Panel Data and Experimental Design

Fiona Burlig, Louis Preonas, and Matt Woerman

August 26, 2019

Appendix: For online publication

Contents

A Derivations and proofs

We are the first to derive analytic power calculation formulas for panel data models under non-i.i.d. error structures, accounting for (i) heterogeneous error structures across panel units, and (ii) random time shocks. In this section, we derive power calculation analytics for cross-sectional, difference-in-differences (DD), collapsed data, and analysis of covariance (ANCOVA) estimators. We present the resulting equations in Section 2 of the main text.

We first re-derive well-known power calculation formulas, for both a cross-sectional experiment and a panel experiment that applies the difference-in-differences estimator, under the assumption that error terms are uncorrelated. We then relax this assumption to consider the difference-in-differences estimator applied to a panel experiment in which the error structure of the data exhibits an arbitrary form of serial correlation. This resembles the most general model derived by Frison and Pocock (1992), while incorporating two features characteristic of panel data in economics: (i) error structures that vary across individual panel units; and (ii) random time series shocks that are common to all panel units. We then consider a collapsed data research design, and demonstrate that collapsing to a single pre/post observation for each panel unit *ex post* does not obviate the need to account for non-constant within-unit serial correlation *ex ante*. We also conduct a parallel series of derivations using the ANCOVA estimator, which underscores the strong assumptions required to achieve analytical tractability in this regression model.

Next, we prove Lemma 1 presented in the main text. We also prove an analogous Lemma A1 for a cross-sectional model. These lemmas show that the above power calculation formulas can be applied even in the presence of cross-sectional correlations, so long as treatment is randomly assigned at the unit level. In particular, Lemma 1 states that the variance estimator we derive, which accounts for non-constant serial correlation, gives an unbiased estimate of the *ex ante* expected variance under unit-level randomization, even when cross-sectional correlations exist. In other words, our newly derived power calculation formula can be applied to any panel experiment setting, regardless of the true error structure of the data, so long as treatment is randomly assigned to units.

Finally, we demonstrate that replacing unit and/or time fixed effects with a Treat_i and/or Post_t dummy, respectively, does not alter the statistical power of the DD estimator, as long as researchers implement the CRVE *ex post*. This analysis complements our simulation results presented in Section 2.2.2 of the main text.

A.1 Cross sectional experiments

Model There are J units randomly assigned a treatment status D_i, with proportion P in treatment ($D_i = 1$) and proportion $(1 - P)$ in control ($D_i = 0$). The units are indexed so $i \in [1, PJ]$ is treated and $j \in [PJ + 1, J]$ is a control. We make standard assumptions for randomized trials:

Assumption A1 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_i = \beta + \tau D_i + \varepsilon_i \tag{A1}$$

where Y_i is the outcome of interest for unit i; β is the expected outcome of non-treated units; τ is the treatment effect which is homogeneous across all units; D_i is a treatment indicator; and ε_i is an idiosyncratic error term distributed i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Assumption A2 (Strict exogeneity)**.** $\mathrm{E}[\varepsilon_i \mid \mathbf{X}] = 0$*, where* $\mathbf{X} = [\mathbf{1}\ \mathbf{D}]$*. In practice, this follows from random assignment of* D_i*.*

Coefficient estimate The coefficient estimates from an OLS regression are

$$\begin{aligned}
\begin{pmatrix} \widehat{\beta} \\ \widehat{\tau} \end{pmatrix} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \begin{pmatrix} J & PJ \\ PJ & PJ \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{J} Y_i \\ \sum_{i=1}^{J} D_i Y_i \end{pmatrix} \\
&= \frac{1}{P(1-P)J^2} \begin{pmatrix} PJ\left(\sum_{i=1}^{J} Y_i - \sum_{i=1}^{J} D_i Y_i\right) \\ J\left(\sum_{i=1}^{J} D_i Y_i - P\sum_{i=1}^{J} Y_i\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{(1-P)J}\sum_{i=PJ+1}^{J} Y_i \\ \frac{1}{PJ}\sum_{i=1}^{PJ} Y_i - \frac{1}{(1-P)J}\sum_{i=PJ+1}^{J} Y_i \end{pmatrix}
\end{aligned}$$

Defining the mean outcome in the treatment and control groups, respectively, as

$$\begin{aligned}
\overline{Y}_T &= \frac{1}{PJ}\sum_{i=1}^{PJ} Y_i \\
\overline{Y}_C &= \frac{1}{(1-P)J}\sum_{i=PJ+1}^{J} Y_i
\end{aligned}$$

gives coefficient estimates of

$$\begin{aligned}
\widehat{\beta} &= \overline{Y}_C \\
\widehat{\tau} &= \overline{Y}_T - \overline{Y}_C
\end{aligned}$$

Variance of coefficient estimate The variance of the estimate of the treatment effect, $\widehat{\tau}$, is

$$\begin{aligned}
\operatorname{Var}(\widehat{\tau} \mid \mathbf{X}) &= \operatorname{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{Y}_T, \overline{Y}_C \mid \mathbf{X}\right) \\
&= \operatorname{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) \qquad \text{(A2)}
\end{aligned}$$

The first term of Equation (A2) is

$$\begin{aligned}
\operatorname{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) &= \operatorname{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} Y_i \mid \mathbf{X}\right) \\
&= \frac{\sigma_\varepsilon^2}{PJ} \qquad \text{(A3)}
\end{aligned}$$

Similarly, the second term of Equation (A2) is

$$\operatorname{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) = \frac{\sigma_\varepsilon^2}{(1-P)J} \qquad \text{(A4)}$$

Substituting Equations (A3) and (A4) into Equation (A2) gives

$$\begin{aligned}\text{Var}(\widehat{\tau} \mid \mathbf{X}) &= \frac{\sigma_\varepsilon^2}{PJ} + \frac{\sigma_\varepsilon^2}{(1-P)J} \\ &= \frac{\sigma_\varepsilon^2}{P(1-P)J}\end{aligned} \tag{A5}$$

The variance estimator produced by an OLS regression of Equation (A1) is an unbiased estimator of this variance.

Minimum detectable effect The minimum detectable effect (MDE), or the smallest treatment effect we have the power to detect, is

$$\begin{aligned}MDE &= \left(t_{1-\kappa}^{J-2} + t_{\alpha/2}^{J-2}\right)\sqrt{\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\ &= \left(t_{1-\kappa}^{J-2} + t_{\alpha/2}^{J-2}\right)\sqrt{\frac{\sigma_\varepsilon^2}{P(1-P)J}}\end{aligned} \tag{A6}$$

where κ is the power of the hypothesis test, α is the significance level, and the critical values are drawn from t-distributions with $J-2$ degrees of freedom.

A.2 Panel experiments

A.2.1 Independent error structure

Model In this model, P proportion of the J units are again randomized into treatment. The researcher collects the outcome Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For units in the treatment group, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for units in the control group, $D_{it} = 0$ in all $(m+r)$ periods.

Assumption A3 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it} \tag{A7}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed i.i.d. $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption A4 (Strict exogeneity)**.** $\text{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J-1$ *unit dummies, and* $(m+r)-1$ *time dummies. This again follows from random assignment of* D_{it}*.*

Assumption A5 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Coefficient estimate The treatment effect, τ, can be estimated by OLS with unit and time fixed effects. In a balanced panel, this is equivalent to de-meaning at both the unit and time levels. Define

$$\ddot{Y}_{it} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{\overline{Y}} \tag{A8}$$

$$\ddot{D}_{it} = D_{it} - \overline{D}_i - \overline{D}_t + \overline{\overline{D}} \tag{A9}$$

$$\ddot{\omega}_{it} = \omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}} \tag{A10}$$

where

$$\overline{Y}_i = \frac{1}{m+r} \sum_{t=-m+1}^{r} Y_{it}$$

$$\overline{Y}_t = \frac{1}{J} \sum_{i=1}^{J} Y_{it}$$

$$\overline{\overline{Y}} = \frac{1}{J(m+r)} \sum_{t=-m+1}^{r} \sum_{i=1}^{J} Y_{it}$$

with $\overline{D}_i$, $\overline{D}_t$, $\overline{\overline{D}}$, $\overline{\omega}_i$, $\overline{\omega}_t$, and $\overline{\overline{\omega}}$ defined analogously. Substituting Equations (A8)–(A10) into Equation (A7) and simplifying gives the de-meaned DGP

$$\ddot{Y}_{it} = \tau \ddot{D}_{it} + \ddot{\omega}_{it}$$

The estimate of the treatment effect is

$$\begin{aligned}
\widehat{\tau} &= (\ddot{\mathbf{D}}'\ddot{\mathbf{D}})^{-1}\ddot{\mathbf{D}}'\ddot{\mathbf{Y}} \\
&= \left(\sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it}^2 \right)^{-1} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it}\ddot{Y}_{it} \\
&= \frac{m+r}{P(1-P)Jmr} \left[\sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} Y_{it} - \sum_{i=1}^{J} \overline{Y}_i \sum_{t=-m+1}^{r} \ddot{D}_{it} - \sum_{t=-m+1}^{r} \overline{Y}_t \sum_{i=1}^{J} \ddot{D}_{it} + \overline{Y} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} \right] \\
&= \frac{m+r}{P(1-P)Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} Y_{it} \\
&= \frac{m+r}{P(1-P)Jmr} \left[\sum_{i=1}^{PJ} \left(\frac{(1-P)m}{m+r} \sum_{t=1}^{r} Y_{it} - \frac{(1-P)r}{m+r} \sum_{t=-m+1}^{0} Y_{it} \right) \right. \\
&\qquad \left. + \sum_{i=PJ+1}^{J} \left(\frac{Pr}{m+r} \sum_{t=-m+1}^{0} Y_{it} - \frac{Pm}{m+r} \sum_{t=1}^{r} Y_{it} \right) \right] \\
&= \left(\overline{Y}_T^A - \overline{Y}_T^B \right) - \left(\overline{Y}_C^A - \overline{Y}_C^B \right)
\end{aligned} \tag{A11}$$

where

$$\overline{Y}_T^A = \frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it}$$

$$\overline{Y}_T^B = \frac{1}{PJm} \sum_{i=1}^{PJ} \sum_{t=-m+1}^{0} Y_{it}$$

$$\overline{Y}_C^A = \frac{1}{(1-P)Jr} \sum_{i=PJ+1}^{J} \sum_{t=1}^{r} Y_{it}$$

$$\overline{Y}_C^B = \frac{1}{(1-P)Jm} \sum_{i=PJ+1}^{J} \sum_{t=-m+1}^{0} Y_{it}$$

Variance of coefficient estimate The variance of the estimate of the treatment effect, $\widehat{\tau}$, is

$$\begin{aligned}
\operatorname{Var}(\widehat{\tau} \mid \mathbf{X}) &= \operatorname{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) \\
&\quad - 2\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) + 2\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^B \mid \mathbf{X}\right) \\
&\quad + 2\operatorname{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^A \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) \\
&= \operatorname{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) \\
&\quad - 2\Bigg[\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) + \operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) \\
&\qquad + \operatorname{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) + \operatorname{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right)\Bigg]
\end{aligned} \tag{A12}$$

The first term of Equation (A12) is

$$\begin{aligned}
\operatorname{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) &= \operatorname{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it} \mid \mathbf{X}\right) \\
&= \operatorname{Var}\left(\frac{1}{PJ} \sum_{i=1}^{PJ} \upsilon_i \mid \mathbf{X}\right) + \operatorname{Var}\left(\frac{1}{r} \sum_{t=1}^{r} \delta_t \mid \mathbf{X}\right) + \operatorname{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} \omega_{it} \mid \mathbf{X}\right) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right)
\end{aligned} \tag{A13}$$

Similarly, the remaining variance terms of Equation (A12) are

$$\operatorname{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) = \frac{1}{PJm}\left(m\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right) \tag{A14}$$

$$\operatorname{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) = \frac{1}{(1-P)Jr}\left(r\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2\right) \tag{A15}$$

$$\operatorname{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)Jm}\left(m\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2\right) \tag{A16}$$

The first covariance component of Equation (A12) is

$$\begin{aligned}\text{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) &= \text{E}\left[\overline{Y}_T^A \overline{Y}_T^B \mid \mathbf{X}\right] - \text{E}\left[\overline{Y}_T^A \mid \mathbf{X}\right] \text{E}\left[\overline{Y}_T^B \mid \mathbf{X}\right] \\ &= \beta(\beta+\tau) + \frac{1}{PJ}\text{E}\left[\upsilon_i^2 \mid \mathbf{X}\right] - \beta(\beta+\tau) \\ &= \frac{\sigma_\upsilon^2}{PJ} \end{aligned} \tag{A17}$$

Similarly, the remaining covariance terms of Equation (A12) are

$$\text{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{r} \tag{A18}$$

$$\text{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{m} \tag{A19}$$

$$\text{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{\sigma_\upsilon^2}{(1-P)J} \tag{A20}$$

Substituting Equations (A13)–(A20) into Equation (A12) gives

$$\begin{aligned}\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= 2\frac{\sigma_\upsilon^2}{P(1-P)J} + 2\frac{(m+r)\sigma_\delta^2}{mr} + \frac{(m+r)\sigma_\omega^2}{P(1-P)Jmr} - 2\left[\frac{\sigma_\upsilon^2}{P(1-P)J} + \frac{(m+r)\sigma_\delta^2}{mr}\right] \\ &= \left(\frac{\sigma_\omega^2}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right) \end{aligned} \tag{A21}$$

The variance estimator produced by an OLS regression of Equation (A7) is an unbiased estimator of this variance.

Minimum detectable effect The MDE is

$$\begin{aligned} MDE &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\ &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\left(\frac{\sigma_\omega^2}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)} \end{aligned}$$

This is the standard Frison and Pocock (1992) result (page 1693), also referenced by McKenzie (2012), and is shown as Equation (3) in the main text. We assume J degrees of freedom to be consistent with the assumptions of the CRVE; alternatively, $J(m+r) - (J+m+r)$ degrees of freedom would be consistent with the assumptions of OLS standard errors. Note that this model differs slightly from Frison and Pocock (1992) and McKenzie (2012), which both assume $\sigma_\delta^2 = 0$.

A.2.2 Serially correlated error structure

Model There are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption A6 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it} \tag{A22}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption A7 (Strict exogeneity)**.** $\mathrm{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*, $J-1$ unit dummies, and $(m+r)-1$ time dummies. This again follows from random assignment of D_{it}.*

Assumption A8 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption A9 (Independence across units)**.** $\mathrm{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0, \ \forall \ i \neq j, \ \forall \ t, s.$

Assumption A10 (Symmetric covariance structures)**.** *Define:*

$$\psi^B \equiv \frac{2}{Jm(m-1)} \sum_{i=1}^{J} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^A \equiv \frac{2}{Jr(r-1)} \sum_{i=1}^{J} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^X \equiv \frac{1}{Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of the same unit, respectively. Define ψ_T^B, ψ_T^A, and ψ_C^X analogously, where we consider only the PJ treated units; also define ψ_C^B, ψ_C^A, and ψ_C^X analogously, where we consider only the $(1-P)J$ control units. Using these definitions, assume that $\psi^B = \psi_T^B = \psi_C^B$; $\psi^A = \psi_T^A = \psi_C^A$; and $\psi^X = \psi_T^X = \psi_C^X$.[1]

Coefficient estimate The treatment effect estimator is the same as Equation (A11):

$$\widehat{\tau} = \left(\overline{Y}_T^A - \overline{Y}_T^B\right) - \left(\overline{Y}_C^A - \overline{Y}_C^B\right)$$

1. We choose the letters "B" to indicate the Before-treatment period, and "A" to indicate the After-treatment period. We index the m pre-treatment periods $\{-m+1, \ldots, 0\}$, and the r post-treatment periods $\{1, \ldots, r\}$. In a randomized setting, $\mathrm{E}\left[\psi^B\right] = \mathrm{E}\left[\psi_T^B\right] = \mathrm{E}\left[\psi_C^B\right]$, $\mathrm{E}\left[\psi^A\right] = \mathrm{E}\left[\psi_T^A\right] = \mathrm{E}\left[\psi_C^A\right]$, and $\mathrm{E}\left[\psi^X\right] = \mathrm{E}\left[\psi_T^X\right] = \mathrm{E}\left[\psi_C^X\right]$, making this a reasonable assumption *ex ante*. However, it is possible for treatment to alter the covariance structure of treated units only.

Variance of coefficient estimate With a serially correlated error structure, Equation (A12) is still correct by Assumption A9, but Equations (A13)–(A20) no longer hold. With serially correlated errors, the first term of Equation (A12) is

$$\begin{aligned}
\operatorname{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) &= \operatorname{Var}\left(\frac{1}{PJr}\sum_{i=1}^{PJ}\sum_{t=1}^{r} Y_{it} \mid \mathbf{X}\right) \\
&= \operatorname{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} \upsilon_i \mid \mathbf{X}\right) + \operatorname{Var}\left(\frac{1}{r}\sum_{t=1}^{r} \delta_t \mid \mathbf{X}\right) + \operatorname{Var}\left(\frac{1}{PJr}\sum_{i=1}^{PJ}\sum_{t=1}^{r} \omega_{it} \mid \mathbf{X}\right) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right) + \frac{2}{(PJr)^2}\sum_{i=1}^{PJ}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r} \operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi^A\right) \qquad \text{(A23)}
\end{aligned}$$

Similarly, with serial correlation, the remaining variance terms of Equation (A12) are

$$\operatorname{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) = \frac{1}{PJm}\left(m\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi^B\right) \qquad \text{(A24)}$$

$$\operatorname{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) = \frac{1}{(1-P)Jr}\left(r\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi^A\right) \qquad \text{(A25)}$$

$$\operatorname{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)Jm}\left(m\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi^B\right) \qquad \text{(A26)}$$

With serial correlation, the first covariance component of Equation (A12) is

$$\begin{aligned}
\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) &= \operatorname{E}\left[\overline{Y}_T^A \overline{Y}_T^B \mid \mathbf{X}\right] - \operatorname{E}\left[\overline{Y}_T^A \mid \mathbf{X}\right]\operatorname{E}\left[\overline{Y}_T^B \mid \mathbf{X}\right] \\
&= \beta(\beta+\tau) + \frac{1}{PJ}\operatorname{E}\left[\upsilon_i^2 \mid \mathbf{X}\right] + \frac{1}{(PJ)^2 mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r} \operatorname{E}[\omega_{it}\omega_{is} \mid \mathbf{X}] - \beta(\beta+\tau) \\
&= \frac{\sigma_\upsilon^2}{PJ} + \frac{1}{(PJ)^2 mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r} \operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) \\
&= \frac{1}{PJ}\left(\sigma_\upsilon^2 + \psi^X\right) \qquad \text{(A27)}
\end{aligned}$$

Similarly, with serial correlation, the remaining covariance terms of Equation (A12) are

$$\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{r} \qquad \text{(A28)}$$

$$\operatorname{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{m} \qquad \text{(A29)}$$

$$\operatorname{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)J}\left(\sigma_\upsilon^2 + \psi^X\right) \qquad \text{(A30)}$$

Substituting Equations (A23)–(A30) into Equation (A12) and simplifying gives

$$\begin{aligned}
\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= \frac{1}{PJr}\left(r\sigma_{\upsilon}^2 + PJ\sigma_{\delta}^2 + \sigma_{\omega}^2 + (r-1)\psi^A\right) \\
&\quad + \frac{1}{PJm}\left(m\sigma_{\upsilon}^2 + PJ\sigma_{\delta}^2 + \sigma_{\omega}^2 + (m-1)\psi^B\right) \\
&\quad + \frac{1}{(1-P)Jr}\left(r\sigma_{\upsilon}^2 + (1-P)J\sigma_{\delta}^2 + \sigma_{\omega}^2 + (r-1)\psi^A\right) \\
&\quad + \frac{1}{(1-P)Jm}\left(m\sigma_{\upsilon}^2 + (1-P)J\sigma_{\delta}^2 + \sigma_{\omega}^2 + (m-1)\psi^B\right) \\
&\quad - 2\left[\frac{1}{PJ}\left(\sigma_{\upsilon}^2 + \psi^X\right) + \frac{\sigma_{\delta}^2}{r} + \frac{\sigma_{\delta}^2}{m} + \frac{1}{(1-P)J}\left(\sigma_{\upsilon}^2 + \psi^X\right)\right] \\
&= \left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right] \quad \text{(A31)}
\end{aligned}$$

Equation (A31) differs from Equation (A21), the variance of $\widehat{\tau}$ that is estimated by an OLS regression, due to the presence of the ψ terms that describe the serially correlated error structure. This error structure alters the true variance of the treatment effect estimator such that the OLS estimator of the variance is not correct. The cluster-robust variance estimator must be used for correct inference.

Minimum detectable effect With serial correlation, the MDE is

$$\begin{aligned}
MDE &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\
&= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}
\end{aligned}$$

This is the serial-correlation-robust (SCR) power calculation formula, found in Equation (2) in the main text. Note that ψ^B, ψ^A, and ψ^X are likely to depend on the length of the pre- and post-treatment periods, m and r; serial correlation often diminishes as time periods become further apart, so larger values of m and r will result in less correlation on average and smaller parameter values. These parameters do not depend in a systematic way on the number of experimental units, J, however.

A.2.3 Serially and cross-sectionally correlated error structure

Randomization at the unit level justifies Assumptions A9 and A10, but we can also characterize the full variance without this random assignment assumption. This highlights the type of correlation structures that a researcher might face when wanting to perform a power calculation for a quasi-experimental design.

Model There are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption A11 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption A12 (Strict exogeneity)**.** $\mathrm{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J-1$ *unit dummies, and* $(m+r)-1$ *time dummies. This again follows from random assignment of* D_{it}*.*

Assumption A13 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption A14 (Independence across units at different times)**.** $\mathrm{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ t \neq s.$

Define ψ_i to be the average serial correlation parameters previously defined in Appendix A.2.2, with the subscript i denoting the correlation is within-unit serial correlation. Also define ψ_t to be the comparable parameters characterizing cross-sectional correlations, with the subscript t denoting the correlation is cross-sectional within a time period. For example, the average cross-sectional covariance among the treated group post-treatment is

$$\psi_{t,T}^A = \frac{2}{PJ(PJ-1)r} \sum_{i=1}^{PJ-1} \sum_{j=i+1}^{PJ} \sum_{t=1}^{r} \mathrm{Cov}(\omega_{it}, \omega_{jt} \mid \mathbf{X})$$

Coefficient estimate The treatment effect estimator is the same as Equation (A11):

$$\widehat{\tau} = \left(\overline{Y}_T^A - \overline{Y}_T^B\right) - \left(\overline{Y}_C^A - \overline{Y}_C^B\right)$$

Variance of coefficient estimate As with the serially correlated error structure in Appendix A.2.2, Equation (A12) is still correct, in this case by Assumption A14. However, neither Equations (A13)–(A20) nor Equations (A23)–(A30) hold. With these arbitrary correlations, the first term of Equation (A12) is

$$\begin{aligned}
\mathrm{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) &= \mathrm{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it} \mid \mathbf{X}\right) \\
&= \mathrm{Var}\left(\frac{1}{PJ} \sum_{i=1}^{PJ} \upsilon_i \mid \mathbf{X}\right) + \mathrm{Var}\left(\frac{1}{r} \sum_{t=1}^{r} \delta_t \mid \mathbf{X}\right) + \mathrm{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} \omega_{it} \mid \mathbf{X}\right) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right) + \frac{2}{(PJr)^2} \sum_{i=1}^{PJ} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \mathrm{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) \\
&\quad + \frac{2}{(PJr)^2} \sum_{t=1}^{r} \sum_{i=1}^{PJ-1} \sum_{j=i+1}^{PJ} \mathrm{Cov}(\omega_{it}, \omega_{jt} \mid \mathbf{X})
\end{aligned}$$

$$= \frac{1}{PJr}\left(r\sigma_v^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi_{i,T}^A + (PJ-1)\psi_{t,T}^A\right) \tag{A32}$$

Similarly, with arbitrary correlations, the remaining variance terms of Equation (A12) are

$$\text{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) = \tfrac{1}{PJm}\left(m\sigma_v^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi_{i,T}^B + (PJ-1)\psi_{t,T}^B\right) \tag{A33}$$

$$\text{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) = \tfrac{1}{(1-P)Jr}\left(r\sigma_v^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi_{i,C}^A + ((1-P)J-1)\psi_{t,C}^A\right) \tag{A34}$$

$$\text{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) = \tfrac{1}{(1-P)Jm}\left(m\sigma_v^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi_{i,C}^B + ((1-P)J-1)\psi_{t,C}^B\right) \tag{A35}$$

With arbitrary correlations, the first covariance component of Equation (A12) is

$$\begin{aligned}
\text{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) &= \text{E}\left[\overline{Y}_T^A\overline{Y}_T^B \mid \mathbf{X}\right] - \text{E}\left[\overline{Y}_T^A \mid \mathbf{X}\right]\text{E}\left[\overline{Y}_T^B \mid \mathbf{X}\right] \\
&= \beta(\beta+\tau) + \frac{1}{PJ}\text{E}\left[\upsilon_i^2 \mid \mathbf{X}\right] + \frac{1}{(PJ)^2mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{E}[\omega_{it}\omega_{is} \mid \mathbf{X}] - \beta(\beta+\tau) \\
&= \frac{\sigma_v^2}{PJ} + \frac{1}{(PJ)^2mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) \\
&= \frac{1}{PJ}\left(\sigma_v^2 + \psi_{i,T}^X\right)
\end{aligned} \tag{A36}$$

Similarly, with arbitrary correlations, the remaining covariance terms of Equation (A12) are

$$\text{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) = \frac{1}{r}\left(\sigma_\delta^2 + \psi_{t,X}^A\right) \tag{A37}$$

$$\text{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{m}\left(\sigma_\delta^2 + \psi_{t,X}^B\right) \tag{A38}$$

$$\text{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)J}\left(\sigma_v^2 + \psi_{i,C}^X\right) \tag{A39}$$

Substituting Equations (A32)–(A39) into Equation (A12) and simplifying gives

$$\begin{aligned}
\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) = &\frac{1}{PJr}\left(r\sigma_v^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi_{i,T}^A + (PJ-1)\psi_{t,T}^A\right) \\
&+ \frac{1}{PJm}\left(m\sigma_v^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi_{i,T}^B + (PJ-1)\psi_{t,T}^B\right) \\
&+ \frac{1}{(1-P)Jr}\left(r\sigma_v^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi_{i,C}^A + ((1-P)J-1)\psi_{t,C}^A\right) \\
&+ \frac{1}{(1-P)Jm}\left(m\sigma_v^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi_{i,C}^B + ((1-P)J-1)\psi_{t,C}^B\right) \\
&- 2\left[\frac{1}{PJ}\left(\sigma_v^2 + \psi_{i,T}^X\right) + \frac{1}{r}\left(\sigma_\delta^2 + \psi_{t,X}^A\right) + \frac{1}{m}\left(\sigma_\delta^2 + \psi_{t,X}^B\right) + \frac{1}{(1-P)J}\left(\sigma_v^2 + \psi_{i,C}^X\right)\right] \\
= &\left(\frac{m+r}{P(1-P)Jmr}\right)\sigma_\omega^2 + \left(\frac{m-1}{PJm}\right)\psi_{i,T}^B + \left(\frac{PJ-1}{PJm}\right)\psi_{t,T}^B \\
&+ \left(\frac{r-1}{PJr}\right)\psi_{i,T}^A + \left(\frac{PJ-1}{PJr}\right)\psi_{t,T}^A + \left(\frac{m-1}{(1-P)Jm}\right)\psi_{i,C}^B
\end{aligned}$$

$$+\left(\frac{(1-P)J-1}{(1-P)Jm}\right)\psi^B_{t,C}+\left(\frac{r-1}{(1-P)Jr}\right)\psi^A_{i,C}+\left(\frac{(1-P)J-1}{(1-P)Jr}\right)\psi^A_{t,C}$$
$$-\frac{2}{PJ}\psi^X_{i,T}-\frac{2}{m}\psi^B_{t,X}-\frac{2}{r}\psi^A_{t,X}-\frac{2}{(1-P)J}\psi^X_{i,C}$$

Minimum detectable effect With arbitrary correlations, the MDE is

$$\begin{aligned}MDE &= \left(t^J_{1-\kappa}+t^J_{\alpha/2}\right)\sqrt{\text{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)}\\ &= \left(t^J_{1-\kappa}+t^J_{\alpha/2}\right)\left[\left(\frac{m+r}{P(1-P)Jmr}\right)\sigma^2_\omega+\left(\frac{m-1}{PJm}\right)\psi^B_{i,T}+\left(\frac{PJ-1}{PJm}\right)\psi^B_{t,T}\right.\\ &\quad+\left(\frac{r-1}{PJr}\right)\psi^A_{i,T}+\left(\frac{PJ-1}{PJr}\right)\psi^A_{t,T}+\left(\frac{m-1}{(1-P)Jm}\right)\psi^B_{i,C}\\ &\quad+\left(\frac{(1-P)J-1}{(1-P)Jm}\right)\psi^B_{t,C}+\left(\frac{r-1}{(1-P)Jr}\right)\psi^A_{i,C}+\left(\frac{(1-P)J-1}{(1-P)Jr}\right)\psi^A_{t,C}\\ &\quad\left.-\frac{2}{PJ}\psi^X_{i,T}-\frac{2}{m}\psi^B_{t,X}-\frac{2}{r}\psi^A_{t,X}-\frac{2}{(1-P)J}\psi^X_{i,C}\right]^{1/2}\end{aligned}\tag{A40}$$

Lemma 1 in Appendix A.3 shows that Equation (A40) is equal in expectation to the SCR power calculation formula when treatment is randomly assigned.

A.2.4 Collapsed dataset

Bertrand, Duflo, and Mullainathan (2004, henceforth BDM) suggest an alternative to the CRVE in order to achieve the correct false rejection rates in the presence of serial correlation: ignore the time-series structure of the data by averaging the pre-treatment data and the post-treatment data for each unit, then estimate a panel DD regression on this two-period collapsed dataset and apply the OLS variance estimator. While this does yield the desired false rejection rate, simply applying the McKenzie formula to a collapsed dataset will *not* yield the desired power.

Using the model from Appendix A.2.2 (under Assumptions A6–A10), suppose that prior to estimation, the data are collapsed to 1 pre- and 1 post-treatment observation per unit (to eliminate serial correlation, as suggested in BDM). The resulting DGP for the collapsed dataset is

$$Y^C_{ip}=\beta+\tau D^C_{ip}+\upsilon_i+\delta^C_p+\omega^C_{ip}\tag{A41}$$

where Y^C_{ip} is the average outcome for unit i for collapsed period p and the other variables are as defined in Appendix A.2.2 or are the collapsed analogs. Note that the τ in Equation (A41) is equivalent to that in Equation (A7). These models will yield the same estimate of the treatment effect, $\widehat{\tau}$, but different estimates of its variance in the presence of a (pre-collapsed) serially correlated error structure.

Equivalence of first-difference model BDM show that applying the OLS variance estimator to Equation (A41) achieves the correct false rejection rate. To see why this is the case, note that this collapsed model can alternatively be expressed as a first-difference model by subtracting each unit's collapsed pre-treatment data from its collapsed post-treatment data. Let ΔY^C_i be this difference

for the outcome of interest, which gives

$$\begin{aligned}\Delta Y_i^C &= Y_{iA}^C - Y_{iB}^C \\ &= \left(\beta + \tau D_{iA}^C + \upsilon_i + \delta_A^C + \omega_{iA}^C\right) - \left(\beta + \tau D_{iB}^C + \upsilon_i + \delta_B^C + \omega_{iB}^C\right) \\ &= \tau\left(D_{iA}^C - D_{iB}^C\right) + \left(\delta_A^C - \delta_B^C\right) + \left(\omega_{iA}^C - \omega_{iB}^C\right)\end{aligned}$$

Defining the other differences variables similarly gives the first-difference DGP of

$$\Delta Y_i^C = \tau \Delta D_i^C + \Delta\delta^C + \Delta\omega_i^C \tag{A42}$$

Equations (A41) and (A42) are isomorphic, so estimating these models yields not only the same estimate of the treatment effect, $\widehat{\tau}$, but also the same estimate of its variance. Note that the first-difference model, Equation (A42), is cross-sectional, so the error terms are i.i.d. and the model meets the assumptions of OLS.[2] As a result, the OLS variance estimator is unbiased for the first-difference model, as well as the isomorphic collapsed model of Equation (A41).

Power Although using a collapsed dataset yields the correct false rejection (or Type I error) rate, experiments will not be correctly powered if the McKenzie formula is applied to a collapsed dataset. To see this, first consider the MDE of an experiment based on the first-difference model of Equation (A42). This is a cross-sectional model, so applying Equations (A5) and (A6) yields:

$$\begin{aligned}\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) &= \frac{\sigma^2_{\Delta\omega^C}}{P(1-P)J} \\ MDE &= \left(t_{1-\kappa}^{J-2} + t_{\alpha/2}^{J-2}\right)\sqrt{\frac{\sigma^2_{\Delta\omega^C}}{P(1-P)J}}\end{aligned} \tag{A43}$$

where $\sigma^2_{\Delta\omega^C}$ is the variance of the error term in the collapsed, first-difference model. This variance can be expressed as a function of the parameters that define the error structure of the *collapsed* data

$$\begin{aligned}\sigma^2_{\Delta\omega^C} &= \text{Var}\left(\omega_{iA}^C - \omega_{iB}^C \mid \mathbf{X}\right) \\ &= \text{Var}\left(\omega_{iA}^C \mid \mathbf{X}\right) + \text{Var}\left(\omega_{iB}^C \mid \mathbf{X}\right) - 2\,\text{Cov}\left(\omega_{iA}^C, \omega_{iB}^C \mid \mathbf{X}\right) \\ &= 2\left(\sigma^2_{\omega^C} - \psi^X_{\omega^C}\right)\end{aligned} \tag{A44}$$

where $\sigma^2_{\omega^C}$ is the variance of the error term of the collapsed model, and $\psi^X_{\omega^C}$ is the average covariance between error terms for the same unit in the collapsed model. Substituting Equation (A44) into Equation (A43) gives the variance of $\hat{\tau}$ in terms of parameters of the collapsed data:

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \left(\frac{2}{P(1-P)J}\right)\left(\sigma^2_{\omega^C} - \psi^X_{\omega^C}\right) \tag{A45}$$

2. There are no cross-sectional error correlations due to Assumption A9, because randomization obviates the need to account for this kind of correlation (see Lemma 1).

This formula is equal to the variance of $\hat{\tau}$ from the SCR formula applied to collapsed data, where $m = r = 1$. Applying the McKenzie formula to collapsed data, however, gives the incorrect variance:

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \left(\frac{2}{P(1-P)J}\right)\sigma^2_{\omega^C} \tag{A46}$$

Equations (A45) and (A46) differ by the $\psi^X_{\omega^C}$ term that characterizes the covariance between the pre- and post-treatment error terms in the collapsed data. This term is omitted from the McKenzie formula that (incorrectly) assumes no serial correlation in the error structure. The error structure parameters of the collapsed data in Equation (A45) can also be expressed as functions of the parameters that define the error structure of the original, uncollapsed data

$$\begin{aligned}
\sigma^2_{\omega^C} &= \frac{1}{2}\text{Var}\left(\omega^C_{iA} \mid \mathbf{X}\right) + \frac{1}{2}\text{Var}\left(\omega^C_{iB} \mid \mathbf{X}\right) \\
&= \frac{1}{2}\text{Var}\left(\frac{1}{r}\sum_{t=1}^{r}\omega_{it} \mid \mathbf{X}\right) + \frac{1}{2}\text{Var}\left(\frac{1}{m}\sum_{s=-m+1}^{0}\omega_{is} \mid \mathbf{X}\right) \\
&= \frac{1}{2r}\left[\sigma^2_\omega + (r-1)\psi^A\right] + \frac{1}{2m}\left[\sigma^2_\omega + (m-1)\psi^B\right] \\
&= \frac{1}{2}\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{r-1}{r}\right)\psi^A + \left(\frac{m-1}{m}\right)\psi^B\right]
\end{aligned} \tag{A47}$$

$$\begin{aligned}
\psi^X_{\omega^C} &= \frac{1}{J}\sum_{i=1}^{J}\text{Cov}\left(\omega^C_{iA}, \omega^C_{iB} \mid \mathbf{X}\right) \\
&= \frac{1}{J}\sum_{i=1}^{J}\text{Cov}\left(\frac{1}{r}\sum_{t=1}^{r}\omega_{it}, \frac{1}{m}\sum_{s=-m+1}^{0}\omega_{is} \mid \mathbf{X}\right) \\
&= \frac{1}{Jmr}\sum_{i=1}^{J}\sum_{t=1}^{r}\sum_{s=-m+1}^{0}\text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) \\
&= \psi^X
\end{aligned} \tag{A48}$$

Substituting Equations (A47) and (A48) into Equation (A45) gives

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]$$

which is equivalent to the variance of $\hat{\tau}$ given in Equation (A31) that is a component of the SCR formula. Hence, the uncollapsed, collapsed, and first-difference models yield (virtually) equivalent MDEs when using the appropriate power calculation formula.[3] By contrast, the McKenzie formula ignores the between-period serial correlation that remains after collapsing serially correlated data.

3. The only difference between these three MDE calculations is the critical values $t^d_{1-\kappa}$ and $t^d_{\alpha/2}$. The uncollapsed model will apply the CRVE *ex post*, which implies $d = J$ degrees of freedom *ex ante*. The collapsed model will apply the OLS variance estimator *ex post* to a panel with $2J$ observations and $J + 2$ regressors, which implies $d = J - 2$ degrees of freedom *ex ante*. The first-difference model will apply the OLS variance estimator *ex post* to a cross-sectional specification with J observations and 2 regressors, which implies $d = J - 2$ degrees of freedom *ex ante*.

A.2.5 Analysis of covariance (ANCOVA)

Model There are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption A15 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a time-invariant unit effect distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption A16 (Strict exogeneity)**.** $\mathrm{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*, and* $J - 1$ *unit dummies. This again follows from random assignment of* D_{it}*.*

Assumption A17 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption A18 (Independence across units)**.** $\mathrm{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ \forall\ t, s.$

Assumption A19 (Uniform covariance structures)**.** *Define:*

$$\psi_i^B \equiv \frac{2}{m(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi_i^A \equiv \frac{2}{r(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi_i^X \equiv \frac{1}{mr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of unit i, respectively. Using these definitions, assume that $\psi^B = \psi_i^B$, $\psi^A = \psi_i^A$, and $\psi^X = \psi_i^X\ \forall i$.

In this section, we consider an analysis of covariance (ANCOVA) regression model:

$$Y_{it} = \alpha + \tau D_i + \theta \overline{Y}_i^B + \varepsilon_{it} \tag{A49}$$

where Y_{it}, τ, and D_i are defined as above; α is an intercept term, $\overline{Y}_i^B = \frac{1}{m} \sum_{t=-m+1}^{0} Y_{it}$ is the mean pre-period value of the outcome variable for unit i, and ε_{it} is an idiosyncratic error term. Note that this regression is estimated using only post-treatment outcomes as the dependent variable, with a unit's pre-treatment mean outcome entering as an independent variable.

Note that this following derivations necessitate two relatively strong assumptions, as we discuss in Section 2.2.2 of the main text. First, we assume zero time shocks in the data generating process, thereby omitting δ_t from Assumption A15. If we included time shocks as in Assumption A3, then the conditional expectations in Equations (A56) and (A57) would each depend on the error terms and pre-period means of all units in the experiment, making the analytic solution intractable.[4] Second, Assumption A19 forces the ψ parameters to be constant across all units. Without this simplifying assumption, Equations (A58)–(A60) would contain many additional terms describing the heterogeneity of ψ parameters, which has a negligible effect in experiments of resaonable size[5] and cannot be estimated from a pre-existing dataset.[6]

Coefficient estimate The coefficient estimates from an OLS regression are

$$\begin{aligned}
\begin{pmatrix} \hat{\alpha} \\ \hat{\tau} \\ \hat{\theta} \end{pmatrix} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \left(\begin{pmatrix} \mathbf{1}' \\ \mathbf{D}' \\ \overline{\mathbf{Y}}^{\mathbf{B}\prime} \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{D} & \overline{\mathbf{Y}}^{\mathbf{B}} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{1}' \\ \mathbf{D}' \\ \overline{\mathbf{Y}}^{\mathbf{B}\prime} \end{pmatrix} \mathbf{Y} \\
&= \underbrace{\begin{pmatrix} Jr & PJr & Jr\overline{Y}^{B} \\ PJr & PJr & PJr\overline{Y}^{B}_{T} \\ Jr\overline{Y}^{B} & PJr\overline{Y}^{B}_{T} & r\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 \end{pmatrix}^{-1}}_{\mathbf{M}} \begin{pmatrix} Jr\overline{Y}^{A} \\ PJr\overline{Y}^{A}_{T} \\ r\sum_{i=1}^{J}\overline{Y}^{A}_{i}\overline{Y}^{B}_{i} \end{pmatrix}
\end{aligned} \tag{A50}$$

The inverse of the $\mathbf{M}$ matrix is

$$\mathbf{M}^{-1} = \frac{PJ^2r^2}{\det(\mathbf{M})} \times \begin{pmatrix} \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - P(\overline{Y}^{B}_{T})^2 & \overline{Y}^{B}\overline{Y}^{B}_{T} - \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 & P\overline{Y}^{B}_{T} - \overline{Y}^{B} \\ \overline{Y}^{B}\overline{Y}^{B}_{T} - \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 & \frac{1}{PJ}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - \frac{1}{P}(\overline{Y}^{B})^2 & \overline{Y}^{B} - \overline{Y}^{B}_{T} \\ P\overline{Y}^{B}_{T} - \overline{Y}^{B} & \overline{Y}^{B} - \overline{Y}^{B}_{T} & 1 - P \end{pmatrix}$$

4. If we include time shocks and time fixed effects in our data generating process and regression model, respectively, the simplest analytic solution includes de-meaning by time period to remove these time shocks, similar to Appendix A.2.1. Then Equations (A56) and (A57) would instead describe the expectation of the product of the de-meaned error terms conditional on the de-meaned pre-period averages, which are a function of the error terms and pre-period means of all units in the experiment. These expressions increase in complexity in proportion to the number of units, making them analytically intractable for any reasonable number of experimental units while allowing for arbitrary serial correlation. An alternate solution method does not de-mean by time period to remove time shocks, but instead solves directly for all parameters in the model. This method requires inverting a $(m + r + 2) \times (m + r + 2)$ matrix, which is analytically intractable for an experiment of arbitrary length.

5. We test this second assumption by simulating datasets with uniform ψ parameters or heterogeneous ψ parameters and then performing Monte Carlo simulations as described in Appendix B.1. We find no appreciable difference in the simulated power of these different datasets.

6. A dataset contains only one realiztion of each error draw, so it is impossible to estimate a covariance between any two of these draws. We can, however, estimate an average covariance over all units by treating each unit's error as a draw from the same distribution, as described in more detail in Appendices D.1 and E.1.

and the determinant of $\mathbf{M}$ is

$$\begin{aligned}
\det(\mathbf{M}) &= PJ^3r^3\left[\frac{1}{J}\sum_{i=1}^{J}(\overline{Y}_i^B)^2 - P(\overline{Y}_T^B)^2\right] + P^2J^3r^3\left[\overline{Y}^B\overline{Y}_T^B - \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}_i^B)^2\right] \\
&\quad + PJ^3r^3\overline{Y}^B\left[P\overline{Y}_T^B - \overline{Y}^B\right] \\
&= P(1-P)J^2r^3\sum_{i=1}^{J}(\overline{Y}_i^B)^2 - P^2(1-P)(\overline{Y}_T^B)^2 - P(1-P)^2(\overline{Y}_C^B)^2 \\
&= P(1-P)J^2r^3\left[\sum_{i=1}^{PJ}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2\right]
\end{aligned}$$

Therefore, the coefficient estimates are

$$\begin{aligned}
\hat{\alpha} &= \frac{\overline{Y}_C^A\left[\sum_{i=1}^{J}(\overline{Y}_i^B)^2 - PJ(\overline{Y}_T^B)^2 - (1-P)J(\overline{Y}_C^B)^2\right]}{\sum_{i=1}^{J}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
&\quad - \frac{\overline{Y}_C^B\left[\sum_{i=1}^{J}\overline{Y}_i^B\overline{Y}_i^A - PJ\overline{Y}_T^B\overline{Y}_T^A - (1-P)J\overline{Y}_C^B\overline{Y}_T^A\right]}{\sum_{i=1}^{J}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
\hat{\tau} &= \frac{(\overline{Y}_T^A - \overline{Y}_C^A)\left[\sum_{i=1}^{J}(\overline{Y}_i^B)^2 - PJ(\overline{Y}_T^B)^2 - (1-P)J(\overline{Y}_C^B)^2\right]}{\sum_{i=1}^{J}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
&\quad - \frac{(\overline{Y}_T^B - \overline{Y}_C^B)\left[\sum_{i=1}^{J}\overline{Y}_i^B\overline{Y}_i^A - PJ\overline{Y}_T^B\overline{Y}_T^A - (1-P)J\overline{Y}_C^B\overline{Y}_T^A\right]}{\sum_{i=1}^{J}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
\hat{\theta} &= \frac{\sum_{i=1}^{J}\overline{Y}_i^B\overline{Y}_i^A - PJ\overline{Y}_T^B\overline{Y}_T^A - (1-P)J\overline{Y}_C^B\overline{Y}_C^A}{\sum_{i=1}^{J}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2}
\end{aligned}$$

We rewrite these terms as

$$\begin{aligned}
\hat{\alpha} &= \overline{Y}_C^A - \hat{\theta}\overline{Y}_C \\
\hat{\tau} &= (\overline{Y}_T^A - \overline{Y}_C^A) - \hat{\theta}(\overline{Y}_T^B - \overline{Y}_C^B) \\
\hat{\theta} &= \frac{\sum_{i=1}^{PJ}(\overline{Y}_i^B - \overline{Y}_T^B)(\overline{Y}_i^A - \overline{Y}_T^A) + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)(\overline{Y}_i^A - \overline{Y}_C^A)}{\sum_{i=1}^{PJ}(\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}_i^B - \overline{Y}_C^B)^2}
\end{aligned}$$

Variance of coefficient estimates The variance of each coefficient estimate is:

$$\text{Var}\left[\begin{pmatrix}\hat{\alpha}\\ \hat{\tau}\\ \hat{\theta}\end{pmatrix} \mid \mathbf{X}\right] = \underbrace{(\mathbf{X}'\mathbf{X})}_{\mathbf{M}}^{-1}\underbrace{\mathbf{X}'\,\mathrm{E}[\varepsilon\varepsilon' \mid \mathbf{X}]\mathbf{X}}_{\mathbf{N}}\underbrace{(\mathbf{X}'\mathbf{X})}_{\mathbf{M}}^{-1}$$

where $\mathbf{M}$ is the same as in Equation (A50) and $\mathbf{N}$ is

$$\begin{aligned}\mathbf{N} &= \begin{pmatrix} \mathbf{1}' \\ \mathbf{D}' \\ \overline{\mathbf{Y}}^{\mathbf{B}\prime} \end{pmatrix} \begin{pmatrix} \mathrm{E}[\varepsilon_{11}\varepsilon_{11} \mid \mathbf{X}] & \dots & \mathrm{E}[\varepsilon_{11}\varepsilon_{Jr} \mid \mathbf{X}] \\ \vdots & \ddots & \vdots \\ \mathrm{E}[\varepsilon_{Jr}\varepsilon_{11} \mid \mathbf{X}] & \dots & \mathrm{E}[\varepsilon_{Jr}\varepsilon_{Jr} \mid \mathbf{X}] \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{D} & \overline{\mathbf{Y}}^{\mathbf{B}} \end{pmatrix} \\ &= \sum_{t=1}^{r}\sum_{s=1}^{r} \begin{pmatrix} \sum_{i=1}^{J}\sum_{j=1}^{J} \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J} \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J} \overline{Y}_i^B \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] \\ \sum_{i=1}^{PJ}\sum_{j=1}^{J} \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{PJ} \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{J} \overline{Y}_j^B \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] \\ \sum_{i=1}^{J}\sum_{j=1}^{J} \overline{Y}_i^B \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ}\sum_{j=1}^{PJ} \overline{Y}_j^B \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{J}\sum_{j=1}^{J} \overline{Y}_i^B\overline{Y}_j^B \mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] \end{pmatrix}\end{aligned}$$

We now focus on the variance of the treatment effect estimator, $\hat{\tau}$. Combining this expression for $\mathbf{N}$ with the expression for $\mathbf{M}^{-1}$ above, this variance is given by

$$\begin{aligned}\mathrm{Var}(\hat{\tau} \mid \mathbf{X}) = \frac{1}{J^2r^2Z^2}\sum_{t=1}^{r}\sum_{s=1}^{r}\Bigg\{ &\frac{1}{P^2}\sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\Bigg[\Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B)\Big) \\ &\times \Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_j^B - \overline{Y}_T^B)\Big)\mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}]\Bigg] \\ &+ \frac{2}{P(1-P)}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\Bigg[\Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B)\Big) \\ &\times \Big(-Z - (1-P)J(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_j^B - \overline{Y}_C^B)\Big)\mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}]\Bigg] \\ &+ \frac{1}{(1-P)^2}\sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\Bigg[\Big(-Z - (1-P)J(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_C^B)\Big) \\ &\times \Big(-Z - (1-P)J(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_j^B - \overline{Y}_C^B)\Big)\mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}]\Bigg]\Bigg\} \qquad \text{(A51)}\end{aligned}$$

where

$$Z = \sum_{k=1}^{PJ}(\overline{Y}_k^B - \overline{Y}_T^B)^2 + \sum_{k=PJ+1}^{J}(\overline{Y}_k^B - \overline{Y}_C^B)^2$$

We consider each pair of summations in Equation (A51) by defining

$$\begin{aligned}V_T \equiv \frac{1}{P^2}\sum_{t=1}^{r}\sum_{s=1}^{r}\sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\Bigg[&\Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B)\Big) \\ &\times \Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_j^B - \overline{Y}_T^B)\Big)\mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}]\Bigg] \qquad \text{(A52)} \\ V_X \equiv \frac{2}{P(1-P)}\sum_{t=1}^{r}\sum_{s=1}^{r}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\Bigg[&\Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B)\Big) \\ &\times \Big(-Z - (1-P)J(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_j^B - \overline{Y}_C^B)\Big)\mathrm{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}]\Bigg] \qquad \text{(A53)}\end{aligned}$$

$$V_C \equiv \frac{1}{(1-P)^2}\sum_{t=1}^{r}\sum_{s=1}^{r}\sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\Bigg[\Big(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_C^B)\Big) \times\Big(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_j^B-\overline{Y}_C^B)\Big)\,\mathrm{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}]\Bigg] \quad \text{(A54)}$$

so Equation (A51) can be rewritten as

$$\mathrm{Var}(\hat{\tau}\mid\mathbf{X}) = \frac{1}{J^2r^2Z^2}\left(V_T+V_X+V_C\right) \quad \text{(A55)}$$

Each of these expressions contains an expectation term, which we rewrite, using the Law of Iterated Expectations, as

$$\mathrm{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}] = \mathrm{E}[\varepsilon_{it}\varepsilon_{js}\mid\overline{Y}_i^B,\overline{Y}_j^B] = \mathrm{E}\left[\varepsilon_{js}\,\mathrm{E}[\varepsilon_{it}\mid\varepsilon_{js},\overline{Y}_i^B,\overline{Y}_j^B]\mid\overline{Y}_i^B,\overline{Y}_j^B\right]$$

Note that ε_{it}, ε_{js}, $\overline{Y}_i^B$, and $\overline{Y}_j^B$ are random variables drawn from a multivariate normal distribution. Consider two cases. First, if $i \neq j$, then the inner expectation term is

$$\mathrm{E}[\varepsilon_{it}\mid\varepsilon_{js},\overline{Y}_i^B,\overline{Y}_j^B] = \mathrm{E}[\varepsilon_{it}\mid\overline{Y}_i^B] = \frac{\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B)}{\mathrm{Var}(\overline{Y}_i^B)}\overline{Y}_i^B$$

and the full expectation term is

$$\mathrm{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}] = \frac{\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B)}{\mathrm{Var}(\overline{Y}_i^B)}\overline{Y}_i^B\,\mathrm{E}[\varepsilon_{js}\mid\overline{Y}_j^B] = \frac{\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B)}{\mathrm{Var}(\overline{Y}_i^B)}\overline{Y}_i^B\frac{\mathrm{Cov}(\varepsilon_{js},\overline{Y}_j^B)}{\mathrm{Var}(\overline{Y}_j^B)}\overline{Y}_j^B \quad \text{(A56)}$$

However, note that

$$\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B) = \mathrm{Cov}(\omega_{it},\overline{\omega}_i^B) - r\psi^X$$

so summing over all r post-treatent periods gives

$$\sum_{t=1}^{r}\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B) = \frac{1}{m}\sum_{t=1}^{r}\sum_{s=-m+1}^{0}\mathrm{Cov}(\omega_{it},\omega_{is}) - r\psi^X = 0$$

As a result, the case of $i \neq j$ will not contribute to V_T, V_X, and V_C. Consider instead the case of $i = j$, in which case the inner expectation term is

$$\begin{aligned}\mathrm{E}[\varepsilon_{it}\mid\varepsilon_{js},\overline{Y}_i^B,\overline{Y}_j^B] &= \mathrm{E}[\varepsilon_{it}\mid\varepsilon_{is},\overline{Y}_i^B] \\ &= \tfrac{\mathrm{Var}(\overline{Y}_i^B)\,\mathrm{Cov}(\varepsilon_{it},\varepsilon_{is})-\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B)\,\mathrm{Cov}(\varepsilon_{is},\overline{Y}_i^B)}{\mathrm{Var}(\varepsilon_{is})\,\mathrm{Var}(\overline{Y}_i^B)-\mathrm{Cov}(\varepsilon_{is},\overline{Y}_i^B)^2}\varepsilon_{is} + \tfrac{\mathrm{Var}(\varepsilon_{is})\,\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B)-\mathrm{Cov}(\varepsilon_{it},\varepsilon_{is})\,\mathrm{Cov}(\varepsilon_{is},\overline{Y}_i^B)}{\mathrm{Var}(\varepsilon_{is})\,\mathrm{Var}(\overline{Y}_i^B)-\mathrm{Cov}(\varepsilon_{is},\overline{Y}_i^B)^2}\overline{Y}_i^B\end{aligned}$$

and the full expectation term is

$$\mathrm{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}] = \mathrm{E}\left[\varepsilon_{is}\,\mathrm{E}[\varepsilon_{it}\mid\varepsilon_{is},\overline{Y}_i^B]\mid\overline{Y}_i^B\right]$$

$$= \text{Cov}(\varepsilon_{it}, \varepsilon_{is}) - \frac{\text{Cov}(\varepsilon_{it}, \overline{Y}_i^B)\,\text{Cov}(\varepsilon_{is}, \overline{Y}_i^B)}{\text{Var}(\overline{Y}_i^B)} + \frac{\text{Cov}(\varepsilon_{it}, \overline{Y}_i^B)\,\text{Cov}(\varepsilon_{is}, \overline{Y}_i^B)}{\text{Var}(\overline{Y}_i^B)^2}(\overline{Y}_i^B)^2 \tag{A57}$$

Substituting Equations (A56) and (A57) into Equations (A52)–(A54) and simplifying gives

$$V_T = \frac{r^2}{P^2}\left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \times \sum_{i=1}^{PJ}\left[Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B))\right]^2 \tag{A58}$$

$$V_X = 0 \tag{A59}$$

$$V_C = \frac{r^2}{(1-P)^2}\left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \times \sum_{i=PJ+1}^{J}\left[-Z - (1-P)J(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_C^B))\right]^2 \tag{A60}$$

where θ is the same as in Equation (A49) and can be expressed as

$$\theta = \frac{m\sigma_\upsilon^2 + m\psi^X}{m\sigma_\upsilon^2 + \sigma_\omega^2 + (m-1)\psi^B}$$

Then substituting Equations (A58)–(A60) into Equation (A55) gives the variance of the treatment effect estimator:

$$\begin{aligned}
\text{Var}(\hat{\tau} \mid \mathbf{X}) &= \frac{1}{J^2r^2Z^2}\left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \\
&\quad \times \left[\frac{r^2}{P^2}\sum_{i=1}^{PJ}\left[Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B))\right]^2\right. \\
&\quad \left. + \frac{r^2}{(1-P)^2}\sum_{i=PJ+1}^{J}\left[-Z - (1-P)J(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_C^B))\right]^2\right] \\
&= \frac{1}{J^2Z^2}\left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \\
&\quad \times \left[\frac{JZ^2}{P(1-P)} + J^2(\overline{Y}_T^B - \overline{Y}_C^B)^2Z\right] \\
&= \left[\frac{1}{P(1-P)J} + \frac{(\overline{Y}_T^B - \overline{Y}_C^B)^2}{Z}\right] \\
&\quad \times \left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]
\end{aligned} \tag{A61}$$

The second component of the first term in Equation (A61), $\frac{(\overline{Y}_T^B - \overline{Y}_C^B)^2}{Z}$, is the only component of this formula that includes data, rather than parameters describing the experimental design or error structure. Additionally, this term decreases in expectation to zero as the number of units in the experiment increases, so it is negligible when J is sufficiently large. Accordingly, we drop this term to yield an approximate expression for the variance of the treatment effect estimator:

$$\text{Var}(\hat{\tau} \mid \mathbf{X}) \approx \frac{1}{P(1-P)J}\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]$$

Minimum detectable effect When using an ANCOVA regression model, the MDE is

$$\begin{aligned} MDE &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\ &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\left[\tfrac{1}{P(1-P)J} + \tfrac{(\overline{Y}_T^B - \overline{Y}_C^B)^2}{Z}\right]^{1/2} \\ &\qquad \times \left[(1-\theta)^2\sigma_v^2 + \left(\tfrac{\theta^2}{m} + \tfrac{1}{r}\right)\sigma_\omega^2 + \tfrac{\theta^2(m-1)}{m}\psi^B + \tfrac{r-1}{r}\psi^A - 2\theta\psi^X\right]^{1/2} \qquad \text{(A62)} \\ &\approx \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\tfrac{1}{P(1-P)J}\left[(1-\theta)^2\sigma_v^2 + \left(\tfrac{\theta^2}{m} + \tfrac{1}{r}\right)\sigma_\omega^2 + \tfrac{\theta^2(m-1)}{m}\psi^B + \tfrac{r-1}{r}\psi^A - 2\theta\psi^X\right]} \end{aligned}$$

This is the serial-correlation-robust ANCOVA (SCR ANCOVA) power calculation formula, found in Equation (11) in the main text.

