Manual for the Synthetic Control Method App

This App has been created by Yujiao Li based on the research in Carling and Li (2016) and Rudholm et al. (2018).

Comparative case studies can easily be performed by using the **Synthetic Control Method (SCM) developed by Abadie et al. (2010)** in this App. This only requires users to upload their data, define what unit is being given a treatment and at which point in time. All results and figures are then automatically generated in green windows. **The procedure has 3 steps:**

1. Click this website

https://wishes.shinyapps.io/intervention_effect_study/

2. User Input

Orange windows require you to import your dataset and set some key parameters.

2.1 Data import

- The App uses data in the following formats, "data.csv" or "data.dta" (Stata datafile)
- The App can manage some missing observations in the predictors variables, but it is sensitive to missing data in the outcome variable. Avoid having string values (text) in any of the numerical variables used in the estimations, and if possible use datasets where all variables have a full set of observations.
- Upload dataset from your local file (presented in "Data" in the left menu bar)
- Input variables which represent the treated unit ID, name and time of treatment.
- The dataset used in our paper "How Does Big-Box Entry Affect Labor Productivity in Durable Goods Retailing? A Synthetic Control Approach" Rudholm et al. (2018) is uploaded as the default dataset. As such, the App can be used to replicate our results, and the examples in this manual show how to replicate the results related to how the IKEA entry in the municipality of Haparanda, Sweden, in 2006 affected labor productivity in durable goods retailing. By adjusting different aspects of the SCM input to the App, the user can also use it together with the default data to investigate how sensitive our results are to changes in the model presented in Rudholm et al. (2018).

2.2 SCM input (Required fields)

- Input the name of the treated unit, the time treatment took place and the outcome variable of interest. For our example, this is then Haparanda, 2006, and Productivity.
- "Matching Variables" lists the predictor variables used to create the synthetic control unit. It should include the outcome variable and some additional variables believed from theory or

previous studies to be correlated to the outcome variable. The variables used in Rudholm et al. (2018) are presented as the default.

- "Matching time" represents the time period used for creating the synthetic control unit and it generally ends the year (time period) before the treatment takes place. As IKEA entered Haparanda in 2006, 2005 is the default.

