Demography (2013) 50:1945–1967
DOI 10.1007/s13524-013-0243-z

Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem

Liying Luo

Published online: 26 September 2013

Abstract In many different fields, social scientists desire to understand temporal variation associated with age, time period, and cohort membership. Among methods proposed to address the identification problem in age-period-cohort analysis, the intrinsic estimator (IE) is reputed to impose few assumptions and to yield good estimates of the independent effects of age, period, and cohort groups. This article assesses the validity and application scope of IE theoretically and illustrates its properties with simulations. It shows that IE implicitly assumes a constraint on the linear age, period, and cohort effects. This constraint not only depends on the number of age, period, and cohort categories but also has nontrivial implications for estimation. Because this assumption is extremely difficult, if not impossible, to verify in empirical research, IE cannot and should not be used to estimate age, period, and cohort effects.

Keywords Age-period-cohort analysis · Intrinsic estimator · Identification problem · Constrained estimator · Linear constraint

Introduction

For more than a century, social scientists have attempted to separate cohort effects from age and period effects on various social phenomena, including mortality, disease rates, and inequality (e.g., Fu 2000; Holford 1983; Mason et al. 1973; O'Brien 2000; Winship and Harding 2008). Whereas *age effects* represent the variation associated with growing older, *period effects* refer to effects due to social and historical shifts—such as economic recessions and prevalent unemployment—that affect all age groups simultaneously. Cohort refers to a group of people who experience an

Electronic supplementary material The online version of this article (doi:10.1007/s13524-013-0243-z) contains supplementary material, which is available to authorized users.

L. Luo (✉)
Department of Sociology, Minnesota Population Center, 909 Social Sciences,
267 19th Avenue S, Minneapolis, MN 55455, USA
e-mail: luoxx054@umn.edu

event—such as birth—at the same age. *Cohort effects* are defined as the formative effects of social events on individuals at a specific period during their life course (Ryder 1965). Age-period-cohort (APC) models, in which the three variables are simultaneously considered in a statistical equation, have been the conventional framework for quantifying age, period, and cohort effects. Unfortunately, such APC models suffer from a logical *identification problem*: when any two of the three variables (age, period, or cohort) are known, the value of the third is determined—that is, because Cohort = Period – Age. Because of this exact linear dependency, there exist no valid estimates of the distinct effects of the three variables.

Various methods have been developed to address this identification problem. For example, Mason et al. (1973) introduced the APC multiple classification model and suggested the constrained generalized linear model (CGLM) as a means of estimating the independent effects of age, period, and cohort. More recently, Fu (2000) and Yang and colleagues (2004) proposed a new APC method, called the *intrinsic estimator* (IE). They recommended IE as "a general-purpose method of APC analysis with potentially wide applicability in the social sciences" (Yang et al. 2008:1699) on the grounds that IE has desirable statistical properties such as unbiasedness and consistency.

However, in this article, I show that IE cannot be used to recover the true age, period, and cohort effects because IE, like CGLM, imposes a constraint on parameter estimation that is difficult to verify using theories or empirical evidence; that is, the validity of IE relies on assumptions that are very difficult to verify in applied practice. In this sense, IE is no better than CGLM. In fact, IE is equivalent to the principal component estimator, an estimator with a potential for bias that was noted by its developers (Kupper et al. 1985). Unfortunately, this has not been understood by the community of demographers, sociologists, and epidemiologists who have used IE in a wide variety of research applications. As I demonstrate in this article, many researchers have misunderstood what IE actually estimates and how IE estimates should be interpreted, resulting in inappropriate applications of IE in empirical research and potentially misleading substantive conclusions.

This article contributes to the literature in two ways: First, although O'Brien (2011a) clarified that IE assumes a special constraint—the null-vector constraint—on parameters, it is challenging for researchers to fully appreciate and evaluate the appropriateness of this constraint when applying IE in substantive studies. In this article, I derive an easily understood form of IE's constraint on the *linear* components of age, period, and cohort effects so that the implications of using IE to estimate the true age, period, and cohort effects can be better understood.[1]

Second, although scholars agree that IE is a constrained estimator, they debate whether IE can provide reliable estimates of the true age, period, and cohort trends (see Fu et al. 2011; O'Brien 2011b). I address this debate using several types of simulated data generated based on social theories. By

[1] One way to characterize the effects of an interval variable like time is to break the effect into two components: linear and nonlinear (curvature or deviations from linearity) trends. It has been known at least since Holford (1983) that the linear components of age, period, and cohort effects cannot be estimated without constraints because they are not identified. In contrast, nonlinear age, period, and cohort trends can be estimated without bias.

comparing IE estimates with the true effects in various circumstances, I show that IE does not work better than CGLM for recovering the true age, period, and cohort trends in empirical research.

This article is organized as follows. I begin with an introduction of the APC multiple classification model and the identification problem. While reviewing the methodological challenge that has hampered APC research for decades, this section establishes a framework for discussing the nature and limitations of different constrained APC estimators, including IE and CGLM. I then review how IE's developers have described IE and how applied researchers have understood and used it in substantive studies: the two are often not the same. As a result, many scholars have misunderstood IE, causing misuse of this technique in empirical research. To clarify this common misunderstanding and avoid further misuse, in the section on the linear constraint implied by IE, I derive the constraint that IE imposes on the *linear* components of age, period, and cohort effects. In the section on IE's application scope that follows the technical discussion of IE's linear constraint, I use simulations to demonstrate how this constraint affects estimation of age, period, and cohort effects. Based on these mathematical derivations and simulation evidence, I conclude that IE cannot and should not be used to estimate true age, period, and cohort effects without theoretical justification.

The Identification Problem

To develop a framework for understanding the nature of IE and other constrained estimators, I first review the identification problem that these methods are intended to address. In APC analysis, researchers have conventionally used an analysis of variance (ANOVA) model to separate the independent age, period, and cohort effects:

$$g\left(E\left(Y_{ij}\right)\right) = \mu + \alpha_i + \beta_j + \gamma_k, \tag{1}$$

for age groups $i = 1, 2, \ldots, a$; periods $j = 1, 2, \ldots, p$; and cohorts $k = 1, 2, \ldots, (a + p - 1)$; where $\sum_{i=1}^{a}\alpha_i=\sum_{j=1}^{p}\beta_j=\sum_{k=1}^{a+p-1}\gamma_k=0$. $E(Y_{ij})$ denotes the expected value of the outcome of interest Y for the ith age group in the jth period of time; g is the "link function"; α_i denotes the mean difference from the global mean μ associated with the ith age category; β_j denotes the mean difference from μ associated with the jth period; and γ_k denotes the mean difference from μ related to the membership in the kth cohort. The usual ANOVA constraint applies where the sum of coefficients for each effect is set to zero.

For a normally distributed outcome Y_{ij}, the ANOVA model above can also be written in a generic regression fashion:

$$\mathbf{Y} = \mathbf{Xb} + \varepsilon, \tag{2}$$

where $\mathbf{Y}$ is a vector of outcomes; $\mathbf{X}$ is the design matrix; $\mathbf{b}$ denotes a parameter vector with elements corresponding to the effects of age, period, and cohort groups; and ε denotes random errors with distribution centered on zero. Then the estimated age, period, and cohort effects can be obtained using the ordinary least squares (OLS) method:

$$\hat{b} = \left(X^T X\right)^{-1} X^T Y. \tag{3}$$

Unfortunately, the inverse $(X^T X)^{-1}$ does not exist because of the age-period-cohort linear dependency, so the parameter vector **b** is inestimable. This is the identification problem in APC analysis: no unique set of coefficients can be obtained because an infinite number of solutions give *identical* fits to the data.

