Ch. 9 and Ch. 10 Responses

Dylan Baker

February 12, 2025

Contents

1 Ch. 9 Question Responses

1.1 Question 1

Suppose you run an experiment to look at long-run outcomes of an intervention. You suspect that the treatment will make subjects less sensitive to shocks. How may this affect your selection of sample size allocated to treatment and control?

The diminished sensitivity will increase autocorrelation, implying relatively less information gained for additional time series observations. We should allocate more observations to the condition where autocorrelation is lower, i.e., the control.

We should set sample sizes according to the formula:

$$\frac{n_1}{n_0} = \sqrt{\frac{1 + (N_{\text{Post}} - 1)\,\rho_1}{1 + (N_{\text{Post}} - 1)\,\rho_0}}$$

1.2 Question 2

A researcher finds in a longitudinal experiment that treatment increases consumption measured at 1 year post-treatment by a precise 30%, and the treatment effect after two years is only 3%. Can we conclude that consumption after one year is not a valid surrogate (does not meet the surrogacy condition) for consumption after two years in this experiment?

Without additional information, this fact alone does not imply that consumption after one year is not a valid surrogate. It could, in principle, be the case that this is the relationship between consumption from year to year in observational samples and that the full causal relationship between the treatment and consumption after two years is captured by consumption after one year – although it may be kind of strange if these conditions are satisfied in this case.

2 Main Takeaways

2.1 Ch. 9 Takeaways

1. Consider the two parameters:

$$\begin{aligned}\tau^{DID} = \tau^{\text{post}} &- \tau^{\text{pre}} \\ &= \sum_{t>0} \left(\mathbb{E}\left[Y_{it} \mid D_{it}=1\right] - \mathbb{E}\left[Y_{it} \mid D_{it}=0\right]\right) \\ &\quad - \sum_{t\leq 0} \left(\mathbb{E}\left[Y_{it} \mid D_{it}=1\right] - \mathbb{E}\left[Y_{it} \mid D_{it}=0\right]\right)\end{aligned}$$

and

$$\tau^{\text{Post}} = \sum_{t>0} \left(\mathbb{E}\left[Y_{it} \mid D_{it}=1\right] - \mathbb{E}\left[Y_{it} \mid D_{it}=0\right]\right)$$

When you have pre-treatment observations, there are instances where it may be beneficial to include them in your specification and instances where it may not be. This depends on how many pre-treatment observations you have and the strength of the autocorrelation parameter.

Under the assumptions of

(a) constant variance across time periods,

(b) equal variances between all pairs of time points

(c) autocorrelation in outcomes does not depend on treatment effects

(d) treatment and control groups are identical in cross-sectional size

we have

$$\text{Var}\left[\tau^{\text{Post}}\right] = \frac{2\sigma^2}{n_d}\left(\frac{1+\left(N_{\text{Post}}-1\right)\rho}{N_{\text{Post}}}\right)$$

and

$$\text{Var}\left[\tau^{DID}\right] = \frac{2\sigma^2}{n_d}\left(\frac{1+\left(N_{\text{Post}}\ -1\right)\rho}{N_{\text{Post}}} - \frac{\left(N_{\text{Pre}}\ +1\right)\rho - 1}{N_{\text{Pre}}}\right)$$

2. The smallest sample size correspond to a given MDE is given by:

$$n_0^* = n_1^* = n^* = \frac{2\left(t_{\alpha/2} + t_\beta\right)^2\sigma^2}{(MDE)^2}\left(\frac{N_{\text{Pre}}\ + N_{\text{Post}}}{N_{\text{Pre}}\ * N_{\text{Post}}}\right)$$

   As the number of periods increases from 4 to 16, the minimum sample size required to detect a given MDE decreases four-fold.

3. When we are interested in the long-run effect of something but may be interested in expanding the policy before we can observe the long-run effect, we may use a surrogate to estimate the long-run effect. In this case, under strong assumptions, we consider the effect of the treatment on the surrogate and the effect of the surrogate on the long-run outcome.

   The key assumptions are:

Assumptions

(a) (A9.3, Comparability): $P_i \perp Y_i \mid Y_i^S$

(b) A linear relationship between the outcome and the surrogate

(c) $\text{Var}\left[Y_i^S \mid D_i\right] = \text{Var}\left[Y_i \mid D_i\right]$

(d) (A9.4, Surrogacy Condition): $D_i \perp Y_i \mid Y_i^S$

   Under these conditions, we get:

$$\tau = \rho_{Y_i, Y_i^S} \cdot \tau^s$$

   If the surrogacy condition doesn't hold, we instead get:

$$\tau = \rho_{Y_i, Y_i^S} \cdot \tau^S + \left(\mathbb{E}\left[Y_i \mid Y_i^S, D_i = 1\right] - \mathbb{E}\left[Y_i \mid Y_i^S, D_i = 0\right]\right)$$

2.2 Ch. 10 Takeaways

1. Additional assumptions required for within-subjects are:

   (a) (A10.1, Balanced Panel): For all $(i,t) \in \mathcal{J} \times \mathcal{T}$, we have $\mathbb{P}\left[R_{it} = 1\right] = 1$, and for all $(i,t) \in \mathcal{J} \times \mathcal{T}$ with $\mathbb{P}\left[R_{it} = 1\right] = 1$ the researcher observes $(Y_{it}, D_{it}, Z_{it}, \boldsymbol{X}_{it})$.

   (b) (A10.2, Temporal Stability): For all $t \in \mathcal{T}, Y_{it}\left(D_{it}, \boldsymbol{D}_{i,-t}, t\right) = Y_{it}\left(D_{it}, \boldsymbol{D}_{i,-\boldsymbol{t}}\right)$.

   - This rules out time-varying effects in the potential outcomes, such as respondent fatigue.

(c) (A10.3, Causal Transience): For all $\boldsymbol{D_i}, Y_{it}\left(D_{it}, \boldsymbol{D_i}, T\right) = Y_{it}\left(D_{it}, T\right)$.

- This implies that the treatment effect does not persist over time when we change the treatment status; the treatment effects do not depend on the order in which they are implemented.

2. Three major threats to temporal stability are:

   - time-specific shocks
   - time trends
   - changing measurement outcome

3. If the assumptions are tenable, within-subjects designs can reach high power levels much faster than between-subjects designs.