MPC
Minnesota Population Center
UNIVERSITY OF MINNESOTA

The Sensitivity of the Intrinsic Estimator to Coding Schemes: A Comment on Yang, Schulhofer-Wohl, Fu, and Land*

Liying Luo
Department of Sociology
Minnesota Population Center, University of Minnesota

James Hodges
Division of Biostatistics, School of Public Health
University of Minnesota

Christopher Winship
Department of Sociology
Harvard University

Daniel Powers
Department of Sociology
University of Texas at Austin

March 2014

Working Paper No. 2014-1
https://doi.org/10.18128/MPC2014-1

*Direct correspondence to Liying Luo, 909 Social Sciences, 267 19th Ave. S., Minneapolis, MN 55455; e-mail: luoxx054@umn.edu. We thank John Robert Warren and Robert O'Brien for their helpful comments.

Abstract

The Intrinsic Estimator (IE) has been proposed to address the age-period-cohort problem and is believed by many to yield robust and reliable estimates. We, however, show that IE estimates are highly sensitive to one's choice of coding scheme or model parameterization. We reanalyze data from published articles to demonstrate that estimation results using one coding scheme (e.g., the zero-to-sum coding) can be dramatically different from those obtained using a different coding scheme (e.g., reference group coding). The results are so different that a researcher would sometimes reach opposite conclusions about the effects of age, period, and/or cohort depending on the seemingly innocuous choice of coding scheme. We provide a nontechnical explanation for this sensitivity; an appendix provides a mathematical proof.

Introduction

In a series of articles, Fu, Yang, and Land (Fu 2000, 2008; Fu and Hall 2006; 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*, vol. 113[6]: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. They argue that the IE produces estimates that approximate well the "true" age, period, and cohort trends, using the General Social Survey data as an example (Yang et al. 2008:1712-1716). 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 of age, period, and cohort in APC models (Yang et al. 2008:1716-1722). The IE now enjoys wide popularity in many disciplines and has been used in multiple empirical applications (e.g., Clark and Einsenstein 2013; Schwadel and Stout 2012; Schwadel 2011; Yang 2008).

O'Brien (2011) and Luo (forthcoming) 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 and they argue that this constraint is essentially arbitrary.

In this comment, we raise additional concerns about the robustness (i.e., sensitivity) and thus 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. We reanalyze data from three published articles to demonstrate that coding the APC model using one coding scheme (e.g., the zero-to-sum/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 often reach opposite conclusions about the effects of age, period, and cohort depending on the choice of coding scheme.

The IE's sensitivity to the coding scheme used in an analysis is in sharp contrast to fully identified statistical models, for which different coding schemes necessarily produce equivalent results after appropriate transformation. We provide a nontechnical explanation for this sensitivity; an appendix provides a mathematical proof. In addition, the appendix shows that for any choice of parameter estimates among the infinite number of solutions for an APC model, there is always a coding scheme in which the IE produces that specific set of estimates. In other words, one can choose any estimate one likes from the possible set of estimates and there will be a coding scheme under which the IE produces that estimate. Because the choice of coding scheme is arbitrary, it follows that the IE's choice of one estimate from the infinite number of solutions for an APC model is also arbitrary.

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 coefficient estimates of age, period, and cohort effects. Because all the possible estimates give the same fitted values, there is no way to use the data to choose among them. The possible estimates all lie on a line, called the "solution line," so if one fixes the value of one parameter estimate at any finite 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, that is, which set of parameter estimates, is privileged.

Traditional approaches to identifying APC models involve either setting some parameter(s) to zero, e.g., assuming there is no period effect, or setting two or more parameters to be equal, e.g., setting adjacent cohorts or periods to have equal effects. 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 1982a, 1982b).

Like traditional estimators, the IE also achieves identification by imposing a constraint (O'Brien 2011; Luo forthcoming), 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. Second, the IE depends on the design matrix in two senses. First, for a given coding scheme (parameterization), the constraint implicit in the IE depends on the number of age and period (and thus cohort) categories. Also, however, 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.

Following Glenn (2005:20), Yang et al. (2008:1699) argue that an APC analysis should be evaluated with respect to its ability to provide correct estimates more often than not, that is, the true parameter estimates, or what O'Brien (2011) calls the data-generating parameters.[1] They conclude that the IE satisfies this criterion (Yang et al. 2008:1732). 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 in detail that the IE is sensitive to the coding scheme used, sometimes dramatically so. As such, there is no basis to the claim that it removes the effect of the design matrix or, given this, that it provides good estimates of the parameters that have generated the data.[2]

Examples: How 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 the three coding schemes is used: sum-to-

[1] Some users of the IE appear to believe that it gives unbiased estimates of the true or data-generating parameters (e.g., see Schwadel and Stout 2012; Keyes and Miech 2011; 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 IE is not distinctive in this respect.

[2] 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. (2008:1707) 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.

zero/ANOVA coding, reference coding with the first group as the reference category, and reference coding with the last group as the reference category.

When working with categorical data, researchers often use different coding schemes, typically choosing a coding scheme because of interpretability or because it highlights a particular empirical 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. In fully identified models, the choice of coding scheme does not affect estimation results. In other words, when appropriately transformed, the parameter estimates are unaffected by the coding—by mathematical necessity, they must be identical. As we will show, in the case of underidentified models like the APC model, this is not the case.

