POLS0012 Causal Analysis: Tutorial Exercises Week 10

In the early 1990s, many African countries went underwent democratisation, ending decades of autocratic post-colonial rule. Economists and political scientists have also spent decades debating whether these democratisations led to higher economic growth. One school of thought (amongst many) holds that democracy is good for growth because it incentivises politicians to invest in projects that boost citizens' wellbeing. Another is that democracy is bad for growth because politicians spend *too much* on inefficient projects that divert money away from more productive private-sector investment. In a recent article, Daniel De Kadt and Steve Wittels[1] investigate this question of whether democratisation boosted growth in African countries using the method of synthetic control. Their original article looked at all cases of democratisation over the period. In this problem set we'll focus on just one of their cases, Mali, which they focus on in particular. They compare Mali's economic growth after its democratisation in 1991 to the growth of a synthetic Mali constructed from a donor pool of all other African countries that remained autocratic over the period, using data from 1975-2008.

You'll need to install the `synth` package for this tutorial. The dataset is called "africa.Rda" and contains the following variables by country and year:

- *Country*: character vector for country
- *wbcode2*: numerical vector for country
- *year*: 1965-2008
- *lngdpmad*: logged real GDP growth, from Angus Maddison historical dataset
- *lnpop*: logged population
- *ki*: investment share of GDP per capita
- *openk*: trade openness (sum of exports and imports divided by GDP)
- *civwar*: =1 if civil war initiated, 0 otherwise
- *civwarend*: =1 if civil war teminated, 0 otherwise
- *pwt_xrate*: exchange rate, relative to US dollar
- *eximdiff*: value of exports - value of imports (index)
- *wbank*: number of world bank projects in place in country-year
- *imfadj*: number of structural adjustment loans from the IMF in place in country-year

[1] De Kadt and Wittels (2017). "Democratization and Economic Output in Sub-Saharan Africa." *Political Science Research and Methods*, Forthcoming

a) Read R's help file for the `dataprep()` function. Then prepare the dataset for synthetic control analysis using `dataprep()`, bearing in mind that democratisation occurred in 1991. You should use the dependent variable and all of the covariates as predictors. Your code should take the following format (fill in everything after the = signs with the relevant names, where appropriate):[2]

```
mali.data <- dataprep(
dataset,
predictors = vector of names of the predictors,
dependent = name of dependent variable,
unit.variable = name of numerical variable identifying each unit,
unit.names.variable = name of character variable identifying each unit,
time.variable = name of variable for time,
treatment.identifier = name of treatment unit,
controls.identifier = c(unique(a$Country[a$Country!="Mali"])),
time.predictors.prior = c(start year:year before treatment),
time.optimize.ssr = c(start year:year of treatment),
time.plot = c(start year:end year)
)
```

Note that I followed the original paper here in beginning estimation in the year 1975, when more data is available. There is nothing particularly wrong with starting in 1965, but you will obtain slightly different answers

```
library(Synth)
 mali.data <- dataprep(a,
                       predictors=c("lngdpmad",
                                    "lnpop",
                                    "ki",
                                    "openk",
                                    "civwar",
                                    "civwarend",
                                    "pwt_xrate",
                                    "eximdiff",
                                    "wbank",
                                    "imfadj"),
                        dependent="lngdpmad",
                        unit.variable="wbcode2",
                        unit.names.variable="Country",
```

[2] You can ignore warnings about missing data

```
time.variable="year",
treatment.identifier="Mali",
controls.identifier=c(unique(a$Country[a$Country!="Mali"])),
time.predictors.prior=c(1975:1990),
time.optimize.ssr=c(1975:1991),
time.plot=c(1975:2008))
```

b) Estimate the synthetic Mali using the `synth()` function, whose only input is the name of the object from a), giving it a name. It will take a minute or two to run

```
synth.mali <- synth(mali.data)
```

c) Using synth's `path.plot()` function, produce a chart comparing Mali's actual GDP growth to that of the synthetic Mali. You code should take the following form:

```
path.plot(
synth.res = name of synthetic mali,
dataprep.res = name of dataprep object,
tr.intake = year of treatment,
Ylab = ,
Xlab =  )
```

See Figure 1

Code:

```
path.plot(synth.res=synth.mali, dataprep.res=mali.data,
          tr.intake=1991,Ylab="Log GDP per capita",Xlab="")
```

d) Read the "Value" section of the help file for `synth()` and use it to find the estimated w vector of country weights. In a table, report the weights assigned for all countries contributing greater than 1% to the synthetic Mali. Which countries contribute the most to the synthetic control?

```
mali.weights <- synth.mali$solution.w
```