A.3 Arbitrary cross-sectional correlations

In this section, we provide proofs of Lemma 1 from the main text. We begin by proving the analogous cross-sectional case, in Lemma A1:[7]

Lemma A1. *In a cross-sectional model with random assignment to treatment,* $\frac{\sigma_\varepsilon^2}{P(1-P)J}$ *is an unbiased estimator of the expectation of* $\text{Var}(\hat{\tau} \mid \mathbf{X})$ *even if* $\text{E}[\varepsilon_i\varepsilon_j \mid \mathbf{X}] \neq 0$ *for some* $i \neq j$.

Proof We wish to demonstrate that the OLS variance estimator, which assumes that errors are independent across units, is an unbiased estimator of the *ex ante* expected variance under non-i.i.d. errors when units are randomly assigned to treatment. To do this, consider a model similar to that of Appendix A.1 but allowing for correlation in error terms. That is:

Model There are J units randomly assigned a treatment status D_i, with proportion P in treatment ($D_i = 1$) and proportion $(1-P)$ in control ($D_i = 0$). The units are indexed so $i \in [1, PJ]$ is treated and $j \in [PJ+1, J]$ is a control. We make standard assumptions for randomized trials:

Assumption A20 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_i = \beta + \tau D_i + \varepsilon_i$$

7. Campbell (1977) provides the first version of this proof, which is cited by Moulton (1986), and which imposes a grouped error structure. Our proof allows for arbitrary cross-sectional error dependence. Athey and Imbens (2017b, 2017a) still recommend using Eicker-Huber-White standard errors in this case, to allow for heteroskedasticity. To our knowledge, no paper discusses power calculations in the presence of heteroskedastic disturbances.

where Y_i is the outcome of interest for unit i; β is the expected outcome of non-treated units; τ is the treatment effect which is homogeneous across all units; D_i is a treatment indicator; and ε_i is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Assumption A21 (Strict exogeneity)**.** $\mathrm{E}[\varepsilon_i \mid \mathbf{X}] = 0$*, where* $\mathbf{X} = [\mathbf{1}\ \mathbf{D}]$*. In practice, this follows from random assignment of* D_i*.*

Coefficient estimate The coefficient estimates from an OLS regression are the same as in Appendix A.1:

$$\begin{aligned}\widehat{\beta} &= \overline{Y}_C \\ \widehat{\tau} &= \overline{Y}_T - \overline{Y}_C\end{aligned}$$

Variance of coefficient estimate The variance of the estimate of the treatment effect, $\hat{\tau}$, is

$$\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) = \mathrm{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) + \mathrm{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) - 2\,\mathrm{Cov}\left(\overline{Y}_T, \overline{Y}_C \mid \mathbf{X}\right) \tag{A63}$$

where the first term of Equation (A63) is

$$\begin{aligned}\mathrm{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) &= \mathrm{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} Y_i \mid \mathbf{X}\right) \\ &= \mathrm{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} \varepsilon_i \mid \mathbf{X}\right) \\ &= \frac{\sigma_\varepsilon^2}{PJ} + \frac{2}{(PJ)^2}\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ} \mathrm{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) \\ &= \frac{\sigma_\varepsilon^2}{PJ} + \frac{PJ-1}{PJ}\psi_T \end{aligned} \tag{A64}$$

where

$$\psi_T \equiv \frac{2}{PJ(PJ-1)}\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ} \mathrm{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)$$

is the average covariance between treated units. Similarly, the second term of Equation (A63) is

$$\mathrm{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) = \frac{\sigma_\varepsilon^2}{(1-P)J} + \frac{(1-P)J-1}{(1-P)J}\psi_C \tag{A65}$$

where

$$\psi_C \equiv \frac{2}{(1-P)J((1-P)J-1)}\sum_{i=PJ+1}^{J-1}\sum_{j=i+1}^{J} \mathrm{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)$$

is the average covariance between control units. The third term of Equation (A63) is

$$\begin{aligned}
-2\operatorname{Cov}\left(\overline{Y}_T, \overline{Y}_C \mid \mathbf{X}\right) &= -2\operatorname{Cov}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} Y_i, \frac{1}{(1-P)J}\sum_{j=PJ+1}^{J} Y_j \mid \mathbf{X}\right) \\
&= \frac{-2}{P(1-P)J^2}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\operatorname{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) \\
&= -2\psi_{TC} \qquad \text{(A66)}
\end{aligned}$$

where

$$\psi_{TC} \equiv \frac{1}{P(1-P)J^2}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)$$

is the average covariance between treatment and control units. Substituting Equations (A64)–(A66) into Equation (A63) yields

$$\begin{aligned}
\operatorname{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= \left[\frac{\sigma_\varepsilon^2}{PJ} + \frac{PJ-1}{PJ}\psi_T\right] + \left[\frac{\sigma_\varepsilon^2}{(1-P)J} + \frac{(1-P)J-1}{(1-P)J}\psi_C\right] - 2\psi_{TC} \\
&= \frac{\sigma_\varepsilon^2}{P(1-P)J} + \frac{PJ-1}{PJ}\psi_T + \frac{(1-P)J-1}{(1-P)J}\psi_C - 2\psi_{TC}
\end{aligned}$$

Note that σ_ε^2, P, and J are constant population or design parameters. With this in mind, taking expectations yields

$$\operatorname{E}\left[\operatorname{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = \frac{\sigma_\varepsilon^2}{P(1-P)J} + \frac{PJ-1}{PJ}\operatorname{E}\left[\psi_T\right] + \frac{(1-P)J-1}{(1-P)J}\operatorname{E}\left[\psi_C\right] - 2\operatorname{E}\left[\psi_{TC}\right] \qquad \text{(A67)}$$

By random assignment to treatment, the expectation terms are

$$\begin{aligned}
\operatorname{E}\left[\psi_T\right] &= \frac{2}{PJ(PJ-1)}\operatorname{E}\left[\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ}\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)\right] \\
&= \frac{2}{PJ(PJ-1)}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j\right)\operatorname{E}\left[\mathbf{1}\{i \in T, j \in T\}\right] \\
&= \frac{2P}{J(PJ-1)}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j\right) \qquad \text{(A68)}
\end{aligned}$$

$$\operatorname{E}\left[\psi_C\right] = \frac{2(1-P)}{J((1-P)J-1)}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j\right) \qquad \text{(A69)}$$

$$\operatorname{E}\left[\psi_{TC}\right] = \frac{2}{J^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\operatorname{Cov}\left(\varepsilon_i, \varepsilon_j\right) \qquad \text{(A70)}$$

Substituting Equations (A68)–(A70) into Equation (A67) yields

$$\begin{aligned}\mathrm{E}\left[\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] &= \frac{\sigma_{\varepsilon}^2}{P(1-P)J} + \frac{2}{J^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\mathrm{Cov}\left(\varepsilon_i, \varepsilon_j\right) + \frac{2}{J^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\mathrm{Cov}\left(\varepsilon_i, \varepsilon_j\right) \\ &\quad - \frac{4}{J^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\mathrm{Cov}\left(\varepsilon_i, \varepsilon_j\right) \\ &= \frac{\sigma_{\varepsilon}^2}{P(1-P)J} \end{aligned} \tag{A71}$$

By Assumptions A20 and A21, the OLS variance estimator is an unbiased estimator of $\frac{\sigma_{\varepsilon}^2}{P(1-P)J}$. That is:

$$\mathrm{E}\left[\widehat{\mathrm{Var}}_{OLS}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = \frac{\sigma_{\varepsilon}^2}{P(1-P)J} \tag{A72}$$

Combining Equations (A71) and (A72) gives

$$\mathrm{E}\left[\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = \mathrm{E}\left[\widehat{\mathrm{Var}}_{OLS}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] \tag{A73}$$

Therefore, the OLS variance estimator is an unbiased estimator of the *ex ante* expected variance under random assignment to treatment, even under non-i.i.d. errors. □

Lemma 1. *In a panel difference-in-differences model with treatment randomly assigned at the unit level,* $\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]$ *is an unbiased estimator of the expectation of* $\mathrm{Var}(\hat{\tau} \mid \mathbf{X})$*, even in the presence of arbitrary within-period cross-sectional correlations.*

Proof We wish to demonstrate that the serial-correlation-robust variance, which assumes that errors are independent across units, is an unbiased estimator of the *ex ante* expected variance under arbitrary within-period correlations, when units are randomly assigned to treatment. To do this, again consider the model of Appendix A.2.3.

Coefficient estimate The coefficient estimate is the same as in Appendix A.2.3:

$$\widehat{\tau} = \left(\overline{Y}_T^A - \overline{Y}_T^B\right) - \left(\overline{Y}_C^A - \overline{Y}_C^B\right)$$

Variance of coefficient estimate The variance of the treatment effect estimator, $\widehat{\tau}$, is also the same as in Appendix A.2.3:

$$\begin{aligned}\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= \left(\frac{m+r}{P(1-P)Jmr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{PJm}\right)\psi_{i,T}^B + \left(\frac{PJ-1}{PJm}\right)\psi_{t,T}^B \\ &\quad + \left(\frac{r-1}{PJr}\right)\psi_{i,T}^A + \left(\frac{PJ-1}{PJr}\right)\psi_{t,T}^A + \left(\frac{m-1}{(1-P)Jm}\right)\psi_{i,C}^B \\ &\quad + \left(\frac{(1-P)J-1}{(1-P)Jm}\right)\psi_{t,C}^B + \left(\frac{r-1}{(1-P)Jr}\right)\psi_{i,C}^A + \left(\frac{(1-P)J-1}{(1-P)Jr}\right)\psi_{t,C}^A\end{aligned}$$

$$- \frac{2}{PJ}\psi^X_{i,T} - \frac{2}{m}\psi^B_{t,X} - \frac{2}{r}\psi^A_{t,X} - \frac{2}{(1-P)J}\psi^X_{i,C}$$

Next, we show that, in expectation, this is equal to the variance component of the SCR formula. We begin by taking expectations:

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = &\left(\frac{m+r}{P(1-P)Jmr}\right)\sigma^2_\omega + \left(\frac{m-1}{PJm}\right)\mathrm{E}\left[\psi^B_{i,T}\right] + \left(\frac{PJ-1}{PJm}\right)\mathrm{E}\left[\psi^B_{t,T}\right] \\
&+ \left(\frac{r-1}{PJr}\right)\mathrm{E}\left[\psi^A_{i,T}\right] + \left(\frac{PJ-1}{PJr}\right)\mathrm{E}\left[\psi^A_{t,T}\right] + \left(\frac{m-1}{(1-P)Jm}\right)\mathrm{E}\left[\psi^B_{i,C}\right] \\
&+ \left(\frac{(1-P)J-1}{(1-P)Jm}\right)\mathrm{E}\left[\psi^B_{t,C}\right] + \left(\frac{r-1}{(1-P)Jr}\right)\mathrm{E}\left[\psi^A_{i,C}\right] + \\
&+ \left(\frac{(1-P)J-1}{(1-P)Jr}\right)\mathrm{E}\left[\psi^A_{t,C}\right] - \frac{2}{PJ}\mathrm{E}\left[\psi^X_{i,T}\right] - \frac{2}{m}\mathrm{E}\left[\psi^B_{t,X}\right] \\
&- \frac{2}{r}\mathrm{E}\left[\psi^A_{t,X}\right] - \frac{2}{(1-P)J}\mathrm{E}\left[\psi^X_{i,C}\right] \qquad \text{(A74)}
\end{aligned}$$

where

$$\mathrm{E}\left[\psi^B_{i,T}\right] = \frac{2}{Jm(m-1)}\sum_{i=1}^{J}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathrm{E}\left[\psi^B_{t,T}\right] = \frac{2P}{J(PJ-1)m}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathrm{E}\left[\psi^A_{i,T}\right] = \frac{2}{Jr(r-1)}\sum_{i=1}^{J}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathrm{E}\left[\psi^A_{t,T}\right] = \frac{2P}{J(PJ-1)r}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathrm{E}\left[\psi^B_{i,C}\right] = \frac{2}{Jm(m-1)}\sum_{i=1}^{J}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathrm{E}\left[\psi^B_{t,C}\right] = \frac{2(1-P)}{J((1-P)J-1)m}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathrm{E}\left[\psi^A_{i,C}\right] = \frac{2}{Jr(r-1)}\sum_{i=1}^{J}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathrm{E}\left[\psi^A_{t,C}\right] = \frac{2(1-P)}{J((1-P)J-1)r}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathrm{E}\left[\psi^X_{i,T}\right] = \frac{1}{Jmr}\sum_{i=1}^{J}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathrm{E}\left[\psi^B_{t,TC}\right] = \frac{2}{J^2m}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathrm{E}\left[\psi^A_{t,TC}\right] = \frac{2}{J^2 r}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathrm{E}\left[\psi^X_{i,C}\right] = \frac{1}{Jmr}\sum_{i=1}^{J}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

Using these terms, we rewrite Equation (A74) as

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)\right] &= \left(\frac{m+r}{P(1-P)Jmr}\right)\sigma^2_\omega + \left(\frac{m-1}{PJm}\right)\psi^B + \frac{2}{J^2m^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&\quad + \left(\frac{r-1}{PJr}\right)\psi^A + \frac{2}{J^2r^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) + \left(\frac{m-1}{(1-P)Jm}\right)\psi^B \\
&\quad + \frac{2}{J^2m^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) + \left(\frac{r-1}{(1-P)Jr}\right)\psi^A \\
&\quad + \frac{2}{J^2r^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) - \left(\frac{2}{PJ}\right)\psi^X \\
&\quad - \frac{4}{J^2m^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) - \left(\frac{2}{(1-P)J}\right)\psi^X \\
&\quad - \frac{4}{J^2r^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&= \left(\frac{m+r}{P(1-P)Jmr}\right)\sigma^2_\omega + \left[\frac{m-1}{PJm} + \frac{m-1}{(1-P)Jm}\right]\psi^B \\
&\quad + \left[\frac{r-1}{PJr} + \frac{r-1}{(1-P)Jr}\right]\psi^A - \left[\frac{2}{PJ} + \frac{2}{(1-P)J}\right]\psi^X \\
&\quad + \left[\frac{2}{J^2m^2} + \frac{2}{J^2m^2} - \frac{4}{J^2m^2}\right]\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&\quad + \left[\frac{2}{J^2r^2} + \frac{2}{J^2r^2} - \frac{4}{J^2r^2}\right]\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&= \frac{1}{P(1-P)J}\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right] \qquad \text{(A75)}
\end{aligned}$$

Recall from Equation (A31) that the variance component of the SCR formula is

$$\mathrm{Var}_{SCR}\left(\widehat{\tau}\mid\mathbf{X}\right) = \frac{1}{P(1-P)J}\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right] \qquad \text{(A76)}$$

Combining Equations (A75) and (A76) gives

$$\mathrm{E}\left[\mathrm{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)\right] = \mathrm{Var}_{SCR}\left(\widehat{\tau}\mid\mathbf{X}\right)$$

Therefore, the SCR variance estimator is an unbiased estimator of the *ex ante* expected variance under random assignment to treatment, even under non-i.i.d. errors. □

A.4 Equivalence of alternative difference-in-difference estimators

In this section, we demonstrate that the DD treatment effect estimator—and therefore, the variance of this estimator—with a full set of unit and time fixed effects is the same as the "simplified" DD estimator where unit fixed effects are replaced with a treatment group dummy and time fixed effects are replaced with a post-period dummy. We begin with the same model, data generating process, and assumptions as Section A.2.1, suppressed here in the interest of parsimony.

Coefficient estimate In Section A.2.1 above, we estimate the treatment effect, τ, using OLS with unit and time fixed effects. Here, we instead estimate the τ using OLS with a treatment group dummy and a post-treatment period dummy (i.e. Equation (8), with new notation):

$$Y_{it} = \beta + \tau D_{it} + \alpha_g + \gamma_p + \varepsilon_{it} \tag{A77}$$

where α_g is a treatment group dummy variable and γ_p is a post-treatment-period dummy variable. In a balanced panel, this is equivalent to de-meaning by the group and post-treatment period. Define:

$$\ddot{Y}_{it} = Y_{it} - \overline{Y}_g - \overline{Y}_p + \overline{\overline{Y}} \tag{A78}$$

$$\ddot{D}_{it} = D_{it} - \overline{D}_g - \overline{D}_p + \overline{\overline{D}} \tag{A79}$$

$$\ddot{\varepsilon}_{it} = \varepsilon_{it} - \overline{\varepsilon}_g - \overline{\varepsilon}_p + \overline{\overline{\varepsilon}} \tag{A80}$$

where:

$$\overline{Y}_T = \frac{1}{PJ(m+r)} \sum_{i=1}^{PJ} \sum_{t=-m+1}^{r} Y_{it}$$

$$\overline{Y}_C = \frac{1}{(1-P)J(m+r)} \sum_{i=PJ+1}^{J} \sum_{t=-m+1}^{r} Y_{it}$$

$$\overline{Y}_A = \frac{1}{Jr} \sum_{i=1}^{J} \sum_{t=1}^{r} Y_{it}$$

$$\overline{Y}_B = \frac{1}{Jm} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} Y_{it}$$

$$\overline{\overline{Y}} = \frac{1}{J(m+r)} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} Y_{it}$$

with $\overline{D}_g$, $\overline{D}_p$, $\overline{\overline{D}}$, $\overline{\varepsilon}_g$, $\overline{\varepsilon}_p$, and $\overline{\overline{\varepsilon}}$ defined analogously. Substituting Equations (A78)–(A80) into Equation (A77) and simplifying gives the de-meaned DGP:

$$
\begin{aligned}
\ddot{Y}_{it} &= (\tau D_{it} + \alpha_g + \gamma_p + \varepsilon_{it}) - (\tau \overline{D}_g + \alpha_g + \overline{\gamma} + \overline{\varepsilon}_g) - (\tau \overline{D}_p + \overline{\alpha} + \gamma_p + \overline{\varepsilon}_p) + (\tau \overline{\overline{D}} + \overline{\alpha} + \overline{\gamma} + \overline{\overline{\varepsilon}}) \\
&= \tau(D_{it} - \overline{D}_g - \overline{D}_p + \overline{\overline{D}}) + (\varepsilon_{it} - \overline{\varepsilon}_g - \overline{\varepsilon}_p + \overline{\overline{\varepsilon}}) \\
&= \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}
\end{aligned}
$$

The estimate of the treatment effect is

$$
\begin{aligned}
\widehat{\tau} &= (\ddot{\mathbf{D}}'\ddot{\mathbf{D}})^{-1}\ddot{\mathbf{D}}'\ddot{\mathbf{Y}} \\
&= \left(\sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it}^2 \right)^{-1} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} \ddot{Y}_{it} \\
&= \frac{m+r}{P(1-P)Jmr} \left[\sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} Y_{it} - \sum_{i=1}^{J} \overline{Y}_i \sum_{t=-m+1}^{r} \ddot{D}_{it} - \sum_{t=-m+1}^{r} \overline{Y}_t \sum_{i=1}^{J} \ddot{D}_{it} + \overline{Y} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} \right] \\
&= \frac{m+r}{P(1-P)Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} Y_{it} \\
&= \frac{m+r}{P(1-P)Jmr} \left[\sum_{i=1}^{PJ} \left(\frac{(1-P)m}{m+r} \sum_{t=1}^{r} Y_{it} - \frac{(1-P)r}{m+r} \sum_{t=-m+1}^{0} Y_{it} \right) \right. \\
&\qquad \left. + \sum_{i=PJ+1}^{J} \left(\frac{Pr}{m+r} \sum_{t=-m+1}^{0} Y_{it} - \frac{Pm}{m+r} \sum_{t=1}^{r} Y_{it} \right) \right] \\
&= \left(\overline{Y}_T^A - \overline{Y}_T^B \right) - \left(\overline{Y}_C^A - \overline{Y}_C^B \right)
\end{aligned}
$$

where

$$
\begin{aligned}
\overline{Y}_T^A &= \frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it} \\
\overline{Y}_T^B &= \frac{1}{PJm} \sum_{i=1}^{PJ} \sum_{t=-m+1}^{0} Y_{it} \\
\overline{Y}_C^A &= \frac{1}{(1-P)Jr} \sum_{i=PJ+1}^{J} \sum_{t=1}^{r} Y_{it} \\
\overline{Y}_C^B &= \frac{1}{(1-P)Jm} \sum_{i=PJ+1}^{J} \sum_{t=-m+1}^{0} Y_{it}
\end{aligned}
$$

which is identical to Equation (A11), meaning that the variance is also identical. □

It is trivial to extend the same logic to variants on the DD estimator in which *either* unit fixed effects are replaced with a treatment group dummy or time fixed effects are replaced with a post-treatment dummy.

B Figures in main text

This section provides further detail on the simulations and power calculations referenced in the main text. We discuss the algorithms and assumptions behind each of the simulation plots, as well as the two analytical power calculation figures.

B.1 Simulated AR(1) data

In Figure 1, we run Monte Carlo simulations where each iteration generates a new simulated dataset with an idiosyncratic error term (ω_{it}) that evolves via an AR(1) process. We vary the following three parameters across sets of 10,000 simulations: the number of pre-treatment periods (m), the number of post-treatment periods (r), and the AR(1) dependence parameter (γ). The remaining parameters (J, P, α, κ, β, μ_υ, σ^2_υ, μ_δ, σ^2_δ, σ^2_ω) are fixed across all simulations.

Step 1: We calculate τ^{McK} and τ^{SCR} for each set of simulations, given a set of parameters values for m, r, and γ. These values are functions of the number of pre-treatment periods m, the number of post-treatment periods r, the number of units J, the proportion of units randomized into treatment P, the desired Type-I error rate α, the desired power κ, the idiosyncratic variance σ^2_ω, and (for τ^{SCR}) the error structure as defined by ψ^B, ψ^A, and ψ^X:

$$\tau^{McK} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\frac{\sigma^2_\omega}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)}$$

$$\tau^{SCR} = (t^J_{1-\kappa} + t^J_{\alpha/2})\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}$$

Note that for both τ^{McK} and τ^{SCR}, we calculate the critical values $t^J_{1-\kappa}$ and $t^J_{\alpha/2}$ assuming J degrees of freedom, which is consistent with applying the CRVE *ex post*, clustering at the unit level with J units. Note also that ψ^B, ψ^A, and ψ^X depend on the correlation structure of the errors, and the AR(1) process enables us to derive closed form expressions for these covariances in terms of γ, σ^2_ω, m, and r. Because we set $m = r$ across all simulations, we can write ψ^B, ψ^A, and ψ^X as:

$$\psi^B = \frac{2\sigma^2_\omega}{(m-1)m}\sum_{z=1}^{m-1}(m-z)\gamma^z \tag{B1}$$

$$\psi^A = \psi^B \tag{B2}$$

$$\psi^X = \frac{\sigma^2_\omega}{m^2}\left[\sum_{z=1}^{m} z\gamma^z + \sum_{z=m+1}^{2m-1}(2m-z)\gamma^z\right] \tag{B3}$$

Step 2: For each simulation, we generate a dataset as specified by the data generating process:

$$Y_{it} = \beta + \upsilon_i + \delta_t + \omega_{it}$$

To do this, we draw J independent values of υ_i from the distribution $N(\mu_\upsilon, \sigma^2_\upsilon)$, and draw $m + r$ independent values of δ_t from the distribution $N(\mu_\delta, \sigma^2_\delta)$. We create the idiosyncratic error $\omega_{it} = \gamma\omega_{i(t-1)} + \xi_{it}$ by simulating an AR(1) process with serial correlation γ and a white noise term ξ_{it}

drawn from the distribution $N(0, \sigma_\xi^2)$, where $\sigma_\xi^2 = \sigma_\omega^2(1-\gamma^2)$.[8]

Step 3: We randomly assign treatment to PJ units. This involves randomly scrambling a vector of PJ ones and $(1-P)J$ zeros and assigning each unit i either a 1 indicating treatment or a 0 indicating control.[9] This allows us to construct a time-varying treatment indicator D_{it}, where $D_{it} = 1$ for all treated units in post-treatment periods only and $D_{it} = 0$ otherwise. We then create three outcome variables by adding treatment effects to the data generated in the previous step:

$$\begin{aligned} Y_{it}^{McK} &\equiv Y_{it} + \tau^{McK} D_{it} \\ Y_{it}^{SCR} &\equiv Y_{it} + \tau^{SCR} D_{it} \\ Y_{it}^{0} &\equiv Y_{it} + \tau^{0} D_{it}, \end{aligned}$$

where $\tau^0 = 0$ is a placebo treatment effect.

Step 4: We separately estimate the following three OLS-fixed effects regressions:

$$\begin{aligned} Y_{it}^{McK} &= \beta + \tau^{McK} D_{it} + \upsilon_i + \delta_t + \omega_{it} \\ Y_{it}^{SCR} &= \beta + \tau^{SCR} D_{it} + \upsilon_i + \delta_t + \omega_{it} \\ Y_{it}^{0} &= \beta + \tau^{0} D_{it} + \upsilon_i + \delta_t + \omega_{it} \end{aligned}$$

Step 5: For each estimated $\hat{\tau}^{McK}$, $\hat{\tau}^{SCR}$, and $\hat{\tau}^0$, we compute both OLS standard errors and CRVE standard errors, clustered at the unit level.

We repeat Steps 2–5 10,000 times, for values of $m = r \in \{1, \ldots, 20\}$ and for values of $\gamma \in \{0, 0.3, 0.5, 0.7, 0.9\}$.[10] After each set of 10,000 simulations, we calculate the percent of simulations where $\hat{\tau}^{McK}$, $\hat{\tau}^{SCR}$, and $\hat{\tau}^0$ reject the null hypothesis of $\tau = 0$ at significance level α, under both OLS and CRVE standard errors.

Figure 1 reports these rejection rates on the vertical axes, with the number of pre- and post-treatment periods ($m = r$) on the horizontal axes, for each value of γ. Reading the top row left to right, we report the rejection rates for τ^{McK} under OLS standard errors, for τ^{McK} under CRVE standard errors, and for τ^{SCR} under CRVE standard errors, respectively. Because these are rejection rates of true effects, we interpret these curves as realized statistical power. Reading the bottom row left to right, we report the rejection rates for τ^0 under OLS standard errors, for τ^0 under CRVE standard errors, and for τ^0 under CRVE standard errors, respectively. Because these are rejection rates of placebo effects, we report these curves as realized false rejection rates. (The bottom-center and bottom-right panels report identical rejection rates, because the center and right columns have the same test for false rejection rates.)

We fix $\alpha = 0.05$ and $\kappa = 0.80$ across all simulations, as these are the critical values commonly used in practice. However, they are essentially arbitrary, and our simulation results would look identical if we had chosen alternative tolerances for Type I vs. Type II errors. (The only difference

8. We allow for a sufficiently long "burn-in period" in this AR(1) process, so that the process starts to evolve many periods before the first period of simulated data.

9. Our code rounds PJ to the nearest integer value, even though PJ is already an integer in our main parameterization. Note that for the τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR} to be precisely calibrated, the effective $\tilde{P}$ (where $\tilde{P} = \text{round}(PJ)/J$) needs to equal the actual parameter value P.

10. We set $m = r$ only for simplicity. However, the results are very similar if we fix $m = 3$ and vary $r \in \{1, \ldots, 20\}$, or vice versa. In Appendix C.1, we present results that vary m and r separately.

would be that the vertical axes would change to reflect these alternative values.) All other fixed parameter values are arbitrary. We have set $J = 500$, $P = 0.5$, $\beta = 1$, $\mu_\upsilon = 100$, $\sigma^2_\upsilon = 80$, $\mu_\delta = 20$, $\sigma^2_\delta = 10$, and $\sigma^2_\omega = 10$. These values of σ^2_υ, σ^2_δ, and σ^2_ω imply an intracluster correlation coefficient of $\rho_\upsilon = 0.8$ and within-period correlation coefficient of $\rho_\delta = 0.1$. Importantly, our simulation results do not depend on any particular combination of these parameters values, because they rely on the internal consistency of the *ex ante* treatment effect calibration and the *ex post* estimation, conditional on a *given* set of parameter values. The only exceptions are for J and P: J must be larger enough to allow us to use the CRVE estimator (i.e., at least 40 clusters), and P must be within a reasonable range (i.e., between 0.1 and 0.9) such that there are a sufficient number of both treated clusters and control clusters.

We report additional simulations for DD power calculations with AR(1) data, each of which slightly tweaks the above algorithm. Figure 2 uses $m \in \{1, 2, 3, 4, 5, 6\}$ and $r \in \{1, 2, 3, 4, 5, 6\}$, varying the number of pre-treatment and post-treatment period separately. Figure 3 alters Step 4 above, by either (i) replacing υ_i fixed effects with a Treat_i dummy, (ii) replacing δ_t fixed effects with a Post_t dummy, or (iii) both. Figure 5 combines this tweak to Step 4 with alternate parameterization of the DGP in Step 2: (i) $\sigma^2_\upsilon = 0$, (ii) $\sigma^2_\delta = 0$, or (iii) $\sigma^2_\upsilon = \sigma^2_\delta = 0$.

Importantly, removing unit fixed effects alters the asymptotic properties of the CRVE, as the number of clusters no longer increases one-for-one with the number of regressors. For all simulations in this paper and appendix, we apply the CRVE estimator that is appropriate to each estimating equation (following Cameron and Miller (2015), pp. 14–15): if the regression includes unit fixed effects and we are clustering by unit, we scale standard errors by $\sqrt{\frac{N-1}{N-J-1}}$, where N the number of observations in the regression and J is the number of units (clusters); if the regression does not include unit fixed effects, we do not apply this correction factor.

