Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The Non-uniqueness Property of the Intrinsic Estimator in APC Models.

Abstract

An important property of the intrinsic estimator is discussed, which has received no attention in literature so far: the age, period, and cohort estimates of the intrinsic estimator are not unique but vary with the parameterization and reference categories chosen for these variables. We give a formal prove of the non-uniqueness property for effect coding and dummy variable coding. For data on female mortality in the U.S. over the years 1960-1999, we show that the variation in the results obtained for different parameterizations and reference categories is substantial and leads to contradictory conclusions. We conclude that the non-uniqueness property is a new argument to not routinely apply the intrinsic estimator.

keywords: APC, identification problem, intrinsic estimator, contrast, reference category, effect coding

B. Pelzer, corresponding author

Assistant professor of Sociology and Social Sciences Research Methods

Radboud University Nijmegen, the Netherlands

P.O. Box 9104

6500 HC Nijmegen

The Netherlands

email: b.pelzer@maw.ru.nl

tel: 0031 243612038

fax: 0031 243612351

M. te Grotenhuis

Assistant professor of Sociology and Social Sciences Research Methods

Radboud University Nijmegen, the Netherlands

R. Eisinga

Professor of Sociology and Social Sciences Research Methods

Radboud University Nijmegen, the Netherlands

A. W. Schmidt-Catran

Research associate

Institute of Sociology and Social Psychology

University of Cologne, Germany

1 Introduction

The December 2013 issue of this journal included a number of contributions which all addressed the Intrinsic Estimator (IE). Fienberg, Held and Riebler, Luo, and O'Brien (2013) all seem to agree that the IE is of rather limited value to simultaneously estimate age, period, and cohort effects. Luo (2013) for instance demonstrated that the IE estimates are biased in case the true parameters of age, period, and cohort show a linear trend that diverges from the one implied by the IE constraint. She concludes that the IE should not be used. In contrast, in their reply to Luo, the IE developers Yang and Land (2013) are convinced about IE's potential for practical research. Which conclusion must one draw from these diverging expert viewpoints on the IE? Should researchers still consider using the IE, given that its application is validated by the three-step procedure proposed by Yang and Land, or should they abandon the whole IE idea right from the start and use other models?

In this paper, we will demonstrate the non-uniqueness of the IE, an important property of the IE that has been overlooked in the aforementioned discussion. This property puts the IE method in a different perspective and may have consequences for future use.

We will first explain the IE, without the use of matrix algebra. Next, we will prove the non-uniqueness property. We finally show the different results obtained by applying different IE solutions to fictitious data as well as to real data on female mortality. Based on our findings, we recommend researchers not to apply the IE routinely to their data, even if the three-step procedure suggested by Yang et al. (2013) would justify its use.

2 The Intrinsic Estimator

To explain how the IE parameters for age, period, and cohort are estimated, we use fictitious data which are limited to three periods, three ages, and five cohorts. With these data, we will explain the IE in an accessible way without relying on matrix notation (readers less interested in the mathematical explanation of the IE may read this introductory paragraph and skip the paragraph where we explain the IE constraint and continue with section 3). We start our explanation with the so-called APC accounting or multiple classification equation, which in regression format reads as:

$$Y = \beta_0 + \sum_{i=1}^{3} \alpha_i A_i + \sum_{j=1}^{3} \beta_j P_j + \sum_{k=1}^{5} \gamma_k C_k + e . \tag{1}$$

In Eq. (1), Y denotes the value of the dependent variable for a given unit (typically a person), and A_i, P_j, and C_k represent independent dummy variables indicating whether or not the unit belongs to age i, period j, and cohort k. Further, β_0 denotes the intercept and e represents the unit's error term. To estimate the parameters α_i, β_j, and γ_k we follow Yang (2004) and apply the following constraints:

$$\sum_{i=1}^{3} \alpha_i = 0 \ , \ \sum_{j=1}^{3} \beta_j = 0, \text{ and } \sum_{k=1}^{5} \gamma_k = 0 .$$

Given the first constraint it follows that in Eq. (1) $\alpha_3 = -\alpha_1 - \alpha_2$ and hence we can rewrite the age effects in Eq. (1) as follows:

$$Y = \beta_0 + \alpha_1 A_1 + \alpha_2 A_2 + (-\alpha_1 - \alpha_2) A_3 + \ ...$$
$$= \beta_0 + \alpha_1 (A_1 - A_3) + \alpha_2 (A_2 - A_3) + \ ...$$

4

Note that in this last equation α_3 is omitted; its value can be directly derived from α_1 and α_2. Also note that the three age dummies in (1) have been replaced by the two differences $A_1 - A_3$ and $A_2 - A_3$. In the same way we can substitute $-\beta_1 - \beta_2$ for β_3 and $-\gamma_1 - \gamma_2 - \gamma_3 - \gamma_4$ for γ_5 in (1). The resulting equation then is:

$$Y = \begin{aligned} & \beta_0^L + \alpha_1^L A_1^L + \alpha_2^L A_2^L + \beta_1^L P_1^L + \beta_2^L P_2^L + \\ & \gamma_1^L C_1^L + \gamma_2^L C_2^L + \gamma_3^L C_3^L + \gamma_4^L C_4^L + e, \end{aligned} \tag{2}$$

where superscript L denotes the elimination of the effects of the Last age, period, and cohort categories (i.e., α_3^L, β_3^L and γ_5^L) from the equation. Note that $A_i^L = A_i - A_3$, $P_j^L = P_j - P_3$ and $C_k^L = C_k - C_5$. The variables A_i^L, P_j^L, and C_k^L take value 1 if a case belongs to the subscripted age, period, and cohort, 0 if not and -1 if the case belongs to the last (omitted) category. According to these codings, the expectation of Y for the three age categories equals $\beta_0^L + \alpha_1^L$, $\beta_0^L + \alpha_2^L$, and $\beta_0^L - \alpha_1^L - \alpha_2^L$, when controlling for period and cohort. The mean of these three expectations equals β_0^L. The mean of the expectations of Y for the three periods and for the five cohorts also equals β_0^L. Consequently, the parameters α_i^L, β_j^L, and γ_k^L represent deviations from the mean, β_0^L, for the given categories of age, period, and cohort. This type of parameterization is known as "effect coding" (Hardy, 1993).

