Methods in Causal Inference Course Assignment

Questions assigned by: Prof. Vidhya Soundararajan
Solutions submitted by: Ashwin Iyenggar (1521001)

January 24, 2017

Question 1 : DiNardo and Pischke (1997)

1a. Find the difference in the mean log wage of computer users and non-computer users using the 1979 data

---

pre code answer

post code answer

1b. What are the assumptions under which this difference would be an unbiased estimate of the effect of computers on wages?

pre code answer

post code answer

1c. Suppose a randomized trial was used to generate this data. This trial assigned some individuals computers and others not. How would you test if the randomization worded? Do a battery of tests to check this using 1979 data.

pre code answer

post code answer

1d. Examining Krueger (1993) table 2, how does the effect of computer usage on wages change when the regressions include nothing else (column 1 and 4) and when the regressions include years of education and other variables (columns 2 and 3, 5 and 6). What does this change tell you about the relation between the schooling of a worker and his or her computer use on the job? which computer coefficient is likely to be a better measure of the causal effect of computer skills?

pre code answer

post code answer

1e. Using the 1985 data in d_p.dta, run a regression similar to Krueger (1993) table 2 (columns 1 and 2 only). Some variables Krueger (1993) uses do not exist in the German data like race or veteran status. Ignore those variables unavailable. Show regression results in a table. Compare your results for Germany to the US results and comment on them.

---

pre code answer

post code answer

1g. How can you perform a formal test of the hypothesis that the association between computer use and wages is the same in 1979 and 1985? clearly state your null and alternate hypothesis. Perform your test in Stata. What is the result of your test?

pre code answer

post code answer

1h (I). Repeat your regressions from 1e for the 1979 data, but additionally include occupation fixed effects. Since there are numerous occupation categories and you may not want ot see a coefficient specific to each occupation, you can use the areg command (absorb option) to see a simplified output. Show regressions in a table. Type "help areg" or "man areg" to obtain help.

pre code answer

post code answer

1h (II). German and math scores, referring to grades in German and Mathematics classes are indicators of ability. The variable father contains occupation code for the father. Use these three as additional variables in the regression for 1979 data, and show results in the same table as you produced for 1h (I), but in anew column. These variables only have few categories, so we can include separate dummies for each of these categories in the regression (use the i. operator).

pre code answer

post code answer

1h (III). Other variables indicating what people do or use on their job are also available: whether they use a calculator or cash register (calc), a telephone (teleph), writing materials (the variable is called pencil), whether they work mostly while sitting (sit), or whether they use blue collar tools like a hammer, screw driver, drill (hammer) etc. Run separate regressions for 1979 and for 1985 where you include these variables one by one in the equation instead of the computer variables. Include again occupation dummies, grades, and family background variables into all of these regressions. Prepare two tables for 1979 and 1985 each. How do the results compare to the computer effects? Do you think these coefficients are associated with the causal effect of a particular skill on earnings? How do these results make you feel about the interpretation of the computer coefficient? How are the non-computer tools affected by this?

pre code answer

post code answer

1h (IV). Write up a conclusion based on your results. What have you learned about the causal effect of computers on wages?

pre code answer

post code answer

Question 2 : Autor (2003)

One of the most common research designs encountered in public policy research is the one with many treated and control groups and many time periods. In a federalist system such as India or the USA, a common situation is for a group of states to pass a policy of interest while another of states either do not pass the policy or adopt the policy only with a lag relative to the other states. One such policy in the United States is the exception to employment at will. Originating in the 19th century, employment at will doctrine states that US firms can fire workers for a just cause or a no cause, at employer's whim or will, and without warning. But some state level courts in the US have allowed exceptions to the employment at will doctrine leading to law suits for unjust dismissal. Autor (2003) was interested in wheter this fear of law suits made firms hire temporary workers through an intermediary agency. Hiring temp workers would make firing costless and simpler because the firm is not directly employing the services of workers and cannot be implicated because of stoppage of their services.