Example 1: US Female Mortality Rate from 1960 to 1999

The first example is mortality rates for US females from 1960 to 1999, used in Yang, Fu, and Land (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 (p. 98). We replicated their estimates for age, period, and cohort effects using the sum-to-zero/ANOVA coding. In Table A1 in the Appendix, the "$\sum=0$" columns show their estimates, with each estimate interpreted 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, shown in Table A1 in the "$\beta_{first}=0$" columns, so each estimate can be interpreted as the difference of each age, period, or cohort group from the first group in each effect. We also computed the IE estimates using a reference coding with the last age, last period, and last cohort category as reference groups, shown in the

"β_{last}=0" columns in Table A1. Finally, we transformed the results of the reference group analyses so that the estimated age effects sum to zero, as do the estimated period and cohort effects, to allow direct comparison of the results from using the IE with the different coding schemes. Fig. 1 graphically presents the IE estimates using these coding schemes.

[Figure 1 about here]

Fig. 1 shows that the IE estimates can substantially differ 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 "β_{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 (identical results shown in Yang et al. (2004:98)) in Fig. 1 indicate an upward mortality trend across time periods from 1960 to 1999, 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 US females in the 1870 to 1874 birth cohort for the sum-to-zero coding (shown in Table A1's "$\sum$=0" column) is $\exp(1.008) = 2.740$ times the global mean, while the estimated morality rate for that birth cohort in the last-category reference coding is only $\exp(0.502) = 1.652$ times the global mean.

Example 2: Verbal Test Scores from 1976 to 2000

The IE estimates also critically depend on coding scheme in the example of vocabulary knowledge used in Yang et al. (2008:1712-16), where the authors were concerned with the age, period, and cohort trends in Americans' vocabularies. The outcome is the variable WORDSUM in the General Social Survey (GSS), 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" (2008: 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 age-period-cohort models, and concluded that the estimated trends are "quite similar" (p. 1716).

Table A2 in the Appendix and Fig. 2 show the IE estimates using the same three coding schemes used in the previous example. As above, we transformed the results using the "$\beta_{first}=0$" and "$\beta_{last}=0$" codings to the sum-to-zero coding so the estimated effects can be compared directly. The age, period, and cohort effects estimated by the IE technique shown in Fig. 2 dramatically differ 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 "$\beta_{first}=0$", they increase initially but decrease starting at the age of 30 to 39. The "$\beta_{last}=0$" coding, by contrast, shows that vocabulary knowledge increases through the age span considered.

[Figure 2 about here]

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 "$\beta_{first}=0$" coding shows a consistent increase, while the "$\beta_{last}=0$" 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 "$\beta_{first}=0$" coding shows strong evidence of an inter-cohort decline, while the "$\beta_{last}=0$" coding shows just the opposite, a consistent increase in vocabulary knowledge from the oldest to the youngest cohorts. Thus a researcher who used the "$\beta_{first}=0$" coding would reach opposite conclusions about the period and cohort trends in vocabulary knowledge to the conclusion by another researcher who happened to choose the "$\beta_{last}=0$" coding scheme.

Example 3: Trust from 1972 to 2010

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). Table A3 in the Appendix and Fig. 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 ("$\beta_{first}=0$") coding, and the last-reference -group ("$\beta_{last}=0$") coding.

[Figure 3 about here]

As Fig. 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 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 inter-cohort 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 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.[3] 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[4]. 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 $T(b_0)$ is, in general, no longer the point closest to the origin among points on

[3] 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.
[4] 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.

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 mathematical appendix, with sum-to-zero coding schemes that have different omitted categories.

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

[Figure 4 about here]

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 seemly 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.

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 resulted in 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.[5] As such, it is perfectly 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 be useful to have formal methods for analyzing the sensitivity of estimates to the

[5] Strictly speaking, there is no variation in the *linear component* of Age; the non-linear components of the three effects are identified (Holford 1983).

constraint (the IE or other) that is used to privilege one set of estimates.[6] This is a topic for future research.

Conclusion

"The Intrinsic Estimator for Age-Period-Cohort Analysis," published in 2008, has been cited 69 times as of September 2013 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 (see e.g., Keyes and Miech 2013; Lanley et al. 2011; Schwadel 2011). The empirical and mathematical results presented in this comment and in the appendix 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

[6] 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 estimates 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."

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. (1985:144-45)

References

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

Clark, April, and Marie Eisenstein. 2013. "Interpersonal Trust: An age–period–cohort Analysis Revisited." *Social Science Research* 42 (2):361-375.

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 Singulah Oesiun 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 & Research* 36 (3):327-361.

Fu, Wenjiang J., and Peter Hall. 2006. "Asymptotic Properties of Estimators in Age-Period-Cohort Analysis." *Statistics & Probability Letters* 76 (17):1925-1929.

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

———. 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):311-324.

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

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):105-111.

Luo, Liying. Forthcoming, “Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem.” *Demography*.

Miech, Richard, Steve Koester, and Brook Dorsey-Holliman. 2011. "Increasing US Mortality due to Accidental Poisoning: The Role of the Baby Boom Cohort." *Addiction* 106 (4):806-815.

O'Brien, Robert M. 2011. "Constrained Estimators and Age-Period-Cohort Models." *Sociological Methods & Research* 40 (3):419-452.

