Online Appendix to "The Child Penalty and an Age-Old Problem"

Neil Thakral
Brown University

Linh T. Tô
Boston University

January 2026*

Abstract

This document contains appendix material for the paper "The Child Penalty and an Age-Old Problem."

*Thakral: Department of Economics, Brown University, Box B, Providence, RI 02912 (email: neil_thakral@brown.edu). Tô: Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215 (email: linhto@bu.edu).

Contents

A Proofs **1**
A.1 Proof of Theorem 1 1
A.2 Proof of Lemma 1 3
A.3 Derivation of bias in age-period panel example 6
A.3.1 Setup 6
A.3.2 Computing the bias 8
A.4 Derivation of bias in cohort-period panel example 9
A.4.1 Setup 9
A.4.2 Computing the bias 11
A.5 Proof of Theorem 2 12
A.6 Proof of cohort FE estimator unbiasedness when DGP has individual FE 14

B Implementing the APC Model with One Linear Restriction **15**

References **17**

C Appendix Figures and Tables **18**

List of Appendix Figures

A1 Simulations: Consecutive cohorts, with a pure control group 19
A2 Simulations: Evenly-spaced cohorts, with a pure control group 20
A3 Simulations: Adding individual fixed effects 21

List of Appendix Tables

A Proofs

A.1 Proof of Theorem 1

Proof. Let $I = \{0, 1, \dots, A-1\}$ and $J = \{0, 1, \dots, P-1\}$, and write each cell $(i, j) \in I \times J$ in terms of age $a = i$, event-time $p = j$, and cohort $c = j - i$. Suppose the distinct observed cohort values are $c_0 < c_1 < \cdots < c_{C-1}$, and let $d_r = c_r - c_{r-1}$

for $r = 1, \ldots, C-1$, with $D = \gcd(d_1, \ldots, d_{C-1})$. Then every observed cell satisfies $j - i \equiv c_0 \pmod{D}$, or equivalently $j \equiv i + c_0 \pmod{D}$.

We work with the saturated model

$$y_{i,j} = \kappa + \alpha_i + \beta_j + \gamma_r + \varepsilon_{i,j},$$

where r is the index with $c_r = j - i$. We use the term *invariance* to refer to a collection $(\Delta\kappa, \{f_i\}, \{g_j\}, \{h_r\})$ satisfying

$$\Delta\kappa + f_i + g_j + h_r = 0$$

whenever $c_r = j - i$. Each independent invariance corresponds to one restriction needed for identification.

First note that if we increase every α_i by $+1$ and decrease κ by 1, then $\kappa + \alpha_i$ is unchanged on all cells; similarly, increasing every β_j by $+1$ and decreasing κ by 1 leaves $\kappa + \beta_j$ invariant, and increasing every γ_r by $+1$ together with $\kappa \mapsto \kappa - 1$ leaves $\kappa + \gamma_r$ unchanged. These three shifts are clearly linearly independent.

Next, because $c_r = j - i$, adding i to each α_i, subtracting j from each β_j, and adding c_r to each γ_r leaves $\alpha_i + \beta_j + \gamma_r$ unchanged on every cell; this is the classical APC linear-trend invariance described in Lemma 1.

Finally, since each observed (i, j) satisfies $i \equiv t \pmod{D}$ and $j \equiv t + c_0 \pmod{D}$ for a unique $t \in \{0, \ldots, D-1\}$, we may for each t increase all α_i with $i \equiv t \pmod{D}$ by $+1$ and decrease all β_j with $j \equiv t + c_0 \pmod{D}$ by 1, leaving $\alpha_i + \beta_j$ unchanged exactly on the observed cells. This construction yields D such invariances, but their sum is zero, so they contribute $D - 1$ independent invariances. We have thus exhibited $3 + 1 + (D - 1) = D + 3$ independent invariances in total.

