Inference in Synthetic Controls with Spillover Effects

Jianfei Cao & Connor Dowd

University of Chicago Booth School of Business

Midwest Econometrics Group
October 26, 2018

Introduction

- Synthetic Controls (SCM) take the basic idea of diff-in-diff, and try to replace the weights of $1/n$ with something data driven.
- Procedure is usually justified as being straightforward in factor model settings.
- The data-driven weighting makes SCM particularly vulnerable to spillover type problems – because it may place a lot of weight on a contaminated observation.

Goals

Our paper:

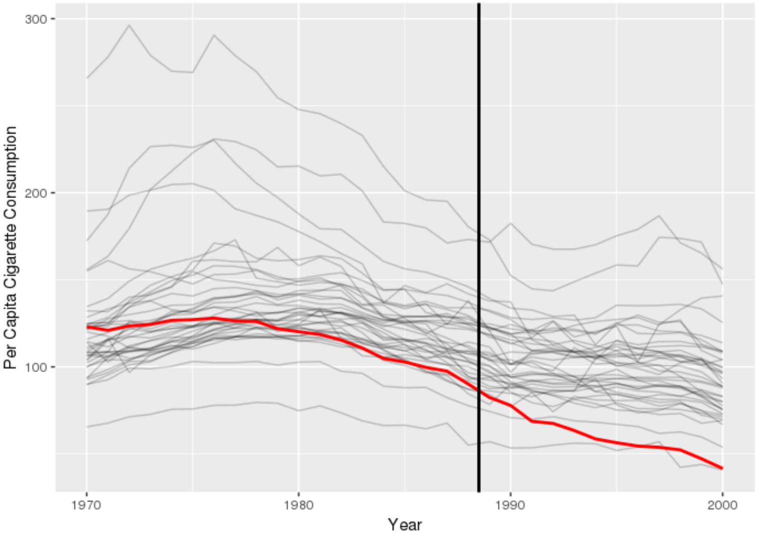
- Estimates treatment effects in the presence of spillovers
- Performs inference on both treatment effects and spillovers
- Give simple conditions for identification and inference

All of which will be in a factor model setting.

Outline

1. Synthetic Controls Framework
2. Motivation for Adding Spillovers
3. Our Model & Procedure
4. Main Results
5. Simulations/Horserace
6. Example

Example: California Prop 9



Synthetic Controls Setting

$$
\begin{array}{cccc|c}
y_{1,1} & y_{1,2} & \cdots & y_{1,T} & y_{1,T+1} \\
\hline
y_{2,1} & y_{2,2} & \cdots & y_{2,T} & y_{2,T+1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
y_{N,1} & y_{N,2} & \cdots & y_{N,T} & y_{N,T+1}
\end{array}
$$

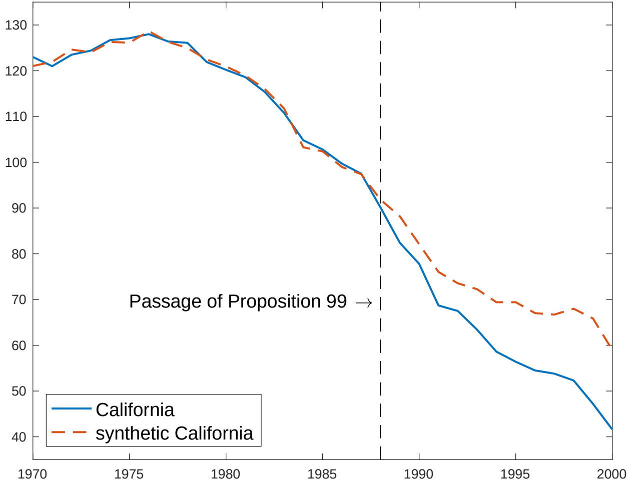
$\}$ treated unit (row 1)

$\}$ control units (rows 2 to N)

$\uparrow$ treatment (between periods T and $T+1$)

$$y_{1,T+1}(1) = y_{1,T+1}(0) + \alpha$$

Abadie, Diamond, and Hainmueller (2010)



Synthetic Control Estimator

Synthetic control weights:

$$\begin{pmatrix} \hat{z}_1 \\ \hat{b}_1 \end{pmatrix} = \underset{(z,b')' \in W}{\arg\min} \sum_{t=1}^{T} (y_{1,t} - z - Y_t' b)^2,$$

where $Y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$ and $W = \mathbb{R} \times \{0\} \times \Delta_{N-1}$

i.e. $\hat{b}_1$ are normalized weights – they sum to one, are non-negative, and we force a weight of 0 on the 'own' observation.

z is included because of the work in Ferman and Pinto (2017) showing an intercept is necessary for unbiased procedures.