The IE Constraint

The APC parameters in Eq. (2) still cannot be estimated due to perfect dependency between the independent variables. For example, C_4^L can be written as a perfect linear combination of other variables in (2):

$$C_4^L = A_1^L - P_1^L + 2C_1^L + C_2^L. \tag{3}$$

If, in Eq. (2), we substitute for C_4^L the expression given in Eq. (3) and rearrange terms, we obtain:

$$\begin{aligned}
Y &= \beta_0^L + (\alpha_1^L + \gamma_4^L)A_1^L + \alpha_2^L A_2^L + (\beta_1^L - \gamma_4^L)P_1^L + \beta_2^L P_2^L + \\
&\quad (\gamma_1^L + 2\gamma_4^L)C_1^L + (\gamma_2^L + \gamma_4^L)C_2^L + \gamma_3^L C_3^L + e\ .
\end{aligned} \tag{4}$$

Equation (4) can be written more compactly:

$$Y = \boldsymbol{\beta}_0^L + \boldsymbol{\alpha}_1^L A_1^L + \boldsymbol{\alpha}_2^L A_2^L + \boldsymbol{\beta}_1^L P_1^L + \boldsymbol{\beta}_2^L P_2^L + \boldsymbol{\gamma}_1^L C_1^L + \boldsymbol{\gamma}_2^L C_2^L + \boldsymbol{\gamma}_3^L C_3^L + e\ , \tag{5}$$

Since in Eq. (5) variable C_4^L has been eliminated, the perfect dependency no longer exists and hence the regression parameters in (5) can be estimated. To emphasize that the parameters in (5) are estimable, we use bold face characters. For the parameters in (4) and (5) the following equalities hold:

$$\hat{\beta}_0^L = \boldsymbol{\beta}_0^L,\ \hat{\alpha}_2^L = \boldsymbol{\alpha}_2^L,\ \hat{\beta}_2^L = \boldsymbol{\beta}_2^L \text{ and } \hat{\gamma}_3^L = \boldsymbol{\gamma}_3^L. \tag{6}$$

The four estimates in (6) are the ones produced by the IE proposed by Yang (2004). Further, $\boldsymbol{\alpha}_1^L$ in (5) equals $\alpha_1^L + \gamma_4^L$ in (4), i.e., $\alpha_1^L = \boldsymbol{\alpha}_1^L - \gamma_4^L$. Similarly, the other three parameters, β_1^L, γ_1^L, and γ_2^L are related to γ_4^L, so in total we have:

$$\alpha_1^L = \boldsymbol{\alpha}_1^L - \gamma_4^L, \; \beta_1^L = \boldsymbol{\beta}_1^L + \gamma_4^L, \; \gamma_1^L = \boldsymbol{\gamma}_1^L - 2\gamma_4^L \text{ and } \gamma_2^L = \boldsymbol{\gamma}_2^L - \gamma_4^L. \tag{7}$$

The expressions in (7) show that, once an estimate of γ_4^L is found, the estimates of α_1^L, β_1^L, γ_1^L, and γ_2^L simply follow, given the estimates of the bold face parameters. However, due to the dependencies in (7) an extra constraint is needed to obtain estimates of α_1^L, β_1^L, γ_1^L, γ_2^L, and γ_4^L. The constraint that the IE applies, consists of the minimization of the sum of squares of these five parameters. For this sum of squared parameters, in terms of their estimates, we can write:

$$(\hat{\alpha}_1^L)^2 + (\hat{\beta}_1^L)^2 + (\hat{\gamma}_1^L)^2 + (\hat{\gamma}_2^L)^2 + (\hat{\gamma}_4^L)^2 =$$
$$(\boldsymbol{\alpha}_1^L - \hat{\gamma}_4^L)^2 + (\boldsymbol{\beta}_1^L + \hat{\gamma}_4^L)^2 + (\boldsymbol{\gamma}_1^L - 2\hat{\gamma}_4^L)^2 + (\boldsymbol{\gamma}_2^L - \hat{\gamma}_4^L)^2 + (\hat{\gamma}_4^L)^2 =$$
$$8(\hat{\gamma}_4^L)^2 - (2\boldsymbol{\alpha}_1^L - 2\boldsymbol{\beta}_1^L + 4\boldsymbol{\gamma}_1^L + 2\boldsymbol{\gamma}_2^L)\hat{\gamma}_4^L + \boldsymbol{\alpha}_1^L + \boldsymbol{\beta}_1^L + \boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L.$$

The minimum value of this sum of squares can be found by setting the first order derivative with respect to $\hat{\gamma}_4^L$ equal to zero, leading to:

$$16\hat{\gamma}_4^L - (2\boldsymbol{\alpha}_1^L - 2\boldsymbol{\beta}_1^L + 4\boldsymbol{\gamma}_1^L + 2\boldsymbol{\gamma}_2^L) = 0,$$

and thus to the following IE estimate of the true value of γ_4:

$$\hat{\gamma}_4^L = (\boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L) / 8. \tag{8}$$

Plugging this estimate of $\hat{\gamma}_4^L$ into the expressions given in (7) for α_1^L, β_1^L, γ_1^L, and γ_2^L leads to the IE estimates for these four remaining parameters. To summarize, the IE estimates of the regression parameters in (2) are those OLS estimates that have the smallest sum of squares for the five collinear variables A_1^L, P_1^L, C_1^L, C_2^L, and C_4^L. In the next section, we will show that the IE estimates are not unique but depend on which categories of age, period, and cohort are omitted and on the type of parameterization that is used.

