Senior Thesis

Econometric implementation

Kyle Coombs (adapted from work by Marc Bellemare and Keith Head)
Bates College | EC/DCS 456

Table of Contents

Prologue

Prologue

- Today I'm going to give some tips on how to implement the econometric models you've been working on.
- How do you:
    1. Pick a model to start
    2. Wrangle the data to work
    3. Maybe transform your variables
    4. Interpret the results

Keep it simple, smartie

- The first thing to remember is that you don't need to use the most complicated model you can think of.
- Start with the simplest version of things because that is likely the easiest to code
- Most complications are just an extension of the simple model

Advice

- If you don't know how to accomplish a method with your data
- Find a case where someone has implemented the method
- Then try to replicate it
- Then move from the replication to your setup iteratively
- Yes, it is slow or tedious, but that is learning
- And in time, you'll be able to do it on your own

Example

- I have spoken to several people about using a triple difference model
- This is a great idea, but it is also a more complicated model
- It is a way to check for how a treatment effect differs for two groups
- But it is also a more complicated model
- If you cannot estimate a simple difference in differences model, you will not be able to estimate a triple difference model
- So start with DiD, then move on to 3DiD

Models

- What do you think is going on in the data?
- What is the treatment?
- What is the outcome?
- What are the potential confounders? (Things that might be correlated with the treatment and the outcome)
- Is there selection bias?
- Take a minute, discuss these wiht a partner

Can you do anything about that?

- Do the confounders fixed over time and you have panel data? Fixed effects could help!
- Are there time trends? Time fixed effects could help!
- Both? Diff-in-diff might be applicable
- Does treatment turn at a specific point in the data? A sharp regression discontinuity might be the way to go!
- Can you observe and control for all the confounders? Regression could work

Assumptions

- No method is perfect and they all rely on assumptions
- Spend two minutes with a partner and name those assumptions
- Report back to the group

Wrangling the data

- Most of the time, the data you have is not in the format you need
- I don't just mean that you need to get it into a dataframe or merge with another dataset
- I mean you need to recode variables to work in your model

Difference in difference

- Diff-in-diff relies on a treatment and control group and a before and after treatment period
- If you are looking at a change to the EITC on women's labor supply, then you know treatment effects moms after the reform
- I'll mostly present this today cause many of you are using a form of this

$$ y{it} = \alpha_i + \delta_t + \beta \delta_t \times Treated_i + \varepsilon{it} $$

Look at the data

Is this data in the right form?

```
od ← causaldata::organ_donations
od %>% head(5)
```

```
## # A tibble: 5 × 4
##   State  Quarter  Rate Quarter_Num
##   <chr>  <chr>   <dbl>       <int>
## 1 Alaska Q42010   0.75           1
## 2 Alaska Q12011   0.77           2
## 3 Alaska Q22011   0.77           3
## 4 Alaska Q32011   0.78           4
## 5 Alaska Q42011   0.78           5
```

Treatment variable

```
od ← od %>%
     mutate(Treated = State == 'California' &
            Quarter %in% c('Q32011','Q42011','Q12012'))
od
```

```
## # A tibble: 162 × 5
##    State   Quarter  Rate Quarter_Num Treated
##    <chr>   <chr>   <dbl>       <int> <lgl>
##  1 Alaska  Q42010  0.75            1 FALSE
##  2 Alaska  Q12011  0.77            2 FALSE
##  3 Alaska  Q22011  0.77            3 FALSE
##  4 Alaska  Q32011  0.78            4 FALSE
##  5 Alaska  Q42011  0.78            5 FALSE
##  6 Alaska  Q12012  0.79            6 FALSE
##  7 Arizona Q42010  0.263           1 FALSE
##  8 Arizona Q12011  0.209           2 FALSE
##  9 Arizona Q22011  0.226           3 FALSE
## 10 Arizona Q32011  0.250           4 FALSE
## # ℹ 152 more rows
```

Doing it in R

```
# feols clusters by the first
# fixed effect by default, no adjustment necessary
clfe ← feols(Rate ~ Treated | State + Quarter,
           data = od)
etable(clfe)
```

```
##                               clfe
## Dependent Var.:               Rate
##
## TreatedTRUE     -0.0225** (0.0061)
## Fixed-Effects:  ------------------
## State                          Yes
## Quarter                        Yes
## _______________ __________________
## S.E.: Clustered          by: State
## Observations                   162
## R2                         0.97932
## Within R2                  0.00922
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Make it pretty in latex