The ANCOVA simulations in Figure 4 follow the a very similar algorithm as those in Figure 1. Each of the above steps is identical, except for the following changes:

Step 1: Instead of using the DD power calculation formulas, we apply the analogous ANCOVA formulas to calculate τ^{McK} and τ^{SCR}.[11]

$$\tau^{McK} = (t^J_{1-k} + t^J_{\alpha/2}) \sqrt{\tfrac{1}{P(1-P)J}\left[(1-\breve{\theta})^2\sigma^2_\upsilon + \left(\tfrac{\breve{\theta}^2}{m} + \tfrac{1}{r}\right)\sigma^2_\omega\right]}$$

where $\breve{\theta} = \dfrac{m\sigma^2_\upsilon}{m\sigma^2_\upsilon + \sigma^2_\omega}$, and

$$\tau^{SCR} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\left[\frac{1}{P(1-P)J} + \frac{(\overline{Y}^B_T - \overline{Y}^B_C)^2}{Z}\right]^{1/2} \times \left[(1-\theta)^2\sigma^2_\upsilon + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma^2_\omega + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]^{1/2}$$

11. To create Figure 4, we use the exact SCR ANCOVA formula of Equation (A62) rather than the approximate formula of Equation (11).

where

$$\theta = \frac{m\sigma_\upsilon^2 + m\psi^X}{m\sigma_\upsilon^2 + \sigma_\omega^2 + (m-1)\psi^B}$$
$$Z = \sum_{i=1}^{PJ} (\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^{J} (\overline{Y}_i^B - \overline{Y}_C^B)^2$$

Step 2: We set $\sigma_\delta^2 = 0$, consistent with the assumptions of our analytic ANCOVA derivations. This assumption would be unrealistic for real-world data, however this Monte Carlo exercise focuses on the internal (in)consistencies of the ANCOVA power calculation formulas.

Step 3: We also collapse all pre-treatment observations into unit-specific unweighted averages:

$$\overline{Y}_i^B \equiv \sum_{t=-m+1}^{0} Y_{it}$$

Step 4: We separately estimate the following ANCOVA regressions:

$$Y_{it}^{McK} = \beta + \tau^{McK} D_i + \theta \overline{Y}_i^B + \varepsilon_{it}$$
$$Y_{it}^{SCR} = \beta + \tau^{SCR} D_i + \theta \overline{Y}_i^B + \varepsilon_{it}$$
$$Y_{it}^{0} = \beta + \tau^{0} D_i + \theta \overline{Y}_i^B + \varepsilon_{it}$$

Note that we do not include time fixed effects δ_t, as we have assumed away time shocks for these simulations.

Step 5: We compute only CRVE standard errors, clustered at the unit level.

Figure 4 reports the resulting rejection rates of the ANCOVA estimator. We do not report the false rejection rates from the τ^0 regressions for the sake of brevity, and they do achieve the desired $\alpha = 0.05$ rejection rate in all cases.

An additional nuance with the ANCOVA simulations is that our simulation results are now sensitive to the intracluster correlation coefficient ρ_υ. This is because the proportion of variance that is unit-specific now affects the precision of the $\hat{\tau}$ estimator, because we have replaced the unit fixed effect (which directly controlled for this variance) with a linear control in the average pre-treatment level of the outcome variable. Figure 4 sets $\rho_\upsilon = 0.8$, and the results are similar for alternative values of ρ_υ.

B.2 Bloom et al. (2015) data

In Figure 6, we run Monte Carlo simulations using data from Bloom et al. (2015). These simulations are analogous to those described above, except that rather than simulating data, we use an actual dataset from a published panel RCT. We downloaded the paper's dataset from the *Quarterly Journal of Economics* website, and focused on the data used to estimate the paper's main results, reported in Column (1) of Table II of the paper. Consistent with the regression model that produced this

result, we base our Monte Carlo analysis on the following DD specification:

$$\text{Performance}_{it} = \alpha \text{Treat}_i \times \text{Experiment}_t + \beta_t + \gamma_i + \varepsilon_{it}$$

Converting the original paper's notation to our notation, and substituting the outcome variable and fixed effects with the names of the variables in the Bloom et al. (2015) dataset, we have:

$$\underbrace{\texttt{perform1}_{it}}_{Y_{it}} = \tau D_{it} + \underbrace{\texttt{year\_week}_t}_{\delta_t} + \underbrace{\texttt{personid}_i}_{\upsilon_i} + \omega_{it}$$

We keep only units in the main sample (i.e. $\texttt{expgroup} \in \{0, 1\}$), only pre-treatment weeks of data (i.e. $\texttt{year\_week} < 201049$), and only individuals with non-missing $\texttt{perform1}_{it}$ values for all weeks of pre-treatment data. This leaves us with a balanced panel of $J = 79$ individuals across 48 weeks. Table B1 provides summary statistics of the resulting dataset.

Table B1: Summary statistics – Bloom et al. (2015)

Mean	Std. Dev.	Min	Max	AR(1) $\hat{\gamma}$	Individuals	Periods	Observations
0.153	0.943	−2.766	3.665	0.233	79	48	3,792

Notes: This table shows summary statistics for worker productivity in the Bloom et al. (2015) data. The data are weekly job performance z-scores, constructed by taking the average of normalized performance measures, where each measure is standardized to have a mean of 0 and standard deviation of 1 across the sample. Our sample consists only of individuals that had no missing observations throughout the entire pre-treatment period, January 1, 2010 through November 28, 2010. We compute $\hat{\gamma}$ by estimating Equation (6) on residuals from this dataset. In doing so, we cluster standard errors at the individual level. The 95% confidence interval is [0.165, 0.300]. For more details on the standardized job performance measures and the actual experimental design, see Bloom et al. (2015).

We conduct simulations on this dataset by varying the number of pre-treatment periods (m) and the number of post-treatment periods (r). As with the simulations described above, we vary the panel length for values $m = r \in \{1, \ldots, 20\}$, iterating 10,000 simulations for each value of $m = r$. We set the parameter values $\sigma^2_\omega = 0.507$ and $\gamma = 0.233$ by regressing Y_{it} on person and week fixed effects, calculating the variance of the resulting residuals $\hat{\omega}_{it}$, and then estimating $\hat{\gamma}$ using

$$\hat{\omega}_{it} = \gamma \hat{\omega}_{i(t-1)} + \xi_{it}$$

Step 1: We calculate τ^{McK} and $\tau^{AR(1)}$ for each set of simulations, given $m = r$:

$$\tau^{McK} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \sqrt{\left(\frac{\sigma^2_\omega}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)}$$

$$\tau^{AR(1)} = (t^J_{1-\kappa} + t^J_{\alpha/2}) \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\breve{\psi}^B + \left(\frac{r-1}{r}\right)\breve{\psi}^A - 2\breve{\psi}^X\right]}$$

where

$$\breve{\psi}^B = \frac{2\sigma^2_\omega}{(m-1)m} \sum_{z=1}^{m-1} (m - z)\gamma^z$$

$$\breve{\psi}^A = \breve{\psi}^B$$
$$\breve{\psi}^X = \frac{\sigma_\omega^2}{m^2}\left[\sum_{z=1}^{m} z\gamma^z + \sum_{z=m+1}^{2m-1}(2m-z)\gamma^z\right]$$

We denote the covariance terms as $\breve{\psi}^B$, $\breve{\psi}^A$, and $\breve{\psi}^X$ to indicate that the AR(1) error assumption is a (poor) representation of the more complex covariance structure of this dataset. For both τ^{McK} and $\tau^{AR(1)}$, we calculate the critical values $t^J_{1-\kappa}$ and $t^J_{\alpha/2}$ assuming J degrees of freedom, which is consistent with applying the CRVE *ex post*, clustering at the individual level with J individuals.

Step 2: We calculate τ^{SCR} given $m = r$, by non-parametrically estimating $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from residuals. Appendix D.1 provides step-by-step details of this estimation algorithm. Rather than impose an AR(1) structure on the serial correlation, this method enables us to flexibly characterize the covariance structure of the Bloom et al. (2015) dataset with just three averaged parameters. This allows us to calculate τ^{SCR} as:

$$\tau^{SCR} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\tfrac{1}{P(1-P)J}\right)\left[\left(\tfrac{m+r}{mr}\right)k_\sigma\sigma^2_{\hat{\omega}} + \left(\tfrac{m-1}{m}\right)k_B\psi^B_{\hat{\omega}} + \left(\tfrac{r-1}{r}\right)k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}}\right]}$$

where

$$k_\sigma = \frac{I(m+r)^2}{2(I-1)mr}$$
$$k_B = \frac{I(m+r)^2}{2(I-1)r^2}$$
$$k_A = \frac{I(m+r)^2}{2(I-1)m^2}$$
$$k_X = 0$$

Appendix E provides a derivation of the coefficients k_σ, k_B, k_A, and k_X, and it proves that this expression for τ^{SCR} as a function of *estimated* variance-covariance parameters is equal (in expectation) to the MDE as a function of the *true* variance-covariance parameters.[12]

Step 3: For each simulation, we randomly select a range of $m+r$ consecutive weeks in the dataset. This subset of weeks will become the $(m+r)$-period panel dataset used in this particular simulation. We randomly assign treatment to PJ individuals. This involves randomly scrambling a vector of PJ ones and $(1-P)J$ zeros and assigning each individual i either a 1 indicating treatment or a 0 indicating control.[13] This allows us to construct a time-varying treatment indicator D_{it}, where $D_{it} = 1$ for all treated units in post-treatment periods only and $D_{it} = 0$ otherwise. We then create three outcome variables by adding treatment effects to the data generated in the previous step:

$$Y^{McK}_{it} \equiv Y_{it} + \tau^{McK}D_{it}$$

12. I denotes the number of units used to estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$ and $\psi^X_{\hat{\omega}}$. This is distinct from the sample size of the experiment J, however these simulations set $I = J = 79$ to include all units in the Bloom et al. (2015) dataset.

13. Our code rounds PJ to the nearest integer value. Note that for the τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR} to be precisely calibrated, the effective $\tilde{P}$ (where $\tilde{P} = \text{round}(PJ)/J$) needs to equal the actual parameter value P.

$$Y_{it}^{AR(1)} \equiv Y_{it} + \tau^{AR(1)} D_{it}$$
$$Y_{it}^{SCR} \equiv Y_{it} + \tau^{SCR} D_{it}$$

Step 4: We separately estimate the following three OLS-fixed effects regressions:

$$Y_{it}^{McK} = \beta + \tau^{McK} D_{it} + \upsilon_i + \delta_t + \omega_{it}$$
$$Y_{it}^{AR(1)} = \beta + \tau^{AR(1)} D_{it} + \upsilon_i + \delta_t + \omega_{it}$$
$$Y_{it}^{SCR} = \beta + \tau^{SCR} D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

Step 5: For each estimated $\hat{\tau}^{McK}$, $\hat{\tau}^{AR(1)}$, and $\hat{\tau}^{SCR}$, we compute CRVE standard errors, clustered at the individual level.

As with the AR(1) simulations above, we repeat Steps 3–5 10,000 times, for values of $m = r \in \{1, \ldots, 12\}$.[14] After each set of 10,000 simulations, we calculate the percent of simulations where $\hat{\tau}^{McK}$, $\hat{\tau}^{AR(1)}$, and $\hat{\tau}^{SCR}$ reject the null hypothesis of $\tau = 0$ at significance level α, under CRVE standard errors. Figure 6 reports these three rejection rates on the vertical axes, with the number of pre- and post-treatment periods ($m = r$) on the horizontal axes. We can interpret these curves as realized statistical power, just as in the top row of Figure 1. We fix $\alpha = 0.05$ and $\kappa = 0.80$ across all simulations, for the reasons discussed above. Besides our arbitrary choices of $P = 0.5$, all other parameters are determined by the Bloom et al. (2015) dataset: $J = I = 79$, $\sigma_\upsilon^2 = 0.243$, $\sigma_\delta^2 = 0.146$, and $\sigma_\omega^2 = 0.507$, implying $\rho_\upsilon = 0.271$ and $\rho_\delta = 0.163$. We do not estimate β, μ_υ, or μ_δ, as these parameters are no longer relevant when simulating on top of an existing dataset.

B.3 Pecan Street data

In Figure 8, we present analogous Monte Carlo results for simulations using the Pecan Street dataset of household electricity consumption (Pecan Street (2016)). These data are publicly available (with a researcher login) at `https://dataport.pecanstreet.org/data/interactive`, and they include 699 households over 26,888 hours. As with the Bloom et al. (2015) simulations, we construct a balanced panel of households and hours, by restricting the full Pecan Street dataset to a sample of households that report non-missing, non-zero electricity consumption for every hour between January 1, 2013 and December 31, 2014. This results in a balanced panel of $J = 97$ households over 17,520 hours, which we collapse to create daily, weekly, and monthly datasets. Table B2 presents basic summary statistics for all four datasets, and Figure B1 displays the time series of data for one randomly selected household in our sample at varying levels of aggregation. Unsurprisingly, while mean electricity use is consistent across different collapses of the data, the standard deviation decreases as we move from higher- to lower-frequency datasets.

These Pecan Street simulations follow an algorithm identical to the Bloom et al. (2015) simulations, and we describe this algorithm in detail above. We repeat the same set of simulations four times, estimating separate rejection rates for τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR}, for each of the hourly, daily, weekly, and monthly datasets. We again set $\alpha = 0.05$, $\kappa = 0.80$, and $P = 0.5$. The other relevant parameters for each dataset are:

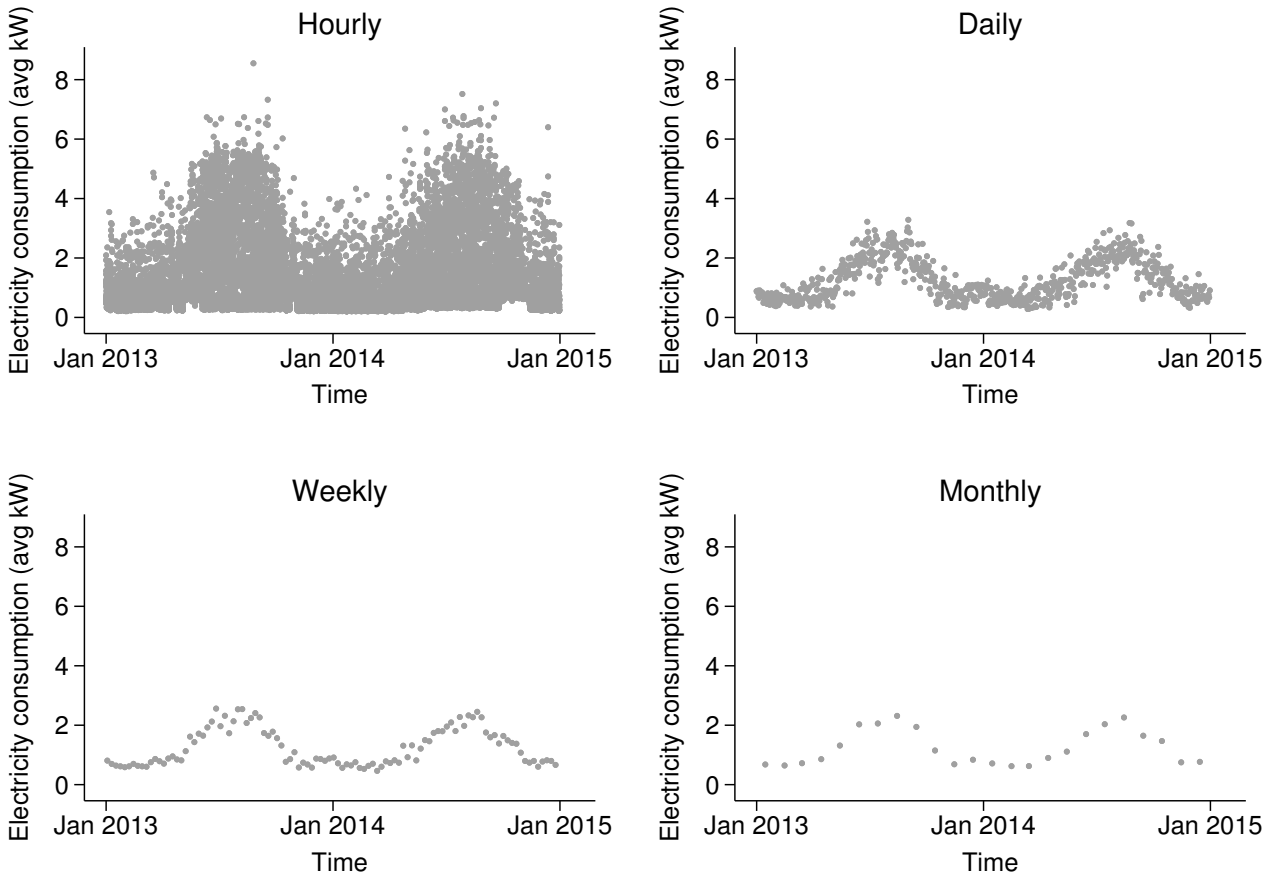
14. As with the AR(1) simulations, we set $m = r$ only for simplicity. However, the results are very similar if we fix $m = 3$ and vary $r \in \{1, \ldots, 20\}$, or vice versa.

Table B2: Summary statistics – Pecan Street

Dataset	Mean	Std. Dev.	Min	Max	AR(1) $\hat{\gamma}$	Households	Periods	Observations
Hourly	1.200	1.164	0.019	13.501	0.628	97	17,520	1,699,440
Daily	1.200	0.789	0.082	6.013	0.651	97	730	70,810
Weekly	1.198	0.739	0.105	5.175	0.713	97	106	10,282
Monthly	1.197	0.712	0.169	4.296	0.654	97	24	2,328

Notes: This table shows summary statistics for electricity consumption in the Pecan Street data. All values are in average kW of electricity consumed. The raw data are at the hourly level, in kWh. To construct the daily, weekly, and monthly dataset, we average hourly kWh consumption data across the relevant time period. We compute $\hat{\gamma}$ by estimating Equation (6) on residuals from this dataset. In doing so, we cluster standard errors at the household level. All are statistically significant at less than the 1% level.

Figure B1: Pecan Street data – Varying levels of aggregation



Notes: This figure shows the time series of Pecan Street electricity consumption data for one randomly selected household in our sample. Each panel displays the data at a different level of aggregation. The data are in units of average kW. These data are highly serially correlated: when we estimate an AR(1) model, Equation (6), on residuals from these datasets, we recover the AR(1) parameters of 0.628, 0.651, 0.713, and 0.654 for the hourly, daily, weekly, and monthly data, respectively.

Table B3: Pecan Street Simulation Parameters

Dataset	J	σ^2_υ	σ^2_δ	σ^2_ω	γ	ρ_υ	ρ_δ
Hourly	97	0.257	0.458	0.642	0.623	0.189	0.337
Daily	97	0.257	0.234	0.135	0.651	0.411	0.373
Weekly	97	0.256	0.211	0.083	0.713	0.465	0.384
Monthly	97	0.256	0.203	0.058	0.654	0.495	0.392

These values are estimated separately from each dataset used in the simulations.

In Appendix C.1, we present simulations analogous to Figures 2 and 3 in the main text, but using the Bloom et al. (2015) and Pecan Street datasets. The former vary m and r separately; the latter replace fixed effects with dummies in the DD estimating equations in Step 4 of Appendix B.2.

B.4 Analytic power calculations

Figure 9 displays the results of analytic power calculations performed using the daily Pecan Street dataset. In other words, we calculate the number of units needed by applying the McKenzie and SCR power calculation formulas, using the data to estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. For each experiment of length $m = r \in \{1, \ldots, 12\}$, we estimate the average values of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ over all possible panels of that length.[15] We assign half of the households to treatment ($P = 0.5$), allow for a 5 percent Type I error rate ($\alpha = 0.05$), and calibrate to 80 percent power ($\kappa = 0.80$). Finally, we rearrange Equations (3) and (D2) to calculate the number of households required to detect MDEs that range from 0 to 15 percent of baseline electricity consumption.

Figure 10 also shows the results of analytic power calculations using the SCR formula. However, instead of parameterizing Equation (2) using estimates from a dataset, we now normalize $\sigma^2_\omega = 1$ and assume an AR(1) correlation structure with $\gamma \in \{0, 0.3, 0.5, 0.7, 0.9\}$. For panel lengths $m = r \in \{1, \ldots, 100\}$, we analytically derive ψ^B, ψ^A, and ψ^X using the formulas from Equations (B1)–(B3). In the left panel, we fix $P = 0.5$, $\alpha = 0.05$, $\kappa = 0.80$, and $J = 100$, and use Equation (2) to solve for MDE as a function of $m = r$ and γ. In the right panel, we fix $P = 0.5$, $\alpha = 0.05$, $\kappa = 0.80$, and $MDE = 1$, and rearrange Equation (2) to solve for J as a function of $m = r$ and γ.

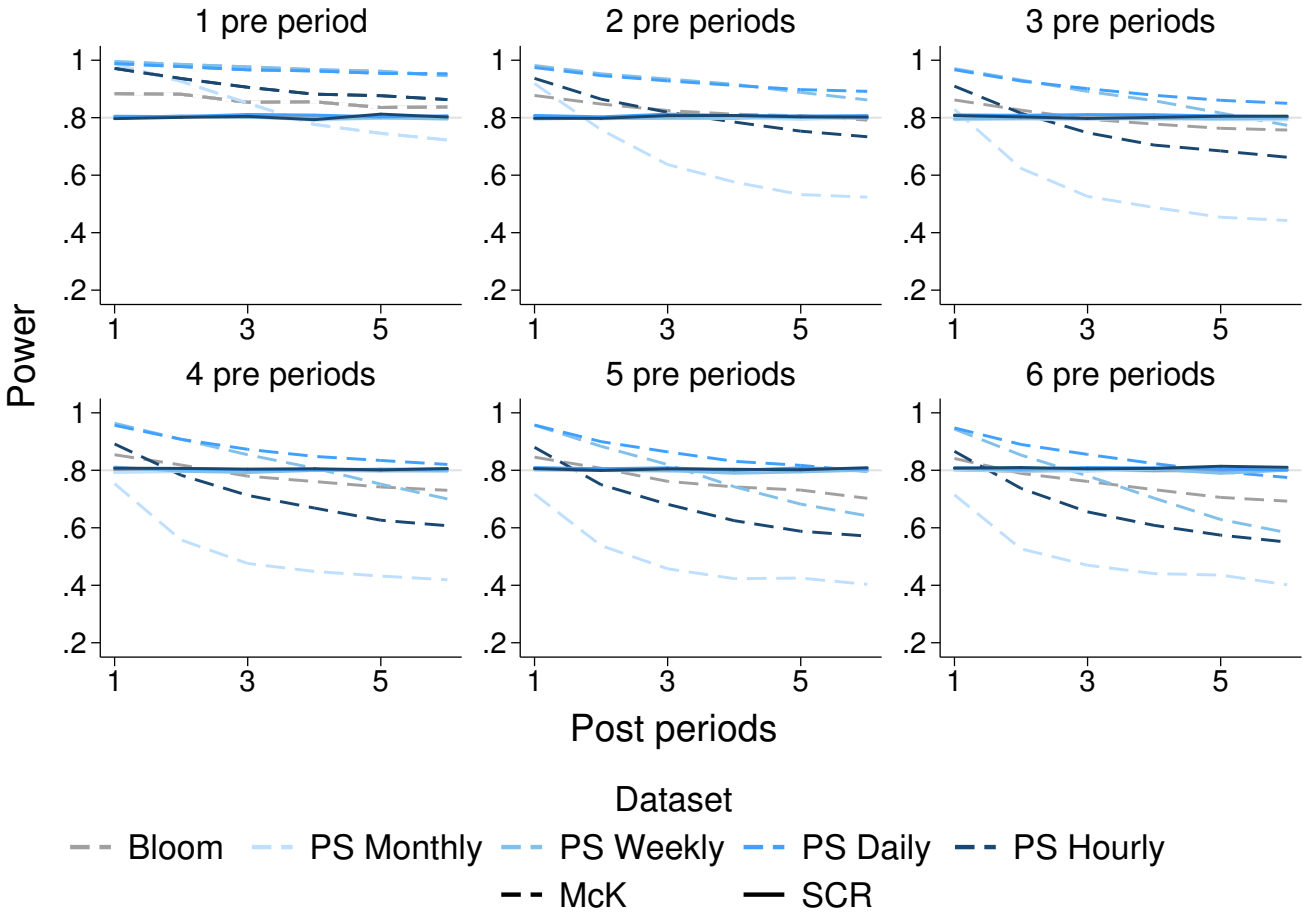
15. We follow the algorithm outlined in Appendix D.1 below. For a given value of $m = r$, we consider each (consecutive) subset S of the daily Pecan Street data with length $2r$. We first residualize this subset of the data with household and day fixed effects, and we calculate $\sigma^2_{\hat{\omega},S}$ from these residuals. We then assign the first m residuals for each household to the pre-treatment period and the remaining r residuals to the post-treatment period, thereby estimating $\psi^B_{\hat{\omega},S}$, $\psi^A_{\hat{\omega},S}$, and $\psi^X_{\hat{\omega},S}$ (by averaging all pairwise covariances for subset S). Averaging $\sigma^2_{\hat{\omega},S}$, $\psi^B_{\hat{\omega},S}$, $\psi^A_{\hat{\omega},S}$, and $\psi^X_{\hat{\omega},S}$ over all subsets S, we arrive at estimates for $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$.

C Additional results

C.1 Sensitivities with real data

In this section, we present extensions of our simulation results from the main text. In Figure C1, we present an analogous version of Figure 2 using real data from Bloom et al. (2015) and Pecan Street. Each panel conducts simulations that are identical to Figures 6 and 8, varying m and r separately (for $m \in \{1, \ldots, 6\}$ and $r \in \{1, \ldots, 6\}$). As with the simulated data, we find that the McKenzie formula typically yields over-powered experiments with either one pre-treatment or one post-treatment period, while the SCR formula yields the desired 80 percent power in all cases.

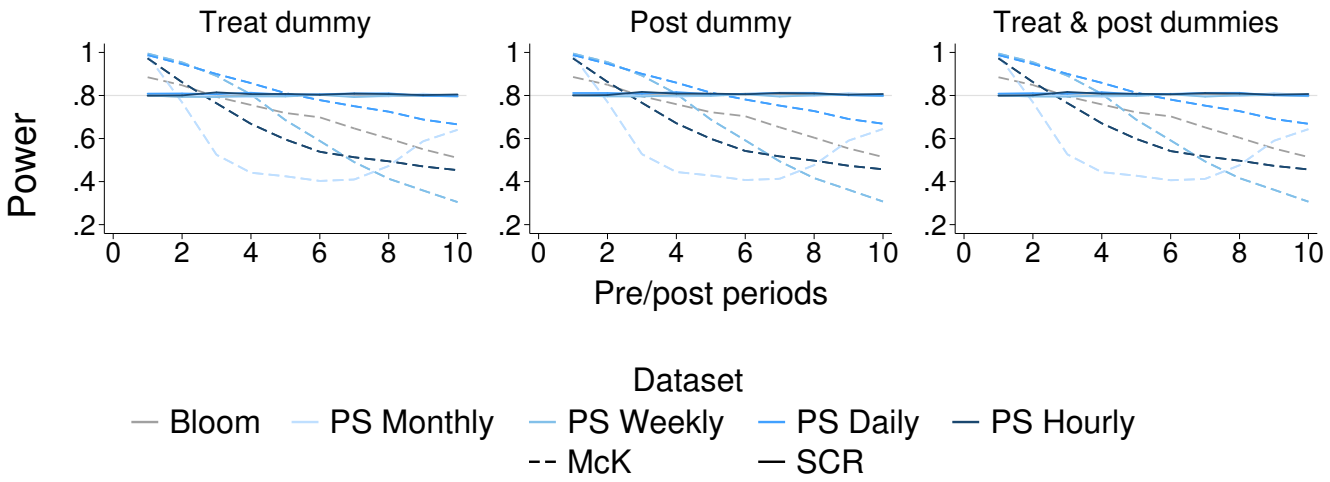
Figure C1: Power in short panels – Real data



Notes: This figure combines Figures 6 and 8 from the main text with the short panels of Figure 2. Simulations using the Bloom et al. (2015) and Pecan Street datasets follow the same algorithm describe in Appendix B.2, but separately varying the number of pre-treatment and post-treatment periods (for $m \in \{1, \ldots, 6\}$ and $r \in \{1, \ldots, 6\}$). For nearly all cases with either one pre-treatment period or one post-treatment period, the McKenzie formula yields over-powered experiments. Hence, experiments that follow the traditional "one baseline, one follow-up" structure will likely be overpowered, having calibrated an excessively large sample size. As the number of pre-treatment periods increases, power decreases monotonically for each dataset. By contrast, the SCR formula is properly powered in all cases.

Next, Figure C2 presents an analogous version of Figure 3 from the main text, using Bloom et al. (2015) and Pecan Street datasets. In the left panel, we replace unit fixed effects in the DD estimating equation with a Treat_i dummy. In the middle panel, we replace time fixed effects in the DD estimating equation with a Post_t dummy. In the right panel, we replace both unit and time fixed effects with $\text{Treat}_i + \text{Post}_t$ dummies. These resulting rejection rates are virtually identical to Figures 6 and 8, demonstrating that these alternative DD estimating equations all yield identical *ex post* power in real data (just as they do with simulated data in Figure 3).

