Qual Quant (2011) 45:1429–1444
DOI 10.1007/s11135-010-9397-6

The age–period–cohort conundrum as two fundamental problems

Robert M. O'Brien

Published online: 12 December 2010

Abstract In general the age–period–cohort (APC) conundrum refers to the problem of separating the effects of age-groups, periods, and cohorts. This formulation, however, fails to differentiate two fundamental problems in APC analysis: (1) the problem of the complete confounding of the linear effects of age with the effects of period and cohort, the linear effects of cohorts with period and age, and the linear effects of period with age and cohort; and (2) the problem of model identification. We elucidate both problems and show how the first problem makes the partitioning of variance between cohort effects, period effects, and age effects and the deviation of their effects from linearity problematic even when these approaches do not suffer from the problems associated with model identification. We conclude by examining the affects of this linear confounding on estimates of the individual effect coefficients for age-groups, periods, and cohorts when a linear constraint it imposed on the matrix of independent variables to produce an identifiable model.

Keywords Age–period–cohort models · Two fundamental problems · Constrained Generalized Linear Models · Intrinsic Estimator

1 Introduction

Recognition and proposed remedies to the age–period–cohort (APC) problem have a long history in the social sciences (e.g., Greenberg et al. 1950; Hobcraft 1982; Mason et al. 1973). In the most general terms we might refer to the APC problem as one of separating the effects of age-groups, periods, and cohorts. When examining unemployment rates we might ask; are age, year (period), and birth cohort all needed to obtain the best predicted values of the age–period-specific rates? Are there particular birth cohorts that are prone to homicide offending? These are important questions and they certainly have to do with the separation of the effects of age-groups, periods and cohorts. This separation formulation, however, fails

R. M. O'Brien (✉)
Department of Sociology, University of Oregon, Eugene, OR 97403, USA
e-mail: bobrien@uoregon.edu

to differentiate two fundamental, and distinct, problems in APC analysis: (1) the problem of the complete confounding of the linear effects of age with the effects of period and cohort, the linear effects of cohorts with period and age, and the linear effects of period with age and cohort; and (2) the problem of model identification.

We first clearly differentiate these two fundamental problems. Second, we show how these two problems impact the conclusions that can be drawn from various "solutions" to the APC problem. In a section on estimable functions, we examine the impact of complete linear confounding on approaches for which model identification is not a problem. Specifically, we examine the following approaches: attributing the unique variance in the dependent variable to age-groups, periods, or cohorts; examining departures of the effects of each cohort, or period, or age-group from linearity; and finally examining approaches that use a proxy variable to characterized the effect of age, period or cohort by using a measurable characteristic of, for example, cohorts. We then address two approaches to the identification problem: the traditional Constrained Generalized Linear Model (CGLM) approach and the recently proposed Intrinsic Estimator (IE) approach (Yang et al. 2004, 2008) and show how they work with data for which age-group, period, and cohort effects are confounded.

1.1 Multiple classification (dummy variable) coding for APC analysis

Age-groups, periods, and cohorts are often coded with dummy variables in APC analysis, because coding in this way does not constrain the functional form of the relationship between the dependent variable and the age-groups, periods, and cohorts. We adhere to this tradition by using the multiple classification model represented in Eq. 1 throughout this paper,[1]

$$Y_{ij} = \mu + \alpha_i + \pi_j + \chi_{a-i+j} \tag{1}$$

where Y_{ij} is the age–period-specific value of the dependent variable, μ represents the intercept, α_i represents the effect of the ith age group, π_j denotes the effects of the jth period, and χ_{a-i+j} represents the effects of the $(a - i + j)$th cohort (where a equals the number of age groups).

Table 1 is an age–period table showing the patterning of the independent variable effects in a standard APC analysis. Age groups are coded by dummy variables for the rows, periods with dummy variables for the columns, and cohorts with dummy variables for the diagonals from the lower left to the upper right (in any analysis, of course, one category serves as a reference category for each of these sets of dummy variables). The cells represent the expected values of the age–period-specific dependent variable denoted in Eq. 1. Table 1 illustrates the unusual additive pattern of effects in the APC model, with the effects of age and period representing the additive effect of the row and column variables and cohorts representing the additive effects of the diagonal variables. It is standard in the analysis of variance, for example, to have the fixed effect of row three and the fixed effect of column 4 contribute to the response variable in cell (3, 4) in an additive manner. What is unusual for the analysis of variance, but standard for APC analysis, is that the fixed effect for the diagonal for cohort 6 contributes to the value for cell (3, 4) in an additive manner. This unusual patterning (seen throughout Table 1) plays a significant role in the confounding of the linear effects of any one factor (age or period or cohort) with the effects of the other factors.

[1] The model can also be estimated as a Poisson regression or as a logistic regression in a straightforward manner (see Yang et al. 2008).

Table 1 The age-period table with the fixed age, period, and cohort parameters that affect the cell values

Age group	Period group (j)				
(i)	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$i = 1$	$\mu + \alpha_1 + \pi_1 + \chi_5$	$\mu + \alpha_1 + \pi_2 + \chi_6$	$\mu + \alpha_1 + \pi_3 + \chi_7$	$\mu + \alpha_1 + \pi_4 + \chi_8$	$\mu + \alpha_1 + \pi_5 + \chi_9$
$i = 2$	$\mu + \alpha_2 + \pi_1 + \chi_4$	$\mu + \alpha_2 + \pi_2 + \chi_5$	$\mu + \alpha_2 + \pi_3 + \chi_6$	$\mu + \alpha_2 + \pi_4 + \chi_7$	$\mu + \alpha_2 + \pi_5 + \chi_8$
$i = 3$	$\mu + \alpha_3 + \pi_1 + \chi_3$	$\mu + \alpha_3 + \pi_2 + \chi_4$	$\mu + \alpha_3 + \pi_3 + \chi_5$	$\mu + \alpha_3 + \pi_4 + \chi_6$	$\mu + \alpha_3 + \pi_5 + \chi_7$
$i = 4$	$\mu + \alpha_4 + \pi_1 + \chi_2$	$\mu + \alpha_4 + \pi_2 + \chi_3$	$\mu + \alpha_4 + \pi_3 + \chi_4$	$\mu + \alpha_4 + \pi_4 + \chi_5$	$\mu + \alpha_4 + \pi_5 + \chi_6$
$i = 5$	$\mu + \alpha_5 + \pi_1 + \chi_1$	$\mu + \alpha_5 + \pi_2 + \chi_2$	$\mu + \alpha_5 + \pi_3 + \chi_3$	$\mu + \alpha_5 + \pi_4 + \chi_4$	$\mu + \alpha_5 + \pi_5 + \chi_5$

