

Section 5

Jensen 2003

Jensen is interested in the effect of regime type on foreign direct investment (FDI)

The design is a TSCS design using regression of 114 countries from 1970-1997.

He finds that "democratic political institutions are associated with higher levels of FDI inflows"

Is this a causal parameter? If not, what would we need to assume or do to make it a causal parameter?

How do we code effective samples?

1. We define our treatment variable and any pre-treatment controls or additional model terms
2. Run a regression of the treatment on the controls
3. Extract the residuals from that regression
4. Square the residuals.
5. Get the average of these residuals² for each country

Packages we'll need

```
library(tidyverse)
library(sf) # for mapping
library(patchwork)
```

For ease of use, we'll use the base `lm()` function to run the regressions. Why is that ok here?

Effective Weights as a function

```
effectiveWeights <- function(arg1, arg2, ...){  
  # Make the OLS formula call  
  
  # Run a regression of the treatment on the controls  
  
  # Extract the residuals from that regression  
  
  # Square the residuals.  
}
```

Effective Weights as a function

```
effectiveWeights <- function(Y, controls, data){  
  # Make the OLS formula call  
  treat_formula <- reformulate(termlabels = c(controls),  
                                response = Y)  
  
  # Run a regression of the treatment on the controls  
  treat.model <- lm(as.formula(treat_formula), data = data)  
  # Extract the residuals from that regression  
  d.tilde <- as.numeric(residuals(treat.model))  
  
  # Square the residuals.  
  weights <- d.tilde^2  
  return(weights)  
}
```

Get our Data

```
world <- st_read("world_countries_boundary_file_world_2002.shp")
```

```
## Reading layer `world_countries_boundary_file_world_2002' from data source
##   `/Users/alexstephenson/Desktop/github/BerkeleyAppliedCausalInference/course_materials/Sections/Section
##   using driver `ESRI Shapefile'
## Simple feature collection with 211 features and 9 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -180.0002 ymin: -90 xmax: 180 ymax: 83.62303
## CRS:            NA
```

```
mapnames <- read_csv("mapnames_filled2.csv")
```

```
jensenData <- read_csv("jensenData.csv")
```

Calculate effective weights

```
X.vars <- c("var5",  
           "market", "lgdppc",  
           "gdpgrowt", "tradeofg",  
           "overallb", "generalg",  
           "country", "d2", "d3")  
  
w <- effectiveWeights(Y = "regime", controls = X.vars, data = jensenData)
```


Step 5: Average of residuals² for each country

```
df <- tibble(weight = w, country = jensenData$country)

weights <- df %>%
  group_by(country)%>%
  summarise(avg = round(mean(weight),4))
```

Apply our weights to our map

```
mapWeights <- mapnames %>%  
  left_join(weights, by = c("jensen"="country"))%>%  
  na.omit()  
  
output <- world %>%  
  left_join(mapWeights, by = c("NAME"="mapname"))%>%  
  filter(NAME != "Antarctica")%>%  
  mutate(weight = if_else(is.na(avg), 0, avg),  
         expW = if_else(is.na(avg), 0, 1))
```

Make plots

```
ns <- ggplot(output)+  
  geom_sf(aes(fill = expW))+  
  scale_fill_gradient(low = "white", high = "black")+  
  theme_void()+  
  theme(legend.position = "none")+  
  ggtitle("Nominal Sample")  
  
es <- ggplot(output)+  
  geom_sf(aes(fill = weight))+  
  scale_fill_gradient(low = "white", high = "black")+  
  theme_void()+  
  theme(legend.position = "none")+  
  ggtitle("Effective Sample")
```

Put our maps on the same plot

```
plots <- ns / es
plots + plot_annotation(
  title = "The difference between Effective and Nominal Samples"
)
```

Put our maps on the same plot

The difference between Effective and Nominal Samples

Nominal Sample



Effective Sample