Figure C2: Sensitivities to estimating equation – Real data



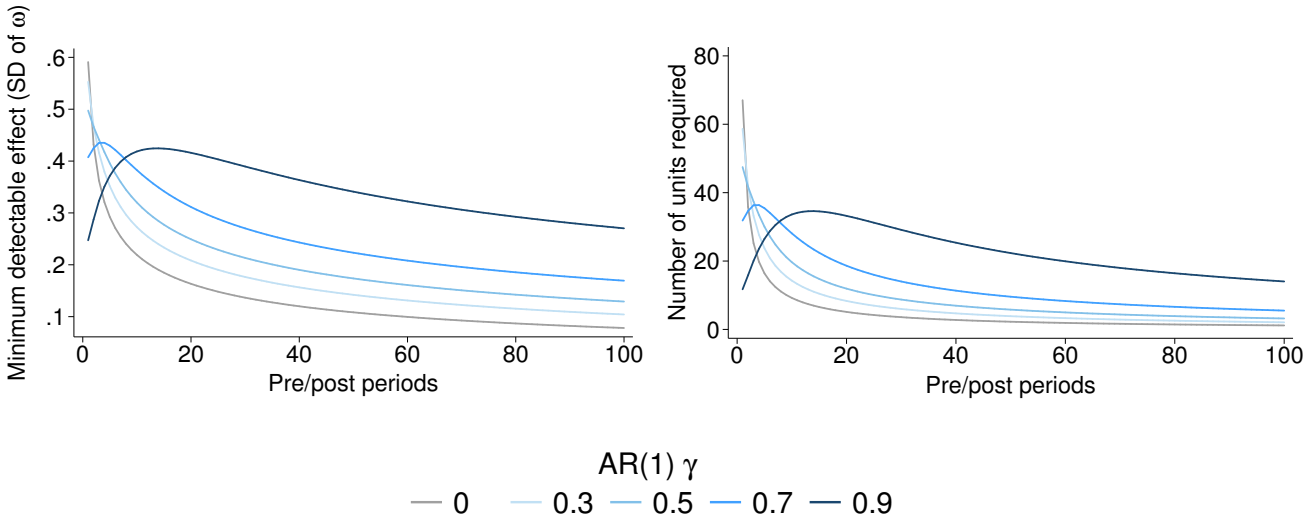
Notes: This figure replicates Figure 3 from the main text, using the Bloom et al. (2015) and Pecan Street datasets. The left panel replaces unit fixed effects in the DD estimating equation with a Treat_i dummy. The middle panel replaces time fixed effects in the DD estimating equation with a Post_t dummy. The right replaces both unit and time fixed effects with $\text{Treat}_i + \text{Post}_t$ dummies. As with simulated data, DD power calculations using real data are incorrectly powered using the McKenzie formula but correctly powered using our SCR formula.

C.2 Tradeoffs between MDE and time periods: ANCOVA

Figure 10 demonstrates that for DD experiments with strong serial correlation, it is possible to increase the MDE (or necessary sample size) by adding pre/post-treatment periods. This follows from Equation (4), which shows that serial correlation can either increase or decrease the variance of the DD estimator. Because DD identifies the treatment effect using differences between pre- and post-treatment outcomes, stronger serial correlation makes differences caused by treatment easier to detect. At the same time, stronger serial correlation means each additional time period provides less information. This intuition also holds for the ANCOVA estimator, as shown by Figure C3. This replicates Figure 10 using the SCR ANCOVA power calculation formula (Equation (11)). As

with DD, we find that adding pre/post periods to an ANCOVA experiment can decrease power, for short panels with strong serial correlation.[16]

Figure C3: Analytical power calculations with increasing panel length – ANCOVA



Notes: This figure replicates Figure 10 from the main text, using the SCR ANCOVA formula (Equation (11)) instead of the SCR DD formula (Equation (2)). The left panel shows the tradeoff between the minimum detectable effect (MDE) and the number of time periods ($m = r$) for varying levels of AR(1) serial correlation, holding the number of units fixed at $J = 100$ and normalizing MDE by the standard deviation of ω_{it}. At low levels of γ, MDE declines monotonically in m and r. However, for higher γ, increasing m and r actually *increases* MDE when $m = r$ is relatively small, and *decreases* MDE when $m = r$ is relatively large. The right panel shows the relationship between the number of units (J) and number of pre/post periods ($m = r$) required to detect an MDE equal to one standard deviation of ω_{it}. Similarly, for low levels of serial correlation, the trade-off between J and $m = r$ is monotonic. However, as γ increases, adding periods in short panels necessitates a greater number of units to achieve the same MDE, while adding periods in longer panels means that fewer units are required to achieve the same MDE.

C.3 Cluster randomization in panel RCTs

We use simulation-based power calculations to compare panel RCTs with unit-level randomization and cluster-level randomization. While cluster randomization remains outside the scope of this paper's analytical framework, it is a common approach for RCTs in development economics. Randomizing at the cluster (e.g. village) level, rather than the unit (e.g. household) level can be less expensive, simplify the logistics of administering treatment, and eliminate concerns about treatment interference (e.g. spillovers from treated to control households within a village). Here, we investigate how cluster-level randomization affects statistical power (relative to unit-level randomization), using the program `pc_simulate` from our STATA package `pcpanel`.[17]

We compare unit-level vs. cluster-level randomization, using simulated panel datasets based on the following DGP:

16. The SCR ANCOVA formula has an additional degree of freedom: we can vary σ_v^2 and σ_ω^2 separately. Here, we set a relatively small σ_v^2 (equal to $0.1\sigma_\omega^2$), which causes ANCOVA to be relatively dissimilar to DD (the two estimators converge as σ_v^2 approaches infinity).

17. We do not incorporate treatment spillovers, since doing so in a simulation context would necessitate (essentially) arbitrary assumptions on the strength and direction of treatment spillovers. Instead, we simulate RCTs with a 100 percent treatment intensity within treated clusters.

Assumption C1 (Group shock data generating process)**.** *The data are generated according to the following model:*

$$Y_{igt} = \beta + \tau D_{igt} + \zeta_g + \upsilon_i + \delta_t + \omega_{igt} \tag{C1}$$

where Y_{igt} is the outcome of interest for unit i in group g at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{igt} is a time-varying treatment indicator; ζ_g is a group-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\zeta)$; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\upsilon)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\delta)$; and ω_{igt} is an idiosyncratic error term with variance σ^2_ω.

We simulate 3 datasets, each containing 50 groups g of 10 units each, where each group is defined by a separate realization of ζ_g.[18] We vary the relative strength of group-specific shocks while holding constant total cross-sectional variation—that is, we vary $\sigma^2_\zeta/(\sigma^2_\zeta + \sigma^2_\upsilon)$ such that groups contribute 25, 50, of 75 percent of the total cross-sectional variation, while holding $\sigma^2_\zeta + \sigma^2_\upsilon$ fixed. For all datasets, we incorporate non-constant serial correlation where ω_{igt} follows an AR(1) process with $\gamma = 0.5$.[19]

Using `pc_simulate`, we conduct 6 sets of simulation-based power calculations, for varying panel lengths ($m = r \in \{1, 2, 3, 4, 5, 6\}$), assuming an estimating equation with unit fixed effects (which subsume group fixed effects) and time fixed effects. While all 6 sets of power calculations hold the sample size fixed at 500 units i, 3 assign treatment at the unit level (ignoring cross-sectional group correlations) and 3 assign treatment at the group level (with 100 percent treatment intensity within treated groups). We hold the proportion of units in treatment and minimum detectable effect constant across all simulations ($P = 0.5$, $MDE = 1.4$). All simulations cluster standard errors at the level of treatment (either unit or group), and calculate average rejection rates over 5,000 iterations to compute power. Using this data generating process, we expect power to be the same for unit-level randomization and group-level randomization, because the unit fixed effects fully control for the group shock.

Next, we repeat this entire procedure with a data generating process including group-by-time shocks:

Assumption C2 (Group-by-time shock data generating process)**.** *The data are generated according to the following model:*

$$Y_{igt} = \beta + \tau D_{igt} + \zeta_g + \upsilon_i + \delta_t + \chi_{gt} + \omega_{igt} \tag{C2}$$

where Y_{igt} is the outcome of interest for unit i in group g at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{igt} is a time-varying treatment indicator; ζ_g is a group-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\zeta)$; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\upsilon)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\delta)$; χ_{gt} is a group-by-time specific disturbance distributed i.i.id. $\mathcal{N}(0, \sigma^2_\chi)$; and ω_{igt} is an idiosyncratic error term with variance σ^2_ω.

For this set of simulations, we set all of the parameters identically to the above, and add $\sigma^2_\chi = 50$. In the `pc_simulate` stage, we set $MDE = 1.7$, in order to ensure that our power

18. We use 50 groups in order to have enough clusters to apply the CRVE.

19. We set the following specific parameters: $\sigma^2_\zeta \in \{20, 40, 60\}$, $\sigma^2_\upsilon \in \{60, 40, 20\}$, $\sigma^2_\zeta + \sigma^2_\upsilon = 80$, $\sigma^2_\delta = 10$, $\sigma^2_\omega = 150$. All parameter values are arbitrary, besides imposing constant $\sigma^2_\zeta + \sigma^2_\upsilon$ (which facilitates comparisons across datasets).

estimates are not truncated by the zero lower bound. Importantly, we still estimate an *ex post* model with unit and time fixed effects, which intentionally misspecifies Assumption C2 by omitting group-by-time fixed effects. Due to this misspecification (which mimics a real-world scenario in which a researcher does not know the true DGP), we now expect that power for the unit-level randomization will be greater than that of the group-level randomization, since the group-by-time shocks are not absorbed by the fixed effects in the model.

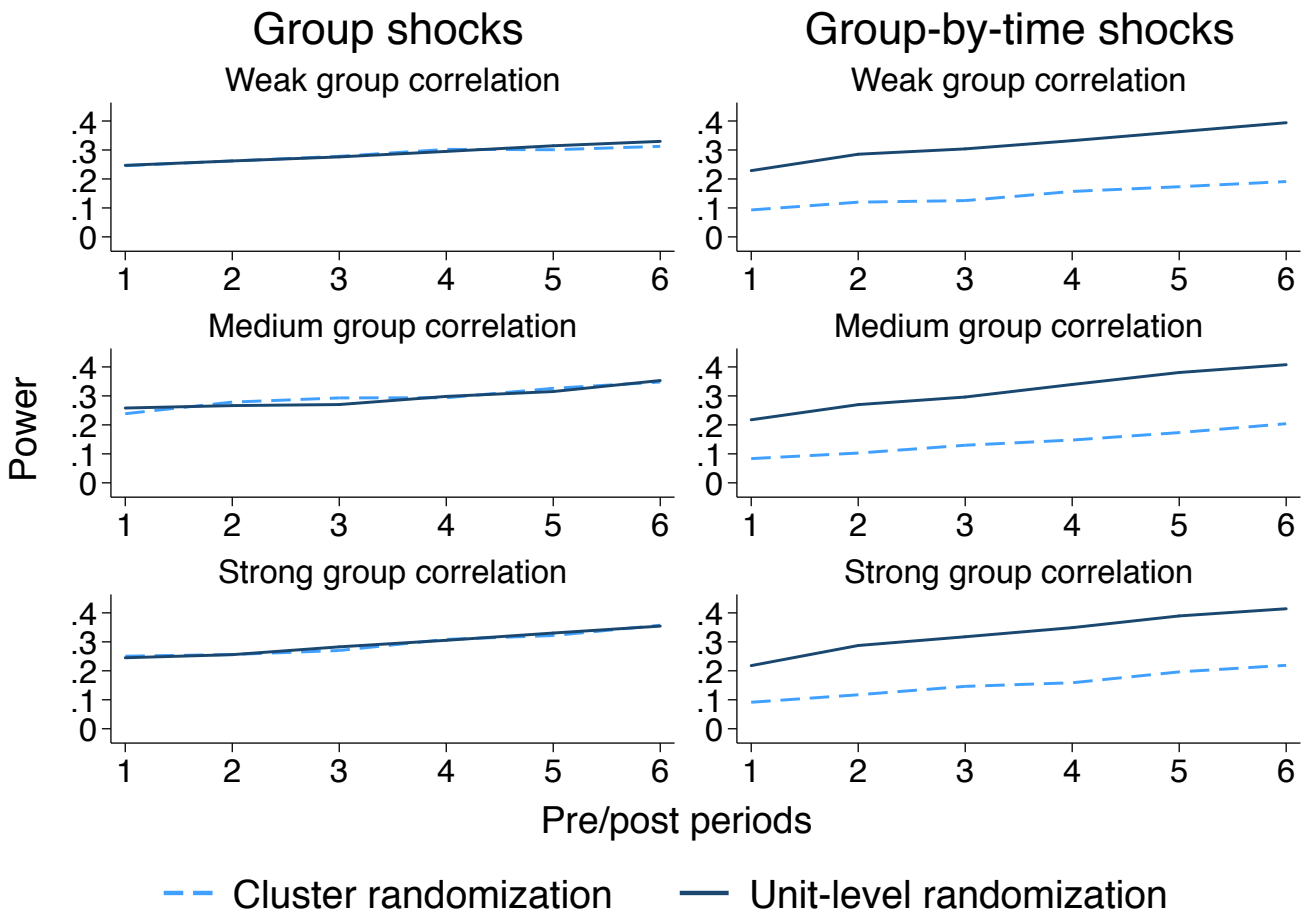
Figure C4 presents the results of this exercise. Each plot reports an "apples-to-apples" comparison for power calculations on the same dataset, randomized at the unit vs. group level. Moving top to bottom, the relative strength of group-specific shocks increases from 25 percent ("weak") to 50 percent ("medium") to 75 percent ("strong"). In the panels on the left, the data generating process includes group shocks only (Assumption C1). In the panels on the right, the data generating process also includes group-by-time shocks (Assumption C2).[20]

For all 3 datasets that only have group shocks, randomizing at the group level yields statistical power that is indistinguishable from that of the same experiment randomized at the unit level. The intuition behind this equivalence lies in how panel estimators control for unit-specific baselines: DD identifies off of (post − pre) differences at the *unit* level, which controls for the remaining within-group variation not exploited when randomizing at the group level. By contrast, when we repeat the same exercise for cross-sectional RCTs, we finds substantially lower power for group-level randomization (relative to unit-level randomization). This is because the cross-sectional treatment effect estimator does not control for within-group variation, meaning that the treatment indicator leaves more unexplained variation when randomized at the group level.

For all 3 datasets that have group-by-time shocks, randomizing at the group level yields statistical power that is substantially lower than that of the same experiment randomized at the unit level. This is because the misspecified model does not account for group-by-time variation from χ_{gt}. Just as non-constant unit-specific serial correlation affects power in panel data under unit-level randomization, non-constant group-specific correlation affects power in panel data under group-level randomization. Extending our analytical framework to include panel RCTs with arbitrary group shocks is beyond the scope of this paper. Given that group shocks can impact power in simulated data, and given that real-world data likely exhibit more complex group-specific correlation structures, we encourage researchers interested in cluster randomized panel RCTs to conduct simulation-based power calculations.

20. Technically, introducing χ_{gt} in Assumption C2 without changing σ^2_ζ and σ^2_υ lowers the composite share of group-specific shocks below {25, 50, 75} percent in the panels on the right.

Figure C4: Power calculations by simulation - Cluster randomization



Notes: This figure reports results from simulation-based power calculations, comparing unit-level randomization vs. group-level (cluster) randomization. Each panel reports power for DD experiments using a single simulated dataset, varying the number of pre- and post-treatment periods ($m = r = \{1, 2, 3, 4, 5, 6\}$). All datasets have 500 units, divided into 50 groups of 10 units each. Moving from top to bottom, datasets have either weak, medium, or strong within-group correlations (either 25, 50, or 75 percent of total cross-sectional variation, respectively). When cluster randomizing, all treated groups have a treatment intensity of 100 percent. All simulations calculate average rejection rates over 5,000 iterations. For the left column, the data generating process includes unit-, time-, and group-specific shocks (Assumption C1), and the MDE is 1.4. For the right column, the data generating process adds group-by-time shocks (Assumption C2), and the ME is 1.7. See text for further details.

C.4 Non-constant correlation: McKenzie (2012) empirical approach

The McKenzie results in the main text apply McKenzie's theoretical formula, which assumes serial correlation is constant across all time periods and is the same for all units. McKenzie (2012, p. 215) acknowledges that this assumption might be unrealistic and suggests an alternative for empirical settings with non-constant serial correlation. He recommends that practitioners use the average autocorrelation across all off-diagonal elements of the correlation matrix. For example, if the error term follows an AR(1) process such that adjacent periods have autocorrelation of 0.8 with geometric

decay in time, then for a panel RCT with 5 periods, McKenzie recommends calculating:

$$\overline{\rho}_t = \frac{0.8 + 0.8^2 + 0.8^3 + 0.8^4}{4 + 3 + 2 + 1} = 0.655 \tag{C3}$$

We run an additional set of simulations in order to compare this method to our SCR method. We adapt the procedures we use to generate Figure 6 (outlined in Appendix B.2) as follows:

Step 1: We generate a panel dataset with 500 units and 10,000 time periods, based on the data generating process[21]:

$$Y_{it} = \beta + \upsilon_i + \delta_t + \omega_{it}$$

We draw independent values of υ_i from the distribution $N(\mu_\upsilon, \sigma_\upsilon^2)$ but set $\sigma_\delta^2 = 0$ to assume deterministic time effects. This aligns with the assumptions of McKenzie (2012) who likewise assumes away stochastic time shocks, and represents a special case of our data generating process in Assumption 1. We simulate an AR(1) process to create the idiosyncratic error $\omega_{it} = \gamma\omega_{i(t-1)} + \xi_{it}$, for AR(1) parameters $\gamma \in \{0, 0.3, 0.5, 0.7, 0.9\}$, and a white noise term ξ_{it} drawn from the distribution $N(0, \sigma_\xi^2)$, where $\sigma_\xi^2 = \sigma_\omega^2(1 - \gamma^2)$.

Step 2: For varying numbers of pre- (m) and post-treatment periods (r), we calculate $\overline{\sigma}_\varepsilon^2$ as the average variance of Y_{it} across all subset panels of length $m + r$. For example, if $m + r = 4$, we calculate

$$\overline{\sigma}_\varepsilon^2 = \text{ mean}\Big\{ \text{Var}\left(Y_{it} \mid 1 \le t \le 4\right),\ \text{Var}\left(Y_{it} \mid 2 \le t \le 5\right),\ \text{Var}\left(Y_{it} \mid 3 \le t \le 6\right),\ \ldots \Big\}. \tag{C4}$$

Each simulated dataset contains many subsets of $m + r$ consecutive periods, and averaging $\text{Var}(Y_{it})$ across all such subsets reduces sampling variation in $\overline{\sigma}_\varepsilon^2$.

Step 3: We calculate $\overline{\rho}$ as the average pairwise within-unit correlation, across all off-diagonal elements of the correlation matrix (as in Equation (C3)). Following Step 2, we average $\overline{\rho}$ across all subset panels of length $m + r$. For example, if $m = r = 4$, we calculate

$$\overline{\rho} = \text{ mean}\Big\{ \left(\overline{\rho}_1 \mid 1 \le t \le 4\right),\ \left(\overline{\rho}_2 \mid 2 \le t \le 5\right),\ \left(\overline{\rho}_3 \mid 3 \le t \le 6\right),\ \ldots \Big\}, \tag{C5}$$

where

$$\overline{\rho}_t = \frac{2}{(m + r)(m + r + 1)} \sum_{s=t}^{t+m+r-1} \sum_{q=s+1}^{t+m+r} \text{Corr}(Y_{is}, Y_{iq})\,. \tag{C6}$$

As with $\overline{\sigma}_\varepsilon^2$, averaging $\overline{\rho}_t$ across all $m + r$ subsets reduces sampling variation in $\overline{\rho}$.

21. Long panels minimize sampling error when subsetting consecutive time periods in Steps 2 and 3 below.

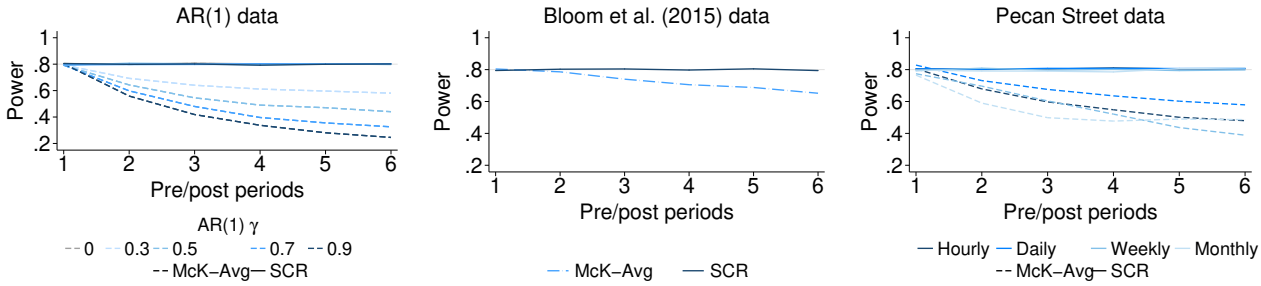
Step 4: Using $\overline{\sigma}^2_\varepsilon$ and $\overline{\rho}_t$, we calculate an MDE using McKenzie's power calculation formula for the difference-in-differences estimator:

$$\tau^{McK\text{-}Avg} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\frac{1}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)\overline{\sigma}^2_\varepsilon(1-\overline{\rho})} \tag{C7}$$

Step 5: We simulate 10,000 experiments, each using a randomly selected subset of $m+r$ consecutive periods, randomly assigning $P = 0.5$ of $J = 500$ units into treament, and applying the treatment effect $\tau^{McK\text{-}Avg}$. For each simulated experiment, we estimate $\hat{\tau}^{McK\text{-}Avg}$ with CRVE standard erorrs clustered at the unit level.

We also repeat Steps 2–5 above using both the Bloom et al. (2015) dataset and the Pecan Street datasets (hourly, daily, weekly, and monthly) in order to test McKenzie's approach on real-world data as well. Figure C5 reports realized power for all three sets of simulations. Using both simulated and real data, the above method yields underpowered experiments in all cases *except* two-period panels with $m = r = 1$. In this special case, calculating the average autocorrelation yields results equivalent to our SCR method.[22]

Figure C5: Power calculations estimating average autocorrelation



Notes: This figure reproduces Figures 1, 6, and 8 from the main text, but uses McKenzie (2012)'s recommended method of calculating average autocorrelation for panels of length $m + r$ (which is described above) to estimate the relevant parameters. Dashed lines report realized power from simulations with MDE calibrated using Equation (C7); solid lines report realized power from simulations using our preferred SCR method. The left panel reports results from simulated datasets with AR(1) errors; the central panel reports results using real data from Bloom et al. (2015); and the right panel reports results using real data from Pecan Street (at varying levels of aggregation). In both simulated and real data, McKenzie's method of averaging autocorrelation for experiments with more than two periods yields underpowered experiments. However, McKenzie's method does achieve the desired 80 percent power when $m = r = 1$.

22. When we fit an AR(1) function to each of these datasets, we recover AR(1) parameters of 0.233 (Bloom et al. (2015)), 0.623 (Pecan Street hourly), 0.651 (Pecan Street daily), 0.713 (Pecan Street weekly), and 0.654 (Pecan Street monthly). Hence, this finding holds for real datasets with both strong and weak non-constant serial correlation.

D A practical guide to power calculations

In this section, we address several practical considerations when conducting power calculations. Most of these challenges involve variance and covariance parameters that must be either estimated or assumed in order to operationalize a power calculation formula. We also outline steps for estimating power calculations via simulation, which is our preferred method.

D.1 Analytical power calculations

The most challenging aspect of analytical power calculations is parameterizing the variance and (if applicable) covariance terms that characterize the data's error structure. In the absence of a representative pre-existing dataset, researchers may struggle to even guess the order of magnitude of the error variance, let alone generate a precise estimate of this key parameter. Our theoretical results demonstrate that power for panel RCTs also hinges on the error *covariance* structure.

As in the paper, we denote the true parameters governing the data generating process as σ^2_ω, ψ^B, ψ^A, and ψ^X. We define $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ to be the parameters that characterize the *residuals* (rather than real errors). If researchers have access to a representative dataset *ex ante*, they can directly estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, and use these values to parameterize power calculations. Appendix E.2 proves that researchers can recover the true MDE using these residual-based parameters. Nevertheless, this process is not trivial, for several reasons.

First, while the idiosyncratic variance σ^2_ω is a population parameter, the three ψ parameters are functions of both the full covariance structure of the population and the specific values of m and r.[23] For a given population and serially correlated outcome variable, experiments with small m and r are likely to exhibit larger ψ^B, ψ^A, and ψ^X parameters than experiments with large m and r. This is because as the number of pre-treatment (post-treatment) periods increases, ψ^B (ψ^A) averages across covariances of time periods that are farther apart. For example, compare ψ^B with $m = 3$ vs. $m = 30$, for an outcome with a covariance structure where adjacent periods are more positively correlated than distant periods. For $m = 3$, ψ^B averages $m(m-1)/2 = 3$ pairwise covariances, 2 of which are for adjacent periods; for $m = 30$, ψ^B averages $m(m-1)/2 = 435$ pairwise covariances, only 29 of which are for adjacent periods. Because ψ^X expands with both m and r, it attenuates relatively faster than ψ^B and ψ^A as panel length grows.

Second, estimating $\text{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is})$ using residuals from an existing dataset is fundamentally impossible, given that each dataset contains only one realization of (Y_{it}, Y_{is}). However, researchers may treat the $(I \times 1)$ vectors of residuals $(\hat{\boldsymbol{\omega}}_t, \hat{\boldsymbol{\omega}}_s)$ as I draws from the distributions of residuals for periods (t,s) and estimate these distributions' covariance. The resulting estimates, which we denote $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$, are unbiased estimators of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$.[24]

Third, if the representative dataset contains a long time series, the residual variance and covariance structure may change throughout the time series. This means if researchers estimate $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ by averaging across the full time series, these estimated parameters may be less representative than if they were estimated from just the end of the time series.[25] Since the residual

23. Deriving the residual-based parameters $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ introduces an additional complexity, as these residual-based parameters are defined by the number of pre-treatment periods (m), post-treatment periods (r) and cross-sectional units (I) used to produce these residuals.

24. Appendix E.1 proves that $\text{E}[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}] = \sigma^2_{\hat{\omega}}$, $\text{E}[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}] = \psi^B_{\hat{\omega}}$, $\text{E}[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}] = \psi^A_{\hat{\omega}}$, and $\text{E}[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}] = \psi^X_{\hat{\omega}}$.

25. Of course, if the researcher expects a certain subset of her data is likely more representative, it would be wise to perform power calculations on this subset alone.

variance is not a function of panel length, it may be tempting to estimate $\tilde{\sigma}^2_{\hat{\omega}}$ using a long vector of residuals, while estimating $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ using only residuals within an $(m+r)$-period range. However, in a time series where the variance-covariance structure is changing, this would produce $\tilde{\psi}_{\hat{\omega}}$ estimates that are inconsistent with $\tilde{\sigma}^2_{\hat{\omega}}$.

Fourth, while $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ are unbiased estimators of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, they are **not** unbiased estimators of σ^2_{ω}, ψ^B, ψ^A, and ψ^X. This is because the residuals from the regression $Y_{it} = \upsilon_i + \delta_t + \omega_{it}$ will have a variance less than the parameter σ^2_{ω} from the data generating process, by the properties of linear projection. In addition, when they are estimated using residuals from shorter panels, $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ have a more severe bias, but these estimates converge to their true values (i.e., σ^2_{ω}, ψ^B, ψ^A, and ψ^X) as the panel length used to estimate these residuals increases.[26] Importantly, for the purposes of power calculations using the SCR formula, we *can* recover an unbiased estimate of the minimum detectable effect with the true parameters using our parameter estimates. That is, $MDE^{est}(\sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}) = MDE(\sigma^2_{\omega}, \psi^B, \psi^A, \psi^X)$. As Appendix E.1 shows, $\mathrm{E}[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}] = \sigma^2_{\hat{\omega}}$, $\mathrm{E}[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}] = \psi^B_{\hat{\omega}}$, $\mathrm{E}[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}] = \psi^A_{\hat{\omega}}$, and $\mathrm{E}[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}] = \psi^X_{\hat{\omega}}$. Combining these two proofs suggests that $MDE^{est}(\mathrm{E}[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}], \mathrm{E}[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}], \mathrm{E}[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}], \mathrm{E}[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}]) = MDE(\sigma^2_{\omega}, \psi^B, \psi^A, \psi^X)$. Therefore, for values of $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ estimated from a pre-existing dataset with I cross-sectional units:

$$\begin{aligned} MDE^{est} &= \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \left\{ \left(\frac{1}{P(1-P)J} \right) \left[\left(\frac{m+r}{mr} \right) k_\sigma \, \mathrm{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right] + \left(\frac{m-1}{m} \right) k_B \, \mathrm{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right] \right. \right. \\ &\qquad \left. \left. + \left(\frac{r-1}{r} \right) k_A \, \mathrm{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right] - 2k_X \, \mathrm{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right] \right] \right\}^{1/2} \qquad \text{(D1)} \\ &= \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \left\{ \left(\frac{1}{P(1-P)J} \right) \left[\left(\frac{m+r}{mr} \right) k_\sigma \sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m} \right) k_B \psi^B_{\hat{\omega}} \right. \right. \\ &\qquad \left. \left. + \left(\frac{r-1}{r} \right) k_A \psi^A_{\hat{\omega}} - 2k_X \psi^X_{\hat{\omega}} \right] \right\}^{1/2} \qquad \text{(D2)} \end{aligned}$$

where

$$\begin{aligned} k_\sigma &= \frac{I(m+r)^2}{2(I-1)mr} \\ k_B &= \frac{I(m+r)^2}{2(I-1)r^2} \\ k_A &= \frac{I(m+r)^2}{2(I-1)m^2} \\ k_X &= 0 \end{aligned}$$

and the expectation of parameters are taken over subsets of the dataset, as described in the next point. Appendix E.2 proves that Equations (D1) and (D2) are equivalent, and derives the above expressions for the coefficients k_σ, k_B, k_A, and k_X.