2 First fundamental problem: the confounding of linear effects

The first fundamental problem of APC analysis involves linear relationships between time and the effects of age-groups, periods or cohorts. Such effects can produce a complete confounding of the linear effects of age, period, and cohort. By linear effects in this context, we refer to the effects of age (or period or cohort) that increase or decrease over time. For example, a period effect that grows steadily from the earliest to the most recent period or an age-group effect that decreases uniformly from the youngest to the oldest age-group. This problem is distinct from the identification problem.[2] We illustrate this problem in Table 2. First, we examine the top entries (bolded italic) in each cell of the age–period table and visualize these data as generated by a linear affect of cohorts that grows over time. Each new cohort is more prone to suicide and the rate increases by 0.5 per 100,000 for each new cohort. The earliest cohort is represented by the earliest period and the oldest age group: the bottom left-hand cell of Table 2. There is only one observation for this cohort, since in the next period (period 2) members of this cohort would be in age-group 6, which is not represented in the table. The next cohort has two observations: one in period1-age4 and another in period2-age5 and the suicide rates are 3.5 for both observations. The next oldest cohort has three observations and they are all 4.0. This pattern is repeated throughout the table with an increment of 0.5 for each cohort. From one perspective this is a pure cohort effect that is linear in terms of time since birth. It is clear also from Table 1 how this pattern is derived with χ_2 having a value 0.5 greater than χ_1 and χ_3 having a value of 0.5 greater than χ_2, and so on ending with χ_9. We could label, $Y_{ij} = 2.5 + (a - i + j) \times 0.5$, as the "generating formula" for these data—a pattern generated by a pure linear cohort effect with no age or period effects.

The depth of the first fundamental problem becomes apparent when we shift perspectives and examine these same age–period-specific rates (bold-italics) from a period or a column perspective. On average the effect of period increases by 0.5 for each period from the earliest to the most recent. From an age (row) perspective, the decrease is 0.5 for each age-group beginning with the youngest age-group to the oldest. Is this an effect of age and period with no cohort effect or a pure cohort effect? After all, we can reproduce the bolded age–period-specific effects in Table 2 using only age and period dummy variables.[3]

[2] Note that the identification problem is a problem even if there is not a linear relationship between the effects of age (or period or cohort) and time.

[3] $Y_{ij} = 5.0 - 0.5\mathrm{d}_{a2} - 1.0\mathrm{d}_{a3} - 1.5\mathrm{d}_{a4} - 2.0\mathrm{d}_{a5} + 0.5\mathrm{d}_{p2} + 1.0\mathrm{d}_{p3} + 1.5\mathrm{d}_{p4} + 2.0\mathrm{d}_{p5}$, where Y_{ij} is the cell value for the ith age group in the jth period and the d's are the zero-one coded dummy variables for the age groups and periods with age-group 1 and period 1 serving as the reference categories.

Table 2 The confounding problem of linear effects in APC models for cohorts, ages, and periods

	Period 1	Period 2	Period 3	Period 4	Period 5
Age 1	***5.0***	***5.5***	***6.0***	***6.5***	***7.0***
	3.0	3.0	3.0	3.0	3.0
	3.0	*3.5*	*4.0*	*4.5*	*5.0*
	5.0	*5.5*	*6.0*	*6.5*	*7.0*
Age 2	***4.5***	***5.0***	***5.5***	***6.0***	***6.5***
	3.5	3.5	3.5	3.5	3.5
	3.0	*3.5*	*4.0*	*4.5*	*5.0*
	4.5	*5.0*	*5.5*	*6.0*	*6.5*
Age 3	***4.0***	***4.5***	***5.0***	***5.5***	***6.0***
	4.0	4.0	4.0	4.0	4.0
	3.0	*3.5*	*4.0*	*4.5*	*5.0*
	4.5	*4.5*	*5.0*	*5.5*	*6.0*
Age 4	***3.5***	***4.0***	***4.5***	***5.0***	***5.5***
	4.5	4.5	4.5	4.5	4.5
	3.0	*3.5*	*4.0*	*4.5*	*5.0*
	4.5	*4.5*	*4.5*	*5.0*	*5.5*
Age 5	***3.0***	***3.5***	***4.0***	***4.5***	***5.0***
	5.0	5.0	5.0	5.0	5.0
	3.0	*3.5*	*4.0*	*4.5*	*5.0*
	4.5	*4.5*	*4.5*	*4.5*	*5.0*

As Table 2 shows, it is a problem for linear effects associated with age or period (as well as for cohort). The second row in each of the cells of Table 2 could represents a pure linear age effect with the response variable increasing 0.5 as age increases from age1 to age2 to age3 to age4 to age5. This linear "age effect," however, can be fully attributed to cohort and period effects (as cohort year of birth increases the suicide rate for cohorts decreases by 0.5 and as period increases the suicide rate increases by 0.5 for each period).[4] The third row in each cell could represent a pure linear period effect. Again, in this case these "period effects" can be fully attributed to age and cohort effects (as age decreases the suicide rate decreases by 0.5 and as birth cohort increases the suicide rate increases by 0.5). It is the structure of the APC data matrix that guarantees the complete confounding of a "linear effect" of age (or cohort or period) with the other two sets of dummy variables of interest.

This first fundamental problem in APC models involves the complete confounding of linear effects. This makes it impossible to separate the linear effects of age, period, and cohort. We refer to this as a Y-side problem, since it is created by the patterning of the response variable. We will see that the first fundamental problem creates difficulties even when the

[4] This equivalence can be difficult to see. Imagine that there is a 0.5 increase in the cohort effect as we move from the youngest to the oldest cohort and that there is a decrease in the period effect as we move from the earliest to the most recent period. For example, for age 5 in the table, we would expect the data be 5.0, 5.5, 6.0, 6.5, and 7.0 as we move across periods reflecting the increase due to cohort replacement. But with a 0.5 decrease moving from the earliest to the most recent periods associated with the period effect, these observed rates for age 5 would all be 5.0 as they are in Table 2 in the second row entries. This same argument can be applied to the other age groups.

approaches used do not require "model identification" of the individual age, period, and cohort effects.

