

# The synth\_runner Package: Utilities to Automate Synthetic Control Estimation Using synth

Sebastian Galiani    Brian Quistorff

Department of Economics  
University of Maryland, College Park

IDB, April 2014

# Introduction

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduced Synthetic Control Methods (SCM), to identifying treatment effects in case-studies.
- Abadie et al. (2010) released synth for Stata to perform a Synthetic Control estimation.
- With the help of the IDB, we have released synth\_runner to automate many of the tasks associated with these types of estimations

# Introduction

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduced Synthetic Control Methods (SCM), to identifying treatment effects in case-studies.
- Abadie et al. (2010) released `synth` for Stata to perform a Synthetic Control estimation.
- With the help of the IDB, we have released `synth_runner` to automate many of the tasks associated with these types of estimations

# Introduction

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduced Synthetic Control Methods (SCM), to identifying treatment effects in case-studies.
- Abadie et al. (2010) released `synth` for Stata to perform a Synthetic Control estimation.
- With the help of the IDB, we have released `synth_runner` to automate many of the tasks associated with these types of estimations

# Outline

## 1 Synthetic Control Methods

- Classic SCM
- Multiple Treatments

## 2 Stata Module

- Single Treatment - Example 1
- Single Treatment - Example 2
- Multiple Treatments - Example 3
- Installation

## 3 Discussion

- Future Work
- Q & A

# Intuition

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduced Synthetic Control Methods (SCM).
- Allows for estimating treatment effects even if treated unit is not on a parallel time trend as the mean of the untreated units.
- Constructs a counterfactual by finding a weighted average of non-treated units that resembles the treated unit in the pre-treatment period.

# Intuition

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduced Synthetic Control Methods (SCM).
- Allows for estimating treatment effects even if treated unit is not on a parallel time trend as the mean of the untreated units.
- Constructs a counterfactual by finding a weighted average of non-treated units that resembles the treated unit in the pre-treatment period.

# Intuition

- Abadie and Gardeazabal (2003) and Abadie et al. (2010) introduced Synthetic Control Methods (SCM).
- Allows for estimating treatment effects even if treated unit is not on a parallel time trend as the mean of the untreated units.
- Constructs a counterfactual by finding a weighted average of non-treated units that resembles the treated unit in the pre-treatment period.

# DGP for Synthetic Controls

Synthetic Controls can be applied in multiple settings, but most common setting where there are theoretical results is:

- Units  $\{1, \dots, J + 1\}$  with 1 being treated and  $J$  untreated “donors”.  $D_{jt}$  is the treatment indicator.  $T_0$  pre-treatment periods and  $T$  total time periods. Factor Model:

$$Y_{jt} = \alpha_{jt} D_{jt} + Y_{jt}^N$$

$$Y_{jt} = \alpha_{jt} D_{jt} + (\theta_t' \mathbf{Z}_j + \delta_t + \lambda_t' \mu_j + \varepsilon_{jt})$$

$\alpha_{jt}$  are time-varying treatment effects,  $Y_{jt}^N$  is the no-treated counterfactual,  $\theta_t$  are unknown parameters,  $\mathbf{Z}_j$  are observed unaffected covariates,  $\delta_t$  is an unknown time factor,  $\lambda_t$  are unknown factors,  $\mu_j$  are unknown factor loadings, and the error  $\varepsilon_{jt}$  is independent across units and time with zero mean.

- Treatment can be correlated w/ unobserved factors, but not  $\varepsilon_{jt}$ .

# DGP for Synthetic Controls

Synthetic Controls can be applied in multiple settings, but most common setting where there are theoretical results is:

- Units  $\{1, \dots, J + 1\}$  with 1 being treated and  $J$  untreated “donors”.  $D_{jt}$  is the treatment indicator.  $T_0$  pre-treatment periods and  $T$  total time periods. Factor Model:

$$Y_{jt} = \alpha_{jt} D_{jt} + Y_{jt}^N$$

$$Y_{jt} = \alpha_{jt} D_{jt} + (\theta_t' \mathbf{Z}_j + \delta_t + \lambda_t' \mu_j + \varepsilon_{jt})$$

$\alpha_{jt}$  are time-varying treatment effects,  $Y_{jt}^N$  is the no-treated counterfactual,  $\theta_t$  are unknown parameters,  $\mathbf{Z}_j$  are observed unaffected covariates,  $\delta_t$  is an unknown time factor,  $\lambda_t$  are unknown factors,  $\mu_j$  are unknown factor loadings, and the error  $\varepsilon_{jt}$  is independent across units and time with zero mean.

- Treatment can be correlated w/ unobserved factors, but not  $\varepsilon_{jt}$ .

# DGP for Synthetic Controls

Synthetic Controls can be applied in multiple settings, but most common setting where there are theoretical results is:

- Units  $\{1, \dots, J + 1\}$  with 1 being treated and  $J$  untreated “donors”.  $D_{jt}$  is the treatment indicator.  $T_0$  pre-treatment periods and  $T$  total time periods. Factor Model:

$$Y_{jt} = \alpha_{jt} D_{jt} + Y_{jt}^N$$

$$Y_{jt} = \alpha_{jt} D_{jt} + (\theta_t' \mathbf{Z}_j + \delta_t + \lambda_t' \mu_j + \varepsilon_{jt})$$

$\alpha_{jt}$  are time-varying treatment effects,  $Y_{jt}^N$  is the no-treated counterfactual,  $\theta_t$  are unknown parameters,  $\mathbf{Z}_j$  are observed unaffected covariates,  $\delta_t$  is an unknown time factor,  $\lambda_t$  are unknown factors,  $\mu_j$  are unknown factor loadings, and the error  $\varepsilon_{jt}$  is independent across units and time with zero mean.

- Treatment can be correlated w/ unobserved factors, but not  $\varepsilon_{jt}$ .

# SC Estimator

- Let  $\mathbf{W}$  be a weight vector over the other units such that  $\mathbf{W} \geq \mathbf{0}$  and  $\sum_j w_j = 1$ .
- Split pre-/post-treatment  $\mathbf{Y}$  as  $(\hat{\mathbf{Y}}_1 \setminus \hat{\mathbf{Y}}_0)$ .
- Let  $\mathbf{Y}_0$  be the  $(T \times J)$  matrix of outcomes for donors (similar for  $\mathbf{Z}_0, \hat{\mathbf{Y}}, \vec{\mathbf{Y}}$ ).
- Suppose a  $\mathbf{W}$  can be found such that the synthetic control matches the treated unit in pre-treatment:

$$\hat{\mathbf{Y}}_1 = \hat{\mathbf{Y}}_0 \mathbf{W}$$

$$\mathbf{Z}_1 = \mathbf{Z}_0 \mathbf{W}$$

and that  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  is non-singular.

- Estimate  $\hat{\mathbf{Y}}_1^N$  as  $\hat{\mathbf{Y}}_0 \mathbf{W}$  so  $\hat{\alpha}_1 = \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}$
- Then  $Bias(\hat{\alpha}) \rightarrow 0$  as  $T_0$  grows large relative to  $\varepsilon_{it}$ .

# SC Estimator

- Let  $\mathbf{W}$  be a weight vector over the other units such that  $\mathbf{W} \geq \mathbf{0}$  and  $\sum_j w_j = 1$ .
- Split pre-/post-treatment  $\mathbf{Y}$  as  $(\hat{\mathbf{Y}}_1 \setminus \hat{\mathbf{Y}}_0)$ .
- Let  $\mathbf{Y}_0$  be the  $(T \times J)$  matrix of outcomes for donors (similar for  $\mathbf{Z}_0, \hat{\mathbf{Y}}, \vec{\mathbf{Y}}$ ).
- Suppose a  $\mathbf{W}$  can be found such that the synthetic control matches the treated unit in pre-treatment:

$$\hat{\mathbf{Y}}_1 = \hat{\mathbf{Y}}_0 \mathbf{W}$$

$$\mathbf{Z}_1 = \mathbf{Z}_0 \mathbf{W}$$

and that  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  is non-singular.

- Estimate  $\hat{\mathbf{Y}}_1^N$  as  $\hat{\mathbf{Y}}_0 \mathbf{W}$  so  $\hat{\alpha}_1 = \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}$
- Then  $Bias(\hat{\alpha}) \rightarrow 0$  as  $T_0$  grows large relative to  $\varepsilon_{it}$ .

# SC Estimator

- Let  $\mathbf{W}$  be a weight vector over the other units such that  $\mathbf{W} \geq \mathbf{0}$  and  $\sum_j w_j = 1$ .
- Split pre-/post-treatment  $\mathbf{Y}$  as  $(\hat{\mathbf{Y}}_1 \setminus \hat{\mathbf{Y}}_0)$ .
- Let  $\mathbf{Y}_0$  be the  $(T \times J)$  matrix of outcomes for donors (similar for  $\mathbf{Z}_0, \hat{\mathbf{Y}}_0, \hat{\mathbf{Y}}_1$ ).
- Suppose a  $\mathbf{W}$  can be found such that the synthetic control matches the treated unit in pre-treatment:

$$\hat{\mathbf{Y}}_1 = \hat{\mathbf{Y}}_0 \mathbf{W}$$

$$\mathbf{Z}_1 = \mathbf{Z}_0 \mathbf{W}$$

and that  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  is non-singular.

- Estimate  $\hat{\mathbf{Y}}_1^N$  as  $\hat{\mathbf{Y}}_0 \mathbf{W}$  so  $\hat{\alpha}_1 = \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}$
- Then  $Bias(\hat{\alpha}) \rightarrow 0$  as  $T_0$  grows large relative to  $\varepsilon_{it}$ .

# SC Estimator

- Let  $\mathbf{W}$  be a weight vector over the other units such that  $\mathbf{W} \geq \mathbf{0}$  and  $\sum_j w_j = 1$ .
- Split pre-/post-treatment  $\mathbf{Y}$  as  $(\hat{\mathbf{Y}}_1 \setminus \hat{\mathbf{Y}}_0)$ .
- Let  $\mathbf{Y}_0$  be the  $(T \times J)$  matrix of outcomes for donors (similar for  $\mathbf{Z}_0, \hat{\mathbf{Y}}_0, \hat{\mathbf{Y}}_1$ ).
- Suppose a  $\mathbf{W}$  can be found such that the synthetic control matches the treated unit in pre-treatment:

$$\hat{\mathbf{Y}}_1 = \hat{\mathbf{Y}}_0 \mathbf{W}$$

$$\mathbf{Z}_1 = \mathbf{Z}_0 \mathbf{W}$$

and that  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  is non-singular.

- Estimate  $\hat{\mathbf{Y}}_1^N$  as  $\hat{\mathbf{Y}}_0 \mathbf{W}$  so  $\hat{\alpha}_1 = \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}$
- Then  $Bias(\hat{\alpha}) \rightarrow 0$  as  $T_0$  grows large relative to  $\varepsilon_{it}$ .

# SC Estimator

- Let  $\mathbf{W}$  be a weight vector over the other units such that  $\mathbf{W} \geq \mathbf{0}$  and  $\sum_j w_j = 1$ .
- Split pre-/post-treatment  $\mathbf{Y}$  as  $(\bar{\mathbf{Y}}_1 \setminus \bar{\mathbf{Y}}_0)$ .
- Let  $\bar{\mathbf{Y}}_0$  be the  $(T \times J)$  matrix of outcomes for donors (similar for  $\mathbf{Z}_0, \bar{\mathbf{Y}}, \bar{\mathbf{Y}}$ ).
- Suppose a  $\mathbf{W}$  can be found such that the synthetic control matches the treated unit in pre-treatment:

$$\bar{\mathbf{Y}}_1 = \bar{\mathbf{Y}}_0 \mathbf{W}$$

$$\mathbf{Z}_1 = \mathbf{Z}_0 \mathbf{W}$$

and that  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  is non-singular.

- Estimate  $\bar{\mathbf{Y}}_1^N$  as  $\bar{\mathbf{Y}}_0 \mathbf{W}$  so  $\hat{\alpha}_1 = \bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_0 \mathbf{W}$
- Then  $Bias(\hat{\alpha}) \rightarrow 0$  as  $T_0$  grows large relative to  $\varepsilon_{it}$ .

# SC Estimator

- Let  $\mathbf{W}$  be a weight vector over the other units such that  $\mathbf{W} \geq \mathbf{0}$  and  $\sum_j w_j = 1$ .
- Split pre-/post-treatment  $\mathbf{Y}$  as  $(\bar{\mathbf{Y}}_1 \setminus \bar{\mathbf{Y}}_0)$ .
- Let  $\bar{\mathbf{Y}}_0$  be the  $(T \times J)$  matrix of outcomes for donors (similar for  $\mathbf{Z}_0, \bar{\mathbf{Y}}, \bar{\mathbf{Y}}$ ).
- Suppose a  $\mathbf{W}$  can be found such that the synthetic control matches the treated unit in pre-treatment:

$$\bar{\mathbf{Y}}_1 = \bar{\mathbf{Y}}_0 \mathbf{W}$$

$$\mathbf{Z}_1 = \mathbf{Z}_0 \mathbf{W}$$

and that  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  is non-singular.

- Estimate  $\bar{\mathbf{Y}}_1^N$  as  $\bar{\mathbf{Y}}_0 \mathbf{W}$  so  $\hat{\alpha}_1 = \bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_0 \mathbf{W}$
- Then  $Bias(\hat{\alpha}) \rightarrow 0$  as  $T_0$  grows large relative to  $\varepsilon_{it}$ .

# SC Estimation

- Let “predictors”  $\hat{\mathbf{X}}$  be comprised of  $\mathbf{Z}$  and some linear combinations of  $\hat{\mathbf{Y}}$ .
- Define  $\|\mathbf{A}\| = \sqrt{\text{Acols}(\mathbf{A})^{-1}\mathbf{A}}$  and  $\|\mathbf{A}\|_{\mathbf{V}} = \sqrt{\mathbf{A}\mathbf{V}^{-1}\mathbf{A}}$ .
  - $\|\hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0\mathbf{W}\| = s_1$  is the root mean squared prediction error (RMSPE).
- As matching may only hold approximately, we need a set of predictor weights  $\mathbf{V}$  that prioritizes which variables to match better.
- Given  $\mathbf{V}$ , then

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_0\mathbf{W}\|_{\mathbf{V}}$$

- Multiple options for  $\mathbf{V}$  (synth includes regression weights and nested optimization  $\mathbf{V}^* = \arg \min_{\mathbf{V}} \left\| \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0\mathbf{W}(\mathbf{V}) \right\| \right)$

# SC Estimation

- Let “predictors”  $\hat{\mathbf{X}}$  be comprised of  $\mathbf{Z}$  and some linear combinations of  $\hat{\mathbf{Y}}$ .
- Define  $\|\mathbf{A}\| = \sqrt{\mathbf{A} \text{cols}(\mathbf{A})^{-1} \mathbf{A}}$  and  $\|\mathbf{A}\|_{\mathbf{V}} = \sqrt{\mathbf{A} \mathbf{V}^{-1} \mathbf{A}}$ .
  - $\|\mathbf{Y}_1 - \mathbf{Y}_0 \mathbf{W}\| = s_1$  is the root mean squared prediction error (RMSPE).
- As matching may only hold approximately, we need a set of predictor weights  $\mathbf{V}$  that prioritizes which variables to match better.
- Given  $\mathbf{V}$ , then

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_{\mathbf{V}}$$

- Multiple options for  $\mathbf{V}$  (synth includes regression weights and nested optimization  $\mathbf{V}^* = \arg \min_{\mathbf{V}} \left\| \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}(\mathbf{V}) \right\| \right)$

# SC Estimation

- Let “predictors”  $\hat{\mathbf{X}}$  be comprised of  $\mathbf{Z}$  and some linear combinations of  $\hat{\mathbf{Y}}$ .
- Define  $\|\mathbf{A}\| = \sqrt{\mathbf{A} \text{cols}(\mathbf{A})^{-1} \mathbf{A}}$  and  $\|\mathbf{A}\|_{\mathbf{V}} = \sqrt{\mathbf{A} \mathbf{V}^{-1} \mathbf{A}}$ .
  - $\|\mathbf{Y}_1 - \mathbf{Y}_0 \mathbf{W}\| = s_1$  is the root mean squared prediction error (RMSPE).
- As matching may only hold approximately, we need a set of predictor weights  $\mathbf{V}$  that prioritizes which variables to match better.
- Given  $\mathbf{V}$ , then