This identification problem can be shown more explicitly. For simplicity, suppose that the data used are perfect, without random or measurement errors, so that $\varepsilon=\mathbf{0}$; then the problem is mathematical rather than statistical, and the regression model is

$$\mathbf{Y} = \mathbf{X}\mathbf{b}. \tag{4}$$

Because of the linear dependency between age, period, and cohort, there exists a nonzero vector $\mathbf{b_0}$—a linear function of the design matrix **X**—such that the product of the design matrix and the vector equals zero:

$$\mathbf{X}\mathbf{b_0} = 0. \tag{5}$$

In other words, $\mathbf{b}_0$ represents the null space of the design matrix **X**, which has dimension equal to 1. (The null space has dimension 1 by the specification of model (1), and the value of $\mathbf{b}_0$ is given in Eq. (10).) It follows that the parameter vector **b** can be decomposed into components:

$$\mathbf{b} = \mathbf{b_1} + s \cdot \mathbf{b_0}, \tag{6}$$

where s is an arbitrary real number corresponding to a specific solution to Eq. (4), and $\mathbf{b_1}$ is a linear function of the parameter vector **b**, corresponding to the projection of **b** on the nonnull space of the design matrix **X**, orthogonal to the null space. Thus, $\mathbf{b_1}$ and $\mathbf{b_0}$ are orthogonal to each other. That is, $\mathbf{b_1}$ is the part of **b** that is in the nonnull space of the design matrix **X**, orthogonal (perpendicular) to the null space, so that $\mathbf{b_0}$ is orthogonal to $\mathbf{b_1}$; that is, $\mathbf{b_1} \cdot \mathbf{b_0} = 0$.

Given Eqs. (4) and (6), the following equation must hold:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{b_1} + s \cdot \mathbf{b_0}) = \mathbf{X}\mathbf{b_1} + s \cdot \mathbf{X}\mathbf{b_0}. \tag{7}$$

However, $\mathbf{X}\mathbf{b_0} = 0$, and thus $s \cdot \mathbf{X}\mathbf{b_0} = 0$, so equation Eq. (7) is true for all values of s. That is, s can be *any* real number, and each distinct value of s gives a distinct solution to Eq. (4). Therefore, an infinite number of possible solutions exists for **b**, and no solution can be deemed the uniquely preferred or "correct" solution without additional constraints on **b**.

To illustrate, suppose the data have three age groups, three periods, and five cohorts, and that error is zero for ease of presentation (and without loss of generality). Table 1 presents three different parameter vectors $\mathbf{b}^T = (\mu,\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3,\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5)$ arising from three different values of s: namely 0, 2, and 10. In the top panel of Table 2, the observed value in each cell is represented in terms of the unknown parameters α_i, β_j, and γ_k. The bottom panel of Table 2 shows the fitted values $\mu + \alpha_i + \beta_j + \gamma_k$ based on Table 1's three different values of s in the same tabular form as in the top panel of Table 2. Note that these three sets of fitted values are *identical*, although the parameter vectors in Table 1 differ. In fact, these parameter vectors are not just different; their age and period

Table 1 Different values of *s* and the corresponding parameters

	Age			Period			Cohort				
s	α_1	α_2	α_3	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	γ_5
0	2	0	–2	–1	0	1	–1	–0.5	0	0.5	1
2	0	0	0	1	0	–1	–5	–2.5	0	2.5	5
10	–8	0	8	9	0	–9	–21	–10.5	0	10.5	21

Notes: *s* is an arbitrary real number corresponding to a specific solution to Eq. (4). Numbers in each row are a set of age, period, and cohort coefficients corresponding to a specific value of *s*.

effects change directions depending on *s*, and the data cannot distinguish between different values of *s*.

Taken together, Tables 1 and 2 show that for a single data set, an infinite number of possible solutions for age, period, and cohort effects exist; and each solution corresponds to a specific value of *s*. Therefore, any solution—or alternatively, none of these solutions—can be viewed as reflecting the "true" effects even though different values of *s* give radically different age, period, and cohort effects. In social science research, data inevitably contain random and/or measurement errors, so researchers will not have the perfect fit of the idealized data here; however, the fundamental identification problem remains. Various methods have been developed to address the identification problem and find a set of uniquely preferred estimates. In the following section, I consider IE and other solutions to the identification problem that impose a constraint on **b**.

The Constrained Approach: IE and CGLM

A large body of literature dating back to the 1970s has addressed the identification problem. Mason et al. (1973) explicated the identification problem in APC analysis

Table 2 Tabular data: Unobserved parameters and fitted values from the three different parameter vectors shown in Table 1

			Period		
			1	2	3
Unobserved Parameters	Age	1	$\mu+\alpha_1+\beta_1+\gamma_3$	$\mu+\alpha_1+\beta_2+\gamma_4$	$\mu+\alpha_1+\beta_3+\gamma_5$
		2	$\mu+\alpha_2+\beta_1+\gamma_2$	$\mu+\alpha_2+\beta_2+\gamma_3$	$\mu+\alpha_2+\beta_3+\gamma_4$
		3	$\mu+\alpha_3+\beta_1+\gamma_1$	$\mu+\alpha_3+\beta_2+\gamma_2$	$\mu+\alpha_3+\beta_3+\gamma_3$
Observed Values	Age	1	11	12.5	14
		2	8.5	10	11.5
		3	6	7.5	9

Note: The bottom panel presents *identical* observed values produced by the three different parameter vectors in Table 1.

and proposed the constrained generalized linear model (CGLM), a coefficient-constrained approach that has been used as a conventional method for APC analysis. This method places at least one identifying restriction on the parameter vector **b** in Eq. (2). For example, the effects of the first two age groups, periods, or cohorts are usually constrained to be equal based on theoretical or external information. With this additional constraint, the APC model becomes just-identified, and unique OLS and maximum likelihood (ML) estimators exist. However, such theoretical information often does not exist or cannot easily be verified. Also, different choices of identifying constraint can produce widely different estimates for age, period, and cohort effects. That is, CGLM estimates are quite sensitive to the choice of constraints (Glenn 2005; Rodgers 1982a, b).

More recently, a group of scholars developed a new APC estimator, called the *intrinsic estimator* (IE). They argued that IE has clear advantages over CGLM[2] and can produce valid estimates of the true age, period, and cohort effects (see Fu 2000, Fu and Hall 2006; Yang et al. 2004, 2008). The most compelling evidence they provided to support this claim is from a simulation in which IE and CGLM estimates were compared with the true effects of age, period, and cohort (see Yang et al. 2008:1718–1719). They concluded that IE outperforms CGLM because IE estimates are closer to the true parameters that generate the data than CGLM (Yang et al. 2008:1719–1722).

This evidence could easily be interpreted as confirmation that IE produces unbiased estimates of the true age, period, and cohort effects. Unfortunately, few clarifications are provided, and the developers of IE themselves are sometimes unclear about what IE actually estimates. For example, they state that

> for a finite number of time periods of data, the IE produces an unbiased estimate of the coefficient vector. (Yang 2008:400)

> Because of its estimability and unbiasedness properties, the IE may provide a means of accumulating reliable estimates of the trends of coefficients across the categories of the APC accounting model. (Yang et al. 2008:1711)

> The IE, by its very definition and construction, satisfies the estimability condition. . . . If other estimators do indeed satisfy the estimability condition, then they also produce unbiased estimates of the A, P, and C effect coefficients. If not, then the estimates they produce are biased. (Yang et al. 2008:1710)

> Perhaps most importantly for empirical applications of APC analysis, the IE produces estimated age, period, and cohort coefficients and their standard errors in a direct way, without the necessity of choosing among a large array of possible constraints on coefficients that may or may not be appropriate for a particular analysis. (Yang et al. 2004:105)

Many researchers conducting substantive APC analyses have interpreted these and other statements to mean that IE produces unbiased estimates of true age, period, and cohort effects. Consequently, they have used IE in empirical research to address substantive issues, including mortality, disease, and religious activity (e.g., Keyes and

[2] Yang et al. (2008) refer to this as "CGLIM."

