

Bounding Analyses of Age-Period-Cohort Effects

Ethan Fosse¹ · Christopher Winship²



Published online: 28 August 2019
© Population Association of America 2019

Abstract

For more than a century, researchers from a wide range of disciplines have sought to estimate the unique contributions of age, period, and cohort (APC) effects on a variety of outcomes. A key obstacle to these efforts is the linear dependence among the three time scales. Various methods have been proposed to address this issue, but they have suffered 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 effects in population-level data. Fundamental to our framework is the recognition that it is only the slopes of an APC model that are unidentified, not the nonlinearities or particular combinations of the linear effects. 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 effects in prostate cancer incidence and homicide rates. We conclude with a discussion of guidelines for further research on APC effects.

Keywords Age-period-cohort (APC) models · Identification problem · Cohort analysis · Causal inference · Bounding analysis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13524-019-00801-6>) contains supplementary material, which is available to authorized users.

✉ Ethan Fosse
ethan.fosse@utoronto.ca

¹ Department of Sociology, University of Toronto, Toronto, ON M5S 2J4, Canada

² Department of Sociology, Harvard University, Cambridge, MA 02138, USA

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) effects¹ on various outcomes.² The core idea is that any temporal change can be attributed to three kinds of processes (Glenn 2005; Yang and Land 2013): (1) changes over the life course of individuals, or *age effects*; (2) changes due to events in the year the outcome has been observed, or *period effects*; and (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), the slopes³ of an APC model cannot be uniquely estimated because of the linear dependency among the age, period, and cohort variables. Researchers have used a variety of techniques in their attempts to overcome the APC identification problem, with recent attempts based on constraining the temporal effects through hierarchical (or multilevel) models or variations of the Moore-Penrose generalized inverse (Fu 2000, 2016; Fu and Land 2015; Fu et al. 2011; Knight and Fu 2000; Land et al. 2016; Yang 2008; Yang and Land 2006, 2013; Yang et al. 2004).

In this article, we propose an alternative approach based on placing bounds on one or more of the temporal effects from an APC model. Our approach has important similarities to Charles Manski's work 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 effects is a classic identification problem because it cannot be resolved by gathering larger samples (Manski 2003). Our approach begins by examining what the data alone, with as few constraints as possible, can tell us. Then, given that full identification of the parameters from an APC model is not possible, we proceed with the goal of partially identifying the APC effects using explicit theoretical considerations that are based on the expected size, direction, or overall shape of one or more of the temporal effects.

The rest of this article is organized as follows. First, we offer a brief history of APC analysis, summarizing recent approaches to the APC identification 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

¹ Following the convention in the APC literature, we use the shorthand of "effects" when referring to age, period, and cohort processes (e.g., Fienberg and Mason 1979; Glenn 1981; Mason et al. 1973; O'Brien 2015a; Yang and Land 2013a:). We discuss the issue of interpreting the coefficients from an APC model in the online appendix.

² For example, researchers have examined verbal ability (Alwin 1991; Hauser and Huang 1997; Wilson and Gove 1999; Yang and Land 2006), social trust (Clark and Eisenstein 2013; Putnam 1995; Robinson and Jackson 2001; Schwadel and Stout 2012), 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).

³ Researchers have used various terms in the literature to refer to the linear and nonlinear effects of an APC model. In this article, we refer to the linear effects as "slopes" or "linear effects"; conversely, we refer to the nonlinear effects as "nonlinearities" or "deviations." By "effects" or "overall effects," we refer to the combination of the linear and nonlinear effects.

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 effects 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 effects 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 effects in prostate cancer incidence and homicide rates. Finally, we conclude with a summary of our framework and an outline for further research on APC effects.

Background

The literature on APC effects has interdisciplinary origins, with seminal works arising in epidemiology and medicine in the 1930s (e.g., Frost 1939/1995) and later in sociology and demography in the 1950s and 1960s (Mannheim 1952; Ryder 1965).⁴ Since these early works, a dizzying array of methods have been proposed to address the APC identification problem (for reviews of the literature, see O'Brien 2015a; Yang and Land 2013). In sociology and demography, these proposed solutions have appeared in two distinct waves of articles and books. The first set of solutions to the identification problem, developed in the 1970s and 1980s (see Mason and Fienberg 1985), consists 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 2013). Among the most widely used of these techniques 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 (Fienberg and Mason 1979; Mason et al. 1973; Mason and Fienberg 1985). As some researchers have noted (Glenn 1976; Palmore 1978; Rodgers 1982a, b), 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, and cohort slopes (O'Brien 2015a).⁵ 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 and often atheoretical ways.⁶ Furthermore, researchers may be misled by failing to recognize that models with differing equality constraints have identical fit statistics (O'Brien 2015a; Yang and Land 2013).

⁴ However, APC analysis arguably dates back to at least the 1860s, predating the eponymous diagrams of Wilhelm Lexis (see Keiding 2011).

⁵ As Rodgers (1982a:785) 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." However, see also the reply by Smith et al. (1982) as well as the rejoinder by Rodgers (1982b).

⁶ For notable examples of researchers identifying APC effects using explicit, theoretically motivated equality constraints, see Mason and Smith (1985) and O'Brien (2015a).

Over the past two decades, a second set of models for disentangling the unique contributions of age, period, and cohort effects has emerged in sociology and demography (Fu 2000; Fu et al. 2011; Yang and Land 2006). These approaches have led to a proliferation 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 generally use either shrinkage (or regularization) of the parameters or some variant of the Moore-Penrose generalized inverse. 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 2011a, 2015b). These techniques share many of the same shortcomings as earlier approaches, as demonstrated by a spate of recent studies (e.g., see Bell and Jones 2014a, 2014b, 2015a, 2015b; Fienberg et al. 2015; Luo et al. 2016; Pelzer et al. 2014).⁷ Like earlier techniques, these methods require researchers to impose implicit constraints on the model parameters, relying on potentially strong assumptions that may fail to consistently estimate the true underlying age, period, and cohort parameters.⁸ 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; Luo 2013; Luo et al. 2016; Pelzer et al. 2014).

In short, some progress has been made since the early works on APC analysis, 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 (Bell and Jones 2015b; Luo et al. 2016; O'Brien 2015a; Rodgers 1982a). Furthermore, 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 2015a). As we show in the next section, this entails an unnecessarily strong assumption about one or more of the APC effects, given that the nonlinear components are fully identifiable.

Modeling Temporal Effects

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 et al. 1973) or *accounting model* (Mason and Fienberg 1985). Following the convention in the APC literature, we let $i = 1, \dots, I$ represent the age groups, $j = 1, \dots, J$ represent the period groups, and $k = 1, \dots, K$ represent the cohort groups with $k = j - i + I$ and $K = I + J - 1$.⁹ With this notation, we can express the C-APC as follows (Mason and Fienberg 1985):

⁷ For alternative perspectives on these techniques, see Fu (2016) and Land et al. (2016).

⁸ For example, with reference to the IE, Yang and Land (2013a:119) noted that “the objective of the IE is not to estimate the unidentifiable regression coefficient vector.” 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 age, period, and cohort effects.

⁹ 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, because this can generate artifactual cyclical patterns. For approaches to estimating temporal effects when age and period intervals are unequal, see Holford (2006).

$$Y_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k + \varepsilon_{ijk}, \tag{1}$$

where Y_{ijk} is the outcome variable to be explained, μ is the intercept, α_i represents the i th age effect, π_j represents the j th period effect, γ_k represents the k th cohort effect, and ε_{ijk} is the error term. To avoid overparameterization, we apply the so-called usual constraints that the parameters sum to 0, such that $\sum_{i=1}^I \alpha_i = \sum_{j=1}^J \pi_j = \sum_{k=1}^K \gamma_k = 0$.¹⁰

The parameterization shown in Eq. (1) is very flexible, allowing the age, period, and cohort effects to be highly nonlinear because there is one parameter for each age, period, and cohort category (Mason et al. 1973). In an age-period array, each cell is represented by a unique set of parameters (O’Brien 2011a). However, the C-APC model suffers from a fundamental identification problem due to linear dependence in the columns of the design matrix (O’Brien 2015a; Yang and Land 2013). As a result, the model in Eq. (1) cannot be estimated using conventional statistical techniques.

An alternative representation of the C-APC orthogonally decomposes the linear from the nonlinear components, clarifying the nature of the identification problem (see Holford 2006). We can specify a *linearized APC (L-APC) model* with the following form:

$$Y_{ijk} = \mu + \alpha(i-i^*) + \pi(j-j^*) + \gamma(k-k^*) + \tilde{\alpha}_i + \tilde{\pi}_j + \tilde{\gamma}_k + \varepsilon_{ijk}, \tag{2}$$

where the asterisks denote midpoint or referent indices $i^* = (I + 1 / 2)$, $j^* = (J + 1 / 2)$, and $k^* = (K + 1 / 2)$. As before, we refer to the linear effects as α for age, π for period, and γ for cohort. However, we now introduce $\tilde{\alpha}$, $\tilde{\pi}$, and $\tilde{\gamma}$ to represent age, period, and cohort nonlinearities, respectively.