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_{\mathbf{V}}$$

- Multiple options for  $\mathbf{V}$  (synth includes regression weights and nested optimization  $\mathbf{V}^* = \arg \min_{\mathbf{V}} \left\| \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}(\mathbf{V}) \right\| \right)$

# SC Estimation

- Let “predictors”  $\hat{\mathbf{X}}$  be comprised of  $\mathbf{Z}$  and some linear combinations of  $\hat{\mathbf{Y}}$ .
- Define  $\|\mathbf{A}\| = \sqrt{\mathbf{A} \text{cols}(\mathbf{A})^{-1} \mathbf{A}}$  and  $\|\mathbf{A}\|_{\mathbf{V}} = \sqrt{\mathbf{A} \mathbf{V}^{-1} \mathbf{A}}$ .
  - $\|\hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}\| = s_1$  is the root mean squared prediction error (RMSPE).
- As matching may only hold approximately, we need a set of predictor weights  $\mathbf{V}$  that prioritizes which variables to match better.
- Given  $\mathbf{V}$ , then

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_0 \mathbf{W}\|_{\mathbf{V}}$$

- Multiple options for  $\mathbf{V}$  (synth includes regression weights and nested optimization  $\mathbf{V}^* = \arg \min_{\mathbf{V}} \left\| \hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}(\mathbf{V}) \right\| \right)$

# SC Estimation

- Let “predictors”  $\hat{\mathbf{X}}$  be comprised of  $\mathbf{Z}$  and some linear combinations of  $\hat{\mathbf{Y}}$ .
- Define  $\|\mathbf{A}\| = \sqrt{\mathbf{A} \text{cols}(\mathbf{A})^{-1} \mathbf{A}}$  and  $\|\mathbf{A}\|_{\mathbf{V}} = \sqrt{\mathbf{A} \mathbf{V}^{-1} \mathbf{A}}$ .
  - $\|\hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}\| = s_1$  is the root mean squared prediction error (RMSPE).
- As matching may only hold approximately, we need a set of predictor weights  $\mathbf{V}$  that prioritizes which variables to match better.
- Given  $\mathbf{V}$ , then

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_0 \mathbf{W}\|_{\mathbf{V}}$$

- Multiple options for  $\mathbf{V}$  (synth includes regression weights and nested optimization  $\mathbf{V}^* = \arg \min_{\mathbf{V}} \|\hat{\mathbf{Y}}_1 - \hat{\mathbf{Y}}_0 \mathbf{W}(\mathbf{V})\|$ )

## DGP II

- Factor model ( $\lambda_t' \mu_j$ ) accommodates many forms:
  - Standard panel model with time and units fixed effects from  $\mu_j = (1, \gamma_j)$  and  $\lambda_t = (\xi_t, 1)'$
  - Unit-specific (non-parallel) time trends:  $\mu_j = (\gamma_j)$  and  $\lambda_t = (t)$
- Need  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  non-singular. Implies:
  - All factors must have been “active” in the past (one can not be all 0).
  - Number of factors can be no more than  $T_0$  (Cauchy-Binet formula).

## DGP II

- Factor model  $(\lambda_t' \mu_j)$  accommodates many forms:
  - Standard panel model with time and units fixed effects from  $\mu_j = (1, \gamma_j)$  and  $\lambda_t = (\xi_t, 1)'$
  - Unit-specific (non-parallel) time trends:  $\mu_j = (\gamma_j)$  and  $\lambda_t = (t)$
- Need  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  non-singular. Implies:
  - All factors must have been “active” in the past (one can not be all 0).
  - Number of factors can be no more than  $T_0$  (Cauchy-Binet formula).

## DGP II

- Factor model ( $\lambda_t' \mu_j$ ) accommodates many forms:
  - Standard panel model with time and units fixed effects from  $\mu_j = (1, \gamma_j)$  and  $\lambda_t = (\xi_t, 1)'$
  - Unit-specific (non-parallel) time trends:  $\mu_j = (\gamma_j)$  and  $\lambda_t = (t)$
- Need  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  non-singular. Implies:
  - All factors must have been “active” in the past (one can not be all 0).
  - Number of factors can be no more than  $T_0$  (Cauchy-Binet formula).

## DGP II

- Factor model ( $\lambda_t' \mu_j$ ) accommodates many forms:
  - Standard panel model with time and units fixed effects from  $\mu_j = (1, \gamma_j)$  and  $\lambda_t = (\xi_t, 1)'$
  - Unit-specific (non-parallel) time trends:  $\mu_j = (\gamma_j)$  and  $\lambda_t = (t)$
- Need  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  non-singular. Implies:
  - All factors must have been “active” in the past (one can not be all 0).
  - Number of factors can be no more than  $T_0$  (Cauchy-Binet formula).

## DGP II

- Factor model  $(\lambda_t' \mu_j)$  accommodates many forms:
  - Standard panel model with time and units fixed effects from  $\mu_j = (1, \gamma_j)$  and  $\lambda_t = (\xi_t, 1)'$
  - Unit-specific (non-parallel) time trends:  $\mu_j = (\gamma_j)$  and  $\lambda_t = (t)$
- Need  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  non-singular. Implies:
  - All factors must have been “active” in the past (one can not be all 0).
  - Number of factors can be no more than  $T_0$  (Cauchy-Binet formula).

## DGP II

- Factor model ( $\lambda_t' \mu_j$ ) accommodates many forms:
  - Standard panel model with time and units fixed effects from  $\mu_j = (1, \gamma_j)$  and  $\lambda_t = (\xi_t, 1)'$
  - Unit-specific (non-parallel) time trends:  $\mu_j = (\gamma_j)$  and  $\lambda_t = (t)$
- Need  $\sum_{t=1}^{T_0} \lambda_t \lambda_t'$  non-singular. Implies:
  - All factors must have been “active” in the past (one can not be all 0).
  - Number of factors can be no more than  $T_0$  (Cauchy-Binet formula).

# Inference

- SCM relies on in-place placebo (permutation) tests for inference.
- Re-estimate synthetic controls on all  $J$  donors to get  $\{\hat{\alpha}_p\}$ . Call this the placebo distribution  $\hat{\alpha}_1^{PL}$ .
- Compare  $\hat{\alpha}_1$  to  $\hat{\alpha}_1^{PL}$ . The two-sided  $p$ -value for  $\hat{\alpha}_{1t}$  is then

$$\begin{aligned} p\text{-value} &= \Pr(|\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}|) \\ &= \frac{\sum_{j \neq 1} 1(|\hat{\alpha}_{jt}| \geq |\hat{\alpha}_{1t}|)}{J} \end{aligned}$$

the one-sided  $p$ -values (for positive effects) are

$$p\text{-value} = \Pr(\hat{\alpha}_{1t}^{PL} \geq \hat{\alpha}_{1t})$$

- Can assess significant of joint post-treatment effects by comparing  $\vec{s}_1$  to  $\{\vec{s}_p\}_p$ .

# Inference

- SCM relies on in-place placebo (permutation) tests for inference.
- Re-estimate synthetic controls on all  $J$  donors to get  $\{\hat{\alpha}_p\}$ . Call this the placebo distribution  $\hat{\alpha}_1^{PL}$ .
- Compare  $\hat{\alpha}_1$  to  $\hat{\alpha}_1^{PL}$ . The two-sided  $p$ -value for  $\hat{\alpha}_{1t}$  is then

$$\begin{aligned} p\text{-value} &= \Pr(|\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}|) \\ &= \frac{\sum_{j \neq 1} 1(|\hat{\alpha}_{jt}| \geq |\hat{\alpha}_{1t}|)}{J} \end{aligned}$$

the one-sided  $p$ -values (for positive effects) are

$$p\text{-value} = \Pr(\hat{\alpha}_{1t}^{PL} \geq \hat{\alpha}_{1t})$$

- Can assess significant of joint post-treatment effects by comparing  $\vec{s}_1$  to  $\{\vec{s}_p\}_p$ .