Miech 2013; Langley et al. 2011; Miech et al. 2011; Schwadel 2011; Winkler and Warnke 2013). These authors seem convinced that IE produces unbiased estimates of age, period, and cohort effects:

> Recent advances in modeling APC effects with repeated cross-sectional data allow age, period, and cohort effects to be simultaneously estimated without making subjective choices requiring constraining data or dropping age, period, or cohort indicators from the model. In particular, APC intrinsic estimator models provide unbiased estimates of regression coefficients for age groups, time periods, and birth cohorts (Fu 2000). (Schwadel 2011:183)

> The intrinsic estimator provides unbiased estimates of age, period, and cohort effects. (Schwadel 2011:184)

> The IE model has been recommended as a better alternative to the widely discussed constrained generalized linear model (CGLM) (Yang et al. 2004). We used the IE model to estimate individual effects of age, period, and cohort for males and females separately. (Langley et al. 2011:106)

> The IE is an approach that places a constraint on the model, but not a constraint that affects the estimation of regression parameters for age, period, and cohort in any way. That is, the regression parameter estimates are unbiased by the constraint placed, and a unique set of regression estimates can be estimated. (Keyes and Miech 2013:2)

Unfortunately, claims of this sort are incorrect. As I demonstrate in this article, IE does impose constraints that are as consequential as those imposed by CGLM. To help researchers better understand the constraint imposed by IE and make informed decisions in choosing an APC estimator, I first derive an easily understood form of IE's constraint. Because an unbiased and consistent estimator is desirable and necessary to produce reliable and valid results, I then address how IE's constraint affects these key properties:

- Unbiasedness: Is the expectation of IE the "true" age, period, and cohort effects?
- Consistency: As the sample size increases, does IE converge to the "true" effects?

The Linear Constraint Implied by IE

To understand IE's constraint and its implications for estimation, it is helpful to review IE's conceptual foundation and computational algorithm. IE can be viewed as an extension of principal component (PC) analysis, a multipurpose technique that can be used to deal with identification problems when explanatory variables are highly correlated. By transforming correlated explanatory variables to a set of orthogonal linear combinations of these variables, called *principal components*, PC analysis can be a useful tool for reducing data redundancy and developing predictive models.

In contrast, the goal of IE is neither data reduction nor prediction, but rather estimation of the effects of, and capturing the general trends of, age, period, and cohort.[3] IE's computational algorithm includes five steps: (1) transform the design matrix **X** to the PC space using its eigenvector matrix; (2) in the PC space, identify the "null eigenvector"—the special eigenvector that corresponds to an eigenvalue of zero—and the corresponding null subspace (with 1 dimension) and nonnull subspace (with $m - 1$ dimensions, where m denotes the number of coefficients to be estimated); (3) in the nonnull subspace of $m - 1$ dimensions, regress the outcome of interest using OLS or ML on the $m - 1$ PCs to obtain $m - 1$ coefficient estimates; (4) extend the $m - 1$ coefficient estimates to the whole PC space of dimension m by adding an element corresponding to the null eigenvector direction and arbitrarily setting it to zero; and (5) use the eigenvector matrix to transform the extended coefficient vector estimated in the PC space, including the added zero element, back to the original age-period-cohort space to obtain estimates for age, period, and cohort effects (see Yang 2008; Yang et al. 2008).[4]

The fourth step—extend the $m - 1$ coefficient estimates to the whole PC space of dimension m by adding an element corresponding to the null eigenvector direction and arbitrarily setting it to zero—carries the key assumption of the IE approach to APC analysis. This assumption is implicit yet has major implications for the validity and application of the IE approach. Specifically, setting the coefficient of the null eigenvector, s, to zero is equivalent to assuming that

$$\mathbf{b} \cdot \mathbf{b_0} = 0. \tag{8}$$

That is, the projection of $\mathbf{b}$ on $\mathbf{b_0}$ is zero, where $\mathbf{b}$ and $\mathbf{b_0}$ are as defined in Eq. (6). Kupper and colleagues (1985) provided a closed-form representation for the eigenvector $\mathbf{b_0}$. Using vector notation,[5]

$$\mathbf{b_0} = (0, A, P, C)^T, \tag{9}$$

where

$$A = \left(1 - \frac{1+a}{2}, \ldots, (a-1) - \frac{1+a}{2}\right)$$

$$P = \left(-1 \cdot \left(1 - \frac{1+p}{2}, \ldots, (p-1) - \frac{1+p}{2}\right)\right)$$

$$C = \left(1 - \frac{a+p}{2}, \ldots, (a+p-2) - \frac{a+p}{2}\right).$$

[3] It is important to distinguish data reduction or prediction from coefficient estimation. Because the identification problem does not prevent obtaining a set of solutions with good fit to the data, one can still make good predictions. The PC technique treats such problems as data redundancy and allows obtaining one solution. However, as noted earlier, none of these solutions is the uniquely preferred solution: the solution that APC techniques, including IE, aim to discover. Therefore, providing a solution for the purpose of prediction is not the same as finding a uniquely preferred solution for estimation of separate age, period, and cohort effects.

[4] Alternatively, Yang (2008:413) described the computational algorithm of IE as follows: after obtaining $r - 1$ coefficients in the PC space $(w_2, \ldots, w_r)$, "set coefficient w_1 equal to 0 and transform the coefficients vector $w = (w_1, \ldots, w_r)^T$," where w_1 corresponds to the null eigenvector direction.

[5] Yang et al. (2004, 2008) used $\mathbf{b_0^*} = \frac{\mathbf{b_0}}{\|\mathbf{b_0}\|}$, where $\|\mathbf{b_0}\|$ is the length of $\mathbf{b_0}$, so $\mathbf{b_0^*}$ has a length of 1. $\mathbf{b_0}$ is used in this article because it is simply a multiple of $\mathbf{b_0^*}$ and is simpler for exposition and computation.

For example, when a = 3 and p = 3 (i.e., for three age groups and three time periods),

$$\mathbf{b_0} = (0,-1,0,1,0,-2,-1,0,1)^T, \tag{10}$$

where A = (–1,0), P = (1,0), and C = (–2,–1,0,1).

What does Eq. (8) mean? What is the specific form of this constraint for data sets with varying number of age, period, and cohort groups? To illustrate, suppose that age, period, and cohort each have effects on the outcome variable that show a linear trend. Denote these trends as k_a, k_p, and k_c, respectively; the intercepts for the three variables as i_a, i_p, and i_c; and the overall mean as μ. Thus, the effects associated with the three age categories are i_a, $i_a + k_a$, and $i_a + 2 \cdot k_a$, respectively. Similarly, the effects related to the three periods are i_p, $i_p + k_p$, and $i_p + 2 \cdot k_p$, respectively. For the five cohorts, the effects are i_c, $i_c + k_c$, $i_c + 2 \cdot k_c$, $i_c + 3 \cdot k_c$, and $i_c + 4 \cdot k_c$, respectively. Then, the parameter vector, **b**, can be written as

$$\mathbf{b} = \left(\mu, i_a, i_a + k_a, i_p, i_p + k_p, i_c, i_c + k_c, i_c + 2 \cdot k_c, i_c + 3 \cdot k_c\right)^T, \tag{11}$$

where the last category of each variable is omitted as the reference group. According to the constraint for age effects in model (1), we know that

$$\sum\nolimits_{i=1}^{a} \alpha_i = i_a + (i_a + k_a) + (i_a + 2 \cdot k_a) = 3 \cdot i_a + 3 \cdot k_a = 0, \tag{12}$$

which implies that

$$i_a = -k_a. \tag{13}$$

Similarly, it can be shown using the constraint for period and cohort effects in model (1) that

$$i_p = -k_p, \tag{14}$$

and

$$i_c = -2 \cdot k_c. \tag{15}$$

Using Eqs. (13), (14), and (15), Eq. (11) can be simplified as

$$\mathbf{b} = \left(\mu, -k_a, 0, -k_p, 0, -2 \cdot k_c, -k_c, 0, k_c\right)^T. \tag{16}$$

Because the constraint that IE implicitly imposes is $\mathbf{b} \cdot \mathbf{b_0} = 0$, by Eqs. (8), (10), and (16), the specific form of IE's linear constraint (LC) for APC data with three age categories, three periods, and five cohorts are

$$\begin{aligned} \mathbf{b} \cdot \mathbf{b_0} = {} & \mu \cdot 0 + (-k_a) \cdot (-1) + 0 \cdot 0 + \left(-k_p\right) \cdot 1 + 0 \cdot 0 + (-2 \cdot k_c) \cdot (-2) + (-k_c) \cdot (-1) \\ & + 0 \cdot 0 + k_c \cdot 1 = k_a - k_p + 6 \cdot k_c = 0. \end{aligned} \tag{17}$$

In other words, when age, period, and cohort show linear trends, IE's implicit constraint is that these linear trends *must* satisfy Eq. (17). If, in fact, the true age, period, and cohort trends do not satisfy this equation, then the implicit LC imposed by IE is incorrect.