The C-APC and L-APC are equivalent representations of temporal data grouped by age, period, and cohort.¹¹ As with the C-APC, each cell in an age-period array is modeled by a unique combination of parameters under the usual zero-sum constraints. For example, the i th age effect in the C-APC is represented in the L-APC by the overall linear age effect along with a unique parameter for the i th age nonlinearity: $\alpha_i = (i-i^*)\alpha + \tilde{\alpha}_i$. 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 $\tilde{\alpha}_i$, 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 L-APC model is that by explicitly separating the slopes from their deviations, it clearly shows that the identification problem is confined to the linear effects. This is reflected in the null vector of the L-APC, which consists of $\{1, -1, 1\}$ for the age, period, and cohort linear effects, respectively, and a set of 0s for the nonlinearities and the intercept.¹² Accordingly, when the nonlinear terms are 0 in the population, the L-APC is equivalent to a basic model with the linear components of age, period, and cohort as inputs:

¹⁰ Alternatively, one could fix the parameters at one of the levels to 0. By convention, researchers typically fix to 0 the first set of levels (e.g., $\alpha_{i=1} = \pi_{j=1} = \gamma_{k=1} = 0$) or the last set (e.g., $\alpha_{i=J} = \pi_{j=J} = \gamma_{k=K} = 0$), although other sets could be used.

¹¹ They are equivalent in the sense that as basis vectors they span the same space.

¹² The null vector is unique up to multiplication by a scalar.

$$\begin{aligned}
 Y_{ijk} &= \mu + \alpha(i-i^*) + \pi(j-j^*) + \gamma(k-k^*) + (0) + (0) + (0) + \varepsilon_{ijk} \\
 &= \mu + \alpha(\text{age}_i) + \pi(\text{period}_j) + \gamma(\text{cohort}_k) + \varepsilon_{ijk},
 \end{aligned} \tag{3}$$

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 $\tilde{\alpha}_i = \tilde{\pi}_j = \tilde{\gamma}_k = 0$.¹³ Because $\text{cohort}_k = \text{period}_j - \text{age}_i$, only two of the three linear components can be included in the L-APC (and, by extension, the C-APC). Thus, as Fienberg (2013:1982) stated, “[t]he APC problem is a linear effects problem.”

Despite widespread use of the APC model in the social and biological sciences, there has been little formal discussion in the literature about how to interpret the coefficients from an APC model. Some researchers appear to interpret these variables as producing causal effects (e.g., Mason and Fienberg 1985; Mason et al. 1973; Rodgers 1990), while others seem to view the variables as providing descriptive comparisons (e.g., Firebaugh 1989; Ryder 1965). Notwithstanding, when researchers talk about age, period, and cohort as having an “effect,” they are generally referring to a set of underlying causal processes that are indexed by the three APC variables (e.g., see Mason and Smith 1985). Putatively, these processes are potentially subject to manipulation and thus have well-defined causal effects in the sense of having relatively unambiguous potential outcomes or counterfactuals (Morgan and Winship 2015).¹⁴

The coefficients of an APC model can be interpreted in two main ways. First, the coefficients can be understood as representing the differences in the group means for each of the three APC variables conditional on the other two. However, because of the linear dependence, we cannot directly observe the three conditional mean differences; rather, we can observe only $\alpha + \pi = \theta_1$ and $\pi + \gamma = \theta_2$. Second, the coefficients can be viewed as representing the causal processes that generate observed mean differences across the cross-classified age, period, and cohort categories. These causal processes are not directly observed but rather are indexed by the APC variables. We provide a more extended discussion of how to interpret the coefficients in the [online appendix](#).

APC Bounding Formulas

Bounds on the APC effects can be represented algebraically or graphically. Readers who prefer a graphical representation may skip to the next section (Graphical Tools for APC Analysis). In this section, we first show how, although the true APC slopes can lie anywhere on the real number line, particular combinations of the slopes are identifiable from the data. From these interdependencies, we derive a set of formal bounds on the size and sign of the slopes. If one is willing to make an informed statement about the

¹³ A simple linear transformation can be used to convert age_i to $i - i^*$ because $i - i^* = (\text{age}_i - \text{age}^*) / (\Delta \text{age})$, 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 the fixed difference between the groups is 5. Thus, we can calculate that $\text{age}_1 = 32$ equals $(32 - 42) / 5 = -2$, which is equivalent to $i - i^* = 1 - 3 = -2$.

¹⁴ There is considerable disagreement in the social science and statistics literature on the causal status of nonmanipulable variables. As Rubin (1986) and Holland (1986) argued, such variables do not themselves have well-defined causal effects. However, in Pearl’s (2009) framework, these variables may be ascribed a causal status, with corresponding counterfactuals, even though they are not manipulable.

size and sign of at least one of the slopes from an APC model, then the data may be quite informative about the range of the remaining parameters.

Interdependencies Among the Slopes

Because 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 nonidentifiability problem is to note that for any particular APC model, we can specify the slopes as (Rodgers 1982a):

$$\begin{aligned} \alpha^* &= \alpha + \nu \\ \pi^* &= \pi - \nu \\ \gamma^* &= \gamma + \nu, \end{aligned} \tag{4}$$

where the asterisk indicates an arbitrary set of estimated slopes from an APC model under a particular constraint, and ν is a scalar fixed to some value. As Eq. (4) 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 . Because ν can take on any value from negative to positive infinity, the true slopes lie at some unknown location on the real number line.

Although individual slopes are unidentified, particular sums of the slopes are known because the value of ν cancels out. In particular, from Eq. (4), the following equations hold:

$$\begin{aligned} (\alpha + \nu) + (\pi - \nu) &= \alpha + \pi = \theta_1 \\ (\gamma + \nu) + (\pi - \nu) &= \gamma + \pi = \theta_2. \end{aligned} \tag{5}$$

Moreover, we also know the differences $\gamma - \alpha = \theta_2 - \theta_1$ and $\alpha - \gamma = \theta_1 - \theta_2$. These combinations of slopes are invariant to any given constraint defined by ν (that is, they are linearly independent) and thus can be estimated from the data. For example, consider the basic model predicting an outcome as a function of age, period, and cohort:

$$Y_{ijk} = \mu + (age_i)\alpha + (period_j)\pi + (cohort_k)\gamma + \varepsilon_{ijk}, \tag{6}$$

where, for ease of exposition, the parameters for the nonlinear terms are assumed to equal 0 in the population (e.g., see Eq. (3)). Substituting $age_i + cohort_k$ for $period_j$, we obtain the following:

$$\begin{aligned} Y_{ijk} &= \mu + age_i(\alpha + \pi) + (cohort_k)(\gamma + \pi) + \varepsilon_{ijk} \\ &= \mu + (age_i)\theta_1 + (cohort_k)\theta_2 + \varepsilon_{ijk}. \end{aligned} \tag{7}$$

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 $(\cdot)_{\min}$ and $(\cdot)_{\max}$ 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 $\theta_2 - \theta_1 = \gamma - \alpha$.

Because age and cohort are not linearly dependent in the absence of period, θ_1 and θ_2 are identified and can be estimated using Eq. (7).¹⁵

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 Eq. (4). 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. (7)), which have the following relationship to the true age, period, and cohort slopes:

$$\begin{aligned} \theta_1 &= \alpha + \nu \\ 0 &= \pi - \nu \\ \theta_2 &= \gamma + \nu. \end{aligned} \tag{8}$$

Because we have constrained the period slope to be 0 in the estimated model, we can meaningfully interpret the arbitrary constant ν in terms of upper and lower bounds on the unknown period slope. That is, if we let $\nu = \pi_{\min}$ and $\nu = \pi_{\max}$, then we have the inequalities

$$\begin{aligned} \theta_1 - \pi_{\min} &\geq \alpha & \theta_1 - \pi_{\max} &\leq \alpha \\ \pi_{\min} &\leq \pi & \text{and} & \pi_{\max} &\geq \pi \\ \theta_2 - \pi_{\min} &\geq \gamma & \theta_2 - \pi_{\max} &\leq \gamma, \end{aligned} \tag{9}$$

respectively. Given that we have set bounds on the period slope, the equations above indicate the corresponding bounds on the age and cohort slopes. In a similar way, we can construct bounding formulas for age and cohort.¹⁶ We have

¹⁵ One could also estimate the θ s using $Y_{ijk} = \mu + (\text{period}_j)(\theta_1) + (\text{cohort}_k)(\theta_2 - \theta_1) + \varepsilon_{ijk}$ or $Y_{ijk} = \mu + (\text{age}_i)(\theta_1 - \theta_2) + (\text{period}_j)(\theta_2) + \varepsilon_{ijk}$.

¹⁶ The estimates when the age slope is constrained to equal 0 are $\alpha^* = 0$, $\pi^* = \theta_1$, and $\gamma^* = \theta_2 - \theta_1$, and the corresponding estimates when the cohort slope is fixed to 0 are $\alpha^* = \theta_1 - \theta_2$, $\pi^* = \theta_2$, and $\gamma^* = 0$.