26. The estimated residuals include both the true idiosyncratic error, ω_{it}, and (attenuating) fixed-effect estimation error. Although both sets of fixed effects, υ_i and δ_t, are unbiased and consistent in T and I, respectively, error in estimating these parameters will always yield residuals that are smaller on average, biasing the estimation of these parameters. The estimation error and resulting biases decrease in T and I.

Fifth, because the estimated variance-covariance terms enter the power calculation under a radical, researchers must be conscious of Jensen's Inequality. If the researcher is estimating $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ by taking the expectation of $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ across a range of $(m+r)$-period subsets, then the correct calculation is:

$$MDE^{est}\left(\mathrm{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right], \mathrm{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right], \mathrm{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right], \mathrm{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right]\right), \text{ not } \mathrm{E}\left[MDE^{est}\left(\tilde{\sigma}^2_{\hat{\omega}}, \tilde{\psi}^B_{\hat{\omega}}, \tilde{\psi}^A_{\hat{\omega}}, \tilde{\psi}^X_{\hat{\omega}}\right) \mid \mathbf{X}\right].$$

Similarly, if Equation (2) is rearranged as a function of κ, it becomes convex in the variance-covariance parameters, and the correct calculation is:

$$\kappa\left(\mathrm{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right], \mathrm{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right], \mathrm{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right], \mathrm{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right]\right), \text{ not } \mathrm{E}\left[\kappa\left(\tilde{\sigma}^2_{\hat{\omega}}, \tilde{\psi}^B_{\hat{\omega}}, \tilde{\psi}^A_{\hat{\omega}}, \tilde{\psi}^X_{\hat{\omega}}\right) \mid \mathbf{X}\right].$$

When solving for sample size J, Equation (2) becomes linear in variance-covariance parameters, meaning that Jensen's Inequality does not affect the estimate of $J\left(\sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}\right)$.

In light of each of these issues, we recommend the following algorithm for estimating the MDE using a pre-existing panel dataset (implemented by our STATA program `pc_dd_covar`):

1. Determine all feasible ranges of experiments with $(m+r)$ periods, given the number of time periods in the pre-existing dataset. For example, if this dataset contains 100 time periods indexed $t = \{1, \ldots, 100\}$, and $m = 5$ and $r = 6$, then there are 90 feasible ranges for an experiment with $(m+r) = 11$ periods (i.e., beginning in periods $t = \{1, \ldots, 90\}$).

2. For each feasible range S:

   (a) Regress the outcome variable on unit and time-period fixed effects, $Y_{it} = \upsilon_i + \delta_t + \omega_{it}$, and store the residuals. (This regression includes all I available cross-sectional units, but only time periods with the specific range S.[27])

   (b) Calculate the variance of the stored residuals, and save as $\tilde{\sigma}^2_{\hat{\omega},S}$.

   (c) For each pair of pre-treatment periods (i.e., the first m periods in range S), calculate the covariance between these periods' residuals. Take an unweighted average of these $m(m-1)/2$ covariances, and save as $\tilde{\psi}^B_{\hat{\omega},S}$.

   For example, if $m = 4$, $r = 2$, and range S begins in period $t = 1$, sum $\mathrm{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$, $\mathrm{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_3)$, $\mathrm{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_4)$, $\mathrm{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_3)$, $\mathrm{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_4)$, and $\mathrm{Cov}(\boldsymbol{\omega}_3, \boldsymbol{\omega}_4)$, and divide by $m(m-1)/2 = 6$.

   (d) For each pair of post-treatment periods (i.e., the last r periods in range S), calculate the covariance between these periods' residuals. Take an unweighted average of these $r(r-1)/2$ covariances, and save as $\tilde{\psi}^A_{\hat{\omega},S}$.

   For example, if $m = 4$, $r = 2$, and range S begins in period $t = 1$, $\tilde{\psi}^A_{\hat{\omega},S}$ is the average of a single post-period covariance, $\mathrm{Cov}(\boldsymbol{\omega}_5, \boldsymbol{\omega}_6)$.

   (e) For each pair of pre- and post-treatment periods (i.e. the first m and the last r periods in range S), calculate the covariance between these periods' residuals. Take an unweighted average of these mr covariances, and save as $\tilde{\psi}^X_{\hat{\omega},S}$.

27. This bears no relationship to the sample size J units to be included in the power calculation. Assuming that all I units in the pre-existing dataset represent the population to be included in the randomization, estimating the variance and covariances using all available units will provide the best estimates of $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ (by the Weak Law of Large Numbers).

For example, if $m = 4$, $r = 2$, and range S begins in period $t = 1$, sum $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_5)$, $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_6)$, $\text{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_5)$, $\text{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_6)$, $\text{Cov}(\boldsymbol{\omega}_3, \boldsymbol{\omega}_5)$, $\text{Cov}(\boldsymbol{\omega}_3, \boldsymbol{\omega}_6)$, $\text{Cov}(\boldsymbol{\omega}_4, \boldsymbol{\omega}_5)$, and $\text{Cov}(\boldsymbol{\omega}_4, \boldsymbol{\omega}_6)$, and divide by $mr = 8$.

3. Calculate the average of $\tilde{\sigma}^2_{\hat{\omega},S}$, $\tilde{\psi}^B_{\hat{\omega},S}$, $\tilde{\psi}^A_{\hat{\omega},S}$, and $\tilde{\psi}^X_{\hat{\omega},S}$ across all ranges S, deflating $\tilde{\sigma}^2_{\hat{\omega},S}$ by $\frac{IT-1}{IT}$, and $\tilde{\psi}^B_{\hat{\omega},S}$, $\tilde{\psi}^A_{\hat{\omega},S}$, and $\tilde{\psi}^X_{\hat{\omega},S}$ by $\frac{I-1}{I}$. These averages are equal in expectation to $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$.

4. Plug these values into Equation (D2) to produce MDE^{est}.

Figure D1 highlights the difference between true and estimated variance-covariance parameters in AR(1) data. In particular, we show true parameter values of σ^2_{ω}, ψ^B, ψ^A, and ψ^X alongside estimated values of these same parameters, calculated according to the procedure outlined above. As expected, $\sigma^2_{\hat{\omega}}$ is biased downwards relative to σ^2_{ω}, but converges towards this value as the panel length increases. This convergence is slower for larger AR(1) parameters, as highly serially correlated errors make it harder to identify the unit fixed effects. Similarly, while the true ψ^X is positive across all panel lengths, $\psi^X_{\hat{\omega}}$ is negative everywhere, and ψ^B and ψ^A also differ from their estimated counterparts. Despite the differences between the true parameters and their estimated values, Appendix E.2 proves that we can recover the MDE based on true underlying parameters using residual-based parameters. In conjunction with the fact that we can estimate the residual-based parameters from real data, this confirms that researchers can use estimated parameters to calibrate power calculations.

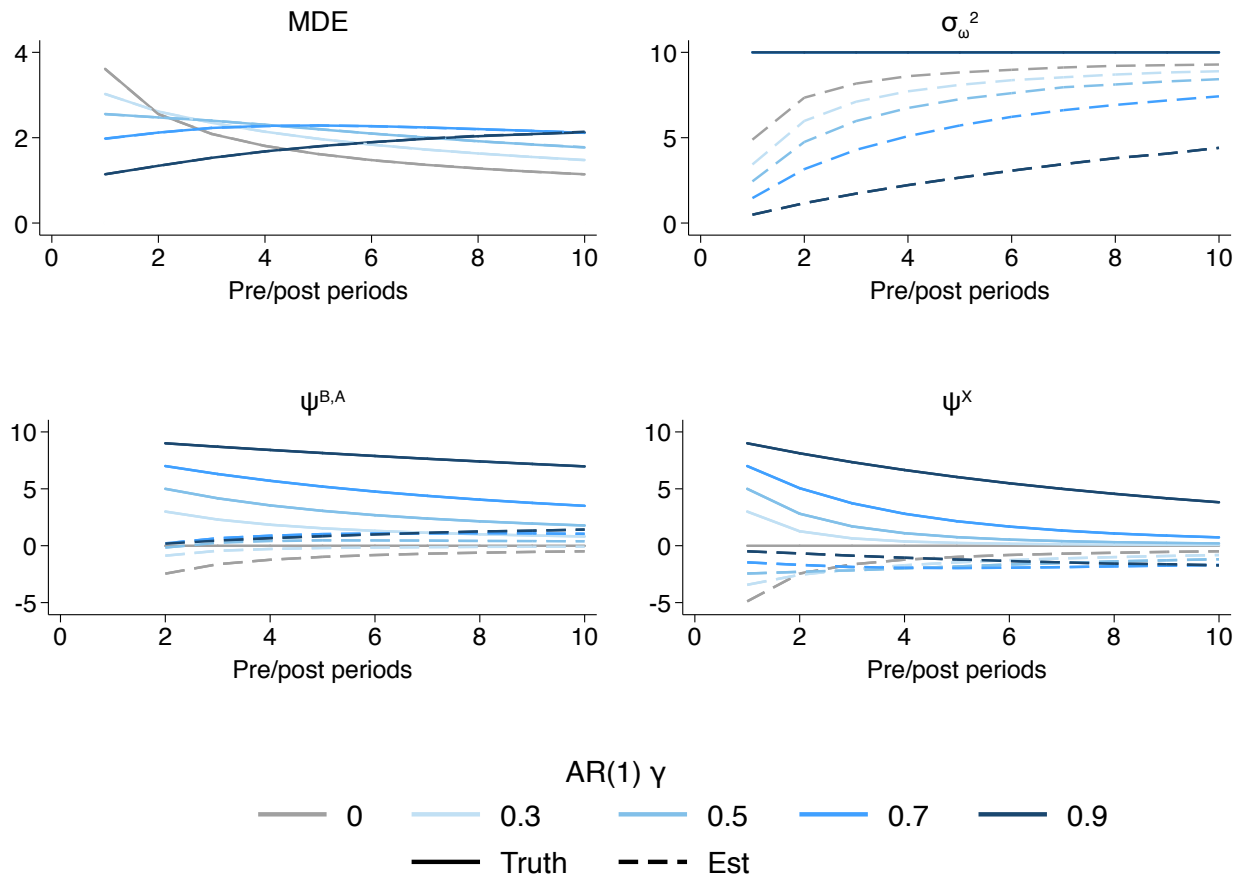
Figure D2 uses the Bloom et al. (2015) dataset to present an analogous comparison between actual vs. estimated σ^2_{ω}, ψ^B, ψ^A, and ψ^X parameters. Here, as in Figure D1, the dotted lines estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ using the above algorithm. However, unlike with simulated AR(1) datasets, the "true" parameters of the Bloom et al. (2015) data generating process are unknown. We estimate these "true" values using residuals from the full 48-period time series, which minimizes the fixed effect estimation error that biases $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ in short panels.[28] This reveals a very similar pattern: "subsetted" $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ estimates are systematically biased downward, but converge to their "full time-series" (i.e. closer to "true") values as panel length increases. As in Figure D1, we show that both sets of estimated variance-covariance parameters yield (virtually) identical MDEs, as long as Equation (2) uses estimated parameters that are internally consistent.

Figure D3 replicates Figure D2 for all four Pecan Street datasets. We see that while the estimated values $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ differ across different levels of aggregation, they follow the same pattern. The subsetted estimates are biased downward, but appear to converge to the full time series estimates (i.e. closer to truth) as panel length increases. In all four cases, the MDE is (virtually) identical when calculated using either all full time series estimates or all subsetted estimates.

Two additional nuances that arise during analytical power calculations are worth noting. First, the critical values $t^d_{1-\kappa}$ and $t^d_{\alpha/2}$ should be drawn from an inverse t-distribution with the same degrees of freedom as the *ex post* regression model. This means that if researchers plan to use CRVE standard errors clustered by experimental unit, they should draw these critical values from

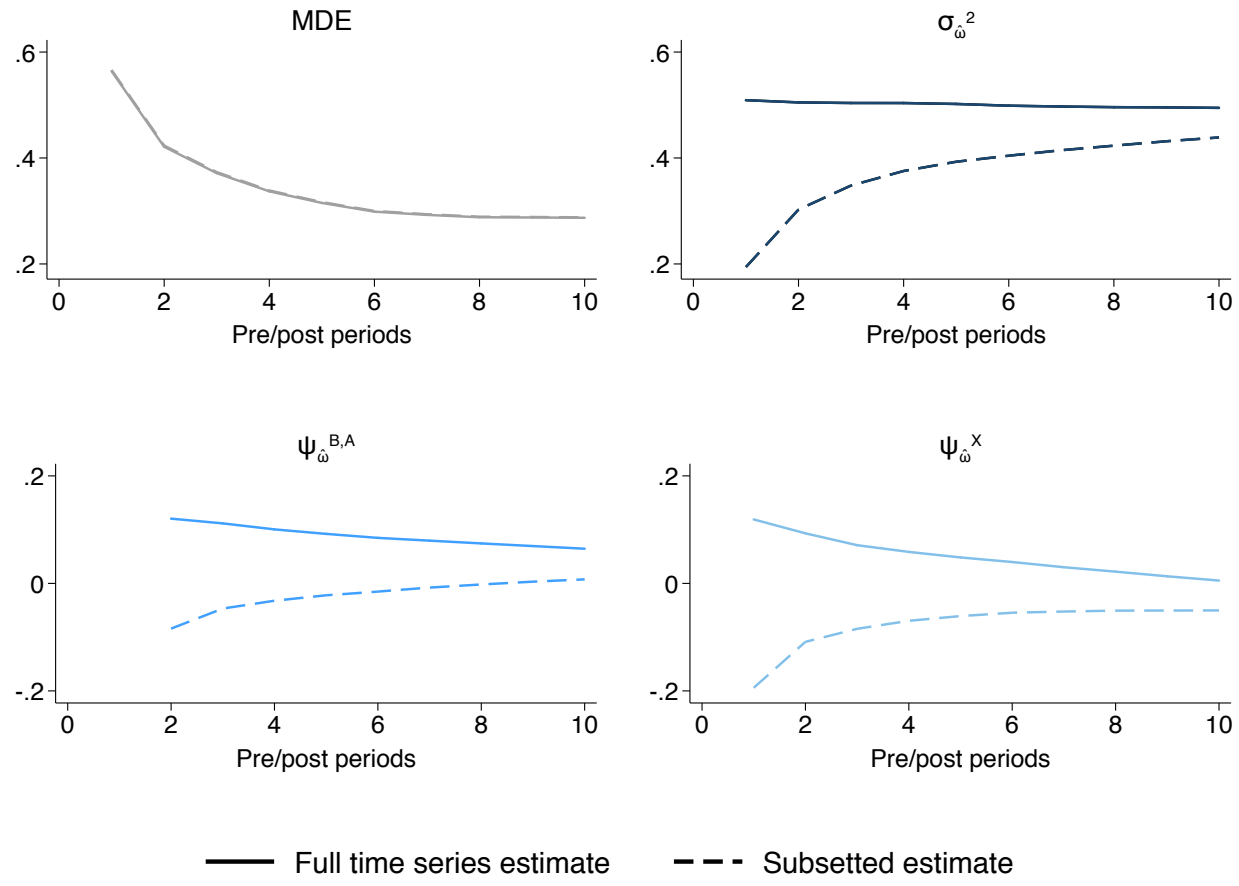
28. These $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ estimates (represented by solid lines in Figure D2) result from the same algorithm as detailed above, except omitting Step 2(a) and estimating a single 48-period set of residuals in Step 1. This provides the closest possible approximation to the "true" variance-covariance structure of these data, and hence the most apples-to-apples comparison to Figure D1.

Figure D1: Actual vs. estimated parameters – AR(1) data



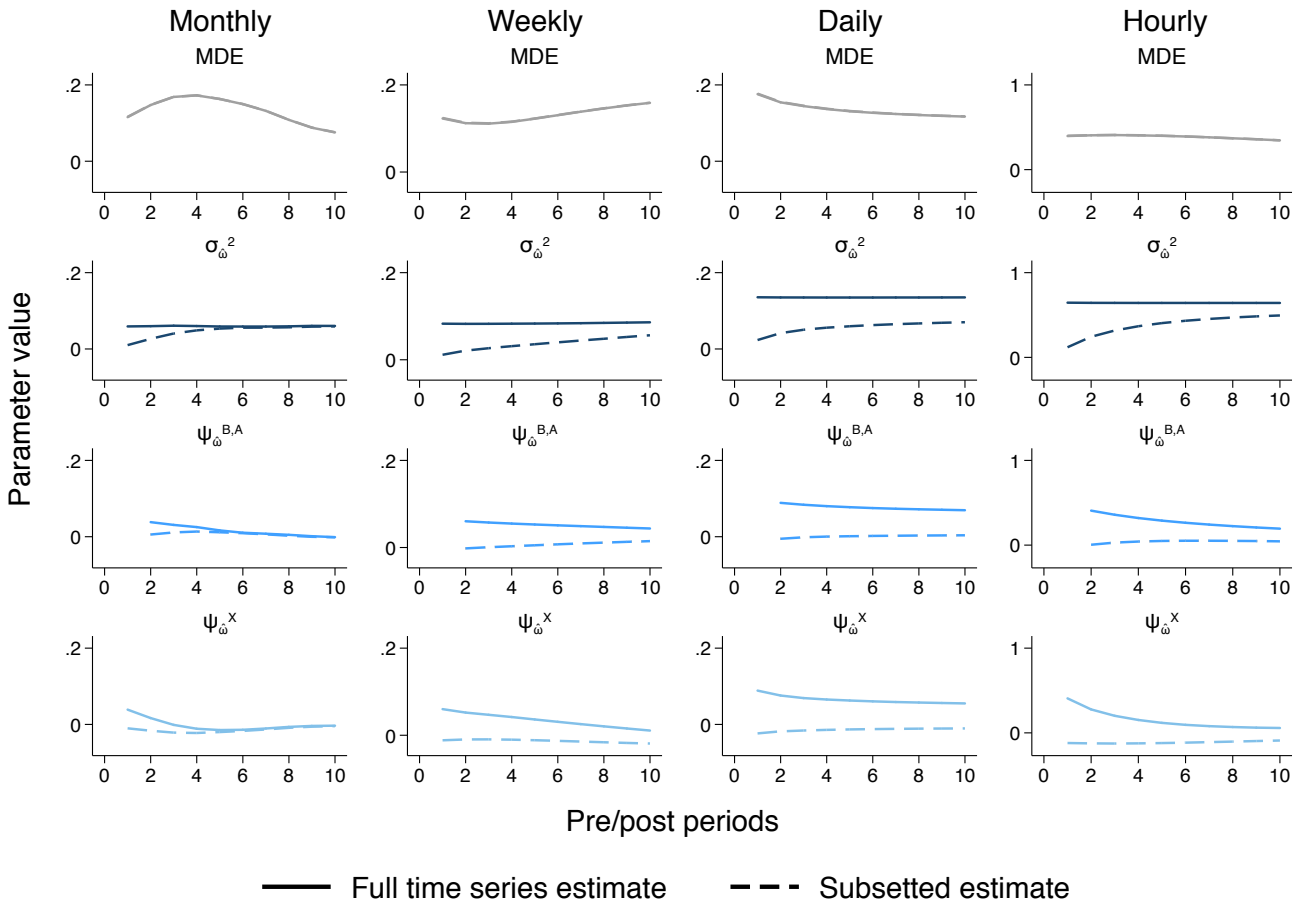
Notes: This figure displays the difference between the true residual variance (σ^2_ω), average pre (post)-period covariance ($\psi^{B,A}$), and average cross-period covariance (ψ^X), and their estimated counterparts over varying panel lengths. It also shows the resulting minimum detectable effect (MDE), calculated using the SCR formula (Equation (2)). These parameters and estimates come from a simulated datasets with AR(1) errors, generated identically to those presented in Figure 1. True parameters are displayed with solid lines (with ψ terms derived analytical using Equations (B1)–(B3)), and estimates are displayed with dashed lines (estimated according to the algorithm described above). As expected, the true σ^2_ω is constant across panel lengths, while $\psi^{B,A}$ declines with the number of pre (post) periods, and ψ^X declines more quickly than $\psi^{B,A}$ as the panel length increases. The higher the AR(1) parameter, the larger are the ψ terms. The estimated parameters behave quite differently from their true counterparts. In short panels, $\sigma^2_{\hat{\omega}}$ is biased downwards, because the regression model's individual fixed effects are inconsistently estimated, and capture more of the true error variance than they should explain in expectation. This has the effect of reducing the estimated covariances $\psi^{B,A}_{\hat{\omega}}$, which scale with $\sigma^2_{\hat{\omega}}$. At the same time, $\psi^X_{\hat{\omega}}$ is mechanically negative, as the estimated fixed effects yield residuals that are negatively correlated within individuals across pre/post-treatment time periods (and some of the variation that should be captured by $\sigma^2_{\hat{\omega}}$ is instead loaded on to the $\psi_{\hat{\omega}}$ terms). As panel length increases, the $\psi_{\hat{\omega}}$ terms converge to the true ψ values. Estimated parameters result in the same MDEs as real parameters after applying the correction factors in Equation (D2), as demonstrated by the top left panel.

Figure D2: Estimated parameters – Bloom et al. (2015) data



Notes: This figure shows two different methods for estimating variance parameters, applied to the Bloom et al. (2015) data, and also depicts the resulting minimum detectable effect (MDE) resulting from both methods. The solid lines plot parameters calculated by running a regression of the outcome variable on unit and time period fixed effects, estimated on the entire time series of data, and then using the residuals from this regression to calculate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ for the average panel of length $m + r$. We then plug these estimates into Equation (2) to calculate the minimum detectable effect. The dashed lines show parameters estimated using the procedure described above, where rather than use residuals from the full time series, we subset the dataset into shorter panels of length $m + r$, calculate the parameters using residuals only from this subset, and average across all possible subsets to arrive at $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. We calculate the MDE by plugging these estimates into Equation (D2). Note that these variance-covariance estimates converge as the panel length increases. Both procedures yield (virtually) identical MDEs, even though the underlying parameter estimates differ substantially.

Figure D3: Estimated parameters – Pecan Street data



Notes: This figure shows two different methods for estimating variance parameters, applied to the Pecan Street data at the four levels of aggregation presented in the main text, and also depicts the resulting minimum detectable effect (MDE) resulting from both methods. Note that the y axis scale differs between the hourly data and the other three datasets; this is because the degree of residual variation left in the hourly data after removing time and individual fixed effects is much greater than in the other datasets. The solid lines show parameters calculated by running a regression of the outcome variable on unit and time period fixed effects, estimated on the entire time series of data, and then using the residuals from this regression to calculate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ for the average panel of length $m + r$. We then plug these estimates into Equation (2) to calculate the minimum detectable effect. By contrast, the dashed lines show parameters estimated using the procedure described above, where rather than use residuals from the full time series, we subset the dataset into shorter panels of length $m + r$, calculate the parameters using residuals only from this subset, and average across all possible subsets to arrive at $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. We calculate the MDE by plugging these estimates into Equation (D2). Note that these variance-covariance estimates converge as the panel length increases. Both procedures yield (virtually) identical MDEs, even though the two procedures' method for estimating the underlying parameters differ substantially.

an inverse t-distribution with J degrees of freedom. To be precise, these critical values should be sensitive to changes in the number of unit/clusters J, although the t degrees of freedom has a very small effect on MDE, relative to other parameters in Equation (2).[29]

Finally, in panel RCTs with CRVE standard errors clustered by unit, the proportion of units treated P cannot be too large or too small. Our simulations have demonstrated that the Equation (2) performs poorly if $P < 0.1$ or $P > 0.9$, because the CRVE requires a sufficient number of clusters that are both treated and control.

D.2 Simulation-based power calculations

In cases where researchers have access to a representative pre-existing dataset, we recommend that they perform power calculations via simulation. This obviates the need to estimate *ex ante* variance-covariance parameters, and it ensures that *ex ante* power calculations assume the same experimental design, regression model, and variance estimator expected to be used in *ex post* analysis. Our accompanying STATA package facilitates simulation-based power calculations using the program `pc_simulate`. This program implements the following algorithm:

1. Choose the following candidate parameters: sample size J, pre-treatment periods m, and post-treatment periods r, treatment ratio P, minimum detectable effect MDE, and significance level α. Let X_{it} denote the outcome variable of interest in the pre-existing dataset.

2. Randomly select J units from the representative dataset, and randomly select a range of $(m + r)$ consecutive time periods. This will serve as a simulated experimental dataset, with sample size J, m pre-treatment periods, and r post-treatment periods.

3. Randomly scramble a $[J \times 1]$ vector of PJ ones and $(1 - P)J$ zeros, rounding PJ to the nearest integer. Assign each of the J units to either treatment ($D = 1$) or control ($D = 0$), based on the order of this scrambled vector.

4. Construct an experimental outcome variable Y_{it}, where $Y_{it} = X_{it} + MDE$ for treated units in post-treatment periods, and $Y_{it} = X_{it}$ otherwise.

5. Using this simulated experimental dataset and the simulated outcome variable Y_{it}, implement the exact regression specification and variance estimator to be used in *ex post* analysis. Record whether this model rejects the null hypotheses of zero treatment effects with significance level α (i.e. $H_0 : \tau = 0$).

6. Repeat Steps 2–5 many times, and calculate the rejection rate across all simulations. This is the experiment's statistical power as a function of J, m, r, P, MDE, and α.

7. Repeat Steps 1–6 for a range of MDEs and design parameters, increasing the number of simulations after narrowing down this range of parameters to more precisely calibrate power.

This algorithm allows users to test alternative regression specifications and alternative standard error assumptions, without needing to formally derive a power calculation expression for each model. If the pre-existing dataset contains fewer cross-sectional units than the desired sample size

29. The STATA program `sampsi` assumes a normal distribution, which is not appropriate for small samples.

J, `pc_simulate` lets users simulate additional units by bootstrapping units with replacement from the existing dataset. Unfortunately, if the pre-existing dataset contains fewer time periods than the desired panel length ($m+r$), an analogous bootstrapping procedure would be much less straightforward (because unlike cross-sectional units, time periods are ordered and have a ordered covariance structure that is not orthogonal to the treatment vector D).

This simulation-based algorithm can only calibrate statistical power κ. Rather than rely on the critical value $t^d_{1-\kappa}$, the algorithm simply estimates realized power as the proportion of simulations where the treatment effect is statistically distinguishable from zero. (By contrast, users may algebraically rearrange (or invert) an analytical power calculation formula to solve for any one of its parameters.) Calibrating simulation-based power calculations for a parameter other than κ necessitates a grid search over candidate parameter values, as described in Step 7 above. For example, to calibrate sample size J by simulation, users may repeat Step 1–6 over a range of candidate J values, narrowing this range (while simultaneously increasing the number of simulations) to calibrate to the desired power.

D.3 Lack of (representative) pre-existing data

To perform accurate *ex ante* power calculations, researchers must either have access to data that is representative (in expectation) of their future experimental data, or be able to parameterize an analytical formula with accurate estimates of the variance and covariance of the error structure. We recommend that simulation-based power calculations (as described above) in cases where researchers have a representative pre-existing dataset with (i) data for the desired outcome (and relevant control variables); (ii) at least as many unique cross-sectional units as the desired experimental sample size; and (iii) a time series at least as long as the desired experimental panel length. Many candidate experiments likely satisfy these criteria, such as when researchers partner with organizations that maintain historical databases on the desired population of experimental subjects.

At the same time, researchers may lack access to representative data *ex ante*. This problem is not unique to panel data, as even the simple cross-sectional power calculation formula (Equation (A6)) hinges on (an estimate of) the variance σ^2_ε. However, power calculations for panel RCT designs require four variance-covariance parameters: σ^2_ω, ψ^B, ψ^A, and ψ^X. While σ^2_ω is fixed in the population, the ψ (and $\psi_{\hat{\omega}}$) terms are endogenous to the panel length of the experiment, which underscores the importance of estimating $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from a representative time series.

In the absence of representative data, researchers may using analytical formulas in conjunction with appropriate sensitivity analyses.[30] Depending on the type of data that *is* available, approximating the parameters σ^2_ω, ψ^B, ψ^A, and ψ^X may be possible. We consider four cases:

1. *Too few units:* If researchers have access to a representative pre-existing dataset with too few cross-sectional units, they may still estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, and apply these values to the (estimation-specific) analytic formula.[31] These variance-covariance parameters do *not* depend on sample size J in the SCR power calculation formula, and estimates of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ derived from residuals are *not* sensitive in expectation to the number of panel units

30. An alternative is to either impose assumptions on some existing dataset or construct a simulated dataset, either of which could be used to conduct power calculations by simulation. However, this process will be much more computationally intensive than simply applying an analytical formula with appropriate parameter sensitivities.