Answer:
See Table 1. The biggest contributors are Burundi, Burkina Faso and Togo

Figure 1: **Synthetic Control Analysis of Mali's Democratisation**

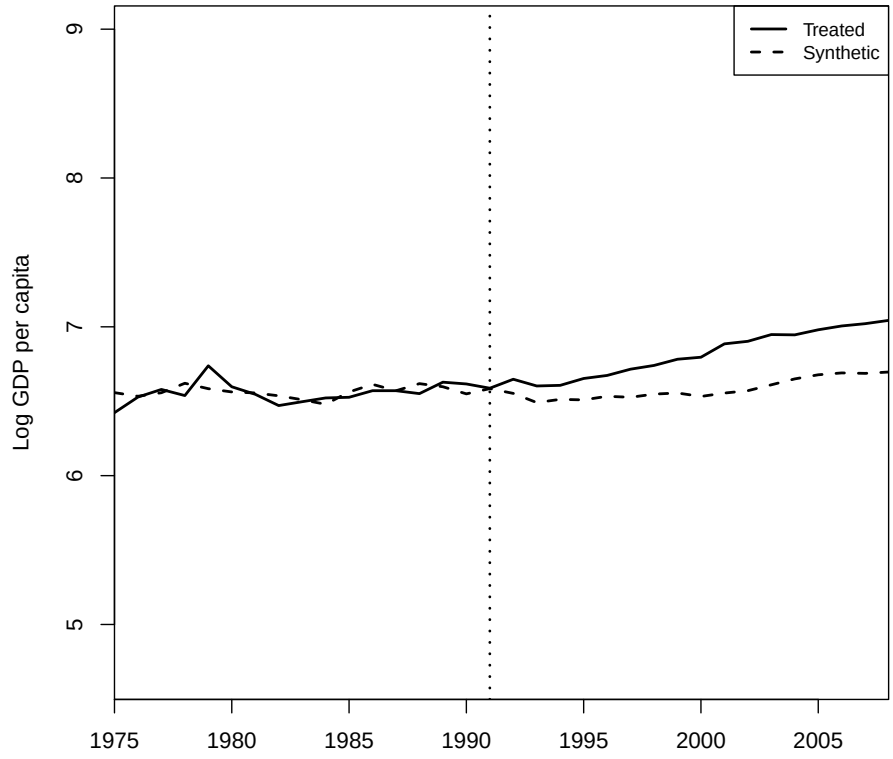


Table 1: **Countries Contributing to Mali's Synthetic Control**

Country	Weight
Burundi	0.39
Burkina Faso	0.24
Niger	0.085
Chad	0.062
Togo	0.23

e) Now find the estimated v vector of weights assigned to each variable and give it a name. In a table, report the weights assigned for all variables contributing greater than 1% to the synthetic Mali. Which variables contribute the most to the synthetic control?

```
mali.vars <- synth.mali$solution.v
```

Answer:
See Table 2. By far the biggest contributor is pre-treatment GDP, with civil war initiation

contributing around 10%. Note that it is very common for the dependent variable to contribute a large proportion to the synthetic control, since the algorithm aims to produce a close fit between the actual and synthetic outcome before the treatment occurs.

Table 2: **Variables Contributing to Mali's Synthetic Control**

Country	Weight
Pre-Treatment GDP	0.82
Civil War Initiation	0.10
Exchange Rate	0.015
IMF Loans	0.047

f) Plot the estimated treatment effect (the difference between the actual and synthetic Mali) over the whole range of the data. To extract Mali's synthetic control, run the following command that uses matrix algebra. "mali.weights" is the w vector from d) and "Y0plot" gives a matrix of the outcome variable where the rows are each year and the columns are each control unit.[3]

```
synthetic.mali <- mali.data$Y0plot %*% mali.weights
```