To see there are no others, suppose $(\Delta\kappa, \{f_i\}, \{g_j\}, \{h_r\})$ is *any* invariance, so that

$$\Delta\kappa + f_i + g_j + h_r = 0$$

whenever $c_r = j - i$. Fix a class $t \in \{0, \ldots, D-1\}$; by construction every (i, j) with $i \equiv t \pmod{D}$ and $j \equiv t + c_0 \pmod{D}$ is observed, and so is $(i + D, j + D)$. Writing the invariance equation at (i, j) and at $(i + D, j + D)$, we have

$$\Delta\kappa + f_i + g_j + h_r = 0$$

and

$$\Delta\kappa + f_{i+D} + g_{j+D} + h_r = 0,$$

where in both cases r satisfies $c_r = j - i$. Subtracting yields

$$f_{i+D} - f_i = -(g_{j+D} - g_j).$$

Because the set of observed (i, j) in this class is connected by successive shifts of D, the difference $f_{i+D} - f_i$ must be the same constant for every (i, j) in class t. Call that constant λD. Then $g_{j+D} - g_j = -\lambda D$. It follows by an argument analogous to that in Lemma A.1 that

$$f_i = \lambda i + \mu_t$$
$$g_j = -\lambda j + \nu_t$$

for all (i, j) in class t, for some μ_t, ν_t. Substituting back into $\Delta\kappa + f_i + g_j + h_r = 0$ shows

$$h_r = -\Delta\kappa - (\mu_t + \nu_t) + \lambda c_r$$

whenever $c_r \equiv c_0 \pmod{D}$. Thus, for each class t, the functions f, g, h are exactly of the form generated by the invariances exhibited earlier. Since this holds for every t, no other independent invariance can exist. Hence the space of all invariances is spanned by the $D + 3$ we constructed, completing the proof.

The space of invariances thus has dimension $D + 3$, and the model requires $D + 3$ restrictions for identification. □

A.2 Proof of Lemma 1

Proof. Consider the APC specification for estimating child penalties:

$$Y_{it} = \alpha_{a(i,t)} + \beta_{p(i,t)} + \gamma_{c(i)} + \delta_t + \kappa + \varepsilon_{it}, \tag{1}$$

where $a(i, t)$ is the age of individual i at time t, $p(i, t) = t - E_i$ is event time relative to childbirth, and $c(i) = a(i, E_i)$ is the childbirth cohort (or age at first childbirth). These variables are mechanically related by the relationship $c(i) = a(i, t) - p(i, t)$ for all i and t.

While the model includes calendar-year fixed effects δ_t, note that these do not pose any identification problem alongside the age, period, and cohort effects. Calendar year t is not mechanically linked to the other variables as long as E_i (the time of childbirth) varies across individuals. In this case, δ_t is not collinear with any combination of $\alpha_{a(i,t)}$, $\beta_{p(i,t)}$, or $\gamma_{c(i)}$. The identification issue arises solely from the perfect linear dependence among age, period, and cohort. Therefore, we omit δ_t in what follows, without loss of generality.

To understand the source of non-identification, consider transforming the fixed effects to

$$\begin{aligned}\alpha'_a &= \alpha_a + f(a)\\ \beta'_p &= \beta_p + g(p)\\ \gamma'_c &= \gamma_c + h(c)\end{aligned}$$

for arbitrary functions f, g, and h. The transformation leaves the model unchanged as long as

$$f(a(i,t)) + g(p(i,t)) + h(c(i)) = 0$$

for all i, t. Since $c(i) = a(i,t) - p(i,t)$, we can rewrite this condition as

$$f(a) + g(p) + h(a-p) = 0 \tag{2}$$

for all a and p.

To solve this functional equation, we start by setting $p = 0$. This implies $c = a$ and hence $f(a) + g(0) + h(a) = 0$, or equivalently,

$$h(a) = -f(a) - g(0).$$

Substituting back into Equation (2) gives $f(a) + g(p) - f(a-p) - g(0) = 0$, or equivalently,

$$f(a) - f(a-p) = g(0) - g(p).$$

By Lemma A.1 below, this equation holds only if f is affine, i.e., there exist constants λ and μ such that

$$f(a) = \lambda a + \mu.$$

This implies

$$f(a) - f(a-p) = \lambda p$$

and hence

$$g(p) = g(0) - \lambda p.$$

Absorbing the constant $g(0)$ into the intercept term κ, without loss of generality, we can write

$$g(p) = -\lambda p.$$

Substituting back into the expression for $h(a)$ gives

$$h(a) = -(\lambda a + \mu) - g(0) = \lambda a - \mu - g(0).$$

Again absorbing the constant into κ, we obtain

$$h(c) = \lambda c.$$

Putting this all together, the model is invariant to the transformation:

$$\begin{aligned}
\alpha_a &= \alpha_a + \lambda a, \\
\beta_p &= \beta_p - \lambda p, \\
\gamma_c &= \gamma_c + \lambda c.
\end{aligned}$$

