`allsynth`: Synthetic Control Bias-Correction Utilities for Stata

Justin C. Wiltshire

University of California, Davis

August 5, 2021

Prepared for 2021 Stata Conference

I introduce a new community-contributed stata package: `allsynth`

`allsynth` is built on the `synth` package and adds functionality

In this presentation:

- Review of synthetic control methodology, including proposed bias-correction for inexact matching (Abadie and L'hour, 2020)
- Describe the functionality added by the `allsynth` package (with examples!):
  - Synthetic control bias-correction for inexact matching on predictor values
  - Calculation of RMPSE-ranked *p*-values from in-space treatment permutations
  - Expanded graphing functionality
  - Diagnostics for whether $\hat{\mathbf{W}}$-weighting matrix is likely unique
- Examples rely on synth_smoking data from `synth` package (Abadie et al., 2010)

Synthetic control literature

Canonical:

- Abadie & Gardeazabal, 2003; Abadie et al., 2010, 2015

Many treated units:

- Cavallo et al., 2013; Dube & Zipperer, 2015; Acemoglu et al., 2016; Abadie & L'Hour, 2020; Abadie, 2021; Ben-Michael et al., 2021b; Wiltshire, 2021a

Inference:

- Abadie et al., 2010, 2015; Doudchenko & Imbens, 2016; Hahn & Shi, 2017; Ferman & Pinto, 2017; Firpo & Possebom, 2018; Chernozhukov et al., 2019

Bias-correction for inexact matching on predictor values:

- Abadie & L'Hour, 2020; Abadie, 2021; Ben-Michael et al., 2021a; Wiltshire, 2021a

Abadie (2021) provides an excellent, current review

Potential outcomes framework

For any unit j at time t:

- Let $Y_{j,t}^{I}$ be the potential outcome under Intervention/treatment
- Let $Y_{j,t}^{N}$ be the potential outcome under Non-intervention/non-treatment
- The observed outcome is: $Y_{j,t} = Y_{j,t}^{N} + \tau_{j,t} D_{j,t}$
  - $\rightarrow$ $D_{j,t}$ is a dummy indicating if j is treated at t
- Define the treatment effect in $\{j, t\}$ as: $\tau_{j,t} = Y_{j,t}^{I} - Y_{j,t}^{N}$
- Let a single unit, $j = 1$, become treated at $T_0 + 1$
- We want to estimate path of treatment effects: $(\tau_{1,T_0+1}, \ldots, \tau_{1,T})$
- We can never observe both $Y_{1,t}^{I}$ *and* $Y_{1,t}^{N}$
- For $t > T_0$, $Y_{1,t} = Y_{1,t}^{I}$ is observable so we only need to estimate $Y_{1,t}^{N}$

Synthetic control estimator with a single treated unit

Suppose we have data on J units over T periods

- $j = 1$ is a treated unit. $j = 2, ..., J + 1$ are untreated "donor pool" units
- T_0 pre-treatment periods, $T - T_0 > 0$ treated periods
- We specify r covariates plus M linear combinations of $Y_{j,t}$ (for $t \leq T_0$)
  $\rightarrow$ $r + M = K$ total predictor variables
- $\mathbf{X}_1$ is a $K \times 1$ vector of predictors of $Y_{1,t}$ in treated unit $j = 1$
- $\mathbf{X}_0$ is a $K \times J$ matrix of predictors of $Y_{j,t}$ in donor pool units $j > 1$

Synthetic control estimator with a single treated unit

Synthetic control estimator identifies a weighted average of donor pool units:

$$\hat{Y}^N_{1,t} = \sum_{j=2}^{J+1} \hat{w}_j Y_{j,t} \quad \forall \quad t$$

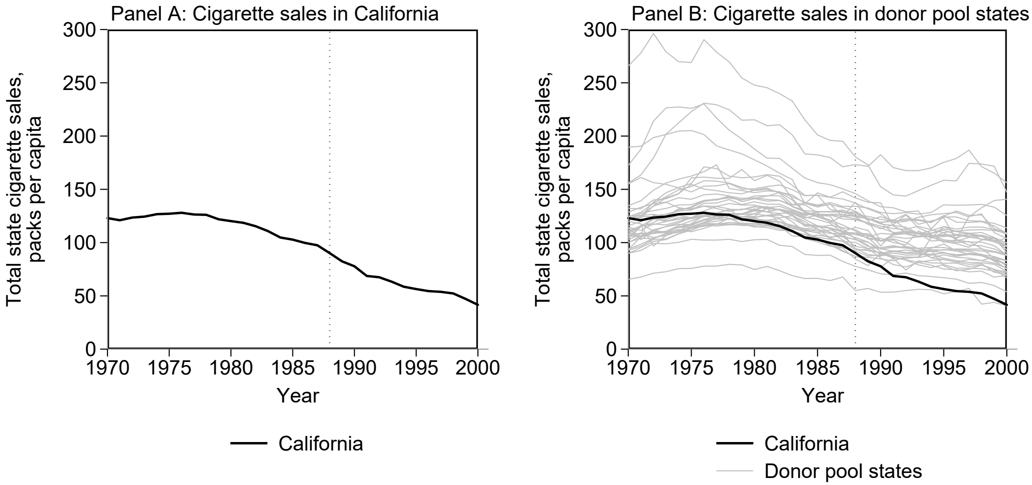
$\rightarrow$ Once we have $\hat{Y}^N_{1,t}$, we can calculate: $\hat{\tau}_{j,t} = Y_{j,t} - \hat{Y}^N_{j,t}$

- $\mathbf{V} = (v_1, ..., v_K)$ is a matrix of weights on the predictor variables
- $\mathbf{W(V)} = (w_2(\mathbf{V}), \ldots, w_{J+1}(\mathbf{V}))'$ is a vector of weights on donor pool units $j > 1$
- Classic synthetic control selects $\hat{\mathbf{V}}$ and $\hat{\mathbf{W}} = \mathbf{W}(\hat{\mathbf{V}})$ to minimize:

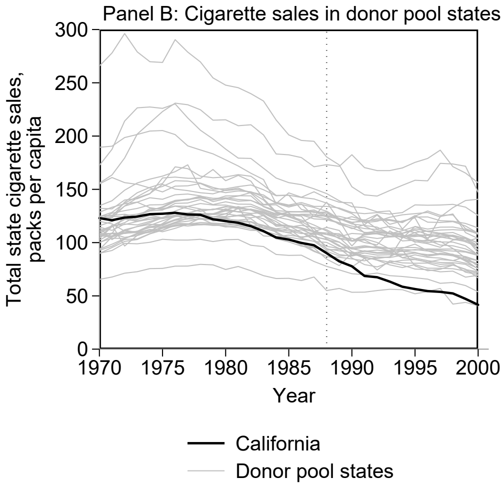
$$\left(\sum_{k=1}^{K} \hat{v}_k (X_{k,1} - w_2 X_{k,2} - ... - w_{J+1} X_{k,J+1})^2 \right)^{1/2}$$

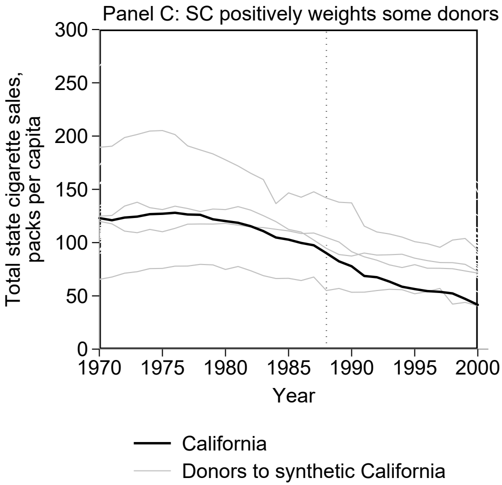
s.t. $\quad \sum_{j=2}^{J+1} w_j = 1, \; w_j \; \geq \; 0 \;\; \forall \;\; j \in \{2, ..., J+1\}$

Example: We observe cigarette sales in California and untreated states (donor pool). In 1989, California increased its cigarette excise tax