3 The non-uniqueness of the intrinsic estimator

In the previous section we have shown that the IE obtains estimates with a minimum sum of squares of those parameters which are not identifiable due to collinearity. The IE estimates however depend on a) the choice of omitted categories and b) the type of parameterization applied. As an example of the dependence on omitted categories, we derive the IE estimates when the first categories of age, period, and cohort are omitted instead of the last ones, the latter being the default in Yang (2004).

Effect Coding with First Categories Omitted

If one omits the first categories of age, period, and cohort, Eq. 8 turns into:

$$\hat{\gamma}_4^F = (-\boldsymbol{\alpha}_3^L + \boldsymbol{\beta}_3^L + \boldsymbol{\gamma}_2^L - 2\boldsymbol{\gamma}_5^L) / 8. \tag{8a}$$

See appendix A for the proof of (8a). For the fourth cohort, the deviation $\hat{\gamma}_4^F$ from the mean differs from the deviation $\hat{\gamma}_4^L$ in case the last categories were omitted. In contrast, the estimate of the mean is the same, no matter whether the last or the first categories are omitted. In identified models, changing the omitted category has no impact on the deviations from the mean, but here it does. As a demonstration, we applied the default IE with the last categories omitted, and the alternative IE, with the first categories omitted, to a fictitious data set for three ages, three periods, and five cohorts. For each of the nine age/period combinations, we simulated data for 100 people.[1] Table 1 contains the mean of the dependent variable for each combination. We used linear regression, the results of which are shown in Fig. 1.

Table 1   Mean values of fictitious data for three ages and three periods.

Fig. 1 IE estimates with effect coding for the fictitious data in Table 1, with last and first categories of age, period, and cohort omitted.

In Fig. 1, the age and period trends of both IE solutions differ when moving from year 2 to year 3, with IE (last) showing an increase in the predicted Y value for age and a decrease for period, whereas opposite trends hold for IE (first). For cohort, moving from year 1 to year 2 results in an increase or a decrease of the prediction, depending on the omitted categories chosen. Notice that for the midpoints of age, period, and cohort, both IE solutions yield the same predictions. This is congruent with Eq. (6), which shows that the estimates for the middle categories are identifiable without additional restrictions.

Because the type of parameterization in the default IE is effect coding with the unweighted mean as reference point, the sensitivity will be less the more categories of age, period, and cohort are available. This is illustrated in Fig. 2 where we use the same dataset on U.S. female mortality as was used by Yang et al. (2004, 2008). These data are from the Berkeley Human Mortality Database and include 19 age categories, 8 periods, and 26 birth cohorts.

---

[1] For the APC models that we discuss in this paper, it is not required to have the same number of units in each combination of age and period.

Fig. 2 IE estimates with effect coding and last categories omitted and IE estimates with lowest and highest slopes of age, period, and cohort (Berkeley Human Mortality Database).

We took all possible 19 x 8 x 26=3952 triplets of omitted categories and for each triplet we calculated the corresponding set of effect coded IE estimates, using the Poisson regression model also employed by Yang et al. (2004, 2008). In Fig. 2 we show the triplets with the lowest and the highest linear trends in age, period, and cohort, as well as the default IE (last categories omitted).[2] To find the triplets with lowest and highest linear trends, we implemented the principal component regression method in the software package R. As Fig. 2 demonstrates, it is of close to no importance which triplet of categories is omitted, the general tendency being more or less the same. In sum, the IE estimates with effect coding do depend on the choice which categories are omitted, but this undesired sensitivity becomes less strong the more categories of age, cohort, and period are being used. This is important to note because due to data limitations researchers may work with a limited set of periods, while the set of age and cohort categories is a lesser problem under normal circumstances. In any event, it seems advisable to check for sensitivity to the chosen omitted categories when using the IE and effect coding.

Dummy Variable Coding

In the preceding paragraph we have shown that the IE estimates depend on the choice of which categories are omitted. In more general terms, the IE estimates depend on the design matrix that is chosen before estimation takes place. This matrix is defined by the choice of omitted categories and the type of parameterization one wishes to apply. For instance, a researcher may be interested in developments from the first period, youngest age and/or oldest cohort onwards. In that case, a dummy variable parameterization may be appropriate, with the first categories as points of reference instead of the mean. In identified models, results from any parameterization can be transformed into the results from any other parameterization. However, for IE models this is not true: a solution obtained with effect coding is different from a solution based on dummy variable coding. Interestingly, although these solutions yield different estimates, they both have the IE

---

[2] We determined the linear trend in, for example, age by a simple OLS regression on the 19 age estimates. We used the highest and lowest linear trends only to demonstrate the variability in IE estimates when different sets of categories are omitted.

properties presented by Yang et al. (2004): each minimizes its own particular sum of squared estimates and the associated standard errors of these estimates are minimal. Appendix B contains the proof that IE estimates with dummy variable coding as a rule differ from the default IE estimates using effect coding.