This shows that the parameters of the APC model are identified only up to an arbitrary linear trend. Thus, an additional constraint (e.g., assuming no net drift in cohort effects or period effects) is necessary for estimation. □

Lemma A.1. *Suppose functions $f\colon \mathbb{Z} \to \mathbb{R}$ and $g\colon \mathbb{Z} \to \mathbb{R}$ satisfy*

$$f(a) - f(a-p) = g(0) - g(p)$$

for all integers a, p. Then f must be affine: there exist constants $\lambda, \mu \in \mathbb{R}$ such that

$$f(a) = \lambda a + \mu$$

for all integers a.

Proof. Write the right-hand side as a function $\Delta(p) = g(0) - g(p)$ depending only on

p. In particular, setting $p = 1$ shows that the first difference is given by

$$f(a) - f(a-1) = \Delta(1),$$

which is independent of a. Define this constant as $\lambda = \Delta(1)$. It follows by a straightforward induction argument that

$$f(a) - f(a-k) = k\lambda$$

for every integer $k \geq 0$. Taking $a = k$, we obtain

$$f(k) = f(0) + k\lambda.$$

Thus, defining $\mu = f(0)$, we conclude $f(a) = \lambda a + \mu$ for all integers a. □

A.3 Derivation of bias in age-period panel example

A.3.1 Setup

Let the true data-generating process be:

$$y = X\theta + Z\gamma + \varepsilon,$$

where:

- X is a 9×5 matrix of regressors consisting of age and period fixed effects.
- $\theta = [\alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1]^\top$ is the vector of age and period coefficients.
- Z is a 9×5 matrix of cohort dummies, and $\gamma = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5]^\top$ is the vector of true cohort effects.

We assume that the model is estimated without cohort effects, so the OLS estimator becomes:

$$\hat{\theta} = (X^\top X)^{-1} X^\top y = \theta + (X^\top X)^{-1} X^\top Z\gamma + (X^\top X)^{-1} X^\top \varepsilon,$$

implying that the bias from omitting $Z\gamma$ is:

$$\hat{\theta} - \theta = (X^\top X)^{-1} X^\top Z\gamma.$$

The panel consists of $A = 3$ ages $a \in \{1, 2, 3\}$ and $P = 3$ periods $p \in \{-1, 0, 1\}$, for a total of 9 observations. The implied cohort is $c = a - p \in \{0, 1, 2, 3, 4\}$.

The observations are:

Age	Period	Cohort
1	-1	2
1	0	1
1	1	0
2	-1	3
2	0	2
2	1	1
3	-1	4
3	0	3
3	1	2

Since we normalize $\beta_{-1} = 0$ and omit the constant, the matrix X of age and period effects is:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The cohort dummy matrix Z is:

$$Z = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

A.3.2 Computing the bias

We first compute $X^\top X$. Using matrix identities, this matrix simplifies to:

$$X^\top X = \begin{bmatrix} 3I_3 & J_{3\times 2} \\ J_{2\times 3} & 3I_2 \end{bmatrix},$$

where I_k is the $k \times k$ identity matrix and $J_{m\times n}$ is an $m \times n$ matrix of ones.

Using the matrix inversion formula for block matrices, the inverse is:

$$(X^\top X)^{-1} = \frac{1}{9}\begin{bmatrix} 3I_3 + 2J_{3\times 3} & -3J_{3\times 2} \\ -3J_{2\times 3} & 3I_2 + 3J_2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 5 & 2 & 2 & -3 & -3 \\ 2 & 5 & 2 & -3 & -3 \\ 2 & 2 & 5 & -3 & -3 \\ -3 & -3 & -3 & 6 & 3 \\ -3 & -3 & -3 & 3 & 6 \end{bmatrix}.$$

Next, we compute $X^\top Z$:

$$X^\top Z = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

Finally, the bias is:

$$(X^\top X)^{-1} X^\top Z = \frac{1}{9} \begin{bmatrix} 2 & 1 & 3 & 1 & 2 \\ -1 & 1 & 3 & 4 & 2 \\ -1 & -2 & 3 & 4 & 5 \\ 0 & 3 & 0 & 0 & -3 \\ 3 & 3 & 0 & -3 & -3 \end{bmatrix}.$$