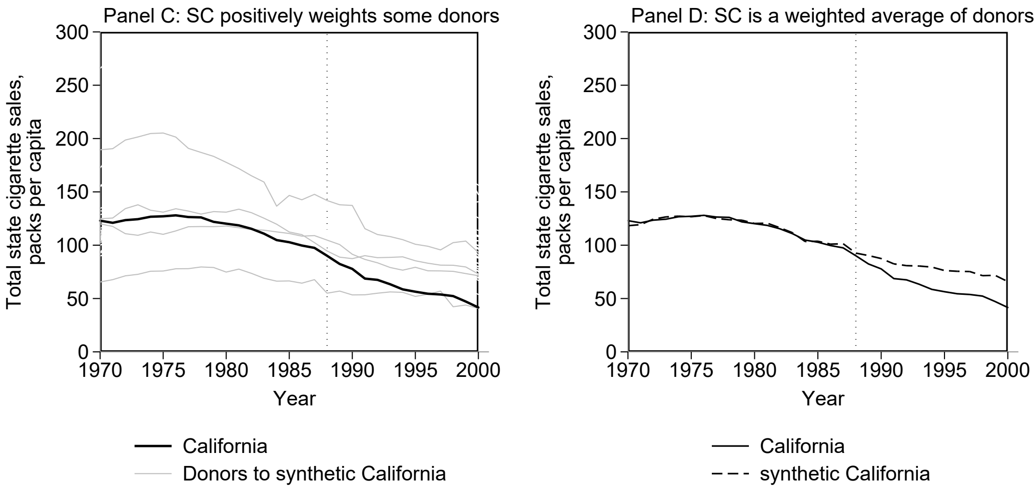


Synthetic control weights predictor variables, then positively weights some untreated states to best match pre-treatment California on those predictors

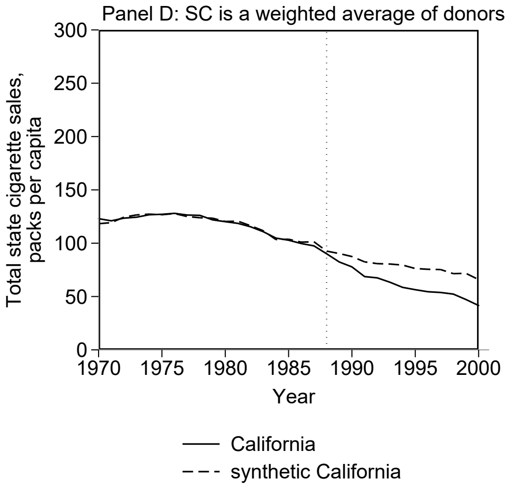


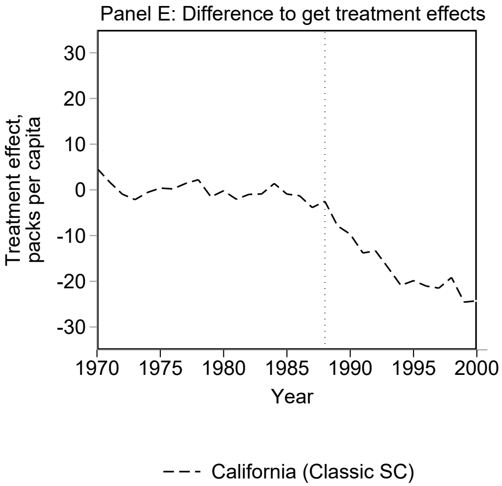


The weighted average of those donors is the synthetic California



California minus synthetic California cig sales is estimated treatment effect





The `allsynth` package: adds functionality to the `synth` package

- Synthetic control bias-correction for inexact matching on predictor values between a treated unit and its synthetic control donors (OLS regression)
- Automated calculation of RMPSE *p*-values from in-space permutation tests
- Expanded graphing functionality
- Uniqueness diagnostics (e.g. warns if the $\hat{\mathbf{W}}$ matrix is unlikely unique)

