Commentary and Debate

To conserve space for the publication of original contributions to scholarship, the comments in this section must be limited to brief critiques; author replies must be concise as well. Comments are expected to address specific substantive errors or flaws in articles published in *AJS*. They are subject to editorial board approval and peer review. Only succinct and substantive commentary will be considered; longer or less focused papers should be submitted as articles in their own right. *AJS* does not publish rebuttals to author replies.

THE SENSITIVITY OF THE INTRINSIC ESTIMATOR TO CODING SCHEMES: COMMENT ON YANG, SCHULHOFER-WOHL, FU, AND LAND[1]

INTRODUCTION

In a series of articles, Wenjiang Fu, Yang Yang, and Kenneth Land (Fu 2000, 2008; Yang, Fu, and Land 2004; Yang, Schulhofer-Wohl, Fu, and Land 2008; Yang 2008) proposed the intrinsic estimator (IE) and argued that it is a general-purpose, robust, reliable, and useful tool for estimating age-period-cohort (APC) and similar models, where identification and estimation are deeply problematic because of exact linear dependence among the explanatory variables. For example, in their article "The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It" (*American Journal of Sociology* 113 [2008]:1697–1736), Yang, Schulhofer-Wohl, Fu, and Land describe the IE and how to use it to disentangle age, period, and cohort effects in empirical research. Using the General Social Survey (GSS) data as an example, they argue that the IE produces estimates that approximate well true age, period, and cohort trends (Yang et al. 2008, pp. 1712–16). They also use simulated data to argue that the IE performs better than the traditional constrained generalized linear model. They conclude that the IE can be used to produce reliable and useful estimates of the underlying independent effects

[1] This is truly a joint work. We thank Robert O'Brien, Tom Pullum, and John Robert Warren and for their helpful comments. Direct correspondence to Liying Luo, Department of Sociology and Criminology, Pennsylvania State University, 714 Oswald Tower, University Park, Pennsylvania 16802. E-mail: liyingluo@psu.edu

0002-9602/2016/12203-0007$10.00

of age, period, and cohort in APC models (Yang et al. 2008, pp. 1716–22). The IE now enjoys wide popularity in many disciplines and has been used in multiple empirical applications (e.g., Clark and Einsenstein 2013; Masters et al. 2014; Schwadel and Stout 2012; Schwadel 2011; Yang 2008).

O'Brien (2011) and Luo (2013*a*, 2013*b*) raise questions about whether the IE is in fact a useful method for estimating the true effects of age, period, and cohort. In particular, they show that, like other APC estimators, the IE involves a constraint that is essentially arbitrary and that it performs badly when the situation does not satisfy this constraint.

In this comment, we raise additional concerns about the robustness (i.e., sensitivity) and thus the usefulness of the IE. Specifically, we show that IE estimates can be highly sensitive to a researcher's choice of coding scheme or model parameterization. For example, suppose that two researchers begin with an identical age-by-period data array of rates, with age factor α_i for $i = 1, \ldots, a$ and period factor β_j for $j = 1, \ldots, p$. They construct the cohort factor γ_k for $k = j - i + a = 1, \ldots, a + p - 1$ from the identity cohort = period − age. Both researchers now wish to carry out an APC analysis on these data but have not settled on a common coding to adopt. One researcher adopts α_1 as the reference (i.e., the first level of the age factor is chosen as the omitted category), and the other adopts α_a as the omitted category (i.e., the last level of the age category as reference). The resulting dummy variable coding pertaining to the age factor is $\alpha_2, \ldots, \alpha_a$ (for researcher 1) and $\alpha_1, \ldots, \alpha_{a-1}$ (for researcher 2) accordingly. Both apply the same respective coding schemes to period, $\beta_2, \ldots, \beta_p$ and $\beta_1, \ldots, \beta_{p-1}$ and cohort, $\gamma_2, \ldots, \gamma_{a+p-1}$ and $\gamma_1, \ldots, \gamma_{a+p-2}$. Because of these different parameterizations of the design matrix, the two researchers will necessarily get different coefficient estimates for the same rate data. In an identified model, they could easily obtain *identical* estimates by transforming both results to a common parameterization such as one in which the effects sum to zero across the a levels of age, p levels of period, and $a + p - 1$ cohort categories. However, even after doing this, the IE estimates that the two researchers get will be different and will often, in fact, lead to opposite findings about the patterning of APC effects.

We reanalyze data from three published articles to demonstrate that coding the APC model using one coding scheme (e.g., the sum-to-zero/ANOVA coding) can give dramatically different results from those obtained using a different coding scheme (e.g., using a reference group). The results are so different that a researcher would reach opposite conclusions about the effects of age, period, and cohort depending on the choice of coding scheme.

INSENSITIVITY TO CODING SCHEMES IN IDENTIFIED MODELS

In most basic regression courses, students are taught that a regression model with categorical variables can be parameterized in different ways, for example, using one category as the reference or base group or alternatively using so-called sum-to-zero coding. They are then shown that these different coding schemes are equivalent in that (1) they give the same predicted or fitted values, (2) one set is easily transformed into another, and most important, (3) the substantive interpretation of estimated effects is invariant to the coding scheme. The last is true in a very precise sense: the age, period, and cohort effects estimated under the different coding schemes are *identical* if they are reexpressed in a common coding scheme. For example, if one coding scheme gives estimates showing an increasing period effect, the alternative coding scheme will also give an identical increasing period effect. Because of this equivalence, the typical advice is to choose that coding scheme within which it is easiest to interpret the results. In all other respects, the choice of coding scheme is innocuous in an identified model.

For example, consider an identified Poisson model that regresses the outcome—the U.S. female mortality rates from 1960 to 1999—on age and period groups. Table 1 reports the estimation results for this age-period model using the sum-to-zero coding with the last group omitted and also for coding schemes using the first group and last group as reference groups. Because the model is identified, these seemingly different effect estimates using different coding schemes are in fact identical after being transformed to a common scale. For example, the effect estimate for the 5–9 age group in the $\beta_{\text{first}} = 0$ column in table 1 can be calculated by subtracting the effect estimate for the 0–4 age group from that for 5–9 age group in column $\beta_{\text{last}} = 0$: $-2.453 = -6.631 - (-4.178)$. Similarly, to get the estimate for the 1990–94 period in the $\sum = 0$ column in table 1 using the estimates in column $\beta_{\text{first}} = 0$, we first calculate the average of the eight period effects shown in the $\beta_{\text{first}} = 0$ and then subtract the average from the estimate for 1990–94 period effect: $-0.413 - (-0.246) = -0.167$.

More formally, given a simple single-factor dummy variable model of the form

$$Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k D_{ik} + \varepsilon_i, \tag{1}$$

where D_{ik} is a dummy variable derived from a factor with K levels, and where any category k can be chosen as the reference (i.e., $\beta_1 = 0$), we can construct the sum-to-zero or centered effects parameterized model as

$$Y_i = (\beta_0 + \bar{\beta}) + \sum_{k=1}^{K} (\beta_k - \bar{\beta}) D_{ik} + \varepsilon_i = \alpha_0 + \sum_{k=1}^{K} \alpha_k D_{ik} + \varepsilon_i, \tag{2}$$

where $\bar{\beta} = \Sigma_{k=1}^{K} \beta_k / K$, and therefore $\Sigma_{k=1}^{K} \alpha_k = 0$. (Alternatively, the same results can be obtained using matrix transformations as described in the appendix.)

TABLE 1
ESTIMATED AGE AND PERIOD EFFECTS ON MORTALITY UNDER THREE CODING SCHEMES

Category/Effects	$\sum=0$	$\beta_{first}=0$	$\beta_{last}=0$
Intercept	−5.343	−5.539	−1.778
		Age	
0–4	−.442	.000	−4.178
5–9	−2.894	−2.453	−6.631
10–14	−2.989	−2.548	−6.726
15–19	−2.236	−1.794	−5.972
20–24	−2.059	−1.618	−5.796
25–29	−1.900	−1.459	−5.637
30–34	−1.607	−1.166	−5.344
35–39	−1.230	−.789	−4.966
40–44	−.812	−.371	−4.549
45–49	−.373	.068	−4.109
50–54	.061	.503	−3.675
55–59	.475	.916	−3.261
60–64	.903	1.345	−2.833
65–69	1.328	1.769	−2.409
70–74	1.779	2.220	−1.958
75–79	2.245	2.687	−1.491
80–84	2.754	3.195	−.983
85–89	3.263	3.704	−.473
90–94	3.736	4.178	.000
		Period	
1960–64	.246	.000	.416
1965–69	.196	−.050	.367
1970–74	.116	−.130	.287
1975–79	−.025	−.270	.146
1980–84	−.083	−.329	.088
1985–89	−.112	−.357	.059
1990–94	−.167	−.413	.004
1995–99	−.171	−.416	.000

NOTE.—Data are from Yang et al. (2004). $\sum=0$: sum-to-zero coding. $\beta_{first} = 0$: reference-group coding with the first group omitted for each effect. $\beta_{last} = 0$: reference-group coding with the last group omitted for each effect.

Because the APC model is not identified, the IE is sensitive to the coding scheme used in an analysis in the sense that different coding schemes (that would give identical results in an identified model after being transformed) can give estimates that imply dramatically different substantive conclusions. In addition, in the appendix, we show that IE will give *any* of the infinite number of solutions that fit the data equally well if the right coding scheme is chosen and IE is performed in that scheme. Therefore, there is no single IE solution, but rather a different IE solution for each coding scheme, and there will always be a coding scheme where the IE produces any of the infinite number of feasible solutions.