To illustrate the implications of IE's LC, I simulate normally distributed data as follows. For those at age i in period j, the mean response is $10+k_a \cdot age_i+k_p \cdot period_j+k_c \cdot cohort_{ij}$, and the standard deviation of error ε equals 0.1. The number of age and period groups is fixed at three each. I consider three sets of true k_a, k_p, and k_c: (1) $k_a=1$, $k_p=7$, $k_c=1$; (2) $k_a=1$, $k_p=7$, $k_c=10$; and (3) $k_a=3$, $k_p=1$, $k_c=4$. For each selection of true k_a, k_p, and k_c, I simulate 1,000 such data sets by drawing random errors. As shown in Table 3, for data set 1, the true effects for the three age categories are –1, 0, and 1, respectively, so k_a (the linear trend in age effects) equals 1. The period effects are –7, 0, and 7, respectively, so k_p is 7. Similarly, since the cohort effects are –2, –1, 0, 1, and 2, k_c is 1. For this data set,

$$k_a - k_p + 6 \cdot k_c = 1 - 7 + 6 \cdot 1 = 0. \quad (18)$$

That is, the relationship between the linear trends in the true age, period, and cohort effects satisfies Eq. (17), the LC implicit in IE. However, for data sets 2 and 3 generated by the other sets of true k_a, k_p, and k_c in Table 3, Eq. (17) does not hold. Specifically, for the second set, $k_a = 1$, $k_p = 7$, and $k_c = 10$, so

$$k_a - k_p + 6 \cdot k_c = 1 - 7 + 6 \cdot 10 = 54 \neq 0. \quad (19)$$

And for the third set, $k_a = 3$, $k_p = 1$, and $k_c = 4$, so

$$k_a - k_p + 6 \cdot k_c = 3 - 1 + 6 \cdot 4 = 26 \neq 0. \quad (20)$$

Table 3 presents IE estimates, averaged over the 1,000 simulated data sets, for the three sets of age, period, and cohort effects. The bias of IE is estimated by the difference between the truth and the averaged IE estimates. Table 3 shows that for data set 1, IE yields good estimates because the true k_a, k_p, and k_c in the data satisfy

Table 3 Simulation results: IE estimates for three data sets

		Data Set 1			Data Set 2			Data Set 3		
		Truth	IE	Bias	Truth	IE	Bias	Truth	IE	Bias
Age	1	–1	–0.997	0.003	–1	5.747	6.747	–3	0.249	3.249
	2	0	–0.002	–0.002	0	0.002	0.002	0	0.000	0.000
	3	1	0.999	–0.001	1	–5.749	–6.749	3	–0.249	–3.249
Period	1	–7	–6.999	0.001	–7	–13.75	–6.750	–1	–4.250	–3.250
	2	0	–0.002	–0.002	0	0.002	0.002	0	–0.002	–0.002
	3	7	7.002	0.002	7	13.748	6.748	1	4.252	3.252
Cohort	1	–2	–2.001	–0.001	–20	–6.497	13.503	–8	–1.500	6.500
	2	–1	–0.998	0.002	–10	–3.253	6.747	–4	–0.750	3.250
	3	0	–0.001	–0.001	0	0.002	0.002	0	0.000	0.000
	4	1	1.004	0.004	10	3.250	–6.750	4	0.750	–3.250
	5	2	1.996	–0.004	20	6.498	–13.502	8	1.500	–6.500

Notes: For each data set, the IE estimates are averaged over 1,000 simulations. The bias of IE is evaluated by the difference between the true effects and the IE estimates, averaged over 1,000 simulations. Equation (17) holds for data set 1 but not for data set 2 and 3.

Eq. (17), the implicit LC that IE imposes. Specifically, the estimated slopes for age, period, and cohort are $\hat{k}_a = 0.999$, $\hat{k}_p = 7.001$, and $\hat{k}_c = 1.000$, respectively. In contrast, IE returns highly biased estimates, very different from the true effects, for the second and third data sets because the true k_a, k_p, and k_c do not satisfy IE's LC. For example, for data sets 2 and 3, the estimated age effects, averaged over the 1,000 simulations, show a downward trend ($\hat{k}_a = -5.750$ for data set 2 and $\hat{k}_a = -2.582$ for data set 3) when the true trend is upward. (The true age slopes are $k_a = 1$ for data set 2 and $k_a = 3$ for data set 3.)

Note that Eq. (17) is derived for the simplest scenario in which the age, period, and cohort trends are purely linear. For more complex scenarios in which these trends are not purely linear, IE's constraint depends on the nonlinear components of the age, period, and cohort effects.[6] For example, suppose that age, period, and cohort each has effects on the outcome of interest that include a linear and a quadratic trend. Denote the quadratic trends as k'_a, k'_p, and k'_c, respectively. Using the same derivation in this section, the specific form of IE's constraint for APC data with three age categories, three periods, and five cohorts is

$$\left(k_a - k_p + 6 \cdot k_c\right) + \frac{5}{3}\left(k'_a - k'_p + 12 \cdot k'_c\right) = 0. \tag{21}$$

That is, when age, period, and cohort effects include quadratic components, these effects must satisfy Eq. (21) in order for IE to yield good estimates. Equation (17) can be viewed as a special case of Eq. (21) when there are no quadratic or higher-order nonlinear components in the age, period, and cohort effects. Alternatively, because the linear dependency between age, period, and cohort does not affect the identification of nonlinear effects, IE's constraint can be said to *bind* only on the linear age, period, and cohort trends; the specific value of the constraint on the linear effects is determined by the nonlinear effects, which are estimable.

For any coefficient-constraint approach, such as CGLM and IE, "the choice of constraint is *the* crucial determinant of the accuracy in the estimated age, period, and cohort effects" (Kupper et al. 1985:822). Because the constraint assumption strongly affects estimation results, no matter what constraint a statistical method assumes, that method produces good estimates only when its assumption approximates the true structure of the data under investigation. It follows that when there are three age groups, three periods, and five cohorts and their effects are purely linear, IE can yield accurate estimates *only* when these linear effects of age, period, and cohort satisfy Eq. (17). Unfortunately, researchers usually have no *a priori* knowledge about true age, period, and cohort effects that would allow them to evaluate whether the constraint implied in Eq. (17) holds. Therefore, researchers cannot assess whether IE produces unbiased estimates of age, period, and cohort effects for their data. Thus, IE is no better than CGLM in this respect.

[6] The constraint imposed by IE depends on how model (2) is parameterized. If the model is parameterized in terms of orthogonal polynomial contrasts for each of the age, period, and cohort effects, as in Holford (1983), then IE imposes a constraint solely on the linear contrasts of age, period, and cohort effects irrespective of any nonlinear trends that are present. The parameterization used here is more common (e.g., Kupper et al. 1985), and in this parameterization, the constraint on the linear components of the age, period, and cohort effects depends on the nonlinear components when both components are present.

More importantly, the exposition in this section indicates that the LC assumed by IE also depends on the design matrix **X**—that is, on the number of age, period, and cohort groups. For example, if one age group is added to the example presented here, such that there are now four age groups, three periods, and six cohorts, then the LC implied by IE is

$$\mathbf{b}\cdot\mathbf{b_0} = 2.75\cdot k_a - k_p + 11.25\cdot k_c = 0, \tag{22}$$

or

$$\mathbf{b}\cdot\mathbf{b_0} = \left(k_a - k_p + 6\cdot k_c\right) + \left(1.75\cdot k_a + 5.25\cdot k_c\right) = 0. \tag{23}$$

Compared with Eq. (17) for the case of three age groups, three periods, and five cohorts, Eqs. (22) and (23) show that adding an age group dramatically changes the constraint so that the true effects satisfying IE's LC with three age categories no longer satisfy this LC when an age category is added. Readers can verify that increasing or reducing the number of periods or cohorts also greatly alters IE's LC.

These examples demonstrate that not only does IE rely on a constraint like CGLM does, but unlike CGLM, in which the constraint (e.g., equal effects for the first two age groups) is explicit and rationalized by theoretical account or side information, the LC of IE is implicit and varies depending on the number of age, period, and cohort groups. Although this constraint has been described as minimal (e.g., Schwadel 2011; Yang et al. 2008), in fact, it can have major implications for the quality of substantive results, as shown.

