B VARIANCE AND COVARIANCE OF ESTIMATORS

B.1 Simplified data-generating setting

Throughout this section, we assume that the underlying data-generating process for the cluster-level outcomes follows a mixed-effects model:

$$Y_{i,j} = \mu + \alpha_i + \theta_j + X_{i,j}\beta + \epsilon_{i,j} \tag{B1}$$

where the α_i are independently distributed with mean 0 and variance τ^2, the $\epsilon_{i,j}$ are independently distributed with mean 0 and variance $\sigma^2_{i,j}$ and $\alpha_i \perp\!\!\!\perp \epsilon_{i',j'}$ for all i, i', j'. Note that this is similar to the risk difference model in Simulation 1 in Section 3, but differs in that we force $\epsilon_{i,j} \perp\!\!\!\perp \alpha_i$, so the mean-variance relationship no longer holds.

Thus:

$$\text{Var}(Y_{i,j}) = \tau^2 + \sigma^2_{i,j}.$$

$$\text{Cov}(Y_{i,j},\ Y_{i',j'}) = \begin{cases} \tau^2, & i = i', j \neq j' \\ 0, & i \neq i' \end{cases}.$$

We assume, as in the simulations in Section 3 that there are J periods, $J-1$ clusters, with cluster i on control for periods $1, \dots, i$ and on intervention for periods $i+1, \dots, J$.

B.2 Variance of the non-parametric within-period estimator

The non-parametric within-period estimator is given by:

$$\hat{\beta}^{NPWP} = \sum_{j=2}^{J-1} v_j \left(\frac{\sum_{i=1}^{j-1} Y_{i,j}}{j-1} - \frac{\sum_{i=j}^{J-1} Y_{i,j}}{J-j} \right) \equiv \sum_{j=2}^{J-1} v_j A_j, \tag{B2}$$

for weights v_j that sum to 1. For ease of variance calculations, we here assume, contrary to the estimator used Section 3, that the weights v_j are fixed and independent of the data (they can still depend on the number of clusters in the intervention and control conditions in period j, however).

$$\begin{aligned}\text{Var}(A_j) &= \text{Var}\left(\frac{\sum_{i=1}^{j-1} Y_{i,j}}{j-1} - \frac{\sum_{i=j}^{J-1} Y_{i,j}}{J-j} \right) = \sum_{i=1}^{j-1} \frac{\text{Var}(Y_{i,j})}{(j-1)^2} + \sum_{i=j}^{J-1} \frac{\text{Var}(Y_{i,j})}{(J-j)^2} \\ &= \frac{\sum_{i=1}^{j-1}(\sigma^2_{i,j} + \tau^2)}{(j-1)^2} + \frac{\sum_{i=j}^{J-1}(\sigma^2_{i,j} + \tau^2)}{(J-j)^2} = \frac{J-1}{(J-j)(j-1)}\tau^2 + \frac{\sum_{i=1}^{j-1} \sigma^2_{i,j}}{(j-1)^2} + \frac{\sum_{i=j}^{J-1} \sigma^2_{i,j}}{(J-j)^2}.\end{aligned}$$

For $j \neq k$ (WLOG, assume $k > j$):

$$\begin{aligned}\text{Cov}(A_j, A_k) &= \text{Cov}\left(\frac{\sum_{i=1}^{j-1} Y_{i,j}}{j-1} - \frac{\sum_{i=j}^{J-1} Y_{i,j}}{J-j},\ \frac{\sum_{i=1}^{k-1} Y_{i,k}}{k-1} - \frac{\sum_{i=k}^{J-1} Y_{i,k}}{J-k} \right) \\ &= \frac{1}{(j-1)(k-1)} \sum_{i=1}^{j-1} \text{Cov}(Y_{i,j},\ Y_{i,k}) - \frac{1}{(J-j)(k-1)} \sum_{i=j}^{k-1} \text{Cov}(Y_{i,j},\ Y_{i,k}) + \frac{1}{(J-j)(J-k)} \sum_{i=k}^{J-1} \text{Cov}(Y_{i,j},\ Y_{i,k}) \\ &= \tau^2 \left(\frac{1}{k-1} - \frac{k-j}{(J-j)(k-1)} + \frac{1}{J-j} \right) = \frac{\tau^2}{(J-j)(k-1)} (J - j - (k-j) + k - 1) \\ &= \tau^2 \frac{J-1}{(J-j)(k-1)}.\end{aligned}$$