In this comment, we reanalyze data from published articles to demonstrate that estimated age, period, and cohort effects using one coding scheme (e.g., the sum-to-zero coding) can be dramatically different from estimated effects obtained using a different coding scheme (e.g., reference-group coding). This difference is particularly problematic because the choice of coding schemes should be completely arbitrary or innocuous, meaning that the resulting estimates should be *identical* after being transformed. The empirical examples below demonstrate that since the choice of coding scheme is arbitrary, any IE solution is arbitrary. Some might give preference to the sum-to-zero coding scheme, *but this choice is entirely conventional and without substantive foundation*. Besides, the sum-to-zero coding scheme itself requires an arbitrary choice of a category to omit and the IE solution is sensitive to that choice as well. Below, we provide a nontechnical explanation for this sensitivity; an appendix provides a mathematical proof.

THE INTRINSIC ESTIMATOR

The IE achieves identification in ways that are both similar to and different from more traditional approaches to estimation of APC models. Because age, period, and cohort are linear functions of each other, there are an infinite number of possible estimates for the APC model, all of which give identical fitted values for the response variable but which can give highly different estimates of age, period, and cohort effects (Fienberg and Mason 1979). Because all the possible estimates give the same fitted values, there is no way to use the data to choose among them. These estimates can be said to lie on a line, called the solution line, and each point on this line corresponds to one set of the infinitely many estimates. As such, if one fixes the value of one parameter estimate at any specific value, the values of all the other parameter estimates are then determined by the data. The problem in doing APC analysis is deciding which set of estimates—that is, which point on the solution line—to privilege. As O'Brien (2011) shows, the IE, like traditional APC estimators, imposes a particular constraint on the parameter estimates that determines which point along the solution line is privileged.

Traditional approaches to identifying APC models involve either setting some parameter(s) to zero, for example, assuming there is no period effect (e.g., see Alwin 1991; Glenn 1994), or setting two or more parameters to be equal, for example, setting adjacent cohorts or periods to have equal effects (e.g., Clark and Eisenstein 2013; Knoke and Hout 1974). The presumption is that such constraints should be based on theoretical assumptions, though in many cases the constraints appear to be *ad hoc* (Glenn 1976; Rodgers 1982).

Like traditional estimators, the IE also achieves identification by imposing a constraint (O'Brien 2011; Luo 2013*a*), but one defined using a different criterion. Specifically, the IE chooses that set of estimates on the solution line that has the smallest variance. (This criterion has a few equivalent forms, one of which is discussed just below.) Thus the IE uses a statistical rather than theoretical or substantive rationale to determine which set of estimates should be privileged.

We make two critical mathematical observations: First, choosing the set of estimates with the smallest variance is equivalent to choosing the set of estimates that gives the smallest value when the individual parameter estimates are squared and summed; that is, that set of estimates that is the shortest distance from the origin.[2] Second, the IE depends on the design matrix in two senses. One, for a given coding scheme (parameterization), the constraint implicit in the IE depends on the number of age and period (and thus cohort) categories (Kupper et al. 1985). Two, as we show below, even with a fixed number of age and period categories, the IE depends on the design matrix through the coding scheme that is used. This directly contradicts Yang et al.'s (2008) critical assertion that the IE is invariant to the choice of design matrix.

Following Glenn (2005, p. 20), Yang et al. (2008, p. 1699) argue that an APC analysis should be evaluated with respect to its ability to provide correct estimates more often than not, that is, to estimate the true parameter estimates or what O'Brien (2011) calls the data-generating parameters.[3] They conclude that the IE satisfies this criterion (Yang et al. 2008, p. 1732). Furthermore, they argue that the essential purpose of the IE is to remove the influence of the coding scheme, or in equivalent terms, the design matrix (p. 1707). Below we show that this is not the case and show in detail that the IE is in fact sensitive to the coding scheme, sometimes dramatically so. As such, there is no basis to Yang et al.'s (2008) claim, critical in the assessment of its robustness

[2] In articles in which IE is introduced, distance is defined for this purpose as Euclidean or L2 norm, as we have described. While this choice is consistent with using least squares as an estimation criterion, it is essentially arbitrary and could be replaced by other norms, e.g., the sum of the absolute values of individual parameter estimates or L1 norm, which gives estimates differing from IE estimates and is consistent with the estimation criterion in some robust estimation methods.

[3] Some users of the IE appear to believe that it gives unbiased estimates of the true or data-generating parameters (e.g., see Keyes and Miech 2013; Masters et al. 2014; Schwadel and Stout 2012; Schwadel 2011). This is false. The IE gives an unbiased estimate of the set of parameter values on the solution line that is closest to the origin. *All* constrained APC estimators give an unbiased estimate of some parameters. Thus the IE is not distinctive in this respect.

and desirability, that the IE removes the effect of the design matrix or, given this, that it provides good estimates of the parameters that have generated the data.[4]

EMPIRICAL EXAMPLES IN WHICH IE ESTIMATES CHANGE WITH CODING SCHEMES

In this section we demonstrate how IE estimates can change with coding schemes by considering three published empirical examples, including studies of mortality (Yang et al. 2004), vocabulary knowledge (Yang et al. 2008), and trust (Schwadel and Stout 2012). In each case, we show that the IE estimates change depending on which of three coding schemes is used. Specifically, we obtain the IE estimates using three different coding schemes, namely, the sum-to-zero/ANOVA coding, reference coding with the first group as the reference category, and reference coding with the last group as the reference category. Because estimates with different codings are necessarily different, we then transform these IE estimates to a common scale, as shown in the figures below, to allow direct comparisons.

When working with categorical data, researchers typically choose a coding scheme because of interpretability or because it highlights a particular result. The sum-to-zero/ANOVA and reference group coding schemes are most popular because of their interpretability, though in principle an infinite number of coding schemes exist. As discussed above, in fully identified models, the choice of coding scheme does not affect estimation results when those results are reexpressed in the same coding scheme. In other words, the parameter estimates are unaffected by the coding—by mathematical necessity, they must be *identical*. As our empirical analyses show, in the case of unidentified models like the APC model, this is not the case.

Example 1

The first example is mortality rates for U.S. females from 1960 to 1999, used in Yang et al. (2004). These authors found that mortality rates increase after age 15, increased in the 1960s and early 1970s and rose again from 1980 to 1999, and decreased steadily across cohorts (Yang et al. 2004, p. 98). We rep-

[4] The IE can be understood as a type of ridge regression estimator (Fu 2000). However, when using dummy variables, the ridge estimator, like the IE, will be sensitive, potentially seriously so, to the coding scheme chosen. Yang et al. (2008, p. 1707) describe the IE as a type of principal component estimator. When the dimension of a factor space is two or greater, there are identification issues that principal components does not solve analogous to those in APC models. Principal components can discover the subspace in which the data lie, but it cannot determine what the axes of that subspace should be.

licated their estimates for age, period, and cohort effects using the sum-to-zero/ANOVA coding. In appendix table A1, the $\sum=0$ columns show their estimates, with each estimate expressed as the difference from the global mean associated with an age, period, or cohort group. We then obtained IE estimates using a reference coding with the first age, first period, and first cohort category as reference groups, and we also computed the IE estimates using a reference coding with the last age, last period, and last cohort category as reference groups. Finally, to allow direct comparison of the two reference group analyses to the sum-to-zero scheme used by Yang et al. (2004), we transformed the results of the reference group analyses using equation (2) above, so that the estimated age effects sum to zero, as do the estimated period and cohort effects. Appendix table A1 shows these transformed results in the $\beta^*_{\text{first}}=0$ and $\beta^*_{\text{last}}=0$ columns. Figure 1 graphically presents the estimates, *transformed to a common coding*, from the three coding schemes.

Figure 1 shows that the IE estimates can change substantially depending on the choice of coding scheme. The IE estimated age and cohort effects are qualitatively similar for the three coding schemes, but the IE estimates for period effects using the $\beta_{\text{first}}=0$ coding are strikingly different from the IE estimates using the sum-to-zero coding. While the IE estimates using the sum-to-zero coding in figure 1 (identical results are shown in Yang et al. [2004, p. 98]) indicate an upward mortality trend across time periods from 1960 to 1970, the IE estimates using the first-reference-group coding show a downward trend over the same periods. Similarly, for the years from 1975 to 1999, the IE estimates under the sum-to-zero coding suggest a sharp increase in death rates, whereas the IE estimates under the first-reference-group coding show a flat trend. Thus, a researcher would reach opposite conclusions about the effects of period depending on the coding scheme he or she happened to choose.

The magnitude of the cohort effects does depend on the choice of coding scheme. For example, the estimated mortality rate for U.S. females in the 1870–74 birth cohort for the sum-to-zero coding (table A1, col. $\sum=0$) is $\exp(1.008) = 2.740$ times the global mean, while the estimated morality rate for that birth cohort in the first-category reference coding is only $\exp(0.502) = 1.652$ times the global mean.

