fig. 5: Age-Period Array of Logged Prostate Cancer Rates in the United States



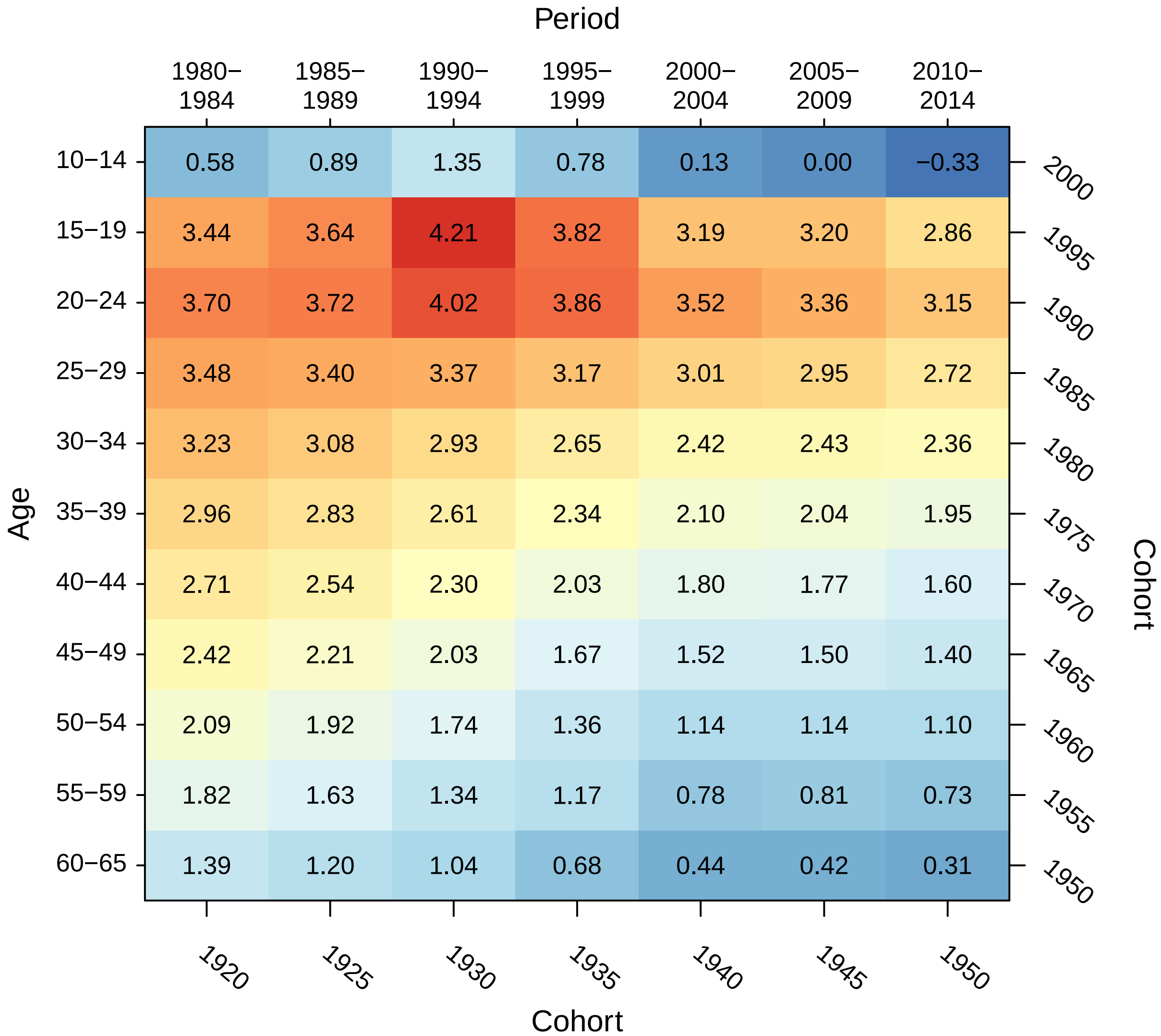
Notes: Each cell shows the logged cancer diagnoses per 100,000 men in the United States. Cohort axes indicate midpoint birth year for each age-period cell.