3 Second fundamental problem: model identification

The identification problem in APC analysis arises because of the linear dependency between age, period, and cohort dummy variables. Using matrix notation of we can denote the multiple classification Eq. 1 as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{Y}$ is an $ap \times 1$ vector of the age–period-specific rates. The $\boldsymbol{X}$-matrix is an $ap \times 2(a + p) - 3$ matrix that contains ones in the first column and "dummy variables" in the remaining columns. The order of the columns can be schematized as $[1, (a - 1), (p - 1), (a + p - 2)]$, where $a - 1$ represents the number of age dummy variables, $p - 1$ represents the number of period dummy variables, and $a + p - 2$ represents the number of cohort dummy variables. Each of these dummy variables has ap entries (zeros and ones) in its column and is coded to correspond to the "cell" in which the age–period-specific value of Y resides.

When a regular inverse exists, solving an equation of the form of (2) for a unique set of least square regression coefficients is trivial:

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

where the superscripted T represents the transpose and the superscripted -1 indicates the inverse. The problem with this solution in the APC case occurs because of a linear dependency in the $\boldsymbol{X}$-matrix, which means that the standard inverse of $X^T X$ does not exist. If we label the number of columns in the $\boldsymbol{X}$ matrix as $m[= 2(a + p) - 3]$, only $m - 1$ of these columns are linearly independent.

Table 3 illustrates this second fundamental problem and is helpful in discussing the IE approach to the identification problem. The first vector is a vector of ones and the other vectors represent the dummy variables for the age-groups, periods, and cohorts (excluding the reference categories). The b's represent coefficients that when multiplied times the column vectors produce the zero-vector on the right hand side of the equal sign. The fact that such a vector of b's exists (not all of which are zero) indicates that there is a linear dependency in the columns of the $\boldsymbol{X}$-matrix. In general, one and only one such vector of bs exists for the APC model when no special constraints are placed on the model (this vector is unique up to multiplication by a scalar). In the language of linear algebra, we say that there is a nontrivial solution to the $a \times p$ homogeneous equations, and we can write $\boldsymbol{X}v = 0$ where $\boldsymbol{X}$ is the $ap \times 2(a + p) - 3$ design matrix and v is a $2(a + p) - 3 \times 1$ vector containing the bs of Table 3. This vector is said to be in the null space of $\boldsymbol{X}$ and is labeled as the null vector. It is the linear combination of columns of $\boldsymbol{X}$ that result in the zero-vector. That there is only one such vector indicates that the rank of the $\boldsymbol{X}$ matrix is just one less than full column rank and that a single linear constraint should allow for *a solution* to the identification problem.

Thus, to find a unique solution to the individual dummy variables, a linear constraint must be chosen.[5] Each such constraint is associated with a generalized inverse that allows

[5] Such a constraint in the case of CGLM would be written, for example, as: $c^T = (0, 1, -1, 0, 0, \ldots, 0)$ for the case where age1 and age2 are assumed to have the same effect in the population (the coding for age1 and age2 being in the second and third columns of the $\boldsymbol{X}$-matrix). The assumption is that c^T times the vector of population effect coefficients is zero.

Table 3 The model identification problem in APC models

$$b_1\begin{bmatrix}1\\1\\\vdots\\1\\1\end{bmatrix}+b_2\begin{bmatrix}x\\x\\\vdots\\x\\x\end{bmatrix}+b_3\begin{bmatrix}x\\x\\\vdots\\x\\x\end{bmatrix}+\cdots+b_{m-1}\begin{bmatrix}x\\x\\\vdots\\x\\x\end{bmatrix}+b_m\begin{bmatrix}x\\x\\\vdots\\x\\x\end{bmatrix}=\begin{bmatrix}0\\0\\\vdots\\0\\0\end{bmatrix}$$

for a solution given the specified linear constraint. We will use the symbol G_c to represent the generalized inverse associated with a particular constraint, rather than the more awkward notation $(X^T X)_c^-$. With the generalized inverse associated with a particular constraint in hand, we multiply $X^T Y$ in (3) by the generalized inverse to obtain *a solution*: $\hat{b}_c = G_c X^T Y$. We emphasize again that this solution is unique given the constraint, that is, G_c is the generalized inverse associated with the constraint and $\hat{b}_c$is vector of parameter estimates for the equation given the constraint.

When the CGLM is used, the constraint typically involves setting two of the coefficients associated with age, period, or cohort to be equal and the choice of this constraint determines a unique generalized inverse that is used to solve for the age-group, period, and cohort coefficients (Mazumdar et al. 1980). The assumption is that $c^T \boldsymbol{\beta} = 0$, where c is the $m \times 1$ vector for the constraint and $\boldsymbol{\beta}$ is the $m \times 1$ vector of population effect coefficients. To the extent that $c^T \boldsymbol{\beta} = 0$ is not true, the estimates will be biased (see Kupper et al. 1983 for an explication of the bias associated with violating this assumption).

In the case of the IE, recently proposed by Yang et al. (2004), the constraint involves v (an $m \times 1$ vector), with the assumption that $v^T \beta = 0$. Specifically, $\boldsymbol{v}$ is the null vector (the vector of coefficients that when multiplied times the columns of the $\boldsymbol{X}$-matrix results in the zero vector). To the extent that this assumption is not true, the estimates associated with this constraint will be biased (again Kupper et al. 1983 explicate the degree of bias associated with violating this assumption). The generalized inverse that is associated with this the constraint used by Yang et al. (2008) is the Moore-Penrose generalized inverse (Mazumdar et al. 1980).

We label the identification problem an X-side problem because involves the X-matrix and constraining the columns of that matrix to provide a unique solution. In theory, one does not have to be concerned with the patterning of the response variable, Y, in making such a constraint. Some would, in fact, discourage such an investigation of the patterning of the response variable as data snooping.

4 Confounding: variance decomposition, deviations from linearity, and cohort characteristics

There are a number of approaches to APC analysis that depend upon estimable functions (functions that are identified). These approaches avoid the identification problem by not focusing on the estimation of the specific age, period, and cohort coefficients (these specific coefficients are not estimable without the imposition of a linear constraint on the X-matrix). Some important information about age, period, and cohort effects can be gained from these estimable functions. For example, we can partition the unique variance in the age–period-specific dependent variable to cohort effects, period effects, or age-group effects, without placing special constraints on the solution space. We can estimate the deviations of the age, period, and cohort effect coefficients from linearity and can estimate the second differences in these effect coefficients. Coming, admittedly, out of a different tradition, one or more cohort characteristics can be used as a proxies for the cohort effects and provide identified effects for the age effects, period effects, and cohort characteristic effect. All of these approaches share the characteristic that the effects they estimate are identifiable. They also illustrate the problems of the complete confounding of linear effects when identification is not a problem.

