Eco 311 Optional Reading: Synthetic Control

(Jing Li, Miami University)

1. We use difference-in-difference (DID) method to estimate treatment effect when there are a lot of treated and untreated units. For instance, building a garbage incinerator could have negative impact on many houses close to it, while more houses far away are unaffected. So both treatment and control groups have many observations, and the DID method is appropriate in such setting.

2. The method of Synthetic Control (SC) is a macro version DID, which can be used when only one aggregate unit (country, state, ...) is exposed to treatment or intervention.

   (a) Example 1: In 1997 the sovereignty of Hong Kong was returned to China. In that year, there was no sovereignty transfer in other Asian countries.

   (b) Example 2: In 2011 Germany decided to shut down its nuclear power plants. Again, no other European countries saw such big change of energy policy.

3. In both examples, we are interested in the whole unit at aggregate level, not their micro-level individual persons, counties... Thus, there is only one unit in the treatment group. Nevertheless, the control or untreated group consists of several aggregate units—we can compare Hong Kong to Singapore, Taiwan, and South Korea (together they are called four little dragons), or we can contrast Germany with European countries that keep their nuclear power plants such as France and Spain.

4. Let us focus on the Hong Kong example. The SC method aims to construct an imaginary or synthetic Hong Kong, which is a weighted average of donor pool made of Singapore, Taiwan, South Korea, and Japan. The weights are data-driven. We hope the synthetic Hong Kong can mimic Hong Kong as much as possible before 1997. In that case, we can use the outcome of synthetic Hong Kong after 1997 as an estimate for the counterfactual (potential outcome) of what would have happened to Hong Kong in the absence of sovereignty transfer. The gap between the two Hong Kongs after 1997 can be viewed as evidence for the treatment effect.

5. The SC method requires panel data in long form, in which there are multiple observations for treated and untreated units, before and after the intervention date. The Hong Kong data span from 1989 to 2008, and for each year, we have observations for

Hong Kong and other four countries. By contrast, the panel data used by DID typically have a big number of units and a small number of time periods. In light of that difference, the panel data for SC research could alternatively be treated as multivariate time series.

6. The other four countries are collectively called donor pool, or control group in general. To ensure apple-to-apple comparison, the donor pool is carefully chosen, and the emphasis is on their similarity to Hong Kong economy. It makes no sense to include in the donor pool, say, Cambodia because it was much poorer than Hong Kong. Macau is excluded in the donor pool as its sovereignty was transferred to China in 1999.

7. We should have enough number of pre-treatment periods in order to capture the common trends. The post-treatment periods should also be long enough to allow for gradual treatment effect over time. It is challenging to interpret the finding of SC analysis if there are more than one treatment occurring in the same period. For instance, there were two shocks—sovereignty transfer and financial crisis that affected Hong Kong in 1997. So the authors of Hong Kong paper need to disentangle the two treatment effects.

8. The key outputs of synthetic control analysis are two sets of weights

   (a) The weights for donor pool units are $w = (w_1, \ldots, w_n)$, an n by 1 vector. For Hong Kong example we let $n = 4$

   (b) The weights for predictors are $v = (v_1, \ldots, v_k)$, where k is the number of predictors. Later we will try $k = 3$

   Those weights are obtained by solving a problem of two nested optimizations.

   (a) For given v, we can create a diagonal matrix V using v. Then w is the solution of minimizing a quadratic form subject to certain restrictions

$$\min(X_1 - X_0 w)^T V (X_1 - X_0 w) \quad st\ w_i \geq 0, \forall i,\ and \sum w_i = 1 \tag{1}$$

   where T denotes transpose, X_1 is a k by 1 vector of values of predictors for Hong Kong, and X_0 is a k by n matrix of values of predictors for donor pool. Basically, we hope the optimal weighted average $X_0 w$ can produce the best approximation for X_1. In some sense, $(X_1 - X_0 w)^T V(X_1 - X_0 w)$ is analogous to the residual sum square (RSS)[1] when we solve the ordinary least squares problem.

(b) The best v is found by minimizing mean squared prediction error (MSPE) subject to restriction

$$\min(Z_1 - Z_0 w(v))^T (Z_1 - Z_0 w(v)) \quad st\ v_i \geq 0, \forall i,\ and\ \sum v_i = 1 \tag{2}$$

where Z_1 is a h by 1 vector of values of outcome variable for Hong Kong, and Z_0 is a h by n matrix of values of outcome variables for donor pool. Again, we hope to find an optimal weighted average $Z_0 w(v)$ that can approximate Z_1 well. R calls the value of $(Z_1 - Z_0 w(v))^T (Z_1 - Z_0 w(v))$ Loss V.

(c) If we do not use matrix notation, the objective functions in (1) and (2) can be rewritten as

$$(X_1 - X_0 w)^T V(X_1 - X_0 w) = \sum_{i=1}^{k} v_i \left(x_{1i} - \sum_{j=1}^{n} w_j x_{0ij} \right)^2 \tag{3}$$

where x is predictor, i is the index for predictor, and j is index for donor pool unit;

$$(Z_1 - Z_0 w(v))^T (Z_1 - Z_0 w(v)) = \sum_{t=1}^{h} \left(y_{1t} - \sum_{j=1}^{n} w_j y_{0tj} \right)^2 \tag{4}$$

where y is the outcome variable. To learn more about the SC method, please read `https://www.aeaweb.org/articles?id=10.1257/jel.20191450`.

(d) It is challenging to find v and w because (i) we need to impose non-negative restriction for each weight and restriction that the sum of weights equals one; (ii) the quadratic form in (1) is not necessarily positive definite. Anyway, the Synth package of R can take care of those technical difficulties. The detail of how to use Synth package can be found at `https://www.jstatsoft.org/article/view/v042i13`.

9. Here, we include 4 units in the donor pool (so $n = 4$), and we choose three predictors (so $k = 3$)—they are averages unemployment rate (un) and average investment share

[1] Using matrix notation we can express RSS as $RSS = (Y - X\beta)^T (Y - X\beta)$. Minimizing RSS is much easier than solving (1) because there is no restriction on β, and in general, the quadratic form is positive definite.

(inv) from 1989 to 1996, and average purchasing power parity adjusted GDP from 1991 to 1993. Keep following in mind when selecting predictors

(a) The value of predictor of treated unit cannot be minimum or maximum among all units. That is why we do not consider inflation rate as the predictor since Hong Kong has the highest inflation rate before 1997 among all units in data. Mathematically, if $a > b > c$, there is no weighted average of b and c that can duplicate a, (i.e., no $0 \leq w \leq 1$ can make $wb + (1 - w)c = a$), or no weighted average of a and b that can duplicate c.

(b) It is important to use lagged values of outcome variable as predictors since they can capture unobserved confounders. That is why we consider the average purchasing power parity adjusted GDP from 1991 to 1993

(c) If the number of predictors k is less than the number of donor pool units n, then the matrix $X_0^T V X_0$ in quadratic form (1) does not have full rank. This can pose extra challenge[2].

10. The outcome variable is purchasing power parity adjusted GDP (pp). Again we need to think carefully when choosing the outcome variable. Suppose Hong Kong has the highest emigration rate, then we cannot use emigration rate as the outcome variable since no weighted average of donor pool units can duplicate it.

11. The panel data is sorted by year (ye) and country index (ci) that equals one for Hong Kong. The ci variable provides index for treated unit ($i = 1$) and untreated units ($i = 2, 3, 4, 5$).

```
library(Synth)
library(readxl)
setwd("/Users/lij14/Dropbox")
d = data.frame(read_xlsx("scmhkdata.xlsx"))
str(d)
names(d) = c("co","ye","un","op","pi","pp","sch","sto","lif","tfp","inv","gr","re")

d$ci = rep(0,length(d$co))
```

[2]We can show $(X_1 - X_0 w)^T V (X_1 - X_0 w) = X_1^T X_1 - 2X_1^T X_0 w + w^T X_0^T V X_0 w$. If we ignore the restriction, this becomes a standard quadratic programming problem, for which a solution exists when $X_0^T V X_0$ has full rank and is non-singular.

```
d$ci[d$co=="hong kong"]=1
d$ci[d$co=="japan"]=2
d$ci[d$co=="singapore"]=3
d$ci[d$co=="south korea"]=4
d$ci[d$co=="taiwan"]=5
d$ci[d$co=="indonesia"]=6
d$ci[d$co=="malaysia"]=7
d$ci[d$co=="philippines"]=8
d$ci[d$co=="thailand"]=9
d$ci = as.numeric(d$ci)
d = d[d$ci<6,]
d = d[order(d$ci,d$ye),]
```

Next, we tell R how to obtain X_0, X_1, Z_0 and Z_1 with dataprep function

```
dataprep_out = dataprep(
  foo = d,
  predictors = c("un","inv"),
  predictors.op = "mean",
  time.predictors.prior = c(1989:1996),
  special.predictors = list(
          list("pp",1991:1993,"mean")
  ),
  dependent = "pp",
  unit.variable = "ci",
  unit.names.variable = "co",
  time.variable = "ye",
  treatment.identifier = 1,
  controls.identifier = c(2,3,4,5),
  time.optimize.ssr = c(1994:1996),
  time.plot = 1989:2008
)
```

Notice that the data for Z_0 spans from 1994 to 1996, so $h = 3$

12. After that, synth function is used to find those optimal weights

```
> synth_out = synth(data.prep.obj = dataprep_out)
X1, X0, Z1, Z0 all come directly from dataprep object.
****************
searching for synthetic control unit
****************
MSPE (LOSS V): 385081.4
solution.v:
 9.13382e-05 0.3100151 0.6898936
solution.w:
 0.523315 0.2117927 0.05383438 0.2110579
```

The MSPE (LOSS V) can be used to pick the best specification that produces the smallest MSPE.

13. The weights for donor pool united are

```
> out.table = synth.tab(dataprep.res=dataprep_out,synth.res=synth_out)
> out.table$tab.w
  w.weights  unit.names unit.numbers
2     0.523       japan            2
3     0.212   singapore            3
4     0.054 south korea            4
5     0.211      taiwan            5
```

So the synthetic Hong Kong is an optimal weighted average of donor pool given by

$$0.523 Japan + 0.212 Singapore + 0.054 SouthKorea + 0.211 Taiwan$$

The weighs for the three predictors are

```
> out.table$tab.v
                     v.weights
un                   0
inv                  0.31
special.pp.1991.1993 0.69
```

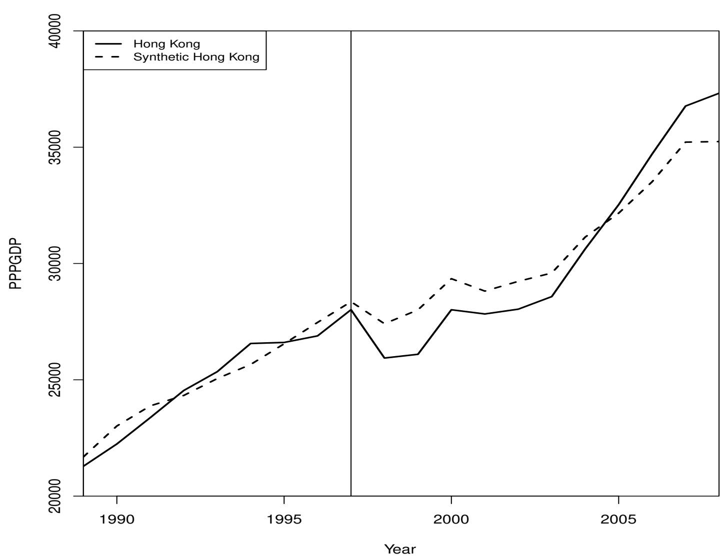
So average of unemployment receives zero weight, and average of purchasing power parity adjusted GDP from 1991 to 1993 receives greater weight than average of investment share. Note that all those weights are data-driven.

14. The synthetic Hong Kong is deemed successful because we find close match of predictors between Hong Kong and synthetic Hong Kong

```
> out.table$tab.pred
                        Treated Synthetic Sample Mean
un                        2.013     2.202       2.113
inv                      35.035    35.036      37.022
special.pp.1991.1993  24420.654 24419.951   20802.893
```

For instance, the average pp from 1991 to 1993 of synthetic Hong Kong is 24419.951, very close to the Hong Kong value 24420.654

15. The punchline of SC research is to plot time series of outcome variable for Hong Kong and its synthetic counterpart



We see a widening gap in PPPGDP after 1997, which indicates the negative treatment effect of soveverinty transfer on Hong Kong economy. In other words, in the absence of sovereignty transfer, Hong Kong economy could have performed better than it actually did (the dash line or potential outcome lies above the solid line or actual outcome after 1997). Also we see that the synthetic Hong Kong traces Hong Kong very well before 1997, another signal for the success of SC method.

16. Other than using dataprep function to prepare the data, we can use following codes to

create X_0 and X_1 by ourselves

```
> r1 = NULL; r2 = NULL; r3 = NULL
> for (j in 1:5) {
+ r1 = c(r1,mean(d$un[d$ye<1997&d$ci==j],na.rm=T))
+ r2 = c(r2,mean(d$inv[d$ye<1997&d$ci==j],na.rm=T))
+ r3 = c(r3,mean(d$pp[d$ye<1994&d$ye>1990&d$ci==j],na.rm=T))
+ }
> X1 = rbind(r1,r2,r3)[,1]
> X0 = rbind(r1,r2,r3)[,-1]
> X1
         r1          r2          r3
    2.01250    35.03502 24420.65437
> X0
           [,1]       [,2]         [,3]        [,4]
r1     2.564583     1.7400     2.439583     1.70750
r2    34.025331    39.8789    43.717947    30.46774
r3 28531.550790 25927.4347 13170.675270 15581.91121
```

They are the same as the output from dataprep

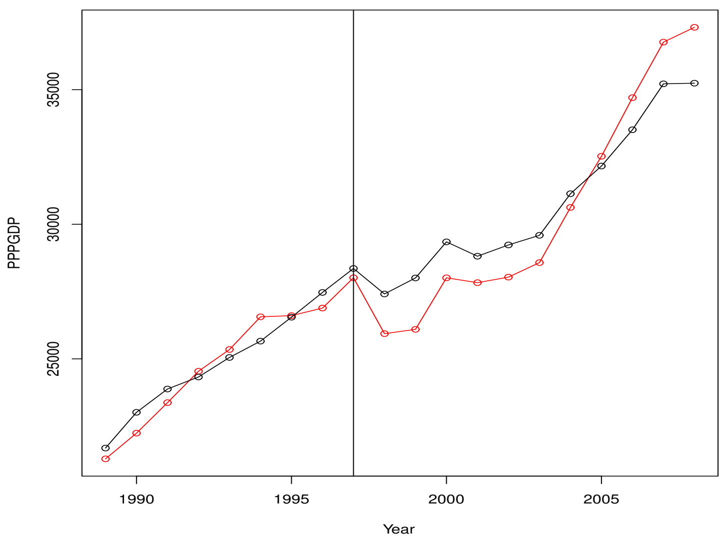
```
> dataprep_out$X1
                                1
un                        2.01250
inv                      35.03502
special.pp.1991.1993 24420.65437
> dataprep_out$X0
                                2          3            4           5
un                       2.564583     1.7400     2.439583     1.70750
inv                     34.025331    39.8789    43.717947    30.46774
special.pp.1991.1993 28531.550790 25927.4347 13170.675270 15581.91121
```

Similarly, we can create Z_0 and Z_1 using following codes

```
> co = NULL
> for (j in 1:5) {
+ co = cbind(co,d$pp[d$ye<1997&d$ye>1993&d$ci==j])
+ }
> Z1 = co[,1]
> Z0 = co[,-1]
> Z0
         [,1]     [,2]     [,3]     [,4]
[1,] 28561.37 29288.16 14825.87 17581.24
[2,] 28970.08 31249.67 15889.18 18542.18
[3,] 29635.03 32874.69 16904.11 19361.19
> Z1
[1] 26561.79 26606.00 26887.97
> dataprep_out$Z0
            2        3        4        5
1994 28561.37 29288.16 14825.87 17581.24
1995 28970.08 31249.67 15889.18 18542.18
1996 29635.03 32874.69 16904.11 19361.19
> dataprep_out$Z1
            1
1994 26561.79
1995 26606.00
1996 26887.97
```

17. We can also manually compute the outcome variable for synthetic Hong Kong, and plot it along with Hong Kong

```
var = "pp"
d.w = cbind(d[[var]][d$ci==2],d[[var]][d$ci==3],d[[var]][d$ci==4],d[[var]][d$ci==5]
syn = d.w%*%synth_out$solution.w
plot(d$ye[d$ci==1],d[[var]][d$ci==1],type="o", col="red",ylab="PPPGDP",xlab="Year")
lines(d$ye[d$ci==1],syn,type="o", col="black")
abline(v=1997)
```


PPPGDP
25000
30000
35000
1990
1995
2000
2005
Year