Fig. 6: 2D-APC Graph of the Solution Line: Logged Prostate Cancer Rates



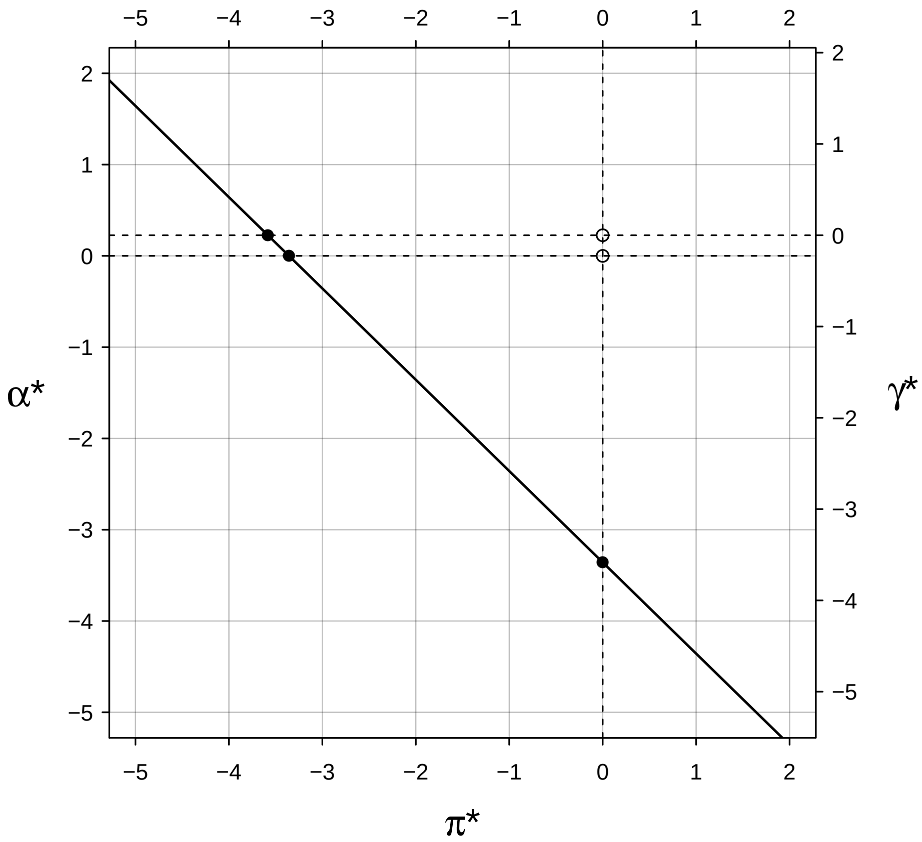
Notes: Solid line gives the solution line for the age, period, and cohort slopes. Dotted lines indicate values in the parameter space with α^*, π^*, or γ^* set to zero. Solid circles indicate points on the solution line in which the age, period, or cohort slopes are set to zero. Hollow circles denote age-period and cohort-period origins, respectively.