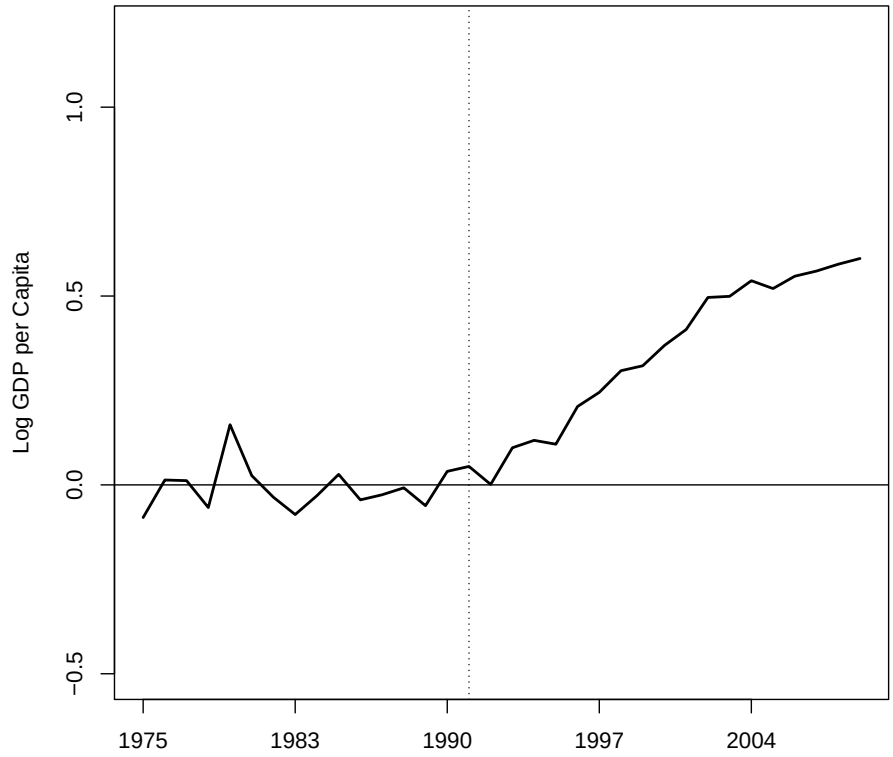
See Figure 2

Code:

```
synthetic.mali <- mali.data$Y0plot %*% mali.weights
actual.mali <- a$lngdpmad[a$Country=="Mali"&a$year>1974]

treatment.effect <- actual.mali - synthetic.mali

plot(treatment.effect,
     type="l",
     lwd=2,
     ylim=c(-0.5,1.2),
     xlab="",
     ylab="Log GDP per Capita",
     xaxt="n")
axis(1,at=c(1,8,15,22,29),labels=c(1975,1983,1990,1997,2004))
abline(v=16,lty=3)
abline(h=0)
```

[3]For those of you who know matrix algebra, "%*%" simply performs matrix multiplication in R. The line of code is multiplying a TxK matrix of GDP by a Kx1 vector of weights to give a Tx1 vector of synthetic GDP, where T is the number of time periods and K is the number of control units

Figure 2: **Estimated Treatment Effect of Mali's Democratisation**



g) Now we'll begin doing some inference by estimating a placebo synthetic control treatment effect for the first control unit, Angola. To do this, repeat steps a), b), d) and f) for Angola, using a dataset called a.angola that excludes Mali, created using the code `a.angola <- subset(a,a$Country!="Mali")`. In the last stage, use the `plot()` and `lines()` commands to place Mali and Angola's treatment effects on the same plot. Interpret the resulting figure.

See Figure 3. In this placebo test, there is no evidence that Angola experienced a positive placebo treatment effect.