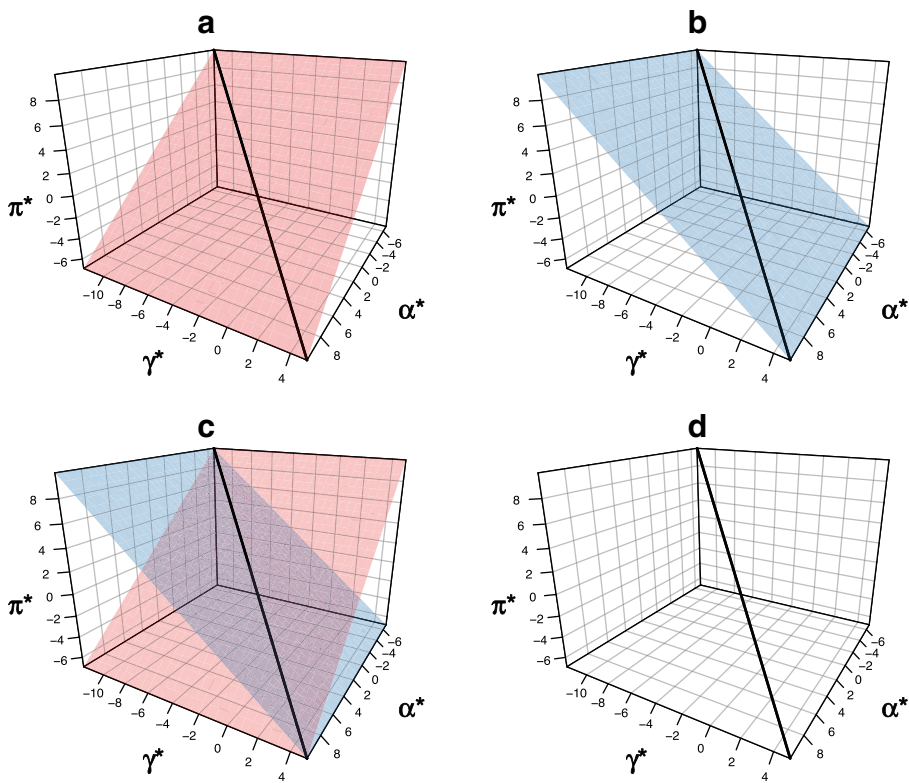


Fig. 1 Geometric derivation of the solution line. Panel a shows an age-period plane defined by $\pi = \theta_1 - \alpha$, where $\theta_1 = 1$. Panel b shows a period-cohort plane defined by $\pi = \theta_2 - \gamma$, where $\theta_2 = -1$. Panel c visualizes the intersection of the age-period and period-cohort planes, and panel d visualizes the solution line based on $\theta_1 = 1$ and $\theta_2 = -1$.

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. Because θ_1 and θ_2 are invariant to the constraint imposed on the APC model, they are identifiable and can be estimated from the data.

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 $\alpha_{\min} = 0$ and an upper bound of $\alpha_{\max} = 2$, then we can use the θ_1 and θ_2 to bound the period and cohort slopes as well. Specifically, by setting $\alpha_{\max} = 2$, we know that the lower bound for the period slope is $\theta_1 - \alpha_{\max} = \alpha + \pi - 2 = 3 - 2 = 1$. Likewise, by setting $\alpha_{\min} = 0$, we know that the upper bound for the period slope must be $\theta_1 - \alpha_{\min} = \alpha + \pi - \alpha_{\min} = 3 - 0 = 3$. Similarly, we can find the bounds for the cohort slopes. By setting $\alpha_{\min} = 0$ and $\alpha_{\max} = 2$, we know that the lower bound of the cohort slope must be $\theta_2 - \theta_1 + \alpha_{\min} = \gamma - \alpha + \alpha_{\min} = -5 + 0 = -5$, and the upper bound must be $\theta_2 - \theta_1 + \alpha_{\max} = \gamma - \alpha + 2 = -5 + 2 = -3$. Thus, by restricting the age slope to lie between 0 and 2, we can conclude from the bounding formulas that the period slope must range from 1 to 3, and the cohort slope must range from -5 to -3.

Graphical Tools for APC Analysis

Although data visualization has a long history in APC modeling (Keiding 2011), there are no widely used graphical tools for evaluating and understanding the slopes of APC models. In this section, 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 2015a).

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 $\theta_2 = -2$. Panel a of Fig. 1 depicts the age-period plane defined by $\theta_1 = 3$, while panel b depicts the period-cohort plane defined by $\theta_2 = -2$. The intersection of these two planes then defines a line, as illustrated in panels c and d. This line is known as the *solution line* because any point on the line is a potentially valid set of estimates for the unknown parameters α , π , and γ (O'Brien 2015a).¹⁷ That is, all points on the solution line give exactly the same predicted values of the outcome, 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.

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 involving only linear components, the data have taken us from a three-dimensional space in which all parameter values are possible to a one-dimensional space in which only certain combinations of estimates lying on a line are consistent with the data. This same extent of reduction holds even if our data have nonlinearities because they are fully identified. As such, the data have gone a long way toward 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 but not as informative as we might like in the sense of providing unique point estimates for the linear effects.

2D-APC Graph of the Solution Line

Because the solution line is a geometric representation of the APC identification problem, we can use panel d of Fig. 1 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, because 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

¹⁷ Figure 1 is a graphical representation of the APC identification problem. If there were no linear dependency, we would have three planes intersecting at a single point in the parameter space (O'Brien 2015a).

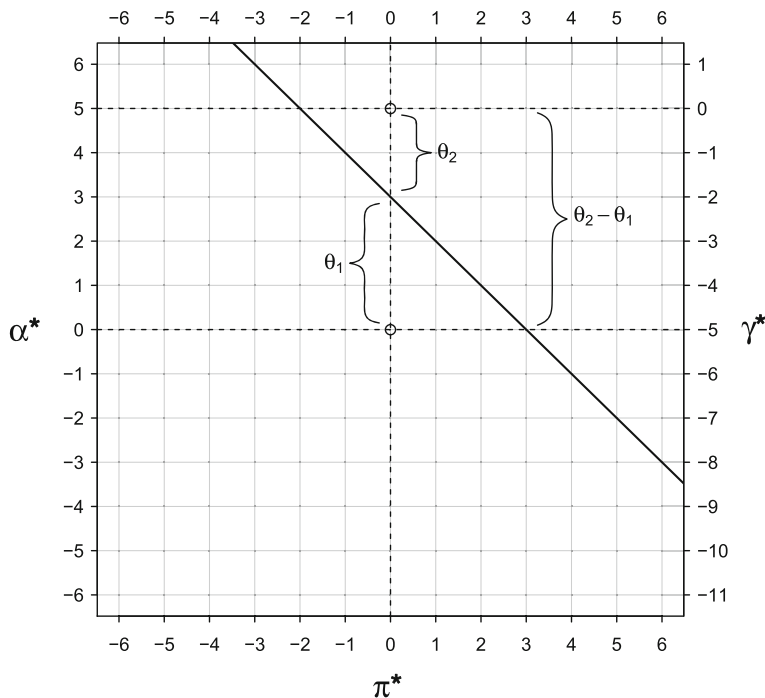


Fig. 2 2D-APC graph of the solution line. The solid black line denotes the solution line, with $\theta_1 = \alpha + \pi$, $\theta_2 = \gamma + \pi$, and $\theta_2 - \theta_1 = \gamma - \alpha$. Dashed lines refer to values in the parameter space when α^* , π^* , or γ^* are set to 0. Hollow circles denote the age-period and period-cohort origins. Simulated data based on values of $\theta_1 = 3$ and $\theta_2 = -2$.

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*.

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 panel d of Fig. 1. The left vertical axis refers to the values of the age slope; the top and bottom horizontal axes, to period; and the 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; the three dashed lines indicate when each of the age, period, and cohort slopes are set to 0, respectively. The point at which each dashed line intersects with the solution line gives the estimates for models that set the age, period, or cohort slope to 0. For example, the bottom horizontal dashed line, indicating when $\alpha = 0$, intersects the solution at (0, 3, -5). If the true slope for age were 0, then we would know that the period slope is 3 and the cohort slope is -5. The three dashed lines intersect at the origin for the age-period dimension and that for the period-cohort dimension.¹⁸ 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.

¹⁸ The age-period-cohort origin is (0, 0, 0), but this is not directly visible on the 2D-APC graph unless $\theta_2 - \theta_1 = 0$.

Importantly, Fig. 2 provides information on three key features of the data: (1) the slope of the solution line, which is always -1 ; (2) the direction and scale of the axes; and (3) the values of θ_1 , θ_2 , and their difference. For all APC data, features (1) and (2) will be the same: the slope of the solution line relating period to age and cohort will always be -1 ; accordingly, the age and cohort axes will always run in the same direction, whereas the period axis will run in the opposite direction.¹⁹ However, depending on the unknown parameters α , π , and γ , the data will generate different values of θ_1 , θ_2 , and their difference. 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.²⁰ Given that $\theta_1 = \alpha + \pi = 3$, we can write $\alpha = 3 - \pi$, which can be thought of as a linear equation relating the scale for age (α) to the scale for period (π) with an intercept of 3 and a slope of -1 . Because the slope is -1 , the age and period scales have flipped signs; and because the intercept is 3, age and period 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 0, then the value of the period (or age) slope must be $\theta_1 = 3$.

Similarly, because $\theta_2 = \pi + \gamma = -2$, 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 of -2 and a slope of -1 . Again, because the slope is negative the cohort and period scales have flipped signs and because the intercept is -2 , cohort and period are offset by two units. As shown in Fig. 2, the vertical (and horizontal) distance from the period-cohort origin $(5, 0, 0)$ to the solution line is given by $\theta_2 = -2$ because if the cohort (or period) slope is fixed to 0, 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 dashed 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 slopes for both age and cohort.

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 $\theta_2 + 2 \leq \pi \leq \theta_2 + 4$. Because we know from the data that $\theta_1 = 3$ and $\theta_2 = -2$, then we can conclude that if the cohort slope ranges from -4 to -2 , and it must also be the case that 1

¹⁹ The -1 slope relating period to age and cohort as well as the differing direction of the period axis relative to the age and cohort axes in the 2D-APC graph are reflected in the opposing sign of v in Eq. (4) for the period slope compared with that for the age and cohort slopes.

²⁰ Likewise, each plane in Fig. 1 can be thought of as a function 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 $\pi = \theta_1 - \alpha$. Similarly, the period-cohort plane can be thought of as a function of $\theta_2 = \pi + \gamma$, with $\pi = \theta_2 - \gamma$.

$\leq \alpha \leq 3$ and $0 \leq \pi \leq 2$. Because the APC identification problem can be represented both mathematically and geometrically, we can visualize the bounds in a 2D-APC graph.