```
etable(clfe,tex=TRUE)
```

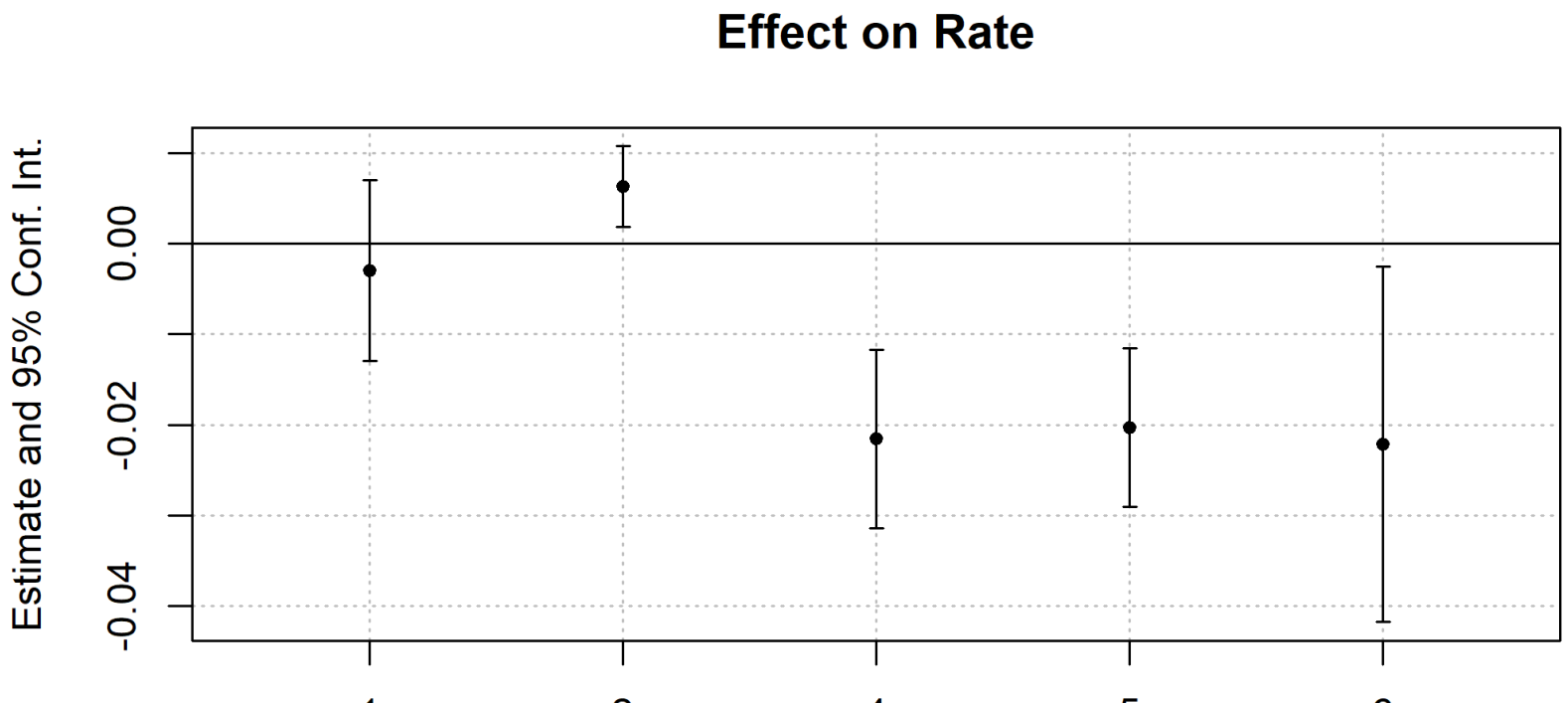
\begingroup \centering \begin{tabular}{lc} \tabularnewline \midrule \midrule Dependent Variable: & Rate\
Model: & (1)\
\midrule \emph{Variables}\ TreatedTRUE & -0.0225$^{}$\
& (0.0061)
\midrule \emph{Fixed-effects}\ State & Yes
Quarter & Yes
\midrule \emph{Fit statistics}\ Observations & 162
R^2 & 0.97932
Within R^2 & 0.00922
\midrule \midrule \multicolumn{2}{l}{\emph{Clustered (State) standard-errors in parentheses}}
\multicolumn{2}{l}{\emph{Signif. Codes: : 0.01, *: 0.05, : 0.1}}\ \end{tabular} \par\endgroup

Assess pre-trends

Remember parallel trends? Let's check for parallel pre-trends

```
# Interact quarter with being in the treated group using
# the fixest i() function, which also lets us specify
# a reference period (using the numeric version of Quarter)
clfe ← feols(Rate ~ i(Quarter_Num, California, ref = 3) |
            State + Quarter_Num, data = od)

# And use coefplot() for a graph of effects
coefplot(clfe)
```



Not perfect

- The pre-trends above are not perfect
- They're short!
- That was a quick example I could get on hand that worked
- Ideally you have tons of pre-periods!

Triple differences

- Generally, you might have some diff-in-diff relationship
- But you might also have a third grouping that varies between the treated/untreated
- For example, within California some people may need organs and others may not
- You can use a triple difference to see how the treatment effect in California varies between these groups
- If you have the right data

Transform variables

- Sometimes you need to transform your variables to make them work in your model
- For example, if you a variable with a long tail, you might want to take the log
    - If you regress a log on a log, you also get a percentage change, basically an elasticity
- Alternatively, you may want to normalize your data to make it easier to interpret
    - Subtract the mean and divide by the standard deviation
    - Then the coefficient is the number of standard deviations the outcome changes for a one unit change in the treatment
- Sometimes you need to create a new variable
    - For example, if you have a treatment that is a dummy variable, you might want to create a variable that is the interaction of the treatment and a continuous variable
- You can and should do these things
- I'll try to show you live now

Next classes: One-on-ones and Student proposal presentation