# Inference

- SCM relies on in-place placebo (permutation) tests for inference.
- Re-estimate synthetic controls on all  $J$  donors to get  $\{\hat{\alpha}_p\}$ . Call this the placebo distribution  $\hat{\alpha}_1^{PL}$ .
- Compare  $\hat{\alpha}_1$  to  $\hat{\alpha}_1^{PL}$ . The two-sided  $p$ -value for  $\hat{\alpha}_{1t}$  is then

$$\begin{aligned} p\text{-value} &= \Pr(|\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}|) \\ &= \frac{\sum_{j \neq 1} 1(|\hat{\alpha}_{jt}| \geq |\hat{\alpha}_{1t}|)}{J} \end{aligned}$$

the one-sided  $p$ -values (for positive effects) are

$$p\text{-value} = \Pr(\hat{\alpha}_{1t}^{PL} \geq \hat{\alpha}_{1t})$$

- Can assess significant of joint post-treatment effects by comparing  $\vec{s}_1$  to  $\{\vec{s}_p\}_p$ .

# Inference

- SCM relies on in-place placebo (permutation) tests for inference.
- Re-estimate synthetic controls on all  $J$  donors to get  $\{\hat{\alpha}_p\}$ . Call this the placebo distribution  $\hat{\alpha}_1^{PL}$ .
- Compare  $\hat{\alpha}_1$  to  $\hat{\alpha}_1^{PL}$ . The two-sided  $p$ -value for  $\hat{\alpha}_{1t}$  is then

$$\begin{aligned} p\text{-value} &= \Pr(|\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}|) \\ &= \frac{\sum_{j \neq 1} 1(|\hat{\alpha}_{jt}| \geq |\hat{\alpha}_{1t}|)}{J} \end{aligned}$$

the one-sided  $p$ -values (for positive effects) are

$$p\text{-value} = \Pr(\hat{\alpha}_{1t}^{PL} \geq \hat{\alpha}_{1t})$$

- Can assess significant of joint post-treatment effects by comparing  $\vec{s}_1$  to  $\{\vec{s}_p\}_p$ .

# Interpretation of Inference

- Interpretation is usually different than with randomization inference. Here it is “what proportion of untreated units have calculated effects as large?”
- Randomization tests have  $J + 1$  in the denominator.
- Could construct confidence intervals though interpretation is again slightly different (shows the variability of the placebo distribution).

# Interpretation of Inference

- Interpretation is usually different than with randomization inference. Here it is “what proportion of untreated units have calculated effects as large?”
- Randomization tests have  $J + 1$  in the denominator.
- Could construct confidence intervals though interpretation is again slightly different (shows the variability of the placebo distribution).

# Interpretation of Inference

- Interpretation is usually different than with randomization inference. Here it is “what proportion of untreated units have calculated effects as large?”
- Randomization tests have  $J + 1$  in the denominator.
- Could construct confidence intervals though interpretation is again slightly different (shows the variability of the placebo distribution).

# Accounting for Match Quality

- Post-treatment differences may be large just because of poor match quality.
- Can take quality of match for pre-treatment period into account:
  - Limit donor pool to cases that matched as well as the treated unit was matched (limit by  $\bar{s}_1$  or some multiple  $m$  of that).
  - Construct “pseudo  $t$ -stats” as in  $\hat{\alpha}_1/\bar{s}_1$  and  $\bar{s}_1/\bar{s}_1$  (similar to Studentizing).
- Inference can then be conducted over four quantities  $(\hat{\alpha}_{jt}, \bar{s}_j, \hat{\alpha}_{jt}/\bar{s}_j, \bar{s}_j/\bar{s}_j)$  and the comparison set can also be limited by the choice of  $m$ .

# Accounting for Match Quality

- Post-treatment differences may be large just because of poor match quality.
- Can take quality of match for pre-treatment period into account:
  - Limit donor pool to cases that matched as well as the treated unit was matched (limit by  $\bar{s}_1$  or some multiple  $m$  of that).
  - Construct “pseudo  $t$ -stats” as in  $\hat{\alpha}_1/\bar{s}_1$  and  $\bar{s}_1/\bar{s}_1$  (similar to Studentizing).
- Inference can then be conducted over four quantities  $(\hat{\alpha}_{jt}, \bar{s}_j, \hat{\alpha}_{jt}/\bar{s}_j, \bar{s}_j/\bar{s}_j)$  and the comparison set can also be limited by the choice of  $m$ .

# Accounting for Match Quality

- Post-treatment differences may be large just because of poor match quality.
- Can take quality of match for pre-treatment period into account:
  - Limit donor pool to cases that matched as well as the treated unit was matched (limit by  $\bar{s}_1$  or some multiple  $m$  of that).
  - Construct “pseudo  $t$ -stats” as in  $\hat{\alpha}_1/\bar{s}_1$  and  $\bar{s}_1/\bar{s}_1$  (similar to Studentizing).
- Inference can then be conducted over four quantities  $(\hat{\alpha}_{jt}, \bar{s}_j, \hat{\alpha}_{jt}/\bar{s}_j, \bar{s}_j/\bar{s}_j)$  and the comparison set can also be limited by the choice of  $m$ .

# Accounting for Match Quality

- Post-treatment differences may be large just because of poor match quality.
- Can take quality of match for pre-treatment period into account:
  - Limit donor pool to cases that matched as well as the treated unit was matched (limit by  $\bar{s}_1$  or some multiple  $m$  of that).
  - Construct “pseudo  $t$ -stats” as in  $\hat{\alpha}_1/\bar{s}_1$  and  $\bar{s}_1/\bar{s}_1$  (similar to Studentizing).
- Inference can then be conducted over four quantities  $(\hat{\alpha}_{jt}, \bar{s}_j, \hat{\alpha}_{jt}/\bar{s}_j, \bar{s}_j/\bar{s}_j)$  and the comparison set can also be limited by the choice of  $m$ .

# Accounting for Match Quality

- Post-treatment differences may be large just because of poor match quality.
- Can take quality of match for pre-treatment period into account:
  - Limit donor pool to cases that matched as well as the treated unit was matched (limit by  $\bar{s}_1$  or some multiple  $m$  of that).
  - Construct “pseudo  $t$ -stats” as in  $\hat{\alpha}_1/\bar{s}_1$  and  $\bar{s}_1/\bar{s}_1$  (similar to Studentizing).
- Inference can then be conducted over four quantities  $(\hat{\alpha}_{jt}, \bar{s}_j, \hat{\alpha}_{jt}/\bar{s}_j, \bar{s}_j/\bar{s}_j)$  and the comparison set can also be limited by the choice of  $m$ .

# Multiple Treatments - Setup

- Cavallo et al. (2013) analyze multiple independent treatments at possibly different time periods.
- Index treated units  $g \in \{1\dots G\}$ . Let  $J$  be those units that never undergo treatment.
- Index effects relative to treatment date.

# Multiple Treatments - Setup

- Cavallo et al. (2013) analyze multiple independent treatments at possibly different time periods.
- Index treated units  $g \in \{1\dots G\}$ . Let  $J$  be those units that never undergo treatment.
- Index effects relative to treatment date.

# Multiple Treatments - Setup

- Cavallo et al. (2013) analyze multiple independent treatments at possibly different time periods.
- Index treated units  $g \in \{1\dots G\}$ . Let  $J$  be those units that never undergo treatment.
- Index effects relative to treatment date.

# Multiple Treatments - Setup

- Cavallo et al. (2013) analyze multiple independent treatments at possibly different time periods.
- Index treated units  $g \in \{1\dots G\}$ . Let  $J$  be those units that never undergo treatment.
- Index effects relative to treatment date.

# Multiple Treatments - Estimation

For each  $g \in \{1, \dots, G\}$ :

- Estimate vector  $\hat{\alpha}_g$
- Estimate set of vectors  $\hat{\alpha}_g^{PL} = \{\hat{\alpha}_j^g\}_{j \in J}$  assuming treatment date same as  $g$
- (If two treated units have the same treatment period, then their placebo sets will be the same.)