In the question below, we will examine some interesting aspects of this paper and understand the multi-periods and multi-group difference-in-difference estimators (more generally fixed effects estimator). In the paper, there are three policy variables indicating if there was a court ruling providing an exception to employment at will doctrine, in each state 's' in each year 't'. y_{st}, the outcome variable is the log of temporary workers in state 's' and year 't'.

2a. Let's begin by understanding the nature of the date and three policy variables. Open autotempworkers.dta, the dataset used in the study. Use the `desc` **command to examine the variables in the dataset. use the following commands to take a peek at how the data is set up.**

```
sort state year
list state abbrev year mico mppa mgfa thsemp, sepby(state)
tab1 mico mppa mgfa
```

Describe very briefly in words how the three key policy variables (mico, mppa, and mgfa) are coded (that is, what does 0 mean and 1 mean for each). Please skim through section I of the paper to intuitively understand the meaning of the three policies namely the implicit contract exception (mico), the public policy exception (mppa), the implied covenant of good faith and fair dealing (mgfa). Appendix Table A1 has details on the court cases used for analysis. You do not have to understand the details of any of these to do this assignment, but just understand the idea of what the policy does.

`mico` is coded 1 for every year from the first year in which an implicit contract exception was adopted in a given state, and coded 0 for all years prior to which the first implicit contract exception was adopted.

`mppa` is similarly coded 0 for all years prior to which a public policy exception was first adopted in a state, and coded 1 for all years from the year in which a public policy exception was adopted in a state for the first time.

`mgfa`, on similar lines is coded 0 for all years prior to which the implied covenant of good faith and fair dealing was first adopted, and coded 1 for all years since (including the year of adoption)

```
desc
sort state year
list state abbrev year mico mppa mgfa thsemp, sepby(state)
tab1 mico mppa mgfa
```

2b. Run a regression of the lob of temporary help employment (log thsemp) on all the three policy variables and a constant term, using the standard 'robust ' option to get robust standard errors. Taking your results at face value, what would you conclude about the impact of these three types of policies on temporary help employment? Why would this be a biased estimator for the true causal effects of these policies?

Table 1 presents the results of this basic regression. If one were to take the results in Table 1 at face value (as suggested by the question), one would conclude that both `mico` and `mppa` have a positive and significant effect on the log of temporary employment. The size of the effect on total temporary employment indicated is 42% and 57% for the adoption implicit contract exception and public policy exception respectively. The results in Table 1 also suggests that the effect of adoption of the implied covenant of good faith and fair dealing (`mgfa`) has a negative and statistically insignificant effect on ln (temporary workers).

The estimates in Table 1 may be biased because state level and year level effects are not controlled for. Specifically it is possible that different states have different starting dispositions that might affect the level of temporary workers. (»>endogeniety...)

```
gen lnthsemp=ln(thsemp)
reg lnthsemp mico mppa mgfa, robust
outreg2 using 'imagepath'2b.tex, title("Basic Regression\label{2b}")
   ctitle("ln(temporary employment)") tex(pretty frag) dec(4)
   replace
```

Table 1:

Basic Regression

VARIABLES	(1) ln(temporary employment)
mico	0.4210***
	(0.1333)
mppa	0.5706***
	(0.1350)
mgfa	-0.1985
	(0.1668)
Constant	8.2248***
	(0.1127)
Observations	850
R^2	0.0548

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

2c. Run the same regression but compute 'cluster robust ' standard errors, using the option 'cluster(state) ' which allows for arbitrary correlation of the error term across years within the same state. Compare the standard errors you get from this approach relative to the one above. Why do you think your estimate of the precision of your estimates changed so much? Use cluster robust standard errors for the remainder of this question.

