Extension of the paper “Carbon Taxes and CO_2 Emissions: Sweden as a Case Study” (Julius J. Andersson, 2019)

Federico Bertoia

May 25, 2023

Outline

1. Overview of the paper
   - Introduction
   - DiD and synthetic control
   - Potential improvement
2. Theoretical Framework
   - DiD revisited
   - SC revisited
   - Synthetic DiD
3. Results
   - No covariates
   - Optimized and projected covariates
   - Inference with SDiD
4. Appendix
   - Regularization parameter
   - Placebo estimates

Overview of the paper

- **Research question**: What is the causal effect of carbon taxes on CO_2 emissions?
- **Case study**: Sweden as treated unit, since it was one of the first countries in the world to implement a carbon tax in 1991.
- **Data**: Annual panel data on per capita CO_2 emissions from transport for the years 1960-2005 for 25 OECD countries, including Sweden. Key predictors: GDP per capita, number of motor veichles, gasoline consumption per capita and percentage of urban population.

Overview of the paper (2)

- **Block treatment assignment**: Sweden is the only treated country and there is a single adoption period. Formally, we can define the assignment matrix as:

$$W = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

where each row represents a time period and each column a country. The assignment matrix leads to the following outcome matrix:

$$Y = \begin{bmatrix} \boldsymbol{Y}_{pre,co} & \boldsymbol{Y}_{pre,tr} \\ \boldsymbol{Y}_{post,co} & \boldsymbol{Y}_{post,tr} \end{bmatrix}$$

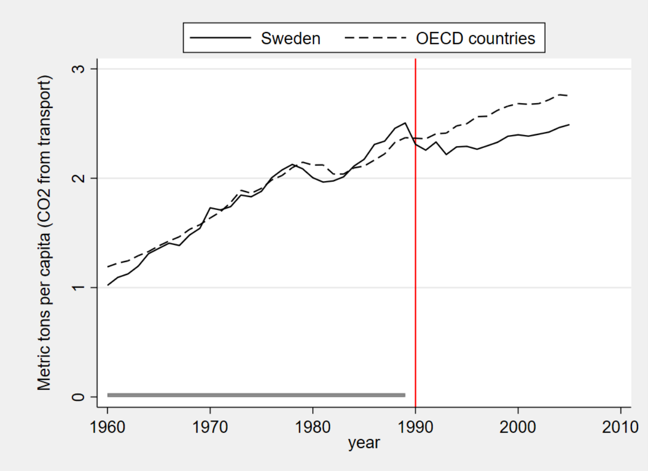
Failure of DiD

The main challenge is to estimate the path of CO_2 emissions in Sweden if it had not been treated. In comparative case studies, a widely used strategy is Difference in Differences, which selects a non treated unit to be taken as a comparison to estimate the average treatment effect. The main assumption of DiD is **parallel trends**, which assumes that:

$$E[Y(0)_{post,co} - Y(0)_{pre,co}] = E[Y(0)_{post,tr} - Y(0)_{pre,tr}]$$

In our setting, we would like to have a country that follows the same trend of Sweden in the pre treatment period. If so, then any difference from that path in the post treatment period can be imputed to the carbon tax.

Failure of DiD (2)



From this figure we can clearly see a failure of the parallel trend assumption!

Synthetic Control

- **J+1**: Number of units. We assume that the first unit (j=1) is the treated unit. The "donor pool", j=2,...,J+1, is a collection of untreated units not affected by the intervention.
- **T**: Time periods. First T_0 periods are before the intervention, while from T_0+1 to T are post treatment periods in which we estimate the causal effect.
- For each unit and time we observe the outcome of interest, Y_{jt}. Moreover, for each unit we also observe a set of k predictors of the outcome, $X_{1j},\ldots,X_{kj}$.
- Y_{jt}^N is the potential outcome in the absence of the intervention.
- Y_{jt}^I is the potential outcome under the intervention.

Synthetic Control (2)

The effect of the intervention for the affected unit in period t (with $t > T_0$) is:

$$\tau_{1t} = Y_{1t}^{I} - Y_{1t}^{N}$$

Since unit j=1 is the treated, Y_{1t}^{N} is counterfactual. The great policy evaluation challenge is to estimate Y_{1t}^{N} for $t > T_0$: how the outcome of interest would have evolved for the affected unit in the absence of the intervention.

