

Supplementary Material

Estimating Treatment Effects on Proportions with Synthetic Controls

Konstantin Bogatyrev, LSE, United Kingdom (✉ k.bogatyrev@lse.ac.uk)

Lukas F. Stoetzer, University of Witten/Herdecke, Germany.

Table of Contents

1	Formalization of Synthetic Control Methods	2
2	Proofs	3
3	Alternative Formalization of the Constant Control Comparison Under the Demeaned Weighting Estimator	5
3.1	Estimator	5
3.2	Constant Control Comparison (CCC)	5
3.3	Proof: Sufficiency of the CCC	6
4	Algorithm: Synthetic Control Method with Common Weights	7
5	Simulation: Example of the data generating process	7
	References	8

1 Formalization of Synthetic Control Methods

We consider a balanced panel with N units and T time periods. Where we denote units with $i \in \{1, \dots, N\}$ and periods with $t \in \{1, \dots, T\}$. The outcome of interest are K proportions, where each proportion value is denoted with $k \in \{1, \dots, K\}$. The outcome for unit i in period t is a multivariate vector of $Y_{it} = [y_{it1}, \dots, y_{itK}]$. With a proportion, it holds that $y_{itk} \in [0, 1]$ for all k and $\sum_{k=1}^K y_{itk} = 1$, forming a K -dimensional simplex.

We further have a binary treatment $D_{it} \in \{0, 1\}$. For now we assume that there is a share of control units that are never exposed to the treatment N_{co} , while there are units N_{tr} that are treated after time T_0 . For the analysis, we arrange the i units such that first all control units appear and then all treatment units $i \in \{1, \dots, N_{co}, N_{co} + 1, \dots, N\}$.

A DID method would estimate separate average treatment effects of exposure τ_k^{did} for each proportion, essentially solving a two-way fixed effects regression problem:

$$(\hat{\tau}_k^{did}, \hat{\mu}_k, \hat{\alpha}_{ik}, \hat{\beta}_{tk}) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (y_{itk} - \mu_k - \alpha_{ik} - \beta_{tk} - D_{it}\tau)^2 \right\} \quad (1)$$

for this specification, the (unweighted) average of the control units essentially serves as the counterfactual outcome for the treatment units after treatment exposure. The key identification parallel trends assumption makes this (unweighted) average a valid counterfactual outcome.

Synthetic methods instead calculate weights $\hat{\omega}_{ik}$ that align pre-exposure trends in the outcomes of the control with the treatment units, such that $\sum_{i=1}^{N_{co}} \hat{\omega}_{ik} y_{itk} \approx \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N y_{itk}$ for $t \in \{1, \dots, T_0\}$. The approach further specifies time-weights $\hat{\lambda}_{tk}$ that balance pre- and post-exposure periods for the different shares. The estimation then proceeds with finding a solution to the minimization problem.

$$(\hat{\tau}_k^{sdid}, \hat{\mu}_k, \hat{\alpha}_{ik}, \hat{\beta}_{tk}) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (y_{itk} - \mu_k - \alpha_{ik} - \beta_{tk} - D_{it}\tau)^2 \hat{\omega}_{ik} \hat{\lambda}_{tk} \right\} \quad (2)$$

The synthetic difference-in-differences estimator can also be applied to classic synthetic method approaches (Arkhangelsky et al. 2021). Here the minimization problem excludes the time-weights and unit fixed effects.

Of central importance for SCMs is the calculation of the weights. In the application to proportions, these would be calculated for each proportion separately. Arkhangelsky et al. (2021) approach for this is an optimization problem with regularization:

$$(\hat{\omega}_{0k}, \hat{\omega}_k) = \arg \min_{\omega_{0k} \in \mathbb{R}, \omega_k \in \Omega} \ell_{unit}(\omega_{0k}, \omega_k) \text{ where} \quad (3)$$

$$\ell_{unit}(\omega_{0k}, \omega_k) = \sum_{t=1}^{T_0} \left(\omega_{0k} + \sum_{i=1}^{N_{co}} \omega_{ik} y_{itk} - \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N y_{itk} \right)^2 + \xi_k^2 T_0 \|\omega_k\|_2^2,$$