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

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

Schwadel, Philip. 2011. "Age, Period, and Cohort Effects on Religious Activities and Beliefs." *Social Science Research* 40 (1):181-192.

Schwadel, Philip, and Michael Stout. 2012. "Age, Period and Cohort Effects on Social Capital." *Social Forces* 91 (1):233-252.

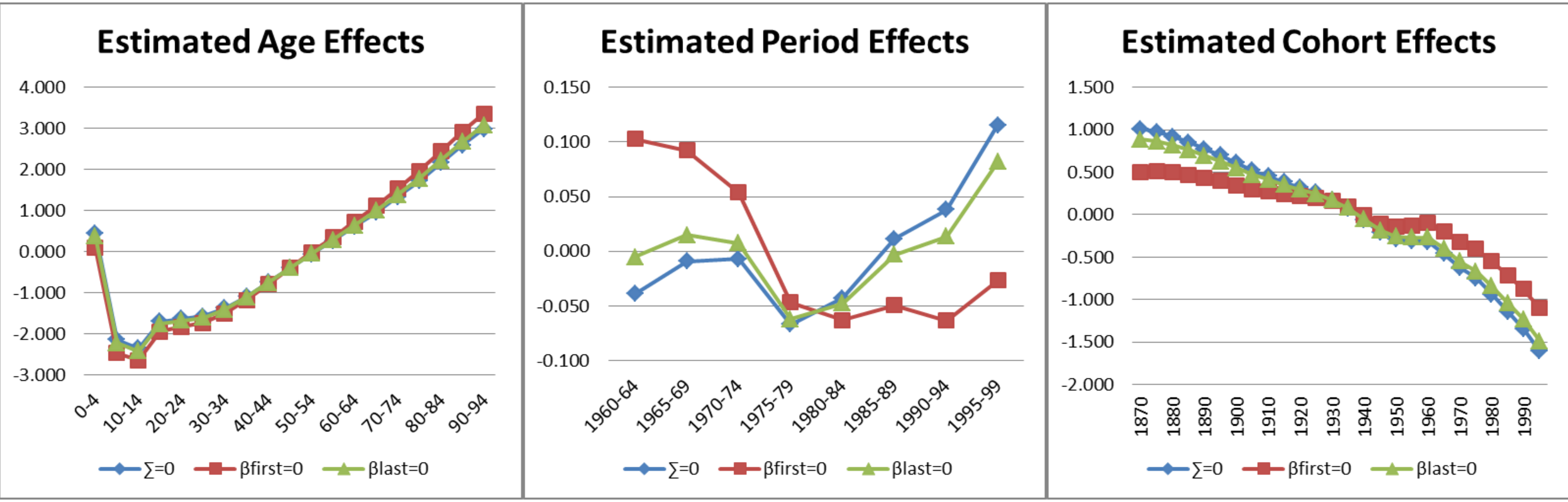
Yang, Yang. 2008. "Trends in U.S. Adult Chronic Disease Mortality, 1960-1999: Age, Period, and Cohort Variations." *Demography* 45 (2):387-416.

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, and Kenneth C. Land. 2006. "A Mixed Models Approach to the Age-Period-Cohort Analysis of Repeated Cross-Section Surveys, with an Application to Data on Trends in Verbal Test Scores." *Sociological Methodology* 36:75-97.

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

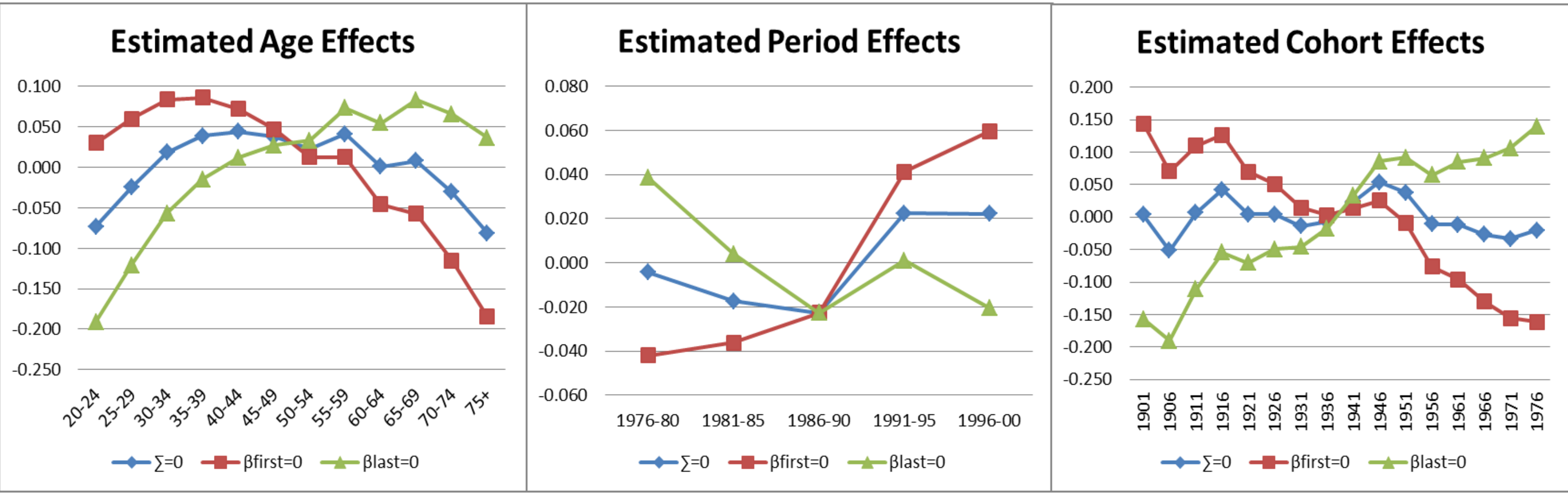
Figure 1. Estimated APC Trends in Mortality Using IE under Three Coding Schemes



NOTES:

1. Data source: Yang, Fu, and Land 2004, *Sociological Methodology;*
2. ∑=0: sum-to-zero coding;
3. βfirst = 0: reference-group coding wit the first group omitted for each effect;
4. βlast = 0: reference-group coding wit the last group omitted for each effect.

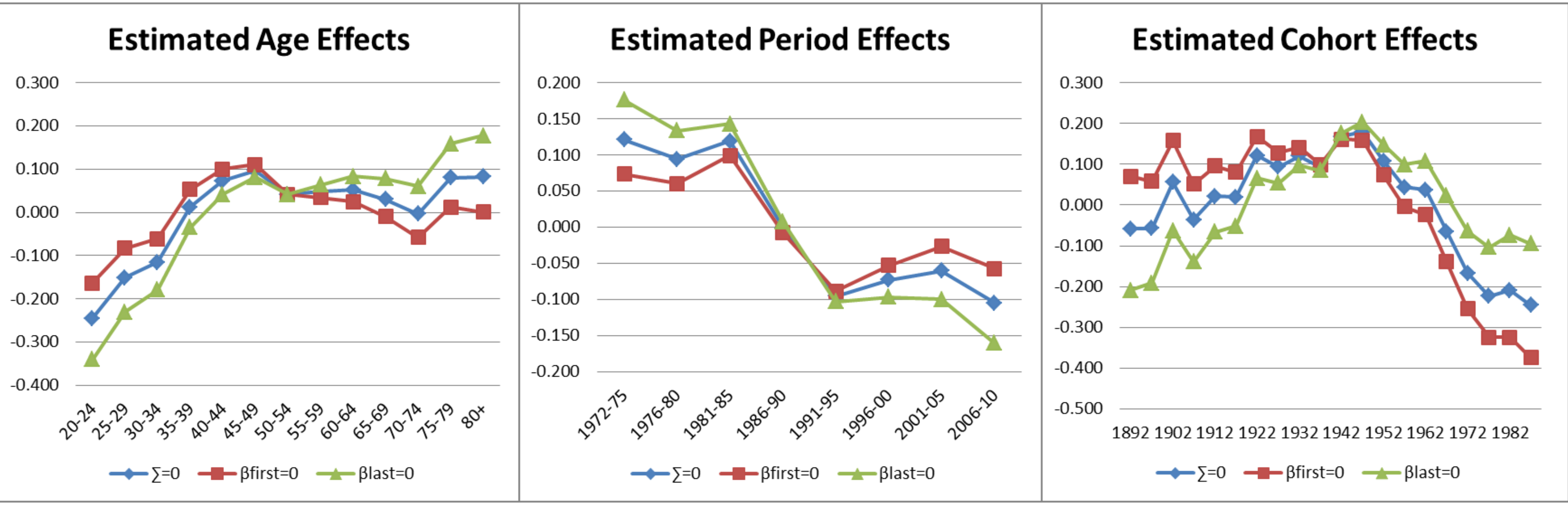
Figure 2. Estimated APC Trends in Vocabularies Using IE under Three Coding Schemes



NOTES:

1. Data source: Yang et al. 2008, *American Journal of Sociology*;
2. ∑=0: sum-to-zero coding;
3. βfirst = 0: reference-group coding wit the first group omitted for each effect;
4. βlast = 0: reference-group coding wit the last group omitted for each effect.

Figure 3. Estimated APC Trends in Trust Using IE under Three Coding Schemes



NOTES:

1. Data source: the General Social Survey, 1972-2010 used in Schwadel and Stout 2012;
2. ∑=0: sum-to-zero coding;
3. βfirst = 0: reference-group coding wit the first group omitted for each effect;
4. βlast = 0: reference-group coding wit the last group omitted for each effect.

