Recitation Period 8: Understanding Fixed Effects

Alexander Dean

March 6, 2025

This proof solves proposition 1 of Gibbons and coauthors' 2023 paper on the bias of Fixed Effects Estimator:

Proposition 1 (Bias of Fixed Effects under Heterogeneous Treatment Effects)**.** *Consider the outcome model*

$$y_i \;=\; x_i\,\beta_{g(i)} \;+\; z_i'\,\gamma \;+\; \varepsilon_i, \tag{1}$$

where

(i) $g(i) \in \{1, 2, \ldots, G\}$ *denotes the group to which observation* i *belongs;*

(ii) each group g *may have its own slope* β_g *(i.e., treatment effect);*

(iii) z_i *is a vector of additional covariates (including group dummies, time dummies, etc.);*

(iv) ε_i *is a mean-zero error term satisfying standard OLS-type assumptions.*

Define the average treatment effect *(ATE) as*

$$\beta_{\text{ATE}} \;=\; \sum_{g=1}^{G}\left(\tfrac{N_g}{N}\right)\beta_g, \quad \textit{where } N_g = \sum_{i:g(i)=g} 1 \textit{ and } N = \sum_{g=1}^{G} N_g.$$

Let $\hat{b}_{\text{FE}}$ *be the usual OLS estimator from a regression that* includes group fixed effects *but* does not interact x_i with group dummies*. Then if* $\beta_g \neq \beta_{g'}$ *for at least one pair* (g, g') *and the within-group variances of* x_i *(after partialing out* z_i*) differ,* $\hat{b}_{\text{FE}}$ *is generally* biased *(and inconsistent) for* β_{ATE}*.*

Proof of Proposition. **Step 1: Residualize with respect to the other covariates z_i.**
1. Let X, Z, and A collectively denote the full set of x_i, z_i, and group membership data. Define the "annihilator matrix"

$$M \;=\; I_N - Z\,(Z'Z)^{-1}Z',$$

which projects any N-vector orthogonally to the column space of Z.

2. For each observation i, define:

$$\tilde{y}_i \;=\; (M\,y)_i, \quad \tilde{x}_i \;=\; (M\,x)_i, \quad \tilde{\varepsilon}_i \;=\; (M\,\varepsilon)_i.$$

Then *by construction* $M\,z = 0$, so the model (1) becomes

$$\tilde{y}_i \;=\; \tilde{x}_i\,\beta_{g(i)} \;+\; \tilde{\varepsilon}_i,$$

because $Mz_i = 0$ for each observation i.

Step 2: Write down the Fixed Effects (FE) estimator in residualized form.
The usual FE approach is equivalent to OLS regression of $\tilde{y}_i$ on $\tilde{x}_i$, i.e.

$$\hat{b}_{\text{FE}} \;=\; \frac{\sum_{i=1}^{N} \tilde{x}_i\,\tilde{y}_i}{\sum_{i=1}^{N} \tilde{x}_i^2} \;=\; \frac{\tilde{x}'\tilde{y}}{\tilde{x}'\tilde{x}}. \tag{2}$$

Since $\tilde{y}_i = \tilde{x}_i\,\beta_{g(i)} + \tilde{\varepsilon}_i$, we have

$$\sum_{i=1}^{N} \tilde{x}_i\,\tilde{y}_i \;=\; \sum_{i=1}^{N} \tilde{x}_i^2\,\beta_{g(i)} \;+\; \sum_{i=1}^{N} \tilde{x}_i\,\tilde{\varepsilon}_i.$$

Taking conditional expectation on (X, Z, A), and assuming $\mathbb{E}[\tilde{x}_i\,\tilde{\varepsilon}_i \mid X, Z, A] = 0$ (the usual exogeneity assumption in OLS), we get

$$\mathbb{E}\big[\hat{b}_{\text{FE}} \;\big|\; X, Z, A\big] \;=\; \frac{\sum_{i=1}^{N} \tilde{x}_i^2\,\beta_{g(i)}}{\sum_{i=1}^{N} \tilde{x}_i^2}. \tag{3}$$

Step 3: Express the sum in terms of group-specific components.
1. Partition the observations into G groups. Let $N_g = \sum_{i=1}^{N} \mathbf{1}_{\{g(i)=g\}}$ be the number of observations in group g. Set $\pi_g = N_g/N$.
2. In each group g, define the sample variance of $\tilde{x}_i$ as

$$\widehat{\text{Var}}\big(\tilde{x}_i \mid g\big) \;=\; \frac{1}{N_g} \sum_{\substack{i=1 \\ g(i)=g}}^{N} \tilde{x}_i^2,$$

assuming that the group-mean of $\tilde{x}_i$ has been subtracted out if necessary. Likewise, define

$$\widehat{\text{Var}}\big(\tilde{x}_i\big) \;=\; \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i^2.$$

Hence

$$\sum_{i=1}^{N} \tilde{x}_i^2 \;=\; N \times \widehat{\text{Var}}(\tilde{x}_i), \qquad \sum_{\{i: g(i)=g\}} \tilde{x}_i^2 \;=\; N_g \times \widehat{\text{Var}}(\tilde{x}_i \mid g).$$

3. Rewrite the numerator of (3):

$$\sum_{i=1}^{N} \tilde{x}_i^2 \, \beta_{g(i)} \;=\; \sum_{g=1}^{G} \sum_{\{i:g(i)=g\}} \tilde{x}_i^2 \, \beta_g \;=\; \sum_{g=1}^{G} \big(N_g \, \widehat{\mathrm{Var}}(\tilde{x}_i \mid g)\big) \, \beta_g.$$

Thus

$$\mathbb{E}\big[\hat{b}_{\mathrm{FE}} \mid X, Z, A\big] \;=\; \frac{\sum_{g=1}^{G} N_g \, \widehat{\mathrm{Var}}(\tilde{x}_i \mid g) \, \beta_g}{N \, \widehat{\mathrm{Var}}(\tilde{x}_i)} \;=\; \sum_{g=1}^{G} \pi_g \, \beta_g \, \frac{\widehat{\mathrm{Var}}(\tilde{x}_i \mid g)}{\widehat{\mathrm{Var}}(\tilde{x}_i)}.$$

Step 4: Compare to β_{ATE}.
Recall that $\beta_{\mathrm{ATE}} = \sum_{g=1}^{G} \pi_g \, \beta_g$. Subtracting, we obtain

$$\mathbb{E}[\hat{b}_{\mathrm{FE}} \mid X, Z, A] \;-\; \beta_{\mathrm{ATE}} \;=\; \sum_{g=1}^{G} \big(\pi_g \, \beta_g\big) \left[\tfrac{\widehat{\mathrm{Var}}(\tilde{x}_i \mid g)}{\widehat{\mathrm{Var}}(\tilde{x}_i)} - 1\right].$$

Therefore, *unless* $\widehat{\mathrm{Var}}(\tilde{x}_i \mid g)$ is the same across all groups g or $\{\beta_g\}$ is the same in every group, the above sum will *not* be zero. That is,

$$\mathbb{E}[\hat{b}_{\mathrm{FE}} \mid X, Z, A] \;\neq\; \beta_{\mathrm{ATE}}.$$

Step 5: Conclude bias and inconsistency.
1. We have shown that in expectation (conditional on X, Z, A), $\hat{b}_{\mathrm{FE}}$ does not recover β_{ATE}, hence there is a *bias*.
2. As $N \to \infty$, by standard laws of large numbers for sample variances, $\widehat{\mathrm{Var}}(\tilde{x}_i \mid g)$ converges to the true within-group variance of $\tilde{x}_i$ (provided that variance exists). Thus the ratio $\frac{\widehat{\mathrm{Var}}(\tilde{x}_i \mid g)}{\widehat{\mathrm{Var}}(\tilde{x}_i)}$ converges to the limit of the ratio of those variances. Consequently, if β_g differs by group and the ratio of within-group variances differs across groups, the limiting value of $\hat{b}_{\mathrm{FE}}$ is not β_{ATE}, showing *inconsistency* as well.
Thus,

$$\hat{b}_{\mathrm{FE}} \;\xrightarrow{p}\; \sum_{g=1}^{G} \pi_g \, \beta_g \, \frac{\mathrm{Var}(\tilde{x}_i \mid g)}{\mathrm{Var}(\tilde{x}_i)} \;\neq\; \sum_{g=1}^{G} \pi_g \, \beta_g \;=\; \beta_{\mathrm{ATE}},$$

whenever $\{\beta_g\}$ and $\mathrm{Var}(\tilde{x}_i \mid g)$ vary across g. This completes the proof. $\square$