The idea is that a combination of units in the donor pool may approximate the characteristics of the treated unit better than any unaffected unit alone. Thus, a synthetic control is defined as a weighted average of the units in the donor pool.

Synthetic Control estimation

Formally, a synthetic control can be represented by a Jx1 vector of weights:

$$\boldsymbol{W} = (w_2, ..., w_{J+1})$$

Given a set of weights $\mathbf{W}$, the synthetic control estimators of Y_{1t}^N and τ_{1t} are:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}$$

and

$$\hat{\tau}_{1t} = Y_{jt}^I - \hat{Y}_{jt}^N$$

To avoid extrapolation, the weights are restricted to be nonnegative and sum to one.
How do the weights should be chosen in practice?

Synthetic Control estimation (2)

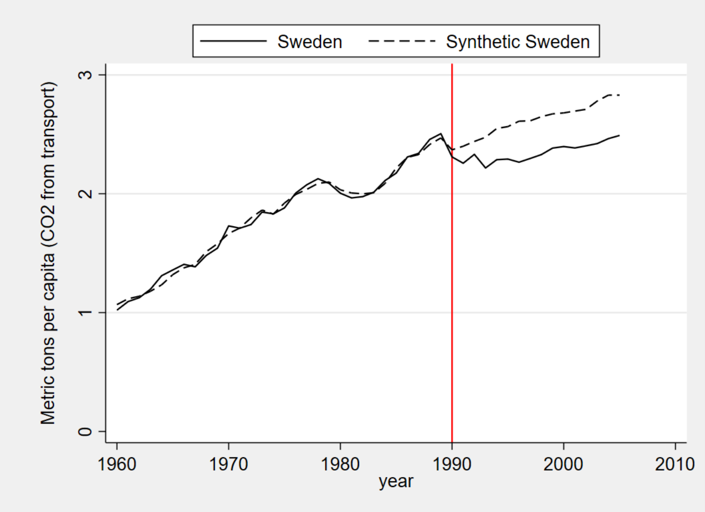
Abadie, Diamond, and Hainmueller (2010) propose to choose $w_2 \ldots w_{J+1}$ so that the resulting synthetic control best resembles the pre-intervention values for the treated unit of predictors of the outcome variables.

$$||\boldsymbol{X}_1 - \boldsymbol{X}_0 \boldsymbol{W}|| = \left(\sum_{h=1}^{k} v_h \Big(X_{h1} - \sum_{j=2}^{J+1} w_j X_{hj} \Big)^2 \right)^{\frac{1}{2}}$$

The positive constants $v_1 \ldots v_k$ reflect the relative importance of each predictor for the treated unit. A way to choose $\mathbf{V}$ is such that the synthetic control $\mathbf{W}(\mathbf{V})$ minimizes the mean squared prediction error (MSPE):

$$\sum_{t \in T_0} (Y_{1t} - w_2(\mathbf{V}) Y_{2t} - \ldots - w_{J+1}(\mathbf{V}) Y_{J+1t})^2$$

Synthetic Control estimation (3)



Potential improvement

Synthetic control seems to work perfectly in this setting. However, is it possible to do even better?
Arkhangelsky et al. (2021) propose the synthetic difference in differences estimator (SDiD), which combines attractive features of both.

- Like SC, it re-weights and matches pre-exposure trends to weaken the reliance on parallel trend assumption.
- Like DiD, it is invariant to additive unit-level shifts and allows for valid large-panel inference.

DiD and SC Revisited

In the canonical framework where two time periods are available, the DiD estimator can be recasted in the Two-Way Fixed-Effects formulation.

$$\hat{\tau}^{did} = \underset{\mu,\alpha,\beta,\tau}{argmin} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - (\mu + \alpha_i + \beta_t + \tau W_{it} \right)^2 \right\}$$

where α_i are the unit fixed effects and β_t are the time fixed effects.

The synthetic control estimator can be recasted as the following optimization problem:

$$\hat{\tau}^{sc} = \underset{\mu,\beta,\tau}{argmin} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \mu - \beta_t - \tau W_{it} \right)^2 \hat{w}_i^{sc} \right\}$$

where we omit the unit fixed effects but we add unit weights.