$$\Omega = \left\{ \omega_k \in \mathbb{R}_+^N : \sum_{i=1}^{N_{co}} \omega_{ik} = 1, \omega_{ik} = \frac{1}{N_{tr}} \forall i \in \{N_{co} + 1, \dots, N\} \right\}$$

where the regularization parameter is pre-specified as follows:

$$\xi_k = (N_{tr}(T - T_0))^{1/4} \hat{\sigma} \text{ with } \hat{\sigma}^2 = \frac{1}{N_{co}(T_0 - 1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_0-1} (\Delta_{itk} - \bar{\Delta}_k)^2, \quad (4)$$

$$\text{where } \Delta_{itk} = y_{i(t+1)k} - y_{itk} \text{ and } \bar{\Delta}_k = \frac{1}{N_{co}(T_0 - 1)} \sum_{i=1}^{N_{co}} \sum_{t=1}^{T_0-1} \Delta_{itk} \quad (5)$$

The time weights are analogously implemented by solving a minimization problem. Following Arkhangelsky et al. (2021), it does not include a penalization parameter:

$$(\hat{\lambda}_{0k}, \hat{\lambda}_k) = \arg \min_{\lambda_{0k} \in \mathbb{R}, \lambda_k \in \Lambda} \ell_{time}(\lambda_{0k}, \lambda_k) \text{ where} \quad (6)$$

$$\ell_{time}(\lambda_{0k}, \lambda_k) = \sum_{i=1}^{N_{co}} \left(\lambda_{0k} + \sum_{t=1}^{T_0} \lambda_{tk} y_{itk} - \frac{1}{T - T_0} \sum_{t=T_0+1}^T y_{itk} \right)^2,$$

$$\Lambda = \left\{ \lambda_k \in \mathbb{R}_+^T : \sum_{t=1}^{T_0} \lambda_t = 1, \lambda_t = \frac{1}{T - T_0} \forall t \in \{T_0 + 1, \dots, T\} \right\}$$

The algorithm for synthetic Difference in Differences proceeds as follows:

- For each proportion $k \in \{1, \dots, K\}$:
 1. Compute regularization parameter ξ_k
 2. Compute unit weights $\hat{\omega}_k$
 3. Compute time weights $\hat{\lambda}_k$
 4. Compute the SDID estimator via the weighted DID regression

$$(\hat{\tau}_k^{sdid}, \hat{\mu}_k, \hat{\alpha}_{ik}, \hat{\beta}_{tk}) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (y_{itk} - \mu_k - \alpha_{ik} - \beta_{tk} - D_{it} \tau)^2 \hat{\omega}_{ik} \hat{\lambda}_{tk} \right\} \quad (7)$$

2 Proofs

Lemma 1: By linearity of expectation, the sum of the treatment effects across all categories is $\sum_{k=1}^K \tau_k = \mathbb{E}[\sum_{k=1}^K y_{ik}(1) - \sum_{k=1}^K y_{ik}(0) \mid D_i = 1]$. Using the compositional constraint, $\sum_{k=1}^K y_{ik}(1) = \sum_{k=1}^K y_{ik}(0) = 1$, this simplifies to $\sum_{k=1}^K \tau_k = \mathbb{E}[1 - 1 \mid D_i = 1] = \mathbb{E}[0 \mid D_i = 1] = 0$, so the sum of the treatment effects is zero.

Theorem 1: Let $\hat{\tau}_k^{sdid}$ be defined by the SDID double difference in Eq. 2 in the article. Under CCC (Definition 2 in the article), $\omega_{jk} \equiv \omega_j$ and $\lambda_{tk} \equiv \lambda_t$ for all k . Summing the SDID estimator over k gives

$$\sum_{k=1}^K \hat{\tau}_k^{sdid} = \underbrace{\left(\frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N \sum_{k=1}^K y_{iT k} \right)}_{\mathcal{Q}_1} - \underbrace{\left(\sum_{j=1}^{N_{co}} \omega_j \sum_{k=1}^K y_{jT k} \right)}_{\mathcal{Q}_2} - \underbrace{\left(\frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N \sum_{t=1}^{T_0} \lambda_t \sum_{k=1}^K y_{itk} \right)}_{\mathcal{Q}_3} + \underbrace{\left(\sum_{j=1}^{N_{co}} \omega_j \sum_{t=1}^{T_0} \lambda_t \sum_{k=1}^K y_{jtk} \right)}_{\mathcal{Q}_4}.$$

First, because outcomes are compositional, for every treated unit i at the period T , $\sum_{k=1}^K y_{iT_k} = 1$. Hence

$$\mathcal{Q}_1 = \frac{1}{N_{\text{tr}}} \sum_{i=N_{\text{co}}+1}^N 1 = 1.$$

Second, for every control unit j at the period T , $\sum_{k=1}^K y_{jT_k} = 1$. Using $\sum_j \omega_j = 1$ from the definition of the estimator,

$$\mathcal{Q}_2 = \sum_{j=1}^{N_{\text{co}}} \omega_j \cdot 1 = 1.$$

Third, compositional data implies $\sum_{k=1}^K y_{itk} = 1$ for each i, t . Using $\sum_t \lambda_t = 1$ from the definition of the estimator,

$$\mathcal{Q}_3 = \frac{1}{N_{\text{tr}}} \sum_{i=N_{\text{co}}+1}^N \sum_{t=1}^{T_0} \lambda_t \cdot 1 = \sum_{t=1}^{T_0} \lambda_t = 1,$$

Fourth, similarly,

$$\mathcal{Q}_4 = \sum_{j=1}^{N_{\text{co}}} \omega_j \sum_{t=1}^{T_0} \lambda_t \cdot 1 = 1.$$

Therefore,

$$\sum_{k=1}^K \hat{\tau}_k^{\text{sdid}} = (1 - 1) - (1 - 1) = 0,$$

which proves the claim.

Generalization to multiple post-treatment periods: If there are $T - T_0 \geq 1$ post periods and SDID uses uniform post weights $1/(T - T_0)$, replace y_{iT_k} by the post average $\frac{1}{T - T_0} \sum_{t=T_0+1}^T y_{itk}$; the same argument goes through since the post weights sum to one.

Example without and with constant control comparisons: Consider $K = 2$ with one treated unit and two control units. The treated unit has $y_{iT_1}(1) = 0.6$, $y_{iT_2}(1) = 0.4$. The control units have: $y_{1T_1}(0) = 0.7$, $y_{1T_2}(0) = 0.3$, $y_{2T_1}(0) = 0.5$, $y_{2T_2}(0) = 0.5$.

We first show the case with unequal weights. The weights are $\omega_{11} = 0.8$, $\omega_{21} = 0.2$ for $k = 1$, and $\omega_{12} = 0.2$, $\omega_{22} = 0.8$ for $k = 2$. The pre-treatment averages are $\bar{y}_{i.1} = 0.7$ and $\bar{y}_{i.2} = 0.3$. The demeaned outcomes for controls are: $\dot{y}_{11} = 0$, $\dot{y}_{21} = -0.2$, $\dot{y}_{12} = 0$, $\dot{y}_{22} = 0.2$. The ATT for $k = 1$: $\hat{\tau}_1 = 0.6 - (0.7 - 0.04) = 0.6 - 0.66 = -0.06$, and for $k = 2$: $\hat{\tau}_2 = 0.4 - (0.3 + 0.16) = 0.4 - 0.46 = -0.06$. The sum of the treatment effects is: $\sum_{k=1}^2 \hat{\tau}_k = -0.12$, which violates the constraint $\sum_{k=1}^K \hat{\tau}_k = 0$ due to unequal weights.