Example 2

The IE estimates also critically depend on coding scheme in the example of vocabulary knowledge used in Yang et al. (2008, pp. 1712–16), where the authors were concerned with age, period, and cohort trends in Americans' vocabularies. The outcome is the variable WORDSUM in GSS data collected from 1976 to 2000. Yang and colleagues reported that "the age effects on vocabularies show a concave pattern, . . . rising to a peak in the forties"

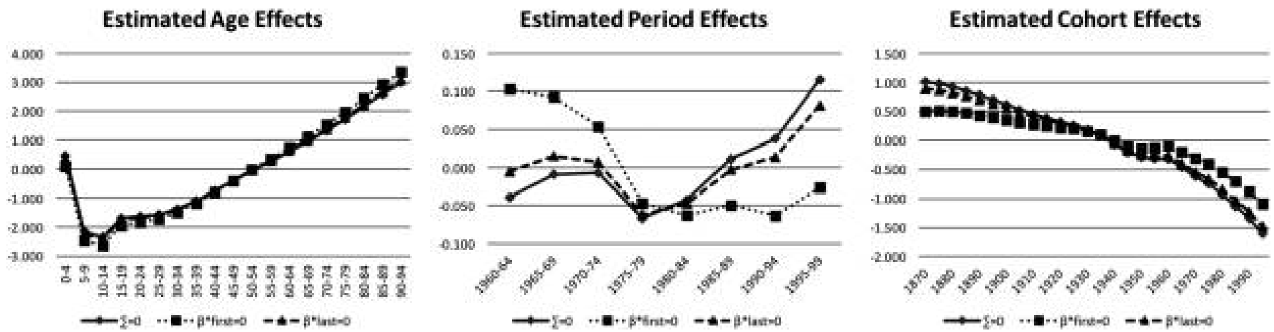

FIG. 1.—Estimated APC trends in mortality using IE under three coding schemes. Data are from Yang et al. (2008). $\sum=0$: sum-to-zero coding. $\beta^*_{\text{first}} = 0$: estimates obtained under the reference-group coding with the first group omitted for each effect, then transformed to the sum-to-zero scale. $\beta^*_{\text{last}} = 0$: estimates obtained under the reference-group coding with the last group omitted for each effect, then transformed to the sum-to-zero scale.

(2008, p. 1714). They also found period and cohort variations in vocabularies, although there were no clear linear patterns (p. 1714). They compared the estimates from the IE with the results from the hierarchical APC models and concluded that the estimated trends are quite similar (p. 1716).

Appendix table A2 and figure 2 show IE estimates using the same three coding schemes used in the previous example. As above, we transformed the results using the $\beta_{\text{first}}=0$ and $\beta_{\text{last}}=0$ codings to the sum-to-zero coding so the estimated effects can be compared directly. The age, period, and cohort effects estimated by IE shown in figure 2 differ dramatically depending on the choice of coding scheme (model parameterization). For example, under the $\sum=0$ coding, vocabulary scores first increase with age but then decrease starting at age 60. Under the $\beta_{\text{first}}=0$, they increase initially but decrease starting at the ages of 30–39. The $\beta_{\text{last}}=0$ coding, by contrast, shows that vocabulary knowledge increases through the age span considered.

The estimated period effects also differ qualitatively depending on the coding scheme. The $\sum=0$ coding shows a modest decrease in vocabulary scores until 1986–90 and then a sharp increase. The $\beta_{\text{first}}=0$ coding shows a consistent increase, while the $\beta_{\text{last}}=0$ coding shows a sharp initial decrease and then a flat trend after 1986–90.

The IE estimates for cohort effects also differ completely depending on the coding scheme. With the $\sum=0$ coding, there is little trend in the estimated effects. The $\beta_{\text{first}}=0$ coding shows strong evidence of an intercohort decline, while the $\beta_{\text{last}}=0$ coding shows just the opposite, a consistent increase in vocabulary knowledge from the oldest to the youngest cohorts. Thus a researcher who used the $\beta_{\text{first}}=0$ coding would reach conclusions about the period and cohort trends in vocabulary knowledge opposite to those reached by another researcher who happened to choose the $\beta_{\text{last}}=0$ coding scheme.

Example 3

The third empirical example considers change in the level of trust among Americans. Schwadel and Stout (2012) applied the IE to the 1972 to 2010 GSS data and showed that the cohorts born before the 1920s are less trusting than those born in the 1920s through 1940s (p. 243). Following these authors, we dichotomized the GSS measure of trust (1 = agree that people can be trusted; 0 = disagree or depends). Appendix table A3 and figure 3 present the IE estimates of the age, period, and cohort effects in trust level using the sum-to-zero ($\sum=0$) coding, the first-reference-group ($\beta_{\text{first}}=0$) coding, and the last-reference-group ($\beta_{\text{last}}=0$) coding.

As figure 3 shows, the IE again yields estimates that depend on the choice of coding, though less so than in the other two examples. For example, the estimated age and period effects have the same general trend in the three

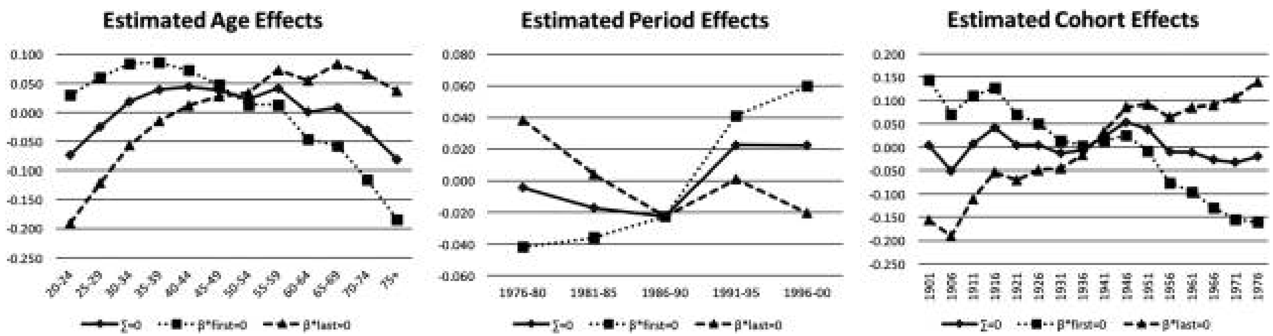

FIG. 2.—Estimated APC trends in vocabularies using IE under three coding schemes. Data are from Yang et al. (2008). $\sum=0$: sum-to-zero coding. $\beta^*_{\text{first}} = 0$: estimates obtained under the reference-group coding with the first group omitted for each effect, then transformed to the sum-to-zero scale. $\beta^*_{\text{last}} = 0$: estimates obtained under the reference-group coding with the last group omitted for each effect, then transformed to the sum-to-zero scale.

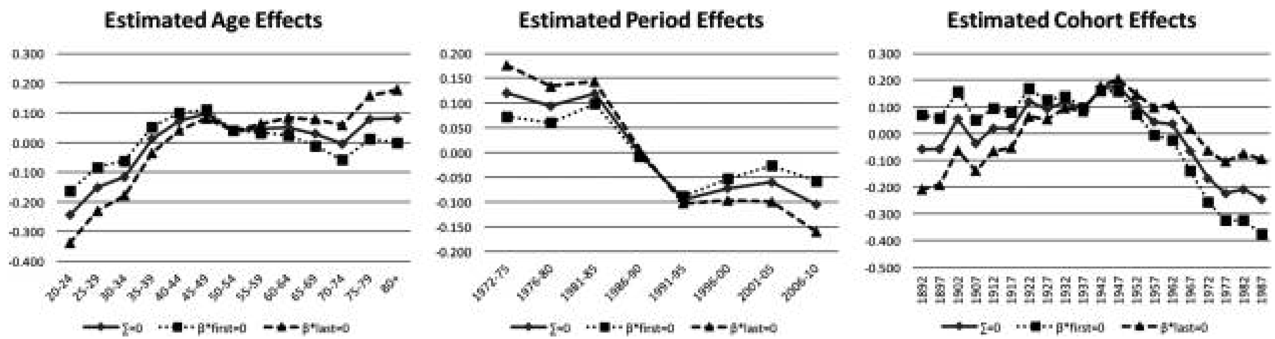

FIG. 3.—Estimated APC trends in trust using IE under three coding schemes. Data are from the GSS, 1972–2010 (Schwadel and Stout 2012). $\sum=0$: sum-to-zero coding. $\beta^*_{\text{first}} = 0$: estimates obtained under the reference-group coding with the first group omitted for each effect, then transformed to the sum-to-zero scale. $\beta^*_{\text{last}} = 0$: estimates obtained under the reference-group coding with the last group omitted for each effect, then transformed to the sum-to-zero scale.

codings but much larger magnitude in the last-reference-group coding. However, the magnitude and general trends in the estimated cohort effects differ qualitatively for the cohorts born in 1942 and earlier, depending on the coding used. For example, contrary to Schwadel and Stout's (2012) conclusion about the intercohort increase in trust for cohorts born between 1892 and 1942, the IE estimates under the first-reference-group coding show a flat pattern across those birth cohorts.

EXPLAINING THE IE'S SENSITIVITY

The above examples show that the results produced by the IE can be highly sensitive to the coding scheme a researcher employs, a choice that is of no consequence with fully identified models.[5] A full understanding of the sensitivity of the IE to coding schemes requires a strong understanding of linear algebra. Here, we attempt to provide an intuitive understanding of the IE's sensitivity. The mathematical appendix provides a more formal treatment.

