Supplementary material

S1 Derivation of $\hat{\tau}_{\mathrm{FE}}$

Proposition S1. *Let y denote a vector of outcomes and X a matrix in which the first column is the treatment assignment and columns 2 to $s+1$ are dummy variables for each of s strata. Let n_j, p_j, n_j^1, $\overline{y}_j$, $\overline{y}_j^0$, and $\overline{y}_j^1$ denote, respectively, the size of stratum j, the share of units in stratum j in treatment, the number of units in stratum j in treatment, the average outcome of units in stratum j, and the stratum j average (observed) outcome among treated and control units, respectively.*

Then, OLS regression of y on X yields:

$$\begin{bmatrix} \hat{\tau}_{\mathrm{FE}} \\ \hat{\alpha}_{\mathrm{FE}}^1 \\ \vdots \\ \hat{\alpha}_{\mathrm{FE}}^s \end{bmatrix} = (X'X)^{-1}X'y = \begin{bmatrix} \dfrac{\sum_j n_j p_j(1-p_j)(\overline{y}_j^1 - \overline{y}_j^0)}{\sum_j n_j p_j(1-p_j)} \\ \overline{y}_1 - p_1 \dfrac{\sum_j n_j p_j(1-p_j)(\overline{y}_j^1 - \overline{y}_j^0)}{\sum_j n_j p_j(1-p_j)} \\ \vdots \\ \overline{y}_s - p_s \dfrac{\sum_j n_j p_j(1-p_j)(\overline{y}_j^1 - \overline{y}_j^0)}{\sum_j n_j p_j(1-p_j)} \end{bmatrix}.$$

Proof. To establish the result, note first that the matrix $X'X$ can be represented as a block matrix:

$$X'X = \begin{bmatrix} n^1 & \mathbf{n}^{1\prime} \\ \mathbf{n}^1 & M, \end{bmatrix},$$

where n^1 is the number of units in treatment, $\mathbf{n}^1 = [n_1^1, n_2^1, \ldots, n_s^1]'$ is the number treated in each stratum, and M is a diagonal matrix reporting the number of units in each stratum.

From the inversion of block matrices (see Eqn 2.8.17 in [1]):

$$(X'X)^{-1} = \begin{bmatrix} (n^1 - \mathbf{n}^{1\prime}M^{-1}\mathbf{n}^1)^{-1} & -(n^1 - \mathbf{n}^{1\prime}M^{-1}\mathbf{n}^1)^{-1}\mathbf{n}^{1\prime}M^{-1} \\ -M^{-1}\mathbf{n}^1(n^1 - \mathbf{n}^{1\prime}M^{-1}\mathbf{n}^1)^{-1} & M^{-1} + M^{-1}\mathbf{n}^1(n^1 - \mathbf{n}^{1\prime}M^{-1}\mathbf{n}^1)^{-1}\mathbf{n}^{1\prime}M^{-1} \end{bmatrix}.$$

Observing that

$$\mathbf{n}^{1\prime}M^{-1} = (p_1, p_2 \ldots p_s)$$

and defining

$$w \coloneqq (n^1 - \mathbf{n}^{1\prime}M^{-1}\mathbf{n}^1)^{-1} = \frac{1}{\sum_j p_j n_j - \sum_j p_j^2 n_j} = \frac{1}{\sum_j n_j p_j(1-p_j)},$$

we have

$$(X'X)^{-1} = w \cdot \begin{bmatrix} 1 & -p_1 & -p_2 & \cdots & -p_s \\ -p_1 & \frac{1}{n_1 w} + p_1^2 & p_1 p_2 & \cdots & p_1 p_s \\ -p_2 & p_2 p_1 & \frac{1}{n_2 w} + p_2^2 & \cdots & p_2 p_s \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -p_s & p_s p_1 & p_s p_2 & \cdots & \frac{1}{n_s w} + p_s^2 \end{bmatrix}.$$

Similarly

$$X'y = \left(\sum_{i:d_i=1} y_i, \sum_{i \in X_1} y_i, \ldots, \sum_{i \in X_s} y_i \right).$$