We will now demonstrate how, with dummy variable coding, the IE estimates vary with the reference categories chosen, using the U.S. female mortality database. Again, all possible 19 x 8 x 26=3952 triplets of omitted categories were examined and for each triplet the corresponding set of dummy variable coded IE estimates was obtained. Next, we selected the sets with the lowest and the highest linear tendency in age, period, and cohort, which were found for the triplets (19, 5, 14) and (10, 4, 26). These two sets of IE estimates are plotted in Fig. 3 together with the estimates in case the first categories act as points of reference.

Fig. 3 IE estimates with dummy variable coding and first categories omitted and IE estimates with lowest and highest slopes of age, period, and cohort (Berkeley Human Mortality Database).

Figure 3 shows that for age and even more so for period and cohort, the differences between the three dummy coded IE solutions are substantial. For age, the dashed curve predicts less mortality at the end of the life cycle then at the beginning, whereas the other two IE's show the opposite. For period, the dotted IE and the IE with the first categories omitted show a significant decline in mortality over the years, whereas for the dashed IE mortality increases. For cohort, the dotted IE estimates show little variation in mortality over the years, whereas the other two IE's suggest a small and big decline in mortality risk respectively. Comparing figures 2 and 3 shows that with dummy variable coding, there is more variability in the estimates of age, period, and cohort than with effect coding. Further, if a researcher would choose the first age, period, and cohort year as (starting) points of reference, he or she would obtain different results than another researcher using the default IE, in particular with respect to the period (fall instead of rise in mortality) and the cohort trend (less steep decline in mortality).

4 Discussion

In this paper we showed that there is no unique set of IE estimates, but that there exist many, each corresponding to a particular type of parameterization and a particular triplet of omitted categories. Using fictitious and real data, we demonstrated that different IE's can lead to different conclusions about age, period, and cohort trends. With many time points, the IE using effect coding seems relatively robust for the choice of omitted categories. However, with a limited number of time points, which in actual research will most likely occur for period, the effect coded IE can lead to quite different results for different omitted categories. The IE based on dummy coding seems more sensitive to the choice of categories that act as points of reference, even if the number of time points under study is large.

A consequence of our findings is that users of the IE may have to consider which IE best fits their needs. From a mathematical point of view it is difficult, if not impossible, to prefer one particular parameterization and set of omitted categories: each IE has the desirable properties that were in detail described by Yang et al. (2004), be it with respect to a different set of parameters, determined by the chosen parameterization and omitted categories. The default IE presented by Yang et al. (2004) is special in the sense that it minimizes the sum of squared deviations from the mean, i.e., the variance of the estimates. Some researchers may consider this a desirable property: if one is completely agnostic about age, period, and cohort influences, one may prefer a "conservative" estimation method, which favors a small variance. Yet, in our opinion, other parameterizations can be equally valid. For example, other researchers may favor estimates that show the smallest (sum of squared) changes compared to some reference year of age, period, and cohort and therefore choose a dummy variable parameterization. Yet other researchers may prefer estimates which show the smallest (sum of squared) changes compared to the immediately preceding time point, and hence use so called "repeatedly" coded dummy variables.

The point we like to make is that, if someone decides to use the IE, he or she has to be aware of the different possibilities which may lead to different results. For instance, Yang et al. (2004:100) state that the reason for the slow or no increase in mortality for period in the middle panel of our Fig. 2 is not clear, and they hypothesize that this may partly be due to increasing cigarette smoking in females. However, with the first categories as references, there is a slight decrease in mortality over the four decades, as shown in the middle panel of Fig. 3, which seems

equally, if not more, plausible. Because of such possibly divergent conclusions, we agree with Yang et al. (2013) that the IE should never be used routinely, but for another additional reason: even if the three steps procedure recommended by Yang et al. is carefully conducted, the question concerning parameterization and omitted categories remains to be answered *before* applying the IE. In this respect the IE is more similar to constrained generalized linear models (CGLIM) than one may think, since both types of models depend on a constraint that has to be chosen *before* analyzing. In CGLIM, each pair of equaled categories corresponds to a different constraint and thus different estimates. In the IE, each choice of parameterization and/or omitted categories corresponds to a different constraint and, hence, different estimates.

Above we noted that an argument to use the default IE is that researchers, from an agnostic point of view, may prefer estimates for the age, period, and cohort categories with the smallest variance possible. The minimization criterion of the default IE, however, does not involve the estimates of the omitted categories. Instead of minimizing $(\hat{\alpha}_1^L)^2 + (\hat{\beta}_1^L)^2 + (\hat{\gamma}_1^L)^2 + (\hat{\gamma}_2^L)^2 + (\hat{\gamma}_4^L)^2$ in our fictitious example, one could minimize $(\hat{\alpha}_1^L)^2 + (\hat{\alpha}_3^L)^2 + (\hat{\beta}_1^L)^2 + (\hat{\beta}_3^L)^2 + (\hat{\gamma}_1^L)^2 + (\hat{\gamma}_2^L)^2 + (\hat{\gamma}_4^L)^2 + (\hat{\gamma}_5^L)^2$ which is the sum of squares of estimates of *all* parameters that are not identified, including the three omitted categories. It's noteworthy that this criterion leads to estimates that are *independent* of the omitted categories, as opposed to the IE criterion.