Theoretically speaking, the limitation of IE results from a misinterpretation of the constraint that IE imposes on parameter estimation. It is true that $\mathbf{b_0}$, the null eigenvector, is determined by the design matrix, but it is incorrect to conclude that therefore $\mathbf{b_0}$ "should not play any role in the estimation of effect coefficients" (Yang et al. 2008:1705). Rather, both the null eigenvector and nonnull eigenvectors (with nonzero eigenvalues) are determined by the design matrix—that is, by the number of age, periods, and cohort groups. To this extent, it is no less likely that the data contain a significant component in the $\mathbf{b_0}$ direction than in the directions of the nonnull eigenvectors. The fact that s, the coefficient for $\mathbf{b_0}$, can be *any* real number without changing the fitted values $\mathbf{Xb}$ simply means that variation in $\mathbf{Y}$ in the direction of $\mathbf{b_0}$ is not estimable. If the data have variation in this direction, IE will mistakenly attribute that variation to other columns in the design matrix, causing significant errors in estimation.

The Implications of IE's Constraint: Is IE an Unbiased and Consistent Estimator?

Because IE imposes a constraint on the linear age, period, and cohort trends, IE yields reliable estimates only when the true trends satisfy its constraint. However, Yang and colleagues argued that "because of its estimability and unbiasedness properties, the IE may provide a means of accumulating reliable estimates of the trends of coefficients across the categories of the APC accounting model" (Yang et al. 2008:1711). In the following discussion, I clarify that IE is not an unbiased estimator of the "true" age, period, and cohort effects. I also use concrete examples to illustrate that IE is not consistent and to explain why IE appears to be converging to the truth in the Yang et al. (2008) article. This section may be particularly helpful for nontechnical researchers.

Biasedness

By definition, an estimator δ is an unbiased estimator of a parameter θ if the expectation of δ over the distribution that depends on θ is equal to θ, or $E_\theta(\delta)=\theta$. It follows that for an unbiased APC estimator, its expectation must be the true effects of age, period, and cohort.[7] Per this definition, if IE is an unbiased estimator, the expected value of IE must be the true age, period, and cohort effects. The following mathematical computation shows, however, that the expectation of the IE estimator is not the true effects unless those true effects happen to satisfy IE's implicit constraint.

As noted in the preceding section, the key computation of IE is to extend the coefficient estimates in the PC space, $\mathbf{b}'$,

$$(\mathbf{b}')^T = (b'_0, b'_1, b'_2, \ldots, b'_{m-1}) \tag{24}$$

by adding a zero element, such that

$$(\mathbf{b}'_{\mathbf{new}})^T = (b'_0, b'_1, b'_2, \ldots, b'_{m-1}, 0), \tag{25}$$

where $\mathbf{b}'_{\mathbf{new}}$ corresponds to the projection of the coefficient vector $\mathbf{b}$ in the nonnull space, that is, $\mathbf{b_1}$ in Eq. (6). IE then transforms the extended coefficient vector $\mathbf{b}'_{\mathbf{new}}$, including the added zero element, back to the original age-period-cohort space to obtain coefficient estimates for age, period, and cohort.

Given that OLS and ML estimators have been proven unbiased in simpler, identifiable problems with normally distributed errors as in Eq. (2), and because IE uses these methods to obtain estimates for $\mathbf{b_1}$, whose projection in the PC space corresponds to the extended coefficient vector $\mathbf{b}'_{\mathbf{new}}$, IE yields unbiased estimates for $\mathbf{b_1}$. In other words,

$$E(\mathbf{b_{IE}}) = \mathbf{b_1}. \tag{26}$$

Based on the preceding discussion of the identification problem, the true parameter space $\mathbf{b}$ can be decomposed into two orthogonal subspaces corresponding to $\mathbf{b_1}$ and $\mathbf{b_0}$ in Eq. (6), which is equivalent to

$$\mathbf{b_1} = \mathbf{b} - s \cdot \mathbf{b_0}. \tag{27}$$

[7] Yang and colleagues have used "unbiasedness" in a different sense to mean that the expectation of IE is equal to $\mathbf{b_1}$, the projection of parameter vector $\mathbf{b}$ onto the nonnull space of design matrix $\mathbf{X}$ (see, e.g., Yang et al. 2008:1709). This is an important distinction because the true parameter vector $\mathbf{b}$ can be very different from its projection $\mathbf{b_1}$ onto the nonnull space, the vector that IE actually estimates. Because APC analysts are usually interested in estimating the true age, period, and cohort effects, the classic concept of unbiasedness is more relevant to APC research than that used by IE's proponents. Thus, I use "unbiasedness" in its classic sense in my discussion.

Substituting Eq. (27) in Eq. (26) results in

$$E(\mathbf{b_{IE}}) = \mathbf{b_1} = \mathbf{b} - s \cdot \mathbf{b_0}. \tag{28}$$

Equation (28) means that the expectation of the IE estimator will be different from the true effects **b** unless $s \cdot \mathbf{b_0} = 0$—that is, unless $s = 0$. IE assumes $s = 0$; thus, IE is a biased estimator when the true value of s is anything but 0. The larger the absolute value of s, the more biased the IE estimates become.

For researchers who wish to investigate age, period, and cohort effects for the purposes of substantive demographic, social, or other applied research, there exists little theoretical or empirical knowledge about the value of s and what $\mathbf{b_0}$—the null eigenvector—may imply about the outcome variable. In specific applications, then, IE must be assumed to be biased, resulting in misleading conclusions about the true age, period, and cohort effects unless proven otherwise.

Note that IE's developers argue that IE satisfies the "estimability criterion" proposed by Kupper et al. (1985), so IE is, in that sense, an unbiased estimator. However, estimability of a function of **b** implies unbiased estimation only of the estimable function of **b**, not necessarily of the true parameter **b** itself. The projection of the parameter vector onto the nonnull space, $\mathbf{b_1}$, is indeed an estimable function of **b**, the true parameter vector, and thus IE is an unbiased estimator of $\mathbf{b_1}$. However, IE is a biased estimator for the true APC effects when $\mathbf{b_1}$ is different from **b**. Therefore, it is not accurate to say that "Kupper et al. (1985) . . . suggested that an estimable function satisfying this condition resolves the identification problem" as claimed in Yang et al. (2008:1703). To emphasize, estimability in the nonnull space does not imply unbiasedness in estimating the *true* age, period, and cohort effects. Discovering a set of estimable functions is not the same as solving the identification problem.

Consistency

In statistics, for an estimator δ to be a consistent estimator of an unknown parameter space θ, δ must converge in probability to θ as the sample size grows. If δ is unbiased, consistency usually follows immediately. A biased estimator can be consistent if its bias decreases as the sample size increases. However, the bias of IE, $s \cdot \mathbf{b_0}$, does not necessarily shrink as the sample size grows. Thus, IE is not a consistent estimator of the coefficient vector **b**.

This theoretical argument can be illustrated with simulations. I simulate normally distributed data using the same function as that for data set 1 in Table 3: for those at age i in period j, the mean response is $10 + 1 \cdot age_i + 7 \cdot period_j + 1 \cdot cohort_{ij}$, and the standard deviation of error is 0.1. I begin with 3 age groups and 3 periods, and then increase the number of periods to 6 and 12, respectively. For each scenario, I simulate 1,000 such data sets by drawing random errors. If IE is a consistent estimator, then as the number of periods increases, the resulting estimates should get closer and closer to the true effects that we know based on the simulation function.

