Optimal Design of Experiments in Panel Data Settings

Nick Doudchenko, Khashayar Khosravi, Jean Pouget-Abadie, Sebastien Lahaie, Miles Lubin, Vahab Mirrokni (Google), Jann Spiess, Guido Imbens (Stanford)

Abstract

We investigate the optimal design of experimental studies that have pre-treatment outcome data available. The average treatment effect is estimated as the difference between the weighted average outcomes of the treated and control units. A number of commonly used approaches fit this formulation, including the difference-in-means estimator and a variety of synthetic-control techniques. We propose several different novel estimators and motivate the choice between them depending on the underlying assumptions the researcher is willing to make. Observing the NP-hardness of the problem, we introduce a mixed-integer programming formulation which selects both the treatment and control sets and unit weightings. We prove that these proposed estimators lead to qualitatively different experimental units being selected for treatment. We use simulations based on publicly available data from the US Bureau of Labor Statistics that show improvements in terms of the mean squared error of the estimates when compared to simple and commonly used alternatives such as randomized trials.

Motivation

Randomized Controlled Trials (RCT's) or A/B tests—as they are commonly called in applied commercial settings—while being the gold standard of *causal inference*, are not always feasible or desirable for the reasons such as:

- **Interference:** Treatment status of one unit may affect the outcome of another and the units need to be combined into larger clusters.
- **Institutional constraints:** Treatment can only be assigned to large geographic areas.
- **Costs:** Only one or just a few units may be treated.
- **Privacy and/or fairness:** Treatment assignment at the individual level may be problematic.

Framework

- For N units in T time periods there are two *potential outcomes*, $(Y_{it}(0), Y_{it}(1))$, associated with an outcome metric of interest, Y:

$$Y_{it}(0) = \mu_{it} + \varepsilon_{it}$$
$$Y_{it}(1) = Y_{it}(0) + \tau_i,$$

where the error terms, ε_{it}, are homoskedastic with mean zero and variance σ^2.

- The observed outcome that depends on the treatment status, $D_i \in \{0,1\}$, is:

$$Y_{it} = Y_{it}(1)D_i + Y_{it}(0)(1 - D_i).$$

- At time T the researcher decides which units should be treated ($D_i = 1$) in period $T+1$.
- Our goal is to estimate the treatment effects (either individually or on average):

$$\tau_i = Y_{i,T+1}(1) - Y_{i,T+1}(0).$$

- Family of estimators of the *(weighted) Average Treatment Effect on the Treated* (ATET):

$$\hat{\tau} = \sum_{i=1}^{N} w_i Y_{i,T+1} D_i - \sum_{i=1}^{N} w_i Y_{i,T+1}(1 - D_i).$$

Results

- Unemployment rate data from the Bureau of Labor Statistics: 50 states in 40 months.
- Select 10-by-10 blocks and apply treatment in the last 3 periods to the units selected either:
  - Optimally, using one of our problems, or
  - Randomly.
- Heterogeneous treatment effects, τ_i, with values from 0 to 0.1 spread linearly across treated units.

Table: Root-mean-square errors of the average and unit-level effect estimates (lowest values in **bold**)

Heterogeneous effects	$K=3$		$K=7$	
	Average	Indiv.	Average	Indiv.
(i) Two-way	7.1	41.6	7.7	34.4
(ii) One-way	**6.6**	41.5	7.3	34.3
(iii) Per-unit	8.2	**13.1**	**7.1**	**14.1**
(iv) Synth. control (random D_i's)	8.1	13.2	9.8	16.3
(v) Diff-in-means (random D_i's)	9.3	42.2	10.1	35.3

Design & Estimation Approach

Simultaneously choose the units intended for treatment (D) and the weights used for estimation (w).