Recall that the IE estimate is the point on the solution line that is closest to the origin. Consider what happens when we change coding schemes. First, the solution line in the original coding scheme is transformed to a new solution line in the new coding scheme. It is the same solution line, but now represented with respect to the new parameterization. Second, in transforming from the original to the new coding scheme, distances between pairs of points change.[6] As a result, the point on the solution line that is closest to the origin changes; that is, the points that are closest to the origin under the two coding schemes are different. In particular, suppose that in the original coding scheme, a point b_0 on the solution line is closer to the origin than any other point on the solution line; after transforming to the new coding scheme, the transformed value $\mathrm{T}(b_0)$ is, in general, no longer the point closest to the origin among points on the transformed solution line. The IE estimate—the closest point to the origin on the solution line—is sensitive to the coding scheme because the *ordering* of points on the solution line according to their distance from the origin is generally not the same in the original and new coding schemes. This is even true, as shown in the mathemat-

[5] There is an important way in which the traditional constrained estimator is superior to the IE estimator. By its very nature, as explained above and in the mathematical appendix, the IE depends on the coding scheme. This is not the case with the traditional constrained estimator. When a coefficient constraint is imposed, the coding scheme has no effect on the estimates because the constraint is invariant to the researcher's choice of coding scheme, unlike the IE.

[6] If the transformation is an orthogonal transformation, then the distances from the origin of points on the solution line are preserved after the transformation. None of the changes of coding scheme considered in previous sections is an orthogonal transformation.

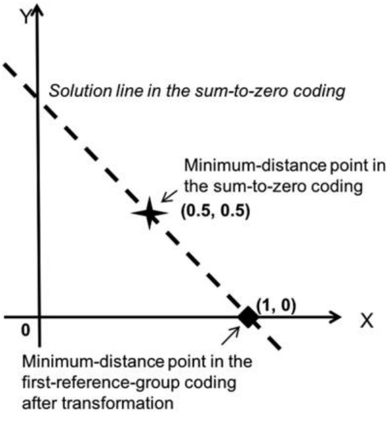

FIG. 4.—IE solution in the sum-to-zero coding

ical appendix, with sum-to-zero coding schemes that have different omitted categories.[7]

The fact that the solution line and the measure of distance change with coding schemes is illustrated in figures 4 and 5, which are necessarily stylized because real APC problems have too many dimensions to show in a two-dimensional figure. In figure 4, the dashed line denotes the solution line in the sum-to-zero coding scheme (parameterization). Transforming to the first-reference-group coding transforms the solution line to figure 5's vertical dashed line. In figure 4, the point on the solution line that is closest to the origin is (0.5, 0.5), but after being transformed to the point (1, −1) in figure 5's first-reference-group coding scheme, it is no longer closest to the origin among points on the solution line.

This is a disturbing result: given that there are infinitely many possible coding schemes (though most would be difficult to interpret), there are infinitely many IE estimates. The seemingly innocuous choice of a coding scheme affects the IE estimates, sometimes very much. As discussed above, because of the identification problem in APC models, producing an estimate amounts to choosing one set of estimates from the solution line, which contains the infinitely many estimates that are consistent with the data. As O'Brien (2011) showed, any constrained estimation procedure, including the IE, simply picks out one particular set of estimates on the solution line.

[7] For nonidentified models, changing coding scheme doesn't inherently change the likelihood (Gelman 2004). However, choosing a unique estimate requires choosing one point from the set that maximizes the likelihood, and that can depend on the coding scheme/parameterization.

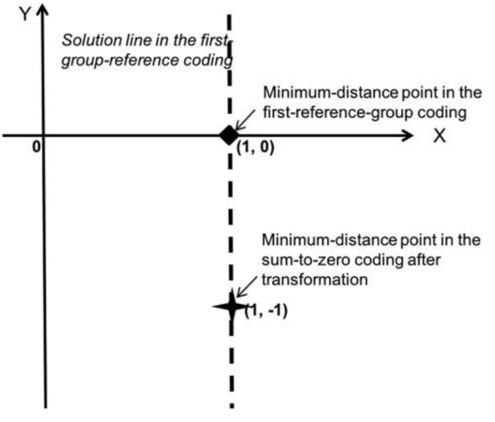

FIG. 5.—IE solution in the first-reference-group coding

For the IE, however, the situation is even worse. In the mathematical appendix, we show that *any* set of estimates on the solution line is the IE estimate for an appropriately chosen coding scheme/design matrix. In other words, one can choose any set of estimates on the solution line that one wants to privilege, and there will be a coding scheme in which the IE estimates are that chosen set. Using the IE, we can privilege any point on the solution line we want simply by choosing the right coding scheme.

Given these mathematical results, it is not surprising that different design matrices produced different estimates in the empirical examples. But how can the IE estimates differ as much as they do in the empirical examples? Exact linear dependence, as in the APC model, can be understood as the most extreme form of multicollinearity. As is well known regarding multicollinearity, small changes in the data or in the model specification can change estimates dramatically. When multicollinearity is present, we simply do not have sufficient variation in a variable of interest, holding other variables constant, to precisely estimate the effect of that variable of interest. In the case of exact linear dependence, as in the APC model, there is no variation at all in, say, age, when the other variables (period and cohort) are held constant.[8] As such, it is entirely reasonable to expect IE's results to be highly unstable.

This comparison of linear dependence to multicollinearity suggests a direction for future research. As the three empirical examples illustrate, the choice of coding scheme, or equivalently the choice of constraint that is imposed on the parameters, affects parameter estimates dramatically in some cases but not in others. As for multicollinearity (Belsley et al. 1980), it would

[8] Strictly speaking, there is no variation in the *linear component* of age; the nonlinear components of the three effects are identified (Holford 1983).

be useful to have formal methods for analyzing the sensitivity of estimates to the constraint (the IE or other) that is used to privilege one set of estimates.[9] This is a topic for future research.

CONCLUSION

"The Intrinsic Estimator for Age-Period-Cohort Analysis" (Yang et al. 2008) has been cited 189 times as of October 2016 and has been used by researchers in different disciplines to address important substantive questions. Many researchers appear convinced that the assumptions implicit in the IE do not affect the IE's ability to estimate, even if only approximately, the "true" age, period, and cohort effects (see, e.g., Keyes and Miech 2013; Langley et al. 2011; Schwadel 2011; Masters et al. 2014). The empirical and mathematical results presented in this comment contradict that optimistic view. Social scientists should be aware that the seemingly innocuous choice of a coding scheme can have a major effect on the estimates produced by the IE and, as a result, on the conclusions they reach.

Social scientists have long looked for statistical methods that will provide assumption-free results revealing the underlying structure of empirical data. As with causal analysis of observational data (Pearl 2009; Morgan and Winship 2007), we believe this is an impossible goal. Heckman and Robb (1985) stated the situation correctly nearly three decades ago: "The age-period-cohort effect identification problem arises because analysts want something for nothing: a general statistical decomposition of data without specific subject matter motivation underlying the decomposition. In a sense it is a blessing for social science that a purely statistical approach to the problem is bound to fail" (pp. 144–45).

Liying Luo
Pennsylvania State University

James Hodges
University of Minnesota

Christopher Winship
Harvard University

Daniel Powers
University of Texas at Austin

[9] Comparing the IE's estimates under just three coding schemes, the sum-to-zero and the two reference-group coding schemes, is unlikely to indicate the true sensitivity of IE es-

APPENDIX

Preliminaries: Defining notation, the IE estimate

Suppose we have an outcome measure Y that we are willing to treat as continuous, for example, WORDSUM score. (This is not necessary but simplifies the presentation.) Suppose also that we have a age groups and p periods, and thus $a + p - 1$ cohorts, and that we have chosen a coding scheme (parameterization) for the APC model, for example, the sum-to-zero coding scheme. Then in the usual APC model analysis, we have a vector of outcomes, $\mathbf{Y}$, that has mean $\mathbf{Xb}$, with design matrix $\mathbf{X}$ and parameter vector $\mathbf{b}$ as follows. The design matrix $\mathbf{X}$ has one row for each observation (i.e., for each element in the vector $\mathbf{Y}$) and one column for each element in $\mathbf{b}$. The parameter vector $\mathbf{b}$ has one element for an intercept, $a - 1$ elements for the age effect, $p - 1$ elements for the period effect, and $a + p - 2$ elements for the cohort effects. Thus $\mathbf{b}$ has $2(a + p) - 3$ elements.

The APC model is not identified in the sense that the design matrix $\mathbf{X}$ has rank less than $2(a + p) - 3$; in particular, its rank is smaller than this by one. Thus, there is exactly one null vector $\mathbf{B}_0$, having $2(a + p) - 3$ elements like $\mathbf{b}$, such that $\mathbf{XB}_0 = 0$ and $\mathbf{B}_0^T\mathbf{B}_0 = 1$, that is, $\mathbf{B}_0$ has Euclidean length 1. For a given data set $\mathbf{Y}$, the ordinary least squares estimates of $\mathbf{b}$ satisfy the equation $\mathbf{X}^T\mathbf{Xb} = \mathbf{X}^T\mathbf{Y}$, but this equation does not have a unique solution because $\mathbf{X}$ is not of full rank. If $\mathbf{b}_1$ is a solution to this equation, then any solution can be written as $\mathbf{b}_1 + r\mathbf{B}_0$ for some real number r. This defines the solution line for this coding scheme and data set $\mathbf{Y}$. In this coding scheme, the IE estimate is given by the value of r that minimizes the (Euclidean) length of $\mathbf{b}_1 + r\mathbf{B}_0$, or equivalently its squared length, which is $(\mathbf{b}_1 + r\mathbf{B}_0)^T(\mathbf{b}_1 + r\mathbf{B}_0)$. Simple calculus shows that the squared length is minimized for $r = -\mathbf{B}_0^T\mathbf{b}_1$, so the IE estimate is

$$\begin{aligned}\mathbf{b}_0 &= \mathbf{b}_1 - (\mathbf{B}_0^T\mathbf{b}_1)\mathbf{B}_0 \\ &= \mathbf{b}_1 - \mathbf{B}_0(\mathbf{B}_0^T\mathbf{b}_1) \text{ because } \mathbf{B}_0^T\mathbf{b}_1 \text{ is a scalar} \\ &= (\mathbf{I} - \mathbf{B}_0\mathbf{B}_0^T)\mathbf{b}_1 \text{ where } \mathbf{I} \text{ is the identity matrix of order } 2(a + p) - 3\,.\end{aligned} \tag{A1}$$