SDiD estimator

Synthetic difference in differences combines both features of DiD and SC in the following way:

$$\hat{\tau}^{sdid} = \underset{\mu,\alpha,\beta,\tau}{argmin} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - (\mu + \alpha_i + \beta_t + \tau W_{it} \right)^2 \hat{w}_i^{sdid} \hat{\lambda}_t^{sdid} \right\}$$

- **Unit weights**: Designed so that the outcome for the treated unit is approximately parallel to the weighted average for control units.
- **Time weights**: Designed so that the average post-treatment outcome for each of the control units differs by a constant from the weighted average of the pre-treatment outcomes for the same control units.

Unit weights in SDiD

Unit weights $\hat{w}_i^{sdid}$ are calculated solving the optimization problem:

$$(\hat{w}_0, \hat{w}^{sdid}) = \underset{w_0 \in \mathbb{R}, w \in \Omega}{argmin} \ \ell_{unit}(w_0, w)$$

where

$$\ell_{unit}(w_0, w) = \sum_{t=1}^{T_{pre}} \left(w_0 + \sum_{i=1}^{N_{co}} w_i Y_{it} - \frac{1}{N_{tr}} \sum_{i=N_{co}+1}^{N} Y_{it} \right)^2 + \zeta^2 T_{pre} \|w\|_2^2$$

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

Unit weights in SDiD (2)

The SDiD weights are closely related to those used in Abadie, Diamond, and Hainmueller (2010), with two differences:

- SDiD allows for an intercept w_0, thus the unexposed pre-trends do not need to perfectly match the other ones, it is sufficient that the weights make the trend parallel.
- SDiD adds a regularization penalty to increase the dispersion. In particular, it adds a ridge regression penalty on the weights[1].

[1] Details in the regularization parameter ζ in the appendix

Time weights in SDiD

Time weights $\hat{\lambda}_t^{sdid}$ are calculated solving the optimization problem:

$$(\hat{\lambda}_0, \hat{\lambda}^{sdid}) = \underset{\lambda_0 \in \mathbb{R}, \lambda \in \Lambda}{argmin}\, \ell_{time}(\lambda_0, \lambda)$$

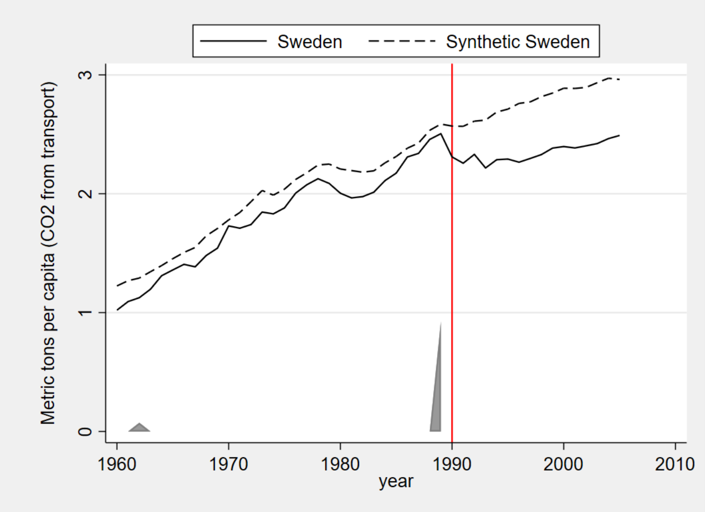
where

$$\ell_{time}(\lambda_0, \lambda) = \sum_{i=1}^{N_{co}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{t=T_{pre}+1}^{T} Y_{it} \right)^2$$