Note: In addition to directly utilizing the `synth` **package, the code for** `allsynth` **draws appreciatively on Jens Hainmueller's code for** `synth` **and slightly on the code for** `synth_runner` **(Galiani and Quistorff, 2018).**

Bias correction for inexact matching on predictor values

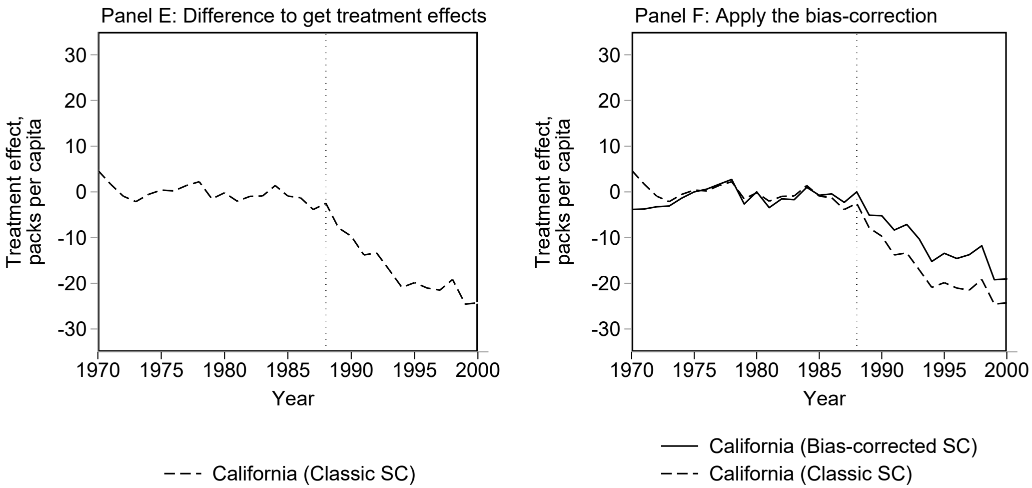
- Abadie and L'Hour (2020) (independently, Ben-Michael et al. (2021a)) propose bias correction analagous to Abadie and Imbens (2011) for matching estimators
  - → Wrote package to implement this in R

Bias correction for inexact matching on predictors (single treated unit)

- First get $\hat{w}_j$ from synthetic control estimation on uncorrected values
- Let $\mu_{0,t}(x) = E[Y|X = x, D = 0]$, and let $\hat{\mu}_{0,t}(x)$ be an estimator of $\mu_{0,t}(x)$
- Estimate $\hat{\mu}_{0,t}(x)$ by regressing $Y_{j,t}$ for untreated $j > 1$, in each $t \leq T$, on the predictor values, $\mathbf{X}_0$, for untreated $j > 1$
- The bias in $\hat{\tau}_{1,t}$ from inexact matching is $\sum_{j=2}^{J+1} \hat{w}_j \hat{\mu}_{0,t}(\mathbf{X}_j) - \hat{\mu}_{0,t}(\mathbf{X}_1)$
- Then the bias-corrected treatment effect at time t is:

$$
\begin{aligned}
\tilde{\hat{\tau}}_{1,t} &= \hat{\tau}_{1,t} - bias_t \\
&= (Y_{j,t} - \sum_{j=2}^{J+1} \hat{w}_j Y_{j,t}) - (\sum_{j=2}^{J+1} \hat{w}_j \hat{\mu}_{0,t}(\mathbf{X}_j) - \hat{\mu}_{0,t}(\mathbf{X}_1)) \\
&= (Y_{1,t} - \hat{\mu}_{0,t}(\mathbf{X}_1)) - \sum_{j=2}^{J+1} \hat{w}_j \left(Y_{j,t} - \hat{\mu}_{0,t}(\mathbf{X}_j)\right)
\end{aligned}
$$

Applying bias-correction to the California cigarette sales $\hat{\tau}_{1,t}$



`allsynth` can be used like `synth` but offers additional functionality

`allsynth` **requires same specifications as** `synth`. **In addition, users may specify**:

- `bcorrect(`*string*`)`
  - One of `nosave`, `merge`, or `replace` must be specified with `bcorrect()`
    - → $\hat{\mu}_{0,t}(x)$ is estimated using OLS regression
    - → Requires at least $K + 2$ donor pool units, K is # of predictors
  - `figure` may additionally be specified. e.g. `bcorrect(replace figure)`
- `pvalues` calculates RMPSE-ranked p-values from in-space placebo runs
  - If specified with `bcorrect()`, calculates classic and bias-corrected p-values
- `placeboskeep` saves the results of the placebo runs estimated for `pvalues`
  - May only be specified when both `keep()` and `pvalues` are also specified
- `gapfigure(`*string*`)`
  - One of `classic`, `bcorrect`, or `placebos` must be specified with `gapfig()`
    - At most two may be specified together
  - `lineback` may additionally be specified. e.g. `gapfig(bcorrect lineback)`
- `allsynth` will always warn if the $\hat{\mathbf{W}}$ matrix unlikely unique

allsynth: Can be used like synth

Same primary specification as in the synth **help file yields same results:**

```
#delimit ;
   allsynth
     cigsale beer(1984(1)1988) lnincome retprice age15to24
     cigsale(1988) cigsale(1980) cigsale(1975),
     trunit(3) trperiod(1989)
   #delimit cr
```

allsynth: Can be used like synth

Also lets you know that you haven't properly specified bias-correction:

Louisiana	0
Maine	0
Minnesota	0
Mississippi	0
Missouri	0
Montana	0
Nebraska	0
Nevada	.245
New Hampshire	0
New Mexico	0
North Carolina	0
North Dakota	0
Ohio	0
Oklahoma	0
Pennsylvania	0
Rhode Island	0
South Carolina	0
South Dakota	0
Tennessee	0
Texas	0
Utah	.369
Vermont	0
Virginia	0
West Virginia	0
Wisconsin	0
Wyoming	0

Predictor Balance:

	Treated	Synthetic
beer(1984(1)1988)	24.28	23.22596
lnincome	10.03176	9.867266
retprice	66.63684	65.40743
age15to24	.1786624	.1825559
cigsale(1988)	90.1	92.6063
cigsale(1980)	120.2	120.3907
cigsale(1975)	127.1	126.7094

```
Plain vanilla -synth- estimates provided. No bias correction or p-value calculations specified or applied
If bcorrect() was specified but no option was input inside the parentheses, bias-correction was not applied.
>  bcorrect() must contain one of -nosave-, -merge-, or -replace- for bias-correction to be applied
```

allsynth: Cautions the user if the $\hat{\mathbf{W}}$ matrix is unlikely unique

e.g. If we specify too few predictor variables, we get a warning message:

```
#delimit ;
    allsynth
      cigsale beer retprice cigsale(1980),
      trunit(3) trperiod(1989)
    #delimit cr
```

$\rightarrow$

```
Warning: the -synth- weighting matrix W for treated unit (state == 3) contains more non-zero weights
> than predictor variables and is likely not unique. Consider adding predictor variables
```

`allsynth`: Can calculate, display, and save classic & bias-corrected "gaps"

Add `keep(smokingresults) replace figure bcorrect(replace figure)`:

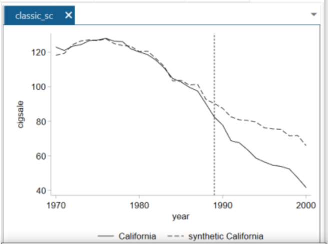
```
#delimit ;
    allsynth
      cigsale beer(1984(1)1988) lnincome retprice age15to24
      cigsale(1988) cigsale(1980) cigsale(1975),
      trunit(3) trperiod(1989)
      keep(smokingresults) replace figure
      bcor(replace figure)
    #delimit cr
```

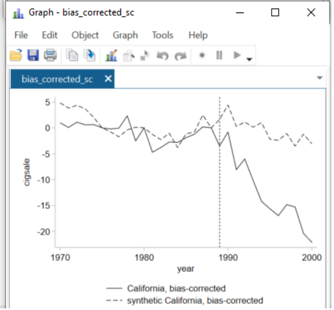
The bias-corrected outcome values are only useful to calculate the bias-corrected gaps!