Note that $\mathbf{b}_0$ and $\mathbf{B}_0$ are orthogonal by construction: $\mathbf{B}_0^T\mathbf{b}_0 = (\mathbf{B}_0^T - \mathbf{B}_0^T)\mathbf{b}_1 = 0$ because $\mathbf{B}_0^T\mathbf{B}_0 = 1$. Any parameter vector $\mathbf{b}$ on the solution line can now be written as $\mathbf{b}_0 + s\mathbf{B}_0$ for $s = \mathbf{B}_0^T\mathbf{b}$.

timates to the coding scheme. The sum-to-zero and reference-group coding schemes are simply common, conventional choices. An alternative approach would be to examine the full set of estimates on the solution line. This would be analogous to what is done in principal components analysis, where various rotations of the axes are tried in order to determine what solution makes the most sense. Of course, this has led people using this approach to be accused of trying to “read tea leaves.”

Reparameterizing can change the ordering of points on the solution line according to their distance from the origin

Suppose we have written the APC model in one coding scheme with the design matrix and parameter vector $\mathbf{X}$ and $\mathbf{b}$, respectively. Suppose now that we want to change to a new coding scheme. Then there is an invertible square matrix $\mathbf{T}$ of dimension $2(a + p) - 3$ that effects the change from the original to the new coding scheme, as follows:

$$\mathbf{Xb} = \mathbf{XT}^{-1}\mathbf{Tb} = \mathbf{X}(\mathbf{T})\mathbf{b}(\mathbf{T}), \tag{A2}$$

where $\mathbf{X}(\mathbf{T}) = \mathbf{XT}^{-1}$ is the design matrix in the new coding scheme and $\mathbf{b}(\mathbf{T}) = \mathbf{Tb}$ is the parameter in the new coding scheme corresponding to $\mathbf{b}$ in the original coding scheme. Below in this appendix we show how to derive $\mathbf{T}$ for any choice of original and new coding schemes.

So suppose we have an original coding scheme, with design matrix $\mathbf{X}$ and null vector $\mathbf{B}_0$. Suppose also we have a data set $\mathbf{Y}$, and that the IE estimate for this data set is $\mathbf{b}_0$, as above. Then as noted, any estimate $\mathbf{b}$ in the solution line for this coding scheme has the form $\mathbf{b}_0 + s\mathbf{B}_0$, for $s = \mathbf{B}_0^T\mathbf{b}$. Any solution $\mathbf{b}$ in the original coding scheme is therefore transformed to $\mathbf{Tb} = \mathbf{T}(\mathbf{b}_0 + s\mathbf{B}_0)$ in the new coding scheme. In the new coding scheme, the squared distance of $\mathbf{Tb}$ to the origin is $\mathbf{b}^T\mathbf{T}^T\mathbf{Tb} = (\mathbf{b}_0 + s\mathbf{B}_0)^T\mathbf{T}^T\mathbf{T}(\mathbf{b}_0 + s\mathbf{B}_0)$. This distance is a quadratic in the scalar s, and simple calculus shows that this squared distance is minimized by

$$s_{\mathbf{T}} = -\mathbf{B}_0^T\mathbf{T}^T\mathbf{Tb}_0/\mathbf{B}_0^T\mathbf{T}^T\mathbf{TB}_0. \tag{A3}$$

Thus, the IE solution in the new coding scheme, back transformed to the original coding scheme, is $\mathbf{b}_0 + s_{\mathbf{T}}\mathbf{B}_0$. This is equal to the IE solution in the original coding scheme if and only if $s_{\mathbf{T}} = 0$. It is easy to show that $s_{\mathbf{T}} = 0$ if (1) $\mathbf{T}$ is an orthogonal matrix, or (2) $\mathbf{T}$ has one row proportional to $\mathbf{B}_0$ and its other rows are orthogonal to $\mathbf{B}_0$. (Orthogonal matrices correspond to rigid transformations such as rotations and reflections, which preserve distances between pairs of points.) It is also easy to show that all other $\mathbf{T}$ giving $s_{\mathbf{T}} = 0$ depend on $\mathbf{b}_0$, that is, on the specific data set $\mathbf{Y}$. In other words, some $\mathbf{T}$ exist for which $s_{\mathbf{T}} = 0$, but they are few and very specific, and they do not include the $\mathbf{T}$ that effect changes between any pair of familiar coding schemes, such as those considered in the main body of this paper. Thus, except for uninteresting cases, changing coding schemes changes the distances between pairs of points. In particular, changing coding schemes changes the *ordering* of points in the original coding scheme's solution line according to their distance from the origin in the coding scheme defined by $\mathbf{T}$. This happens because the transformation $\mathbf{T}$ is not rigid, which means that a vector $\mathbf{b}$ is stretched by different amounts in different directions when it is transformed to $\mathbf{Tb}$. If $\mathbf{T}$ has singular value decomposition $\mathbf{T} = \mathbf{UDV}^T$, for $\mathbf{U}$

and **V** orthogonal matrices and **D** diagonal, then **D**'s diagonal elements describe the differential stretching applied to directions defined by $\mathbf{V}^T$. The following section gives an example.

For any estimate b on the solution line, there exists a coding scheme such that b is the IE estimate in that coding scheme, back transformed to the original coding scheme

Suppose we have a age groups and p periods and data **Y**, and that we have chosen a coding scheme, which we will call the original coding scheme. Then this implies a design matrix **X**, a null vector $\mathbf{B}_0$, and the IE estimate $\mathbf{b}_0$. Any other solution to the equation $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{Y}$ has the form $\mathbf{b}_0 + s\mathbf{B}_0$, for some real number s. The burden of this section is to show that for any real number t, there is an invertible square matrix **T** of dimension $r = 2(a + p) - 3$ and a new coding scheme **Tb** such that the IE estimate in the new coding scheme, back transformed to the original coding scheme, is $\mathbf{b}_0 + t\mathbf{B}_0$. First we prove this main claim; then we prove a closely related secondary claim, which is stated below.

Proof of the main claim

This proof uses the fact that in any given coding scheme, IE's estimate minimizes, among points on the solution line, the squared distance from the estimate to the origin. As noted above, for a given transformation (recoding) **T**, the IE estimate in the new coding scheme, back transformed to the original coding scheme, has

$$s_{\mathbf{T}} = -\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{b}_0/\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{B}_0. \tag{A4}$$

We need to prove that for any real number t, we can choose a **T** such that

$$t = s_{\mathbf{T}} = -\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{b}_0/\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{B}_0. \tag{A5}$$

If $\mathbf{b}_0$ is the zero vector, then the IE solution in all coding schemes is also the zero vector. This case is so unlikely that it is of no interest, so we assume that $\mathbf{b}_0$ is not the zero vector.

$\mathbf{T}^T\mathbf{T}$ is positive definite and symmetric of dimension r, so it has spectral decomposition $\mathbf{T}^T\mathbf{T} = \mathbf{G}\mathbf{D}\mathbf{G}^T$, where **G** is an orthogonal matrix of dimension r and **D** is diagonal with r positive diagonal entries; by convention, **D**'s diagonal entries d_i are sorted in decreasing order, so $d_1 \geq d_2 \ldots \geq d_r$. Choosing **T** is equivalent to choosing **G** and **D**.

For any legal **G** and **D**, $\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{b}_0 = \Sigma_i a_i c_i d_i$, where $\mathbf{B}_0^T\mathbf{G} = (a_1, a_2, \ldots a_r)$ and $\mathbf{b}_0^T\mathbf{G} = (c_1, c_2, \ldots c_r)$ and the sum runs over i = 1, . . . r. (Note that $\Sigma_i a_i^2 = \mathbf{B}_0^T\mathbf{G}\mathbf{G}^T\mathbf{B}_0 = 1$ because $\mathbf{B}_0$ has length 1, and $\Sigma_i c_i^2 = \mathbf{b}_0^T\mathbf{b}_0$.) With these definitions, $\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{B}_0 = \Sigma_i a_i^2 d_i$. Thus, we need to choose $\mathbf{G}$—that is, choose the a_i and c_i—and choose $\mathbf{D}$—that is, choose the d_i—so that

$$t = -\sum_i a_i c_i d_i / \sum_i a_i^2 d_i \,. \tag{A6}$$

If $d_2, \ldots d_r$, are fixed at some values and d_1 is made very large, then $-\Sigma_i a_i c_i d_i / \Sigma_i a_i^2 d_i$ becomes arbitrarily close to $-c_1/a_1$. Our proof is finished if we choose $\mathbf{G}$ so that $-c_1/a_1 = t$; then we let d_1 grow very large and $s_\mathbf{T}$ becomes arbitrarily close to t, as needed. To choose such a $\mathbf{G}$, define $\boldsymbol{\beta}_0 = \mathbf{b}_0(\mathbf{b}_0^T\mathbf{b}_0)^{-0.5}$, so $\boldsymbol{\beta}_0^T\boldsymbol{\beta}_0 = 1$ and $\boldsymbol{\beta}_0^T\mathbf{B}_0 = 0$. Then let the first column of $\mathbf{G}$ be $\mathbf{G}_1 = \alpha\varphi\boldsymbol{\beta}_0 + (1 - \alpha)\mathbf{B}_0$, where α is between 0 and 1, and φ is -1 if $t > 0$ and 1 if $t < 0$. Then $a_1 = 1 - \alpha$ and $c_1 = (\mathbf{b}_0^T\mathbf{b}_0)^{0.5}\,\alpha\varphi$, so $-c_1 / a_1 = -\varphi(\mathbf{b}_0^T\mathbf{b}_0)^{0.5}\,\alpha / (1 - \alpha)$. Set $\alpha = |t| / ((\mathbf{b}_0^T\mathbf{b}_0)^{0.5} + |t|)$; then $-c_1 / a_1 = t$.