Table 2 presents the results of this basic regression with clustered standard errors. We learn from theory that standard errors of coefficient estimates are crucial in determining the statistical significance of the result, and in determining the outcomes of the hypothesis testing. In simpler words, the standard error determines how accurate the estimates are. In observing the standard error of the estimates in Table 2, we observe that the standard error values are higher than those in Table 1. The higher standard error values in Table 2 suggest that the effects of the three policy variables are all not significant. The reason for the significant differences in standard errors is that it may have been quite likely that observations from different states had different variances in the values of temporary workers - violating the iid (identically independently distributed) assumption. While robust standard errors control for heteroskedasticity in the distribution of log temporary workers in the sample as a whole , adjusting for clustering based on state allows for all observations within a state to be clustered so that similar effects that affect a given state may be considered state by state. Having adjusted for state level effects, the aggregate standard errors are found to be higher. This may imply that the distribution across states (when state effect is not clustered) may have been more uniform, but that there was higher variance

Table 2: Regression Results

Regression with Robust Cluster SE

VARIABLES	(1) ln(temporary employment)
mico	0.4210
	(0.3491)
mppa	0.5706
	(0.3877)
mgfa	-0.1985
	(0.5737)
Constant	8.2248***
	(0.3368)
Observations	850
R^2	0.0548

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

within states. When clustered standard effects were applied, the independently distributed assumption was applied only across states and not within it, and we therefore received a higher standard error indicating the higher within state variance in log temporary workers.

```
reg lnthsemp mico mppa mgfa, robust cluster(state)
outreg2 using 'imagepath'2c.tex, title("Regression with Robust
   Cluster SE\label{2c}") ctitle("ln(temporary employment)") tex(
   pretty frag) dec(4) replace
```

2d. Run two regressions with the same variables as b and c, but include in addition: (i) year dummy variables only, and (ii) state dummy variables only. Explain what sources of bias are being controlled for in each regression i and ii.

Table 3 presents the results for the two models including the year dummies and state dummies respectively. The model with year dummies controls for bias in estimates due to any events that affected all observations in a certain year but not others. The model with state dummies controls for bias in estimates due to any state specific effects that does not affect observations in other states. The resulting estimates are expected to be controlled for bias from such year or state effects, and therefore a truer estimation of the real effect of the policy variables.

Table 3: Regression with Year and State Controls

	(1) lnthsemp	(2) lnthsemp
mico	-0.117	0.812***
	(0.369)	(0.125)
mppa	0.192	0.808***
	(0.380)	(0.138)
mgfa	-0.298	0.545**
	(0.611)	(0.199)
_cons	7.917***	6.651***
	(0.257)	(0.125)
N	850	850
Model	Year Dummies	State Dummies

Standard errors in parentheses

Reference Year is 79 for Year Dummies Model

Reference State is 11 (ME) for State Dummies Model

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
levelsof state, local(lstate)
foreach ls of local lstate {
      gen dstate'ls' = 1 if state=='ls'
      replace dstate'ls' = 0 if missing(dstate'ls')
}

levelsof year, local(lyear)
foreach ly of local lyear {
      gen dyear'ly' = 1 if year=='ly'
      replace dyear'ly' = 0 if missing(dyear'ly')
}
```

```
eststo clear
reg lnthsemp mico mppa mgfa dyear80-dyear95, robust cluster(state)
estadd local Model "Year Dummies"
est store model1
reg lnthsemp mico mppa mgfa dstate12-dstate95, robust cluster(state)
estadd local Model "State Dummies"
est store model2
esttab model1 model2 using 'imagepath'2d.tex, title("Regression with
    Year and State Controls\label{2d}") longtable se not drop(d*)
    scalars("Model") addn("Reference Year is 79 for Year Dummies
    Model" "Reference State is 11 (ME) for State Dummies Model")
    replace
```

2e. Replicate columns (7) and (8) of table 3 in the Autor (2003). Describe an omitted factor that might lead to bias in the specification in column (7) but would not in column (8) i.e., what sources of bias does adding state speciific trends to the model control for? Hint: to create variables for state specific time trends, crate a variable gen t=year-78, and then use a command like xi i.state i.year i.state*t. You might want to use the 'list ' command to look at the data to be sure you understand what this variable looks like.

Table 4 replicates the results in columns (7) and (8) of Table 3 in Autor (2003).

The specification without controlling for state * time trend (column 1 in Table 4) could potentially be biased due to missing the effect of time within the state. While state dummies controlled for state level effects, and year dummies controlled for year effects, there may have still been a time varying state effect which would not have been previously captured. By including the state * time trend variables, the effects of time within a state are controlled for. This leaves us with a much better estimate for the causal effect of the policy intervention on lag wages.