$\hat{\tau}_{\mathrm{FE}}$ is then the inner product of the first row of $(X'X)^{-1}$ and $X'y$:

$$\hat{\tau}_{\mathrm{FE}} = w\left(\sum_{i:d_i=1} y_i - p_1 \sum_{i \in X_1} y_i - p_2 \sum_{i \in X_2} y_i - \ldots - p_s \sum_{i \in X_s} y_i \right).$$

To simplify, observe that $\sum_{i:d_i=1} y_i = \sum_j n_j p_j \overline{y}_j^1$ and $\sum_{i \in j} y_i = n_j (p_j \overline{y}_j^1 + (1 - p_j) \overline{y}_j^0)$, and so

$$\begin{aligned} \hat{\tau}_{\mathrm{FE}} &= w\left(\sum_j n_j p_j \overline{y}_j^1 - \sum_j n_j p_j (p_j \overline{y}_j^1 + (1 - p_j) \overline{y}_j^0) \right) \\ &= \frac{\sum_j n_j p_j (1 - p_j)(\overline{y}_j^1 - \overline{y}_j^0)}{\sum_j n_j p_j (1 - p_j)}. \end{aligned}$$

In the same way

$$\begin{aligned} \hat{\alpha}_{\mathrm{FE}}^j &= -w p_j \left(\sum_{i:d_i=1} y_i - p_1 \sum_{i \in X_1} y_i - p_2 \sum_{i \in X_2} y_i - \ldots - p_s \sum_{i \in X_s} y_i \right) - w\left(-\frac{1}{n_j w} \right) \sum_{i \in X_j} y_i \\ &= \overline{y}_j - p_j \hat{\tau}_{\mathrm{FE}}. \end{aligned}$$

□

S2 Omitted steps in proof of Proposition 1

The proof for Proposition 1 relies on an equivalence between

$$\sum_j \frac{p_j(1 - p_j) w_j}{\sum_j p_j (1 - p_j) w_j} \tau_j \leq \sum_j \frac{p_j w_j}{\sum_j p_j w_j} \tau_j \tag{S1}$$

and

$$\sum_j \left(\frac{p_j w_j}{\sum_j p_j w_j} - \frac{p_j^2 w_j}{\sum_j p_j^2 w_j} \right) \tau_j \leq 0. \tag{S2}$$

To see this equivalence, define $\alpha = \sum_j p_j w_j$ and $\beta = \sum_j p_j^2 w_j$.

Condition (S1) can then be written as:

$$\sum_j p_j w_j \tau_j \frac{1-p_j}{\alpha-\beta} \leq \sum_j p_j w_j \tau_j \frac{1}{\alpha}$$
$$\leftrightarrow$$
$$\sum_j p_j w_j \tau_j \left(\frac{1-p_j}{\alpha-\beta} - \frac{1}{\alpha}\right) \leq 0$$
$$\leftrightarrow$$
$$\sum_j p_j w_j \tau_j \left(\frac{\beta-\alpha p_j}{(\alpha-\beta)\alpha}\right) \leq 0$$
$$\leftrightarrow$$
$$\sum_j p_j w_j \tau_j \left(\frac{1}{\alpha} - \frac{p_j}{\beta}\right) \leq 0,$$

where the last step results from multiplying across by $\frac{\alpha-\beta}{\beta} > 0$.

Resubstituting for α and β yields the result.

S3 Derivation of equation (11)

In the linear case, monotonicity is satisfied and we can therefore write

$$\tau_{\mathrm{FE}} = \lambda \tau_{\mathrm{ATT}} + (1-\lambda)\tau_{\mathrm{ATC}}.$$

Substituting from equations (5), (6), and (8)

$$\frac{\sum_j p_j(1-p_j)w_j\tau_j}{\sum_j p_j(1-p_j)w_j} = \lambda \frac{\sum_j p_j w_j \tau_j}{\sum_j p_j w_j} + (1-\lambda)\frac{\sum_j (1-p_j)w_j\tau_j}{\sum_j (1-p_j)w_j}.$$