Secondary claim

Suppose that in the original coding scheme, the true value of the parameter is $\mathbf{b}$. Then the IE estimate is an unbiased estimate of $\mathbf{b}_{0u} = (\mathbf{I} - \mathbf{B}_0\mathbf{B}_0^T)\mathbf{b}$, where the subscript u indicates true, referring to the true $\mathbf{b}$. For any real number t, there is an invertible square matrix $\mathbf{T}$ of dimension $r = 2(a + p) - 3$ and a new coding scheme $\mathbf{Tb}$ such that the IE estimate in the new coding scheme, back transformed to the original coding scheme, is unbiased for $\mathbf{b}_{0u} + t\mathbf{B}_0$. The proof follows.

Again, if $\mathbf{b}_{0u}$ is the zero vector, then $\mathbf{T}\mathbf{b}_{0u}$ is also the zero vector for all $\mathbf{T}$. As before, this case so unlikely that it is of no interest, so we assume that $\mathbf{b}_{0u}$ is not the zero vector. The difference between $\mathbf{b}_{0u}$ and the back-transformed IE estimand in the new coding scheme is

$$\begin{aligned}&(\mathbf{I} - \mathbf{B}_0\mathbf{B}_0^T)\mathbf{b} - \mathbf{T}^{-1}(\mathbf{I} - \mathbf{T}\,\mathbf{B}_0\mathbf{B}_0^T\mathbf{T}^T)\mathbf{T}\mathbf{b}\\ &= \mathbf{B}_0[\mathbf{B}_0^T(\mathbf{T}^T\mathbf{T} - \mathbf{I})\mathbf{b}],\end{aligned} \tag{A7}$$

where the expression in square brackets is a scalar and $\mathbf{I}$ is the identity matrix of dimension $2(a + p) - 3$. The burden of this proof is to show how to choose $\mathbf{T}$ so $t = \mathbf{B}_0^T(\mathbf{T}^T\mathbf{T} - \mathbf{I})\mathbf{b}$. Recalling that $\mathbf{b} = \mathbf{b}_{0u} + s\,\mathbf{B}_0$ for a particular scalar s, we need

$$t = \mathbf{B}_0^T(\mathbf{T}^T\mathbf{T} - \mathbf{I})\mathbf{b} = \mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{b}_{0u} + s\mathbf{B}_0^T\mathbf{T}^T\mathbf{T}\mathbf{B}_0 - s, \tag{A8}$$

where $\mathbf{B}_0$, $\mathbf{b}_{0u}$, and s are fixed and $\mathbf{B}_0^T\mathbf{b}_{0u} = 0$.

As above, choosing $\mathbf{T}$ is equivalent to choosing an orthogonal matrix $\mathbf{G}$ and a diagonal matrix $\mathbf{D}$ with all diagonal elements positive so that

$\mathbf{T}^T\mathbf{T} = \mathbf{GDG}^T$, and using the notation defined in proving the main claim, we need to choose a_i, c_i, and d_i, $i = 1, \ldots r$ so that

$$t = \sum_i a_i c_i d_i + s\sum_i a_i^2 d_i - s. \tag{A9}$$

As in the earlier proof, we do so by fixing $d_2, \ldots, d_r$ at some values (which do not matter) and adjusting d_1 to get the desired result. To do this, define the function $g(d_1) = \Sigma_i a_i c_i d_i + s\Sigma_i a_i^2 d_i - s$; then g's derivative with respect to d_1 is

$$g'(d_1) = a_1 c_1 + s a_1^2, \tag{A10}$$

which does not depend on d_1. To get the desired result, we need only show that we can pick $\mathbf{G}$ so that $a_1 c_1 + s a_1^2$ is positive for the given s and that we also can pick $\mathbf{G}$ so that $a_1 c_1 + s a_1^2$ is negative for the given s. Then we can fix a_i, c_i, and d_i for $i = 2, \ldots, r$ at any values, pick $\mathbf{G}$ so that $g'(d_1)$ has the appropriate sign, and increase d_1 until $g(d_1) = t$.

If $s = 0$, $g'(d_1) = a_1 c_1$. As in the proof of the main claim, let the first column of $\mathbf{G}$ be $\mathbf{G}_1 = \boldsymbol{\alpha\varphi\beta}_0 + (1 - \boldsymbol{\alpha})\mathbf{B}_0$, and set $\alpha = 0.5$. Then $g'(d_1) = a_1 c_1 = 0.25\,(\mathbf{b}_0^T\mathbf{b}_0)^{0.5}\varphi$, which is made positive or negative by choosing positive or negative φ, respectively. Now suppose $s \neq 0$. Then $g'(d_1) = a_1 c_1 + s a_1^2 > 0$ if $c_1 a_1 > -s$, and $g'(d_1) = a_1 c_1 + s a_1^2 < 0$ if $c_1 / a_1 > -s$. Either of these inequalities can be satisfied as in the proof of the main claim, by defining $\mathbf{G}_1$ as above and selecting α and φ as needed.

Constructing the transformation matrix T for any change in coding scheme

Recall that $\mathbf{T}$ transforms the $(2(a + p) - 3)$-vector $\mathbf{b}$ in the original coding scheme to the $(2(a + p) - 3)$-vector $\mathbf{Tb}$ in the new coding scheme. The $(2(a + p) - 3) \times (2(a + p) - 3)$ matrix $\mathbf{T}$ can be constructed as follows. First, construct the $(2(a + p) \times (2(a + p) - 3)$ matrix $\mathbf{T}_1$ that transforms $\mathbf{b}$ in the original coding scheme to a $2(a + p)$-vector $\mathbf{T}_1\mathbf{b}$ in the full (redundant) coding scheme with a parameters for the age effect, p for the period effect, and $a + p - 1$ for the cohort effect, along with the intercept. Second, construct the $(2(a + p) - 3) \times 2(a + p)$ matrix $\mathbf{T}_2$ that transforms a $2(a + p)$-vector in the full (redundant) coding scheme to a $(2(a + p) - 3)$-vector in the new coding scheme. Then $\mathbf{T} = \mathbf{T}_2\mathbf{T}_1$.

For example, suppose we have three age groups and four periods, so $a = 3$ and $p = 4$. Suppose the original coding scheme is sum-to-zero with the last group omitted for each of the age, period, and cohort effects, so the parameter vector $\mathbf{b}$ has 11 elements: the intercept, the first two age group effects, the first three period effects, and the first five cohort effects. Then $\mathbf{T}_1$ is the 14×11 matrix

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

Suppose the new coding scheme is the first-category reference group scheme, so that the parameter vector **Tb** has 11 elements: the intercept, the last two age group effects minus the first age group effect, the last three period effects minus the first period effect, and the last five cohort effects minus the first cohort effect. Then $\mathbf{T}_2$ is the 11×14 matrix

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Then $\mathbf{T}$ is the 11×11 matrix $\mathbf{T} = \mathbf{T}_2\mathbf{T}_1 =$

$$\begin{pmatrix}
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & -2 & -1 & -1 & -1 & -1
\end{pmatrix}$$

This matrix is invertible; its 11 singular values range in absolute value from 0.78 to 3.27, so the ratio of $\mathbf{b}$'s maximally and minimally stretched directions is $3.27/0.78 = 4.20$.

The IE estimate is sensitive to the coding scheme, which is even true with sum-to-zero coding schemes that have different omitted categories

Consider an example of three age groups and three periods, so $a = 3$ and $p = 3$. Suppose the original coding scheme is sum-to-zero with the last group omitted for each of the age, period, and cohort effects, so the parameter vector $\mathbf{b}$ has nine elements: the intercept, the first two age group effects, the first two period effects, and the first four cohort effects. As shown in the section above, the 12×9 matrix $\mathbf{T}_1$ that transforms $\mathbf{b}$ in the original coding scheme to the full coding is

$$
\mathbf{T}_1 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1
\end{pmatrix}
$$

Suppose the new coding scheme is the sum-to-zero coding scheme with the first category of each effect omitted, so that the parameter vector **Tb** has nine elements: the intercept, the last two age group effects, the last two period effects, and the last four cohort effects. Then $\mathbf{T}_2$ is a 9×12 matrix

$$
\mathbf{T}_2 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

Then $\mathbf{T}$ is a 9×9 matrix, where $\mathbf{T} = \mathbf{T}_2\mathbf{T}_1 =$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \end{pmatrix}$$

$\mathbf{T}$ is invertible but not orthogonal; its nine singular values range in absolute value from 0.46 to 2.19, so the ratio of $\mathbf{b}$'s maximally and minimally stretched directions is $2.19/0.46 = 4.76$.