Fig. 7: Age-Period Array of Logged Homicide Rates in the United States



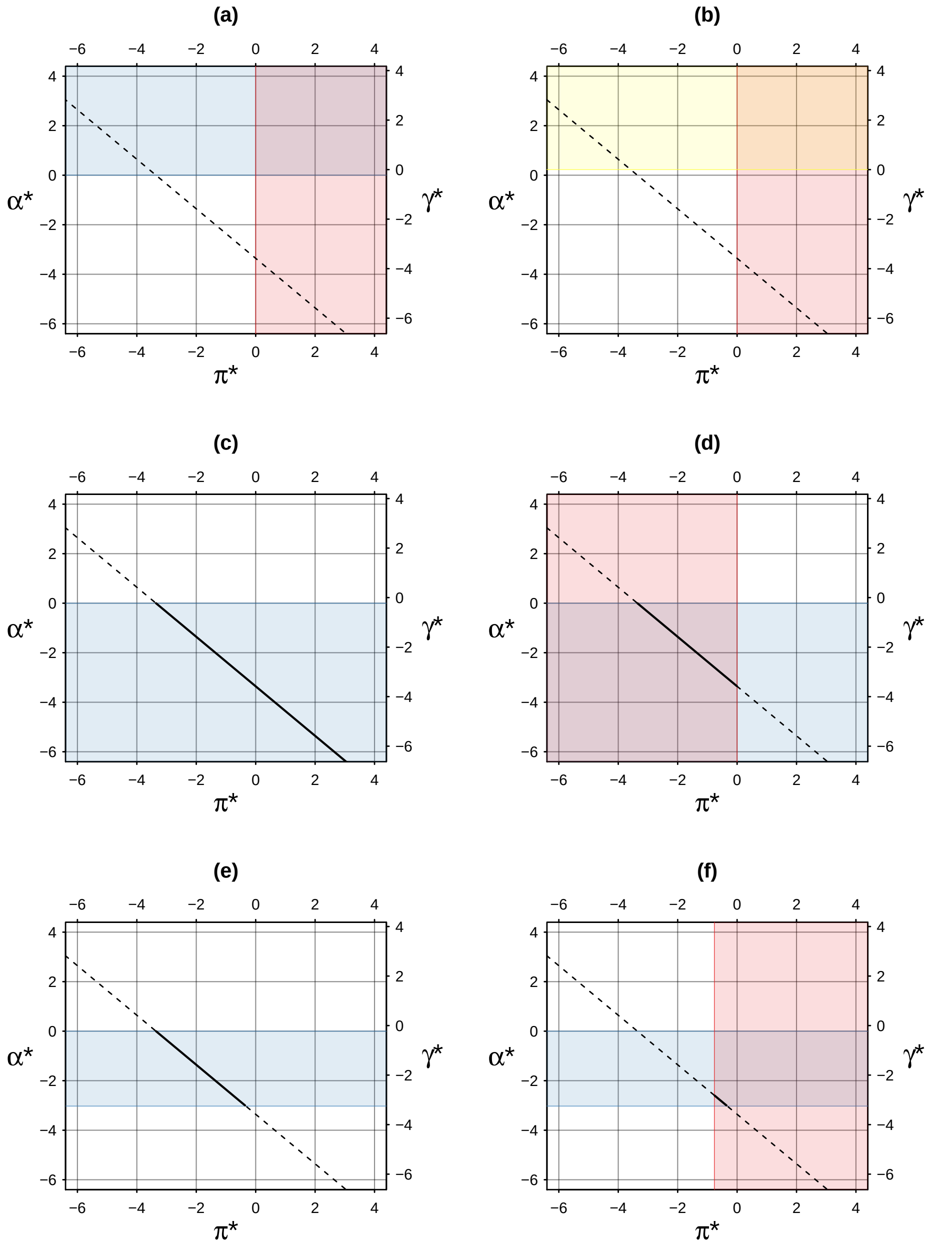
Notes: Each cell shows the logged number of homicide arrest per 100,000 men in the United States. Cohort axes indicate midpoint birth year for each age-period cell.

Fig. 8: 2D-APC Graph of the Solution Line: Logged Homicide Rates



Notes: Solid line gives the solution line for the age, period, and cohort slopes. Dotted lines indicate values in the parameter space with α^*, π^*, or γ^* set to zero. Solid circles indicate points on the solution line in which the age, period, or cohort slopes are set to zero. Hollow circles denote age-period and cohort-period origins, respectively.

Fig. 9: 2D-APC Graphs with Bounds
on Age and Period Linear Components: Homicide Arrest Rate



Notes: Panels **(a)** and **(b)** show that, based on the data alone, we can reject claims that the slopes for both age and period are positive or both period and cohort are positive. Panel **(c)** gives the bounds for a non-positive age slope, while panel **(d)** shows the bounds with the additional assumption that the period slope is non-positive. Panel **(e)** illustrates the bounds under the assume of an age-crime curve, while panel **(f)** incorporates the additional assumption of no monotonic decrease from the 1980s to mid-1990s.

References

Alwin, D. F. (1991). Family of origin and cohort differences in verbal ability. *American Sociological Review*, *56*(5), 625–638.

Ames, B. N., Shigenaga, M. K., & Hagen, T. M. (1993). Oxidants, antioxidants, and the degenerative diseases of aging. *Proceedings of the National Academy of Sciences*, *90*(17), 7915–7922. pmid: 8367443

Andvord, K. F. (1930). Hvad kan vi lære ved å følge tuberkulosens gang fra generasjon til generasjon (What can we learn by following the development of tuberculosis from one generation to another)? *Norsk Magazin for Lægevidenskapen*, *91*, 642–660. Reprinted in *International Journal of Tuberculosis and Lung Disease*, 2002, 6(7): 562–568.

Bartels, L. M. & Jackman, S. (2014). A generational model of political learning. *Electoral Studies*, *33*, 7–18.

Bell, A. & Jones, K. (2014a). Another 'futile quest'? A simulation study of Yang and Land's Hierarchical Age-Period-Cohort model. *Demographic Research*, *30*(11), 333–360.

Bell, A. & Jones, K. (2014b). Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser, and Yang's (2009) age-period-cohort study of obesity. *Social Science & Medicine*, *101*, 176–180.