Panel a of Fig. 3 shows the bounds for the cohort slope ($-4 \leq \gamma \leq -2$) in three-dimensional space. The shaded box denotes the region of the parameter space corresponding to a range of -4 to -2 . The solution line cuts through this region. Because we have ruled out cohort slopes greater than -2 and less than -4 , we have restricted the set of feasible APC slopes to those lying on the line in the feasible region (i.e., to those points lying on the solid line rather than the dashed line in panel a of Fig. 3).

However, the consequences of assuming these particular bounds for the cohort slope are not clear in panel a of Fig. 3 because of the inherent difficulties of interpreting a three-dimensional visualization on a two-dimensional surface. Panel b of Fig. 3 depicts the same information in a 2D-APC graph, which clarifies the nature of our assumptions. Restricting the cohort slope to range from -4 to -2 reduces the parameter space to the shaded rectangle. Because each of the three APC slopes must lie on the solution line, the data inform 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 1 to 3 and that the period slope must range from 0 to 2. 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. In the next section, using the bounding formulas and our 2D-APC graph, we provide a more detailed overview of various strategies that can be used to produce informative bounds from an APC model.

Bounding Strategies

The bounding formulas and 2D-APC graph provide tools for evaluating the consequences of restricting the range of one or more slopes in an APC model. In practice, a researcher may not have enough information to specify the size and sign of one or more of the slopes. Moreover, we have said little so far about the constraints involving the nonlinearities, which are identified. Other strategies are possible, some of which may entail considerably weaker

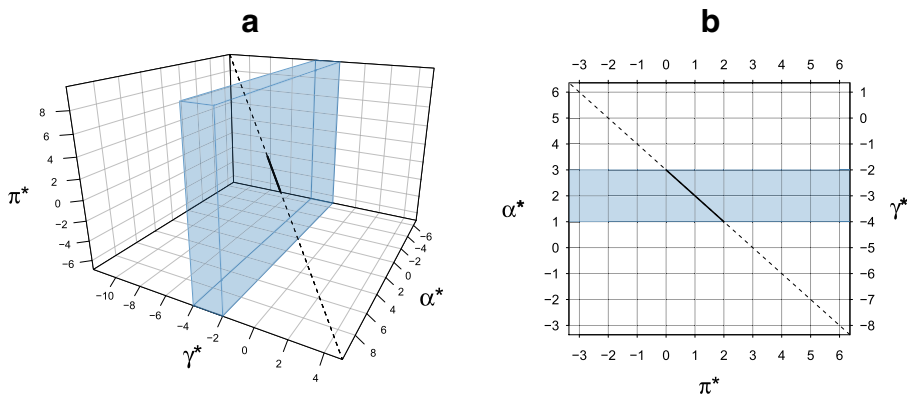


Fig. 3 Solution line with bounds on cohort slope: 3D- vs. 2D-APC graph. Panels a and b give 3D and 2D representations of specified bounds on the solution line. Shaded areas reflect the assumption that $-4 \leq \gamma \leq -2$. Simulated data based on values of $\theta_1 = 3$ and $\theta_2 = -2$.

assumptions. In this section, we discuss two such strategies: (1) restricting only the direction of one or more APC slopes; and (2) constraining the shape of one or more overall effects, which entails incorporating the nonlinear components. We review these strategies in turn.

Constraints on the Sign

In many empirical examples, a relatively weak assumption is to constrain just the direction of one or more slopes. These bounds are necessarily one-sided, with one end of the interval set at 0 and the other at either positive or negative infinity. Remarkably, by constraining the direction of one temporal slope, we can in many cases 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 only the sign of a single linear effect is considered.

To illustrate the application of bounds based only on 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 nonpositive. Using the one-sided bounding formulas in Table 2, we know that $\theta_1 \leq \pi < +\infty$ and $-\infty < \gamma \leq \theta_2 - \theta_1$. Because $\theta_2 - \theta_1 = -2$, we can conclude that the period slope is $1 \leq \pi < +\infty$ and the cohort slope is $-\infty < \gamma \leq -2$. In other words, by positing the direction of the age linear effect, we can conclude that the cohort slope must be negative. Moreover, even though we just constrained the sign of a single slope, we can make conclusions about the magnitude of the cohort slope; specifically, we know that the cohort slope must be less than or equal to -2 if the age slope is nonpositive. Likewise, we can conclude that the period slope must be greater than or equal to 1.

One drawback of setting the sign of a single slope is that the final set of bounds is not finite. However, we obtain finite bounds for the slopes only when one-sided

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 $(\cdot)_{\min}$ and $(\cdot)_{\max}$ denoting minimum and maximum values of the bounds. We donate $\theta_1 = \alpha + \pi$, $\theta_2 = \pi + \gamma$, $\theta_1 - \theta_2 = \alpha - \gamma$, and $\theta_2 - \theta_1 = \gamma - \alpha$.

intervals operate in the opposite directions. Because 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 nonnegative and that we know from our data that $\theta_1 = 1$ and $\theta_2 = -1$. By assuming the age slope is nonnegative, we can conclude that $0 \leq \alpha < +\infty$, $-\infty < \pi \leq 1$, and $-2 \leq \gamma < +\infty$. By assuming that the cohort slope is nonnegative, we know that $-2 \leq \alpha < +\infty$, $-\infty < \pi = -1$, 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 nonnegative. The formulas in Table 2 tell us that $0 \leq \alpha \leq 1$, $0 \leq \pi \leq 1$, and $-2 \leq \gamma \leq -1$. Consequently, we can conclude that if the age and period slopes are both nonnegative, then the age slope must lie between 0 and 1, the period slope must lie between 0 and 1, and the cohort slope must lie between -2 and -1 . Conversely, if we assume that the age and cohort slopes are both nonpositive, then we know that one or both of these assumptions must be incorrect: if both were true, we would obtain $\theta_1 \leq \alpha \leq 0$, which is impossible because $\theta_1 = 1$.

The key to our approach entails using the bounding formulas and a 2D-APC graph to analyze the consequences of assumptions regarding the size and sign of the slopes from an APC model. However, so far we have considered only the case in which age, period, and cohort consist solely of linear effects. This is a worst-case scenario because in reality temporal effects are composed of both linear and nonlinear components, the latter of which are fully identified and will help us narrow the bounds even further.

Shape Constraints

So far, we have considered bounding only the linear effects. However, as shown in Eq. (2), a more general APC model partitions the overall temporal effects 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 monotonically increasing (m.i.) age slope, and let $\alpha_{m.d.}$ denote a monotonically decreasing (m.d.) age slope. If we have an APC model with only linear components, the situation is simple: a monotonically increasing age effect is equivalent to a positive slope, and a monotonically decreasing effect 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., nonpositive or nonnegative), resulting in a considerably expanded set of possible bounding strategies.²¹

²¹ As a reviewer noted, some of these nonlinearities could be noise. One way to address possible noise is to smooth out the nonlinearities by setting the parameters for the higher-order nonlinearities to 0. Alternatively, one could use natural cubic splines, treating age, period, and cohort as continuous rather than categorical variables (see Heuer 1997).

One possibility is to specify that the overall age effect 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 effect 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 an effect 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).

Nonmonotonicity constraints are also possible, as in the assumption that an age effect is not monotonically increasing ($-\infty < \alpha < \alpha_{m.i.}$) or not monotonically decreasing ($\alpha_{m.d.} < \alpha < +\infty$). For example, one might assume that the probability of developing a particular cancer

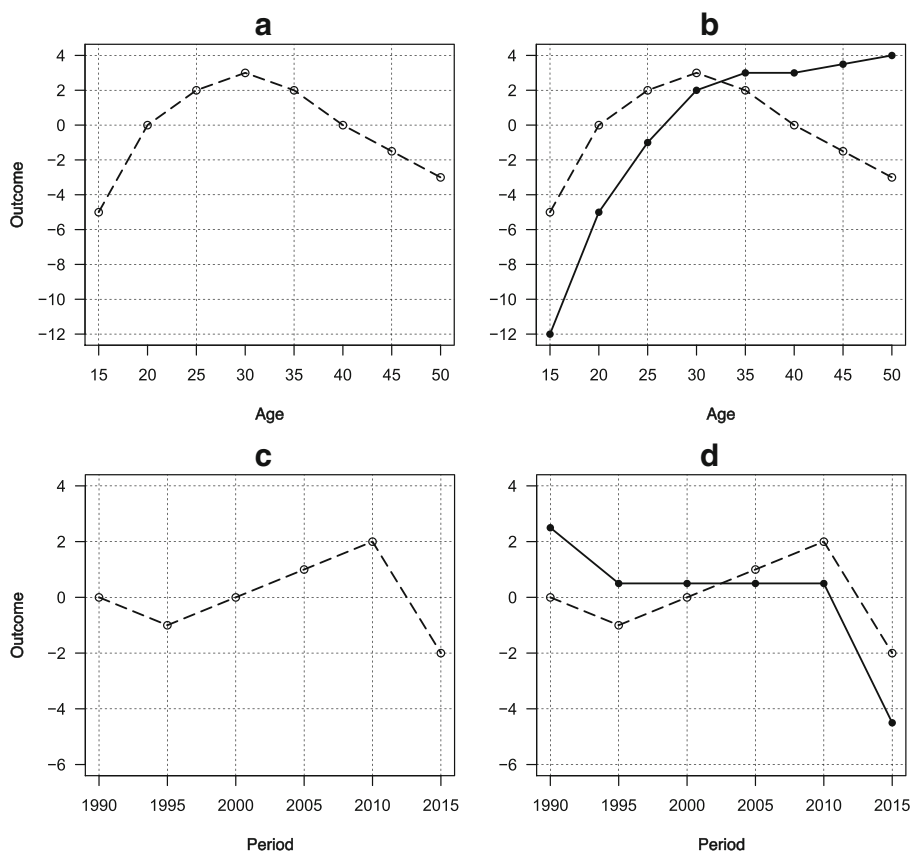


Fig. 4 Monotone constraints on nonlinear effects of age and period. The dashed lines in panels a and c show simulated nonlinearities for age and period. The solid line in panel c reflects the assumption that $\alpha = 2$, which is the minimum age slope required for a monotonically increasing overall age effect. The solid line in panel d reflects the assumption that $\pi = -1$, which is the maximum period slope required for a monotonically decreasing overall period effect.