Multiplying by the vector $\gamma = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5]^\top$, we obtain:

$$\begin{bmatrix} \widehat{\alpha}_1 - \alpha_1 \\ \widehat{\alpha}_2 - \alpha_2 \\ \widehat{\alpha}_3 - \alpha_3 \\ \widehat{\beta}_0 - \beta_0 \\ \widehat{\beta}_1 - \beta_1 \end{bmatrix} = \begin{bmatrix} \frac{2}{9}\gamma_1 + \frac{1}{9}\gamma_2 + \frac{1}{3}\gamma_3 + \frac{1}{9}\gamma_4 + \frac{2}{9}\gamma_5 \\ -\frac{1}{9}\gamma_1 + \frac{1}{9}\gamma_2 + \frac{1}{3}\gamma_3 + \frac{4}{9}\gamma_4 + \frac{2}{9}\gamma_5 \\ -\frac{1}{9}\gamma_1 - \frac{2}{9}\gamma_2 + \frac{1}{3}\gamma_3 + \frac{4}{9}\gamma_4 + \frac{5}{9}\gamma_5 \\ \frac{1}{3}\gamma_2 - \frac{1}{3}\gamma_5 \\ \frac{1}{3}\gamma_1 + \frac{1}{3}\gamma_2 - \frac{1}{3}\gamma_4 - \frac{1}{3}\gamma_5 \end{bmatrix}.$$

A.4 Derivation of bias in cohort-period panel example

We now compute the omitted variable bias when cohort effects are excluded in a cohort-period panel with uniform coverage across cohorts and periods.

A.4.1 Setup

In this design, we observe $C = 3$ cohorts $c \in \{1, 2, 3\}$ at $P = 3$ event times $p \in \{-1, 0, 1\}$, leading to 9 total observations. Age is computed as $a = c+p \in \{0, 1, 2, 3, 4\}$.

We consider the following linear model:

$$y = X\theta + Z\gamma + \varepsilon,$$

where:

- X is a 9×7 matrix with columns for age fixed effects $(\alpha_0, \ldots, \alpha_4)$ and period fixed effects (β_0, β_1), omitting $\beta_{-1} = 0$ for normalization.
- $\theta = [\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_0, \beta_1]^\top$ is the parameter vector.
- Z is a 9×3 matrix of cohort dummies, and $\gamma = [\gamma_1, \gamma_2, \gamma_3]^\top$ are the true cohort effects.

The omitted variable bias is:

$$\hat{\theta} - \theta = (X^\top X)^{-1} X^\top Z\gamma.$$

There are 9 observations from $C = 3$ cohorts ($c \in \{1, 2, 3\}$) and $P = 3$ periods ($p \in \{-1, 0, 1\}$). Age is defined as $a = c + p$, resulting in possible ages from 0 to 4.

Each cohort appears at the same three periods and at three corresponding ages:

Cohort	Period	Age $= c + p$
1	-1	0
1	0	1
1	1	2
2	-1	1
2	0	2
2	1	3
3	-1	2
3	0	3
3	1	4

Compared to the age-period panel, this structure ensures that each cohort is observed over the same number of ages and periods, eliminating edge effects and asymmetries in exposure.

Since we normalize $\beta_{-1} = 0$ and omit the constant, the matrix X of age and period effects is:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The cohort dummy matrix Z is:

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

A.4.2 Computing the bias

The cross-product matrix is:

$$X^\top X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 3 \end{bmatrix}.$$

The inverse of $X^\top X$ is:

$$(X^\top X)^{-1} = \frac{1}{6} \begin{bmatrix} 6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 4 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 3 & 0 & -3 & -3 \\ 0 & 0 & 0 & 0 & 6 & 0 & -1 \\ 0 & -3 & -2 & -3 & 0 & 5 & 2 \\ 0 & 0 & -2 & -3 & -1 & 2 & 4 \end{bmatrix}.$$

Now compute:

$$X^\top Z = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Then:

$$(X^\top X)^{-1} X^\top Z = \frac{1}{6} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \\ -6 & -3 & -3 \\ -1 & -1 & 5 \\ 2 & -1 & 2 \\ 4 & 1 & 0 \end{bmatrix}.$$