Table 4 presents the IE estimates, averaged over 1,000 data sets, for the three scenarios in which the number of periods is set at 3, 6, and 12, respectively. It shows that the IE estimates are not converging to the truth, and the bias appears to increase

Table 4 Simulation results: Inconsistent IE estimates as the number of periods increases

		Periods = 3			Periods = 6			Periods = 12		
		Truth	IE	Bias	Truth	IE	Bias	Truth	IE	Bias
Age	1	–1	–0.997	0.003	–1	–2.144	–1.144	–1	–3.016	–2.016
	2	0	–0.002	–0.002	0	0.000	0.000	0	–0.000	–0.000
	3	1	0.999	–0.001	1	2.144	1.144	1	3.017	2.017
Period	1	–7	–6.999	0.001	–17.5	–14.642	2.858	–38.5	–27.407	11.093
	2	0	–0.002	–0.002	–10.5	–8.783	1.717	–31.5	–22.427	9.073
	3	7	7.002	0.002	–3.5	–2.931	0.569	–24.5	–17.441	7.059
	4				3.5	2.928	–0.572	–17.5	–12.462	5.038
	5				10.5	8.785	–1.715	–10.5	–7.470	3.030
	6				17.5	14.643	–2.857	–3.5	–2.493	1.007
	7							3.5	2.494	–1.006
	8							10.5	7.473	–3.027
	9							17.5	12.455	–5.045
	10							24.5	17.441	–7.059
	11							31.5	22.425	–9.075
	12							38.5	27.412	–11.088
Cohort	1	–2	–2.001	–0.001	–3.5	–7.497	–3.997	–6.5	–19.612	–13.112
	2	–1	–0.998	0.002	–2.5	–5.362	–2.862	–5.5	–16.590	–11.090
	3	0	–0.001	–0.001	–1.5	–3.214	–1.714	–4.5	–13.575	–9.075
	4	1	1.004	0.004	–0.5	–1.071	–0.571	–3.5	–10.558	–7.058
	5	2	1.996	–0.004	0.5	1.074	0.574	–2.5	–7.542	–5.042
	6				1.5	3.215	1.715	–1.5	–4.525	–3.025
	7				2.5	5.353	2.853	–0.5	–1.506	–1.006
	8				3.5	7.502	4.002	0.5	1.509	1.009
	9							1.5	4.529	3.029
	10							2.5	7.542	5.042
	11							3.5	10.559	7.059
	12							4.5	13.575	9.075
	13							5.5	16.589	11.089
	14							6.5	19.604	13.104

as the number of periods increases from 3 to 12. Specifically, when p (the number of periods) equals 6 and 12, although IE correctly captures the direction of the age, period, and cohort trends, there is no evidence that these estimates are converging to the truth; the estimated age, period, and cohort slopes are $\hat{k}_a = 2.144$, $\hat{k}_p = 5.857$, and $\hat{k}_c = 2.144$, respectively, when $p = 6$; and $\hat{k}_a = 3.017$, $\hat{k}_p = 4.983$, and $\hat{k}_c = 3.017$, respectively, when p increases to 12. In fact, even with an unrealistically large number of periods (e.g., 100 periods), as I show in Fig. S1 in Online Resource 1, the IE estimates do not appear to converge to the truth.

The developers of IE correctly noted that the estimation of period and cohort effects will not improve with more time periods because "adding a period to the data set does not add information about the previous periods or about cohorts not present in the period just added" (Yang et al. 2008:1718). However, when they simulated data, the IE estimates for age effects did appear to become closer and closer to the true values as the number of periods increased. They simulated data using the following function:

$$y_{ij} \sim Poisson\left\{\exp\left[0.3+0.1\left(age_i-5\right)^2+0.1\sin\left(period_j\right) + 0.1\cos\left(cohort_{ij}\right)+0.1\sin\left(10\cdot cohort_{ij}\right)\right]\right\}. \tag{29}$$

It appears that IE estimates of the age effects converge to the true effects in this simulation as the number of periods increases because IE's implicit LC is not satisfied by the true age, period, and cohort effects in the simulation mechanism Eq. (29) with five periods ($\mathbf{b}\cdot\mathbf{b_0} = -0.339$), but the true effects do approximately satisfy the LC ($\mathbf{b}\cdot\mathbf{b_0} = -0.036$) when the number of periods increases to 50. In other words, IE appears to perform better as the number of periods increases, *not* because IE is a consistent procedure but because the true effects used in the data-generating function Eq. (29) conform better to IE's implicit LC as the number of periods increases.

For demographic or social data in which the linear trends in the three variables are unknown, adding more periods or cohorts promises nothing about the accuracy of the coefficient estimation for age, period, or cohort effects. That is, even with a sufficiently large sample, researchers using IE to estimate the true age, period, and cohort effects are not guaranteed to have desirable results that are close to the true values.

Application Scope: IE Versus CGLM

The preceding discussions of IE's linear constraint (LC) and statistical properties are fairly technical. In this section, I use several types of simulated data to illustrate how the implicit LC of IE affects its ability to recover the underlying age, period, and cohort effects in social science research.[8] This exercise is important because scholars have debated the application scope of IE in empirical research. As Fu et al. (2011:455) suggested, "the important statistical issue about APC modeling is how to identify the trend that helps to resolve the real-world problem for a given APC data set." Thus, I examine whether compared

[8] Yang and colleagues have used empirical data, in which the true effects are unknown, to assess the properties and performance of IE (Yang et al. 2008:1712–1716). However, it is logically impossible to assess the performance of an estimator when the true effects are unknown. If such a cross-model validation of IE for a specific empirical data set were to show that IE yields reasonable estimates, this can only depend on having selected examples that are consistent with the IE's constraint. Therefore, cross-model comparisons using empirical data are not an appropriate method to validate IE.

with CGLM, IE yields better (if not unbiased) estimates of the true age, period, and cohort patterns that may be observed in empirical research.

IE's developers provided simulations in which IE estimates are closer to the true age, period, and cohort effects than CGLM results. This, they argued, supports their conclusion that IE has clear advantages over CGLM. However, as noted earlier, the true age, period, and cohort effects in Yang et al.'s (2008) simulation in fact approximately satisfy the LC that IE imposes ($\mathbf{b} \cdot \mathbf{b_0} = -0.036$).[9] For age, period, and cohort effects that do not satisfy IE's implicit constraint, IE will not necessarily perform better than CGLM and may perform much worse. Thus, IE is no better than CGLM because the restriction that IE imposes is essentially no different from the constraints assumed in CGLM.

To illustrate, I show simulations, as Yang and colleagues did, to compare the CGLM and IE estimates. However, here the data-generating mechanisms satisfy the constraint assumed by CGLM but not the constraint assumed by IE. Moreover, I simulate data from four models that embody specific social theories and thus conform to empirical reality. The first data set is simulated to represent the observation that overall health for adults deteriorates as they grow older, and that while recent developments in health knowledge and technology have improved health conditions for the entire population, people born in more recent years are likely to be healthier than older cohorts. On the other hand, the demographic literature has also suggested that age, period, or cohort effects may not all exist (Alwin 1991; Fabio et al. 2006; Preston and Wang 2006; Winship and Harding 2008). Accordingly, the other three simulations approximate likely empirical situations in which one of the three variables has little impact on the outcome variable.

Specifically, I fix the number of age groups at 9 and the number of periods at 50 in all these simulations with little loss of generality. I then generate 1,000 data sets from each of the following four models:

$$y_{ij} \sim Normal\left\{10 + 2 \cdot age_i - 0.5 \cdot age_i^2 + 1 \cdot period_j - 0.015 \cdot period_j^2 + 0.15 \cdot cohort_{ij} + 0.03 \cdot cohort_{ij}^2, \sigma = 0.1\right\} \tag{30}$$

$$y_{ij} \sim Normal\left\{10 + 1 \cdot period_j - 0.015 \cdot period_j^2 + 0.15 \cdot cohort_{ij} + 0.03 \cdot cohort_{ij}^2, \sigma = 0.1\right\} \tag{31}$$

$$y_{ij} \sim Normal\left\{10 + 2 \cdot age_i - 0.5 \cdot age_j^2 + 0.15 \cdot cohort_{ij} + 0.03 \cdot cohort_{ij}^2, \sigma = 0.1\right\} \tag{32}$$

$$y_{ij} \sim Normal\left\{10 + 2 \cdot age_i - 0.5 \cdot age_j^2 + 1 \cdot period_j - 0.015 \cdot period_{ij}^2, \sigma = 0.1\right\}. \tag{33}$$

[9] Although Yang and colleagues correctly pointed out that IE estimates the *projection* of the "true" effects onto the nonnull space, they compared IE estimates with the "true" *parameters*, not to the projection (Yang et al. 2008:1718–1722). This is key because the true parameter vector can be very different from its projection onto the nonnull space (the vector that IE actually estimates). That is, what IE actually estimates can be very different from the true APC effects if the true effects do not at least approximately satisfy the LC implicit in IE.