Thus:

$$\begin{aligned}\text{Var}(\hat{\beta}^{NPWP}) &= \text{Cov}\left(\sum_{j=2}^{J-1} v_j A_j,\ \sum_{k=2}^{J-1} v_k A_k \right) = \sum_{j=2}^{J-1} v_j^2 \text{Var}(A_j) + 2 \sum_{j=2}^{J-2} \sum_{k=j+1}^{J-1} v_j v_k \text{Cov}(A_j,\ A_k) \\ &= \sum_{j=2}^{J-1} v_j^2 \left[\frac{J-1}{(J-j)(j-1)}\tau^2 + \frac{\sum_{i=1}^{j-1} \sigma^2_{i,j}}{(j-1)^2} + \frac{\sum_{i=j}^{J-1} \sigma^2_{i,j}}{(J-j)^2} \right] + 2 \sum_{j=2}^{J-2} \sum_{k=j+1}^{J-1} v_j v_k \tau^2 \frac{J-1}{(J-j)(k-1)}.\end{aligned} \tag{B3}$$

Note that the variance increases as any $\sigma^2_{i,j}$ or τ^2 increase, holding all else constant.

B.3 Variance of the crossover estimator

The crossover estimator is given by:

$$\hat{\beta}^{CO} = \sum_{j=2}^{J-1} w_j \left[Y_{j-1,j} - Y_{j-1,j-1} - \frac{1}{J-j} \sum_{\ell=j}^{J-1} (Y_{\ell,j} - Y_{\ell,j-1}) \right] \equiv \sum_{j=2}^{J-1} w_j B_j, \tag{B4}$$

for weights w_j that sum to 1.

Note that for all i, $Y_{i,j} - Y_{i,j'} = \epsilon_{ij} - \epsilon_{ij'} + \beta I(X_{ij} \neq X_{ij'})$. Hence:

$$\mathrm{Var}(B_j) = \mathrm{Var}\left(\epsilon_{j-1,j} - \epsilon_{j-1,j-1} + \beta - \frac{1}{J-j} \sum_{\ell=j}^{J-1} (\epsilon_{\ell,j} - \epsilon_{\ell,j-1}) \right) = \sigma^2_{j-1,j} + \sigma^2_{j-1,j-1} + \frac{1}{(J-j)^2} \sum_{\ell=j}^{J-1} (\sigma^2_{\ell,j} + \sigma^2_{\ell,j-1}).$$

$$\begin{aligned} \mathrm{Cov}(B_j, B_{j+1}) &= \mathrm{Cov}\left(\epsilon_{j-1,j} - \epsilon_{j-1,j-1} + \beta - \frac{1}{J-j} \sum_{\ell=j}^{J-1} (\epsilon_{\ell,j} - \epsilon_{\ell,j-1}),\ \epsilon_{j,j+1} - \epsilon_{j,j} + \beta - \frac{1}{J-j-1} \sum_{m=j+1}^{J-1} (\epsilon_{m,j+1} - \epsilon_{m,j}) \right) \\ &= \frac{1}{J-j} \left[\sigma^2_{j,j} - \sum_{\ell=j+1}^{J-1} \frac{\sigma^2_{\ell,j}}{J-j-1} \right]. \\ \mathrm{Cov}(B_j, B_k) &= \mathrm{Cov}\left(\epsilon_{j-1,j} - \epsilon_{j-1,j-1} + \beta - \frac{1}{J-j} \sum_{\ell=j}^{J-1} (\epsilon_{\ell,j} - \epsilon_{\ell,j-1}),\ \epsilon_{k-1,k} - \epsilon_{k-1,k-1} + \beta - \frac{1}{J-k} \sum_{m=k}^{J-1} (\epsilon_{m,k} - \epsilon_{m,k-1}) \right) \\ &= 0 \text{ for } k > j+1. \end{aligned}$$

Thus:

$$\begin{aligned} \mathrm{Var}(\hat{\beta}^{CO}) &= \mathrm{Cov}\left(\sum_{j=2}^{J-1} w_j B_j,\ \sum_{k=2}^{J-1} w_k B_k \right) = \sum_{j=2}^{J-1} w_j^2 \mathrm{Var}(B_j) + 2 \sum_{j=2}^{J-2} w_j w_{j+1} \mathrm{Cov}(B_j,\ B_{j+1}) \\ &= \sum_{j=2}^{J-1} w_j^2 \left[\sigma^2_{j-1,j} + \sigma^2_{j-1,j-1} + \frac{1}{(J-j)^2} \sum_{\ell=j}^{J-1} (\sigma^2_{\ell,j} + \sigma^2_{\ell,j-1}) \right] + 2 \sum_{j=2}^{J-2} w_j w_{j+1} \frac{1}{J-j} \left(\sigma^2_{j,j} - \sum_{\ell=j+1}^{J-1} \frac{\sigma^2_{\ell,j}}{J-j-1} \right). \end{aligned} \tag{B5}$$

Note that, unlike the variance of the NPWP estimator, this variance does not depend on τ^2.