To illustrate, we simulate a data set with three age groups, three periods, and five cohorts as follows:

$$y_{ij} \sim \{10 + 2 \times \text{age}_i - 0.5 \times \text{age}_i^2 - 1 \times \text{period}_j - 0.5 \times \text{period}_j^2 + 1 \times \text{cohort}_{ij} + 0.5 \times \text{cohort}_{ij}^2, \sigma = 0\}.$$

For each age-by-period combination, there is one observation, so the total sample size is nine. Table A4 and figure A1 show IE estimates for the simulated data using two different sum-to-zero coding schemes, namely, the sum-to-zero coding with the last category of each effect omitted and the same coding with the first category omitted; note the estimates in table A4 and figure A1 are obtained using the sum-to-zero coding with the first category omitted and then transformed to the same coding with the first category omitted. The resulting two sets of IE estimates are different. For example, the estimated cohort effect for the first cohort is 0.75 under the first type of sum-to-zero coding, whereas the estimated effect is 1.75 under the second type after being transformed.

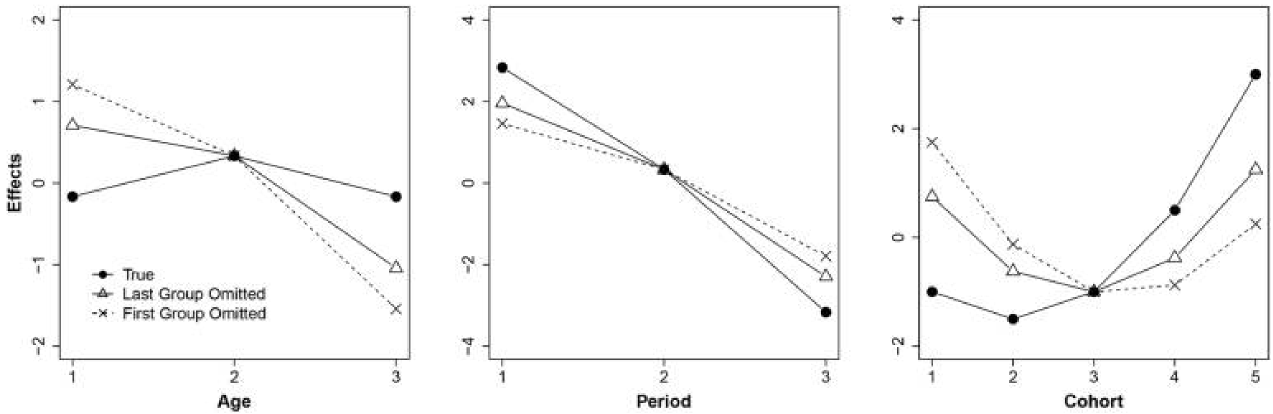

FIG. A1.—IE estimates under two different sum-to-zero coding schemes for a simulated data set

TABLE A1
ESTIMATED AGE, PERIOD, AND COHORT EFFECTS ON MORTALITY
UNDER THREE CODING SCHEMES

Category/Effects	$\sum=0$	$\beta_{first}=0$	$\beta^*_{first}=0$	$\beta_{last}=0$	$\beta^*_{last}=0$
Intercept	−5.400	−4.708	−5.400	−3.729	−5.400
			Age		
0–4	.453	...	.088	−2.710	.366
5–9	−2.144	−2.556	−2.468	−5.297	−2.221
10–14	−2.354	−2.726	−2.637	−5.497	−2.421
15–19	−1.704	−2.035	−1.947	−4.838	−1.762
20–24	−1.630	−1.921	−1.833	−4.754	−1.678
25–29	−1.571	−1.821	−1.733	−4.685	−1.610
30–34	−1.377	−1.587	−1.499	−4.482	−1.406
35–39	−1.091	−1.260	−1.172	−4.186	−1.111
40–44	−.751	−.879	−.791	−3.836	−.760
45–49	−.398	−.486	−.398	−3.473	−.398
50–54	−.057	−.104	−.016	−3.123	−.047
55–59	.266	.259	.347	−2.790	.286
60–64	.610	.644	.732	−2.436	.639
65–69	.956	1.030	1.118	−2.081	.995
70–74	1.331	1.446	1.534	−1.696	1.380
75–79	1.724	1.879	1.967	−1.293	1.782
80–84	2.157	2.352	2.440	−.851	2.224
85–89	2.590	2.826	2.914	−.408	2.667
90–94	2.988	3.265	3.353	...	3.076
			Period		
1960–64	−.039	...	.103	−.087	−.005
1965–69	−.009	−.011	.092	−.067	.015
1970–74	−.007	−.049	.054	−.074	.007
1975–79	−.067	−.150	−.047	−.144	−.062
1980–84	−.043	−.166	−.063	−.129	−.047
1985–89	.011	−.152	−.049	−.085	−.003
1990–94	.038	−.166	−.063	−.068	.014
1995–99	.115	−.129	−.026	...	.082
			Cohort		
1870	1.008	...	.502	2.374	.887
1875	.977	.009	.511	2.352	.866
1880	.922	−.006	.496	2.306	.820
1885	.853	−.033	.468	2.247	.761
1890	.776	−.071	.431	2.179	.693
1895	.698	−.107	.394	2.112	.626
1900	.610	−.155	.347	2.034	.548
1905	.522	−.203	.299	1.954	.468
1910	.455	−.229	.273	1.898	.412
1915	.383	−.261	.241	1.835	.349
1920	.317	−.286	.216	1.779	.293
1925	.262	−.301	.201	1.733	.247
1930	.178	−.344	.158	1.659	.173
1935	.077	−.404	.097	1.568	.082

TABLE A1 (*Continued*)

Category/Effects	$\sum$=0	β_{first}=0	β^*_{first}=0	β_{last}=0	β^*_{last}=0
			Period		
1940	−.067	−.508	−.006	1.434	−.052
1945	−.204	−.605	−.103	1.306	−.180
1950	−.287	−.647	−.145	1.233	−.253
1955	−.312	−.631	−.129	1.218	−.268
1960	−.319	−.597	−.096	1.221	−.266
1965	−.460	−.699	−.197	1.088	−.398
1970	−.620	−.818	−.316	.939	−.547
1975	−.748	−.905	−.403	.821	−.665
1980	−.934	−1.051	−.549	.644	−.842
1985	−1.137	−1.213	−.712	.450	−1.036
1990	−1.342	−1.378	−.876	.255	−1.231
1995	−1.607	−1.602	−1.100	...	−1.486

NOTE.—Data are from Yang et al. (2004). $\sum$=0: sum-to-zero coding. $\beta_{\text{first}} = 0$: reference-group coding with the first group omitted for each effect. $\beta_{\text{last}} = 0$: reference-group coding with the last group omitted for each effect. $\beta^*_{\text{first}} = 0$: $\beta_{\text{first}} = 0$ estimates transformed to the sum-to-zero scale. $\beta^*_{\text{last}} = 0$: $\beta_{\text{last}} = 0$ estimates transformed to the sum-to-zero scale.

TABLE A2

ESTIMATED AGE, PERIOD, AND COHORT EFFECTS ON VOCABULARIES UNDER THREE CODING SCHEMES

Category/Effects	$\sum$=0	β_{first}=0	β^*_{first}=0	β_{last}=0	β^*_{last}=0
Intercept	−2.820	−2.688	−2.820	−2.664	−2.820
			Age		
20–24	−.073	...	.030	−.227	−.191
25–29	−.025	.030	.059	−.158	−.121
30–34	.018	.054	.084	−.093	−.057
35–39	.039	.056	.086	−.051	−.015
40–44	.044	.042	.072	−.025	.012
45–49	.038	.017	.047	−.009	.027
50–54	.022	−.017	.013	−.004	.033
55–59	.041	−.017	.013	.036	.073
60–64	.001	−.076	−.046	.018	.054
65–69	.008	−.088	−.058	.046	.083
70–74	−.031	−.145	−.115	.029	.066
75+	−.081	−.214	−.184	...	.036
			Period		
1976–80	−.004	...	−.042	.059	.038
1981–85	−.017	.006	−.036	.025	.004
1986–90	−.023	.019	−.023	−.002	−.023
1991–95	.022	.083	.041	.022	.001
1996–00	.022	.102	.060	...	−.021

TABLE A2 (*Continued*)

Category/Effects	$\sum=0$	$\beta_{first}=0$	$\beta^*_{first}=0$	$\beta_{last}=0$	$\beta^*_{last}=0$
			Cohort		
1901	.003	. . .	.144	−.297	−.157
1906	−.052	−.074	.070	−.331	−.191
1911	.006	−.034	.110	−.251	−.111
1916	.042	−.018	.126	−.194	−.054
1921	.004	−.074	.070	−.210	−.070
1926	.004	−.093	.051	−.189	−.049
1931	−.014	−.130	.014	−.186	−.046
1936	−.007	−.141	.003	−.157	−.017
1941	.023	−.130	.014	−.106	.034
1946	.054	−.119	.025	−.054	.086
1951	.038	−.153	−.009	−.048	.091
1956	−.011	−.220	−.076	−.076	.064
1961	−.011	−.240	−.096	−.055	.085
1966	−.027	−.274	−.130	−.049	.090
1971	−.033	−.299	−.155	−.034	.106
1976	−.021	−.305	−.161	. . .	.140

NOTE.—Data are from Yang et al. (2008). $\sum=0$: sum-to-zero coding. $\beta_{first}=0$: reference-group coding with the first group omitted for each effect. $\beta_{last}=0$: reference-group coding with the last group omitted for each effect. $\beta^*_{first}=0$: $\beta_{first}=0$ estimates transformed to the sum-to-zero scale. $\beta^*_{last}=0$: $\beta_{last}=0$ estimates transformed to the sum-to-zero scale.