```
gen age=year-78
foreach var of varlist dstate* {
  gen tt`var' = `var'*age
}
local imagepath /Users/aiyenggar/OneDrive/code/articles/mci-
   assignment-images/
reg lnthsemp mico mppa mgfa dyear80-dyear95 dstate12-dstate95, robust
    cluster(state)
outreg2 using `imagepath'2e.tex, drop (dyear* dstate* tt*) tex(pretty
    frag) dec(3) addtext(State and year dummies, Yes, State * time
   trends, No) addn("Reference Year is 79, Reference State is 11 (ME
   )") label replace

reg lnthsemp mico mppa mgfa dyear80-dyear95 dstate12-dstate95
   ttdstate12-ttdstate95, robust cluster(state)
outreg2 using `imagepath'2e.tex, drop (dyear* dstate* ttdstate*) tex(
   pretty frag) dec(3) addtext(State and year dummies, Yes, State *
   time trends, Yes) label append
```

Table 4: The Estimated Impact of Common Law Exceptions to Employment at Will on THS Employment, 1979-95

VARIABLES	(1) Log(Temporary Employment)	(2) Log(Temporary Employment)
Implied contract	0.096	0.137**
	(0.099)	(0.062)
Public policy	0.126	-0.023
	(0.094)	(0.058)
Good faith	0.100	-0.079
	(0.113)	(0.093)
Constant	6.596***	6.117***
	(0.099)	(0.068)
Observations	850	850
R^2	0.969	0.988
State and year dummies	Yes	Yes
State * time trends	No	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Reference Year is 79, Reference State is 11 (ME)

2f. Replicate columns 1 in table 5. Explain the purpose of columns (1) through (4) in table 5, which includes log non-farm employment (log anneemp) as an additional explanatory variable. What is the author trying to convince you of, by showing that the results are robust with the addition of non-farm employment as an additional variable?

Table 5 replicates the results in columns (1) and (2) of table 5 in Autor (2003). By adding a control for the total non-farm employment, the author demonstrates that his results are robust to the trends in the broader job market. In other words, having controlled for the total number of new jobs created the effects of the policy measures on temporary employment may no longer be assumed to have been due just to the effect of the larger job market.

```
gen lnannemp = ln(annemp)
label variable lnannemp "Log(Non-Farm Employment)"

foreach var of varlist dstate* {
  gen tt2'var' = 'var'*age*age
}
local imagepath /Users/aiyenggar/OneDrive/code/articles/mci-
   assignment-images/
reg lnthsemp mico lnannemp mppa mgfa dyear80-dyear95 dstate12-
   dstate95 ttdstate12-ttdstate95, robust cluster(state)
outreg2 using 'imagepath'2f.tex, drop (mppa mgfa dyear* dstate* tt*)
   tex(pretty frag) dec(3) addtext(State and year dummies, Yes,
   State * time trends, Yes, State * time2 trends, No) addn("
   Reference Year is 79, Reference State is 11 (ME)") label replace