To summarize, we showed that the intrinsic estimator has a non-uniqueness property, which raises the question which IE to choose. In case one values minimum variance, the IE with effect coding would be the obvious choice, which however does not provide the smallest variance across all APC categories as the omitted categories are not part of the constraint. In case one values smallest (squared) deviations from some reference year, the dummy variable coded IE would be more appropriate, an application of which we demonstrated in this paper. The findings in this article may have implications for past research in which the IE has been applied, and for future research considering the usage of the IE.

We end by noting that in this paper, we did not go into the issue of the IE being biased with respect to the "true" data generating parameters. This has been discussed thoroughly in other contributions (see the December 2013 issue of this journal). Our findings demonstrate this bias in the sense that different IE's lead to different results.

References

Fienberg, S.E. (2013). Cohort Analysis' Unholy Quest: A Discussion. *Demography*, 50 (6), 1981-1984.

Hardy, M.A. (1993). Regression with Dummy Variables. Newbury Park, CA: Sage.

Held, L. & Riebler, A. (2013). Comment on "Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort (APC) Problem". *Demography*, 50 (6), 1977-1979.

Luo, L. (2013). Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem. *Demography*, 50 (6), 1945-1967.

O'Brien, R.M. (2013). Comment of Liying Luo's Article, "Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem". *Demography*, 50 (6), 1973-1975.

Yang, Y.C., Fu, W.J., & Land, K.C. (2004). A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models. *Sociological Methodology*, 34 (1), 75-110.

Yang, Y.C., Schulhofer-Wohl, S., Fu, W.J., & Land, K.C. (2008). The Intrinsic Estimator for Age-Period-Cohort Analysis: What it is and how to use it. *American Journal of Sociology*, 113 (6), 1697-1736.

Yang, Y.C. & Land, K.C. (2013). Misunderstandings, Mischaracterizations, and the Problematic Choice of a Specific Instance in Which the IE Should Never Be Applied. *Demography*, 50 (6), 1969-1971.

Appendix A: proof of the non-uniqueness of the IE in case of effect coding.

To arrive at Eq. (2), we used effect coding with the last categories of age, period, and cohort omitted from the equation. In this section, we will show that the IE yields different estimates when the first categories are omitted. Equation (2) then changes into:

$$Y = \beta_0^F + \alpha_2^F A_2^F + \alpha_3^F A_3^F + \beta_2^F P_2^F + \beta_3^F P_3^F + \\ \gamma_2^F C_2^F + \gamma_3^F C_3^F + \gamma_4^F C_4^F + \gamma_5^F C_5^F + e \; , \tag{2a}$$

where superscript F denotes that the first categories of age, period, and cohort are omitted. The independent variables in (2a) differ from those in (2), since they now take value -1 for cases in the first category of age, period, or cohort. Again, the variable for the fourth cohort can be expressed in terms of other variables in (2a):

$$C_4^F = -A_3^F + P_3^F + C_2^F - 2C_5^F \; . \tag{3a}$$

Substituting into (2a) the expression for C_4^F given in (3a) yields:

$$Y = \beta_0^F + \alpha_2^F A_2^F + (\alpha_3^F - \gamma_4^F)A_3^F + \beta_2^F P_2^F + (\beta_3^F + \gamma_4^F)P_3^F + \\ (\gamma_2^F + \gamma_4^F)C_2^F + \gamma_3^F C_3^F + (\gamma_5^F - 2\gamma_4^F)C_5^F + e \; . \tag{4a}$$

The above equation can be represented more compactly as:

$$Y = \boldsymbol{\beta}_0^F + \boldsymbol{\alpha}_2^F A_2^F + \boldsymbol{\alpha}_3^F A_3^F + \boldsymbol{\beta}_2^F P_2^F + \boldsymbol{\beta}_3^F P_3^F + \boldsymbol{\gamma}_2^F C_2^F + \boldsymbol{\gamma}_3^F C_3^F + \boldsymbol{\gamma}_5^F C_5^F + e \; . \tag{5a}$$

The parameters in (5a) are identified due to the fact that C_4^F is not part of the equation. Following the same line of reasoning as in the main text for the last categories omitted, we now

have to minimize the sum of squares $(\hat{\alpha}_3^F)^2 + (\hat{\beta}_3^F)^2 + (\hat{\gamma}_2^F)^2 + (\hat{\gamma}_4^F)^2 + (\hat{\gamma}_5^F)^2$ which finally leads

to the following IE estimator for γ_4^F:

$$\hat{\gamma}_4^F = (-\boldsymbol{\alpha}_3^F + \boldsymbol{\beta}_3^F + \boldsymbol{\gamma}_2^F - 2\boldsymbol{\gamma}_5^F)\,/\,8\,. \tag{8a}$$

Recall that, with the last categories omitted, we found the estimator given in Eq. (8):

$$\hat{\gamma}_4^L = (\boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L)\,/\,8\,. $$

Apparently, the estimate $\hat{\gamma}_4^F$ depends on the value of $-\boldsymbol{\alpha}_3^F + \boldsymbol{\beta}_3^F + \boldsymbol{\gamma}_2^F - 2\boldsymbol{\gamma}_5^F$ whereas the estimate

$\hat{\gamma}_4^L$ depends on the value of $\boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L$. In general, with actual data