TABLE A3

ESTIMATED AGE, PERIOD, AND COHORT EFFECTS ON TRUST UNDER THREE CODING SCHEMES

Category/Effects	$\sum=0$	$\beta_{first}=0$	$\beta^*_{first}=0$	$\beta_{last}=0$	$\beta^*_{last}=0$
Intercept	−.996	−1.015	−.996	−1.074	−.996
			Age		
20–24	−.245	. . .	−.163	−.517	−.340
25–29	−.151	.080	−.083	−.407	−.230
30–34	−.116	.102	−.062	−.356	−.179
35–39	.013	.217	.053	−.212	−.035
40–44	.073	.264	.100	−.136	.041
45–49	.097	.274	.111	−.096	.081
50–54	.041	.204	.041	−.136	.041
55–59	.047	.197	.034	−.114	.063
60–64	.052	.188	.025	−.093	.084
65–69	.030	.153	−.010	−.099	.078
70–74	−.004	.105	−.058	−.117	.060
75–79	.080	.175	.012	−.018	.159
80+	.082	.164	.001	−	.177

TABLE A3 (*Continued*)

Category/Effects	$\sum=0$	$\beta_{first}=0$	$\beta^*_{first}=0$	$\beta_{last}=0$	$\beta^*_{last}=0$
			Period		
1972–75	.121	–	.073	0.337	.176
1976–80	.094	−.013	.060	.294	.134
1981–85	.119	.026	.099	0.304	.143
1986–90	−.001	−.081	−.008	.168	.007
1991–95	−.095	−.162	−.088	.057	−.103
1996–00	−.073	−.126	−.053	.064	−.097
2001–05	−.060	−.100	−.026	.061	−.100
2006–10	−.105	−.131	−.058	–	−.161
			Cohort		
1892	−.058	–	.071	−.114	−.209
1897	−.057	−.012	.059	−.097	−.192
1902	.056	.088	.158	.033	−.062
1907	−.036	−.019	.052	−.045	−.139
1912	.021	.025	.096	.029	−.066
1917	.020	.010	.081	.043	−.052
1922	.121	.098	.168	.160	.065
1927	.093	.057	.127	.149	.054
1932	.120	.070	.141	.192	.097
1937	.092	.029	.099	.179	.085
1942	.168	.091	.161	.271	.176
1947	.179	.088	.159	.298	.203
1952	.108	.003	.074	.242	.147
1957	.044	−.074	−.004	.194	.099
1962	.037	−.095	−.024	.203	.108
1967	−.065	−.210	−.139	.117	.023
1972	−.167	−.326	−.255	.031	−.064
1977	−.223	−.396	−.325	−.009	−.104
1982	−.209	−.395	−.324	.021	−.074
1987	−.245	−.445	−.374	...	−.095

NOTE.—Data are from Schwadel and Stout (2012). $\sum=0$: sum-to-zero coding. $\beta_{first}=0$: reference-group coding with the first group omitted for each effect. $\beta_{last}=0$: reference-group coding with the last group omitted for each effect. $\beta^*_{first}=0$: $\beta_{first}=0$ estimates transformed to the sum-to-zero scale. $\beta^*_{last}=0$: $\beta_{last}=0$ estimates transformed to the sum-to-zero scale.

TABLE A4

INTRINSIC ESTIMATOR ESTIMATES UNDER TWO DIFFERENT SUM-TO-ZERO CODING SCHEMES FOR A SIMULATED DATA SET

Category/Effects	last $\sum=0$	first $\sum=0$*
Intercept	8.333	8.333
	Age	
1	.708	1.208
2	.333	.333
3	−1.042	−1.542

TABLE A4 (*Continued*)

Category/Effects	last $\sum=0$	first $\sum=0$*
	Period	
1	1.958	1.458
2	.333	.333
3	−2.292	−1.792
	Cohort	
1	.750	1.750
2	−.625	−.125
3	−1.000	−1.000
4	−.375	−.875
5	1.250	.250

NOTE.—Last $\sum=0$: estimates obtained under the sum-to-zero coding with the last group omitted for each effect. First $\sum=0$*: estimates obtained under the sum-to-zero coding with the first group omitted for each effect, then transformed to the sum-to-zero coding with the last group omitted scale.

REFERENCES

Alwin, Duane F. 1991. "Family of Origin and Cohort Differences in Verbal-Ability." *American Sociological Review* 56 (5): 625–38.

Belsley, David. A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons.

Clark, April, and Marie Eisenstein. 2013. "Interpersonal Trust: An Age-Period-Cohort Analysis Revisited." *Social Science Research* 42 (2): 361–75.

Fienberg, Stephen E., and William M. Mason. 1979. "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data." *Sociological Methodology* 10:1–67.

Fu, Wenjiang J. 2000. "Ridge Estimator in Singular Design with Application to Age-Period-Cohort Analysis of Disease Rates." *Communications in Statistics-Theory and Methods* 29 (2): 263.

———. 2008. "A Smoothing Cohort Model in Age-Period-Cohort Analysis with Applications to Homicide Arrest Rates and Lung Cancer Mortality Rates." *Sociological Methods and Research* 36 (3): 327–61.

Gelman, Andrew. 2004. "Parameterization and Bayesian Modeling." *Journal of the American Statistical Association* 99 (466): 537–45.

Glenn, Norval D. 1976. "Cohort Analysts' Futile Quest: Statistical Attempts to Separate Age, Period and Cohort Effects." *American Sociological Review* 41 (5): 900–904.

———. 1994. "Television Watching, Newspaper Reading, and Cohort Differences in Verbal-Ability." *Sociology of Education* 67 (3): 216–30.

———. 2005. *Cohort Analysis*. Thousand Oaks, Calif.: Sage Publications.

Heckman, James, and Richard Robb. 1985. ''Using Longitudinal Data to Estimate Age, Period, and Cohort Effects in Earnings Equations.'' Pp. 137–50 in *Cohort Analysis in Social Research*, edited by W. M. Mason and S. E. Fienberg. New York: Springer-Verlag.

Holford, Theodore R. 1983. "The Estimation of Age, Period and Cohort Effects for Vital Rates." *Biometrics* 39 (2): 311–24.

Keyes, Katherine, and Richard Miech. 2013. “Age, Period, and Cohort Effects in Heavy Episodic Drinking in the U.S. from 1985 to 2009.” *Drug and Alcohol Dependence*. doi: 10.1016/j.drugalcdep.2013.01.019.

Knoke, David, and Michael Hout. 1974. “Social and Demographic Factors in American Political Party Affiliations, 1952–72.” *American Sociological Review* 39 (5): 700–713.

Kupper Lawrence L., Joseph M. Janis, Azza Karmous, and Bernard G. Greenberg. 1985. “Statistical Age-Period-Cohort Analysis: A Review and Critique.” *Journal of Chronic Diseases* 38 (10): 811–30.

Langley, John, Ari Samaranayaka, J. Davie, and A. J. Campbell. 2011. “Age, Cohort and Period Effects on Hip Fracture Incidence: Analysis and Predictions from New Zealand Data, 1974–2007.” *Osteoporosis International* 22 (1): 105–11.

Luo, Liying. 2013*a*. “Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem.” *Demography* 50 (6): 1945–67.

———. 2013*b*. “Paradigm Shift in Age-Period-Cohort Analysis: A Response to Yang and Land, O’Brien, Held and Riebler, and Fienberg.” *Demography* 50 (6): 1985–88.

Masters, Ryan K., Robert A. Hummer, Daniel A. Powers, Audrey Beck, Shih-Fan Lin, and Brian Karl Finch. 2014. “Long-Term Trends in Adult Mortality for U.S. Blacks and Whites: An Examination of Period- and Cohort-Based Changes.” *Demography* (51): 2047–73.

Morgan, Stephen, and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

O’Brien, Robert M. 2011. “Constrained Estimators and Age-Period-Cohort Models.” *Sociological Methods and Research* 40 (3): 419–52.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference,* 2d ed. Cambridge: Cambridge University Press.

Rodgers, Willard L. 1982. “Estimable Functions of Age, Period, and Cohort Effects.” *American Sociological Review* 47 (6): 774–87.

Schwadel, Philip. 2011. “Age, Period, and Cohort Effects on Religious Activities and Beliefs.” *Social Science Research* 40 (1):181–92.

Schwadel, Philip, and Michael Stout. 2012. “Age, Period and Cohort Effects on Social Capital.” *Social Forces* 91 (1): 233–52.

Yang, Yang. 2008. “Trends in U.S. Adult Chronic Disease Mortality, 1960–1999: Age, Period, and Cohort Variations.” *Demography* 45 (2): 387–416.

Yang, Yang, Sam Schulhofer-Wohl, Wenjiang J. Fu, Kenneth C. Land. 2008. “The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It.” *American Journal of Sociology* 113 (6):1697–1736.

Yang, Yang, Wenjiang Fu, and Kenneth C. Land. 2004. “A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models.” *Sociological Methodology* 34: 75–110.

———. 2008. “Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Fixed or Random Effects?” *Sociological Methods and Research* 36 (3): 297–326.