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 effect is neither monotonically increasing nor monotonically decreasing, such that $\alpha_{m.d.} < \alpha < \alpha_{m.i.}$. For example, when examining unemployment rates, it might make sense to rule out any monotonic effect across periods because it may violate theoretical assumptions regarding a business cycle (e.g., Burns and Mitchell 1946).

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 $\tilde{\alpha} = \{-5, 0, 2, 3, 2, 0, -1.5, -3\}$ and are displayed as a dashed line in panel a of Fig. 4. Assume that we believe that the overall effect in age is monotonically increasing. As can be seen in panel a, the nonlinearities are increasing at some intervals and decreasing at other intervals, relative to an unknown linear age effect. We want to specify a value for the linear age effect that ensures that between any two adjacent age categories, the effect is flat at the minimum. To do so, we need only find that pair of neighboring age categories in which the downward effects are the most negative. This can be computed using the following equation:

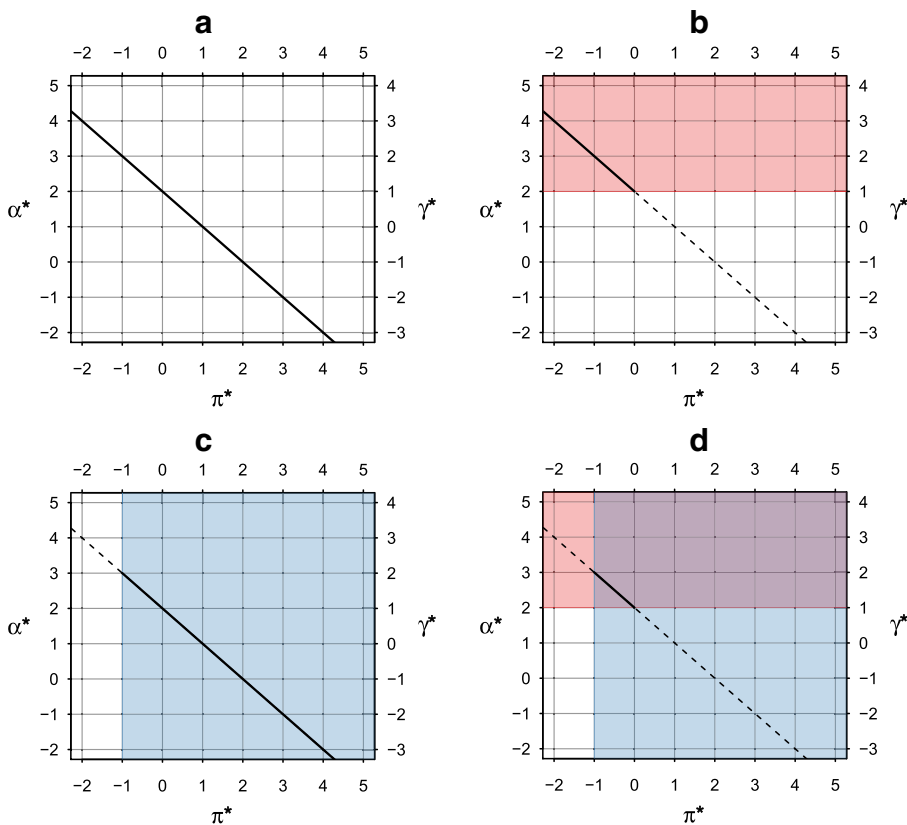


Fig. 5 2D-APC graphs with bounds on age and period linear effects. Panel a gives the 2D-APC graph of the solution line for simulated data with $\theta_1 = 2$ and $\theta_2 = 1$. Panel b shows the bounds for $2 \leq \alpha < +\infty$, reflecting the assumption that the overall age effect is monotonically increasing. Panel c shows bounds for $-1 < \pi < +\infty$, reflecting the assumption that the overall period effect is not monotonically decreasing. Panel d shows the bounds for $2 \leq \alpha < +\infty$ as well as $-1 < \pi < +\infty$.

$$\alpha_{m.i.} = \begin{cases} (-1) \times \min(\Delta\tilde{\alpha}) & \text{if } \min(\Delta\tilde{\alpha}) < 0, \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

where the Δ notation indicates forward differences (i.e., $\tilde{\alpha}_{i+1} - \tilde{\alpha}_i$), and the subscript m.i. indicates the (nonnegative) age slope required for a monotonically increasing overall effect. The forward differences of the age nonlinearities in panel a of Fig. 4 are $\Delta\tilde{\alpha}_i = \{5, 2, 1, -1, -2, -1.5, -1.5\}$. The minimum of these differences is -2 , which is the difference between ages 35 and 40. To counter this downward deviation, the parameter value for the linear age term must be greater than or equal to 2. Panel b of Fig. 4 illustrates what happens to the overall age effect when the slope is set to 2, which is shown by the solid line. The overall effect is now monotonically increasing. Any slope less than 2 will result in an overall effect that is either nonmonotonic or monotonically decreasing over the full set of age groups. Alternatively, any slope greater than 2 will be monotonically increasing; this is the *minimum* monotonically increasing age slope consistent with our data.

We can determine the implications of a monotonically increasing age effect by using our 2D-APC graph. The results for age are shown in panel b of Fig. 5. By assuming the age effect increases monotonically, we can conclude that the period slope must be negative and that 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 that the overall period effect is *not* monotonically decreasing. This is a weaker assumption than assuming the effects are monotonically increasing because we allow downward deviations as long as they are not *always* decreasing or staying the same across adjacent period groups.

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

$$\pi_{m.d.} = \begin{cases} (-1) \times \max(\Delta\tilde{\pi}) & \text{if } \max(\Delta\tilde{\pi}) > 0, \\ 0 & \text{otherwise} \end{cases}, \quad (11)$$

where the Δ notation again indicates forward differences (i.e., $\tilde{\pi}_{j+1} - \tilde{\pi}_j$), and the subscript m.d. indicates the (nonpositive) slope needed for a monotonically decreasing effect. Using Eq. 11 we can calculate the period slope corresponding to an overall monotonically decreasing effect. The forward differences for the period nonlinearities are $\Delta\tilde{\pi} = \{-1, 1, 1, 1, -4\}$, and the maximum of these differences is 1. To counter this upward deviation, the parameter value for the period slope must be less than or equal to -1 . When we set the slope to -1 , we obtain the solid line in panel d of Fig. 4, which corresponds to the maximum period slope required for a monotonically decreasing overall period effect. Any slope less than -1 will result in a steeper downward period effect.

However, in our case, we would like to assume that the overall period effect is *not* monotonically decreasing. Thus, the solid line in panel c of Fig. 4 is the most downward effect we are willing to consider; that is, we are assuming that the period slope is greater than -1 . The consequences of this assumption are shown in panel c of

Fig. 5, which restricts the age slope to be less than 3 and the cohort slope to be less than 2. 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, as shown in panel d of Fig. 4. By assuming the overall age effect is monotonically increasing and the period effect is not monotonically decreasing, we can conclude that the age slope is between 2 and 3, the period slope is between -1 and 0 , and the cohort slope is between 1 and 2.

The bounding strategies outlined here are flexible and can be combined in a multitude of ways. In the worst-case scenario in which we have no nonlinearities, we can still obtain finite bounds under certain conditions by specifying only the direction of two or more slopes. With the more realistic case of incorporating nonlinearities, we can include assumptions about the shape, size, and direction of the overall effects for specified ranges of each variable, greatly expanding the variety of bounding strategies available.

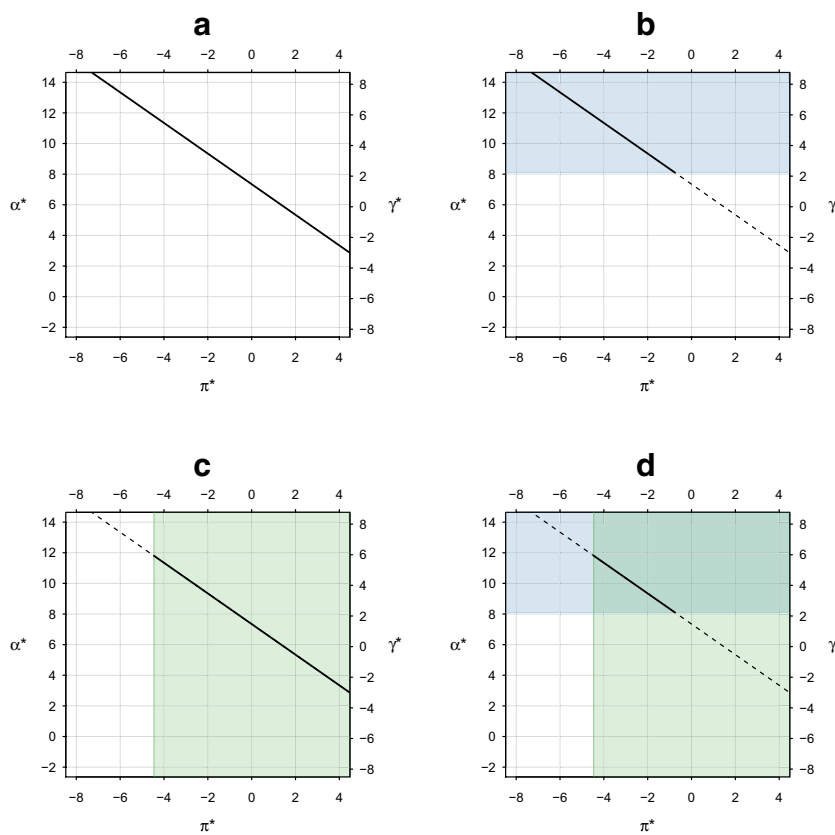


Fig. 6 2D-APC graphs with bounds on age and period linear effects: prostate cancer incidence. Panel a gives the 2D-APC graph of the solution line with $\theta_1 = 7.344$ and $\theta_2 = 1.473$. Panel b shows the bounds for $8.104 \leq \alpha < +\infty$, reflecting the assumption that the overall age effect is monotonically increasing from ages 40 to 85. Panel c shows bounds for $-4.458 \leq \pi < +\infty$, reflecting the assumption that the overall period effect is not monotonically decreasing. Panel d shows the bounds for $8.104 \leq \alpha < +\infty$ as well as $-4.458 \leq \pi < +\infty$.