Multiplying by $\gamma = [\gamma_1, \gamma_2, \gamma_3]^\top$, we obtain the bias:

$$\begin{bmatrix} \hat{\alpha}_0 - \alpha_0 \\ \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \\ \hat{\alpha}_3 - \alpha_3 \\ \hat{\alpha}_4 - \alpha_4 \\ \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ -\frac{1}{2}\gamma_3 \\ 0 \\ -\gamma_1 - \frac{1}{2}\gamma_2 - \frac{1}{2}\gamma_3 \\ -\frac{1}{6}\gamma_1 - \frac{1}{6}\gamma_2 + \frac{5}{6}\gamma_3 \\ \frac{1}{3}\gamma_1 - \frac{1}{6}\gamma_2 + \frac{1}{3}\gamma_3 \\ \frac{2}{3}\gamma_1 + \frac{1}{6}\gamma_2 \end{bmatrix}.$$

A.5 Proof of Theorem 2

Proof. Write $y_{i,t}$ for the outcome of individual i at time t, let $a(i,t)$ be that individual's age and $p(i,t)$ the calendar year at (i,t). Define the childbirth-cohort $c(i) = a(i,t) - p(i,t)$ and the own-birth-cohort $b(i) = t - a(i,t)$. Consider the model

$$y_{i,t} = \alpha_{a(i,t)} + \beta_{p(i,t)} + \zeta_i + \delta_t + \kappa + \varepsilon_{i,t}.$$

Any alternative parameterization $(\alpha'_a, \beta'_p, \zeta'_i, \delta'_t)$ that leaves every fitted value unchanged must satisfy two overlapping APC-type invariances. First, tilting age, period and childbirth-cohort effects by a common slope λ, i.e.,

$$\begin{aligned}
\alpha'_a &= \alpha_a + \lambda a \\
\beta'_p &= \beta_p - \lambda p \\
\zeta'_i &= \zeta_i + \lambda c(i) \\
\delta'_t &= \delta_t,
\end{aligned}$$

leaves $\alpha_{a(i,t)} + \beta_{p(i,t)} + \zeta_i + \delta_t + \kappa$ invariant. Second, tilting age, year and own-birth-cohort effects by another slope μ, i.e.,

$$\begin{aligned}
\alpha'_a &= \alpha_a + (\lambda + \mu)a \\
\beta'_p &= \beta_p - \lambda p \\
\zeta'_i &= \zeta_i + \lambda c(i) + \mu b(i) \\
\delta'_t &= \delta_t - \mu t,
\end{aligned}$$

again leaves every fitted value unchanged. No other smooth relabeling is possible, so the full set of equivalent solutions is indexed by the two parameters λ, μ.

To avoid perfect multicollinearity with the constant term, we first normalize one category of each fixed-effect series, for example $\alpha_{a_0} = \beta_{p_0} = \gamma_{i_0} = \delta_{t_0} = 0$. Since the slope indeterminacy (λ, μ) remains unaffected, we must impose two additional restrictions; for instance, $\alpha_{a_1} = 0$ and $\beta_{p_1} = 0$ with $a_1 \neq a_0$, $p_1 \neq p_0$.

We now show that if *one* of these slope normalizations is consistent with the data generating process, then that entire series is recovered exactly, regardless of whether the other slope restriction is wrong.

Suppose first that $\alpha_{a_1} = 0$ is consistent with the true data generating process. Then requiring the fitted $\alpha'_{a_1} = 0$ forces $(\lambda + \mu)a_1 = 0$, hence $\mu = -\lambda$. At the same time imposing $\beta'_{p_1} = 0$ yields $\beta_{p_1} - \lambda p_1 = 0$, so $\lambda = \beta_{p_1}/p_1$. Substituting gives $\alpha'_a = \alpha_a$ for all a, while β'_p, δ'_t, and the implicit cohort effects are identified up to a linear trend, tilted by λ. Thus the age effects $\{\alpha_a\}$ are exactly recovered even if the period restriction was incorrect.

Alternatively, if $\beta_{p_1} = 0$ is consistent with the true data generating process, then imposing $\beta'_{p_1} = 0$ gives $\lambda = 0$. This gives $\beta'_p = \beta_p$ for all p, so the period series is

exactly identified, even if the age, year, and implicit cohort effects may be tilted by μ.