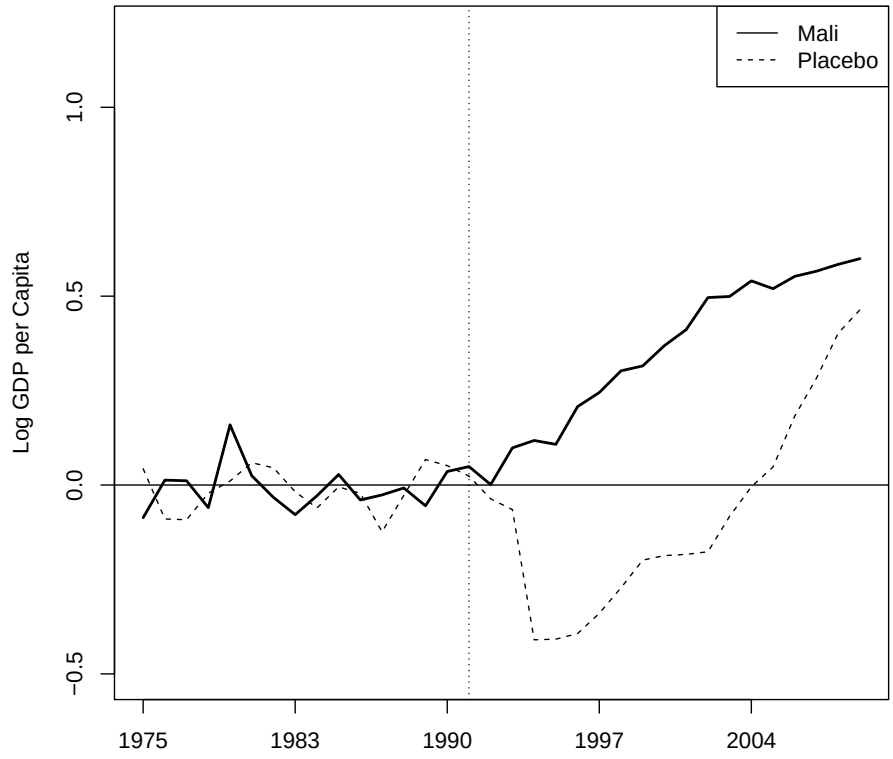
Code:

```
a.angola <- subset(a,a$Country!="Mali")
angola.data <- dataprep(a.angola,
                        predictors=c("lngdpmad",
                                     "lnpop",
                                     "ki",
```

```
                                    "openk",
                                    "civwar",
                                    "civwarend",
                                    "pwt_xrate",
                                    "eximdiff",
                                    "wbank",
                                    "imfadj"),
                        dependent="lngdpmad",
                        unit.variable="wbcode2",
                        unit.names.variable="Country",
                        time.variable="year",
                        treatment.identifier="Angola",
                        controls.identifier=c(unique(a.angola$Country[a.angola$Country!="A
                        time.predictors.prior=c(1975:1990),
                        time.optimize.ssr=c(1975:1991),
                        time.plot=c(1975:2008))

synth.angola <- synth(angola.data)
angola.weights <- synth.angola$solution.w

synthetic.angola <- angola.data$Y0plot %*% angola.weights
actual.angola <- a$lngdpmad[a$Country=="Angola" & a$year>1974]
treatment.effect.angola <- actual.angola - synthetic.angola

plot(treatment.effect,
     type="l",
     lwd=2,
     ylim=c(-0.5,1.2),
     xlab="",
     ylab="Log GDP per Capita",
     xaxt="n")
axis(1,at=c(1,8,15,22,29),labels=c(1975,1983,1990,1997,2004))
abline(v=16,lty=3)
abline(h=0)
lines(treatment.effect.angola,lty=2,lwd=1)
legend("topright",c("Mali","Placebo"),lty=c(1,2))
```

h) **[Challenging Question]** Conduct full inference by using a for() loop to estimate and plot placebo treatment effects for all of the control units on a single figure, comparing them to the actual estimated effect for Mali. What do you conclude?

Figure 3: **Estimated Synthetic Control Treatment Effect of Mali's Democratisation vs. Placebo Effect for Angola**



Hints: Begin by plotting the effect for Mali, then add a new line to this plot with the `lines()` command in each iteration of your loop

See Figure 4. Mali stands out as having by far the largest positive treatment effect, at least over the long run. The likelihood that the result for Mali arose by chance alone is very low.

Code:

```
plot(treatment.effect,
     type="l",
     lwd=2,
     ylim=c(-0.5,1.2),
     xlab="",
     ylab="Log GDP per Capita",
     xaxt="n")
```

```
axis(1,at=c(1,8,15,22,29),labels=c(1975,1983,1990,1997,2004))
abline(v=16,lty=3)
abline(h=0)
legend("topright",c("Mali","Placebo"),lty=c(1,2))


a.controls <- subset(a,a$Country!="Mali")
for(i in 1:length(unique(a.controls$Country))){
  my.country <- unique(a.controls$Country)[i]

  controls.data <- dataprep(a.controls,
                            predictors=c("lngdpmad",
                                         "lnpop",
                                         "ki",
                                         "openk",
                                         "civwar",
                                         "civwarend",
                                         "pwt_xrate",
                                         "eximdiff",
                                         "wbank",
                                         "imfadj"),
                            dependent="lngdpmad",
                            unit.variable="wbcode2",
                            unit.names.variable="Country",
                            time.variable="year",
                            treatment.identifier=my.country,
                            controls.identifier=c(unique(a.controls$Country[a.controls$Cou
                            time.predictors.prior=c(1975:1990),
                            time.optimize.ssr=c(1975:1991),
                            time.plot=c(1975:2008))

  synth.control <- synth(controls.data)
  control.weights <- synth.control$solution.w

  synthetic.control <- controls.data$Y0plot %*% control.weights
  actual.control <- a$lngdpmad[a$Country==my.country & a$year>1974]
  treatment.effect.control <- actual.control - synthetic.control

  lines(treatment.effect.control,lty=2,lwd=1)
}
```

Figure 4: **Estimated Treatment Effect of Mali's Democratisation vs. Placebo Effects for All Controls**