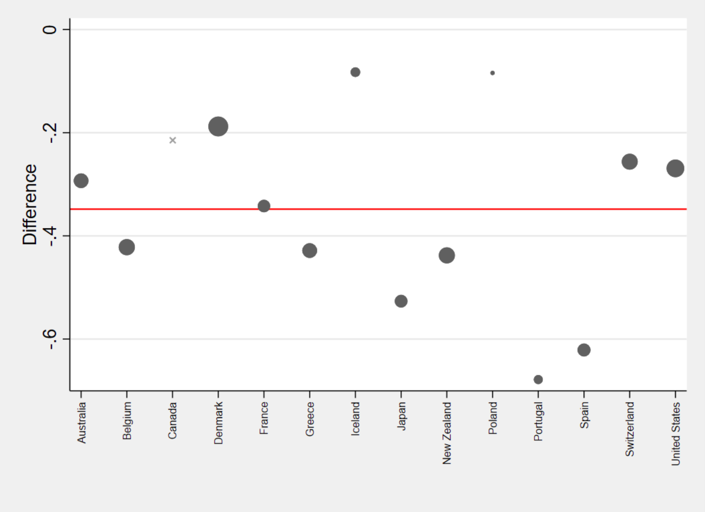
$$\Lambda = \left\{ \lambda \in \mathbb{R}_+^T : \sum_{t=1}^{T_{pre}} \lambda_t = 1, w_i = \frac{1}{T_{post}} \forall\ t = T_{pre} + 1, ..., T \right\}$$

Notice that there is no regularization here: this is because it allows for correlated observations within periods for the same units, but no across units within a time period.

SDiD estimation and time weights



SDiD unit weights



SDiD with covariates

So far I considered the case of no covariates, however it is possible to extend the synthetic difference in differences by conditioning on exogenous time-varying covariates X_{it}. Arkhangelsky et al. (2021) proposes to apply SDiD algorithm to the residuals calculated as:

$$Y_{it}^{res} = Y_{it} - X_{it}\hat{\beta}$$

where $\hat{\beta}$ comes from a regression of Y_{it} on X_{it}.

- **Abadie et al. (2010)**: When covariates are included the synthetic control is chosen to ensure that these covariates are as closely matched as possible between treated and synthetic control units.
- **Arkhangelsky et al. (2021)**: Covariate adjustment is viewed as a pre-processing task, which removes the impact of changes in covariates from the outcome Y_{it} prior to calculating the synthetic control.

Projected covariates

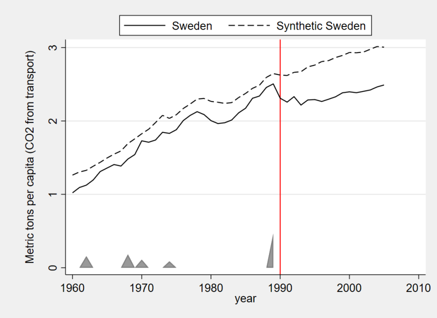
- **Kranz (2022)**: If covariates are corralated with unit and time effects, then the method proposed by Arkhangelsky et al. (2021) may not yield a consistent estimator. Kranz proposal is to first estimate a two way fixed effect regression using a subsample that omits any unit in which treatment takes place:

$$Y_{it} = \alpha_i + \gamma_t + X_{it}\beta + u_{it}$$

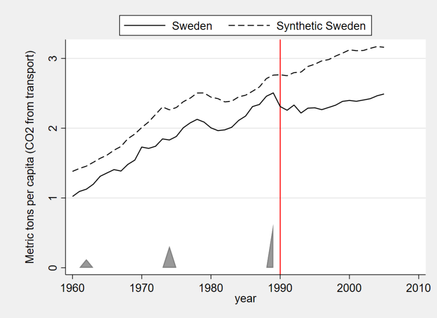
and then apply the SDiD estimator to the adjusted outcomes:

$$Y_{it}^{adj} = Y_{it} - X_{it}\hat{\beta}$$

SDiD results with covariates

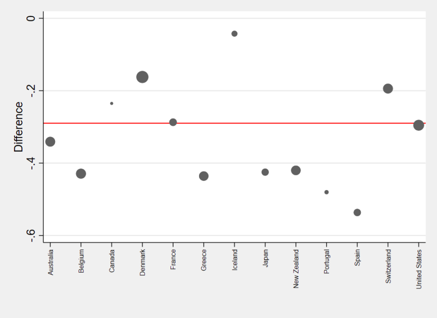


Optimized

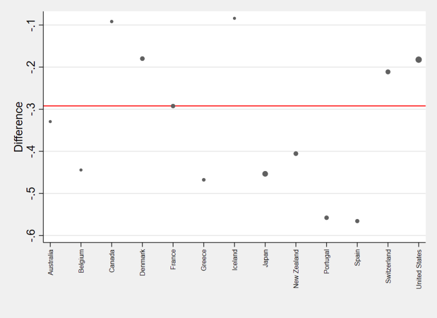


Projected

SDiD weight with covariates



Optimized



Projected

Inference following Arkhangelsky et al. (2021)

In the paper the authors show that, under some conditions, we have a consistent estimator for the asymptotic variance V_τ and we can use conventional confidence intervals:

$$\tau \in \hat{\tau}^{sdid} \pm z_{\alpha/2}\sqrt{\hat{V}_\tau}$$

In the original paper three methods are proposed to estimate V_τ:

- Bootstrap
- Jackknife
- Placebo

However, only placebo can be used if $N_{tr} = 1$.