allsynth: Can calculate, display, and save classic & bias-corrected "gaps"

Treated unit results:

	state	year	gap	gap_bc
1.	3	1970	4.714999	-3.847018
2.	3	1971	1.701101	-3.750713
3.	3	1972	-.9330038	-3.256549
4.	3	1973	-2.134596	-3.091832
5.	3	1974	-.5483028	-1.377421
6.	3	1975	.3905984	0
7.	3	1976	.2115016	.5977706
8.	3	1977	1.404701	1.673252
9.	3	1978	2.2231	2.723592
10.	3	1979	-1.531299	-2.652859
11.	3	1980	-.1907034	0
12.	3	1981	-2.0393	-3.449938
13.	3	1982	-.9850991	-1.524845
14.	3	1983	-.8995969	-1.701223
15.	3	1984	1.391301	.995675
16.	3	1985	-.9249966	-.7384964
17.	3	1986	-1.280606	-.4435956
18.	3	1987	-3.860799	-2.29259
19.	3	1988	-2.5063	0
20.	3	1989	-7.887098	-5.108345
21.	3	1990	-9.693599	-5.212851
22.	3	1991	-13.8027	-8.347837
23.	3	1992	-13.344	-7.124086
24.	3	1993	-17.0624	-10.32301
25.	3	1994	-20.8943	-15.22401
26.	3	1995	-19.8568	-13.43382
27.	3	1996	-21.0405	-14.57536
28.	3	1997	-21.4914	-13.74036
29.	3	1998	-19.1642	-11.78082
30.	3	1999	-24.554	-19.20885
31.	3	2000	-24.2687	-19.05358





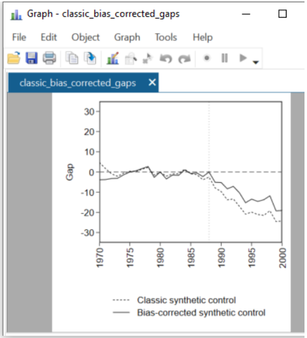
allsynth: Can generate nice graphs of the classic and bias-corrected gaps

Also add `gapfig(classic bcorrect lineback)`**. Drop** `fig` **and** `bcor(figure)`**:**

```
#delimit ;
    allsynth
      cigsale beer(1984(1)1988) lnincome retprice age15to24
      cigsale(1988) cigsale(1980) cigsale(1975),
      trunit(3) trperiod(1989)
      keep(smokingresults) replace
      bcorrect(replace) gapfig(classic bcorrect lineback)
    #delimit cr
```

allsynth: Can generate nice graphs of the classic and bias-corrected gaps

	state	year	gap	gap_bc
1.	3	1970	4.714999	-3.847018
2.	3	1971	1.701101	-3.750713
3.	3	1972	-.9330038	-3.256549
4.	3	1973	-2.134596	-3.091832
5.	3	1974	-.5483028	-1.377421
6.	3	1975	.3905984	0
7.	3	1976	.2115016	.5977706
8.	3	1977	1.404701	1.673252
9.	3	1978	2.2231	2.723592
10.	3	1979	-1.531299	-2.652859
11.	3	1980	-.1907034	0
12.	3	1981	-2.0393	-3.449938
13.	3	1982	-.9850991	-1.524845
14.	3	1983	-.8995969	-1.701223
15.	3	1984	1.391301	.995675
16.	3	1985	-.9249966	-.7384964
17.	3	1986	-1.280606	-.4435956
18.	3	1987	-3.860799	-2.29259
19.	3	1988	-2.5063	0
20.	3	1989	-7.887098	-5.108345
21.	3	1990	-9.693599	-5.212851
22.	3	1991	-13.8027	-8.347837
23.	3	1992	-13.344	-7.124086
24.	3	1993	-17.0624	-10.32301
25.	3	1994	-20.8943	-15.22401
26.	3	1995	-19.8568	-13.43382
27.	3	1996	-21.0405	-14.57536
28.	3	1997	-21.4914	-13.74036
29.	3	1998	-19.1642	-11.78082
30.	3	1999	-24.554	-19.20885
31.	3	2000	-24.2687	-19.05358

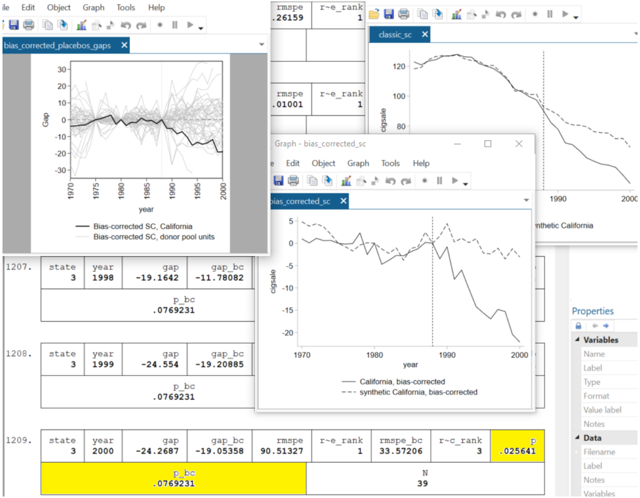


`allsynth`: Can run placebo tests, calculate p-values, and graph permutation distributions

Instead add `gapfig(bcorrect placebos lineback) pvalues placeboskeep`:

```
#delimit ;
    allsynth
      cigsale beer(1984(1)1988) lnincome retprice age15to24
      cigsale(1988) cigsale(1980) cigsale(1975),
      trunit(3) trperiod(1989)
      bcor(replace figure) gapfig(bcorrect placebos lineback)
      pval plac keep(smokingresults) rep
    #delimit cr
```

allsynth: Can run placebo tests, calculate p-values, and graph permutation distributions



Installing `allsynth` package for Stata: currently Version 0.0.7 BETA

In Stata, type:

```
net from https://justinwiltshire.com/s
net install allsynth, replace
help allsynth
```

There are nine examples in help file to teach the functionality of `allsynth`

Version 0.0.5 BETA contained a critical bug. Please update to the latest version

The `allsynth` package is a free contribution to the research community. Please cite it:

Wiltshire, Justin C. 2021b. allsynth: Synthetic Control Bias-correction Utilities for Stata. Working paper.

Email comments and questions: jcwiltshire@ucdavis.edu

References

Abadie, A and J L'Hour. 2020. A Penalized Synthetic Control Estimator for Disaggregated Data. *Working paper.*

Abadie, Alberto. 2021. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature* 59(2): 391–425.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. 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.

———. 2015. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science* 59(2): 495–510.

Abadie, Alberto and Javier Gardeazabal. 2003. The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* 93(1): 113–132.

Abadie, Alberto and Guido W Imbens. 2011. Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business Economic Statistics* 29(1): 1–11.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2021. The Augmented Synthetic Control Method. *NBER Working paper.*

———. 2021. Synthetic Controls with Staggered Adoption. *NBER Working paper.*

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. Catastrophic Natural Disasters and Economic Growth. *Review of Economics and Statistics* 95(5): 1549–1561.

Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu. 2019. An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. *Working paper.*

Doudchenko, Nikolay and Guido W Imbens. 2016. Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis. *NBER Working paper.*

Dube, A and B Zipperer. 2015. Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies. *IZA Discussion Papers 8944.*

Ferman, Bruno and Cristine Pinto. 2017. Placebo Tests for Synthetic Controls. *Working paper.*

Firpo, Sergio and Vitor Possebom. 2018. Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets. *Journal of Causal Inference* 6(2).

Galiani S, Quistorff B. 2017. The synth_runner Package: Utilities to Automate Synthetic Control Estimation Using Synth. *The Stata Journal* 17(4): 834-49.

Hahn, Jinyong and Ruoyao Shi. 2017. Synthetic Control and Inference. *Econometrics* 5(4): 52.

Wiltshire, Justin C. 2021a. Walmart Supercenters and Monopsony Power: How a Large, Low-Wage Employer Impacts Local Labor Markets. *Working paper.*