Spillover effects in synthetic control

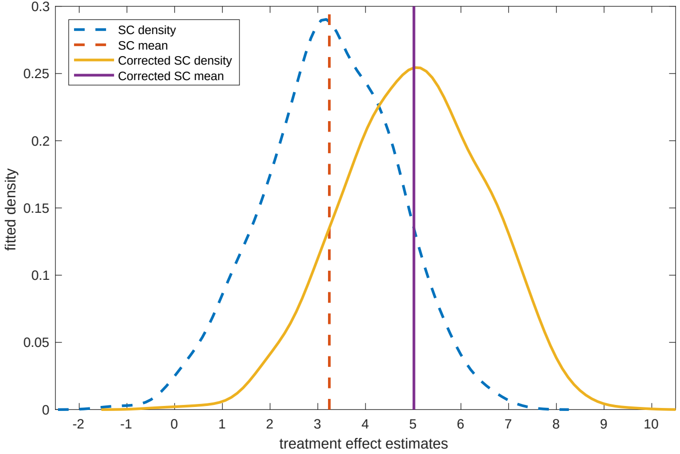
The synthetic control estimator

$$\hat{\alpha} = y_{1,T+1}(1) - \hat{y}_{1,T+1}(0) = y_{1,T+1} - Y'_{T+1}\hat{b}_1 - \hat{z}$$

can be severely biased in the presence of spillover effects:

- As in a diff-in-diff setting, if treatment can affect your control group - your ATE may be biased.
- SCM is particularly vulnerable as it may put extra weight on the same units that are 'contaminated'.
- This can be bad luck – but more worryingly, similar units may actually be prone to spillovers.
- Regularization properties of simplex are suddenly a potential downside.

True treatment effect = 5



Our paper

Idea: calculating synthetic control weights for all units. With a specified spillover pattern, we can build an unbiased estimator.

Results of this paper (large T, fixed N)

- asymptotically unbiased estimators for treatment and spillover effects
- a testing procedure that has asymptotically correct sizes

Model

$$
\begin{array}{cccc|c}
y_{1,1} & y_{1,2} & \cdots & y_{1,T} & y_{1,T+1} \\
\hline
y_{2,1} & y_{2,2} & \cdots & y_{2,T} & y_{2,T+1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
y_{N,1} & y_{N,2} & \cdots & y_{N,T} & y_{N,T+1}
\end{array}
\quad
\begin{array}{l}
\} \text{ treated unit} \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \text{ control units}
\end{array}
$$

$\uparrow$ treatment

DGP:

$$
y_{i,t} = \begin{cases} y_{i,t}(1), & \text{if treated} \\ y_{i,t}(0), & \text{otherwise} \end{cases}
$$

and

$$
\begin{cases} y_{i,t}(0) = \delta_t + \lambda_t' \mu_i + \epsilon_{i,t} \\ y_{i,T+1}(1) = y_{i,T+1}(0) + \alpha_i \end{cases}
$$

Linear Spillover effects

Remember: $\alpha_i = y_{i,T+1}(1) - y_{i,T+1}(0)$
i.e. In the counterfactual, had treatment not occurred anywhere, how different was unit i?

Let

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}$$

Assume linear spillover effects: $\alpha = A\gamma$

- A known
- γ unknown parameters

Examples of A

- Spillovers Exponentially decreasing in distance: $\begin{bmatrix} 1 & 0 \\ 0 & \exp(-d_2) \\ \vdots & \vdots \\ 0 & \exp(-d_N) \end{bmatrix}$
- All but one unit equally hit by spillover: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$
- Range of spillovers is known: $\begin{bmatrix} 1 & 0_{1\times p} \\ 0_{p\times 1} & I_p \\ 0_{(N-p-1)\times 1} & 0_{(N-p-1)\times p} \end{bmatrix}$

Synthetic control weights for all units

We estimate the weight vectors and intercepts for each unit.
For each i,

$$\begin{pmatrix} \hat{z}_i \\ \hat{b}_i \end{pmatrix} = \underset{(z,b)\in W_i}{\arg\min} \sum_{t=1}^{T} (y_{i,t} - z - Y_t b')^2,$$

with i-th entry of $\hat{b}_i$ being 0, and $\hat{b}_i \in \Delta_N$ (non-negative, sum to one).
$Y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$

Define $z_i = \text{plim } \hat{z}_i$ and $b_i = \text{plim } \hat{b}_i$. Let

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}, \; B = \begin{bmatrix} b'_1 \\ \vdots \\ b'_N \end{bmatrix}$$

Estimation

Estimator for γ:

$$\hat{\gamma} = \arg\min_{g \in \mathbb{R}^k} \|(I - \hat{B})(Y_{T+1} - Ag) - \hat{Z}\|_2$$
$$= (A'\hat{M}A)^{-1}A'(I - \hat{B})'((I - \hat{B})Y_{T+1} - \hat{Z}),$$

with $\hat{M} = (I - \hat{B})'(I - \hat{B})$.

Estimator for treatment and spillover effects:

$$\hat{\alpha} = A\hat{\gamma}$$

Model with stationary common factors

Recall the model:

$$y_{i,t}(0) = \delta_t + \lambda_t' + \epsilon_{i,t}$$

Stacking i:

$$Y_t(0) = \delta_t \iota_{N\times 1} + [\mu_1, \mu_2, \ldots, \mu_N]'\lambda_t + \epsilon_t$$

Condition ST

(i) $\{(\delta_t, \lambda_t, \epsilon_t)\}_{t\geq 1}$ *is stationary, ergodic, and has finite* $(2+\delta)$*-th moment*
(ii) $cov[Y_t(0)]$ *is finite, positive definite.*

Identification Condition

$A'MA$ *is non-singular.*

$M = (I-B)'(I-B)$ is a population parameter. A is chosen by researcher.

Identification

This is crucial.

Identification Condition

$A'MA$ is non-singular.

As we know $M = (I - B)'(I - B)$.

That is to say, B identifies α. In order for that to happen, we need non-singularity.

B is a matrix of weights with each row summing to 1.
Thus $I - B$ sums to 0 across each row. This makes $I - B$ and subsequently M singular.

Thus non-singularity requires some degree of dimension reduction occurring in A.

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \checkmark$$

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{X}$$

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$ X

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ X

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$ X

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ X

$\rightarrow$

$$B= \begin{bmatrix} 0 & w & w & 0 \\ w & 0 & w & 0 \\ w & w & 0 & 0 \\ w & w & w & 0 \end{bmatrix}$$

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \text{ X}$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ X}$$

$\longrightarrow$

$$B = \begin{bmatrix} 0 & w & w & 0 \\ w & 0 & w & 0 \\ w & w & 0 & 0 \\ w & w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \checkmark$$

Identification Examples

$$B = \begin{bmatrix} 0 & w & w \\ w & 0 & w \\ w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{X}$$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{X}$$

$\rightarrow$

$$B = \begin{bmatrix} 0 & w & w & 0 \\ w & 0 & w & 0 \\ w & w & 0 & 0 \\ w & w & w & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \checkmark$$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{X}$$

Main results

Theorem 1

Under Condition ST and the Identification Condition
$\hat{\alpha} - (\alpha + Gu_{T+1}) \to 0$ *as* $T \to \infty$
where $G = A(A'MA)^{-1}A'(I - B)'$.
Moreover, $E[Gu_{T+1}] = 0$.

Unbiasedness of estimator – but not consistency.

Andrews' Test (P-Test)

We will use a slightly modified version of Andrews' end of sample instability test. (Andrews 2003, Andrews & Kim 2006)

In essence, by using each pre-treatment year as a hold-out sample we can create a null distribution of 'effect estimates' to compare to.

Andrews Test – Demonstration

Year	CA	AL	AR	CO	...	
1970	123	89.8	100.3	124.8	...	
1971	121	95.4	104.1	125.5	...	
1972	123.5	101.1	103.9	134.3	...	
1973	124.4	102.9	108	137.9	...	← Estimate weights
⋮	⋮	⋮	⋮	⋮	⋮	with this data
1987	97.5	114	122.3	102.4	...	
1988	90.1	112.1	121.5	94.6	...	
1989	82.4	105.6	118.3	88.8	...	← Estimate effects here

Andrews Test – Demonstration

Year	CA	AL	AR	CO	…
1970	123	89.8	100.3	124.8	…
1971	121	95.4	104.1	125.5	…
1972	123.5	101.1	103.9	134.3	…
1973	124.4	102.9	108	137.9	…
⋮	⋮	⋮	⋮	⋮	⋮
1987	97.5	114	122.3	102.4	…
1988	90.1	112.1	121.5	94.6	…
1989	82.4	105.6	118.3	88.8	…

Andrews Test – Demonstration

Year	CA	AL	AR	CO	...	
1970	123	89.8	100.3	124.8	...	$\leftarrow$ Holdout $\Longrightarrow \alpha_{-1970}$
1971	121	95.4	104.1	125.5	...	
1972	123.5	101.1	103.9	134.3	...	
1973	124.4	102.9	108	137.9	...	
⋮	⋮	⋮	⋮	⋮	⋮	
1987	97.5	114	122.3	102.4	...	
1988	90.1	112.1	121.5	94.6	...	
1989	82.4	105.6	118.3	88.8	...	

Andrews Test – Demonstration

Year	CA	AL	AR	CO	...	
1970	123	89.8	100.3	124.8	...	
1971	121	95.4	104.1	125.5	...	$\leftarrow$ Holdout $\Longrightarrow \alpha_{-1971}$
1972	123.5	101.1	103.9	134.3	...	
1973	124.4	102.9	108	137.9	...	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
1987	97.5	114	122.3	102.4	...	
1988	90.1	112.1	121.5	94.6	...	
1989	82.4	105.6	118.3	88.8	...	

Andrews Test – Demonstration

Year	CA	AL	AR	CO	...	
1970	123	89.8	100.3	124.8	...	
1971	121	95.4	104.1	125.5	...	
1972	123.5	101.1	103.9	134.3	...	$\leftarrow$ Holdout $\Longrightarrow \alpha_{-1972}$
1973	124.4	102.9	108	137.9	...	
⋮	⋮	⋮	⋮	⋮	⋮	
1987	97.5	114	122.3	102.4	...	
1988	90.1	112.1	121.5	94.6	...	
1989	82.4	105.6	118.3	88.8	...	

Andrews Test – Demonstration

Year	CA	AL	AR	CO	...	
1970	123	89.8	100.3	124.8	...	
1971	121	95.4	104.1	125.5	...	
1972	123.5	101.1	103.9	134.3	...	
1973	124.4	102.9	108	137.9	...	
⋮	⋮	⋮	⋮	⋮	⋮	
1987	97.5	114	122.3	102.4	...	
1988	90.1	112.1	121.5	94.6	...	$\leftarrow$ Holdout $\Longrightarrow \alpha_{-1988}$
1989	82.4	105.6	118.3	88.8	...	

Andrews' Test (P-Test)

Now we have our null distribution:

$$\begin{bmatrix} \hat{\alpha}_{-1970} \\ \hat{\alpha}_{-1971} \\ \vdots \\ \hat{\alpha}_{-1988} \end{bmatrix}$$

As well as our estimate $\hat{\alpha}$. We can compare $\hat{\alpha}$ to the null distribution to test various hypotheses.

Testing Spillovers

Consider $H_0 : C\alpha = d$.
For some matrix W_T which is p.s.d. and symmetric:
Let

$$\hat{P}_t = \hat{u}_t' \hat{G}' C' W_T C \hat{G} \hat{u}_t,$$

and

$$\hat{P}_{T+1} = (C\hat{\alpha} - d)' W_T (C\hat{\alpha} - d)$$

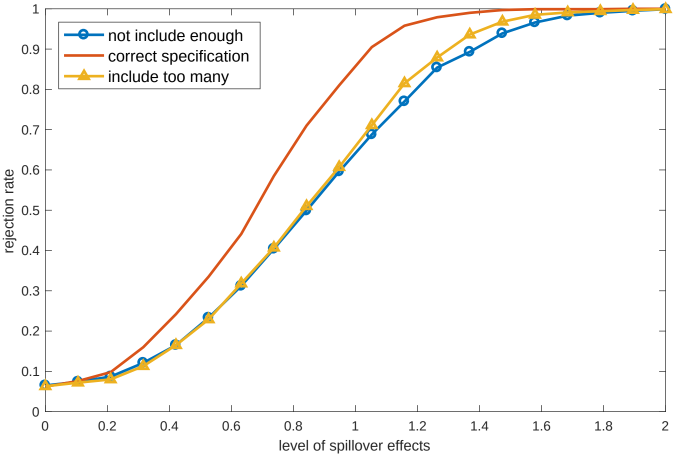
Set our critical value cv as the $(1-q)$-quantile of $\{\hat{P}_1, ..., \hat{P}_T\}$.
Reject H_0 if $\hat{P}_{T+1} > cv$.