4.1 Variance decomposition

Although the individual age, period, and cohort effect coefficients are not estimable in the APC model because the X-matrix is singular; the least squares predicted values for the dependent variable values are estimable functions of the age, period, and cohort dummy variables (Searle 1971; Kupper et al. 1985; O'Brien and Stockard 2009). No matter which generalized inverse is used (which linear constraint we place on the X matrix); we obtain the same estimated values for the age–period-specific rates. Each generalized inverse uses all of the information available in the X variables to predict the Y values, since one of the X variables provides only redundant information. These predicted values allow the researcher to estimate the total variance attributable to age-groups, periods, and cohorts. Then the variance attributable uniquely to one of the factors is this total variance minus the variance attributable to the other two factors (a value found by using just two of the sets of dummy variables in the regression). Using significance tests allows us to determine whether age-groups, periods, and cohorts each make a significant unique contribution to the dependent variable (O'Brien and Stockard 2009).

This variance decomposition approach is a powerful technique for establishing whether all three factors (the set of age dummy variables, period dummy variables, and cohort dummy variables) are needed to model the dependent variable values (typically age–period-specific rates or counts). Using a variety of dependent variables from our own and others published research, we have found that there are unique effects associated with age-groups, periods, and cohorts in most of these data sets. This answers an important question about whether these three factors are each independently associated with the dependent variable or whether any apparent association *might* due to their being confounded with the other two factors.

With *complete* linear confounding, however, *all* of the linear effects of one factor are associated with the effects of the other two sets of dummy variables. For example, in the case of the pure linear effects of cohorts all of the variance associated with cohorts can be explained by the age-group and period dummy variables. In a regression context, if we use the age and period dummy variables as the independent variables they explain 100% of the variance associated with cohorts. A naïve analyst might conclude that there are no cohort effects—but more appropriately would conclude that there are no effects of cohorts that are linearly independent of the age and period dummy variables. In this situation, we might reasonably conclude that there are either no cohort effects or only pure linear cohort effects. The complete linear confounding does not allow us to distinguish between these two possibilities. Similar statements can be made about the pure linear effects of age and of period.

Not surprisingly, effects that are not purely linear—but have a linear relationship with time are also confounded, although not completely confounded. This can be seen by creating a partially linear cohort effect that is not purely linear. If we set the earliest cohort value to 4.5 and the next three oldest cohort values to 4.5, then cohorts 1–4 have the same values and the cohort effect is not purely linear. This corresponds to the entries that comprise the bottom cell entries in Table 2. Using regression analysis with this dependent variable, the cohort dummy variables are associated with 100% of the variance, as they must be for this cohort generated data. But if we use the age dummy variables or the period dummy variables alone each set is associated with 48.3% of the variance, and if we use both sets together they are associated with 96.6% of the variance. Here the variance uniquely associated with cohorts is only 3.4% of the variance associated with age-groups, periods, and cohorts. This confounding of linear effects (the first fundamental problem) makes it difficult to attribute variance to age, period, or cohort sets of dummy variables.

4.2 Deviations from linearity and second differences

Another set of approaches utilizing estimable functions and depending on non-linear effects involves estimating the deviations from linearity of the age-group, period, and cohort effects. Holford (1983, 1985) estimates these deviations from linearity of the age-group, period, and cohort coefficients by controlling for the linear trends of age, period, and cohorts and using orthogonal polynomials to estimate the deviations of the individual age, period, and cohort effects from linearity. These deviations are estimable and allow us to gauge the shape of the deviations of (for example) cohorts from linearity. Importantly, however, this is not at all the same as knowing the trajectory of cohort effects. For example, the deviations from linearity of cohorts might be well represented by a shallow inverted U-shape. But if the inestimable linear cohort trend rose steeply over time, the inverted dish shape deviations from linearity could represent a rise throughout the series in the cohort effects that is less steep with time. Clayton and Schifflers (1987) show that the second differences of the age, period, and cohort coefficients are estimable; that is, we can determine whether the rate of change in changes is increasing or decreasing. Is the slope getting steeper or less steep over time?

The estimates of deviations from linearity and changes in the rate of change in the effect coefficients, however, depend on nonlinear effects in order to detect the effects of deviations or changes in the rate of change. By definition the linear effects of cohorts, periods, or age-groups will not be detected by these methods.

4.3 Cohort characteristics

There is a long tradition in sociology of using cohort characteristics to avoid the identification problem and to potentially explain the mechanism behind the cohort effect (e.g., Kahn and Mason 1987; Heckman and Robb 1985; O'Brien et al. 1999). This approach uses a variable that is tied to cohorts as a proxy for cohort effects. For example, the researcher may theorize that the relative cohort size or the proportion of the cohort growing up in single parent families is directly related to the cohort's suicide rates or trust in government on opinion polls when controlling for age and period. Since the cohort characteristic is not linearly dependent on the age and period dummy variables, the identification problem is resolved.[6] The problem of linear confounding, however, affects this approach in much the same way as it does the approaches of variance decomposition and deviations from linearity. Since this method includes the age-group and period dummy variables any linear effects due to cohort time of birth are controlled for. Thus the cohort characteristics cannot detect linear effects of cohort, but only deviations from linearity.

The variance decomposition approach is based on attributing unique variance to age-groups or to periods or to cohorts that are not associated with the other two factors. The

[6] One can use a period characteristic or age characteristic, if a plausible one is available: Farkas (1977) provides a nice analysis with a period characteristic.

approach can only detect effects of cohorts, or periods, or age-groups that are not linearly associated with time (of birth, period, or cohort year of birth). Any linear effects of cohorts with time will be "absorbed" by age and period, any linear effect of period with time will be absorbed by age and cohort, and any linear effect of age-groups with time will be absorbed by period and cohort. The other approaches described in this section also depend upon deviations from linearity to find significant effects. Otherwise, the only conclusions that can be reached using these approaches are that there are *no non-linear effects*.