Bell, A. & Jones, K. (2015a). Age, period and cohort processes in longitudinal and life course analysis: A multilevel perspective. In *A life course perspective on health trajectories and transitions* (4, pp. 197–213). Life Course Research and Social Policies. Springer International Publishing.

Bell, A. & Jones, K. (2015b). Bayesian informative prios with Yang and Land"s age-period-cohort model. *Quality & Quantity*, *49*(1), 255–266.

Burns, A. F. & Mitchell, W. C. (1946). *Measuring business cycles*. New York: National Bureau of Economic Research.

Case, R. A. M. (1956). Cohort analysis of mortality rates as an historical or narrative technique. *British Journal of Preventive & Social Medicine*, *10*(4), 159–171.

Chaves, M. (1989). Secularization and religious revival: Evidence from U.S. church attendance rates, 1972-1986. *Journal for the Scientific Study of Religion*, *28*(4), 464–477.

Chen, X., Li, G., Unger, J. B., Liu, X., & Johnson, C. A. (2003). Secular trends in adolescent never smoking from 1990 to 1999 in California: An age-period-cohort analysis. *American Journal of Public Health*, *93*(12), 2099–2104.

Clark, A. K. & Eisenstein, M. A. (2013). Interpersonal trust: An age–period–cohort analysis revisited. *Social Science Research*, *42*(2), 361–375.

Clayton, D. & Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine*, *6*(4), 469–481.

Costa, P. T. & McCrae, R. R. (1982). An approach to the attribution of aging, period, and cohort effects. *Psychological Bulletin*, *92*(1), 238–250.

Demartini, J. R. (1985). Change agents and generational relationships: A reevaluation of Mannheim's problem of generations. *Social Forces*, *64*(1), 1–16.

Derrick, V. P. A. (1927). Observations on (1) errors of age in the population statistics of England and Wales, and (2) the changes in mortality indicated by the national records. *Journal of the Institute of Actuaries*, *58*, 117–159.

Diouf, I., Charles, M. A., Ducimetière, P., Basdevant, A., Eschwege, E., & Heude, B. (2010). Evolution of obesity prevalence in France: An age-period-cohort analysis. *Epidemiology*, *21*(3), 360.

Farrington, D. P. (1986). Age and crime. *Crime and Justice: An Annual Review of Research*, *7*, 189–250.

Fienberg, S. E. (2013). Cohort analysis' unholy quest: A discussion. *Demography*, *50*(6), 1981–1984.

Fienberg, S. E., Hodges, J. S., & Luo, L. (2015). Letter to the editor. *Journal of the American Statistical Association*, *110*(509), 457–457.

Fienberg, S. E. & Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data. *Sociological Methodology*, *10*, 1–67.

Firebaugh, G. (1989). Methods for estimating cohort replacement effects. *Sociological Methodology*, *19*, 243–262.

Firebaugh, G. & Harley, B. (1991). Trends in U.S. church attendance: Secularization and revival or merely lifecycle effects? *Journal for the Scientific Study of Religion*, *30*(4), 487–500.

Frost, W. H. (1939). The age selection of mortality from tuberculosis in successive decades. *American Journal of Hygiene*, *30*, 90–06. Reprinted in *American Journal of Epidemiology*, 1995, 41: 4–9.

Fu, Q. & Land, K. C. (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), 901–921.

Fu, W. J. (2000). Ridge estimator in singular design with application to age-period-cohort analysis of disease rates. *Communications in Statistics - Theory and Methods*, *29*(2), 263–278.

Fu, W. J. (2008). A smoothing cohort model in age-period-cohort analysis with applications to homicide arrest rates and lung cancer mortality rates. *Sociological Methods & Research*, *36*(3), 327–361.

Fu, W. J., Land, K. C., & Yang, Y. (2011). On the intrinsic estimator and constrained estimators in age-period-cohort models. *Sociological Methods & Research*, *40*(3), 453–466.

Ghitza, Y. & Gelman, A. (2014). The great society, Reagan's revolution, and generations of presidential voting. Unpublished manuscript, Columbia University, New York, NY.

Glenn, N. D. (1976). Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. *American Sociological Review*, *41*(5), 900–904.

Glenn, N. D. (1981). The utility and logic of cohort analysis. *The Journal of Applied Behavioral Science*, *17*(2), 247–257.

Glenn, N. D. (2005). *Cohort analysis*. Thousand Oaks, CA: Sage.