Optimization Problems

❶ **Per-unit problem:**

$$\min_{\{D_i, \{w_j^i\}_{j=1}^N\}_{i=1}^N} \frac{1}{K}\sum_{i=1}^{N} D_i \left[\frac{1}{T}\sum_{t=1}^{T}\left(Y_{it} - \sum_{j=1}^{N} w_j^i (1 - D_j) Y_{jt}\right)^2 + \sigma^2 \sum_{j=1}^{N}\left(w_j^i\right)^2\right]$$
$$\text{s.t.} \quad \sum_{i=1}^{N} D_i = K, \quad w_j^i \geq 0, \quad \sum_{j=1}^{N} w_j^i = 1$$

❷ **Two-way global problem:**

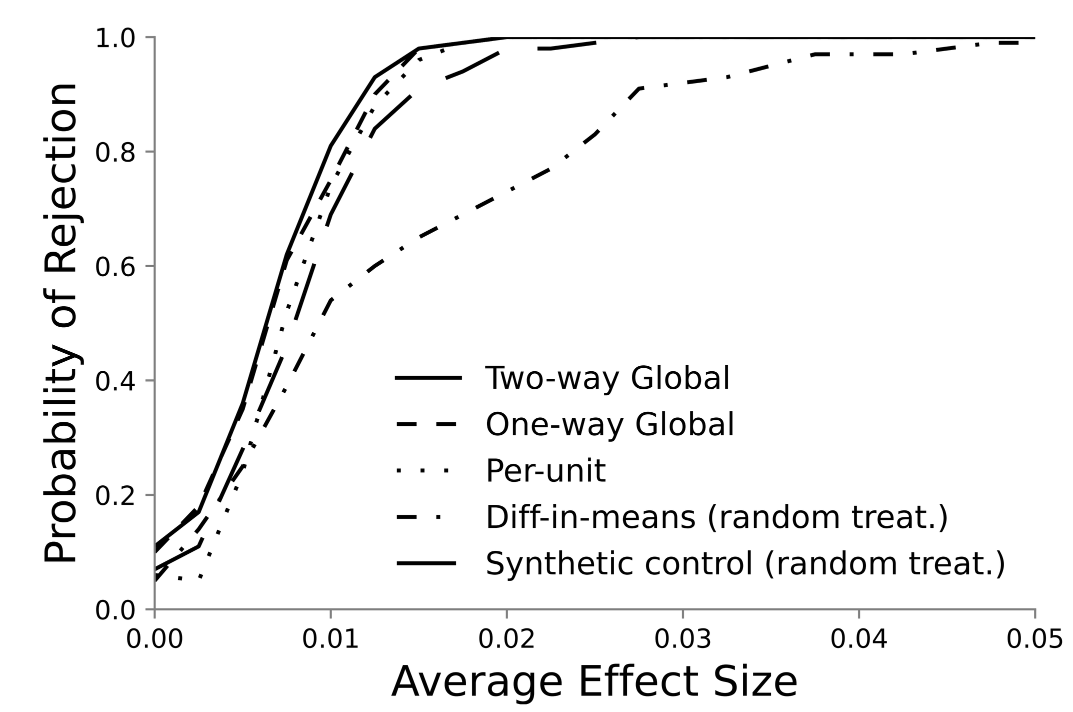
$$\min_{\{D_i, w_i\}_{i=1}^N} \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} w_i D_i Y_{it} - \sum_{i=1}^{N} w_i (1 - D_j) Y_{it}\right)^2 + \sigma^2 \sum_{i=1}^{N} w_i^2$$
$$\text{s.t.} \quad \sum_{i=1}^{N} D_i = K, \quad w_i \geq 0, \quad \sum_{i=1}^{N} D_i w_i = 1, \quad \sum_{i=1}^{N} (1 - D_i) w_i = 1$$

❸ **One-way global problem:** Same as above, but set the weights of the treated units to $1/K$.

Inference

Inference is hard to do in synthetic-control-type settings. We use the permutation-based approach from Chernozhukov, Wüthrich and Zhu (2021) for testing a sharp H_0: $\forall i$: $\tau_i = 0$:

1. Permute the time periods.
2. Re-estimate the treatment effects.
3. Obtain the distribution of the treatment effects under the null.
4. Compare the original estimate to the quantiles of the constructed distribution.



Limitations & Future Work

- Generalization:
  - We focus on the ATET and do not say anything about the ATE on the control units, or the population in general.
- Theory is lacking:
  - No theoretical guarantees for the estimators.
  - The inference results are only proven to be valid under fairly trivial conditions (data *i.i.d.* across time periods).
- Scalability:
  - Currently can handle up to about 100 units.
  - Approximate algorithms are likely required.
- Applications:
  - Please, reach out to us if you think your experiment could benefit from similar ideas and you would like to collaborate!

References

Abadie and Zhao (2021). Synthetic Controls for Experimental Design. https://arxiv.org/abs/2108.02196