Now assume equal weights for the outcome categories but different weights across units: $\omega_{11} = \omega_{12} = 0.5$ for Control 1, $\omega_{21} = \omega_{22} = 0.5$ for Control 2. The pre-treatment averages are $\bar{y}_{i.1} = 0.7$ and $\bar{y}_{i.2} = 0.3$. The demeaned outcomes for controls are: $\dot{y}_{11} = 0$, $\dot{y}_{21} = -0.2$, $\dot{y}_{12} = 0$, $\dot{y}_{22} = 0.2$. The ATT for $k = 1$ is: $\hat{\tau}_1 = 0.6 - (0.7 + 0.5 \cdot 0 + 0.5 \cdot (-0.2)) = 0.6 - 0.6 = 0$, and for $k = 2$: $\hat{\tau}_2 = 0.4 - (0.3 + 0.5 \cdot 0 + 0.5 \cdot 0.2) = 0.4 - 0.4 = 0$. Thus, the sum of the treatment effects is: $\sum_{k=1}^2 \hat{\tau}_k = 0 + 0 = 0$. With equal weights for categories, the compositional constraint $\sum_{k=1}^K \tau_k = 0$ is satisfied.

3 Alternative Formalization of the Constant Control Comparison Under the Demeaned Weighting Estimator

To see whether synthetic control methods satisfy Lemma 1 in the article, we adopt the demeaned weighting estimator (Doudchenko and Imbens 2016).

3.1 Estimator

Definition 1. For each outcome k , the demeaned weighting estimator first computes the pre-treatment averages as $\bar{y}_{i.k} = \frac{1}{T_0} \sum_{t=1}^{T_0} y_{itk}$ and the demeaned outcomes as $\dot{y}_{itk} = y_{itk} - \bar{y}_{i.k}$. The counterfactual outcome for each treatment unit $i \in \{N_{co} + 1, \dots, N\}$ is then estimated as a linear combination of the pre-treatment averages and weighted post-treatment demeaned outcomes from the control units:

$$\hat{y}_{iT_k}(0) = \bar{y}_{i.k} + \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jT_k}, \quad (8)$$

where ω_{jk} are of non-negative weights that sum to one: $\sum_{j=1}^{N_{co}} \omega_{jk} = 1$ for all k . **The demeaned weighting estimator of the ATT** is given by the difference between the average of the observed outcomes for the treated units and the average of the estimated counterfactual outcomes of the treated units:

$$\hat{\tau}_k = \left(\frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N y_{iT_k} \right) - \left(\frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N \left(\bar{y}_{i.k} + \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jT_k} \right) \right). \quad (9)$$

3.2 Constant Control Comparison (CCC)

In case of compositional data, cross-outcome weight invariance is crucial to whether the ATT sum constraint holds. For standard DID, the control weights are the same across outcomes by construction: $\omega_{jk} = \omega_{jk'} = 1/N_{co}$ for all $j \in \{1, \dots, N_{co}\}$ and $k, k' \in \{1, \dots, K\}$. However, SCMs choose weights to match the pre-treatment outcomes, as formally discussed in SM 1. Separate applications of SCMs to proportions can yield weights that can differ across outcomes for the same unit: there exist $k, k' \in \{1, \dots, K\}$ with $k \neq k'$ such that $\omega_{ik} \neq \omega_{ik'}$.

In contrast, we define the constant control comparison for compositional data as follows.

Definition 2. Constant control comparison imposes the same weights on control units for all proportional outcomes $\omega_{ik} = \omega_{ik'}$ for all $k, k' \in \{1, \dots, K\}$ with $k \neq k'$ in the estimation of the counterfactual expectation $\hat{y}_{iT_k}(0)$ using the demeaned weighting estimator (see Definition 1).

The constant control comparison enables the use of the same weights for all proportional outcomes of a unit, denoted as ω_j in the demeaned weighting estimator, rather than specific weights for the proportional outcome ω_{jk} . These common control unit weights across outcomes are a sufficient (though not necessary) condition that the constraint on ATTs for proportional outcomes is met:

Theorem 1 (Sufficiency of constant control comparison). *The constant control comparison ensures that the constraint on the ATTs for proportions holds (see Lemma 1 in the article). Given Definition 2 it holds that the sum of the demeaned weighting estimates of the ATT (see Definition 1) are*

equal to zero:

$$\sum_{k=1}^K \hat{\tau}_k = \sum_{k=1}^K \left(\left(\frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N y_{iTk} \right) - \left(\frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N \left(\bar{y}_{i.k} + \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jTk} \right) \right) \right) = 0 \quad (10)$$

