Cross Section Bias: Age, Period and Cohort Effects

James J. Heckman
University of Chicago

Econ 312, Spring 2023

THE UNIVERSITY OF
CHICAGO

$$\ln W_i = \alpha_0 + \alpha_1 a_i + \alpha_2 y$$
$$\uparrow \qquad \uparrow$$
$$\text{age} \quad \text{year}$$

$$\alpha_3 e_i \qquad + \qquad \alpha_4 s_i \qquad + \qquad \alpha_5 c_i \qquad + \qquad u_i$$
$$\uparrow \qquad\qquad \uparrow \qquad\qquad\quad \uparrow$$

experience   schooling   vintage (birth   cohort)

Two Identities

$$e_i = a_i - s_i \qquad \text{``experience''} \qquad (1)$$

$$y = a_i + c_i \qquad c_i = \text{birth year} \qquad (2)$$

- Solve out for c_i and a_i to get estimable combinations.

- Take the simpler case first:

$$\ln W(a, y, c) = \beta_0 + \underset{(\text{age})}{\beta_1 a_i} + \underset{(\text{year})}{\beta_2 y_i} + \underset{(\text{cohort})}{\beta_3 c_i} + u_i$$
$$y_i = a_i + c_i,$$

where y_1 is the current year, and c_i is the year of birth.

- Obviously, we get an exact linear dependence:

$$(\beta_0, \beta_1, \beta_2, \beta_3)$$

THE UNIVERSITY OF
CHICAGO

- Substitute $c_i = y_i - a_i$.

- $$\ln W_i = \alpha_0 + \beta_1 a_i + \beta_2 y_i + \beta_3 (y_i - a_i) + u_i$$
  $$= \alpha_0 + (\beta_1 - \beta_3) a_i + (\beta_2 + \beta_3) y_i + u_i$$

  can identify only combinations of coefficients.

- In a cross section, y_i is the same for everyone. The intercept is

$$[\alpha_0 + (\beta_2 + \beta_3) y_i].$$

THE UNIVERSITY OF CHICAGO

- We can estimate $(\beta_1 - \beta_3)$: age minus cohort effect.

- If $\beta_3 > 0$, we underestimate true β_1.

- Will longitudinal data rescue us? — Not necessarily.

- With panels, y_i moves with time. Recall that $y_i = a_i + c_i$.

- So we still have exact linear dependence. This is true if we have dummy variables in place of continuous variables (verify). Panel data will rescue us — if we have no year effects.

- We acquire similar problems in models with nonlinear terms:

$$y = a + c$$

$$\left.\begin{array}{l} y^2 = a^2 + 2ac + c^2 \\ ay = a^2 + ac \\ cy = ca + c^2 \end{array}\right\} \text{3 linear dependencies in these set-ups}$$

- Thus when we write

$$\begin{aligned} \ln W &= \beta_0 + \beta_1 a + \beta_2 y + \beta_3 c + \beta_4 a^2 + \beta_5 ac \\ &\quad + \beta_6 ay + \beta_7 cy + \beta_8 c^2 + \beta_9 y^2 + u, \end{aligned}$$

we cannot identify all of the parameters (only 3 second order parameters are estimable out of 6 total.

THE UNIVERSITY OF
CHICAGO

Theorem. In a model with interactions of order k with j variables and one linear restriction among the j variables, then of the $\binom{j+k-1}{k}$ coefficients of order k, only $\binom{j+k-2}{k}$ are estimable. (Heckman and Robb, in S. Feinberg and W. Mason, *Age, Period and Cohort Effects: Beyond the Identification Problem*, Springer, 1986).

E.g. $k = 2$, $j = 3$; 6 coefficients and 3 are estimable, as in the preceding example.

Theorem. In a model with ℓ restrictions on the j variables, then $\binom{j+k-\ell-1}{k}$ kth order coefficients are estimable (Heckman and Robb, 1986).

Question: Generalize this analysis for the case of polychotomous variables for age period and cohort effects.

THE UNIVERSITY OF
CHICAGO

- Return to the more general case. Substitute out for c_i and a_i, using (1) and (3):

$$\ln W_i = \alpha_0 + (\alpha_2 + \alpha_5)y + (\alpha_1 + \alpha_3 - \alpha_5)e_i$$
$$+ (\alpha_1 + \alpha_4 - \alpha_5)s_i + u_i.$$

- In a single cross section, y is the same for everyone. The intercept is then $\alpha_0 + (\alpha_2 + \alpha_5)y$, where y is year of cross section.