Figure 4a. IE Solution in the Sum-to-Zero Coding

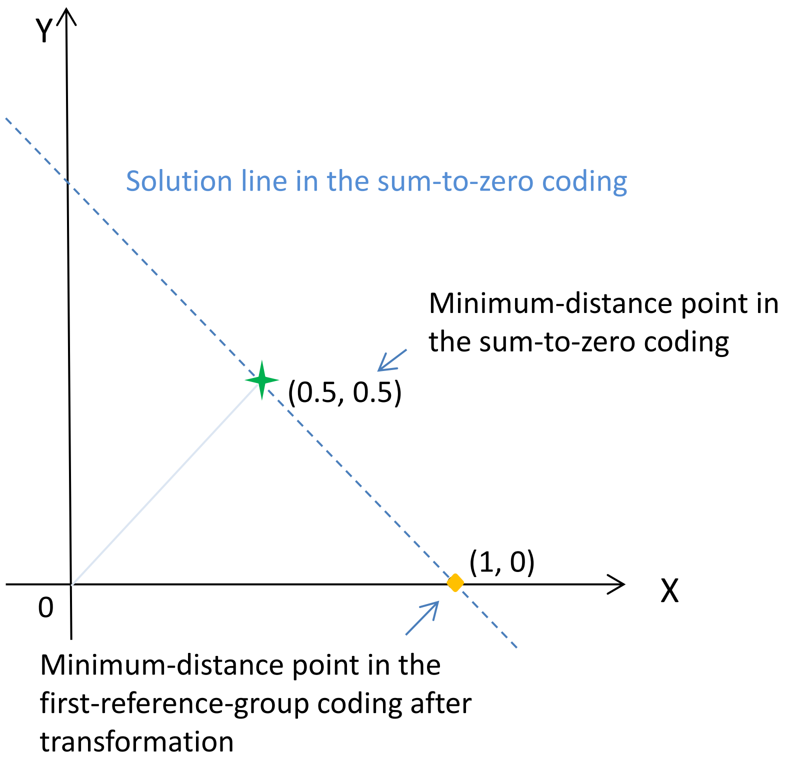
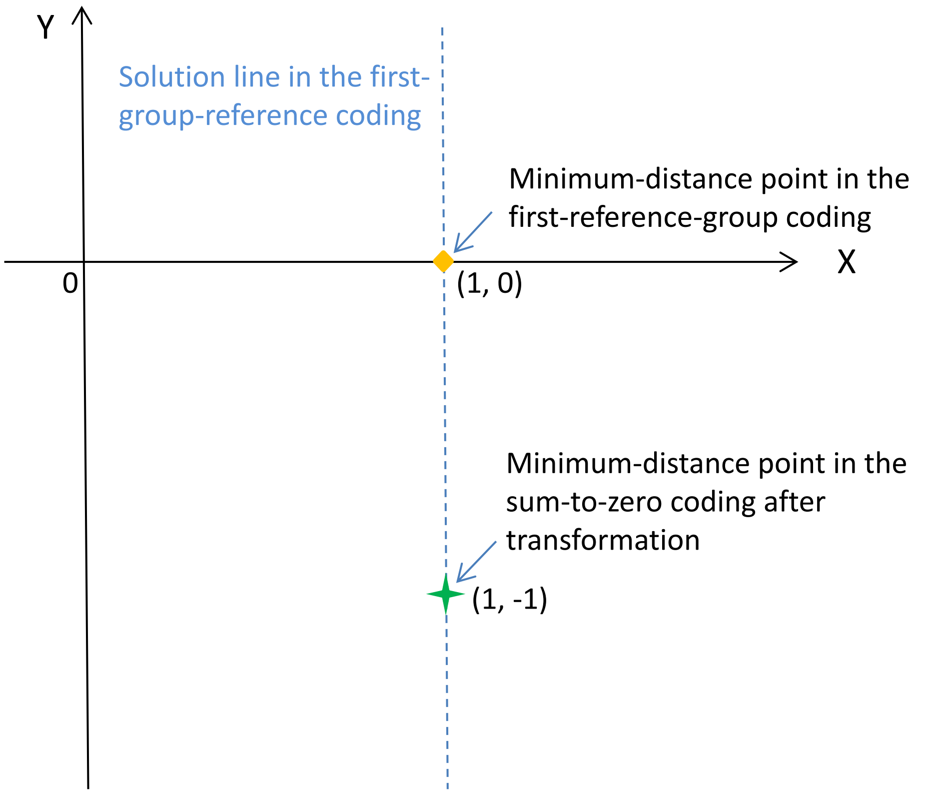


Figure 4b. IE Solution in the First-Reference-Group Coding



Mathematical Appendix

1. Preliminaries: Defining notation; the IE estimate

Suppose we have an outcome measure y that we are willing to treat as continuous, e.g., 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 $\mathbf{B}_0'\mathbf{B}_0 = 1$, i.e., $\mathbf{B}_0$ has Euclidean length 1. For a given dataset $\mathbf{y}$, the ordinary least-squares estimates of $\mathbf{b}$ satisfy the equation $\mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{y}$, 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 dataset $\mathbf{y}$. 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 $(\mathbf{b}_1 + r\mathbf{B}_0)'(\mathbf{b}_1 + r\mathbf{B}_0)$. Simple calculus shows that the squared length is minimized for $r = -\mathbf{B}_0'\mathbf{b}_1$, so the IE estimate is

$$\begin{aligned} \mathbf{b}_0 &= \mathbf{b}_1 - (\mathbf{B}_0'\mathbf{b}_1)\mathbf{B}_0 \\ &= \mathbf{b}_1 - \mathbf{B}_0(\mathbf{B}_0'\mathbf{b}_1) \quad \text{because } \mathbf{B}_0'\mathbf{b}_1 \text{ is a scalar} \qquad (1) \\ &= (\mathbf{I} - \mathbf{B}_0\mathbf{B}_0')\mathbf{b}_1 \quad \text{where } \mathbf{I} \text{ is the identity matrix of order } 2(a+p)-3. \end{aligned}$$