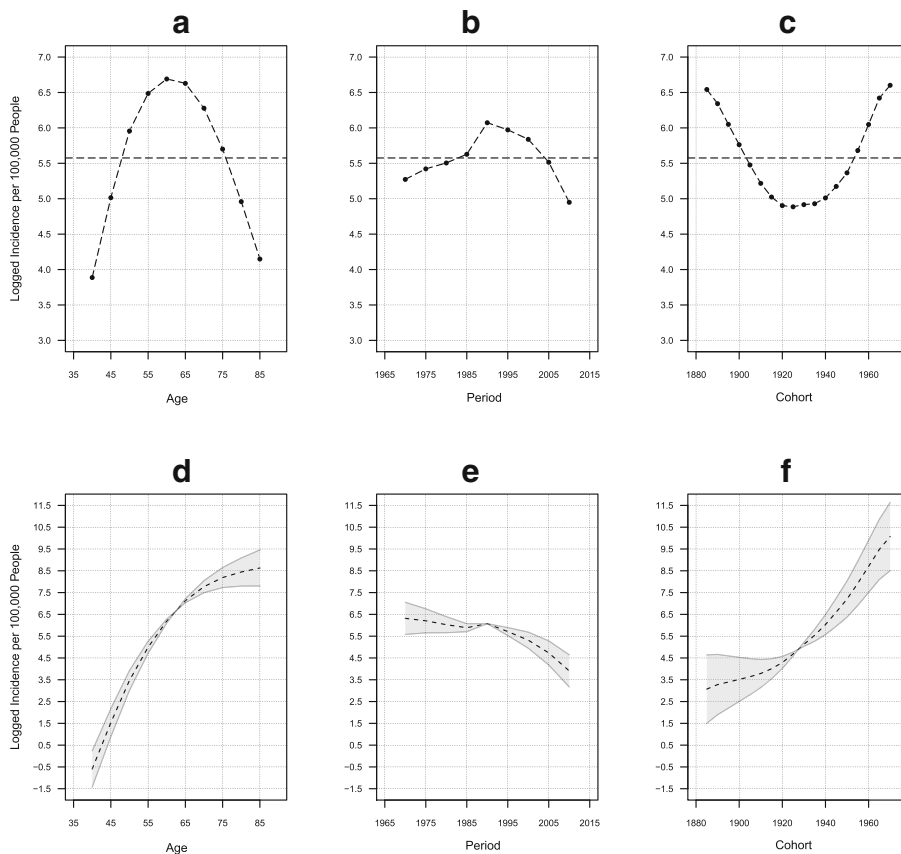


Fig. 7 Temporal effects with and without bounds: prostate cancer incidence. Panels a, b, and c show the identifiable nonlinearities, with the horizontal dashed lines denoting the overall (or grand) mean in the data. Panels d, e, and f show the overall temporal effects under the assumption that the overall age effect is monotonically increasing from ages 40 to 85 and the overall period effect is non-negative. 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 $\alpha = 9.955$, $\pi = -2.611$, and $\gamma = 4.084$.

Empirical Examples

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

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 it is generally thought to be due to an accumulation of exposure risks combined with a loss of effectiveness in cellular repair mechanisms. Untangling temporal effects 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 ages. To examine prostate cancer, 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). 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, and 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 2010–2014.²² The outcome variable is the number of malignant prostate cancer diagnoses per 100,000 men for each age-period cell. We logged the outcome, given that it is highly skewed.

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. (2)).²³ From this model, we estimate that $\theta_1 = \alpha + \pi = 7.344$ and $\theta_2 = \pi + \gamma = 1.473$.²⁴ As discussed in the previous sections, the values of θ_1 and θ_2 define a solution line, which is visualized in Fig. 6, panel a. The nonlinearities are visualized in panels a–c of Fig. 7, with the horizontal dashed lines denoting the overall (or grand) mean in the data.²⁵ The nonlinearities can be difficult to interpret because they are deviations from unknown linear effects. The easiest way to interpret these deviations is by recognizing that these are the temporal effects we would observe in the data if all the slopes were 0.²⁶

At this point, we have exhausted what can be known from the data and we must use additional constraints to split the estimates of θ_1 and θ_2 into three separate age, period, and cohort slopes. An initial theoretical assumption might be that prostate cancer incidence increases monotonically from ages 40 to 85, which results in a minimum age slope of 8.104. This putatively weak constraint restricts the parameter space of the linear effects to the shaded region shown in Fig. 6, panel b. Under this assumption, we can conclude that the cohort linear effect is positive and that the period linear effect is negative. Importantly, this initial bounding analysis informs us that more recent birth

²² The raw data are yearly, ranging from 1973 to 2013.

²³ 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.

²⁴ Estimated values of both θ_1 and θ_2 are statistically significant at the conventional threshold of .05.

²⁵ F tests indicate that all three time scales have statistically significant nonlinearities at the conventional threshold of .05.

²⁶ However, this interpretation is not strictly correct because we can estimate the values of θ_1 and θ_2 . For example, if we assume that the period slope is 0, then the age slope must be 7.344 rather than 0.

cohorts are generally 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 monotonically decreasing because screening tests for prostate cancer have improved over calendar time, thereby enabling earlier treatment. Most notably in the United States, the prostate-specific antigen (PSA) blood test introduced in the 1980s may have reduced the apparent (but not actual) incidence of prostate cancer because of increased detection. The assumption of a period slope that is not monotonically decreasing results in a minimum period slope greater than or equal to -4.458 , with corresponding bounds visualized in Fig. 6, panel c. Combining the assumption of a monotonically increasing age effect from 40 to 85 with a period slope that is not monotonically decreasing, we obtain the bounds shown in Fig. 6, panel d. With these two assumptions, we have restricted the temporal slopes to $8.110 \leq \alpha \leq 11.800$, $-4.456 \leq \pi \leq -0.766$, and $2.239 \leq \gamma \leq 5.929$.

When we incorporate the constrained slopes with the identifiable nonlinearities, we obtain the overall effects shown in panels d–f of Fig. 7. The shaded areas in these panels indicate the range of estimates consistent with our assumptions; the dashed lines denote the average of the upper and lower bounds for each of the age, period, and cohort slopes. These dashed lines correspond to values of $\alpha = 9.955$, $\pi = -2.611$, and $\gamma = 4.084$, 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 effect, prostate cancer incidence remained relatively flat until the early 1990s, after which it clearly decreased. Second, the cohort effect 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, given that 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.

However, there are two important caveats regarding these results. First, our APC analysis examines prostate cancer incidence, not prostate cancer mortality, which may lead to different results using the same set of constraints (cf. Braga et al. 2017; Chang et al. 1997; Niclis et al. 2011). For instance, increasing incidence may reflect improved screening techniques, which may, in fact, coincide with an overall decline in prostate cancer mortality due to earlier treatment. Second, our findings are not inconsistent with studies that examine just one or two of the time effects in prostate cancer incidence. For example, Herget et al. (2015) showed that, across recent periods, the incidence of prostate cancer in the United States has decreased. However, because they did not adjust for both age and cohort, the observed differences across periods were confounded with aging and cohort effects. To our knowledge, no recent studies examining prostate cancer incidence have explicitly modeled all three temporal variables.

Homicide Rates

We now turn to our second empirical example: temporal effects 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 more than 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 $\theta_1 = \alpha + \pi = -3.200$ and $\theta_2 = \pi + \gamma = -1.774$.²⁷ Both of these estimates suggest declines in the homicide rate across age and cohort groups, but these linear effects 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 panels a–c of Fig. 8.²⁸ As with the previous example, these can be difficult to interpret because they are deviations from an unknown linear effect, but the shape of the age nonlinearities is consistent with an underlying age-crime curve. As visualized in panel a of Fig. 8, if the age slope were 0, then we would indeed observe a sharp increase in the homicide arrest rate followed by a steady decline.