$-\boldsymbol{\alpha}_3^F + \boldsymbol{\beta}_3^F + \boldsymbol{\gamma}_2^F - 2\boldsymbol{\gamma}_5^F$ will not be equal to $\boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L$. To see this, note that the bold

face parameters in (8) and (8a) represent deviations from the means $\boldsymbol{\beta}_0^L$ and $\boldsymbol{\beta}_0^F$ in equations (5)

and (5a), respectively. Both (5) and (5a) are estimable due to the constraint that the deviation for

the fourth cohort is equal to zero, i.e., $\gamma_4^L = 0$ and $\gamma_4^F = 0$. As a consequence of using the same

constraint in (5) and (5a) all $\boldsymbol{\alpha}$ estimates with the same subscript are equal in both equations; the

same holds for all $\boldsymbol{\beta}$ estimates with the same subscript and all $\boldsymbol{\gamma}$ estimates with the same

subscript. For example, $\boldsymbol{\alpha}_2^L = \boldsymbol{\alpha}_2^F$. Also, $\boldsymbol{\alpha}_3^L$, to be derived as $-\boldsymbol{\alpha}_1^L - \boldsymbol{\alpha}_2^L$, is equal to $\boldsymbol{\alpha}_3^F$ as

estimated with Eq. (5a). The estimates of the means $\boldsymbol{\beta}_0^L$ and $\boldsymbol{\beta}_0^F$ are equal as well. To prove that

$-\boldsymbol{\alpha}_3^F + \boldsymbol{\beta}_3^F + \boldsymbol{\gamma}_2^F - 2\boldsymbol{\gamma}_5^F \neq \boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L$ therefore boils down to proving that

$-\boldsymbol{\alpha}_3^L + \boldsymbol{\beta}_3^L + \boldsymbol{\gamma}_2^L - 2\boldsymbol{\gamma}_5^L \neq \boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L$. In this last inequality the expression to the left of

the $\neq$ sign contains different deviations from the mean $\boldsymbol{\beta}_0^L$ than the expression to the right. As a

consequence, the values of $\hat{\gamma}_4^F$ and $\hat{\gamma}_4^L$ will usually differ for actual data. Also, the parameter

estimates that depend on the values of $\hat{\gamma}_4^F$ and $\hat{\gamma}_4^L$ will in general be different, for example $\hat{\alpha}_1^L = \boldsymbol{\alpha}_1^L - \hat{\gamma}_4^L$ whereas $\hat{\alpha}_1^F = -\hat{\alpha}_2^F - \hat{\alpha}_3^F = -\boldsymbol{\alpha}_2^F - (\boldsymbol{\alpha}_3^F - \hat{\gamma}_4^F) = \boldsymbol{\alpha}_1^F + \hat{\gamma}_4^F = \boldsymbol{\alpha}_1^L + \hat{\gamma}_4^F$.[3]

Appendix B

In this appendix we show why in general, with dummy variable coding, the IE estimates differ from the "default" IE estimates presented by Yang et al. (2004). Instead of standard 0 and 1 coded dummy variables, we subtract 1/k from these dummies, with k denoting the number of categories, i.e., 3 for age and period, and 5 for cohort. Subtracting the constant 1/k from the original 0 and 1 coded dummies, does not change the interpretation of their regression coefficients, i.e., the deviation from the omitted category. If we omit the last age category, the dummies variables A_1^D and A_2^D for the first two ages have the coding scheme shown in Table 2.

Table 2 Dummy variable coding for age after subtracting 1/3

Note that in Table 2, for each of the dummy variables the sum over the three ages (column sum) equals zero, just as with effect coding. As a result, the intercept in Eq. (2b) below, is equal to the intercept in Eq. (2), both representing the unweighted mean of the three predicted values for age 1, 2, and 3 (controlling for period and cohort). With the last categories omitted, we obtain the following equation to be estimated:

$$
\begin{aligned}
Y = {} & \beta_0^D + \alpha_1^D A_1^D + \alpha_2^D A_2^D + \beta_1^D P_1^D + \beta_2^D P_2^D + \\
& \gamma_1^D C_1^D + \gamma_2^D C_2^D + \gamma_3^D C_3^D + \gamma_4^D C_4^D + e.
\end{aligned}
\tag{2b}
$$

---

[3] The Stata user added routine "apc_ie" only provides deviation contrast IE estimates with the last categories omitted. To obtain estimates with the first categories omitted , one could "mirror" the three APC variables so that the highest age, period, and cohort values becomes lowest.

To keep notation parsimonious, we do not use superscript DL to explicitly indicate that the last dummy variable coded category is the reference. Like equations (2) and (2a), the new Eq. (2b) also suffers from the identification problem. For example, for C_4^D we can write:

$$C_4^D = -2A_1^D - A_2^D + 2P_1^D + P_2^D - 4C_1^D - 3C_2^D - 2C_3^D. \tag{3b}$$

If we substitute in (2b) for C_4^D the expression given in (3b) we obtain:

$$\begin{aligned}
Y = \ & \beta_0^D + (\alpha_1^D - 2\gamma_4^D)A_1^D + (\alpha_2^D - \gamma_4^D)A_2^D + \\
& (\beta_1^D + 2\gamma_4^D)P_1^D + (\beta_2^D + \gamma_4^D)P_2^D + \\
& (\gamma_1^D - 4\gamma_4^D)C_1^D + (\gamma_2^D - 3\gamma_4^D)C_2^D + (\gamma_3^D - 2\gamma_4^D)C_3^D + e.
\end{aligned} \tag{4b}$$