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

2. Re-parameterizing 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:

$$\mathbf{Xb} = \mathbf{XT}^{-1}\mathbf{Tb} = \mathbf{X(T)b(T)} \qquad (2)$$

where $\mathbf{X(T)} = \mathbf{XT}^{-1}$ is the design matrix in the new coding scheme and $\mathbf{b(T)} = \mathbf{Tb}$ is the parameter in the new coding scheme corresponding to $\mathbf{b}$ in the original coding scheme. Section 4 of this Appendix shows 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 dataset $\mathbf{y}$, and that the IE estimate for this dataset 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 $s = \mathbf{B}_0'\mathbf{b}$. 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 $\mathbf{b}'\mathbf{T}'\mathbf{Tb} = (\mathbf{b}_0 + s\mathbf{B}_0)'\mathbf{T}'\mathbf{T}(\mathbf{b}_0 + s\mathbf{B}_0)$. This distance is a quadratic in the scalar s, and simple calculus shows that this squared distance is minimized by

$$s_{\mathbf{T}} = -\mathbf{B}_0{}'\mathbf{T}'\mathbf{T}\mathbf{b}_0 / \mathbf{B}_0{}'\mathbf{T}'\mathbf{T}\mathbf{B}_0. \tag{3}$$

Thus, the IE solution in the new coding scheme, backtransformed 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 (a) $\mathbf{T}$ is an orthogonal matrix, or (b) $\mathbf{T}$ 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 $\mathbf{b}_0$, i.e., on the specific dataset $\mathbf{y}$. 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 $\mathbf{T} = \mathbf{UDV}'$, for $\mathbf{U}$ and $\mathbf{V}$ orthogonal matrices and $\mathbf{D}$ diagonal, then $\mathbf{D}$'s diagonal elements describe the differential stretching applied to directions defined by $\mathbf{V}'$. Section 4 below gives an example.

3. 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 $\mathbf{y}$, and that we have chosen a coding scheme, which we will call the original coding scheme. Then this implies a design matrix $\mathbf{X}$, a null vector $\mathbf{B}_0$, and the IE estimate $\mathbf{b}_0$. Any other solution to the equation $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$ 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 $\mathbf{T}$ of dimension $r = 2(a + p) - 3$ and a new coding scheme $\mathbf{Tb}$ such that the IE estimate in the new coding scheme, backtransformed 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.

<u>Proof of the main claim</u>. 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 in Section 2, for a given transformation (re-coding) $\mathbf{T}$, the IE estimate in the new coding scheme, back-transformed to the original coding scheme, has

$$s_{\mathbf{T}} = -\mathbf{B}_0{}'\mathbf{T}'\mathbf{T}\mathbf{b}_0 / \mathbf{B}_0{}'\mathbf{T}'\mathbf{T}\mathbf{B}_0. \tag{4}$$

We need to prove that for any real number t, we can choose a $\mathbf{T}$ such that

$$t = s_{\mathbf{T}} = -\mathbf{B}_0{}'\mathbf{T}'\mathbf{T}\mathbf{b}_0 / \mathbf{B}_0{}'\mathbf{T}'\mathbf{T}\mathbf{B}_0. \tag{5}$$

Note: 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.

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

For any legal $\mathbf{G}$ and $\mathbf{D}$, $\mathbf{B}_0'\mathbf{T}'\mathbf{T}\mathbf{b}_0 = \sum_i a_i c_i d_i$, where $\mathbf{B}_0'\mathbf{G} = (a_1, a_2, \ldots, a_r)$ and $\mathbf{b}_0'\mathbf{G} = (c_1, c_2, \ldots, c_r)$ and the sum runs over $i = 1, \ldots, r$. (Note that $\sum_i a_i^2 = \mathbf{B}_0'\mathbf{G}\mathbf{G}'\mathbf{B}_0 = 1$ because $\mathbf{B}_0$ has length 1, and $\sum_i c_i^2 = \mathbf{b}_0'\mathbf{b}_0$.) With these definitions, $\mathbf{B}_0'\mathbf{T}'\mathbf{T}\mathbf{B}_0 = \sum_i a_i^2 d_i$. Thus, we need to choose $\mathbf{G}$ — i.e., choose the a_i and c_i — and choose $\mathbf{D}$ — i.e., choose the d_i — so that

$$t = -\sum_i a_i c_i d_i \,/\, \sum_i a_i^2 d_i. \tag{6}$$

If $d_2, \ldots, d_r$ are fixed at some values and d_1 is made very large, then $-\sum_i a_i c_i d_i \,/\, \sum_i a_i^2 d_i$ 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 $\boldsymbol{\beta}_0 = \mathbf{b}_0(\mathbf{b}_0'\mathbf{b}_0)^{-0.5}$, so $\boldsymbol{\beta}_0'\boldsymbol{\beta}_0 = 1$ and $\boldsymbol{\beta}_0'\mathbf{B}_0 = 0$. Then let the first column of $\mathbf{G}$ be $\mathbf{G}_1 = \alpha\phi\boldsymbol{\beta}_0 + (1-\alpha)\mathbf{B}_0$, 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 $c_1 = (\mathbf{b}_0'\mathbf{b}_0)^{0.5}\alpha\phi$, so $-c_1 / a_1 = -\phi(\mathbf{b}_0'\mathbf{b}_0)^{0.5}\alpha / (1-\alpha)$. Set $\alpha = |t| / ((\mathbf{b}_0'\mathbf{b}_0)^{0.5} + |t|)$: then $-c_1 / a_1 = t$.