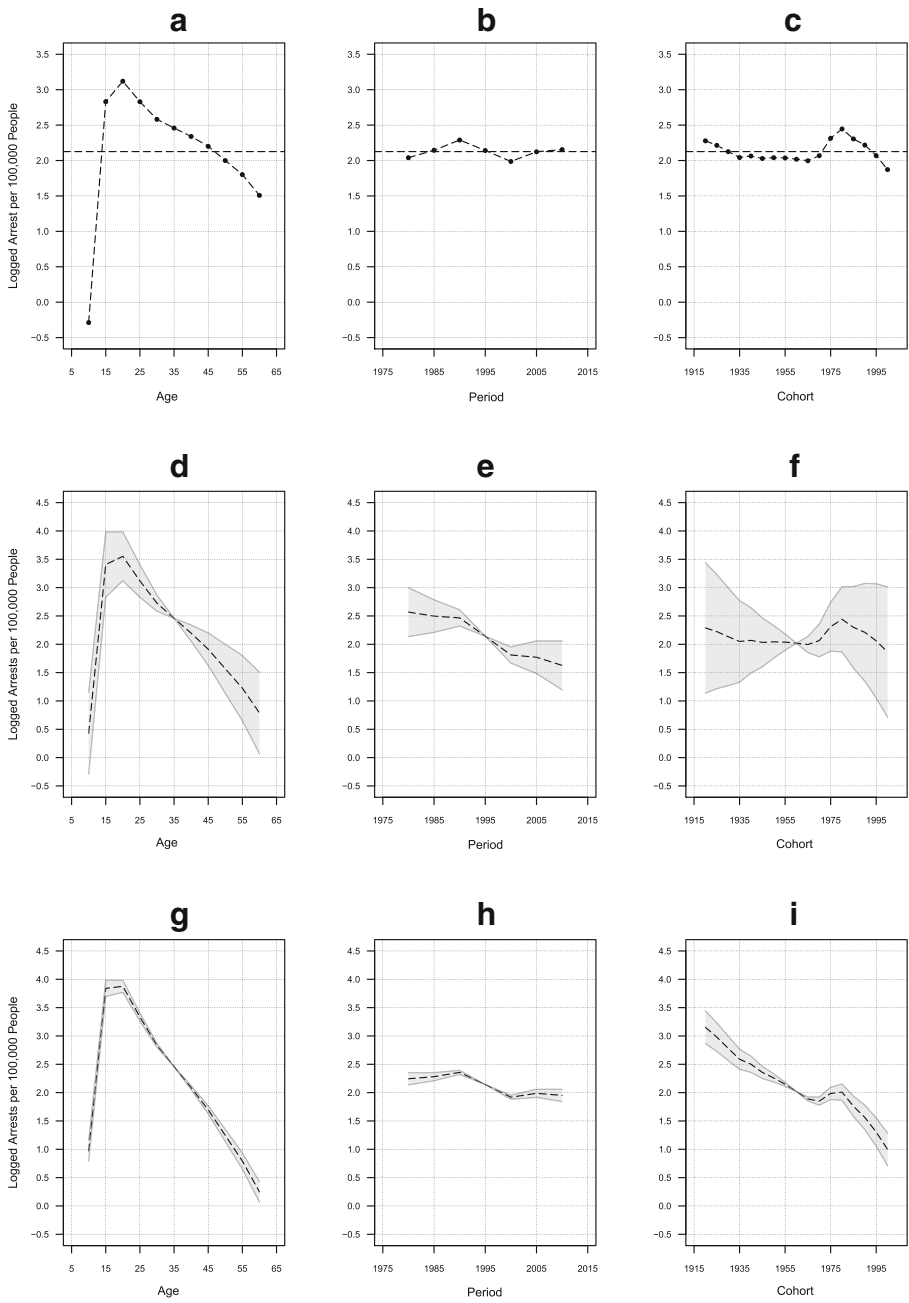
As discussed previously, violent crime is thought to decrease after young adulthood because of a variety of processes, including declining testosterone, greater economic resources, cognitive and analytical skill development, increased legal costs for criminal behavior, changing social norms, and involvement in long-term relationships (see Ulmer and Steffensmeier 2014). That is, we assume the existence of an age-crime curve, such that the crime rate increases monotonically to age 20–24 and decreases monotonically thereafter. This constraint is grounded in social, biological, and cultural theory, and is only as reliable as the theoretical assumptions on which it is based. Assuming that there is an age-crime curve of this form, then the age slope lies between -2.885 to 0 , inclusive.²⁹ If we assume that homicide arrest rates among men follow an age-crime curve, then we can rule out substantial sections of the parameter space. Specifically, we can conclude that the period slope must range from -3.200 to -0.321 , and the cohort slope must range from -1.453 to 1.427 . The corresponding overall effects are shown in panels d–f of Fig. 8. The shaded regions indicate the upper and lower bounds for each of the temporal effects, and the dashed lines indicate the average of the upper and lower bounds for each of the effects. Specifically, the dashed lines correspond to values of $\alpha = -1.440$, $\pi = -1.761$, and $\gamma = -0.013$, respectively. Several important conclusions follow from this initial bounding analysis of homicide arrest rates. As shown in panels e and f of Fig. 8, we can conclude that the overall period effect did *not* increase from 1980 to 2010 and that homicide rates did not monotonically decrease or increase across cohorts.

So far we have examined bounding only the overall age effect for homicide arrest rates. We can obtain narrower bounds using additional assumptions. The emergence of crack cocaine is widely thought to have increased violent crime rates in the 1980s to early 1990s, including homicides (Fryer et al. 2013). An additional assumption, then, is

²⁷ Estimated values of both θ_1 and θ_2 are statistically significant at the conventional threshold of .05.

²⁸ F tests show that at the conventional threshold of .05, all three time scales exhibit statistically significant nonlinearities.

²⁹ The 2D-APC graph of the solution line is available in the online appendix.



that homicide arrest rates from the 1980s to mid-1990s did not uniformly decrease. 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, reflecting the assumption that the crime rate increased in some periods but decreased in others. We further weaken this

Fig. 8 Temporal effects with and without bounds: homicide arrest rate. 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 effects under the assumption that the overall age effect 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 estimated values of $\alpha = -1.440$, $\pi = -1.761$, and $\gamma = -0.013$. Panels g, h, and i show the overall temporal effects under the assumption of a monotonic increase up to the 20–24 age group as well as no monotonic decrease from the 1980–1984 to 1990–1994 periods. Dashed lines correspond to $\alpha = -2.520$, $\pi = -0.681$, and $\gamma = -1.093$.

assumption by restricting this constraint to the periods 1980–1984 to 1990–1994. Thus, we are making no assumptions directly about periods after the 1980–1994 epoch. With this assumption, as well as the assumption that the age effect is monotonically increasing up to those aged 20–24, we can again conclude that the slopes are negative for all three temporal variables. Specifically, we know that $-2.885 \leq \alpha \leq -1.990$, $-2.151 \leq \pi \leq -0.316$, and $-1.458 \leq \gamma \leq -0.724$. When this information is combined with the nonlinearities (see Fig. 8, panels g, h, and i), we can 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 because of large differences between later and earlier birth cohorts.

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 address the APC identification problem, most recently the hierarchical age-period-cohort (HAPC) (or multilevel) models and estimators based on the Moore-Penrose generalized inverse. In this article, we outline an alternative framework that entails placing bounds on the temporal effects. Our approach begins with identifying what can be known from the data alone with as few constraints as possible. Because 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 effects.

We make several important contributions in this article. 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 provide us with identifiable linear combinations of the slopes even though each particular slope is unknown. These linear combinations 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 slopes. Finally, we present a variety of new bounding strategies, including not only those based on fixing the size and direction of one or more slopes but also those based on constraining just the sign of one or more slopes or applying a shape constraint to the overall effects.

These techniques offer a considerably wider range of identification approaches than commonly used in the literature.

There are, however, two caveats for applied researchers interested in using bounds to partially identify APC effects. First, like all APC methods, the constraints outlined here are not verifiable or falsifiable from the data at hand and can be justified only by appealing to social, biological, or cultural theory or the inclusion of additional data (e.g., see Winship and Harding 2008). To reiterate, the recovery of the true, unknown APC effects is only as reliable as the theoretical assumptions on which the constraints are based. Theories of the underlying temporal processes may be fundamentally flawed, thereby leading to mistaken conclusions about APC effects. There is, in this sense, no ultimate resolution of the APC identification problem. Second, because of our focus on model identification, this article says only little about quantifying uncertainty due to sampling variability. Although the data sets APC analysts use typically have large samples, failing to account for sampling variability will nonetheless render bounds that are too narrow. Especially when using small samples, researchers should use bootstrapping techniques to quantify uncertainty due to random sampling.

Notwithstanding, our approach is superior to existing APC methods in at least four ways. First, our constraints are explicit rather than implicit. As Luo (2013) and O'Brien (2015a) have noted, many currently popular methods for APC analysis, such as the intrinsic estimator (IE) or HAPC model, use constraints that are implicit, resulting in confusion and misuse by methodologists and applied researchers. Second, our method is superior in that it is more general and flexible than other approaches. Rather than specifying just one type of restriction, such as equality constraints, we show how various constraints on the size, shape, or sign on one or more of the parameters can be used to identify the temporal effects. Third, our constraints entail weaker theoretical assumptions. We demonstrate that specifying the sign, size, or shape can yield bounds that are quite narrow yet based on general theoretical assumptions about life course or period effects. Finally, our approach provides a novel visualization of the identification problem that facilitates the comparison of multiple estimators. Because all APC methods must address the identification problem (Fienberg 2013; Mason and Smith 1985), other methods—including the IE and HAPC—produce point estimates that lie within our 2D-APC graphs.