An analogous argument applies if the normalization consistent with the DGP is placed on the year or individual fixed effects. □

A.6 Proof of cohort FE estimator unbiasedness when DGP has individual FE

Proposition A.1. *Suppose the true data-generating process satisfies*

$$y_{i,t} = \alpha_{a(i,t)} + \beta_{p(i,t)} + \zeta_i + \delta_t + \varepsilon_{i,t},$$

with $\mathbb{E}\left[\varepsilon_{i,t} \,\middle|\, a(i,t), p(i,t), t, i\right] = 0$*, where* $a(i,t)$ *denotes individual* i*'s age at time* t*,* $p(i,t)$ *denotes* i*'s event time at time* t*, and* $c(i) = a(i,t) - p(i,t)$ *is the birth cohort of individual* i*. Consider estimating the cohort fixed-effects regression*

$$y_{i,t} = \alpha_{a(i,t)} + \beta_{p(i,t)} + \gamma_{c(i)} + \delta_t + u_{i,t}.$$

If

$$\mathbb{E}[\zeta_i \,|\, p(i,t) = p, c(i) = c, t] = \mathbb{E}[\zeta_i \,|\, c(i) = c] \tag{3}$$

for every (c, p, t) *in the support of the data, then OLS delivers unbiased estimates of the identified coefficients in* $(\{\alpha_a\}, \{\beta_p\}, \{\gamma_c\}, \{\delta_t\})$*.*

Proof. Write the cohort fixed-effects regression residual as

$$u_{i,t} = [\zeta_i - \gamma_{c(i)}] + \varepsilon_{i,t}.$$

Equation (3) implies

$$\mathbb{E}\left[\zeta_i - \gamma_{c(i)} \,\middle|\, c(i,t), p(i,t), t\right] = 0.$$

Since $a(i,t) = c(i) + p(i,t)$, the regressors $\alpha_{a(i,t)}$, $\beta_{p(i,t)}$, $\gamma_{c(i)}$, and δ_t all depend only on $(c(i), p(i,t), t)$, so

$$\mathbb{E}\left[\zeta_i - \gamma_{c(i)} \,\middle|\, a(i,t), p(i,t), t\right] = 0.$$

Together with the assumption $\mathbb{E}\left[\varepsilon_{i,t} \,\middle|\, a(i,t), p(i,t), t, i\right] = 0$, this yields

$$\mathbb{E}\left[u_{i,t} \,\middle|\, a(i,t), p(i,t), t\right] = 0.$$

OLS therefore satisfies the standard orthogonality condition for the cohort fixed-effects regression and yields unbiased estimates of the identified coefficients. □

B Implementing the APC Model with One Linear Restriction

Throughout, we assume the presence of a constant term, so that each of the effects—α_i (age), β_j (period), and γ_k (cohort)—is defined relative to a baseline value.

To resolve the identification problem inherent in the APC model—particularly in settings with consecutive cohorts—one additional linear restriction must be imposed (see Theorem 1).

As an example, we may choose to place this restriction on a subset of the age effects,

$$S = \left\{\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_k}\right\}.$$

However, an equivalent restriction could be applied to a subset of the period or cohort effects instead.

For instance, we can require that the linear regression line (in the least-squares sense) through the points

$$\left\{(i_1, \alpha_{i_1}), (i_2, \alpha_{i_2}), \dots, (i_k, \alpha_{i_k})\right\}$$

has slope s. That is:

$$s = \frac{\sum_{j=1}^{k}(i_j - \bar{i})(\alpha_{i_j} - \bar{\alpha})}{\sum_{j=1}^{k}(i_j - \bar{i})^2},$$

where

$$\bar{i} = \frac{1}{k}\sum_{j=1}^{k} i_j, \quad \bar{\alpha} = \frac{1}{k}\sum_{j=1}^{k} \alpha_{i_j}.$$

Rearranging yields the equivalent linear constraint:

$$\sum_{j=1}^{k}(i_j - \bar{i})(\alpha_{i_j} - \bar{\alpha}) = s\sum_{j=1}^{k}(i_j - \bar{i})^2,$$

or, solving for one coefficient (e.g., α_{i_1}):

$$\alpha_{i_1} = \frac{1}{i_1 - \bar{i}} \left(s \sum_{j=1}^{k} (i_j - \bar{i})^2 - \sum_{j=2}^{k} (i_j - \bar{i}) \alpha_{i_j} \right).$$

We now consider a regression model of the form:

$$y = \sum_{j=1}^{k} \alpha_{i_j} \cdot d_{i_j} + (\text{other terms}),$$

and aim to enforce the slope restriction on the subset S by substituting the expression for α_{i_1} directly into the regression.