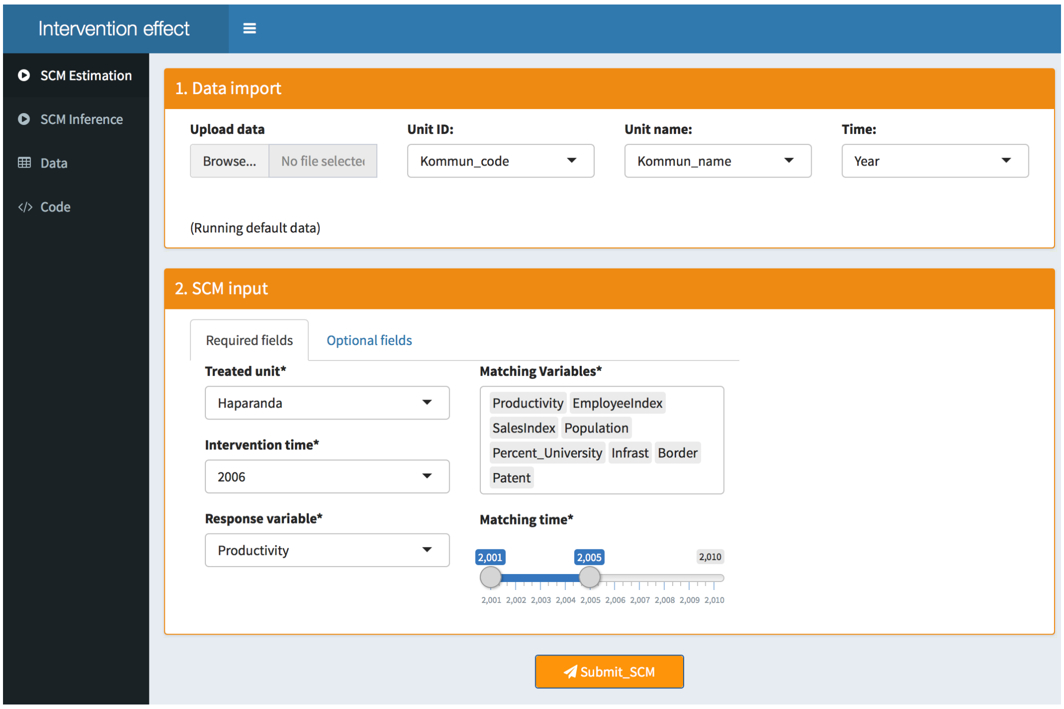


Figure 1: Required fields.

2.3 SCM input (Optional fields)

This step is necessary to replicate the results for Haparanda from Rudholm et al. (2018), but can otherwise be skipped unless you have any of the following special requirements:

- You need exclude special units from your donor pool. Special units can for example consist of other units receiving similar treatments at the same point in time. In our example, we exclude the three other municipalities that had an IKEA entry at approximately the same time as Haparanda (Kalmar, Karlstad and Göteborg).
- You need refine the donor pool because there are too many units in the donor pool for the SCM to work well. Since there are 290 Swedish municipalities this is the case for us. Although not being a fixed limit, SCM works best if the number of possible control units is around 30, but kept below 50. See Carling and Li (2016) and Rudholm et al. (2018) for discussions regarding the optimal size of the donor pool and how to choose cluster variables. In most cases one

can use the same variables for the clustering as the predictors chosen for the SCM. The same predictors as for the SCM are presented in the default settings along with the years 2003, 2004 and 2005, since this replicate Rudholm et al. (2018).

- You want to set some special predictors and their matching periods to make the estimation more precise. Without setting special predictors, SCM estimation will treat the outcome variable and other predictors as equally important when creating the synthetic control unit. However, we would prefer the synthetic control municipality to be more similar to Haparanda with respect to the development of Productivity in the period before IKEA entry, rather than in other aspects. Designating Productivity as Special predictor does that, while restricting the years to 2003, 2004 and 2005 mean that we consider the development of Productivity close to the treatment period in 2006 to be more important than observations longer back in time. In other words, assuming that there are two alternative synthetic control units with the exact same goodness of fit over the whole time period leading up to treatment, setting the years 2003, 2004 and 2005 will make the App choose the synthetic control unit with better goodness of fit in these years rather than over the whole time period leading up to treatment. To replicate Rudholm et al. (2018), use Productivity and the years 2003, 2004 and 2005, which is also presented as the default.