For instance, in Eq. (30), the outcomes for people with age i in period j are normally distributed with mean $(10+2\cdot age_i-0.5\cdot age_j^2+1\cdot period_j-0.015\cdot period_j^2+0.15\cdot cohort_{ij}+0.03\cdot cohort_{ij}^2)$, and standard deviation $\sigma = 0.1$. In Eqs. (31), (32), and (33), one of the age, period, and cohort effects is not present, while the effects for the other two variables are the same as in Eq. (30). Note that none of these models satisfies IE's constraint. Specifically, for the first model, $\mathbf{b}\cdot\mathbf{b_0}=115.01$; and for the second, third, and last model, $\mathbf{b}\cdot\mathbf{b_0}=115.72$, 130.41, and 16.12, respectively.

Figure 1 compares, for the simulated data from the four models, IE estimates and CGLM estimates using two different constraints. The IE estimates, averaged over 1,000 data sets, are largely different from the true effects for all models because for all four models, the constraint that IE assumes is not satisfied. For example, in Scenario 3 in Fig. 1, when there is no period effect in the data-generating mechanism Eq. (32), the IE estimates suggest a substantially positive period trend on top of inaccurate estimates for age and cohort effects. In contrast, the CGLM assuming equal age effects for the first and third age groups produces close estimates for all four models. It is equally important to note that the performance of the CGLM estimator also depends on whether its assumption approximates the truth. For instance, in Scenario 4, whereas the CGLM that assumes equal age effects for the first and third group yields good estimates, the same method with a different constraint—that is, the age effects are the same for the first and second groups—results in biased estimates.

In sum, it must be concluded that (1) if there is *a priori* information or theoretical justification, the constrained solution that corresponds to such information (e.g., CGLM estimates assuming equal effects for the first and third age groups in data-generating functions Eqs. (30) to (33)) will yield better estimates than IE; and (2) without such *a priori* knowledge, IE is not necessarily better than other constrained estimators including CGLM. Without such knowledge, neither IE nor CGLM results are valid.

Conclusion and Discussion

In this article, I focus on the intrinsic estimator (IE), a statistical method intended to separate the independent effects of age, period, and cohort on various outcomes. I discuss the nature and application scope of IE theoretically and illustrate it with simulated data. This article shows that IE assumes a specific constraint on the linear age, period, and cohort effects. This assumption not only depends on the number of age, period, and cohort groups, but also is extremely difficult, if not impossible, to verify in empirical research. This feature of IE is no different from the constraint assumed in CGLM except that the CGLM constraint does not change automatically as the numbers of age, period, and cohort groups change. The conclusion is that IE is not an unbiased or consistent estimator of the "true" age, period, and cohort effects. Therefore, for demographers and social scientists whose goal is to understand the "true, simultaneously independent effects" of age, period, and cohort, IE's strategy of circumventing the identification problem can yield biased and potentially misleading estimates.

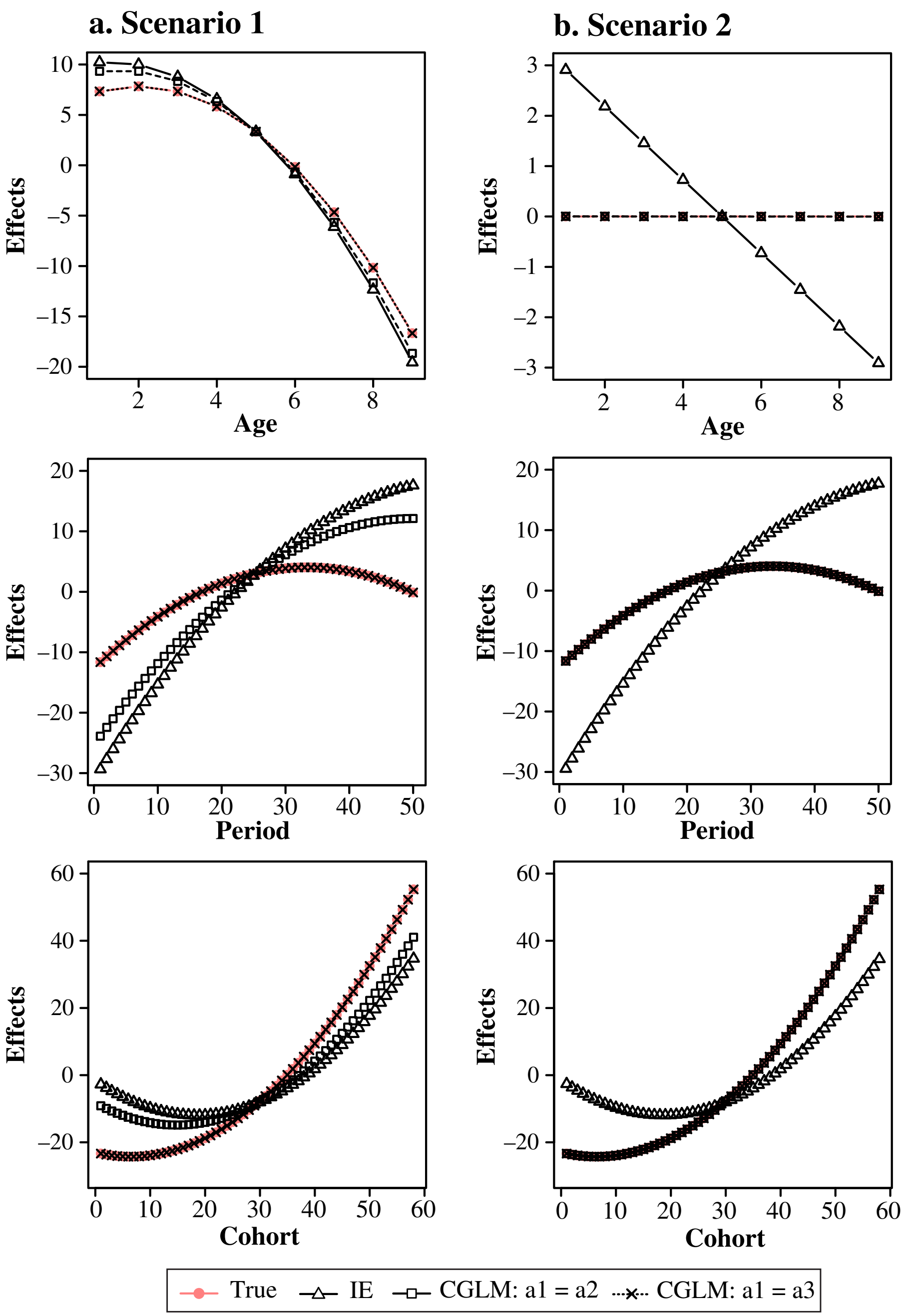

Fig. 1 Simulation results: IE versus CGLM estimates for age, period, and cohort effects

There is no doubt that Yang and associates have revitalized APC research and inspired many scholars. However, IE is nothing new in APC analysis. Kupper and his