Substituting gives:

$$= \left(\frac{1}{i_1 - \bar{i}} \left(s \sum_{j=1}^{k} (i_j - \bar{i})^2 - \sum_{j=2}^{k} (i_j - \bar{i}) \alpha_{i_j} \right) \right) \cdot d_{i_1} + \sum_{j=2}^{k} \alpha_{i_j} \cdot d_{i_j}.$$

Rewriting:

$$= s \cdot \frac{\sum_{j=1}^{k} (i_j - \bar{i})^2}{i_1 - \bar{i}} \cdot d_{i_1} + \sum_{j=2}^{k} \alpha_{i_j} \cdot \left(d_{i_j} - \frac{i_j - \bar{i}}{i_1 - \bar{i}} \cdot d_{i_1} \right).$$

Define:

$$C = \frac{\sum_{j=1}^{k} (i_j - \bar{i})^2}{i_1 - \bar{i}}, \quad r_j = \frac{i_j - \bar{i}}{i_1 - \bar{i}},$$

so the transformed regression becomes:

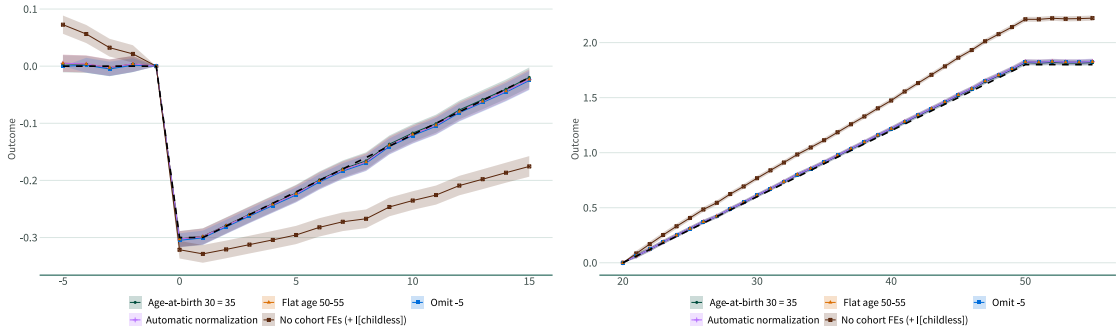
$$y = s \cdot C \cdot d_{i_1} + \sum_{j=2}^{k} \alpha_{i_j} \cdot \left(d_{i_j} - r_j \cdot d_{i_1} \right) + (\text{other terms}).$$

This reparameterized model enforces the desired slope constraint by construction. The transformation only modifies the regressors corresponding to the constrained subset S, and leaves other coefficients—including those for age effects not in S, as well as any period or cohort fixed effects—unaffected.

References

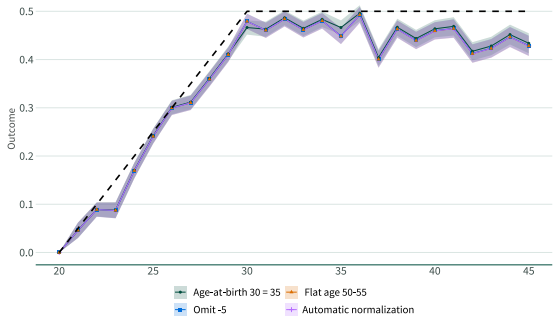
C Appendix Figures and Tables

Figure A1: Simulations: Consecutive cohorts, with a pure control group



(a) Period (event time)