B.4 Covariance of the non-parametric within-period estimator and the crossover estimator

We begin by noting that:

$$\begin{aligned} A_j &= \frac{1}{j-1} \sum_{i=1}^{j-1} (\mu + \alpha_i + \theta_j + \beta + \epsilon_{i,j}) - \frac{1}{J-j} \sum_{i=j}^{J-1} (\mu + \alpha_i + \theta_j + \epsilon_{i,j}) \\ &= \mu + \theta_j + \beta + \frac{1}{j-1} \sum_{i=1}^{j-1} (\alpha_i + \epsilon_{i,j}) - (\mu + \theta_j) - \frac{1}{J-j} \sum_{i=j}^{J-1} (\alpha_i + \epsilon_{i,j}) \\ &= \beta + \frac{1}{j-1} \sum_{i=1}^{j-1} (\alpha_i + \epsilon_{i,j}) - \frac{1}{J-j} \sum_{i=j}^{J-1} (\alpha_i + \epsilon_{i,j}). \end{aligned}$$

Hence, dropping the β terms in A_j and B_j since they do not contribute any variability:

$$\begin{aligned} \mathrm{Cov}(A_j,\ B_j) &= \frac{1}{j-1} \sum_{i=1}^{j-1} \mathrm{Cov}\left(\alpha_i + \epsilon_{i,j},\ \epsilon_{j-1,j} - \epsilon_{j-1,j-1} - \frac{1}{J-j} \sum_{\ell=j}^{J-1} (\epsilon_{\ell,j} - \epsilon_{\ell,j-1}) \right) \\ &\quad - \frac{1}{J-j} \sum_{i=j}^{J-1} \mathrm{Cov}\left(\alpha_i + \epsilon_{i,j},\ \epsilon_{j-1,j} - \epsilon_{j-1,j-1} - \frac{1}{J-j} \sum_{\ell=j}^{J-1} (\epsilon_{\ell,j} - \epsilon_{\ell,j-1}) \right) \\ &= \frac{1}{j-1} \sigma^2_{j-1,j} + \frac{1}{(J-j)^2} \sum_{i=j}^{J-1} \sigma^2_{i,j}. \end{aligned}$$

$$\text{Cov}(A_j,\ B_k) = \frac{1}{j-1}\sum_{i=1}^{j-1}\text{Cov}\left(\alpha_i + \epsilon_{i,j},\ \epsilon_{k-1,k} - \epsilon_{k-1,k-1} - \frac{1}{J-k}\sum_{\ell=k}^{J-1}(\epsilon_{\ell,k} - \epsilon_{\ell,k-1})\right) - \frac{1}{J-j}\sum_{i=j}^{J-1}\text{Cov}\left(\alpha_i + \epsilon_{i,j},\ \epsilon_{k-1,k} - \epsilon_{k-1,k-1} - \frac{1}{J-k}\sum_{\ell=k}^{J-1}(\epsilon_{\ell,k} - \epsilon_{\ell,k-1})\right).$$

For $j = k-1$:

$$\text{Cov}(A_j,\ B_k) = \frac{\sigma^2_{j,j}}{J-j} - \frac{1}{(J-j)(J-j-1)}\sum_{\ell=j+1}^{J-1}\sigma^2_{\ell,j}.$$

For $j < k-1$ or $j > k$, $\text{Cov}(A_j,\ B_k) = 0$.

Therefore:

$$\begin{aligned}\text{Cov}\left(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}\right) &= \text{Cov}\left(\sum_{j=2}^{J-1} v_j A_j,\ \sum_{k=2}^{J-1} w_k B_k\right) = \sum_{j=2}^{J-1}\sum_{k=2}^{J-1} v_j w_k \text{Cov}(A_j,\ B_k) \\ &= \sum_{j=2}^{J-1} v_j w_j\left[\frac{1}{j-1}\sigma^2_{j-1,j} + \frac{1}{(J-j)^2}\sum_{i=j}^{J-1}\sigma^2_{i,j}\right] + \sum_{j=2}^{J-2} v_j w_{j+1}\left[\frac{\sigma^2_{j,j}}{J-j} - \frac{1}{(J-j)(J-j-1)}\sum_{i=j+1}^{J-1}\sigma^2_{i,j}\right].\end{aligned} \tag{B6}$$

Note that, like the variance of the crossover estimator, the covariance here does not depend on τ^2.