<u>Secondary claim:</u> 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 $\mathbf{b}_{0u} = (\mathbf{I} - \mathbf{B}_0\mathbf{B}_0')\mathbf{b}$, 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 $\mathbf{Tb}_{0u}$ 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

$$\begin{aligned} &\left(\mathbf{I} - \mathbf{B}_0\mathbf{B}_0'\right)\mathbf{b} - \mathbf{T}^{-1}\left(\mathbf{I} - \mathbf{T}\mathbf{B}_0\mathbf{B}_0'\mathbf{T}'\right)\mathbf{Tb} \\ &= \mathbf{B}_0\left[\mathbf{B}_0'\left(\mathbf{T}'\mathbf{T} - \mathbf{I}\right)\mathbf{b}\right], \end{aligned} \tag{7}$$

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 $t = \mathbf{B}_0'(\mathbf{T}'\mathbf{T} - \mathbf{I})\mathbf{b}$. Recalling that $\mathbf{b} = \mathbf{b}_{0u} + s\mathbf{B}_0$ for a particular scalar s, we need

$$t = \mathbf{B}_0'\left(\mathbf{T}'\mathbf{T} - \mathbf{I}\right)\mathbf{b} = \mathbf{B}_0'\mathbf{T}'\mathbf{T}\mathbf{b}_{0u} + s\,\mathbf{B}_0'\mathbf{T}'\mathbf{T}\mathbf{B}_0 - s \tag{8}$$

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

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 $\mathbf{T}'\mathbf{T} = \mathbf{GDG}'$, 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

$$t = \sum_i a_i c_i d_i + s\sum_i a_i^2 d_i - s. \tag{9}$$

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 $g(d_1) = \sum_i a_i c_i d_i + s\sum_i a_i^2 d_i - s$; then g's derivative with respect to d_1 is

$$g'(d_1) = a_1 c_1 + s a_1^2 \tag{10}$$

which does not depend on d_1. To get the desired result, we need only show that we can pick $\mathbf{G}$ so that $a_1c_1 + sa_1^2$ is positive for the given s and that we also can pick $\mathbf{G}$ so that $a_1c_1 + sa_1^2$ is negative for the given s. Then we can fix a_i, c_i, and d_i for $i = 2,\ldots,4$ 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 $s = 0$, $g'(d_1) = a_1c_1$. As in the proof of the main claim, let the first column of $\mathbf{G}$ be $\mathbf{G}_1 = \alpha\phi\boldsymbol{\beta}_0 + (1-\alpha)\mathbf{B}_0$, and set $\alpha = 0.5$. Then $g'(d_1) = a_1c_1 = 0.25(\mathbf{b}_0{}'\mathbf{b}_0)^{0.5}\phi$, which is made positive or negative by choosing positive or negative ϕ respectively. Now suppose $s \neq 0$. Then $g'(d_1) = a_1c_1 + sa_1^2 > 0$ if $c_1 / a_1 > -s$, and $g'(d_1) = a_1c_1 + sa_1^2 < 0$ 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.

4. 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 $\mathbf{T_1b}$ 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 3 age groups and 4 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 $\mathbf{Tb}$ 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.

5. 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 **b** has 9 elements: the intercept, the first 2 age group effects, the first 2 period effects, and the first 4 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 9 elements: the intercept, the last 2 age group effects; the last 2 period effects; and the last 4 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 **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 9 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. Fig. A1 reports 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 effects omitted and the same coding with the first category omitted. The resulting two sets of IE estimates are different. For example, the estimated cohort effects for the first cohort is 0.75 under the first type of sum-to-zero coding, whereas the estimated effects is 1.75 under the second type.

Table A1. Estimated Age, Period, and Cohort Effects on Mortality under Three Coding Schemes.

Intercept Estimates			**Age Effects Estimates**				**Period Effects Estimates**				**Cohort Effects Estimates**			
∑=0	βfirst=0	βlast=0	Category	∑=0	βfirst=0	βlast=0	Category	∑=0	βfirst=0	βlast=0	Category	∑=0	βfirst=0	βlast=0
-5.400	-5.400	-5.400	0-4	0.453	0.088	0.366	1960-64	-0.039	0.103	-0.005	1870	1.008	0.502	0.887
			5-9	-2.144	-2.468	-2.221	1965-69	-0.009	0.092	0.015	1875	0.977	0.511	0.866
			10-14	-2.354	-2.637	-2.421	1970-74	-0.007	0.054	0.007	1880	0.922	0.496	0.820
			15-19	-1.704	-1.947	-1.762	1975-79	-0.067	-0.047	-0.062	1885	0.853	0.468	0.761
			20-24	-1.630	-1.833	-1.678	1980-84	-0.043	-0.063	-0.047	1890	0.776	0.431	0.693
			25-29	-1.571	-1.733	-1.610	1985-89	0.011	-0.049	-0.003	1895	0.698	0.394	0.626
			30-34	-1.377	-1.499	-1.406	1990-94	0.038	-0.063	0.014	1900	0.610	0.347	0.548
			35-39	-1.091	-1.172	-1.111	1995-99	0.115	-0.026	0.082	1905	0.522	0.299	0.468
			40-44	-0.751	-0.791	-0.760					1910	0.455	0.273	0.412
			45-49	-0.398	-0.398	-0.398					1915	0.383	0.241	0.349
			50-54	-0.057	-0.016	-0.047					1920	0.317	0.216	0.293
			55-59	0.266	0.347	0.286					1925	0.262	0.201	0.247
			60-64	0.610	0.732	0.639					1930	0.178	0.158	0.173
			65-69	0.956	1.118	0.995					1935	0.077	0.097	0.082
			70-74	1.331	1.534	1.380					1940	-0.067	-0.006	-0.052
			75-79	1.724	1.967	1.782					1945	-0.204	-0.103	-0.180
			80-84	2.157	2.440	2.224					1950	-0.287	-0.145	-0.253
			85-89	2.590	2.914	2.667					1955	-0.312	-0.129	-0.268
			90-94	2.988	3.353	3.076					1960	-0.319	-0.096	-0.266
											1965	-0.460	-0.197	-0.398
											1970	-0.620	-0.316	-0.547
											1975	-0.748	-0.403	-0.665
											1980	-0.934	-0.549	-0.842
											1985	-1.137	-0.712	-1.036
											1990	-1.342	-0.876	-1.231
											1995	-1.607	-1.100	-1.486