- Experience coefficient $= \alpha_1 + \alpha_3 - \alpha_5 = \alpha_3 + (\alpha_1 - \alpha_5)$ if later vintages get higher skills, $\alpha_5 > 0$ and downward bias (*e.g.* higher quality of schooling). If there is an aging effect (> 0, *e.g.* maturation) cannot separate. Produces upward bias for α_3.

Schooling Coefficient

- $\alpha_1 + \alpha_4 - \alpha_5 = \alpha_4 + (\alpha_1 - \alpha_5)$

- Vintage (cohort) effects lead to downward bias.

- Age effects, upward bias.

- Observe that from the
experience coefficient $-$ schooling coefficient:

$$(\alpha_1 + \alpha_3 - \alpha_5) - (\alpha_1 + \alpha_4 - \alpha_5) = \alpha_3 - \alpha_4.$$

- Can estimate difference in "returns" to experience net of
schooling.

- Observe that even if $\alpha_1 = 0$ (no aging effect), still can't estimate these coefficients.

- Is the solution **longitudinal data** (observations n the same people over time)—or **repeated cross section data** (observations on the same population over time but sampling different persons)?

- If $\alpha_2 = 0$,(no year effects), we can estimated α_5.

- Alternatively, for each c_i we can estimate $\alpha_1 + \alpha_3$, and hence we can estimate α_5.

- We also know $\alpha_1 + \alpha_4$. If $\alpha_1 = 0$, then α_3, α_4, α_5 identified.

- Observe the weakness in the procedure.

- If year effects are present, we have that there is no gain to going to longitudinal or repeated cross section data.

- We gain a parameter when we move to the panel or repeated cross sectional data.

Solutions in Literature

1. Redefine vintage (cohort) e.g. vintage fixed over period of years (e.g. a cohort of Depression babies).

- Then $\ln W = (\alpha_0 + \alpha_5 c) + \alpha_1 a + \alpha_2 y + \alpha_3 e + \alpha_4 s + u$.

- In single cross section, c and y are fixed.

- Substitute for e:

$$e = a_i - s_i$$

- Then

$$\ln W = [\alpha_0 + \alpha_5 c + \alpha_2 y] + (\alpha_1 + \alpha_3)a_i + (\alpha_4 - \alpha_3)s_i.$$

- We can estimate $\alpha_1 + \alpha_3$ and $\alpha_4 - \alpha_3$, and thus $\alpha_1 + \alpha_4$.

- Successive time periods for the same vintage gives us α_2 directly [since c doesn't move].

- If no age effect , we get $\alpha_3, \alpha_4, \alpha_2$, and from successive vintage estimations, we get α_5.

THE UNIVERSITY OF
CHICAGO

2. If we measure experience, $a_i \neq e_i + s_i$ (non-market breaks), we get break in linear dependence.

- Cost: better proxies may be endogenous.

- E.g. experience = cumulated hours.

- Results carry over in an obvious way to nonlinear models.

Example of Interpretive Pitfall

1. Johnson and Stafford (AER, 1974)
2. Weiss and Lillard (JPE, 1979)

- **Fact:** Disparity in real wages between recent Ph.D. entrants and experienced workers rose in *physics* and *mathematics* in the late 60s and early 70s. Not observed in the *social sciences*.

- **Why?** — Johnson-Safford story.

- Supplies of Ph.D.s enlarged by federal grants whil emand for scientific personnel declined. Wage rigidity at the top end motivated by specific human capital. Spot market / entrant market bears the brunt of the burden.

THE UNIVERSITY OF CHICAGO

- Weiss & Lillard: "experience – vintage" interaction (ec).

- Ignore age effect:

$$\begin{aligned}
\ln W(e, c, s, y) &= \varphi_0 + \varphi_1 e + \varphi_2 c + \varphi_3 y + \varphi_4 s \\
&\quad + \varphi_5 e^2 + \varphi_6 c^2 + \varphi_7 ec \\
&\quad + \varphi_8 ey + \varphi_9 cy + \varphi_{10} y^2
\end{aligned}$$

- Assume other powers and interactions are zero. Assume $\varphi_{10} = 0$.

- Johnson-Stafford: $\varphi_8 > 0$ or $\varphi_9 < 0$

- Weiss-Lillard: $\varphi_7 > 0$

- Recall that $y = e + s + c$.

THE UNIVERSITY OF CHICAGO

- Weiss-Lillard ignore year effects.