5 Estimating age, period, and cohort coefficients: CGLM and IE approaches

The previous section demonstrated how a linear relationship between time and cohorts, or time and periods, or aging and age-groups can make the separation of age, period and cohort effects in terms of variance attributed to each of these components problematic even when the these components are estimable. The same is true of techniques that examine deviations from linearity or that use cohort characteristics to examine systematic differences between cohorts on the dependent variable. The goals of the techniques discussed in the previous section are modest in comparison to the CGLM and IE approaches that are designed to estimate each of the age effect, period effect, and cohort effect coefficients: the specific effects of each age-group, each period, and each cohort, not just their deviations from linearity or the unique variance accounted for by each of these sets of dummy variables.

The CGLM and IE approaches are both based on placing a linear restriction on the columns of the rank deficient X-matrix. In each case, this constraint is associated with a generalized inverse and that generalized inverse provides a unique solution to the age, period, and cohort coefficients given that constraint. The generalized inverse associated with the IE estimator, the Moore-Penrose generalized inverse, has some special technical properties that make it arguably the best choice for a generalized inverse in the absence of other information.[7] If for some reason a researcher had outside information that "assured" her that two of the coefficients among the age, period, and cohort coefficients were equal in the population; then the choice of the CGLM approach might well be preferable.

Below, we show some results using both of these approaches to the identification of the effect coefficients (the X-side problem) when the data are completely confounded due to the pattern of the Y-side response variable. We then provide an example where this linear confounding is not so severe. For the first example, we examine APC data generated by the pure linear cohort effect represented by the bolded-italicized elements in Table 2. Beginning with the CGLM approach we first fix the effects of age 2 and age 3 to be equal to each other.[8] Table 4 presents the results of this analysis in the first column (no error).

In this situation the CGLM returns the effect coefficient estimates that reflect the cohort generating process; after all, the constraint is correct (in terms of how the data were generated

[7] The Moore-Penrose generalized inverse (sometimes referred to as the pseudo-inverse) has all of the following properties for matrices consisting of real numbers: $AA^{-}A = A$; $A^{-}AA^{-} = A$; $(AA^{-})^{T} = AA^{-}$; and $(A^{-}A)^{T} = A^{-}A$, where A^{-} is the Moore–Penrose generalized inverse of A. These properties do not protect the solutions that it produces from often producing biased estimates of the data generating parameters, as we will see.

[8] We use constrained linear regression in STATA (cnsreg) to estimate the models with a single constraint that just identifies the model. The results based on these models do not differ substantively from the results based on using a regular least squares regression program (regress in STATA) to compute the results for the constrained linear models. When using regular regression, we need to set two of the age coefficients or two of the period coefficients or two of the cohort coefficients equal to zero. Making one of the variables constrained to be equal be the reference category.

Table 4 Analyses using the pure linear cohort effect of Table 2

	a2 = a3			c2 = c3			c4 = p2			Intrinsic estimator			Null vector
	No error	Simulation with error		No error	Simulation with error		No error	Simulation with error		No error	Simulation with error		
	b	b	SE	b	b	SE	b	b	SE	b	b	SE	
a1	0.000*	0.000*		0.000*	0.000*		0.000*	0.000*		0.786	0.785	0.098	−0.267
a2	0.000	0.001	0.262	−0.500	−0.502	0.269	−0.375	−0.370	0.189	0.393	0.398	0.093	−0.134
a3	0.000	0.001	0.262	−1.000	−1.015	0.482	−0.750	−0.742	0.259	0.000	−0.004	0.096	0.000
a4	0.000	−0.006	0.473	−1.500	−1.522	0.664	−1.125	−1.122	0.336	−0.393	−0.390	0.095	0.134
a5	0.000	−0.004	0.621	−2.000	−2.029	0.874	−1.500	−1.491	0.378	−0.786	−0.788	0.092	
p1	0.000*	0.000*		0.000*	0.000*		0.000*	0.371	0.117	−0.786	−0.786	0.087	0.267
p2	0.000	−0.001	0.219	0.500	0.510	0.263	0.375	0.745	0.178	−0.393	−0.393	0.096	0.134
p3	0.000	0.002	0.347	1.000	1.021	0.432	0.750	0.000*		0.000	0.000	0.096	0.000
p4	0.000	0.013	0.475	1.500	1.522	0.658	1.125	1.128	0.251	0.393	0.394	0.091	−0.134
p5	0.000	0.011	0.621	2.000	2.038	0.874	1.500	1.498	0.349	0.786	0.784	0.103	
c1	0.000*	0.000*		0.000*	0.000*		0.000*	0.119	0.247	−0.429	−0.425	0.178	−0.535
c2	0.500	0.491	0.342	0.000	−0.021	0.434	0.125	0.256	0.196	−0.321	−0.323	0.150	−0.401
c3	1.000	0.999	0.447	0.000	−0.021	0.434	0.250	0.371	0.117	−0.214	−0.215	0.138	−0.267
c4	1.500	1.486	0.544	0.000	−0.026	0.693	0.375	0.502	0.223	−0.107	−0.112	0.126	−0.134
c5	2.000	1.989	0.673	0.000	−0.031	0.895	0.500	0.000*		0.000	−0.001	0.105	0.000
c6	2.500	2.486	0.822	0.000	−0.042	1.104	0.625	0.627	0.279	0.107	0.107	0.117	0.134
c7	3.000	2.980	0.964	0.000	−0.052	1.309	0.750	0.749	0.379	0.214	0.215	0.123	0.267
c8	3.500	3.484	1.072	0.000	−0.059	1.538	0.875	0.881	0.483	0.321	0.322	0.132	0.401
c9	4.000	3.984	1.268	0.000	−0.062	1.765	1.000	1.010	0.617	0.429	0.432	0.241	
Cons	3.000	3.009	0.663	5.000	5.028	0.903	4.500	4.496	0.300	5.000	5.002	0.056	0.000

with age 2 and age 3 being equal). The next two columns present results after we have added a random error component to these data. That random component is from a normal distribution and is set to be equal to 5.0% of the total variance of the age–period-specific response variable.[9] We ran 1,000 such simulations and report the results in terms of the mean effect coefficients and mean standard errors of these effects. Not surprisingly, since the errors are random and the results were averaged over 1,000 simulations, the effect coefficients for the linear cohort data both with and without the error are nearly identical. The mean standard errors show that none of the age or period coefficients would typically be statistically significantly different from zero. In general, no matter which of the two age or period coefficients are set equal to each other, we obtain the same results for the data with no error—the age and period effects are zero and the cohort effects are the same as those based on the cohort data generating process. When we add error to the process, we find substantively similar results.