Harman, D. (1956). Aging: A theory based on free radical and radiation chemistry. *Journal of Gerontology*, *11*(3), 298–300.

Hauser, R. M. & Huang, M.-H. (1997). Verbal ability and socioeconomic success: A trend analysis. *Social Science Research*, *26*(3), 331–376.

Hirschi, T. & Gottfredson, M. (1983). Age and the Explanation of Crime. *American Journal of Sociology*, *89*(3), 552–584.

Hobcraft, J., Menken, J., & Preston, S. (1982). Age, period, and cohort in demography: A review. *Population Index*, *48*(1), 4–43.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, *39*(2), 311–324.

Holford, T. R. (1985, January 1). An alternative approach to statistical age-period-cohort analysis. *Journal of Chronic Diseases*, *38*(10), 831–836.

Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine*, *25*(6), 977–993.

Hout, M. & Knoke, D. (1975). Change in voting turnout, 1952–1972. *Public Opinion Quarterly*, *39*(1), 52–68.

Inglehart, R. (1971). The silent revolution in Europe: Intergenerational change in post-industrial societies. *American Political Science Review*, *65*, 991–1017.

James, I. R. & Segal, M. R. (1982). On a method of mortality analysis incorporating age-year interaction, with application to prostate cancer mortality. *Biometrics*, *38*(2), 433. JSTOR: 2530456?origin=crossref

Keiding, N. (2011). Age-period-cohort analysis in the 1879s: Diagrams, stereograms, and the basic differential Equation. *Canadian Journal of Statistics*, 405–420.

Kermack, W. O., McKendrick, A. G., & McKinlay, P. L. (1934). Death-rates in Great Britain and Sweden: Expression of specific mortality rates as products of two factors, and some consequences thereof. *Journal of Hygiene*, *34*, 433–457.

Kerr, W. C., Greenfield, T. K., Bond, J., Ye, Y., & Rehm, J. (2004). Age, period and cohort influences on beer, wine and spirits consumption trends in the U.S. National Alcohol Surveys. *Addiction*, *99*(9), 1111–1120.

Kupper, L. L., Janis, J. M., Karmous, A., & Greenberg, B. G. (1985). Statistical age-period-cohort analysis: A review and critique. *Journal of Chronic Diseases*, *38*(10), 811–830.

Lavori, P. W., Klerman, G. L., Keller, M. B., Reich, T., Rice, J., & Endicott, J. (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), 23–35.

Lee, H. A. & Park, H. (2012). Trends in ischemic heart disease mortality in Korea, 1985-2009: An age-period-cohort analysis. *Journal of Preventive Medicine & Public Health*, *45*(5), 323–328.

Lewinsohn, P. M., Rohde, P., Seeley, J. R., & Fischer, S. A. (1993). Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology*, *102*(1), 110–120.

Liu, S., Semenciw, R., Ugnat, A. M., & Mao, Y. (2001). Increasing thyroid cancer incidence in Canada, 1970–1996: Time trends and age-period-cohort effects. *British Journal of Cancer*, *85*(9), 1335.

Loeber, R. & Farrington, D. P. (2014). Age-crime curve. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of Criminology and Criminal Justice* (pp. 12–18). Springer New York.

Luo, L. (2013). Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography*, *50*(6), 1945–1967.

Luo, L., Hodges, J., Winship, C., & Powers, D. (2016). The sensitivity of the intrinsic estimator to coding schemes: Comment on Yang, Schulhofer-Wohl, Fu, and Land. *American Journal of Sociology*, *122*(3), 930–961.

Mannheim, K. (1952). The sociological problem of generations. In P. Kecskemeti (Ed.), *Essays on the sociology of knowledge* (pp. 276–322). Routledge & Kegan Paul London.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, *80*(2), 319–323.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Manski, C. F. (2003). Identification problems in the social sciences and everyday life. *Southern Economic Journal*, *70*(1), 11–21.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review*, *38*(2), 242–258.

Mason, W. M. & Fienberg, S. E. (Eds.). (1985). *Cohort analysis in social research*. New York: Springer.

Mason, W. M. & Smith, H. L. (1985). Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis. In *Cohort analysis in social research* (pp. 151–227). New York: Springer.

Morgan, S. L. & Winship, C. (2014). *Counterfactuals and causal inference*. Cambridge, UK: Cambridge University Press.

O'Brien, R. (2015). *Age-period-cohort models: Approaches and analyses with aggregate data*. Boca Raton, FL: CRC Press.

