IPSA Multi-Method: Lab 6

Due on Sixth Day

Jason Seawright

Problem 1

Inequality and Democracy, with Matching

In this exercise, we will use matching to set up a multi-method study of whether democracy causes (or reduces) economic inequality. Create a dichotomous version of the Polity regime indicator:

```
inequality$polbin <- 1*(inequality$Polity > 0)
```

Next, load software for matching into R:

```
install.packages("Matching")
library(Matching)
```

Now, build an initial propensity score for democracy:

```
dempropensitylogit <- glm(polbin ~ log(GDP) +
     Industry + FuelExports + CommunistLegacy,
     data=inequality, family=binomial(link=logit))
inequality$pscores <- NA
inequality$pscores[as.numeric(
     names(dempropensitylogit$fitted))]
     <- dempropensitylogit$fitted
```

Now, we discard missing data and carry out the matching:

```
ineqtrim <- inequality[!is.na(inequality$Gini) &
     !is.na(inequality$pscores),]
ineqmatch <- Match(Y=ineqtrim$Gini, Tr=ineqtrim$polbin,
     X=ineqtrim$pscore, estimand="ATE")
summary(ineqmatch)
```

What do you conclude from these results?

What are the key assumptions being made to get to causal inference with this analysis? Design qualitative components to test as many of those assumptions as possible. Select one or more appropriate cases and, using the resources available to you today, carry out as many of those qualitative components as you can. What do you conclude about the original analysis, and how if at all would you redesign it?

For case selection, you can modify the sample code found in the matchingtable R code file.

Problem 2

Legislative Inclusion of Women

Load the data set on women's legislative representation.

```
fourelec <- read.csv("S:/2J Multimethod Research/fourelec.csv")
```

Fit the following logit:

```
maj.glm <- glm(maj ~ prop65 +
     I(log(lyp)) + gastil + federal +
     col_uka + laam,
     family=binomial(link=logit),
     data=fourelec)

summary(maj.glm)
```

Now we will use the logit we estimated above to extract propensity scores:

```
fourelec$pscore <- NA
fourelec$pscore[as.integer(names(maj.glm$fitted.values))] <-
          maj.glm$fitted.values
```

Now we need to create a data set that contains the dependent variable (women's share of legislative seats, or wom1st), the main causal variable (maj), the propensity score, and the variables that went into forming the propensity score. This data set also needs to drop any cases that have missing data on any of those variables:

```
fourelectrim <- na.omit(data.frame(wom_1st=fourelec$wom_1st,
                         maj=fourelec$maj, pscore=fourelec$pscore,
                         prop65=fourelec$prop65, lyp=fourelec$lyp,
                         gastil=fourelec$gastil, federal=fourelec$federal,
                         col_uka=fourelec$col_uka, laam=fourelec$laam,
                         country=fourelec$country))
```

We are now ready to carry out matching:

```
install.packages("Matching")

library("Matching")
firstmatch <- with(fourelectrim,Match(
     Y=wom_1st,Tr=maj,X=pscore,est="ATT"))
summary(firstmatch)
```

We can test the quality of our matching using balance testing:

```
with(fourelectrim,MatchBalance(maj~prop65
     + I(log(lyp)) + gastil + federal
     + col_uka + laam,
     match.out=firstmatch))
```

Let's identify the matched pair that makes the largest contribution to the estimate of ATT:

```
fourelectrim$wom_1st[firstmatch$index.treated] -
     fourelectrim$wom_1st[firstmatch$index.control]

mostimportantmatch <- which.max(
     fourelectrim$wom_1st[firstmatch$index.treated] -
     fourelectrim$wom_1st[firstmatch$index.control])


fourelectrim$country[firstmatch$index.treated[
     mostimportantmatch]]

fourelectrim$country[firstmatch$index.control[
     mostimportantmatch]]
```

Does changing the logit specification alter the matching results? The case selection? Try one or two different logit models, and compare the results.

What are the key moments in this research process where qualitative components could be added to test assumptions?