When we set two of the cohort coefficients equal to each other, however, we obtain estimates that are not in line with how the data were putatively generated. This makes sense because in the data generating process none of the cohort effects were the same. Table 4 displays the effect coefficients obtain when we set c2 and c3 to be equal. The cohort effects are all zero while the age effects decrease by 0.50 with age and the period effects increase by 0.50 with time. Again in the no error case 100% of the variation in the dependent variables is associated with the age, period, and cohort coefficients, so we do not report standard errors for these coefficients. Because of the complete confounding of the linear effects of cohort, we know that even if the cohort effects are estimated to be zero that any linear effect of cohorts have been absorbed by the age and period effect coefficients.

In the next column we again present effect coefficient estimates when we have added error variance to the data. In the data simulation with error, the average estimated value of the coefficients for age and period are often twice the size of their standard errors and the estimated cohort coefficients are close to zero. A researcher faced with such results might well conclude that the cohort effects are near zero and the period and age-group effects are substantively important. This result hinges on setting two of the cohort effect coefficients equal to each other. Although this is incorrect given the data generating process, it would be correct if these same effects were generated entirely by age and period. That is, since pure linear effects of cohorts can be fully explained by age and period dummy variables, the data generating process could have been based solely on age and period effects and it would be appropriate to fix two of the cohort effects equal to each other.

The third example in Table 4 shows what happens when we constrain c4 to equal p2. We choose this constraint because the observed values for cohort 4 are all equal to 4.5 and the average effect for period 2 is 4.5 in Table 2. In this sense, it might appear from the data that these two effects are the same in the population. Using this constraint, we find a pattern of estimates in which the pure linear effects for cohorts of the generating process are attributed

[9] In our simulations the expected value of the corrected R^2 is 0.950: $\tilde{R}^2$ =var(T)/[var(T) + var(e)], where var(T) is the variance of the response variable without error and var(e) is the variance of the random error term. This means that the expected value of the uncorrected R^2 is 0.983 (this occurs because there are 25 age–period-specific rates to be predicted with 16 independent variables). We choose this level of R^2 as a reasonable one given some other data sets—but certainly it will vary from data set to data set. For example, the age–period data on female deaths in Table 1 of Yang et al. (2008) has an uncorrected R^2 of greater than 0.99; the verbal test data in their Table 5 has an uncorrected R^2 of 0.91; and the homicide data in McCall and Land (2004) has an uncorrected R^2 of 0.98. Kupper et al. (1985, p. 820) note: "[I]t is typically the case when fitting by least squares the multiple classification model…to APC data that R^2 values extremely close to 1 are obtained." Obviously the amount of random error introduced will affect which effect coefficients are "statistically significant" in our simulations—but it will not affect the expected values of the effects.

Table 5 Analyses using the partial linear cohort effect of Table 2

	a2 = a3			c4 = c5			c6 = p4			Intrinsic estimator			Null vector
	No error	Simulation with error		No error	Simulation with error		No error	Simulation with error		No error	Simulation with error		
	b	b	SE	b	b	SE	b	b	SE	b	b	SE	
a1	0.000*	0.000*		0.000*	0.000*		0.000*	0.000*		0.476	0.475	0.074	−0.267
a2	0.000	0.001	0.196	−0.500	−0.501	0.154	−0.125	−0.121	0.124	0.238	0.242	0.069	−0.134
a3	0.000	0.001	0.196	−1.000	−1.003	0.249	−0.250	−0.244	0.144	0.000	−0.003	0.072	0.000
a4	0.000	−0.005	0.354	−1.500	−1.511	0.358	−0.375	−0.372	0.165	−0.238	−0.236	0.071	0.134
a5	0.000	−0.003	0.465	−2.000	−2.011	0.473	−0.500	−0.493	0.161	−0.476	−0.478	0.069	
p1	0.000*	0.000*		0.000*	0.000*		0.000*	0.000*		−0.476	−0.476	0.065	0.267
p2	0.000	0.000	0.164	0.500	0.502	0.167	0.125	0.122	0.103	−0.238	−0.238	0.072	0.134
p3	0.000	0.002	0.260	1.000	1.006	0.267	0.250	0.247	0.104	0.000	0.000	0.072	0.000
p4	0.000	0.010	0.356	1.500	1.516	0.376	0.375	0.377	0.093	0.238	0.239	0.068	−0.134
p5	0.000	0.009	0.465	2.000	2.016	0.473	0.500	0.498	0.141	0.476	0.475	0.077	
c1	0.000*	0.000*		0.000*	0.000*		0.000*	0.000*		0.119	0.122	0.133	−0.535
c2	0.000	−0.007	0.256	−0.500	−0.509	0.258	−0.125	−0.129	0.206	−0.119	−0.120	0.112	−0.401
c3	0.000	0.000	0.334	−1.000	−1.004	0.328	−0.250	−0.245	0.178	−0.357	−0.358	0.104	−0.267
c4	0.000	−0.011	0.407	−1.500	−1.516	0.471	−0.375	−0.378	0.150	−0.595	−0.599	0.094	−0.134
c5	0.500	0.491	0.504	−1.500	−1.516	0.471	0.000	0.002	0.132	−0.333	−0.334	0.079	0.000
c6	1.000	0.989	0.615	−1.500	−1.521	0.634	0.375	0.377	0.093	−0.071	−0.072	0.087	0.134
c7	1.500	1.484	0.721	−1.500	−1.528	0.746	0.750	0.750	0.146	0.190	0.191	0.092	0.267
c8	2.000	1.988	0.802	−1.500	−1.526	0.857	1.125	1.131	0.175	0.452	0.453	0.099	0.401
c9	2.500	2.488	0.949	−1.500	−1.528	0.965	1.500	1.509	0.261	0.714	0.717	0.180	
Cons	4.500	4.507	0.496	6.500	6.515	0.504	5.000	4.996	0.161	5.333	5.335	0.042	0.000

to age-groups, periods, and cohorts. After adding a random error component to the model, many of the coefficients for ages, periods, and cohorts are (on average) significantly different from each other.