With treatment effects linear in p_j we have for some β

$$\frac{\sum_j p_j(1-p_j)w_j\beta p_j}{\sum_j p_j(1-p_j)w_j} = \lambda \frac{\sum_j p_j w_j \beta p_j}{\sum_j p_j w_j} + (1-\lambda)\frac{\sum_j (1-p_j)w_j\beta p_j}{\sum_j (1-p_j)w_j}.$$

Dividing across by β, and gathering terms gives

$$\frac{\sum_j p_j^2(1-p_j)w_j}{\sum_j p_j(1-p_j)w_j} = \lambda \frac{\sum_j p_j^2 w_j}{\sum_j p_j w_j} + (1-\lambda)\frac{\sum_j p_j(1-p_j)w_j}{\sum_j (1-p_j)w_j}.$$

Solving for λ then yields

$$\lambda = \frac{\frac{\sum_j p_j^2(1-p_j)w_j}{\sum_j p_j(1-p_j)w_j} - \frac{\sum_j p_j(1-p_j)w_j}{\sum_j (1-p_j)w_j}}{\frac{\sum_j p_j^2 w_j}{\sum_j p_j w_j} - \frac{\sum_j p_j(1-p_j)w_j}{\sum_j (1-p_j)w_j}}.$$

S4 Code illustration

There are many approaches that can be used to generate unbiased estimates in the presence of heterogeneous but known propensities across strata. I illustrate by using the `R` package `DeclareDesign` to simulate data from a version of Example 1 and show the performance of five estimation strategies: pooled OLS, OLS with stratum dummies (fixed-effects), inverse propensity weighting (IPW), OLS but with interactions between treatment and demeaned stratum dummies, following the study by Lin [2], and blocked differences in means. This code draws from material in the study by Blair et al. [3].

Code to compare estimators for Example 1

```
library(DeclareDesign)
prob <- c(.067, .5, .933)
design <-
   declare_model(
     block = add_level(N = 3, p = prob, tau = c(3, -3, 3)),
     unit = add_level(N = 1000, Y0 = 10*(p + rnorm(N)), Y1 = Y0 + tau)) +
   declare_inquiry(ATE = mean(Y1 - Y0)) +
   declare_assignment(Z = block_ra(blocks = block, block_prob = prob)) +
   declare_measurement(
     ipw = 1/(Z*p + (1-Z)*(1-p)),
     Y = Z*Y1 + (1-Z)*Y0) +
   declare_estimator(Y ~ Z, .method = lm_robust,
     label = "Pooled") +
   declare_estimator(Y ~ Z + block, .method = lm_robust,
     label = "Fixed effects") +
   declare_estimator(Y ~ Z, blocks = block, .method = difference_in_means,
     label = "Blocked differences in means") +
   declare_estimator(Y ~ Z, covariates = block, .method = lm_lin,
     label = "Interactions (Lin approach)") +
   declare_estimator(Y ~ Z, .method = lm_robust, weight = ipw,
     label = "Inverse propensity weights")
simulate_design(design) |>
   group_by(estimator) |>
   summarize(
     SE_bias = mean(std.error - sd(estimate)),
     ATE_bias = mean(estimate - estimand) )
```

The code is shown in Box D. Table S1 shows the results, highlighting the poor performance of both the pooled approach and the fixed-effects approach in this setting. IPW, the interaction model, and blocked differences in means are all unbiased, though they differ in the performance of standard errors – assessed here as the difference between the estimated average standard error and the estimated standard deviation of the sampling distribution of estimates under each estimator.

Table S1: Performance of five strategies to estimate average treatment effects

Estimator	Bias of standard errors	Bias of estimates
Pooled	0.02	5
Fixed effects	0.00	−2
Inverse propensity weights	0.04	0
Interactions (Lin approach)	0.00	0
Blocked differences in means	0.00	0

References

[1] Lin W. Agnostic notes on regression adjustments to experimental data: Reexamining freedmanas critique. Ann Appl Stat. 2013;7(1):295–318.
[2] Bernstein DS. Matrix mathematics: theory, facts, and formulas. Princeton University Press; 2009.
[3] Blair G, Coppock A, Humphreys M.. The trouble with controlling for blocks; 2018. Accessed: 2024-10-14. https://declaredesign.org/blog/posts/biased-fixed-effects.html.