Recall that for effect coding with the last category omitted, we used the constraint that $\gamma_4^L = 0$ in Eq. (2) leading to the estimable Eq. (5). This assumption implies that the deviation of cohort 4 from the mean is zero. Since the deviation of cohort 5 from the mean equals γ_5^L, the assumption $\gamma_4^L = 0$ implies that the deviation of cohort 4 from cohort 5 equals $-\gamma_5^L$ (the value of $-\gamma_5^L$ can be derived from the estimates of Eq. (5): $-\gamma_5^L = \gamma_1^L + \gamma_2^L + \gamma_3^L + \gamma_4^L = \gamma_1^L + \gamma_2^L + \gamma_3^L$). In terms of the regression coefficients of Eq. (2b), the above implies: $\gamma_4^D = -\gamma_5^L$. If, in Eq. (2b), we plug in $-\gamma_5^L$ for γ_4^D, substitute the expression for C_4^D given in (3b), and rearrange terms we obtain the estimable equation:[4]

$$\begin{aligned}
Y = \ & \boldsymbol{\beta}_0^D + (\boldsymbol{\alpha}_1^D + 2\boldsymbol{\gamma}_5^L)A_1^D + (\boldsymbol{\alpha}_2^D + \boldsymbol{\gamma}_5^L)A_2^D + \\
& (\boldsymbol{\beta}_1^D - 2\boldsymbol{\gamma}_5^L)P_1^D + (\boldsymbol{\beta}_2^D - \boldsymbol{\gamma}_5^L)P_2^D + \\
& (\boldsymbol{\gamma}_1^D + 4\boldsymbol{\gamma}_5^L)C_1^D + (\boldsymbol{\gamma}_2^D + 3\boldsymbol{\gamma}_5^L)C_2^D + (\boldsymbol{\gamma}_3^D + 2\boldsymbol{\gamma}_5^L)C_3^D + e.
\end{aligned} \tag{5b}$$

---

[4] To estimate the parameters in Eq. (5b) with superscript D, one could first estimate:
$$Y = \boldsymbol{\beta}_0^D + \boldsymbol{\alpha}_1^* A_1^D + \boldsymbol{\alpha}_2^* A_2^D + \boldsymbol{\beta}_1^* P_1^D + \boldsymbol{\beta}_2^* P_2^D + \boldsymbol{\gamma}_1^* C_1^D + \boldsymbol{\gamma}_2^* C_2^D + \boldsymbol{\gamma}_3^* C_3^D + e$$
and next, to obtain for example the value of $\boldsymbol{\gamma}_3^D$ calculate $\boldsymbol{\gamma}_3^D = \boldsymbol{\gamma}_3^* + 2\boldsymbol{\gamma}_5^L$.

Due to the fact that both Eq. (5) and (5b) are based on the same constraint with respect to cohort 4, the effect coding parameters in (5) can be translated into deviations from the reference categories resulting from estimating (5b). This is relevant when we compare the effect coded IE with the dummy coded IE later in this section. The parameters in (4b) and (5b) are related as follows:

$$
\begin{aligned}
&\beta_0^D = \boldsymbol{\beta}_0^D, \\
&a_1^D = \boldsymbol{\alpha}_1^D + 2\boldsymbol{\gamma}_5^L + 2\gamma_4^D, \ \alpha_2^D = \boldsymbol{\alpha}_2^D + \boldsymbol{\gamma}_5^L + \gamma_4^D, \\
&\beta_1^D = \boldsymbol{\beta}_1^D - 2\boldsymbol{\gamma}_5^L - 2\gamma_4^D, \ \beta_2^D = \boldsymbol{\beta}_2^D - \boldsymbol{\gamma}_5^L - \gamma_4^D, \\
&\gamma_1^D = \boldsymbol{\gamma}_1^D + 4\boldsymbol{\gamma}_5^L + 4\gamma_4^D, \ \gamma_2^D = \boldsymbol{\gamma}_2^D + 3\boldsymbol{\gamma}_5^L + 3\gamma_4^D, \ \gamma_3^D = \boldsymbol{\gamma}_3^D + 2\boldsymbol{\gamma}_5^L + 2\gamma_4^D.
\end{aligned}
$$

Having an estimate for the parameter of the fourth cohort leads to estimates for the remaining parameters (except for β_0^D). The IE now employs the criterion of minimizing the following sum of squares:

$$
(\hat{\alpha}_1^D)^2 + (\hat{\alpha}_2^D)^2 + (\hat{\beta}_1^D)^2 + (\hat{\beta}_2^D)^2 + (\hat{\gamma}_1^D)^2 + (\hat{\gamma}_2^D)^2 + (\hat{\gamma}_3^D)^2 + (\hat{\gamma}_4^D)^2.
$$

This is a completely different criterion than the one used in the effect coded IE, proposed by Yang et al. (2004). Not only are more parameters (eight instead of five) involved in the sum of squares to be minimized, but the parameters also have a different meaning, i.e., distances to the reference category instead of distances to the mean. For the above sum of squares we can write:

$$
\begin{aligned}
&(\boldsymbol{\alpha}_1^D + 2\boldsymbol{\gamma}_5^L + 2\hat{\gamma}_4^D)^2 + (\boldsymbol{\alpha}_2^D + \boldsymbol{\gamma}_5^L + \hat{\gamma}_4^D)^2 + \\
&(\boldsymbol{\beta}_1^D - 2\boldsymbol{\gamma}_5^L - 2\hat{\gamma}_4^D)^2 + (\boldsymbol{\beta}_2^D - \boldsymbol{\gamma}_5^L - \hat{\gamma}_4^D)^2 + \\
&(\boldsymbol{\gamma}_1^D + 4\boldsymbol{\gamma}_5^L + 4\hat{\gamma}_4^D)^2 + (\boldsymbol{\gamma}_2^D + 3\boldsymbol{\gamma}_5^L + 3\hat{\gamma}_4^D)^2 + (\boldsymbol{\gamma}_3^D + 2\boldsymbol{\gamma}_5^L + 2\hat{\gamma}_4^D)^2 + (\hat{\gamma}_4^D)^2.
\end{aligned}
$$