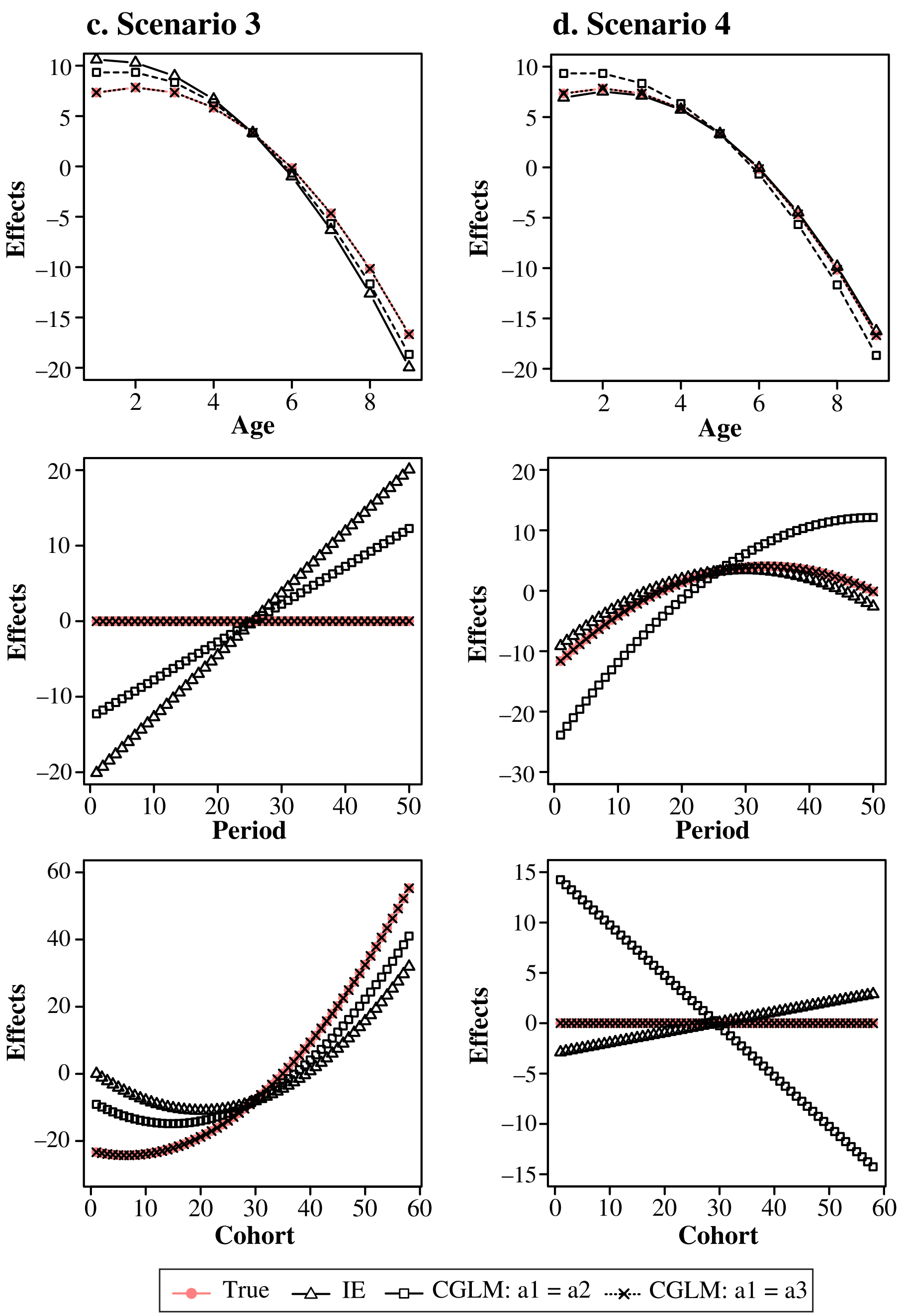

Fig. 1 (continued)

colleagues introduced the IE solution to APC analysts, calling this solution the principal component estimator (PCE) (Kupper et al. 1983:2795–2797). As O'Brien

(2011a:420) noted, such an estimator "produces coefficients identical to those of the recently introduced intrinsic estimator." However, instead of concluding that IE is preferable to CGLM, Kupper et al. (1983:2797) clearly stated that PCE (that is, IE) "could lead to more bias than the use of some other constraints." As a result, Kupper and associates did not advocate PCE/IE as a general solution, then or subsequently.

Generally speaking, PCE/IE or any other constrained estimator provides just *one* possible solution from the infinite number of solutions for an underdetermined problem (i.e., the rank deficiency problem in APC analysis). That said, the PCE/IE solution should not be regarded as *the true solution* or *the uniquely preferred solution* without theoretical justification. In fact, the statistical literature has recognized a variety of constrained estimators, including other types of generalized inverse solutions. Demographers and sociologists should understand that the PCE/IE estimates are not necessarily better (i.e., closer to the true parameters) than other constrained estimators.

What should well-intentioned researchers who wish to investigate the age, period, and cohort patterns do? On the one hand, several alternative methods have been developed, some of which are more theoretically driven, taking external information into account,[10] and some of which are statistical approaches.[11] Although each method has advantages and limitations and a thorough examination is a topic for future research, I caution that purely statistical techniques are unlikely to yield accurate estimates. The methodological problem of IE and its nontrivial implications for empirical research identified in this article are not unique to IE. The biostatistics literature shows that use of the APC model (1), regardless of estimation technique, precludes valid estimation and meaningful interpretations of the linear components of age, period, and cohort effects (see, e.g., Holford 1983; Kupper et al. 1985). Therefore, my position is to encourage development of APC models that are informed by social theories and thus different from model (1) in basic structure.

On the other hand, although the statistical difficulty in quantifying independent effects of age, period, and cohort was recognized long ago, decades of effort has only resulted in unsatisfactory solutions. Thus, it is not unreasonable to ask, Is this unusual challenge suggesting a problem that is not statistical but theoretical in nature? In other words, is the identification problem pointing to a more fundamental problem in the theoretical framework of APC analysis? Should the answers to these questions be positive, the identification problem inherent in model (1) "is a blessing for social science" (Heckman and Robb 1985:144) because it warns scientists that they want something—a general statistical decomposition of data—for nothing.

Acknowledgments I am grateful to James Hodges, John Robert Warren, Robert O'Brien, Christopher Winship, Daniel Powers, Carolyn Liebler, Samir Soneji, Ann Meier, Ian Ross Macmillan, Caren Arbeit, Julia Drew, Catherine Fitch, Julian Wolfson, and Wenjie Liao for their helpful comments. I also thank the Maryland Population Research Center for support. A version of this article was presented at the 2012 meeting of the Population Association of America, San Francisco, CA.

[10] Examples include age-period-cohort characteristic models developed by O'Brien (2000) and the mechanism-based approach proposed by Winship and Harding (2008).

[11] Examples are the cross-classified random-effects models created by Yang and Land (2006, 2008).

References

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

Fabio, A., Leober, R., Balasubramani, G. K., Roth, J., Fu, W., & Farrington, D. P. (2006). Why some generations are more violent than others: Assessment of age, period, and cohort effects. *American Journal of Epidemiology, 164,* 151–160.

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

Fu, W. J., & Hall, P. (2006). Asymptotic properties of estimators in age-period-cohort analysis. *Statistics & Probability Letters, 76,* 1925–1929.

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.

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

Heckman, J., & Robb, R. (1985). Using longitudinal data to estimate age, period, and cohort effects in earnings equations. In W. M. Mason & S. E. Fienberg (Eds.), *Cohort analysis in social research* (pp. 137–150). New York: Springer-Verlag.

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

Keyes, K., & Miech, R. (2013). Age, period, and cohort effects in heavy episodic drinking in the US from 1985 to 2009. *Drug and Alcohol Dependence*. Advance online publication. doi:10.1016/j.drugalcdep.2013.01.019

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

Kupper, L. L., Janis, J., Salama, I. A., Yoshizawa, C. N., Greenberg, B. G., & Winsborough, H. H. (1983). Age-period-cohort analysis: An illustration of the problems in assessing interaction in one observation per cell data. *Communications in Statistics—Theory and Methods, 12,* 2779–2807.

Langley, J., Samaranayaka, A., Davie, J., & Campbell, A. J. (2011). Age, cohort and period effects on hip fracture incidence: Analysis and predictions from New Zealand data 1974–2007. *Osteoporosis International, 22,* 105–111.

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

Miech, R., Koester, S., & Dorsey-Holliman, B. (2011). Increasing US mortality due to accidental poisoning: The role of the baby boom cohort. *Addiction, 106,* 806–815.

O'Brien, R. M. (2000). Age period cohort characteristic models. *Social Science Research, 29,* 123–139.

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

O'Brien, R. M. (2011b). Intrinsic estimators as constrained estimators in age-period-cohort accounting models. *Sociological Methods & Research, 40,* 467–470.

Preston, S. H., & Wang, H. (2006). Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography, 43,* 631–646.

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.

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

Schwadel, P. (2011). Age, period, and cohort effects on religious activities and beliefs. *Social Science Research, 40,* 181–192.

Winkler, R., & Warnke, K. (2013). The future of hunting: An age-period-cohort analysis of deer hunter decline. *Population Environment, 34,* 460–480.

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). Trends in U.S. adult chronic disease mortality, 1960–1999: Age, period, and cohort variations. *Demography, 45,* 387–416.

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. (2008). Age–period–cohort analysis of repeated cross-section surveys: Fixed or random effects? *Sociological Methods & Research, 36,* 297–326.

Yang, Y., Schulhofer-Wohl, S., Fu, W. J., & Land, K. C. (2008). The intrinsic estimator for age-period-cohort analysis: What it is and how to use it. *American Journal of Sociology, 113,* 1697–1736.