reg lnthsemp mico lnannemp mppa mgfa dyear80-dyear95 dstate12-
   dstate95 ttdstate12-ttdstate95 tt2dstate12-tt2dstate95, robust
   cluster(state)
outreg2 using 'imagepath'2f.tex, drop (mppa mgfa dyear* dstate* tt*)
   tex(pretty frag) dec(3) addtext(State and year dummies, Yes,
   State * time trends, Yes, State * time2 trends, Yes) label append
```

Table 5: The Estimated Impact of Common Law Exceptions to Employment at Will on THS Employment, 1979-95, Controlling for State Demographics

VARIABLES	(1) Log(Temporary Employment)	(2) Log(Temporary Employment)
Implied contract	0.148** (0.057)	0.132** (0.063)
Log(Non-Farm Employment)	1.552*** (0.427)	1.588** (0.642)
Constant	-14.012** (5.517)	-14.480* (8.277)
Observations	850	850
R^2	0.989	0.990
Public policy dummy	Yes	Yes
Good faith dummy	Yes	Yes
State and year dummies	Yes	Yes
State * time trends	Yes	Yes
State * time2 trends	No	Yes

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Reference Year is 79, Reference State is 11 (ME)

2g. In table 5, what is the purpose of including region by year dummy variables? Why doesn't the author include state by year dummies?

Table 5 in Autor (2003) is intended at demonstrating the effects of implied contract exception to employment at will to temporary employment numbers, controlling for demographics. One reason for including region by year dummy variables for such may be that the demographic data may have either been collected at a regional level, or because several states in a region may share similar demographic patterns.

2h. Is column 7 in table 5 devoid of bias of all kinds? Why or why not?

It may be very hard to argue against the results in column 7 of table 5 in Autor (2003) as the author has controlled for region, year, state and time trend effects.

2i. Replicate column (2) of Table 7 using the leads and lags variables already created for you in the data set (e.g., admico 2 is a two year lead of the implicit contract exception policy whereas admico2 is a two year lag; and mico4 is a dummy that switches on 4 years or more after policy, etc.). The main takeaway from this table is that future policy variables are not significantly affecting current temporary help employment but only that the current and lagged (past) policy variables are affecting current temporary employment. Why was it important for the author to show this, and how does this result strengthen the claim of causality in the paper?

Table 6 replicates the results in column (2) of table 7 in Autor (2003).

```
local imagepath /Users/aiyenggar/OneDrive/code/articles/mci-
    assignment-images/
label variable admico_2 "Law change t+2"
label variable admico_1 "Law change t+1"
label variable admico0 "Law change t0"
label variable admico1 "Law change t-1"
label variable admico2 "Law change t-2"
label variable admico3 "Law change t-3"
label variable mico4 "Implied contract law t-4 forward"
reg lnthsemp lnannemp admico* mico4 admppa* mppa4 admgfa* mgfa4
    dyear80-dyear95 dstate12-dstate95 ttdstate12-ttdstate95, robust
    cluster(state)
outreg2 using 'imagepath'2i.tex, keep(admico_2 admico_1 admico0
    admico1 admico2 admico3 mico4) tex(pretty frag) dec(3) addtext(
    State and year dummies, Yes, State * time trends, Yes, State *
    time2 trends, No) addn("Reference Year is 79, Reference State is
    11 (ME)") nocon label replace
```

Table 6: The Estimated Impact of Common Law Exceptions to Employment at Will on THS Employment, 1979-95, Controlling for State Demographics

VARIABLES	(1) Log(Temporary Employment)
Law change t+2	-0.017
	(0.052)
Law change t+1	-0.001
	(0.058)
Law change t0	0.108
	(0.080)
Law change t-1	0.147*
	(0.085)
Law change t-2	0.228**
	(0.104)
Law change t-3	0.144
	(0.107)
Implied contract law t-4 forward	0.196
	(0.125)
Observations	850
R^2	0.989
State and year dummies	Yes
State * time trends	Yes
State * time2 trends	No

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Reference Year is 79, Reference State is 11 (ME)

References

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21(1):1–42.

DiNardo, J. E. and Pischke, J.-S. (1997). The returns to computer use revisited: Have pencils changed the wage structure too? *The Quarterly Journal of Economics*, 112(1):291–303.

Krueger, A. B. (1993). How computers have changed the wage structure: Evidence from microdata, 1984-1989. *The Quarterly Journal of Economics*, 108(1):33–60.