Define vector  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$

# Multiple Treatments - Estimation

For each  $g \in \{1, \dots, G\}$ :

- Estimate vector  $\hat{\alpha}_g$
- Estimate set of vectors  $\hat{\alpha}_g^{PL} = \{\hat{\alpha}_j^g\}_{j \in J}$  assuming treatment date same as  $g$
- (If two treated units have the same treatment period, then their placebo sets will be the same.)

Define vector  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$

# Multiple Treatments - Estimation

For each  $g \in \{1, \dots, G\}$ :

- Estimate vector  $\hat{\alpha}_g$
- Estimate set of vectors  $\hat{\alpha}_g^{PL} = \{\hat{\alpha}_j^g\}_{j \in J}$  assuming treatment date same as  $g$
- (If two treated units have the same treatment period, then their placebo sets will be the same.)

Define vector  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$

# Multiple Treatments - Estimation

For each  $g \in \{1, \dots, G\}$ :

- Estimate vector  $\hat{\alpha}_g$
- Estimate set of vectors  $\hat{\alpha}_g^{PL} = \{\hat{\alpha}_j^g\}_{j \in J}$  assuming treatment date same as  $g$
- (If two treated units have the same treatment period, then their placebo sets will be the same.)

Define vector  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$

# Multiple Treatments - Estimation II

Defining  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$  smooths out noise in the estimate so one needs to do the same with the elements of the placebo set.

- Define a vector of choices from each placebo set  $\tilde{j} = (j_1, \dots, j_G)$  and the set of all such choices  $\tilde{J}$ .
  - $|\tilde{J}| = J^G$  unless limited by  $\vec{s}$ , in which case  $\prod_{g=1}^G J_g$
- An average of picks is  $\bar{\alpha}(\tilde{j}) = G^{-1} \sum_{g \in G} \hat{\alpha}_{\tilde{j}[g]}^g$
- Define set of vectors  $\bar{\alpha}^{PL} = \{\bar{\alpha}(\tilde{j}) | \tilde{j} \in \tilde{J}\}$

# Multiple Treatments - Estimation II

Defining  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$  smooths out noise in the estimate so one needs to do the same with the elements of the placebo set.

- Define a vector of choices from each placebo set  $\tilde{j} = (j_1, \dots, j_G)$  and the set of all such choices  $\tilde{J}$ .
  - $|\tilde{J}| = J^G$  unless limited by  $\vec{s}$ , in which case  $\prod_{g=1}^G J_g$
- An average of picks is  $\bar{\alpha}(\tilde{j}) = G^{-1} \sum_{g \in G} \hat{\alpha}_{\tilde{j}[g]}^g$
- Define set of vectors  $\bar{\alpha}^{PL} = \{\bar{\alpha}(\tilde{j}) | \tilde{j} \in \tilde{J}\}$

# Multiple Treatments - Estimation II

Defining  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$  smooths out noise in the estimate so one needs to do the same with the elements of the placebo set.

- Define a vector of choices from each placebo set  $\tilde{j} = (j_1, \dots, j_G)$  and the set of all such choices  $\tilde{J}$ .
  - $|\tilde{J}| = J^G$  unless limited by  $\vec{s}$ , in which case  $\prod_{g=1}^G J_g$
- An average of picks is  $\bar{\alpha}(\tilde{j}) = G^{-1} \sum_{g \in G} \hat{\alpha}_{\tilde{j}[g]}^g$
- Define set of vectors  $\bar{\alpha}^{PL} = \{\bar{\alpha}(\tilde{j}) | \tilde{j} \in \tilde{J}\}$

# Multiple Treatments - Estimation II

Defining  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$  smooths out noise in the estimate so one needs to do the same with the elements of the placebo set.

- Define a vector of choices from each placebo set  $\tilde{j} = (j_1, \dots, j_G)$  and the set of all such choices  $\tilde{J}$ .
  - $|\tilde{J}| = J^G$  unless limited by  $\vec{s}$ , in which case  $\prod_{g=1}^G J_g$
- An average of picks is  $\bar{\alpha}(\tilde{j}) = G^{-1} \sum_{g \in G} \hat{\alpha}_{\tilde{j}[g]}^g$
- Define set of vectors  $\bar{\alpha}^{PL} = \{\bar{\alpha}(\tilde{j}) | \tilde{j} \in \tilde{J}\}$

# Multiple Treatments - Estimation II

Defining  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$  smooths out noise in the estimate so one needs to do the same with the elements of the placebo set.

- Define a vector of choices from each placebo set  $\tilde{j} = (j_1, \dots, j_G)$  and the set of all such choices  $\tilde{J}$ .
  - $|\tilde{J}| = J^G$  unless limited by  $\vec{s}$ , in which case  $\prod_{g=1}^G J_g$
- An average of picks is  $\bar{\alpha}(\tilde{j}) = G^{-1} \sum_{g \in G} \hat{\alpha}_{\tilde{j}[g]}^g$
- Define set of vectors  $\bar{\alpha}^{PL} = \{\bar{\alpha}(\tilde{j}) | \tilde{j} \in \tilde{J}\}$

# Multiple Treatments - Estimation II

Defining  $\bar{\alpha} = G^{-1} \sum_{g \in G} \hat{\alpha}_g$  smooths out noise in the estimate so one needs to do the same with the elements of the placebo set.

- Define a vector of choices from each placebo set  $\tilde{j} = (j_1, \dots, j_G)$  and the set of all such choices  $\tilde{J}$ .
  - $|\tilde{J}| = J^G$  unless limited by  $\vec{s}$ , in which case  $\prod_{g=1}^G J_g$
- An average of picks is  $\bar{\alpha}(\tilde{j}) = G^{-1} \sum_{g \in G} \hat{\alpha}_{\tilde{j}[g]}^g$
- Define set of vectors  $\bar{\alpha}^{PL} = \{\bar{\alpha}(\tilde{j}) | \tilde{j} \in \tilde{J}\}$

# synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

# synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

# synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

# synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

# synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

# synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

## synth\_runner Module

Abadie et al. (2010) introduce `synth` which conducts a single estimation. Building on that, `synth_runner`:

- Runs placebo tests and outputs *p*-values and confidence intervals
- Allows for matching on trends in the outcome variable rather than on the level
- Handles multiple-treatments
- Automates the process of splitting pre-treatment periods into “training” and “validation” sections
- Provides diagnostics
- Graphs results

# Run estimation similar to ADH10

```
*Data setup the same for all examples
sysuse smoking
tsset state year

tempfile keepfile
*similar syntax to -synth-
synth_runner cigsale beer(1984(1)1988) lnincome(1972(1)
1988) retprice age15to24 cigsale(1988) cigsale(1980)
cigsale(1975), trunit(3) trperiod(1989) keep(`keepfile'
')
```