- We get Weiss-Lillard by substituting for y:

$$\begin{aligned}
\ln W(e, c, s) &= \varphi_0 + (\varphi_1 + \varphi_3)e + (\varphi_3 + \varphi_4)s \\
&\quad + (\varphi_2 + \varphi_3)c + (\varphi_5 + \varphi_8)e^2 \\
&\quad + \varphi_8 es + (\varphi_7 + \varphi_8 + \varphi_9)ec \\
&\quad + (\varphi_6 + \varphi_8)c^2
\end{aligned}$$

- Note that if $\varphi_7 = 0$ but $\varphi_9 > 0$, we get ec interaction, but it is "really" a year effect. If entry level wages fall relative to wages of experienced workers, the wage / experience profile is steeper in more recent cross-sections.

THE UNIVERSITY OF
CHICAGO

- Looking at social scientists where no interaction appears favors Johnson-Stafford.

- Moral: auxiliary evidence and theory break the identification problem.

THE UNIVERSITY OF CHICAGO

Cohort vs. Cross-Section Internal Rate of Return

- Take a cohort rate of return.

  1. $Y_{a,c}^h$ is the earnings of a high school graduate of cohort c at age a.

  2. $Y_{a,c}^d$ is the earnings of a droupout of cohort c at age a.

  3. $\rho_c = IRR_c$ (cohort internal rate of return).

  4.
  $$\sum_{a=0}^{A} \frac{Y_{a,c}^h - Y_{a,c}^d}{(1+\rho_c)^a} = 0.$$

- The cross-section consists of a set of member of different cohorts.

- Start with $c = 1$ as the youngest age group and proceed.

- At a point in time, we have $a = 0 \Longrightarrow c = 1$; $c + a = t$..

- The cross-section internal rate of return is

$$\sum_{a=0}^{A} \frac{\left(Y_{a,1-a}^{h} - Y_{a,1-a}^{d} \right)}{(1 + \rho_t)^a} = 0,$$

where $A + 1$ is the maximum age in the population.

THE UNIVERSITY OF
CHICAGO

- When can $\rho_c = \rho_t$?

- This can occur if the environment is stationary.

- With steady growth in differentials, it cannot help explain $\rho_c = \rho_t$.

- The case

$$\Delta_{a,c}^{h,d} = Y_{a,c}^h - Y_{a,c}^d \qquad (3)$$
$$\Delta_{a,c+j}^{h,d} = \left(\Delta_{a,c}^{h,d}\right)(1+g)^j$$

will not work.

- With constant growth, g cannot explain $\rho_t = \rho_c$ (!):

$$c = 0, 1 \qquad t = a + c.$$

- Consider a model with 2 cohorts, focus on cohort $c = 0$. ρ_c is the root of

$$0 = Y_{0,0}^h - Y_{0,0}^d + \frac{Y_{1,0}^h - Y_{1,0}^d}{1 + \rho_c}.$$

- Cross-section at $t = 1$, when cohort c enters, is

$$0 = Y_{0,0}^h - Y_{0,0}^d + \frac{Y_{1,-1}^h - Y_{1,-1}^d}{1 + \rho_t} text.$$

- In general, $\rho_c \neq \rho_t$. More generally, for cohort $\bar{c}$, the benchmark cohort, $\rho_{\bar{c}}$ is the IRR that solves

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^h - Y_{a,\bar{c}}^d \right)}{(1 + \rho_{\bar{c}})^a} = 0.$$

- Cross section in year $t = \bar{c}$ produces the equation

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}-a}^h - Y_{a,\bar{c}-a}^d\right)}{(1+\rho_t)^a} = 0,$$

where ρ_t is the root.

- If growth rates across cohorts are benchmarked against $\bar{c}$, we obtain

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^h - Y_{a,\bar{c}}^d\right)(1+g)^{-a}}{(1+\rho_t)^a} = 0$$

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^h - Y_{a,\bar{c}}^d\right)}{[(1+\rho_t)(1+g)]^a} = 0,$$

so clearly $\rho_t < \rho_c$.

THE UNIVERSITY OF
CHICAGO

- Suppose that there are no cohort effects but that there are smooth time effects, say, $1 + \varphi$.

- Then the cohort rate of return is calculated as the root of the following equation in which the choice of a cohort $\bar{c}$ as a benchmark is innocuous:

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^{h} - Y_{a,\bar{c}}^{d}\right)(1+\varphi)^{a}}{(1+\rho_{\bar{c}})^{a}} = 0$$

- The cross-section rate at time $t = \bar{c}$ is

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}}^{h} - Y_{a,\bar{c}}^{d}\right)}{(1+\rho_{t})^{a}} = 0, \qquad t = \bar{c},$$

where clearly if $\varphi > 0$, then $\rho_{\bar{c}} > \rho_{t}$.