We can also use this framework as a short-cut to select most-similar cases:

```
fourelectrim$pscore[firstmatch$index.treated] -
     fourelectrim$pscore[firstmatch$index.control]

mostsimilarcases <- which.min(
     fourelectrim$pscore[firstmatch$index.treated] -
     fourelectrim$pscore[firstmatch$index.control])

fourelectrim$country[firstmatch$index.treated[
     mostsimilarcases]]

fourelectrim$country[firstmatch$index.control[
     mostsimilarcases]]
```

Alternatively, you can construct a case-selection table like the one we used in lecture, using the matching table code.R file.

Problem 3

Economic Effects of Brexit

Can we use a synthetic controls strategy to come up with a mixed-method approach to understanding the economic consequences of Britain's exit from the European Union? Provide a proposed research design complete with necessary data and estimator, explain the required assumptions, and discuss the strengths and weaknesses of your proposal. Use the qualitative information available to you from your background knowledge and research available on the internet to justify the assumptions you are making to the extent feasible.

If you have the time and can get it to work, provide results, as well. A good data source might be the QoG dataset (`https://www.gu.se/en/quality-government/qog-data/data-downloads/standard-dataset`).

```
install.packages("Synth")
library(Synth)
```

```
dataprep_out␣<-
dataprep(
foo␣=␣NAME_of_DATA,
predictors␣␣␣␣=␣c("INDEPENDENTVAR1","INDEPENDENTVAR1","ETC"),
dependent␣␣␣␣␣=␣"OUTCOME",
unit.variable␣=␣index,
time.variable␣=␣"TIMEPERIOD",

treatment.identifier␣=␣7,
#Switch␣the␣7␣to␣list␣the␣case␣in␣which␣the␣treatment␣happens

controls.identifier␣=␣unique(NAME_of_DATA$index)[-7],
#Switch␣the␣7␣to␣lost␣the␣case
#where␣the␣treatment␣happens

time.predictors.prior␣=␣1981:1990,
#Switch␣these␣years␣to␣list␣the␣time␣periods␣just
#before␣the␣treatment␣happens

time.optimize.ssr␣=␣1960:1989,
#List␣all␣the␣time␣periods␣before␣the␣treatment␣happens

unit.names.variable␣=␣"country",
#Put␣the␣variable␣storing␣names␣of␣the␣units,␣if␣any

time.plot␣=␣1960:2003
#List␣all␣the␣time␣periods␣in␣the␣whole␣dataset
)

synth_out␣<-␣synth(dataprep_out)

path.plot(synth.res␣=␣synth_out,
dataprep.res␣=␣dataprep_out,
tr.intake␣=␣1990,␣#Put␣the␣time␣period␣when␣the␣treatment␣happens
Ylab␣=␣"NAME␣OF␣THE␣OUTCOME",
Xlab␣=␣"NAME␣OF␣THE␣TREATMENT",
Legend␣=␣c("NAME␣OF␣THE␣CASE␣OF␣INTEREST",␣"Synthetic␣CASE␣NAME"),
Main␣=␣"CASE␣OF␣INTEREST␣vs␣Synthetic␣CASE␣OF␣INTEREST")

gaps.plot(synth.res␣=␣synth_out,
dataprep.res␣=␣dataprep_out,
tr.intake␣=␣1990,␣#Put␣the␣time␣period␣when␣the␣treatment␣happens
Ylab␣=␣"Effect",
Xlab␣=␣"Time␣Period",
Main␣=␣"EFFECT␣OF␣INTEREST␣in␣CASE␣OF␣INTEREST␣and␣its␣synthetic␣version")
```

Problem 4

Multi-Method Matching in Practice

For a research topic that interests you, find a published article that uses matching. Describe the qualitative evidence used in that article, if any. What untested assumptions are made to get to causal inference with this matching analysis? Design qualitative research components to test as many of those assumptions as possible; specify your design as thoroughly as possible, including data sources, what to look for in carrying out the analysis, and so forth.