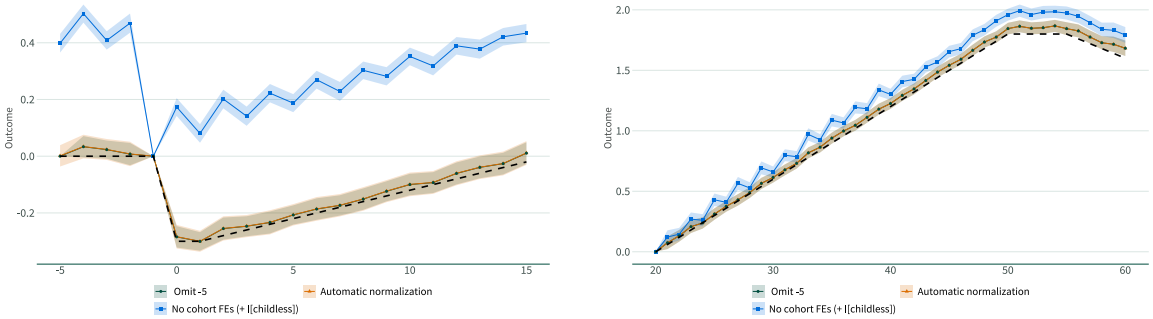
(b) Age (life-cycle) profile



(c) Cohort (age-at-birth) profile

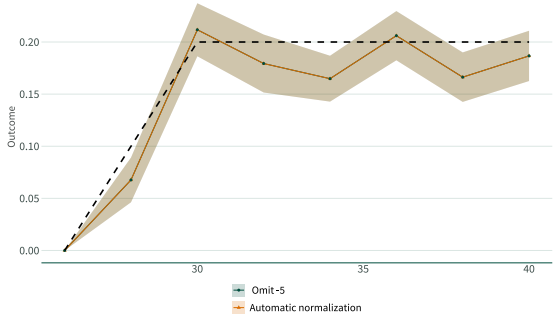
Note: The figures show simulation results in a setting with consecutive cohorts and when childless individuals are included in the sample. (The counterpart of this figure when childless individuals are not included is Figure 1.) True effects are shown in dashed black lines, while estimated effects across estimation strategies are shown in solid lines. The data generating process is described in Section 3.1. The estimation strategies are explained in Section 3.2. 95% confidence intervals are shown in shaded areas. Panel (a) shows the period (event time) profile relative to event time $t = -1$, panel (b) shows the age (life-cycle) profile relative to age 20, and panel (c) shows the cohort (age-at-birth) profile relative to cohort with age-at-first-birth of 20.

Figure A2: Simulations: Evenly-spaced cohorts, with a pure control group



(a) Period (event time)

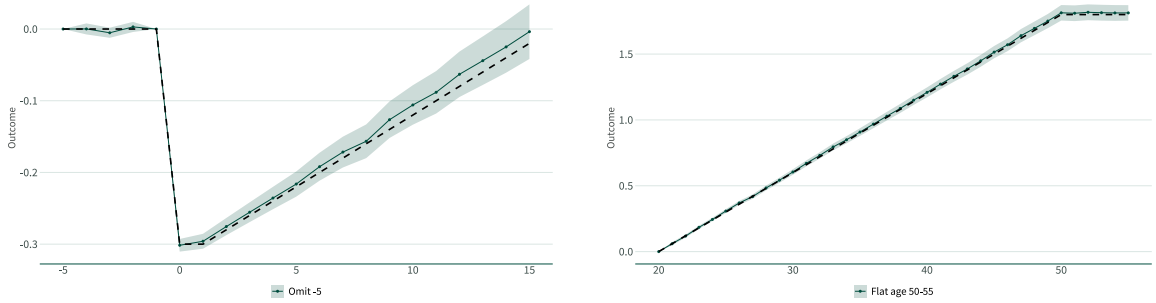
(b) Age (life-cycle) profile



(c) Cohort (age-at-birth) profile

Note: The figures show simulation results in a setting with a cohort every other year and when childless individuals are included in the sample. (The counterpart of this figure when childless individuals are not included is Figure 2.) True effects are shown in dashed black lines, while estimated effects across estimation strategies are shown in solid lines. The data generating process is described in Section 3.1. The estimation strategies are explained in Section 3.2. 95% confidence intervals are shown in shaded areas. Panel (a) shows the period (event time) profile relative to event time $t = -1$, panel (b) shows the age (life-cycle) profile relative to age 20, and panel (c) shows the cohort (age-at-birth) profile relative to cohort with age-at-first-birth of 20.

Figure A3: Simulations: Adding individual fixed effects



(a) Period (event time), when restricting is on period slope

(b) Age (life-cycle) profile, when restricting is on age slope

Note: The figure shows simulation results for the same setting as in Figure 1, but with individual fixed effects included. Panel (a) shows the period (event time) profile relative to event time $t = -1$ when the identifying restriction is put on period, and panel (b) shows the age (life-cycle) profile relative to age 20 when the identifying restriction is on age.