3.3 Proof: Sufficiency of the CCC

We begin with the compositional constraints: $\sum_{k=1}^K y_{itk} = 1$ and $\sum_{k=1}^K \bar{y}_{i.k} = \frac{1}{T_0} \sum_{t=1}^{T_0} \sum_{k=1}^K y_{itk} = 1$. Additionally, we have $\sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jtk} = 0$, which can be shown by expanding the terms: $\sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jtk} = \sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} (y_{jtk} - \bar{y}_{j.k}) = \sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} y_{jtk} - \sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} \bar{y}_{j.k}$. Since the weights ω_{jk} are common across outcome categories k , we can simplify: $\sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} y_{jtk} = \sum_{j=1}^{N_{co}} \omega_j$ and $\sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} \bar{y}_{j.k} = \sum_{j=1}^{N_{co}} \omega_j \sum_{k=1}^K \bar{y}_{j.k} = \sum_{j=1}^{N_{co}} \omega_j$. This leaves us with: $\sum_{k=1}^K \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jtk} = \sum_{j=1}^{N_{co}} \omega_j - \sum_{j=1}^{N_{co}} \omega_j = 0$. Now, we apply these results to the sum of the ATT estimates:

$$\sum_{k=1}^K \hat{\tau}_k = \frac{1}{N_{tr}} \sum_{k=1}^K \left(\sum_{i=N_{co}+1}^N y_{iTk} \right) - \frac{1}{N_{tr}} \sum_{k=1}^K \left(\sum_{i=N_{co}+1}^N \left(\bar{y}_{i.k} + \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jTk} \right) \right).$$

The first term for the treated group: $\frac{1}{N_{tr}} \sum_{k=1}^K \left(\sum_{i=N_{co}+1}^N y_{iTk} \right) = 1$, since $\sum_{k=1}^K y_{iTk} = 1$ for each treated unit. We expand the second term for the control group as: $\frac{1}{N_{tr}} \sum_{k=1}^K \sum_{i=N_{co}+1}^N \bar{y}_{i.k} + \frac{1}{N_{tr}} \sum_{k=1}^K \sum_{i=N_{co}+1}^N \sum_{j=1}^{N_{co}} \omega_{jk} \dot{y}_{jTk}$. The first part simplifies to 1, since $\sum_{k=1}^K \bar{y}_{i.k} = 1$. The second part is 0, as shown earlier. Thus, we have: $\sum_{k=1}^K \hat{\tau}_k = 1 - 1 - 0 = 0$.

4 Algorithm: Synthetic Control Method with Common Weights

Algorithm 1 The Synthetic Control (SC) Estimator with Common Weights

- 1: Compute regularization parameter ξ as in Eq. (4)
- 2: Compute unit weights $\hat{\omega}$ solving minimization:

$$(\hat{\omega}) = \arg \min_{\hat{\omega} \in \mathbb{R}, \hat{\omega} \in \Omega} \ell_{unit}(\omega) \text{ where} \quad (11)$$

$$\ell_{unit}(\omega) = \sum_{k=1}^K \left(\sum_{t=1}^{T_0} \left(\sum_{i=1}^{N_{co}} \omega_i y_{itk} - \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^N y_{itk} \right)^2 \right) + \xi^2 T_0 \|\omega\|_2^2,$$

$$\Omega = \left\{ \omega \in \mathbb{R}_+^N : \sum_{i=1}^{N_{co}} \omega_i = 1, \omega_i = \frac{1}{N_{tr}} \forall i \in \{N_{co} + 1, \dots, N\} \right\}$$

- 3: **for** each outcome $k \in \{1, \dots, K\}$ **do**
- 4: Compute the SC estimate via the weighted regression:

$$(\hat{\tau}_k^{sc}, \hat{\mu}_k, \hat{\beta}_{tk}) = \arg \min_{\tau, \mu, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (y_{itk} - \mu_k - \beta_{tk} - D_{it}\tau)^2 \hat{\omega}_i \right\} \quad (12)$$

- 5: **end for**
-

5 Simulation: Example of the data generating process

Figure [SI1](#) illustrates the average outcomes for the treated and control units over time from an example data set with 400 units, of which 200 are treated at the final time point of the five observed. The figure indicates a treatment effect in proportion one. It also shows that treatment selection based on trends leads to a clear violation of the parallel trends assumption and that with compositional data, the effects we induce for proportion one affect the processes for the other proportional outcomes.