THE UNIVERSITY OF
CHICAGO

- Better notation — distinguish outcomes at age a, cohort c, period t:

$$Y_{a,c,t}^h; \; Y_{a,c,t}^d$$
$$\Delta_{a,c,t}^{h,d} \;\; = \;\; Y_{a,c,t}^h - Y_{a,c,t}^d.$$

- No cohort effects means $Y_{a,c,t}^j = Y_{a,-,t}^j \; \forall c$. "–" sets the argument to a constant.

THE UNIVERSITY OF
CHICAGO

Pure Time Effects

- Take cohort $c = 0$ at time t:

$$\sum_{a=0}^{A} \frac{\left(Y_{a,0,t+a}^h - Y_{a,0,t+a}^d\right)}{\left(1 + \rho_c\right)^a} = 0$$

- Cross section at $t = 0$ for $c = 0$:

$$\sum_{a=0}^{A} \frac{\left(Y_{a,-a,t}^h - Y_{a,-a,t}^d\right)}{\left(1 + \rho_t\right)^a} = 0, \qquad t = 0$$

- No time effects means $Y_{a,c,t}^j = Y_{a,c,-}^j \ \forall t$.

THE UNIVERSITY OF
CHICAGO

- A model with pure cohort effects and no time effects writes, for cohort $\bar{c}$,

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c},-}^h - Y_{a,\bar{c},-}^d\right)}{(1 + \rho_{\bar{c}})^a} = 0.$$

- This defines a cohort rate of return.

- The cross-section at time $t = \bar{c}$ writes

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c},\bar{c}+a}^h - Y_{a,\bar{c},\bar{c}+a}^d\right)(1 + g)^{\bar{c}}}{(1 + \rho_{\bar{c}})^a} = 0.$$

- So if $g > 0$, then $\rho_{\bar{c}} > \rho_t$ $(t = \bar{c})$.

THE UNIVERSITY OF
CHICAGO

- A model with pure time effects $(1 + \varphi)$ writes, for time $t = \bar{c}$, the cohort return for entry cohort $\bar{c}$ as

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c},\bar{c}+a}^{h} - Y_{a,\bar{c},\bar{c}+a}^{d}\right)(1 + g)^{\bar{c}}}{(1 + \rho_{\bar{c}})^{a}} = 0 text.$$

- Benchmarking on the $c = 0$ cohort,

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c},\bar{c}}^{h} - Y_{a,\bar{c},\bar{c}}^{d}\right)(1 + \varphi)^{a}(1 + g)^{\bar{c}}}{(1 + \rho_{\bar{c}})^{a}} = 0.$$

THE UNIVERSITY OF
CHICAGO

- The cross-section return at time $\bar{c}$ is

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}-a,\bar{c}}^{h} - Y_{a,\bar{c}-a,\bar{c}}^{d}\right)}{\left(1+\rho_t\right)^a} = 0,$$

where $Y_{a,\bar{c}-a,\bar{c}}^{h} = Y_{a,c^*,\bar{c}}^{h}$ for all c^*, $t = \bar{c}$, if there are only pure time effects.

THE UNIVERSITY OF
CHICAGO

- Suppose we have both time and cohort effects. Then we have that the cross-section is

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c}-a,\bar{c}}^{h} - Y_{a,\bar{c}-a,\bar{c}}^{d}\right)}{\left(1 + \rho_t\right)^a} = 0.$$

- These can be written at time $t = \bar{c}$ as

$$\sum_{a=0}^{A} \frac{\left(Y_{a,\bar{c},\bar{c}}^{h} - Y_{a,\bar{c},\bar{c}}^{d}\right)\left(1 + g\right)^{\bar{c}-a}}{\left(1 + \rho_t\right)^a} = 0.$$

- Thus, if the cohort rate $(1 + g)^{\bar{c}-a} = (1 + \varphi)^a (1 + g)^{\bar{c}}$ for all $\bar{c}$, we can get the result.

THE UNIVERSITY OF
CHICAGO

- This requires that

$$1 + g = \frac{1}{1 + \varphi} \Rightarrow g = \frac{-\varphi}{1 + \varphi}.$$

- This seems to characterize the IRR for high school vs. dropouts. Cohort growth rate factor is the inverse of the time rate.

THE UNIVERSITY OF
CHICAGO