O'Brien, R. M. (2011). Constrained estimators and age-period-cohort models. *Sociological Methods & Research*, *40*(3), 419–452.

O'Brien, R. M. (2011, October). The age-period-cohort conundrum as two fundamental problems. *Quality and Quantity*, *45*(6), 1429–1444.

O'Brien, R. M. (2015). Age–period–cohort models and the perpendicular solution. *Epidemiologic Methods*, *4*(1), 87–99.

O'Malley, P. M., Bachman, J. G., & Johnston, L. D. (1984). Period, age, and cohort effects on substance use among American youth, 1976-82. *American Journal of Public Health*, *74*(7), 682–688.

Palmore, E. (1978). When can age, period, and cohort be separated? *Social Forces*, *57*(1), 282–295.

Pelzer, B., te Grotenhuis, M., Eisinga, R., & Schmidt-Catran, A. W. (2014). The non-uniqueness property of the intrinsic estimator in APC models. *Demography*, *52*(1), 315–327.

Pilcher, J. (1994). Mannheim's sociology of generations: An undervalued legacy. *The British Journal of Sociology*, *45*(3), 481–495. JSTOR: 591659

Powers, D. A. (2013). Black–white differences in maternal age, maternal birth cohort, and period effects on infant mortality in the U.S. (1983–2002). *Social Science Research*, *42*(4), 1033–1045.

Pryor, W. A. (1982). Free radical biology: Xenobiotics, cancer, and aging. *Annals of the New York Academy of Sciences*, *393*(1), 1–22.

Putnam, R. D. (1995). Tuning in, tuning out: The strange disappearance of social capital in America. *PS: Political Science & Politics*, *28*, 664–683.

Reither, E. N., Hauser, R. M., & Yang, Y. (2009). Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Social Science & Medicine*, *69*(10), 1439–1448.

Robinson, R. V. & Jackson, E. F. (2001). Is trust in others declining in America? An age–period–cohort analysis. *Social Science Research*, *30*(1), 117–145.

Rodgers, W. L. (1982a). Estimable functions of age, period, and cohort effects. *American Sociological Review*, *47*(6), 774–787.

Rodgers, W. L. (1982b). Reply to comment by Smith, Mason, and Fienberg. *American Sociological Review*, *47*(6), 793–796.

Ryder, N. B. (1965). The cohort as a concept in the study of social change. *American Sociological Review*, *30*(6), 843–861.

Schaie, K. W. (1986). Beyond calendar definitions of age, time, and cohort: The general developmental model revisited. *Developmental Review*, *6*(3), 252–277.

Schwadel, P. & Stout, M. (2012). Age, period and cohort effects on social capital. *Social Forces*, *91*(1), 233–252.

Siegel, R. L., Fedewa, S. A., Anderson, W. F., Miller, K. D., Ma, J., Rosenberg, P. S., & Jemal, A. (2017). Colorectal cancer incidence patterns in the United States, 1974–2013. *JNCI: Journal of the National Cancer Institute*, *109*(8).

Simirenko, A. (1966). Mannheim's generational analysis and acculturation. *The British Journal of Sociology*, *17*(3), 292–299.

Smith, H. L., Mason, W. M., & Fienberg, S. E. (1982). Estimable functions of age, period, and cohort effects: More chimeras of the age-period-Cohort accounting framework: Comment on Rodgers. *American Sociological Review*, *47*(6), 787–793.

Tilley, J. & Evans, G. (2014). Ageing and generational effects on vote choice: Combining cross-sectional and panel data to estimate APC effects. *Electoral Studies*, *33*, 19–27.

Vedøy, T. 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*, 35–47.

Wilson, J. A. & Gove, W. R. (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), 287–302.

Yang, Y. (2008). Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review*, *73*(2), 204–226.

Yang, Y. C. & Land, K. C. (2013a). Misunderstandings, mischaracterizations, and the problematic choice of a specific instance in which the IE should never be applied. *Demography*, *50*(6), 1969–1971.

Yang, Y., Fu, W. J., & Land, K. C. (2004). A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. *Sociological Methodology*, *34*(1), 75–110.

Yang, Y. & Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology*, *36*(1), 75–97.

Yang, Y. & Land, K. C. (2013b). *Age-period-cohort analysis: New models, methods, and empirical applications*. Boca Raton, FL: CRC Press.

Yang, Y., Schulhofer-Wohl, S., Fu, W., & Land, K. (2008, May 1). The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It. *American Journal of Sociology*, *113*(6), 1697–1736.