Placebo Variance Estimation

The main idea of such placebo evaluations is to consider the behavior of SDiD estimation when we replace the unit that was exposed to the treatment with different units that were not exposed. Then, assuming B to be the number of placebos, we get:

$$\hat{V}_{\tau}^{placebo} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}^{(b)} - \frac{1}{B}\sum_{b=1}^{B}\hat{\tau}^{(b)}\right)^{2}$$

Notice that the validity relies fundamentally on homoskedasticity across units.

Placebo Variance Estimation: a personal consideration

When making inference, in Abadie et al. (2010) they discard those countries with a pre treatment RMSPE x times higher than California, the treated state. That's because if the synthetic control method is not able to find a convex combination that manages to reconstruct the treated unit, then those results may not be credible. What I would like to highlight here is that, in calculating the standard errors for each of the methods (DiD, SC and SDiD), we **shouldn't include** the results of the units for which the method clearly does not work.

Results Table

Results Table

Method	ATT	SE	SE Revisited
DiD	-0.21372	0.27489	0.14250
SC	-0.27122	0.40887	0.15091
SDiD	-0.34822	0.30675	0.14936
opt SDiD	-0.28977	0.20226	0.13316
proj SDiD	-0.29216	0.24204	0.15072

Notes: Excluded countries for SE Revisited

- DiD (3 7 9 10)
- SC (3 7 10 11 15)
- SDiD (3 7 10 14)
- opt SDiD (3 7 14)
- proj SDiD (7 14)

Appendix. Regularization parameter

$$\zeta = (N_{tr} T_{post})^{1/4} \hat{\sigma}$$

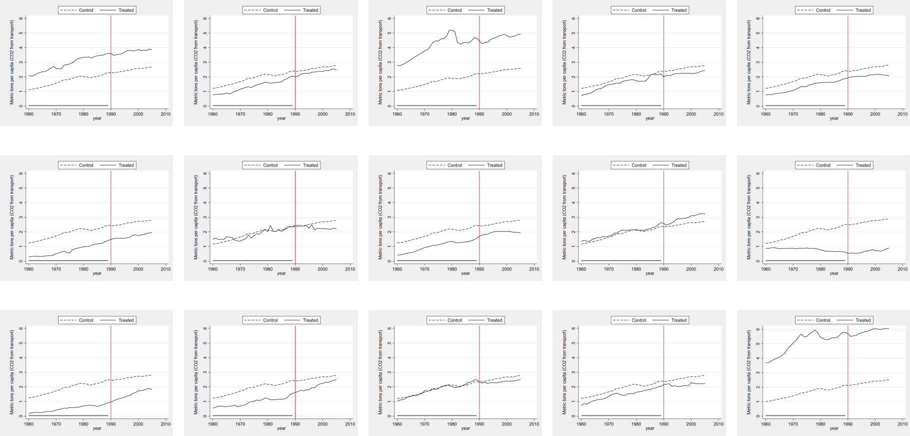
with

$$\hat{\sigma}^2 = \frac{1}{N_{co}(T_{pre} - 1})\sum_{i=1}^{Nco} \sum_{t=1}^{T_{pre}-1} \left(\Delta_{it} - \bar{\Delta}\right)^2$$