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 of the parameters. For example, one might use a gamma distribution as a prior for the age slope, reflecting the assumption that the slope ranges from 0 to positive infinity with some decreasing probability. To date, Bayesian APC analyses have instead focused primarily on using first- or second-order random walk priors on the APC effects, which can improve forecasts of disease rates (Riebler and Held 2017; Smith and Wakefield 2016; see also Nakamura 1986). Second, one can use our approach to conduct a sensitivity analysis of an APC model that includes mechanisms or proxy variables. Note that a sensitivity analysis and a bounds analysis are fundamentally identical (Morgan and Winship 2015). In general, a bounds analysis begins with weak assumptions and then uses stronger assumptions to narrow the bounds (in principle, up to a

point estimate). In contrast, a sensitivity analysis begins with strong assumptions (a point estimate) and then weakens those assumptions, yielding a set of upper and lower bounds. Although mechanisms and proxy variables introduce additional information into a traditional APC analysis, thereby allowing for point 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 sensitivity analysis 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), 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. (p. 3)

References

- Alwin, D. F. (1991). Family of origin and cohort differences in verbal ability. *American Sociological Review*, *56*, 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*, 7915–7922.
- 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*, 333–360. <https://doi.org/10.4054/DemRes.2014.30.11>
- 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 C. Burton-Jeangros, S. Cullati, A. Sacker, & D. Blane (Eds.), *A life course perspective on health trajectories and transitions* (pp. 197–213). Cham, Switzerland: Springer.
- Bell, A., & Jones, K. (2015b). Bayesian informative priors with Yang and Land’s age-period-cohort model. *Quality & Quantity*, *49*, 255–266.
- Braga, S. F. M., de Souza, M. C., & Cherchiglia, M. L. (2017). Time trends for prostate cancer mortality in Brazil and its geographic regions: An age-period-cohort analysis. *Cancer Epidemiology*, *50*, 53–59.
- Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. Cambridge, MA: National Bureau of Economic Research.
- Chang, C.-K., Yu, H.-J., Chan, K.-W. A., & Lai, M.-K. (1997). Secular trend and age-period-cohort analysis of prostate cancer mortality in Taiwan. *Journal of Urology*, *158*, 1845–1848.
- Chaves, M. (1989). Secularization and religious revival: Evidence from U.S. church attendance rates, 1972–1986. *Journal for the Scientific Study of Religion*, *28*, 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*, 2099–2104.

- Clark, A. K., & Eisenstein, M. A. (2013). Interpersonal trust: An age–period–cohort analysis revisited. *Social Science Research, 42*, 361–375.
- Clayton, D., & Schifflers, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine, 6*, 469–481.
- 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*, 360–365.
- Farrington, D. P. (1986). Age and crime. *Crime and Justice: A Review of Research, 7*, 189–250.
- Fienberg, S. E. (2013). Cohort analysis' unholy quest: A discussion. *Demography, 50*, 1981–1984.
- Fienberg, S. E., Hodges, J. S., & Luo, L. (2015). Letter to the editor. *Journal of the American Statistical Association, 110*, 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*, 487–500.
- Frost, W. H. (1995). The age selection of mortality from tuberculosis in successive decades. *American Journal of Epidemiology, 41*, 4–9. (Reprinted from *American Journal of Hygiene, 30*, 91–96).
- Fryer, R. G. Jr., Heaton, P. S., Levitt, S. D., & Murphy, K. M. (2013). Measuring crack cocaine and its impact. *Economic Inquiry, 51*, 1651–1681.
- 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*, 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*, 263–278.
- Fu, W. J. (2016). Constrained estimators and consistency of a regression model on a lexis diagram. *Journal of the American Statistical Association, 111*, 180–199.
- 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*, 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*, 900–904.
- Glenn, N. D. (1981). The utility and logic of cohort analysis. *Journal of Applied Behavioral Science, 17*, 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*, 298–300.
- Hauser, R. M., & Huang, M.-H. (1997). Verbal ability and socioeconomic success: A trend analysis. *Social Science Research, 26*, 331–376.
- Herget, K. A., Patel, D. P., Hanson, H. A., Sweeney, C., & Lowrance, W. T. (2015). Recent decline in prostate cancer incidence in the United States, by age, stage, and Gleason score. *Cancer Medicine, 5*, 136–141. <https://doi.org/10.1002/cam4.549>
- Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics, 53*, 161–177.
- Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology, 89*, 552–584.
- Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics, 39*, 311–324.
- Holford, T. R. (1985). An alternative approach to statistical age-period-cohort analysis. *Journal of Chronic Diseases, 38*, 831–836.
- Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals. *Statistics in Medicine, 25*, 977–993.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–960.
- Hout, M., & Knoke, D. (1975). Change in voting turnout, 1952–1972. *Public Opinion Quarterly, 39*, 52–68.
- 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*, 433–443.
- Keiding, N. (2011). Age-period-cohort analysis in the 1870s: Diagrams, stereograms, and the basic differential equation. *Canadian Journal of Statistics, 39*(Special Issue), 405–420.
- 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*, 1111–1120.

- Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28, 1356–1378.
- Land, K. C., Fu, Q., Guo, X., Jeon, S. Y., Reither, E. N., & Zang, E. (2016). Playing with the rules and making misleading statements: A response to Luo, Hodges, Winship, and Powers. *American Journal of Sociology*, 122, 962–973.
- 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, 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 and Public Health*, 45, 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, 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, 1335–1339.
- Loeber, R., & Farrington, D. P. (2014). Age-crime curve. In G. Bruinsma & D. Weisburd (Eds.), *Encyclopedia of criminology and criminal justice* (pp. 12–18). New York, NY: Springer.
- Luo, L. (2013). Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem. *Demography*, 50, 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, 930–961.
- Mannheim, K. (1952). The sociological problem of generations. In P. Kecskemeti (Ed.), *Essays on the sociology of knowledge* (pp. 276–322). London, UK: Routledge & Kegan Paul.
- Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review: Papers & Proceedings*, 80, 319–323.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60, 531–542.
- Manski, C. F. (2003). Identification problems in the social sciences and everyday life. *Southern Economic Journal*, 70, 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, 242–258.
- Mason, W. M., & Fienberg, S. E. (Eds.). (1985). *Cohort analysis in social research*. New York, NY: Springer.
- Mason, W. M., & Smith, H. L. (1985). Age-period-cohort analysis and the study of deaths from pulmonary tuberculosis. In W. M. Mason & S. E. Fienberg (Eds.), *Cohort analysis in social research* (pp. 151–227). New York, NY: Springer.
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Nakamura, T. (1986). Bayesian cohort models for general cohort table analyses. *Annals of the Institute of Statistical Mathematics*, 38, 353–370.
- Niclis, C., Pou, S. A., Bengió, R. H., Osella, A. R., & Díaz, M. d. P. (2011). Prostate cancer mortality trends in Argentina 1986–2006: An age-period-cohort and joinpoint analysis. *Cadernos de Saúde Pública/Ministério da Saúde, Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública (Public Health Notebooks)*, 27, 123–130. <https://doi.org/10.1590/S0102-311X2011000100013>
- O'Brien, R. M. (2011a). The age-period-cohort conundrum as two fundamental problems. *Quality and Quantity*, 45, 1429–1444.
- O'Brien, R. M. (2011b). Constrained estimators and age-period-cohort models. *Sociological Methods & Research*, 40, 419–452.
- O'Brien, R. M. (2015a). *Age-period-cohort models: Approaches and analyses with aggregate data*. Boca Raton, FL: CRC Press.
- O'Brien, R. M. (2015b). Age-period-cohort models and the perpendicular solution. *Epidemiological Methods*, 4, 87–99. <https://doi.org/10.1515/em-2014-0006>
- 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, 682–688.
- Palmore, E. (1978). When can age, period, and cohort be separated? *Social Forces*, 57, 282–295.
- Pearl, J. (2009). *Causality*. Cambridge, UK: Cambridge University Press.
- 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, 315–327.
- 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, 1033–1045.
- Pryor, W. A. (1982). Free radical biology: Xenobiotics, cancer, and aging. *Annals of the New York Academy of Sciences*, 393, 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, 1439–1448.
- Riebler, A., & Held, L. (2017). Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations. *Biometrical Journal*, 59, 531–549.
- Robinson, R. V., & Jackson, E. F. (2001). Is trust in others declining in America? An age-period-cohort analysis. *Social Science Research*, 30, 117–145.
- Rodgers, W. L. (1982a). Estimable functions of age, period, and cohort effects. *American Sociological Review*, 47, 774–787.
- Rodgers, W. L. (1982b). Reply to comment by Smith, Mason, and Fienberg. *American Sociological Review*, 47, 793–796.
- Rodgers, W. L. (1990). Interpreting the components of time trends. *Sociological Methodology*, 20, 421–438.
- Rubin, D. B. (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81, 961–962.
- Ryder, N. B. (1965). The cohort as a concept in the study of social change. *American Sociological Review*, 30, 843–861.
- Schwadel, P., & Stout, M. (2012). Age, period and cohort effects on social capital. *Social Forces*, 91, 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), djw322. <https://doi.org/10.1093/jnci/djw322>
- 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, 787–793.
- Smith, T. R., & Wakefield, J. (2016). A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science*, 31, 591–610.
- 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.
- Ulmer, J. T., & Steffensmeier, D. (2014). The age and crime relationship: Social variation, social explanations. In K. M. Beaver, J. C. Barnes, & B. B. Boutwell (Eds.), *The nurture versus biosocial debate in criminology: On the origins of criminal behavior and criminality* (pp. 377–396). Newbury Park, CA: Sage.
- 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, 287–302.
- Winship, C., & Harding, D. J. (2008). A mechanism-based approach to the identification of age-period-cohort models. *Sociological Methods & Research*, 36, 362–401.
- Yang, Y. (2008). Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review*, 73, 204–226.
- 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, 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, 75–97.
- Yang, Y., & Land, K. C. (2013). *Age-period-cohort analysis: New models, methods, and empirical applications*. Boca Raton, FL: CRC Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.