# Returned Values

## Post-treatment joint significance

```
e(pval_joint_post) = .1315789473684211  
e(pval_joint_post_t) = 0
```

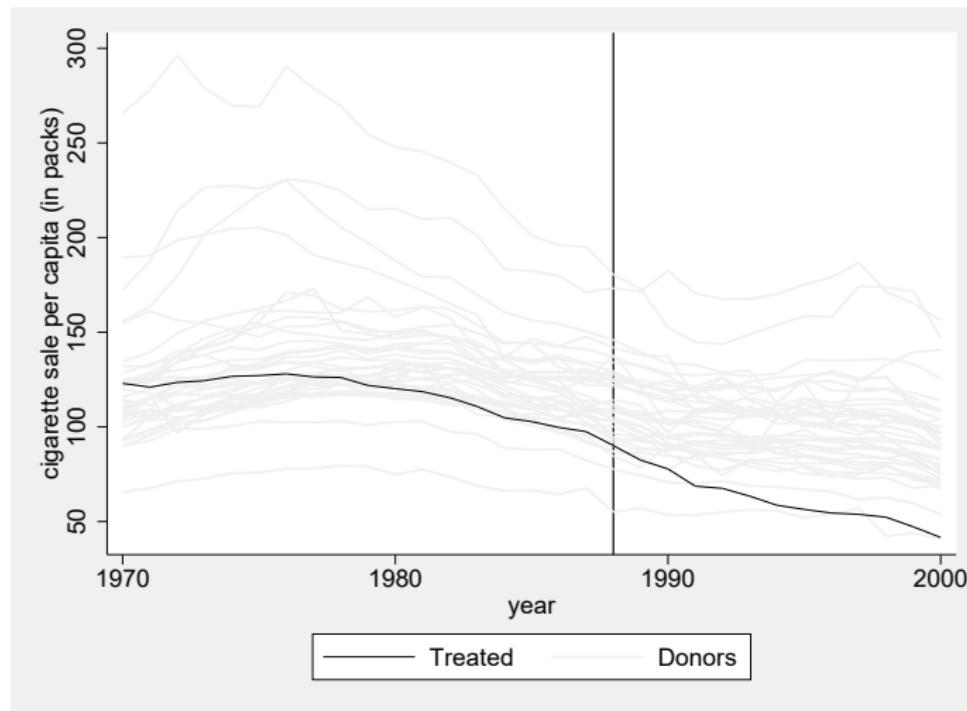
Diagnostics. If too small then SCM may not be appropriate for this unit. If conducting over multiple treatments, good to assess individual match quality and discard some units.

```
e(avg_pre_rmspe_p) = .9210526315789474
```

# Make Graphs

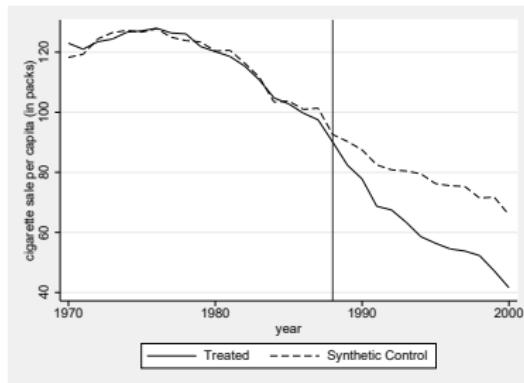
```
*`keepfile' now contains the differences between units and  
    their synthetic controls  
merge 1:1 state year using `keepfile', nogenerate  
gen cigsale_synth = cigsale-effect  
single_treatment_graphs, depvar(cigsale) trunit(3)  
    trperiod(1989) effects_ylabels(-30(10)30) effects_ymax  
    (35) effects_ymin(-35)  
effect_graphs , depvar(cigsale) depvar_synth(cigsale_synth  
    ) trunit(3) trperiod(1989) effect_var(effect)  
pval_graphs
```

# Outcome Trends

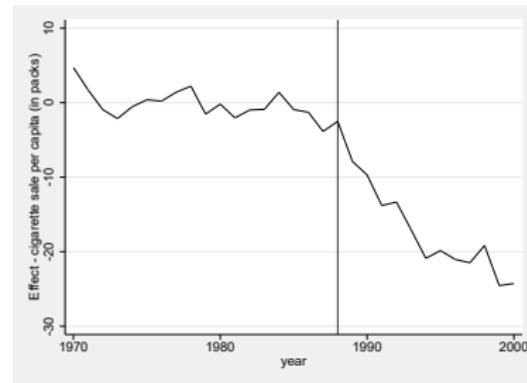


NB: Not a full test of parallel trends assumption as does not account for regression adjustment done in DiD.

# Treated and Synthetic Control

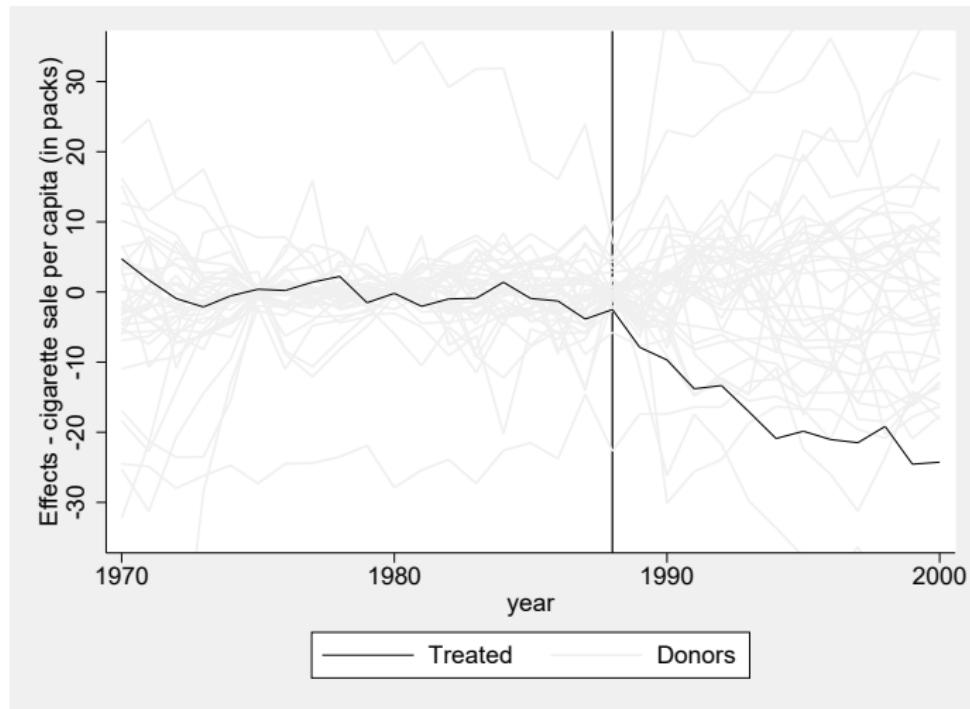


(a) Treated and Control

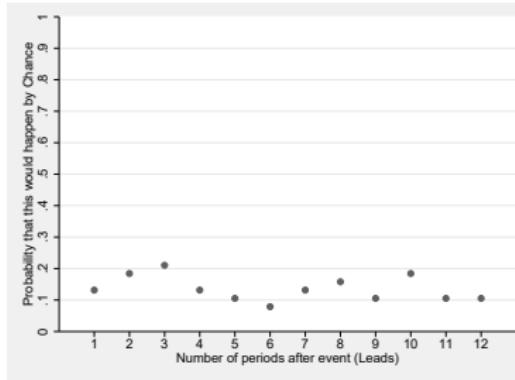


(b) Difference

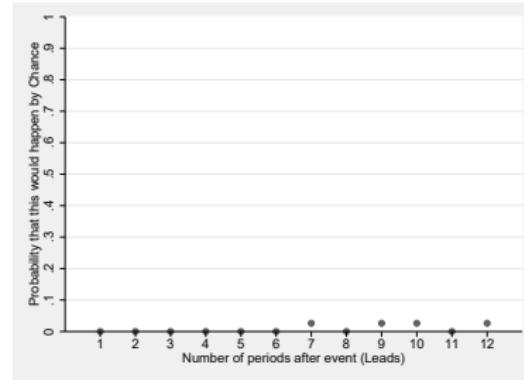
# Differences



# P-Values



(c) Effect



(d) Studentized Effect

# More complicated example

```
gen byte D = (state==3 & year>=1989)
tempfile keepfile2
synth_runner cigsale beer(1984(1)1988) lnincome(1972(1)
1988) retprice age15to24, trunit(3) trperiod(1989)
trends training_propr(`=13/18') pre_limit_mult(10)
keep(`keepfile2')
```

# Returned Values

Post-treatment joint significance

```
e(pval_joint_post) = .0263157894736842  
e(pval_joint_post_t) = 0
```