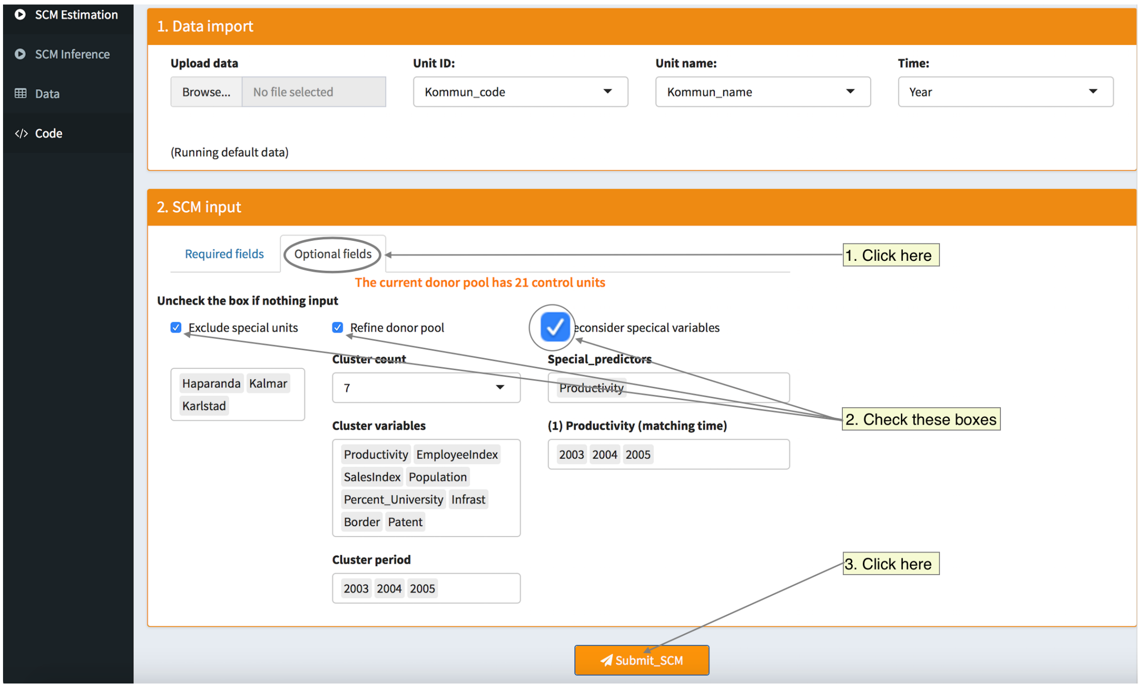


Figure 2: Optional fields.

- Click the "Submit_SCM" button and you will get the following output (see Figure 3). The procedure might take some time so be patient. The output shows how the treated unit (Haparanda) and the created synthetic control unit compare with regard to the variables used in the matching procedure, as well as what units from the donor pool of potential controls are used to create the synthetic control, in our case 68% Strömstad and 32% Krokom. Note that these graphs and the composition of control units used to create the synthetic control unit are exactly as reported in Rudholm et al. (2018) if following the instructions above. Also note that although there are 21 units in the donor pool (see Figure above), only two of these are used to create the synthetic control unit.

- If you need to re-run the SCM procedure for any reason, you should always refresh the web-page.

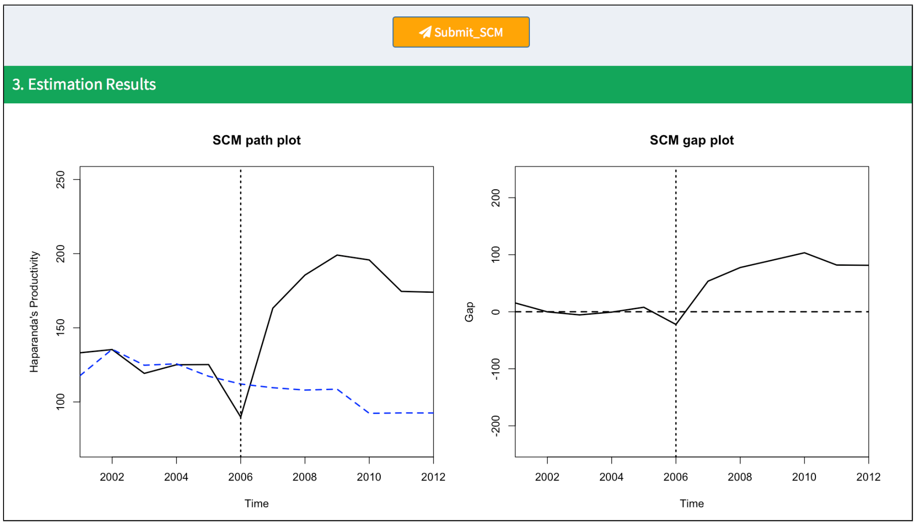


Figure 3: Submit output.