The final set of results in Table 4 use the IE approach to estimate the effect coefficients for the data generated by the pure linear effect of cohorts in Table 2. Whereas we can have a variety of estimates depending upon which constraint we choose in the CGLM approach, this is not the case with the IE approach. Here only one constraint resulting in a single generalized inverse is appropriate.[10] The constraint produces a solution that is orthogonal to null vector in the null space of the X-matrix. This single null vector reflects the fact that the X-matrix has a rank that is just one less than being of full column rank. There is a long history of some advocacy for the generalized inverse that is associated with this constraint being a good choice of a generalized inverse when there are no theoretically or empirically compelling reasons to use a different generalized inverse. Not surprisingly, under this constraint, the IE procedure distributes the effects of the pure linear cohort effect across the ages, periods, and cohorts. There is no way for the IE approach to determine how these data were generated: that is, as a pure linear effect of cohorts or as the linear effects of periods and age groups. The IE requires that the solution vector be orthogonal to the null vector and it is. We can test to see if this constraint is achieved by taking the transpose of the null vector (last column in Table 4) and multiplying this times the solution vector (the vector of estimated coefficients in Table 4).[11] The result is zero (as it should be for orthogonal vectors).

The next two columns report the mean estimated effect coefficients and the mean of the standard errors after adding random error to the dependent variable. Since these are based on 1,000 simulated cases, it is not surprising that the average of the estimated effect coefficients is quite close to those without error. Substantively many of the average effects of the age, period, and cohort coefficients are two or more standard errors (on average) greater than zero and, thus, would typically be statistically significant.

We could present similar analyses for the pure linear effects due to age groups and the pure linear effects due to periods displayed in Table 2—but little would be gained. Whether the CGLM provides estimates that are correct in terms of the “generating process,” depends on whether the constraint is consistent or inconsistent with that process. When the constraint is inconsistent then the estimates will not be in line with the generating mechanism. Some constraints will give results that attribute all of the effects to just one factor (age or period or cohort coefficients) and others will attribute all of the effects to the other two factors. This extreme instability is due to the complete confounding of the linear effects, but instability is always a potential when the researcher sets substantively different constraints.

Although we generated the first three data sets in Table 2 as pure linear effects of cohorts, pure linear effects of age-groups, and pure linear effect of periods, respectively, a different researcher might have generated them as effects of the other two factors. There is no way to distinguish between these quite different explanations when confronted with pure linear

[10] The constraint means that the solution must be orthogonal to the null vector, and we have placed the null vector as the last column of both Tables 4 and 5. Since this vector depends on the X-matrix only, it is the same for both tables. The null vector consists of the b coefficients that multiply the corresponding column vectors to produce the zero vector in Table 3. This null vector is based on the dummy variable coding used by Yang et al. (2008), where the reference variables are coded as minus 1 rather than as zero. We could create a similar constraint for conventional dummy variable coding that would not change the substantive results. We use their coding since we use their program to produce the estimates presented in this paper. The program is an add on file for calculating the IE in STATA (cited in Yang et al. 2008).

[11] Given the coding used by Yang et al. (2008), which does not set the reference category to zero, we substitute zeros into the null vector for the reference categories (age5, period5, and cohort9) before checking the solution for orthogonality.

effects. We note also that there is no empirical way to differentiate these various solutions in terms of explained variance. Each of the models (without error) in Table 4 account for all of the variance in the dependent variable. In the models with error, the expected value of the adjusted R^2 for each model is 0.95. If a researcher chooses to use a different form of analysis, such as a Poisson or Logistic Regression, the results will not distinguish these competing models in terms of model fit (e.g., -2 times the log-likelihood, the Bayesian Information Criterion, or Akaike's Information Criterion).

The pure linear effects simulations are the "worst situations" in terms of the confounding of linear effects, because they involve those effects that are completely confounded, but partial linear effects also cause confounding. We turn to the partial linear confounding depicted in the bottom row of the cell entries in Table 5, where cohorts 1–4 have the same effects (4.5) and the remainder of the cohorts have a linear increase of 0.5 from cohort 5 to cohort 9.

The results in Table 5 are based on this partial linear cohort effects data displayed as the bottom cell entries in Table 2. The first three sets of results in Table 5 shows how selected equality constraints in the CGLM approach produce different estimated parameters for the age, period, and cohort effects. When we set constraints in such a way that they are consistent with the data generating process for age or period (for example a2 equal to a3 in Table 5), we obtain coefficients consistent with the data generating equation. The results using this constraint are the first set of results in Table 5. The effect coefficients for the age and period dummy variables are all zero. Taking the constant into account the predicted values for cohorts 1–4 are all 4.50 and then the cohort effects coefficients add an additional 0.5 for each additional increase in cohort year of birth. The simulated data with error follow this pattern and many of the cohort effect coefficients are (on average) significantly different from one another. The correct estimation of the effect coefficients occurs no matter which age coefficients are set equal to each other or which period coefficients are set equal to each other. When we set c1 equal to c2 or c3 or c4; c2 equal to c3 or c4; c3 equal to c4, which are each consistent with the generating process, we again obtain estimates that are consistent with the data generating process. These results are not reported in Table 5.

The next set of results in Table 5 show what happens when we set an equality constraint on two of the cohort coefficients that are not equal in the data generating process: here, c4 equals c5. We obtain coefficients for cohorts that are equal for all of those that increased by 0.5 in the data generating process and cohort effect coefficients that differ by 0.5 for all of those that were equal in the data generating equation. Additionally, we obtain increasing effects for periods with time and decreasing effect for age-groups with age. The simulations with an error component show that many of these coefficients would typically be statistically significantly different from one another. Importantly, we see a dramatic swing in the effect coefficients when setting a2 equal to a3, a constraint that is consistent the generating process, and when setting c4 equal to c5, a constraint that is inconsistent with the generating process. In the former case, the "action" is in the cohort effect coefficients; in the latter case, the "action" resides mainly with differences in the effect coefficients for age-groups and periods, but also in the first four cohort coefficients. We can expect such wildly divergent results when we have nearly linear effects of age-groups or periods or cohorts.

Setting p4 (which has a mean value of 5.5) and c6 (which has all 5.5 entries) equal to each other, might appeal again to someone who is trying to set the constraint by examining the data. Using this constraint we obtain estimates that do not reflect the data generating process, even though setting this equality is in some sense consistent with the data. For the data with error many of the age, period, and cohort effect coefficients would be (on average) twice their standard errors and would typically be statistically significant. Further, as with the other constraints, there is no way to empirically determine that the data were not generated by this process. For any of these just identified models, which result by placing a single constraint on the data, each model explains the same amount of variance in the dependent variable no matter what constraint is used.