Table A2. Estimated Age, Period, and Cohort Effects on Vocabularies under Three Coding Schemes

Intercept Estimates			**Age Effects Estimates**				**Period Effects Estimates**				**Cohort Effects Estimates**			
∑=0	βfirst=0	βlast=0	Category	∑=0	βfirst=0	βlast=0	Category	∑=0	βfirst=0	βlast=0	Category	∑=0	βfirst=0	βlast=0
-2.820	-2.820	-2.820	20-24	-0.073	0.030	-0.191	1976-80	-0.004	-0.042	0.038	1901	0.003	0.144	-0.157
			25-29	-0.025	0.059	-0.121	1981-85	-0.017	-0.036	0.004	1906	-0.052	0.070	-0.191
			30-34	0.018	0.084	-0.057	1986-90	-0.023	-0.023	-0.023	1911	0.006	0.110	-0.111
			35-39	0.039	0.086	-0.015	1991-95	0.022	0.041	0.001	1916	0.042	0.126	-0.054
			40-44	0.044	0.072	0.012	1996-00	0.022	0.060	-0.021	1921	0.004	0.070	-0.070
			45-49	0.038	0.047	0.027					1926	0.004	0.051	-0.049
			50-54	0.022	0.013	0.033					1931	-0.014	0.014	-0.046
			55-59	0.041	0.013	0.073					1936	-0.007	0.003	-0.017
			60-64	0.001	-0.046	0.054					1941	0.023	0.014	0.034
			65-69	0.008	-0.058	0.083					1946	0.054	0.025	0.086
			70-74	-0.031	-0.115	0.066					1951	0.038	-0.009	0.091
			75+	-0.081	-0.184	0.036					1956	-0.011	-0.076	0.064
											1961	-0.011	-0.096	0.085
											1966	-0.027	-0.130	0.090
											1971	-0.033	-0.155	0.106
											1976	-0.021	-0.161	0.140

Table A3. Estimated Age, Period, and Cohort Effects on Trust under Three Coding Schemes.

Intercept Estimates

∑=0	βfirst=0	βlast=0
-0.996	-0.996	-0.996

Age Effects Estimates

Category	∑=0	βfirst=0	βlast=0
20-24	-0.245	-0.163	-0.340
25-29	-0.151	-0.083	-0.230
30-34	-0.116	-0.062	-0.179
35-39	0.013	0.053	-0.035
40-44	0.073	0.100	0.041
45-49	0.097	0.111	0.081
50-54	0.041	0.041	0.041
55-59	0.047	0.034	0.063
60-64	0.052	0.025	0.084
65-69	0.030	-0.010	0.078
70-74	-0.004	-0.058	0.060
75-79	0.080	0.012	0.159
80+	0.082	0.001	0.177

Period Effects Estimates

Category	∑=0	βfirst=0	βlast=0
1972-75	0.121	0.073	0.176
1976-80	0.094	0.060	0.134
1981-85	0.119	0.099	0.143
1986-90	-0.001	-0.008	0.007
1991-95	-0.095	-0.088	-0.103
1996-00	-0.073	-0.053	-0.097
2001-05	-0.060	-0.026	-0.100
2006-10	-0.105	-0.058	-0.161

Cohort Effects Estimates

Category	∑=0	βfirst=0	βlast=0
1892	-0.058	0.071	-0.209
1897	-0.057	0.059	-0.192
1902	0.056	0.158	-0.062
1907	-0.036	0.052	-0.139
1912	0.021	0.096	-0.066
1917	0.020	0.081	-0.052
1922	0.121	0.168	0.065
1927	0.093	0.127	0.054
1932	0.120	0.141	0.097
1937	0.092	0.099	0.085
1942	0.168	0.161	0.176
1947	0.179	0.159	0.203
1952	0.108	0.074	0.147
1957	0.044	-0.004	0.099
1962	0.037	-0.024	0.108
1967	-0.065	-0.139	0.023
1972	-0.167	-0.255	-0.064
1977	-0.223	-0.325	-0.104
1982	-0.209	-0.324	-0.074
1987	-0.245	-0.374	-0.095

Fig. A1. IE Estimates under Two Different Sum-To-Zero Coding Schemes for Simulated Data