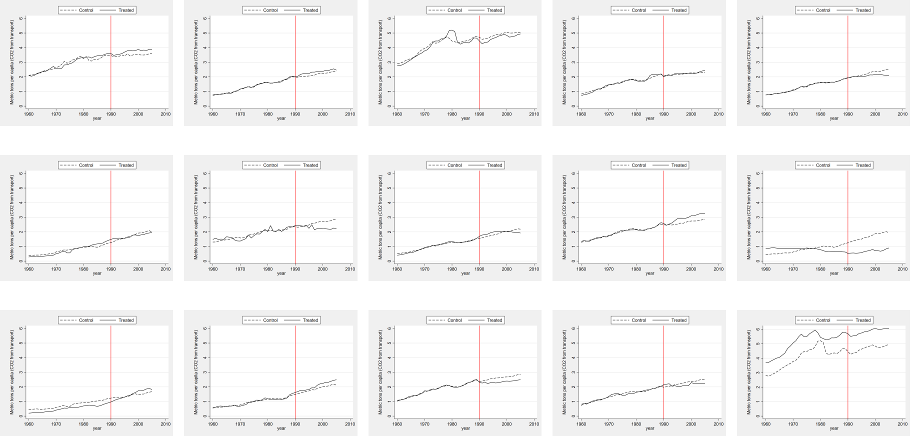
$$\Delta_{it} = Y_{i(t+1)} - Y_{it}$$

$$\bar{\Delta} = \frac{1}{N_{co}(T_{pre} - 1)} \sum_{i=1}^{Nco} \sum_{t=1}^{T_{pre}-1} \Delta_{it}$$

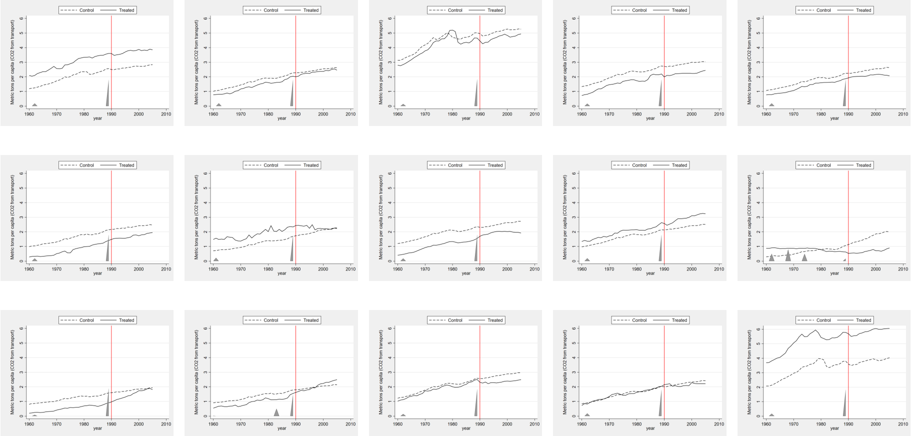
DiD Placebo


Control
Treated
Metric tons per capita (CO2 from transport)
year
0
1
2
3
4
5
6
1960
1970
1980
1990
2000
2010

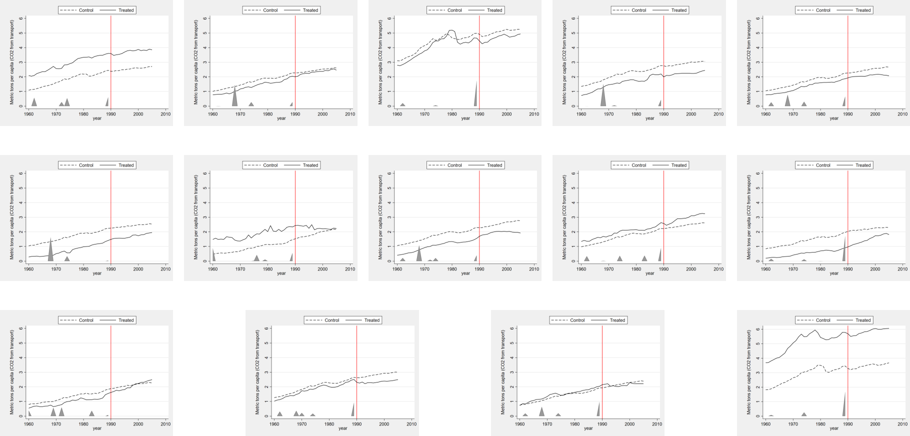
SC Placebo


Control
Treated
Metric tons per capita (CO2 from transport)
year
1960 1970 1980 1990 2000 2010
0 1 2 3 4 5 6

SDiD Placebo



opt SDiD Placebo



proj SDiD Placebo