Taking the first order derivative of this sum of squares with respect to $\hat{\gamma}_4^D$ and setting it to zero, finally leads to the following IE estimate for γ_4^D:

$$\hat{\gamma}_4^D = (-4\boldsymbol{\alpha}_1^D - 2\boldsymbol{\alpha}_2^D + 4\boldsymbol{\beta}_1^D + 2\boldsymbol{\beta}_2^D - 8\boldsymbol{\gamma}_1^D - 6\boldsymbol{\gamma}_2^D - 4\boldsymbol{\gamma}_3^D - 78\boldsymbol{\gamma}_5^L) \,/\, 80\,. \tag{8b}$$

Since the bold face parameters in this expression for $\hat{\gamma}_4^D$ are compatible with the ones in the expression given earlier for $\hat{\gamma}_4^L$, it's possible to formulate $\hat{\gamma}_4^D$ in terms of the effect coding parameters of Eq. (5). For example coefficient $\boldsymbol{\alpha}_1^D$, which is the deviation of age 1 from reference category age 3, can be written as the difference of the two corresponding effect coding parameters, namely $\boldsymbol{\alpha}_1^L - \boldsymbol{\alpha}_3^L$.[5] Building on this relation between the coefficients of effect and dummy coding, we can write the above estimate of $\hat{\gamma}_4^D$ in terms of the effect coding parameters of Eq. (5):

$$\begin{aligned}\hat{\gamma}_4^D = (&-4(\boldsymbol{\alpha}_1^L - \boldsymbol{\alpha}_3^L) - 2(\boldsymbol{\alpha}_2^L - \boldsymbol{\alpha}_3^L) + 4(\boldsymbol{\beta}_1^L - \boldsymbol{\beta}_3^L) + 2(\boldsymbol{\beta}_2^L - \boldsymbol{\beta}_3^L) \\ &-8(\boldsymbol{\gamma}_1^L - \boldsymbol{\gamma}_5^L) - 6(\boldsymbol{\gamma}_2^L - \boldsymbol{\gamma}_5^L) - 4(\boldsymbol{\gamma}_3^L - \boldsymbol{\gamma}_5^L) - 78\boldsymbol{\gamma}_5^L) \,/\, 80\end{aligned} \quad .$$

Using the fact that $\boldsymbol{\alpha}_3^L = -\boldsymbol{\alpha}_1^L - \boldsymbol{\alpha}_2^L$, $\boldsymbol{\beta}_3^L = -\boldsymbol{\beta}_1^L - \boldsymbol{\beta}_2^L$ and $\boldsymbol{\gamma}_5^L = -\boldsymbol{\gamma}_1^L - \boldsymbol{\gamma}_2^L - \boldsymbol{\gamma}_3^L$ we finally get:

$$\hat{\gamma}_4^D = (-\boldsymbol{\alpha}_1^L - 0.8\boldsymbol{\alpha}_2^L + \boldsymbol{\beta}_1^L + 0.8\boldsymbol{\beta}_2^L + 5.2\boldsymbol{\gamma}_1^L + 5.4\boldsymbol{\gamma}_2^L + 5.6\boldsymbol{\gamma}_3^L) \,/\, 8\,. \tag{8b*}$$

If we compare this IE estimator of $\hat{\gamma}_4^D$ with $\hat{\gamma}_4^L = (\boldsymbol{\alpha}_1^L - \boldsymbol{\beta}_1^L + 2\boldsymbol{\gamma}_1^L + \boldsymbol{\gamma}_2^L)/8$ of Eq. (8), it is obvious that the IE estimate for the fourth cohort effect will usually differ between the effect coded IE as proposed by Yang et al. (2004) and the dummy variable coded IE (with the last categories as the omitted ones for both codings). The same holds for the IE estimates of the remaining categories of the three APC variables. Similar to the IE with effect coding, each triplet

---

[5] For any two ages i and j, the predicted values (controlling for period and cohort) based on Eq. 5 are $\boldsymbol{\beta}_0^L + \boldsymbol{\alpha}_i^L$ and $\boldsymbol{\beta}_0^L + \boldsymbol{\alpha}_j^L$. The difference between these predictions equals $\boldsymbol{\alpha}_i^L - \boldsymbol{\alpha}_j^L$, which represents the deviation of age i from age j and hence is equal to $\boldsymbol{\alpha}_i^D$.

of reference categories of age, period, and cohort, leads to different estimates when dummy variable coding is used, which we do not elaborate here.

Tables

Table 1

Period	Age	Mean	N
1	1	-5.89	100
	2	-12.54	100
	3	-10.34	100
2	1	12.31	100
	2	-8.73	100
	3	-0.79	100
3	1	19.66	100
	2	1.25	100
	3	-7.25	100

Table 2

age	A_1^D	A_2^D
1	$+\frac{2}{3}$	$-\frac{1}{3}$
2	$-\frac{1}{3}$	$+\frac{2}{3}$
3	$-\frac{1}{3}$	$-\frac{1}{3}$

Figures

Fig 1

Fig 2

Fig 3