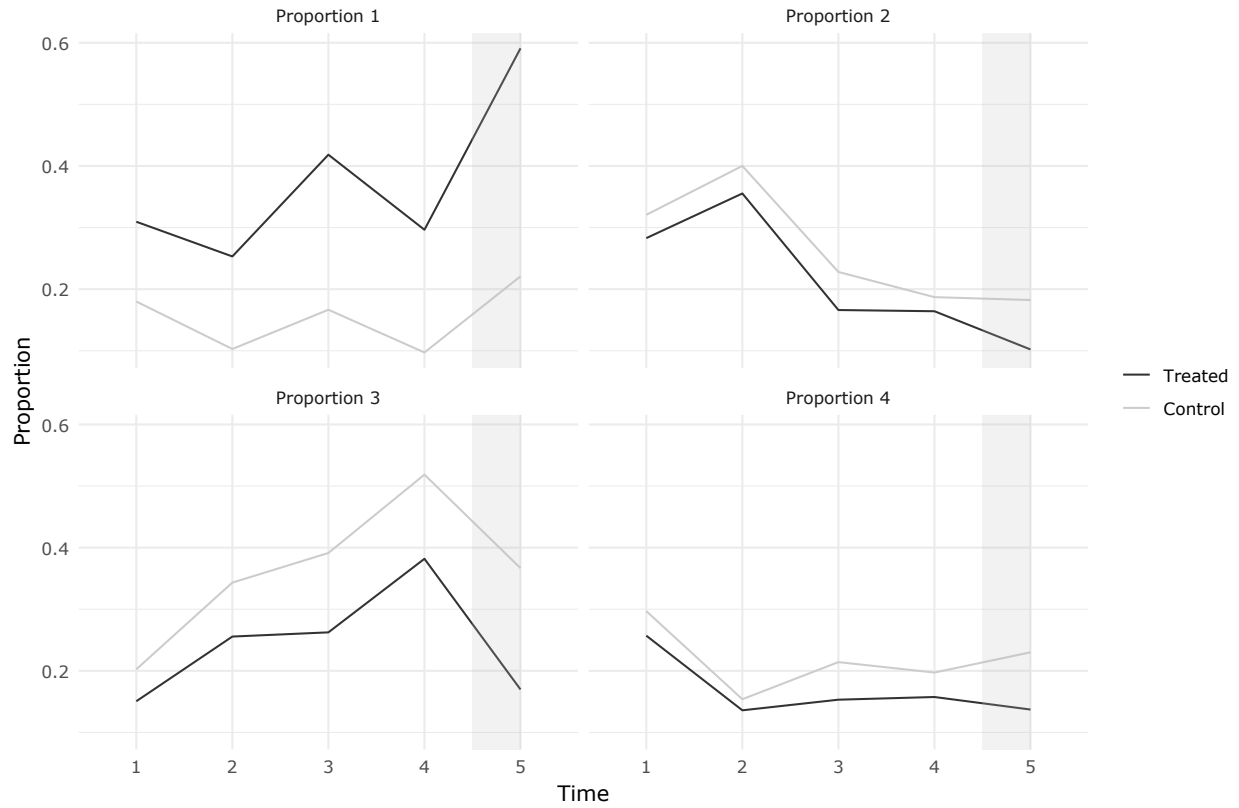


Figure SI1: Example simulation data from the data generating process (DGP). We analyze a situation with 4 proportions and a positive treatment effect on the first proportion. The example is for 400 units (of which 200 are treated) and 5 times points, where in the last time points the treatment is introduced. The first proportion exhibits a local trend, and selection is based on the local trend.

References

- Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager. 2021. “Synthetic Difference-in-Differences.” *American Economic Review* 111 (12): 4088–4118. doi:[10.1257/aer.20190159](https://doi.org/10.1257/aer.20190159).
- Doudchenko, N., and G. W. Imbens. 2016. “Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis.” doi:[10.3386/w22791](https://doi.org/10.3386/w22791). <https://www.nber.org/papers/w22791>. NBER Working Paper.