Theorem 3

Suppose the spillover effects are correctly specified.
Under Condition ST and the Identification Condition
the test converges to the correct size under H_0.

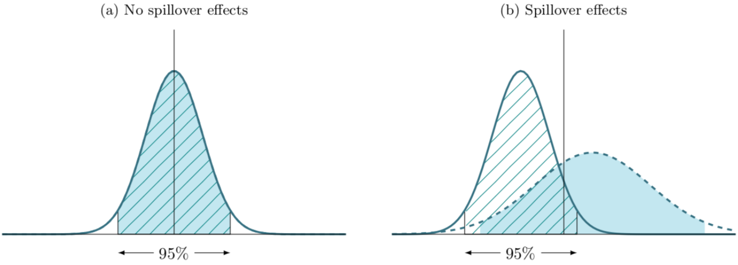
This gives us test validity for a variety of hypotheses about the spillover (and treatment) effects.

Rej. rate of spillover effects



Placebo test

Placebo test: exploits variations of $\{\hat{u}_{i,T+1}\}_{i=1}^{N}$ (across units at $t = T + 1$)



Rej. rate of treatment effect under null

	$N = 30$	
	$T = 50$	200
No spillover effects		
Placebo	0.053	0.062
CWZ	0.082	0.065
→ P-test	0.064	0.052
Concentrated spillover effects		
Placebo	0.046	0.116
CWZ	0.279	0.346
→ P-test	0.069	0.061

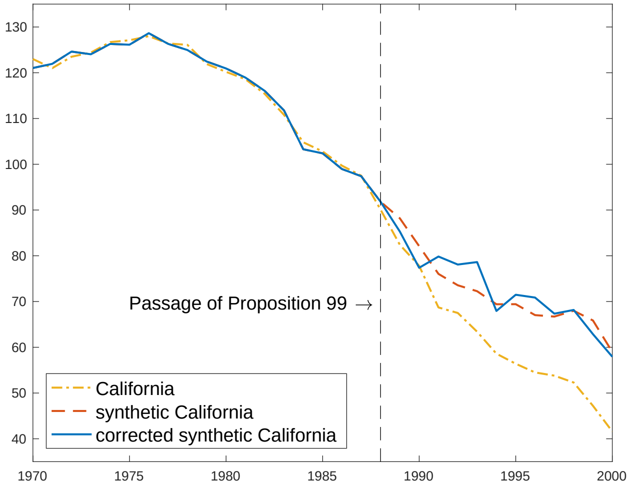
* Blue background : $\geq 5\%$ size error

Rej. rate of treatment effect under alternative

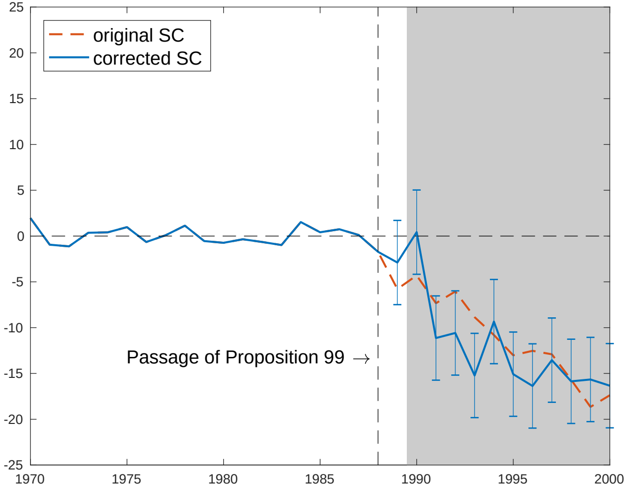
	$N = 30$	
	$T = 50$	200
No spillover effects		
Placebo	0.939	0.966
CWZ	0.901	0.983
→ P-test	0.937	0.965
Concentrated spillover effects		
Placebo	0.502	0.448
CWZ	0.754	0.542
→ P-test	0.918	0.967

* Blue background : $< 60\%$ power

Abadie, Diamond, and Hainmueller (2010)



Abadie, Diamond, and Hainmueller (2010)



Dark Grey Indicates test for presence of Spillovers rejects Null.
CIs are 90%.

Conclusion

We proposed an estimator and a testing procedure

- asymptotically unbiased
- require assumptions of spillover effects, which can be fairly weak
- deal with both stationary and co-integrated cases
- applied to Abadie, Diamond, and Hainmueller (2010)

Questions?

Thanks for listening!