The final set of results in Table 5 are based on the IE approach to identifying the model. The results using this method distribute the partial linear cohort trend effect among all of the coefficients for age-groups, periods, and cohorts. It produces a steady decrease in the cohort effects until cohort 5 and then an upturn in cohort effects for the remainder of the cohorts as they become increasingly more recent. The effect coefficients decrease with age (most of them would be significantly different from each other in the simulations with error) and the period coefficients increase with time (most of them would be statistically significantly different from each other in the simulations with error). These coefficients do not reflect the putative generating procedure. They are the solutions that are constrained to be orthogonal to the null vector, which can again be verified by premultiplying the transpose of the null vector (the last column in Tables 4 or 5) times this solution vector.

6 Conclusions

There are two fundamental problems in APC analysis. The Y-side problem involves the complete confounding of the linear effects of cohort with age and period; age with period and cohort; and period with age and cohort. This problem results from the response variable and the specific patterning of linear effects in any one of the factors (age, period, or cohort). This structure is well illustrated in the age–period matrix with pure linear effect generated by cohort or age or period (see Table 2) and affects approaches that do not depend on identification of the individual effect coefficients. The X-side problem has to do with identification. This problem involves the existence an infinite number of solutions for the age, period, and cohort effect coefficients when no special constraints are placed on the model. No matter what the pattern of the response, an infinite number of solutions for the effect coefficients exists.

The pure linear effects depicted in Table 2 contain some of the most difficult patterns of the response variables to model using any of a variety of APC methods. We choose to illustrate the difficulties encountered in modeling such effects using cohort generated effects. These linear effects play havoc with methods that are designed to examine the variance uniquely attributable to the three sets of dummy variables: age, period, and cohort and methods designed to estimate deviations of age, period, and cohort effects from linearity. While these approaches do not suffer from problems of underidentification, the completed confounding of the linear effects of cohort with the effects of period and age, the linear effects of periods with the effects of age and cohorts, and the linear effects of age with period and cohort, make these procedures far less useful than when such effects are absent or, more realistically, at least not so great.

Approaches that attempt to estimate the individual effect coefficients for age-groups, periods, and cohorts are much more ambitious than those seeking to apportion unique variance accounted for or estimating deviations from linearity. They confront the identification problem. The approaches we examined try to solve this problem by placing *a* constraint on the solution space. If the constraint is consistent with the data generating process then the results will mirror the data generating process. If not the results will not mirror the data generating process. To the extent that the constraint, represented by the vector c' times the vector of population effect coefficients is not zero ($c'\beta \neq 0$), then the estimates of the effect coefficients are biased (see Kupper et al. 1983 for an explication of the bias associated with violating this assumption).

Based on the results from our analyses using data that were generated by a pure linear cohort effect in Table 4 and a partial linear cohort effect in Table 5, we find that if we set the constraint in a way that is consistent with the generating process we obtain coefficients that are consistent with the generating process. When we set the constraint in a manner inconsistent with the generating process we obtain coefficients that are inconsistent with the generating process.[12] This is true for both the CGLM and IE approaches.

References

Clayton, D., Schifflers, E.: Models for temporal variation in cancer rates II: age–period–cohort models.. Stat. Med. **6**, 469–481 (1987)

Farkas, G.: Cohort, age, and period effects upon the employment of white females: evidence for 1957–1968. Demography **14**, 33–42 (1977)

Greenberg, B.G., Wright, J.J., Sheps, C.G.: A technique for analyzing some factors affecting the incidence of syphilis. J. Am. Stat. Assoc. **45**, 373–399 (1950)

Heckman, J., Robb, R.: Using longitudinal data to estimate age, period, and cohort effects in earning equations. In: William Mason, M., Stephen Fienberg, E. (eds.) Cohort Analysis in Social Research: Beyond the Identification Problem, pp. 137–150. Springer, New York (1985)

Hobcraft, J., Menken, J., Preston, S.: Age, period, and cohort as sources of variation in demography. Popul. Index **48**, 4–43 (1982)

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

Holford, T.R.: An alternative approach to statistical age–period–cohort analysis. J. Chronic Dis. **38**, 831–836 (1985)

Kahn, J.R., Mason, W.M.: Political alienation, cohort size, and the Easterlin hypothesis. Am. Sociol. Rev. **52**, 155–169 (1987)

Kupper, L.L., Janis, J.M., Salama, I.A., Yoshizawa, C.N., Greenberg, B.G.: Age–period–cohort analysis: an illustration of the problems in assessing interaction in one observation per cell data. Commun. Stat. Theory Method **12**, 2779–2807 (1983)

Kupper, L.L., Janis, J.M., Karmous, A., Greenberg, B.G.: Statistical age–period–cohort analysis: a review and critique. J. Chronic Dis. **38**, 811–830 (1985)

Mason, K.O., Winsborough, H.H., Mason, W.M., Poole, W.K.: Some methodological issues in cohort analysis of archival data. Am. Sociol. Rev. **38**, 242–258 (1973)

Mazumdar, S., Li, C.C., Bryce, G.R.: Correspondence between a linear restriction and a generalized inverse in linear model analysis. Am. Stat. **34**, 103–105 (1980)

McCall, P.L., Land, K.C.: Trends in environmental lead exposure and troubled youth, 1960–1995: an age–period–cohort-characteristic analysis. Soc. Sci. Res. **33**, 339–359 (2004)

O'Brien, R.M., Stockard, J., Isaacson, L.: The enduring effects of cohort characteristics on age-specific homicide rates, 1960–1995. Am. Sociol. Rev. **104**, 1061–1095 (1999)

O'Brien, R.M., Stockard, J.: Can cohort replacement explain changes in the relationship between age and homicide offending?. J. Quant. Criminol. **25**, 79–101 (2009)

Searle, S.R.: Linear Models. Wiley, New York (1971)

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

Yang, Y., Schulehoffer-Wohl, S., Fu, W.J., Land, K.C.: The intrinsic estimator for age–period–cohort analysis: what it is and how to use it. Am. J. Sociol. **113**, 1697–1736 (2008)

[12] The "putative data generating process" and other ways in which the data could have been generated can be at odds.