31. Researchers using *estimated* parameters should apply Equation (D2) rather than Equation (2).

J to be used in the experiment.[32] Alternatively, we recommend that researchers bootstrap units by sampling existing units with replacement, and use this expanded dataset (including simulated units) to conduct power calculations by simulation, as described above.[33]

2. *Too few time periods:* If the pre-existing dataset contains too few time periods, researchers may still estimate $\tilde{\sigma}^2_{\hat{\omega}}$ using residuals from a regression with fewer than $m+r$ periods (because σ^2_{ω} does not depend on panel length). However, the ψ terms do depend on panel length, and they cannot be estimated directly from a dataset with fewer than $m+r$ periods. One strategy is to simply estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ using the longest possible panel (i.e., all available time periods in the pre-existing dataset), even if it is shorter than $m+r$ periods. The resulting $\psi_{\hat{\omega}}$ estimates are likely to be upper bounds (in absolute value) on the $\psi_{\hat{\omega}}$ estimates for longer panels, because as the panel length increases, the ψ terms incorporate more covariances between time periods that are further apart (which tend to become less correlated in distance). Another strategy is to attempt to extend the time series for each unit, analogous to the approach of bootstrapping units. As a rule of thumb, researchers often approximate time series data as an AR(k) process with $k \geq \sqrt[3]{T}$, where T is the full time series length. To extend short panels, researchers may estimate this AR(k) process using (residuals from) the existing dataset, and then simulate forward for each unit's outcome realization. Neither of these strategies is perfect, and we recommend conducting appropriate sensitivity analysis in either case.

3. *No data, standard cross-sectional or DD model:* In the complete absence of data, power calculations will be challenging. At the very least, we recommend that researchers search for estimates of the residual variance in the existing literature, noting that panel fixed effects models are likely to yield lower residual variances than cross-sectional models with similar outcome data. If this is not possible, researchers may iterate analytical power calculations over a range of parameter choices. If researchers are able to guess a reasonable value of σ^2_{ω}, they my test a range of AR(1) parameters for plausible values of $\psi^{B,A,X}$. As a rule of thumb, $\psi^{B,A,X}$ are likely to be positive in the absence of a strong prior of negative serial correlation. In absolute value, $\psi^{B,A,X}$ should not exceed σ^2_{ω}, and they should decrease monotonically in panel length. To provide a sense of what plausible values of ψ^B, ψ^A, and ψ^X (and their residual-based counterparts) may be, we plot estimates from a range of panel lengths using simulated AR(1) data, the Bloom et al. (2015) data, and Pecan Street data, in Figures D1, D2, and D3, respectively.

4. *No data, other models:* For more complicated RCT designs (or ANCOVA in the presence of time shocks), analytical power calculation formulas might not align with the desired *ex post* estimating equation. This would render our above Case 3 recommendation inactionable. An alternative strategy (in the absence of data) is to simulate data based on a simple DGP, incorporating AR(1) serial correlation into the idiosyncratic error term. Researchers can perform simulation-based power calculations using these simulated data, iterating over multiple simulated datasets to test for sensitivity to DGP (and AR(1)) assumptions.

32. Estimates of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ are sensitive to the number of cross-sectional units I used to estimate $\hat{\omega}_{it}$, but this is not related to the sample size parameter J. In Equation (D2), the comparative static $d\psi_{\hat{\omega}}/dJ = 0$.

33. The `bootstrap` option in `pc_simulate` does exactly this, drawing units with replacement until reaching the desire sample size.

D.4 STATA package `pcpanel`

To facilitate user implementation of the methods described above, we have built an accompanying STATA package `pcpanel`. This package contains three programs, which we introduce in Section 4.2 of the main text. We describe each function in detail below:

`pc_dd_analytic`

This program performs analytical power calculations using our SCR formula for difference-in-differences (Equation (2)). Users may input any two of the options {`mde` (effect size MDE), `n` (sample size J), `power` (κ)}, and the program will solve Equation (2) for the third option. Users may also adjust four additional experimental design options: treatment ratio `p` (P), number of pre-treatment periods `pre` (m), number of post-treatment periods `post` (r), and false rejection rate `alpha` (α). `pc_dd_analytic` parameterizes the variance (σ^2_ω) and covariance (ψ^B, ψ^A, ψ^X) in Equation (2) in one of two ways.

First, users may allow the subprogram `pc_dd_covar` to nonparametrically estimate the idiosyncratic residual variance/covariance papers. To do this, they input the outcome variable of interest from the dataset in memory (using option `depvar`), as well as variables that identify units and time periods (using options `i` and `t`, respectively). This is our preferred means of parameterizing Equation (2) in practice, using pre-existing data.

Second, users may manually input *either* an idiosyncratic residual variance (option `variance` (σ^2_ω)) *or* an idiosyncratic residual standard deviation (option `sd` (σ_ω)). They may also incorporate non-i.i.d. idiosyncratic errors in 1 of 3 ways: (1) input assumed AR(1) parameter(s) (option `ar1` (γ)) which `pc_dd_analytic` uses to calculate average covariances (ψ^B, ψ^A, ψ^X); (2) input assumed average covariances themselves (option `avgcov`); or (3) input assumed average correlations (option `avgcor`), which the program multiplies by the variance to convert to covariances. Specifying `var` (or `sd`) without specifying `ar1`, `avgcov`, or `avgcor` will cause `pc_dd_analytic` to default to assuming $\psi^B = \psi^A = \psi^X = 0$.[34] For users accustomed to `sampsi`, which accepts the *composite* residual standard deviation and intracluster correlation (σ_ε and ρ, in the notation of McKenzie (2012)), the `pc_dd_analytic` option `var` combines these two terms into a single input (σ^2_ω, in our notation) where $\sigma^2_\omega = \sigma^2_\varepsilon(1-\rho)$.[35]

`pc_dd_covar`

This program nonparametrically estimates $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from an existing dataset (i.e. the dataset in memory), for a given number of pre-treatment and post-treatment periods. Users must input the outcome variable in question, as well as the number or pre-treatment periods (option `pre` (m)), post-treatment periods (option `post` (r)), unit identifier (option `i`) and time period identifier (option `t`). Given these inputs, `pc_dd_covar` operationalizes the procedure outlined in Appendix D.1. Most users will not need to directly apply this program, as it is designed to operate "under the hood" of `pc_dd_analytic`.

34. Given that estimating ψ^B, ψ^A, ψ^X is nontrivial (see Appendix D.1), users should apply the options `avgcov` and `avgcor` with caution.

35. In the absence of time shocks (i.e. $\sigma^2_\delta = 0$, as in McKenzie (2012)), this equality holds. However, if the data generating process includes time shocks, then $\sigma^2_\omega = \sigma^2_\varepsilon(1-\rho_\upsilon-\rho_\delta)$, where ρ_υ is the intracluster correlation within units and ρ_δ is the intracluster correlation within time periods.

`pc_simulate`

This program performs power calculations by simulation, following the algorthim in Appendix D.2. it supports four types of RCTs:

- `ONESHOT` (one wave of post-treatment data)
- `POST` (multiple waves of post-treatment data)
- `DD` (pre-treatment and post-treatment data)
- `ANCOVA` (post-treatment data, conditioning on pre-treatment data)

Users may input six basic experimental design options: effect size `mde` (MDE), sample size `n` (J), treatment ratio `p` (P), number of pre-treatment periods `pre` (m), number of post-treatment periods `post` (r), and false rejection rate `alpha` (α). Given these options, `pc_simulate` calculates the average rejection rate (i.e. power (κ)) of the treatment effect estimator, across all simulations.

The program is sufficiently flexible to allow for linear controls (option `controls`), fixed effects (options `absorb`, `absorbfactor`), regression weights (option `weight`) and standard errors (option `vce`). Each of these options, when specified, is passed through the regression function `reghdfe` in each simulation.[36] Toggling the option `bootstrap` causes the program to sample J units with replacement for each simulation, which facilitates power calculations with a greater number of units than are present in the dataset in memory. If `bootstrap` is not toggled, each simulation samples J units without replacement. Users may also restrict the range of time periods over which simulations occur (option `tstart`), and the number of simulations for each set of parameters (option `nsim`).[37]

Instead of applying the CRVE, Bertrand, Duflo, and Mullainathan (2004) also recommend collapsing panel datasets to a single pre-treatment and post-treatment observation. `pc_simulate` lets users conduct power calculations on collapsed panel estimators, by toggling the option `collapse`. For `POST` and `ANCOVA` models, each simulation collapses the estimating equation to a cross-sectional model with one observation per unit (where the dependent variable is averaged within units across all r time periods). For the `DD` model, each simulation collapses the estimating equation to a two-period panel, with one pre-treatment and one post-treatment observation per unit (where the dependent variable is averaged within units across all m or r time periods).

`pc_simulate` supports stratified randomization via the option `stratify`. When users specify one or more categorical variables (e.g. `stratify(gender race)`), the program separately randomizes P units within each `gender-race` cell. For stratified randomization, the option `n` governs the number of units within each *cell* (as opposed to the total number of units in the full sample). The program continues to estimate a single *pooled* average treatment effect. If researchers are interested in heterogeneous treatment effects by cell, they can iterate `pc_simulate` separately for each stratification cell—leaving `stratify` unspecified, and conditioning on stratification cells using the option `if` (e.g. `pc_simulate y if gender=="male" & race=="white", ...`).

`pc_simulate` also supports cluster randomization via the option `idcluster`. When users specify a group identifier (e.g. `idcluster(village_id)`), the program randomizes at the group level, where whole groups are randomly assigned to treatment/control. For cluster randomization, the

36. `pc_simulate` simulates regressions using the function `reghdfe` (`http://scorreia.com/software/reghdfe/`).

37. The program defaults users to 500 simulations. Increasing the number of simulations will improve accuracy at the expense of runtime.

option `n` governs the number of *groups* (e.g. villages), as opposed to the total number of individual units (e.g. households). The program continues to estimate a single *pooled* average treatment effect at the unit level. `pc_simulate` also supports two cluster randomization suboptions. The suboption `sizecluster` governs the number of units sampled within each cluster; if not specified, the program defaults to the size of each cluster in the existing dataset. The suboption `pcluster` governs the intensity of treatment within treated clusters: (i) if `pcluster` is not specified, the program defaults to a treatment intensity of 1 in all treated clusters (i.e. all units in treated clusters receive treatment); (ii) if `pcluster` is specified with a single value (e.g. `pcluster(0.5)`), the program randomly assigns treatment to this proportion (i.e. 50 percent) of units within each treated cluster (all units in control clusters remain untreated); (iii) if `pcluster` is specified with multiple values (e.g. `pcluster(0.33 0.67 1)`), the program varies the intensity of treatment in equal proportions across treated clusters (i.e. one third of treated clusters receive `pcluster(0.33)`, etc.).

E Estimation-related proofs

In this section, we prove that researchers may calculate unbiased power calculations by estimating the variance-covariance parameters from a real dataset, where the parameters governing the data generating process is unknown.

E.1 Recovering estimated parameters

Here, we demonstrate that the procedure described in Appendix D.1 recovers unbiased estimates of the variance and covariance parameters governing the residuals $\hat{\omega}_{it}$ from a regression of Y_{it} on unit and time fixed effects, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ (though these do *not* represent unbiased estimates of the *true* parameters σ^2_{ω}, ψ^B, ψ^A, and ψ^X). We denote our procedure for computing these parameters with a $\sim$. Note that throughout this section, we are considering I units in the sample used to estimate $\hat{\omega}_{it}$, which may be distinct from the sample size J units used in the ensuing power calculations. Note also that because we are estimating the variance and covariance of a *population* of residuals, we use the population variance/covariance estimators as opposed to the (unbiased) sample variance/covariance estimators.[38]

In order to estimate the variance of the residuals, $\sigma^2_{\hat{\omega}}$, we define:

$$\tilde{\sigma}^2_{\hat{\omega}} \equiv \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \left(\hat{\omega}_{it} - \overline{\overline{\hat{\omega}}} \right)^2$$

where $\overline{\overline{\hat{\omega}}} = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \hat{\omega}_{it} = 0$. Taking expectations of both sides:

$$\begin{aligned} \mathrm{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X} \right] &= \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \mathrm{E}\left[\hat{\omega}^2_{it} \mid \mathbf{X} \right] \\ &= \sigma^2_{\hat{\omega}} \end{aligned}$$

To estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, we define the $[I \times 1]$ vector of residuals for period t as $\hat{\boldsymbol{\omega}}_t$ and calculate the average covariance between any two vectors in the relevant range of time periods. For the pre-period, we define:

$$\begin{aligned} \tilde{\psi}^B_{\hat{\omega}} &\equiv \frac{2}{Im(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \sum_{i=1}^{I} \left(\hat{\omega}_{it} - \overline{\hat{\omega}}_t \right) \left(\hat{\omega}_{is} - \overline{\hat{\omega}}_s \right) \\ &= \frac{2}{Im(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \sum_{i=1}^{I} \left(\hat{\omega}_{it}\hat{\omega}_{is} - \hat{\omega}_{it}\overline{\hat{\omega}}_s - \hat{\omega}_{is}\overline{\hat{\omega}}_t + \overline{\hat{\omega}}_t\overline{\hat{\omega}}_s \right) \end{aligned}$$

38. This means that to calculate $\tilde{\sigma}^2_{\hat{\omega}}$, we deflate the sample variance estimate by $\frac{IT-1}{IT}$, and to calculate the $\tilde{\psi}_{\hat{\omega}}$ terms, we deflate the sample covariance estimates by $\frac{I-1}{I}$. This distinction is ultimately innocuous, and the following derivations simply rely on a consistent decision to use either the population or sample variance/covariance estimators.

where $\overline{\hat{\omega}}_t = \frac{1}{I}\sum_{i=1}^{I}\hat{\omega}_{it} = 0$. Taking expectations yields:

$$\begin{aligned}
\mathrm{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right] &= \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\mathrm{E}\left[\hat{\omega}_{it}\hat{\omega}_{is} \mid \mathbf{X}\right] \\
&= \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\mathrm{Cov}(\hat{\omega}_{it},\hat{\omega}_{is} \mid \mathbf{X}) \\
&= \psi^B_{\hat{\omega}}
\end{aligned}$$

Similarly:

$$\begin{aligned}
\tilde{\psi}^A_{\hat{\omega}} &\equiv \frac{2}{Ir(r-1)}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\sum_{i=1}^{I}\left(\hat{\omega}_{it}-\overline{\hat{\omega}}_t\right)\left(\hat{\omega}_{is}-\overline{\hat{\omega}}_s\right) \\
&= \frac{2}{Ir(r-1)}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\sum_{i=1}^{I}\left(\hat{\omega}_{it}\hat{\omega}_{is}-\hat{\omega}_{it}\overline{\hat{\omega}}_s-\hat{\omega}_{is}\overline{\hat{\omega}}_t+\overline{\hat{\omega}}_t\overline{\hat{\omega}}_s\right)
\end{aligned}$$

and therefore:

$$\mathrm{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right] = \psi^A_{\hat{\omega}}$$

Applying the same steps to $\psi^X_{\hat{\omega}}$:

$$\begin{aligned}
\tilde{\psi}^X_{\hat{\omega}} &\equiv \frac{1}{Imr}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\sum_{i=1}^{I}\left(\hat{\omega}_{it}-\overline{\hat{\omega}}_t\right)\left(\hat{\omega}_{is}-\overline{\hat{\omega}}_s\right) \\
&= \frac{1}{Imr}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\sum_{i=1}^{I}\left(\hat{\omega}_{it}\hat{\omega}_{is}-\hat{\omega}_{it}\overline{\hat{\omega}}_s-\hat{\omega}_{is}\overline{\hat{\omega}}_t+\overline{\hat{\omega}}_t\overline{\hat{\omega}}_s\right)
\end{aligned}$$

Taking expectations of both sides:

$$\begin{aligned}
\mathrm{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right] &= \frac{1}{Imr}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\sum_{i=1}^{I}\mathrm{E}\left[\hat{\omega}_{it}\hat{\omega}_{is} \mid \mathbf{X}\right] \\
&= \frac{1}{Imr}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\sum_{i=1}^{I}\mathrm{Cov}\left(\hat{\omega}_{it},\hat{\omega}_{is} \mid \mathbf{X}\right) \\
&= \psi^X_{\hat{\omega}}
\end{aligned}$$

Hence, we can recover unbiased estimates of the parameters $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ (defined over residuals $\hat{\omega}_{it}$, rather than errors ω_{it}) by calculating the averages of the estimated $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$, respectively.

E.2 Estimating MDE from residual-based parameters

To calculate the MDE using the SCR formula, we must know the true parameters that characterize the variance and covariance of the error structure, σ^2_ω, ψ^B, ψ^A, and ψ^X. We cannot calculate these parameters directly from a real dataset, however, because we do not observe the true error structure or data generating process. Instead, we estimate a residual for each observation and calculate the residual-based analogs of these parameters, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. In this section, we derive an expression for MDE^{est} in terms of these residual-based parameters that is equivalent to MDE from the SCR formula as defined in terms of true variance-covariance parameters:

$$MDE^{est}\left(\sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}\right) = MDE\left(\sigma^2_\omega, \psi^B, \psi^A, \psi^X\right)$$

Model While estimating the variance and covariance parameters of a dataset does not require a treatment, we assume that all other features of this model are identical to the model that generates the serial-correlation-robust power calculation formula, Equation (2).

That is, there are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods. We restate the remaining assumptions from Appendix A.2.2 here for convenience:

Assumption (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\upsilon)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\delta)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma^2_\omega)$.

Assumption (Strict exogeneity)**.** $\mathrm{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J-1$ *unit dummies, and* $(m+r)-1$ *time dummies. This again follows from random assignment of* D_{it}*.*

Assumption (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption (Independence across units)**.** $\mathrm{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ \forall\ t, s.$

Assumption (Symmetric covariance structures)**.** *Define:*

$$\psi^B \equiv \frac{2}{Jm(m-1)} \sum_{i=1}^{J} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^A \equiv \frac{2}{Jr(r-1)} \sum_{i=1}^{J} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \mathrm{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^X \equiv \frac{1}{Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of the same unit, respectively. Define ψ_T^B, ψ_T^A, *and* ψ_C^X *analogously, where we consider only the* PJ *treated units; also define* ψ_C^B, ψ_C^A, *and* ψ_C^X *analogously, where we consider only the* $(1-P)J$ *control units. Using these definitions, assume that* $\psi^B = \psi_T^B = \psi_C^B$; $\psi^A = \psi_T^A = \psi_C^A$; *and* $\psi^X = \psi_T^X = \psi_C^X$.

Parameters We first need to estimate the residuals of this model. To do this, we regress Y_{it} on a constant and fixed effects at the unit and time levels. For a balanced panel, the estimated coefficients are

$$\begin{aligned}
\hat{\beta} &= \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it} \\
\hat{\upsilon}_i &= \frac{1}{T} \sum_{t=-m+1}^{r} Y_{it} - \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it} \\
\hat{\delta}_t &= \frac{1}{I} \sum_{i=1}^{I} Y_{it} - \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it}
\end{aligned}$$

Then the residual is defined as

$$\begin{aligned}
\hat{\omega}_{it} &= Y_{it} - \hat{Y}_{it} \\
&= (\beta + \upsilon_i + \delta_t + \omega_{it}) - \left(\hat{\beta} + \hat{\upsilon}_i + \hat{\delta}_t\right) \\
&= \omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}}
\end{aligned}$$

where

$$\begin{aligned}
\overline{\omega}_i &= \frac{1}{T} \sum_{t=-m+1}^{r} \omega_{it} \\
\overline{\omega}_t &= \frac{1}{I} \sum_{i=1}^{I} \omega_{it} \\
\overline{\overline{\omega}} &= \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \omega_{it}
\end{aligned}$$

We can now use this definition of residuals to derive expressions for $\sigma_{\hat{\omega}}^2$, $\psi_{\hat{\omega}}^B$, $\psi_{\hat{\omega}}^A$, and $\psi_{\hat{\omega}}^X$. We first derive an expression for $\sigma_{\hat{\omega}}^2$, the average variance of a residual:

$$\sigma_{\hat{\omega}}^2 = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \text{Var}(\hat{\omega}_{it} \mid \mathbf{X})$$

$$
\begin{aligned}
&= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \Big[\operatorname{Var}(\omega_{it} \mid \mathbf{X}) + \operatorname{Var}\left(\overline{\omega}_i \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{\omega}_t \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{\overline{\omega}} \mid \mathbf{X}\right) \\
&\qquad\qquad - 2\operatorname{Cov}\left(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\omega_{it}, \overline{\omega}_t \mid \mathbf{X}\right) + 2\operatorname{Cov}\left(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}\right) \\
&\qquad\qquad + 2\operatorname{Cov}\left(\overline{\omega}_i, \overline{\omega}_t \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}\right) \Big]
\end{aligned}
$$

Calculating each of these terms, in turn, gives

$$
\begin{aligned}
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}(\omega_{it} \mid \mathbf{X}) &= \sigma_\omega^2 \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\overline{\omega}_i \mid \mathbf{X}\right) &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is} \mid \mathbf{X}\right) \\
&= \frac{1}{T}\sigma_\omega^2 + \frac{m(m-1)}{T^2}\psi^B + \frac{r(r-1)}{T^2}\psi^A + \frac{2mr}{T^2}\psi^X \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\overline{\omega}_t \mid \mathbf{X}\right) &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt} \mid \mathbf{X}\right) \\
&= \frac{1}{I}\sigma_\omega^2 \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\overline{\overline{\omega}} \mid \mathbf{X}\right) &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{s=-m+1}^{r}\omega_{js} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}\right) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Cov}\left(\omega_{it}, \frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is} \mid \mathbf{X}\right) \\
&= \frac{2}{T}\sigma_\omega^2 + \frac{2m(m-1)}{T^2}\psi^B + \frac{2r(r-1)}{T^2}\psi^A + \frac{4mr}{T^2}\psi^X \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\omega_{it}, \overline{\omega}_t \mid \mathbf{X}\right) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Cov}\left(\omega_{it}, \frac{1}{I}\sum_{j=1}^{I}\omega_{jt} \mid \mathbf{X}\right) \\
&= \frac{2}{I}\sigma_\omega^2 \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}\right) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Cov}\left(\omega_{it}, \frac{1}{IT}\sum_{j=1}^{I}\sum_{s=-m+1}^{r}\omega_{js} \mid \mathbf{X}\right) \\
&= \frac{2}{IT}\sigma_\omega^2 + \frac{2m(m-1)}{IT^2}\psi^B + \frac{2r(r-1)}{IT^2}\psi^A + \frac{4mr}{IT^2}\psi^X \\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\overline{\omega}_i, \overline{\omega}_t \mid \mathbf{X}\right) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Cov}\left(\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is}, \frac{1}{I}\sum_{j=1}^{I}\omega_{jt} \mid \mathbf{X}\right) \\
&= \frac{2}{IT}\sigma_\omega^2 + \frac{2m(m-1)}{IT^2}\psi^B + \frac{2r(r-1)}{IT^2}\psi^A + \frac{4mr}{IT^2}\psi^X
\end{aligned}
$$

$$\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}\right) = \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Cov}\left(\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is}, \frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp} \mid \mathbf{X}\right)$$
$$= \frac{2}{IT}\sigma_\omega^2 + \frac{2m(m-1)}{IT^2}\psi^B + \frac{2r(r-1)}{IT^2}\psi^A + \frac{4mr}{IT^2}\psi^X$$
$$\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}\right) = \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}, \frac{1}{IT}\sum_{k=1}^{I}\sum_{s=-m+1}^{r}\omega_{ks} \mid \mathbf{X}\right)$$
$$= \frac{2}{IT}\sigma_\omega^2 + \frac{2m(m-1)}{IT^2}\psi^B + \frac{2r(r-1)}{IT^2}\psi^A + \frac{4mr}{IT^2}\psi^X$$

Combining these terms and simplifying yields

$$\sigma_{\hat{\omega}}^2 = \left(\tfrac{(I-1)(T-1)}{IT}\right)\sigma_\omega^2 - \left(\tfrac{(I-1)m(m-1)}{IT^2}\right)\psi^B - \left(\tfrac{(I-1)r(r-1)}{IT^2}\right)\psi^A - \left(\tfrac{2(I-1)mr}{IT^2}\right)\psi^X \tag{E1}$$

We next derive an expression for $\psi_{\hat{\omega}}^A$:

$$\begin{aligned}
\psi_{\hat{\omega}}^A &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}}, \omega_{is} - \overline{\omega}_i - \overline{\omega}_s + \overline{\overline{\omega}} \mid \mathbf{X}) \\
&= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\Big[\operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \operatorname{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) - \operatorname{Cov}(\omega_{it}, \overline{\omega}_s \mid \mathbf{X}) + \operatorname{Cov}(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad - \operatorname{Cov}(\overline{\omega}_i, \omega_{is} \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i, \overline{\omega}_i \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i, \overline{\omega}_s \mid \mathbf{X}) - \operatorname{Cov}(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad - \operatorname{Cov}(\overline{\omega}_t, \omega_{is} \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t, \overline{\omega}_i \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t, \overline{\omega}_s \mid \mathbf{X}) - \operatorname{Cov}(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad + \operatorname{Cov}(\overline{\overline{\omega}}, \omega_{is} \mid \mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\omega}_i \mid \mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\omega}_s \mid \mathbf{X}) + \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\overline{\omega}} \mid \mathbf{X})\Big]
\end{aligned}$$

We again calculate each term:

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) = \psi^A$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) = \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it}, \frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip} \mid \mathbf{X}\right)$$
$$= \frac{2}{r(r-1)IT}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it}, \omega_{ip} \mid \mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \overline{\omega}_s \mid \mathbf{X}) = \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it}, \frac{1}{I}\sum_{j=1}^{I}\omega_{js} \mid \mathbf{X}\right)$$
$$= \frac{1}{I}\psi^A$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it},\overline{\overline{\omega}}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it},\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\omega_{is}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\omega_{is}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)IT}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}\left(\omega_{it},\omega_{ip}\mid\mathbf{X}\right)$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_i\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq}\mid\mathbf{X}\right)$$
$$=\frac{1}{T}\sigma_\omega^2+\frac{m(m-1)}{T^2}\psi^B+\frac{r(r-1)}{T^2}\psi^A+\frac{2mr}{T^2}\psi^X$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_s\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{I}\sum_{j=1}^{I}\omega_{js}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\overline{\overline{\omega}}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{IT}\sum_{j=1}^{I}\sum_{q=-m+1}^{r}\omega_{jq}\mid\mathbf{X}\right)$$
$$=\frac{1}{IT}\sigma_\omega^2+\frac{m(m-1)}{IT^2}\psi^B+\frac{r(r-1)}{IT^2}\psi^A+\frac{2mr}{IT^2}\psi^X$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\omega_{is}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\omega_{is}\mid\mathbf{X}\right)$$
$$=\frac{1}{I}\psi^A$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\overline{\omega}_i\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\overline{\omega}_s\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{I}\sum_{k=1}^{I}\omega_{ks}\mid\mathbf{X}\right)$$
$$=\frac{1}{I}\psi^A$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\overline{\overline{\omega}}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{IT}\sum_{k=1}^{I}\sum_{p=-m+1}^{r}\omega_{kp}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\omega_{is}\mid\mathbf{X}) = \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\omega_{is}\mid\mathbf{X}\right)$$
$$= \frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\overline{\omega}_{i}\mid\mathbf{X}) = \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq}\mid\mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\overline{\omega}_{s}\mid\mathbf{X}) = \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{I}\sum_{k=1}^{I}\omega_{ks}\mid\mathbf{X}\right)$$
$$= \frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\overline{\overline{\omega}}\mid\mathbf{X}) = \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\right.$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X$$

Combining these terms and simplifying yields

$$\psi_{\hat{\omega}}^A = -\left(\tfrac{I-1}{IT}\right)\sigma_\omega^2 + \left(\tfrac{(I-1)m(m-1)}{IT^2}\right)\psi^B + \left(\tfrac{(I-1)(m^2+2m+r)}{IT^2}\right)\psi^A - \left(\tfrac{2(I-1)m^2}{IT^2}\right)\psi^X \tag{E2}$$

By symmetry, the expression for $\psi_{\hat{\omega}}^B$ is

$$\psi_{\hat{\omega}}^B = -\left(\tfrac{I-1}{IT}\right)\sigma_\omega^2 + \left(\tfrac{(I-1)(r^2+2r+m)}{IT^2}\right)\psi^B + \left(\tfrac{(I-1)r(r-1)}{IT^2}\right)\psi^A - \left(\tfrac{2(I-1)r^2}{IT^2}\right)\psi^X \tag{E3}$$