B.5 Conditions for the ensemble estimator to have lower variance

If we moreover assume that $\sigma^2_{ij} = \sigma^2$ for all i, j, then:

$$\begin{aligned}\text{Var}(\hat{\beta}^{NPWP}) &= (\tau^2 + \sigma^2)\sum_{j=2}^{J-1} v_j^2\left[\frac{J-1}{(J-j)(j-1)}\right] + 2\tau^2\sum_{j=2}^{J-2}\sum_{k=j+1}^{J-1} v_j v_k \frac{J-1}{(J-j)(k-1)}, \\ &= (\tau^2 + \sigma^2)\sum_{j=2}^{J-1} v_j^2\left(\frac{1}{j-1} + \frac{1}{J-j}\right) + 2\tau^2\sum_{j=2}^{J-2}\sum_{k=j+1}^{J-1} v_j v_k \frac{J-1}{(J-j)(k-1)}.\end{aligned}$$

$$\text{Var}(\hat{\beta}^{CO}) = 2\sum_{j=2}^{J-1} w_j^2\sigma^2\left(1 + \frac{1}{J-j}\right).$$

$$\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) = \sum_{j=2}^{J-1} v_j w_j \sigma^2\left(\frac{1}{j-1} + \frac{1}{J-j}\right) < \sum_{j=2}^{J-1} v_j w_j \sigma^2\left(1 + \frac{1}{J-j}\right).$$

$$\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) = \sum_{j=2}^{J-1} v_j w_j \sigma^2\left(\frac{1}{j-1} + \frac{1}{J-j}\right) < \sigma^2\sum_{j=2}^{J-1} v_j w_j\left(\frac{1}{j-1} + \frac{1}{J-j}\right) + 2\tau^2\sum_{j=2}^{J-2}\sum_{k=j+1}^{J-1} v_j v_k\frac{J-1}{(J-j)(k-1)}.$$

If $v_j \leq 2w_j$ for all $j = 1, \ldots, n$, then $\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) < \text{Var}(\hat{\beta}^{CO})$. If $\sigma^2 w_j < (\tau^2 + \sigma^2)v_j$ for all $j = 1, \ldots, n$, then $\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) < \text{Var}(\hat{\beta}^{NPWP})$.

Now consider the ensemble estimator $\hat{\beta}^{ENS} = \frac{1}{2}\hat{\beta}^{NPWP} + \frac{1}{2}\hat{\beta}^{CO}$. Then:

$$\text{Var}(\hat{\beta}^{ENS}) = \frac{1}{4}\text{Var}(\hat{\beta}^{NPWP}) + \frac{1}{4}\text{Var}(\hat{\beta}^{CO}) + \frac{1}{2}\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}).$$

So a necessary and sufficient condition for $\text{Var}(\hat{\beta}^{ENS}) < \min\left(\text{Var}(\hat{\beta}^{NPWP}),\ \text{Var}(\hat{\beta}^{CO})\right)$ is:

$$2\,\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) < 3\min\left(\text{Var}(\hat{\beta}^{NPWP}),\ \text{Var}(\hat{\beta}^{CO})\right) - \max\left(\text{Var}(\hat{\beta}^{NPWP}),\ \text{Var}(\hat{\beta}^{CO})\right).$$

For $\tau^2 = 0$, $\text{Var}(\hat{\beta}^{NPWP} < \hat{\beta}^{CO})$ for some values of v_j, w_j, and σ^2 (e.g., for $v_j = w_j$ for all σ^2). Take such a set of values of v_j, w_j, and σ^2. Then, since $\text{Var}(\hat{\beta}^{NPWP})$ is an increasing function of τ^2 and $\text{Var}(\hat{\beta}^{CO})$ is independent of τ^2, there is a τ^2 such that $\text{Var}(\hat{\beta}^{NPWP}) = \text{Var}(\hat{\beta}^{CO})$. So the condition simplifies to:

$$2\,\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) < 2\,\text{Var}(\hat{\beta}^{CO}).$$

For $v_j \leq 2w_j$ for all $j = 2, \ldots, J-1$, $\text{Cov}(\hat{\beta}^{NPWP},\ \hat{\beta}^{CO}) < \text{Var}(\hat{\beta}^{CO})$. Hence, in this setting, the ensemble estimator has a lower variance than either single estimator.

This suggests that when τ^2 is low enough that the crossover estimator and the non-parametric within-period estimator to have similar variances, and the v_j and w_j are close enough for the covariance to be lower than either of these variances, the ensemble method has lower variance than either single estimator.