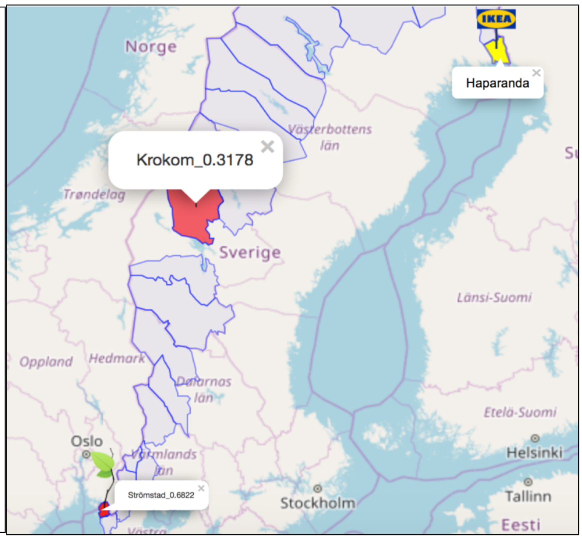
Treated unit vs. Synthetic control unit

Variable	Treated	Synthetic	Sample.Mean
Productivity	127.60	124.20	88.48
EmployeeIndex	15.91	59.40	22.98
SalesIndex	20.32	77.42	21.12
Population	10291.20	12190.32	9941.09
Percent_University	0.05	0.07	0.05
Patent	39.20	309.20	93.14
special.Productivity.2003.2005	123.17	122.60	88.87

Weight of control units

Knname	w.weight
Strömstad	0.68
Krokom	0.32

(a) Synthetic control units weights.



(b) Synthetic control units map.

Figure 4: Synthetic control units.

3. SCM inference

3.1 Click SCM inference in the left side menu bar.

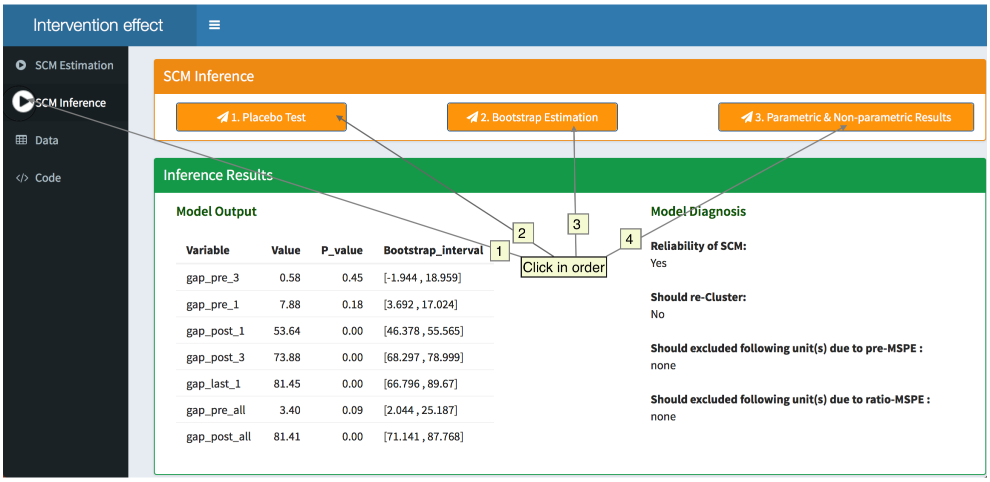


Figure 5: SCM Inference menu bar option.

3.2 Sequentially click the three orange buttons.

- Click button **"1. Placebo Test"** and wait until placebo tables and plots are shown in the green output windows. This might take some time so be patient. When doing this the App also presents model diagnostics indicating if the SCM is reliable, if the clustering of the donor pool should be re-done, and if one should exercise the option to exclude some specific units from the donor pool.
- Click button **"2. Bootstrap Estimation"** and wait until bootstrap results are presented. This might take some time so be patient.
- Click button **"3. Parametric & Non-parametric Results"** and then parametric & non-parametric estimation graphs and tables will be shown as below. This might take some time so be patient.

The model output gives the following information: Gap_pre_3 and Gap_pre_1 gives the difference between Productivity in Haparanda and the synthetic control unit 3 and 1 year before the intervention, Gap_post_1 and Gap_post_3 gives the difference 1 and 3 year after the intervention, while Gap_last_1 is the difference the last year under study. Based on the placebo tests there are no significant differences between Haparanda and the synthetic control at any time before IKEA entry[1]