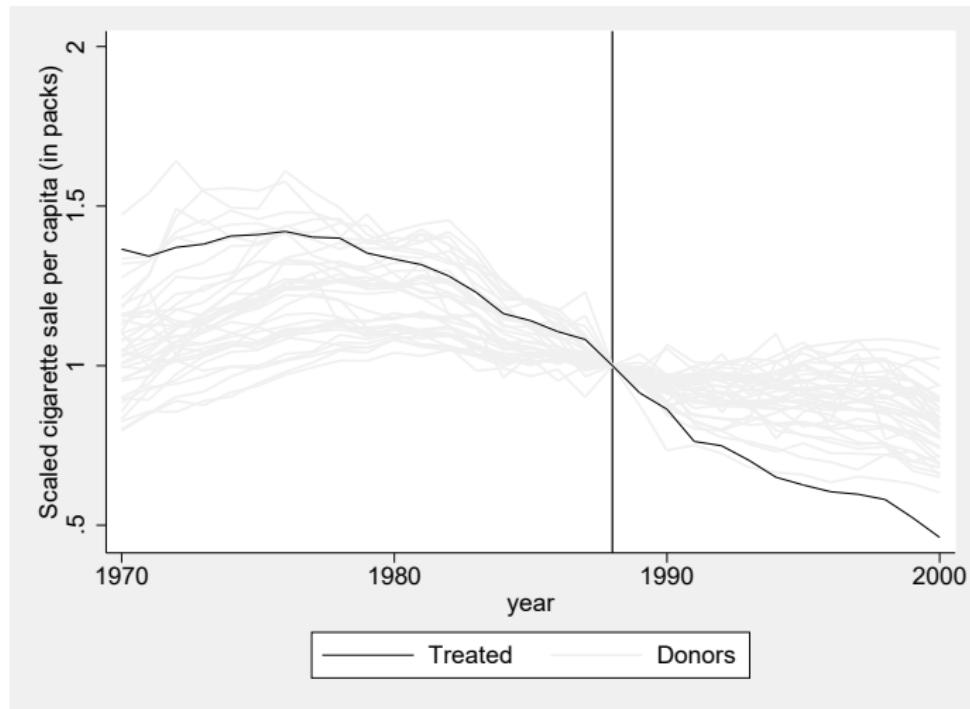
Diagnostics (want large values)

```
e(avg_pre_rmspe_p) = .631578947368421  
e(avg_val_rmspe_p) = .8421052631578947
```

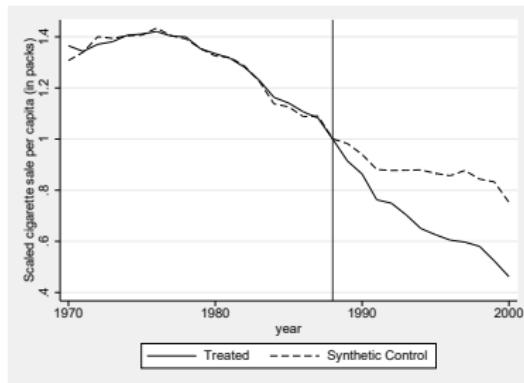
# Make Graphs

```
merge 1:1 state year using `keepfile2', nogenerate  
gen cigsale_scaled_synth = cigsale_scaled - effect_scaled  
single_treatment_graphs, depvar(cigsale_scaled) effect_var  
(effect_scaled) trunit(3) trperiod(1989)  
effect_graphs , depvar(cigsale_scaled) depvar_synth(  
cigsale_scaled_synth) effect_var(effect_scaled) trunit  
(3) trperiod(1989)  
pval_graphs
```

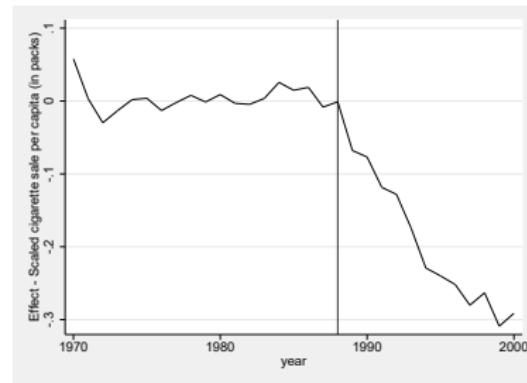
# Outcome Trends



# Mean Treated and Synthetic Control

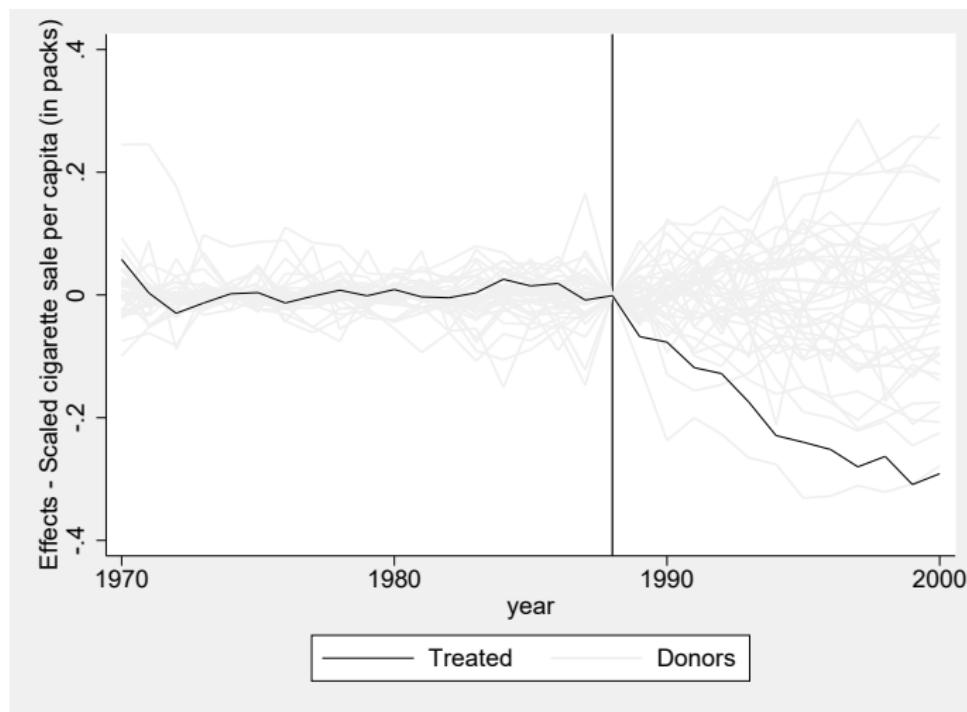


(e) Treated and Control

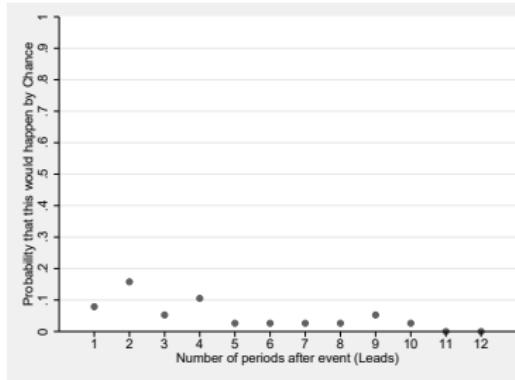


(f) Difference

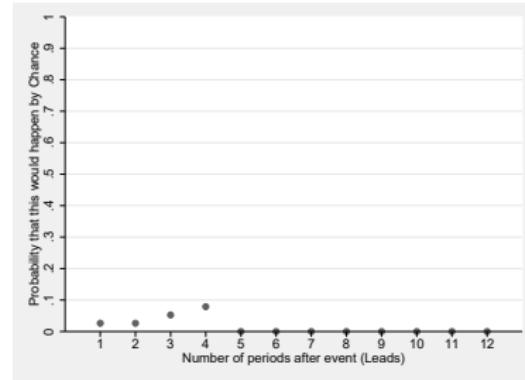
# Differences



# P-Values



(g) Effect



(h) Studentized Effect

# Multiple Treatments

## Placebo Set

- The number of SC estimates grows linearly ( $J \cdot G$ ) while the size of the placebo set grows exponentially ( $J^G$ )
- Inference can quickly become infeasible
- By default, there is a maximum number ( $max$ ) of averages computed for inference (1,000,000)
  - If  $J^G < max$  then all  $J^G$  are used for inference
  - If  $J^G > max$  then  $max$  are drawn at random (with replacement)
  - ( $max$  is modifiable and can be set to  $\infty$ )
- Space requirements scale exponentially if confidence intervals as desired, otherwise scales linearly.