We finally derive an expression for $\psi_{\hat{\omega}}^X$:

$$\psi_{\hat{\omega}}^X = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\operatorname{Cov}(\hat{\omega}_{it},\hat{\omega}_{is}\mid\mathbf{X})$$
$$= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\operatorname{Cov}(\omega_{it}-\overline{\omega}_i-\overline{\omega}_t+\overline{\overline{\omega}},\omega_{is}-\overline{\omega}_i-\overline{\omega}_s+\overline{\overline{\omega}}\mid\mathbf{X})$$
$$= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\Big[\operatorname{Cov}(\omega_{it},\omega_{is}\mid\mathbf{X}) - \operatorname{Cov}(\omega_{it},\overline{\omega}_i\mid\mathbf{X}) - \operatorname{Cov}(\omega_{it},\overline{\omega}_s\mid\mathbf{X}) + \operatorname{Cov}(\omega_{it},\overline{\overline{\omega}}\mid\mathbf{X})$$
$$- \operatorname{Cov}(\overline{\omega}_i,\omega_{is}\mid\mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_i\mid\mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_s\mid\mathbf{X}) - \operatorname{Cov}(\overline{\omega}_i,\overline{\overline{\omega}}\mid\mathbf{X})$$
$$- \operatorname{Cov}(\overline{\omega}_t,\omega_{is}\mid\mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t,\overline{\omega}_i\mid\mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t,\overline{\omega}_s\mid\mathbf{X}) - \operatorname{Cov}(\overline{\omega}_t,\overline{\overline{\omega}}\mid\mathbf{X})$$
$$+ \operatorname{Cov}(\overline{\overline{\omega}},\omega_{is}\mid\mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}},\overline{\omega}_i\mid\mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}},\overline{\omega}_s\mid\mathbf{X}) + \operatorname{Cov}(\overline{\overline{\omega}},\overline{\overline{\omega}}\mid\mathbf{X})\Big]$$

We again calculate each term:

$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\omega_{is},\omega_{it}\mid\mathbf{X})=\psi^X$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\omega_{it},\overline{\omega}_i\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it},\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}\mid\mathbf{X}\right)\\&=\frac{1}{T}\sigma_\omega^2+\frac{m-1}{T}\psi^B+\frac{r}{T}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\omega_{it},\overline{\omega}_s\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it},\frac{1}{I}\sum_{j=1}^{I}\omega_{js}\mid\mathbf{X}\right)\\&=\frac{1}{I}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\omega_{it},\overline{\overline{\omega}}\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it},\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}\mid\mathbf{X}\right)\\&=\frac{1}{IT}\sigma_\omega^2+\frac{m-1}{IT}\psi^B+\frac{r}{IT}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i,\omega_{is}\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\omega_{is}\mid\mathbf{X}\right)\\&=\frac{1}{T}\sigma_\omega^2+\frac{r-1}{T}\psi^A+\frac{m}{T}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i,\overline{\omega}_i\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq}\mid\mathbf{X}\right)\\&=\frac{1}{T}\sigma_\omega^2+\frac{m(m-1)}{T^2}\psi^B+\frac{r(r-1)}{T^2}\psi^A+\frac{2mr}{T^2}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i,\overline{\omega}_s\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{I}\sum_{j=1}^{I}\omega_{js}\mid\mathbf{X}\right)\\&=\frac{1}{IT}\sigma_\omega^2+\frac{r-1}{IT}\psi^A+\frac{m}{IT}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i,\overline{\overline{\omega}}\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{IT}\sum_{j=1}^{I}\sum_{q=-m+1}^{r}\omega_{jq}\mid\mathbf{X}\right)\\&=\frac{1}{IT}\sigma_\omega^2+\frac{m(m-1)}{IT^2}\psi^B+\frac{r(r-1)}{IT^2}\psi^A+\frac{2mr}{IT^2}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_t,\omega_{is}\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\omega_{is}\mid\mathbf{X}\right)\\&=\frac{1}{I}\psi^X\end{aligned}$$

$$\begin{aligned}\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_t,\overline{\omega}_i\mid\mathbf{X})&=\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}\mid\mathbf{X}\right)\\&=\frac{1}{IT}\sigma_\omega^2+\frac{m-1}{IT}\psi^B+\frac{r}{IT}\psi^X\end{aligned}$$

$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\omega}_t,\overline{\omega}_s \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{I}\sum_{k=1}^{I}\omega_{ks} \mid \mathbf{X}\right)$$
$$= \frac{1}{I}\psi^X$$
$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\omega}_t,\overline{\overline{\omega}} \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{IT}\sum_{k=1}^{I}\sum_{p=-m+1}\omega_{kp} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m-1}{IT}\psi^B + \frac{r}{IT}\psi^X$$
$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\overline{\omega}},\omega_{is} \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\omega_{is} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{r-1}{IT}\psi^A + \frac{m}{IT}\psi^X$$
$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\overline{\omega}},\overline{\omega}_i \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X$$
$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\overline{\omega}},\overline{\omega}_s \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{I}\sum_{k=1}^{I}\omega_{ks} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{r-1}{IT}\psi^A + \frac{m}{IT}\psi^X$$
$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\overline{\omega}},\overline{\overline{\omega}} \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{IT}\sum_{k=1}^{I}\sum_{q=-m+1}^{r}\omega_{kq} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X$$

Combining these terms and simplifying yields

$$\psi_{\hat{\omega}}^X = -\left(\tfrac{I-1}{IT}\right)\sigma_\omega^2 - \left(\tfrac{(I-1)r(m-1)}{IT^2}\right)\psi^B - \left(\tfrac{(I-1)m(r-1)}{IT^2}\right)\psi^A + \left(\tfrac{2(I-1)mr}{IT^2}\right)\psi^X \tag{E4}$$

To summarize, Equations (E1), (E2), (E3), and (E4) express the residual-based parameters as functions of the true parameters that define the error structure. Rearranging these four equations:

$$\sigma_{\hat{\omega}}^2 = \left(\frac{I-1}{IT^2}\right)\left(T(T-1)\sigma_\omega^2 - m(m-1)\psi^B - r(r-1)\psi^A - 2mr\psi^X\right)$$
$$\psi_{\hat{\omega}}^B = \left(\frac{I-1}{IT^2}\right)\left(-T\sigma_\omega^2 + (r^2+2r+m)\psi^B + r(r-1)\psi^A - 2r^2\psi^X\right)$$
$$\psi_{\hat{\omega}}^A = \left(\frac{I-1}{IT^2}\right)\left(-T\sigma_\omega^2 + m(m-1)\psi^B + (m^2+2m+r)\psi^A - 2m^2\psi^X\right)$$
$$\psi_{\hat{\omega}}^X = \left(\frac{I-1}{IT^2}\right)\left(-T\sigma_\omega^2 - r(m-1)\psi^B - m(r-1)\psi^A + 2mr\psi^X\right)$$

In matrix notation, these four equations become:

$$\begin{bmatrix} \sigma^2_{\hat{\omega}} \\ \psi^B_{\hat{\omega}} \\ \psi^A_{\hat{\omega}} \\ \psi^X_{\hat{\omega}} \end{bmatrix} = \mathbf{\Gamma} \begin{bmatrix} \sigma^2_{\omega} \\ \psi^B \\ \psi^A \\ \psi^X \end{bmatrix} \tag{E5}$$

where

$$\mathbf{\Gamma} = \frac{I-1}{IT^2} \begin{bmatrix} T(T-1) & -m(m-1) & -r(r-1) & -2mr \\ -T & r^2+2r+m & r(r-1) & -2r^2 \\ -T & m(m-1) & m^2+2m+r & -2m^2 \\ -T & -r(m-1) & -m(r-1) & 2mr \end{bmatrix}$$

Minimum detectable effect We are ultimately interested in deriving an expression for the MDE of an experiment as a function of the residual-based parameters, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, rather than the true parameters, σ^2_{ω}, ψ^B, ψ^A, and ψ^X. Recall that:

$$MDE = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_{\omega} + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}$$

Having solved for $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ as linear functions of the true parameters σ^2_{ω}, ψ^B, ψ^A, and ψ^X, we can define k_σ, k_B, k_A, and k_X as coefficients on the residual-based parameters $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. These coefficients will allow us to use residual-based parameters in the SCR formula in place of the true parameters. In other words, k_σ, k_B, k_A, and k_X must satisfy the following equation:[39]

$$\begin{aligned} &\left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)k_\sigma\sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right)k_B\psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right)k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}}\right]} \\ &\quad = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_{\omega} + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]} \end{aligned}$$

or more simply:

$$\begin{aligned} &\left(\frac{m+r}{mr}\right)k_\sigma\sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right)k_B\psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right)k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}} \\ &\quad = \left(\frac{m+r}{mr}\right)\sigma^2_{\omega} + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X \end{aligned}$$

39. This assumes k_σ, k_B, k_A, and k_X are functions of m, r, and I only, and do not themselves depend on σ^2_{ω}, ψ^B, ψ^A, or ψ^X. We show this to be true below.

We can express this equality in matrix form as:[40]

$$\begin{bmatrix} \frac{m+r}{mr} & \frac{m-1}{m} & \frac{r-1}{r} & -2 \end{bmatrix} \begin{bmatrix} \sigma^2_\omega \\ \psi^B \\ \psi^A \\ \psi^X \end{bmatrix} = \begin{bmatrix} \left(\frac{m+r}{mr}\right) k_\sigma & \left(\frac{m-1}{m}\right) k_B & \left(\frac{r-1}{r}\right) k_A & -2k_X \end{bmatrix} \begin{bmatrix} \sigma^2_{\hat{\omega}} \\ \psi^B_{\hat{\omega}} \\ \psi^A_{\hat{\omega}} \\ \psi^X_{\hat{\omega}} \end{bmatrix}$$

Substituting Equation (E5):

$$\Rightarrow \quad \begin{bmatrix} \frac{m+r}{mr} & \frac{m-1}{m} & \frac{r-1}{r} & -2 \end{bmatrix} \begin{bmatrix} \sigma^2_\omega \\ \psi^B \\ \psi^A \\ \psi^X \end{bmatrix} = \begin{bmatrix} \left(\frac{m+r}{mr}\right) k_\sigma & \left(\frac{m-1}{m}\right) k_B & \left(\frac{r-1}{r}\right) k_A & -2k_X \end{bmatrix} \mathbf{\Gamma} \begin{bmatrix} \sigma^2_\omega \\ \psi^B \\ \psi^A \\ \psi^X \end{bmatrix} \tag{E6}$$

where $\mathbf{\Gamma}$ is defined as above:

$$\mathbf{\Gamma} = \frac{I-1}{IT^2} \begin{bmatrix} T(T-1) & -m(m-1) & -r(r-1) & -2mr \\ -T & r^2+2r+m & r(r-1) & -2r^2 \\ -T & m(m-1) & m^2+2m+r & -2m^2 \\ -T & -r(m-1) & -m(r-1) & 2mr \end{bmatrix}$$

$\mathbf{\Gamma}$ is a singular matrix and cannot be inverted. However, we can show that, rather than having no solutions, Equation (E6) is instead overdetermined and there are infinite solutions. We are simply interested in one such solution. To find one set of k_σ, k_B, k_A, and k_X coefficients that solve this system, we iteratively solve for each k coefficient as a function of the remaining k coefficients and then substitute into the subsequent equations. That is, we use the first equation of this system to solve for k_σ as a function of k_B, k_A, and k_X and substitute this expression in place of k_σ in the subsequent equations in the system, and we repeat for the remaining coefficients and equations. This iterative procedure yields:

$$k_\sigma = \frac{I(m+r)^2 + (I-1)[r(m-1)k_B + m(r-1)k_A - 2mrk_X]}{(I-1)(m+r)(m+r-1)}$$

$$k_B = \frac{I(m+r)^2(m-r+mr+r^2) + (I-1)[m^2(r-1)(2-m-r)k_A + 2mr(r-m-mr-r^2)k_X]}{(I-1)r^2(3m+r+mr+r^2-2)}$$

$$k_A = \frac{I(m+r)^2 - 2(I-1)mrk_X}{2(I-1)m^2}$$

$$k_X = 0$$

We iteratively substitute each coefficient into the expressions for the remaining coefficients. That is, we first substitute this value of k_X into the expressions for the other three coefficients, then substitute k_A, and so on. This yields expressions for k_σ, k_B, k_A, and k_X in terms of m, r, and I.

$$k_\sigma = \frac{I(m+r)^2}{2(I-1)mr}$$

40. Note that if $m = 1$ (or $r = 1$), the corresponding ψ^B and $\psi^B_{\hat{\omega}}$ (or ψ^A and $\psi^A_{\hat{\omega}}$) parameters are undefined and no longer enter the system. Similarly, the corresponding row(s) and column(s) are removed from $\mathbf{\Gamma}$.

$$k_B = \frac{I(m+r)^2}{2(I-1)r^2}$$
$$k_A = \frac{I(m+r)^2}{2(I-1)m^2}$$
$$k_X = 0$$

We can now express the MDE of an experiment as a function of the residual-based parameters:

$$MDE^{est} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \left\{ \left(\frac{1}{P(1-P)J}\right) \left[\left(\frac{m+r}{mr}\right) \left(\frac{I(m+r)^2}{2(I-1)mr}\right) \sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right) \left(\frac{I(m+r)^2}{2(I-1)r^2}\right) \psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right) \left(\frac{I(m+r)^2}{2(I-1)m^2}\right) \psi^A_{\hat{\omega}} \right] \right\}^{1/2}$$

E.3 Estimation with ANCOVA

Here, we extend Appendices E.1 and E.2, which discuss estimating the *MDE* from residual-based parameters, for the ANCOVA model. As above, we do not observe the true parameters characterizing the error structure (σ^2_υ, σ^2_ω, ψ^B, ψ^A, and ψ^X). We cannot calculate these from a dataset, so we turn to residual-based versions of these parameters ($\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$), which we can use to define $MDE^{est}_{ANCOVA}(\sigma^2_{\hat{\upsilon}}, \sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}) = MDE_{ANCOVA}(\sigma^2_\upsilon, \sigma^2_\omega, \psi^B, \psi^A, \psi^X)$ such that the MDE from estimated parameters is equivalent to the SCR formula's MDE derived from true parameters.

Model As in Appendix E.2, we assume this model is identical to the model that generates the relevant power calculation formula; in this case, SCR ANCOVA given in Equation (11).

That is, there are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a time-invariant unit effect distributed i.i.d. $\mathcal{N}(0, \sigma^2_\upsilon)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma^2_\omega)$.

Assumption (Strict exogeneity)**.** $\mathrm{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*, and* $J-1$ *unit dummies. This again follows from random assignment of* D_{it}*.*

Assumption (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption (Independence across units)**.** $\mathrm{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0, \ \forall\ i \neq j, \ \forall\ t, s.$

Assumption (Uniform covariance structures)**.** *Define:*

$$\psi_i^B \equiv \frac{2}{m(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$
$$\psi_i^A \equiv \frac{2}{r(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$
$$\psi_i^X \equiv \frac{1}{mr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of unit i, respectively. Using these definitions, assume that $\psi^B = \psi_i^B$, $\psi^A = \psi_i^A$, and $\psi^X = \psi_i^X$ $\forall i$.

Parameters We begin by estimating the fixed effects and residuals. Note that even though we are ultimately interested in performing a power calculation for the ANCOVA estimator, we generate residuals for this data generating process by regressing Y_{it} on a constant and unit fixed effects:

$$Y_{it} = \beta + \upsilon_i + \omega_{it}$$

For a balanced panel, the estimated coefficients are:

$$\hat{\beta} = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it}$$
$$\hat{\upsilon}_i = \frac{1}{T} \sum_{t=-m+1}^{r} Y_{it} - \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it}$$

Then the estimated fixed effects and the residuals are

$$\hat{\upsilon}_i = (\upsilon_i - \overline{\upsilon}) + (\overline{\omega}_i - \overline{\overline{\omega}})$$
$$\hat{\omega}_{it} = \omega_{it} - \overline{\omega}_i$$

where

$$\overline{\upsilon} = \frac{1}{I} \sum_{i=1}^{I} \upsilon_i$$
$$\overline{\omega}_i = \frac{1}{T} \sum_{t=-m+1}^{r} \omega_{it}$$
$$\overline{\overline{\omega}} = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \omega_{it}$$

We first use the estimated fixed effects to derive an expression for $\sigma^2_{\hat{\upsilon}}$:[41]

$$\begin{aligned}
\sigma^2_{\hat{\upsilon}} &= \frac{1}{I}\sum_{i=1}^{I} \text{Var}(\hat{\upsilon}_i \mid \mathbf{X}) \\
&= \frac{1}{I}\sum_{i=1}^{I} \left[\text{Var}(\upsilon_i \mid \mathbf{X}) + \text{Var}(\overline{\upsilon} \mid \mathbf{X}) - 2\,\text{Cov}(\upsilon_i, \overline{\upsilon} \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X}) + \text{Var}(\overline{\overline{\omega}} \mid \mathbf{X}) - 2\,\text{Cov}(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X})\right] \\
&= \left(\frac{I-1}{IT^2}\right)\left[T^2\sigma^2_{\upsilon} + T\sigma^2_{\omega} + (m(m-1))\psi^B + (r(r-1))\psi^A + 2mr\psi^X\right]
\end{aligned}$$

We also use the definition of residuals to derive expressions for $\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$:

$$\begin{aligned}
\sigma^2_{\hat{\omega}} &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \text{Var}(\hat{\omega}_{it} \mid \mathbf{X}) \\
&= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \left[\text{Var}(\omega_{it} \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X}) - 2\,\text{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X})\right] \\
&= \left(\frac{T-1}{T}\right)\sigma^2_{\omega} - \left(\frac{m(m-1)}{T^2}\right)\psi^B - \left(\frac{r(r-1)}{T^2}\right)\psi^A - \left(\frac{2mr}{T^2}\right)\psi^X \\
\psi^A_{\hat{\omega}} &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r} \text{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r} \left[\text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \text{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) - \text{Cov}(\omega_{is}, \overline{\omega}_i \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X})\right] \\
&= \frac{-1}{T}\sigma^2_{\omega} + \left(\frac{m(m-1)}{T^2}\right)\psi^B + \left(\frac{m^2+2m+r}{T^2}\right)\psi^A - \left(\frac{2m^2}{T^2}\right)\psi^X \\
\psi^B_{\hat{\omega}} &= \frac{-1}{T}\sigma^2_{\omega} + \left(\frac{r^2+2r+m}{T^2}\right)\psi^B + \left(\frac{r(r-1)}{T^2}\right)\psi^A - \left(\frac{2r^2}{T^2}\right)\psi^X \\
\psi^X_{\hat{\omega}} &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r} \text{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r} \left[\text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \text{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) - \text{Cov}(\omega_{is}, \overline{\omega}_i \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X})\right] \\
&= \frac{-1}{T}\sigma^2_{\omega} - \left(\frac{r(m-1)}{T^2}\right)\psi^B - \left(\frac{m(r-1)}{T^2}\right)\psi^A + \left(\frac{2mr}{T^2}\right)\psi^X
\end{aligned}$$

Minimum detectable effect Having solved for $\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ as linear functions of the true parameters σ^2_{υ}, σ^2_{ω}, ψ^B, ψ^A, and ψ^X, we can define k_{υ}, k_{ω}, k_B, k_A, and k_X as coefficients on the residual-based parameters $\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. These coefficients will allow us to use residual-based parameters in the SCR formula in place of the true parameters. In other words, k_{υ},

41. Note that, as stated in the strict exogeneity assumption, $\mathbf{X}$ is a full rank matrix of regressors, including a constant, the treatment indicator $\mathbf{D}$, and $J-1$ unit dummies. That is, $\mathbf{X}$ includes the covariates of the data generating process and residualizing regression, not of the ANCOVA model.

k_ω, k_B, k_A, and k_X must satisfy the following equation:

$$\begin{aligned}(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X \\ = k_\upsilon\sigma_{\hat{\upsilon}}^2 + k_\omega\sigma_{\hat{\omega}}^2 + k_B\psi_{\hat{\omega}}^B + k_A\psi_{\hat{\omega}}^A + k_X\psi_{\hat{\omega}}^X\end{aligned}$$

Following the solution procedure described in Appendix E.2, we solve for a set of k coefficients that satisfy this equality:

$$\begin{aligned}k_\upsilon &= \left(\frac{I}{I-1}\right)(1-\theta)^2 \\ k_\omega &= \left(\frac{m+\theta r}{2m^2r^2}\right)((m+r)(m+\theta r) + (1-\theta)(mr^2 - m^2r)) \\ k_B &= \left(\frac{m+\theta r}{2mr^2}\right)(m-1)(m+\theta r - (1-\theta)mr) \\ k_A &= \left(\frac{m+\theta r}{2m^2r}\right)(r-1)(m+\theta r + (1-\theta)mr) \\ k_X &= 0\end{aligned}$$

Note, however, that these coefficients are functions of not only m, r, and I, as in Appendix E.2. These coefficients are also a function of θ, which is itself a function of true parameters:

$$\theta = \frac{m\sigma_\upsilon^2 + m\psi^X}{m\sigma_\upsilon^2 + \sigma_\omega^2 + (m-1)\psi^B}$$

As a result, in order to use these coefficients in a residual-based SCR ANCOVA formula, we must also derive a residual-based expression for θ. We define a set of coefficients to correspond to the numerator of θ: k_υ^N, k_ω^N, k_B^N, k_A^N, and k_X^N; we also define a set of coefficients to correspond to the denominator of θ: k_υ^D, k_ω^D, k_B^D, k_A^D, and k_X^D. We seek to find the sets of k^N and k^D coefficients that satisfy the following equality:

$$\frac{m\sigma_\upsilon^2 + m\psi^X}{m\sigma_\upsilon^2 + \sigma_\omega^2 + (m-1)\psi^B} = \frac{k_\upsilon^N\sigma_{\hat{\upsilon}}^2 + k_\omega^N\sigma_{\hat{\omega}}^2 + k_B^N\psi_{\hat{\omega}}^B + k_A^N\psi_{\hat{\omega}}^A + k_X^N\psi_{\hat{\omega}}^X}{k_\upsilon^D\sigma_{\hat{\upsilon}}^2 + k_\omega^D\sigma_{\hat{\omega}}^2 + k_B^D\psi_{\hat{\omega}}^B + k_A^D\psi_{\hat{\omega}}^A + k_X^D\psi_{\hat{\omega}}^X}$$

We again follow the solution procedure described in Appendix E.2 as we separately solve for a set of k^N coefficients and k^D coefficients. One such solution for the numerator coefficients is:

$$\begin{aligned}k_\upsilon^N &= \left(\frac{I}{I-1}\right)m \\ k_\omega^N &= \left(\frac{-1}{4r}\right)(m(m-r+2) + r(r-m+2)) \\ k_B^N &= \left(\frac{-m}{4r}\right)(m-1)(m-r+2) \\ k_A^N &= \left(\frac{-1}{4}\right)(r-1)(r-m+2) \\ k_X^N &= 0\end{aligned}$$

and one such solution for the denominator coefficients is:

$$
\begin{aligned}
k_{\upsilon}^{D} &= \left(\frac{I}{I-1}\right) m \\
k_{\omega}^{D} &= \left(\frac{1}{2m}\right) (m(m+1) - r(m-1)) \\
k_{B}^{D} &= \left(\frac{1}{2}\right) (m+1)(m-1) \\
k_{A}^{D} &= \left(\frac{-r}{2m}\right) (m-1)(r-1) \\
k_{X}^{D} &= 0
\end{aligned}
$$

Combining these two sets of coefficients, we express θ as a function of residual-based parameters:

$$
\begin{aligned}
\theta = & \frac{m\left[\left(\frac{I}{I-1}\right) 4mr\sigma_{\hat{\upsilon}}^{2} - (m(m-r+2) + r(r-m+2))\sigma_{\hat{\omega}}^{2}\right]}{2r\left[\left(\frac{I}{I-1}\right) 2m^{2}\sigma_{\upsilon}^{2} + (m(m+1) - r(m-1))\sigma_{\hat{\omega}}^{2} + m(m-1)(m+1)\psi_{\hat{\omega}}^{B} - r(m-1)(r-1)\psi_{\hat{\omega}}^{A}\right]} \\
& + \frac{m\left[-m(m-1)(m-r+2)\psi_{\hat{\omega}}^{B} - r(r-1)(r-m+2)\psi_{\hat{\omega}}^{A}\right]}{2r\left[\left(\frac{I}{I-1}\right) 2m^{2}\sigma_{\upsilon}^{2} + (m(m+1) - r(m-1))\sigma_{\hat{\omega}}^{2} + m(m-1)(m+1)\psi_{\hat{\omega}}^{B} - r(m-1)(r-1)\psi_{\hat{\omega}}^{A}\right]}
\end{aligned}
$$

We can now express the SCR ANCOVA formula using only residual-based parameters:

$$
\begin{aligned}
MDE_{ANCOVA}^{est} \approx (t_{1-\kappa}^{J} + t_{\alpha/2}^{J}) \times \Bigg\{ & \left(\frac{1}{P(1-P)J}\right) \\
& \times \Bigg[\left(\frac{I}{I-1}\right) (1-\theta)^{2}\sigma_{\hat{\upsilon}}^{2} \\
& + \left(\frac{m+\theta r}{2m^{2}r^{2}}\right) ((m+r)(m+\theta r) + (1-\theta)(mr^{2} - m^{2}r))\sigma_{\hat{\omega}}^{2} \\
& + \left(\frac{m+\theta r}{2mr^{2}}\right) (m-1)(m+\theta r - (1-\theta)mr)\psi_{\hat{\omega}}^{B} \\
& + \left(\frac{m+\theta r}{2m^{2}r}\right) (r-1)(m+\theta r + (1-\theta)mr)\psi_{\hat{\omega}}^{A} \Bigg] \Bigg\}^{1/2}
\end{aligned}
$$

where θ is as given above.

Estimated parameters We finally show that we can recover unbiased estimates of the residual-based parameters, as in Appendix E.1. We begin with the variance of the fixed effects, $\sigma_{\hat{\upsilon}}^{2}$, and define:

$$
\tilde{\sigma}_{\hat{\upsilon}}^{2} \equiv \frac{1}{I}\sum_{i=1}^{I}\left(\hat{\upsilon}_{i} - \overline{\hat{\upsilon}}\right)^{2}
$$

where $\overline{\hat{v}} = \frac{1}{I}\sum_{i=1}^{I} \hat{v} = 0$. Taking expectations of both sides:

$$\begin{aligned}
\mathrm{E}\left[\tilde{\sigma}^2_{\hat{v}} \mid \mathbf{X}\right] &= \frac{1}{I}\sum_{i=1}^{I} \mathrm{E}\left[\hat{v}^2 \mid \mathbf{X}\right] \\
&= \sigma^2_{\hat{v}}
\end{aligned}$$

Next, to estimate the variance of the residuals, $\sigma^2_{\hat{\omega}}$, we define:

$$\tilde{\sigma}^2_{\hat{\omega}} \equiv \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\left(\hat{\omega}_{it} - \overline{\overline{\hat{\omega}}}\right)^2$$

where $\overline{\overline{\hat{\omega}}} = \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\hat{\omega}_{it} = 0$. Taking expectations of both sides:

$$\begin{aligned}
\mathrm{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right] &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\mathrm{E}\left[\hat{\omega}^2_{it} \mid \mathbf{X}\right] \\
&= \sigma^2_{\hat{\omega}}
\end{aligned}$$

To estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, we define the $[I \times 1]$ vector of residuals for period t as $\hat{\boldsymbol{\omega}}_t$ and calculate the average covariance between any two vectors in the relevant range of time periods. For the pre-period, we define:

$$\begin{aligned}
\tilde{\psi}^B_{\hat{\omega}} &\equiv \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\left(\hat{\omega}_{it} - \overline{\hat{\omega}}_t\right)\left(\hat{\omega}_{is} - \overline{\hat{\omega}}_s\right) \\
&= \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\left(\hat{\omega}_{it}\hat{\omega}_{is} - \hat{\omega}_{it}\overline{\hat{\omega}}_s - \hat{\omega}_{is}\overline{\hat{\omega}}_t + \overline{\hat{\omega}}_t\overline{\hat{\omega}}_s\right) \\
&= \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\left(\frac{I-1}{I}\right)\hat{\omega}_{it}\hat{\omega}_{is}
\end{aligned}$$

Taking expectations yields:

$$\begin{aligned}
\mathrm{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right] &= \left(\frac{I-1}{I}\right)\frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\mathrm{E}\left[\hat{\omega}_{it}\hat{\omega}_{is} \mid \mathbf{X}\right] \\
&= \left(\frac{I-1}{I}\right)\frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\mathrm{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \left(\frac{I-1}{I}\right)\psi^B_{\hat{\omega}}
\end{aligned}$$

Similarly:

$$\mathrm{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right] = \left(\frac{I-1}{I}\right)\psi^A_{\hat{\omega}}$$
$$\mathrm{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right] = \left(\frac{I-1}{I}\right)\psi^X_{\hat{\omega}}$$

Hence, we can recover unbiased estimates of the parameters $\sigma^2_{\hat{v}}$ and $\sigma^2_{\hat{\omega}}$ by calculating the averages of the estimated $\tilde{\sigma}^2_{\hat{v}}$ and $\tilde{\sigma}^2_{\hat{\omega}}$, respectively. However, to recover unbiased estimates of the parameters $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, we must calculate $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$, respectively, and then inflate these averages by $\frac{I}{I-1}$.

Appendix References

Athey, Susan, and Guido Imbens. 2017a. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31 (2): 3–32.

Athey, Susan, and Guido W. Imbens. 2017b. "The Econometrics of Randomized Experiments." *Handbook of Economic Field Experiments* 1:73–140.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–275.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *The Quarterly Journal of Economics* 130 (1): 165–218.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–372.

Campbell, Cathy. 1977. "Properties of Ordinary and Weighted Least Square Estimators of Regression Coefficients for Two-Stage Samples." *Proceedings of the Social Statistics Section, American Statistical Association:* 800–805.

Frison, L., and S. J. Pocock. 1992. "Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and its Implications for Design." *Statistics in Medicine* 11 (13): 1685–1704.

McKenzie, David. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–221.

Moulton, Brent. 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32 (3): 385–397.

Pecan Street. 2016. "Dataport." `https://dataport.pecanstreet.org/`.