[1] To guarantee that Haparanda have a comparable synthetic control unit, this requires a similarity test in terms of their characteristics before the intervention, i.e. before IKEA entry. The p-value is presented in the App for this purpose, where the null hypothesis is that the intervention unit and it's synthetic control unit are not different. So if we get a p-value < 0.05, it means that Haparanda and the synthetic control unit are significantly different from each other. In our case, $p > 0.05$ (p=0.45 and 0.18), which shows that Haparanda and its synthetic control unit are not different in the period before IKEA entry, i.e. they are comparable. After IKEA entry this changes as there clearly are statistically significant differences between Haparanda and the synthetic control unit based on the p-values.

while the bootstrap interval show a significant difference in the last year before IKEA entry, possibly indicating that some firms adjusted to IKEA prior to actual entry. For the period after IKEA entry, all reported results show a significant increase in Productivity compared to the synthetic control unit, irrespectively if we use placebo testing or bootstrap confidence intervals. Gap_pre_all presents the average difference in Productivity between Haparanda and its synthetic control unit over all pre-intervention years, while Gap_post_all presents the average over all post-intervention years. Note that although the qualitative results are very similar, the bootstrap inference results presented are not exactly the same as those presented in Rudholm et al. (2018). The differences are due to new random samples being drawn to create the bootstrap intervals each time the procedure is run.

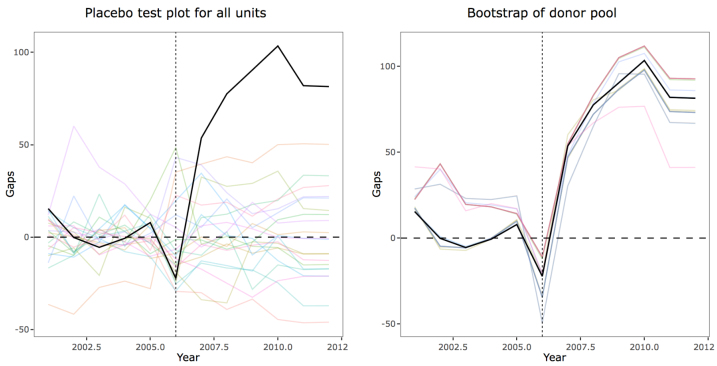
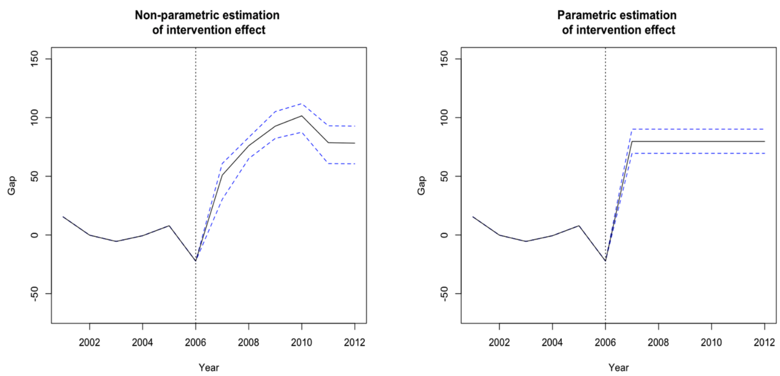


Figure 6: inference plots.



Non-parametric	2.5%	mean	97.5%
2007	30.3198	50.9343	60.8514
2008	65.0914	76.0567	83.1545
2009	82.2625	92.7202	105.1346
2010	87.5509	101.5196	111.8943
2011	60.7957	78.6757	93.046
2012	60.6607	78.2408	92.7443

Parametric	2.5%	mean	97.5%
2007	69.55	79.69	90.14
2008	69.55	79.69	90.14
2009	69.55	79.69	90.14
2010	69.55	79.69	90.14
2011	69.55	79.69	90.14
2012	69.55	79.69	90.14

Figure 7: Non-parametric and parametric estimation of treatment effect.

References

Abadie, A., Diamond, A. and Hainmueller, J. (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.

Carling, K. and Li, Y. (2016), The power of the synthetic control method, Working papers in transport, tourism, information technology and microdata analysis No.2016:10, Dalarna University.

Rudholm, N., Li, Y. and Carling, K. (2018), How does big-box entry affect labor productivity in durable goods retailing? a synthetic control approach, HUI Working Paper No. 130, HUI Research.