# Multiple Treatments

## Placebo Set

- The number of SC estimates grows linearly ( $J \cdot G$ ) while the size of the placebo set grows exponentially ( $J^G$ )
- Inference can quickly become infeasible
- By default, there is a maximum number ( $max$ ) of averages computed for inference (1,000,000)
  - If  $J^G < max$  then all  $J^G$  are used for inference
  - If  $J^G > max$  then  $max$  are drawn at random (with replacement)
  - ( $max$  is modifiable and can be set to  $\infty$ )
- Space requirements scale exponentially if confidence intervals as desired, otherwise scales linearly.

# Multiple Treatments

## Placebo Set

- The number of SC estimates grows linearly ( $J \cdot G$ ) while the size of the placebo set grows exponentially ( $J^G$ )
- Inference can quickly become infeasible
- By default, there is a maximum number ( $max$ ) of averages computed for inference (1,000,000)
  - If  $J^G < max$  then all  $J^G$  are used for inference
  - If  $J^G > max$  then  $max$  are drawn at random (with replacement)
  - ( $max$  is modifiable and can be set to  $\infty$ )
- Space requirements scale exponentially if confidence intervals as desired, otherwise scales linearly.

# Multiple Treatments

## Placebo Set

- The number of SC estimates grows linearly ( $J \cdot G$ ) while the size of the placebo set grows exponentially ( $J^G$ )
- Inference can quickly become infeasible
- By default, there is a maximum number (*max*) of averages computed for inference (1,000,000)
  - If  $J^G < max$  then all  $J^G$  are used for inference
  - If  $J^G > max$  then *max* are drawn at random (with replacement)
    - (*max* is modifiable and can be set to  $\infty$ )
- Space requirements scale exponentially if confidence intervals as desired, otherwise scales linearly.

# Multiple Treatments

## Placebo Set

- The number of SC estimates grows linearly ( $J \cdot G$ ) while the size of the placebo set grows exponentially ( $J^G$ )
- Inference can quickly become infeasible
- By default, there is a maximum number ( $max$ ) of averages computed for inference (1,000,000)
  - If  $J^G < max$  then all  $J^G$  are used for inference
  - If  $J^G > max$  then  $max$  are drawn at random (with replacement)
  - ( $max$  is modifiable and can be set to  $\infty$ )
- Space requirements scale exponentially if confidence intervals as desired, otherwise scales linearly.

# Multiple Treatments

## Placebo Set

- The number of SC estimates grows linearly ( $J \cdot G$ ) while the size of the placebo set grows exponentially ( $J^G$ )
- Inference can quickly become infeasible
- By default, there is a maximum number ( $max$ ) of averages computed for inference (1,000,000)
  - If  $J^G < max$  then all  $J^G$  are used for inference
  - If  $J^G > max$  then  $max$  are drawn at random (with replacement)
  - ( $max$  is modifiable and can be set to  $\infty$ )
- Space requirements scale exponentially if confidence intervals as desired, otherwise scales linearly.

# Estimate with Multiple Treatments

```
gen byte D = (state==3 & year>=1989) | (state==7 & year  
    >=1988)  
synth_runner cigsale beer(1984(1)1987) lnincome(1972(1)  
    1987) retprice age15to24, d(D) trends training_propr  
    (`=13/18')  
*Graphs  
effect_graphs , multi depvar(cigsale)  
pval_graphs
```

# Returned Values

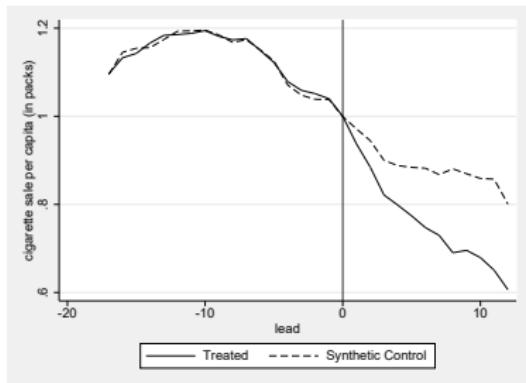
Post-treatment joint significance

```
e(pval_joint_post) = .0423666910153397  
e(pval_joint_post_t) = 0
```

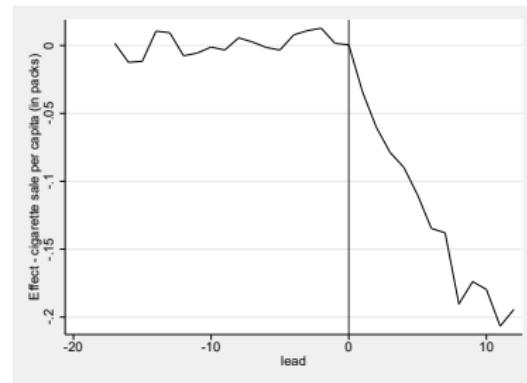
Diagnostics (want large values)

```
e(avg_pre_rmspe_p) = .9298758217677137  
e(avg_val_rmspe_p) = .9598246895544192
```

# Mean Treated and Synthetic Control

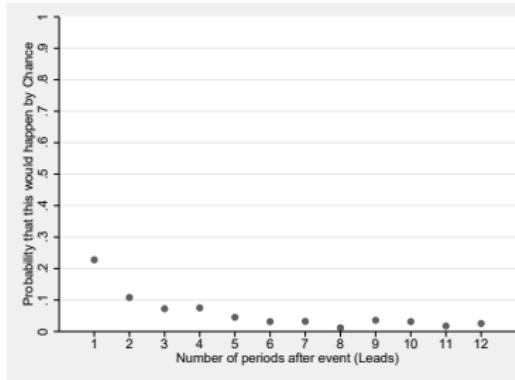


(i) Mean Treated and Control

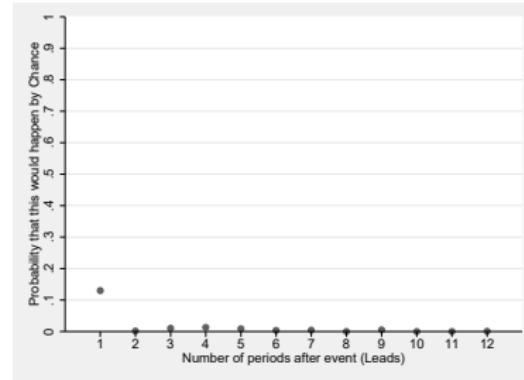


(j) Mean Difference

# P-Values



(k) Effect



(l) Studentized Effect

# Installation

- Stable Version:

<http://econweb.umd.edu/~galiani/code.html>

- Development:

[https://github.com/bquistorff/synth\\_runner](https://github.com/bquistorff/synth_runner)

- Can submit bug reports and patches.

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

# Future Work for synth\_runner

- Cross-validation for picking regressors (Dube and Zipperer, 2015).
- Graph the ratio of post/pre RMSPE (as in Abadie et al. 2014)
- Leave-one-observation-out robustness:
  - Small number: Graph as in Abadie et al. 2014. Would also want to show changing significance levels.
  - Large number: Standard deviation of the change of the estimate  $\{(\hat{\alpha} - \hat{\alpha}_{-j})\}_j$  (Athey and Imbens, 2015). What proportion lose/gain significance?
- Allow per-treatment donor sets e.g. to limit interpolation bias (non-linear DGP) or spill-overs.
- CV for picking  $V$  (mentioned in Abadie et al. 2010 and used in Abadie et al. 2014) using training period

## Q &amp; A

# Appendix Slides

- Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1):113–132, March 2003.  
doi:10.1257/000282803321455188.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505, June 2010.  
doi:10.1198/jasa.2009.ap08746.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, April 2014.  
doi:10.1111/ajps.12116.
- Susan Athey and Guido Imbens. A measure of robustness to misspecification. *American Economic Review*, 105(5):476–480, May 2015. doi:10.1257/aer.p20151020.

- Eduardo Cavallo, Sebastian Galiani, Ilan Noy, and Juan Pantano.  
Catastrophic natural disasters and economic growth. *Review of  
Economics and Statistics*, 95(5):1549–1561, December 2013.  
doi:10.1162/rest\_a\_00413.
- Arindrajit Dube and Ben Zipperer. Pooling multiple case studies  
using synthetic controls: An application to minimum wage policies.  
Technical Report 8944, IZA, March 2015.