Article: Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets
Authors: Sergio Firpo (*Insper*) and Vitor Possebom (*Yale University*)

Confidence Sets for the Synthetic Control Method R and Stata functions - Read Me

Firpo & Possebom (2017) developed confidence sets for the Synthetic Control Method, proposed by Abadie & Gardeazabal (2003), Abadie et al. (2010) and Abadie et al. (2015). We provide a function in *R* and a program in *Stata* to compute these confidence sets. We stress that both codes are beta, i.e., they are preliminary and are still being tested. For this reason, if you encounter any problem with it, email us (Vitor Possebom - vitoraugusto.possebom@yale.edu; Sergio Firpo - firpo@insper.edu.br) so that we can help you solve it and update our codes.

For each software, we made available two types of files:

1. The file whose name starts with *function* is a function in *R* or a program in *Stata* that you can use to compute confidence sets for your applied question after estimating the treatment and the placebo effects using either the R package *Synth* (Abadie et al. 2011) or the Stata command *synth*.

2. The file whose name starts with *california* is one empirical example (Abadie et al. 2010) that illustrates how to save the estimated synthetic weights and, then, compute the confidence sets for the Synthetic Control Estimator.

Both files have detailed comments in order to make any required adaptation easier for you. For example, we compute our confidence sets using the test statistic *RMPSE*, proposed by Abadie et al. (2015). If you want to use a different test statistic, it is easy to make changes to our code.

Below, we describe the required arguments of our Stata program:

- 'Ymat' is a matrix that contains the realized outcome of interest. Each column represents a region and each row represents a time period.

- 'weightsmat' is a matrix that contains the estimated weights for the synthetic control unit of each region. Each columns represents a placebo region and each row, a comparison unit. Pay attention that the order of the regions in matrix 'weightsmat' must be identical to the order of the regions in matrix 'Ymat'. 'weightsmat' is a matrix with J rows and J+1 columns, where J+1 is the number of observed regions.

- 'treated' is the column in matrix 'Ymat' associated with the treated region. It must be a natural number less than or equal to the number of columns in matrix 'Y'.

- 'T0' is the row in matrix 'Ymat' associated with the last pre-intervention period. It must be a natural number less than or equal to the number of rows in matrix 'Ymat'.

- 'phi' is the sensitivity parameter defined either in step 6 or in step 7 of section 3 of Firpo & Possebom (2017). It has to be a positive real number. If you only want to construct the standard confidence interval under the assumption that each region is equally likely to receive treatment, set 'phi' to zero.

- 'v' is the worst (best) case scenario vector defined in step 6 (step 7) of section 3 of Firpo & Possebom (2017). It is a row vector whose length is equal to the number of observed regions, i.e., J+1. The elements of this vector must be equal to 0 or 1. If you only want to construct the standard confidence interval under the assumption that each region is equally likely to receive treatment, let 'v' be a zero vector.

- 'precision' is a natural number. A larger value for 'precision' makes the estimation of the confidence sets more precise, requiring more computing time. A value between 20 and 30 is reasonable.

- 'etype' is a character vector equal to "linear", "constant" or "uniform".[1] 'etype' defines which confidence subset is implemented.

- 'significance' is the significance level in decimal form. It must be a number in the following set: $\left\{\frac{1}{J+1}, \frac{2}{J+1}, \ldots, \frac{J+1}{J+1}\right\}$.

Our R function has only one extra argument:

- 'plot' is a logical value that determines whether a plot with the computed confidence set will be provided or not. The user may want to manually change the graphical parameters at the end of this code.

Any questions, suggestions, comments, critics can be sent to Vitor Possebom or Sergio Firpo by email.

References

Abadie, A., Diamond, A. & Hainmueller, J. (2010), 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program', *Journal of the American Statiscal Association* **105**(490), 493–505.

Abadie, A., Diamond, A. & Hainmueller, J. (2011), 'Synth: An R Package for Synthetic Control Methods in Comparative Case Studies', *Journal of Statistical Software* **42**(13), 1–17.

[1]The uniform confidence set was proposed by Ferman et al. (2020) and contains all function that that are deviations from the estimated treatment effect by an additive and constant factor and are not rejected by the placebo test.

Abadie, A., Diamond, A. & Hainmueller, J. (2015), 'Comparative Politics and the Synthetic Control Method', *American Journal of Political Science* **59**(2), 495–510.

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

Ferman, B., Pinto, C. & Possebom, V. (2020), 'Cherry Picking with Synthetic Controls', *Journal of Policy Analysis and Management* **39**(2), pp. 510–532.

Firpo, S. & Possebom, V. (2017), Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets.