

Methods in Causal Inference Course Assignment

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

January 24, 2017

Question 1 : [DiNardo and Pischke \(1997\)](#)

Does being able to work with a computer raise productivity and earnings? In an important paper, [Krueger \(1993\)](#) included a measure indicating whether a worker uses computers at work in standard wage regressions. The paper's Table II (page 38) contains the main results. The paper found a considerable wage differential for using computers as opposed to not using computers. This has been taken as evidence in favour of the view that introduction of new technologies improve productivity and wages. This question asks you to evaluate this claim. Open `d_p.dta`, the dataset used in another study ([DiNardo and Pischke, 1997](#)). This is based on pooled cross sections of the West German Qualification and Career Survey. Use `desc` and `tab` commands to understand the variables.

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

Table 1 presents the results of a t-test on 1979 log wages between computer and non-computer users. We note that the mean log wages for computer users is 2.553225 and that for non computer users is 2.419198, yielding a difference = 0.1340268. This difference may also be obtained by regressing log wages on computer usage for the year 1979 as determined in Table 2. The instructions specified in the question have been executed in the following Stata code.

```
estpost ttest lnw if year==1979, by(computer)
esttab using 'imagepath'1a.tex, title("T-test of log wages in 1979 by computer
      usage\label{1a}") mtitle("Mean Difference") se replace

reg lnw computer if year==1979
outreg2 using 'imagepath'1a1.tex, title("Regression of log wages in 1979 on
      computer usage\label{1a1}") ctitle("Log Hourly Wage") tex(pretty frag)
      replace
```

Table 1: T-test of log wages in 1979 by computer usage

	(1)
	Mean Difference
Log (Wage)	-0.134*** (0.0170)
Observations	8905

Standard errors in parentheses

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

Table 2:

Regression of log wages in 1979 on computer usage

VARIABLES	(1) Log (Wage)
Computer User (Dummy)	0.134*** (0.0170)
Constant	2.419*** (0.00486)
Observations	8,905
R^2	0.007

Standard errors in parentheses

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

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

We would need random assignment of computer and non computer users in the sample for the above to provide us with an unbiased estimate of the effect of computer usage on log wages. In other words we would require that none of the priors to assignment of computer usage or non-usage must result in a significant difference in average wages. Only if this were true would the coefficient on computer when regressing log wages on computer give us an unbiased estimate of the effect of computer usage on log wages.

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 worked? Do a battery of tests to check this using 1979 data.

Table 3 presents the results of a t-test of the all other variables by computer usage for the 1979 data. For randomization to have worked, we would have needed that none of the mean differences were statistically significant. However it turns out from Table 3 that almost all of the other variables demonstrate statistically significant differences between computer users and non-computer users. We may therefore conclude that computer usage is not randomly assigned in the 1979 data.

The instructions specified in the question have been executed in the following Stata code.

```
estpost ttest german math school exp female married lnw sit pencil teleph calc
      hammer city civser father occ, by(computer)
esttab using 'imagepath'1c.tex, title("T-tests by computer usage\label{1c}")
      mtitle("Mean Difference") replace
```

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?

Table 2 of Krueger (1993) demonstrates across both the October 1984 and October 1989 samples that when years of education, experience, and other variables are added to the model, the effect of computer usage on log wages decreases. The drop in the coefficient estimate indicates that there may be unmeasured variables missing from our model, so that when other variables are added it is really the effect of the missing variable that is causing some of the effect to be taken away from the computer usage variable. Models (3) and (6) are likely to be a better measure of the causal effect of computer skills on log wages since it accounts for many more variables. However a true causal measure many still remain elusive in the absence of either random assignment or a natural experiment.

Table 3: T-tests by computer usage

	(1) Mean Difference
German Grade	-0.250*** (0.0326)
Math Grade	-0.303*** (0.0348)
Years of Schooling	-0.839*** (0.0811)
Experience	3.287*** (0.447)
Female (Dummy)	-0.0751*** (0.0171)
Married (Dummy)	0.0413* (0.0180)
Log (Wage)	-0.134*** (0.0170)
Sedantary Worker (Dummy)	-0.291*** (0.0173)
Pencil User (Dummy)	-0.252*** (0.0191)
Telephone User (Dummy)	-0.294*** (0.0187)
Calculator User (Dummy)	-0.112*** (0.0150)
Physical Tools User (Dummy)	0.284*** (0.0175)
1 = lives in a city	-0.0714*** (0.0181)
1 = civil servant	-0.0178 (0.0127)
Father Occupation	0.538 (0.564)
Occupation	-1192.0*** (88.28)
Observations	8905

Standard errors in parentheses

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

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.

Column (1) of Table 4 displays the OLS regression estimates of the effect of computer use on log wages for 1985 in the German data. The coefficient estimates and standard errors for the coefficients in Table 4 are strikingly similar to those in Table 2 in Krueger (1993). For example, the estimate is 0.155 in Table 4 while it is 0.170 in Table 2 in Krueger (1993). The signs on the coefficient estimate for the square of experience are both negative. Similarly for the coefficients on female, on married and on the interaction between female and married. Finally, the R^2 value in Krueger (1993) Table 2 is 0.446 while it is 0.308 in Table 4. In summary, we may conclude that the coefficient estimates on the 1985 German data is quite similar to that on the US data in 1984.

The instructions specified in the question have been executed in the following Stata code.

```
gen exp2=(exp^2)/100
gen femmar=female*married
reg lnw computer school exp exp2 female married femmar if year==1985
outreg2 using 'imagepath'1e.tex, title("The effect of computer use on log
wages\label{1e}") ctitle("Log Hourly Wage") nocons tex(pretty frag) replace
```

1f. Use the entire dataset d_p.dta - which contains a pooled cross-section from 1979 and 1985. Run the same regression as in 1e, but now include a dummy variable for year 1985 in addition to the other covariates and an interaction term between computer use and your dummy variable for 1985. Add your results in a new column to the table you prepared for 1e

The results of this regression have been added to Table 4 in column (2). We note that the coefficient of the dummy variable d1985 is positive and significant, suggesting therefore that there may be a year effect on wages in addition to those of computer usage, years in school and experience.

The instructions specified in the question have been executed in the following Stata code.

```
gen d1985 = 1 if year==1985
replace d1985 = 0 if missing(d1985)
gen computer1985 = computer * d1985
reg lnw computer school exp exp2 female married femmar d1985 computer1985
outreg2 using 'imagepath'1e.tex, title("The effect of computer use on log
wages\label{1e}") ctitle("Log Hourly Wage") nocons tex(pretty frag) append
```

Table 4:

The effect of computer use on log wages

VARIABLES	(1) Log (Wage)	(2) Log (Wage)
Computer User (Dummy)	0.155*** (0.00990)	0.124*** (0.0145)
Years of Schooling	0.0637*** (0.00161)	0.0682*** (0.00124)
Experience	0.0363*** (0.00145)	0.0326*** (0.00103)
Experience Squared	-0.0603*** (0.00301)	-0.0549*** (0.00217)
Female (Dummy)	-0.104*** (0.0136)	-0.108*** (0.00981)
Married (Dummy)	0.114*** (0.0111)	0.112*** (0.00825)
Female * Married (Dummy)	-0.127*** (0.0177)	-0.130*** (0.0130)
1985 (Dummy)		0.208*** (0.00592)
Computer User * 1985 (Dummy)		0.0219 (0.0174)
Observations	9,733	18,638
R^2	0.308	0.343

Standard errors in parentheses

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

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?

We can use the regression results in column (2) of Table 4 to perform a formal test of hypothesis of the association between computer use and wage between 1979 and 1985. Formally, consider the following model:

$$\ln w = \beta_0 + \beta_1 * \text{computer} + \beta_2 * \text{exp} + \beta_3 * \text{exp}^2 + \beta_4 * \text{female} + \beta_5 * \text{married} + \beta_6 * \text{female} * \text{married} + \beta_7 * d1985 + \beta_8 * \text{computer} * d1985$$

Null Hypothesis $H_0 : \beta_8 = 0$

Alternate Hypothesis $H_a : \beta_8 \neq 0$

From Table 4 we fail to reject H_0 since the coefficient estimate $\hat{\beta}_8$ is not statistically significant at even the 10% significance level (the t statistic value is 1.25).

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 to 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.

Column (1) of Table 5 depicts the results for 1979 effects of computer usage after controlling for occupation effects.

The instructions specified in the question have been executed in the following Stata code.

```
areg lnw computer school exp exp2 female married femmar if year==1979, absorb(
    occ)
outreg2 using 'imagepath'1hI.tex, nocons tex(pretty frag) addtext(Occupation
    dummies, Yes, German and Math Score Dummies, No, Father Occupation Dummies,
    No) label replace
```

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 a new 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).

Column (2) of Table 5 displays the results of this regression.

The instructions specified in the question have been executed in the following Stata code.

```
xi: areg lnw computer school exp exp2 female married femmar i.german i.math i.
    father if year==1979, absorb(occ)
outreg2 using 'imagepath'1hI.tex, drop(*) nocons tex(pretty frag) addtext(
    Occupation dummies, Yes, German and Math Score Dummies, Yes, Father
    Occupation Dummies, Yes) label append
```


Table 5: 1979 effects of computer usage on log wages, with occupation controls

VARIABLES	(1) Log (Wage)	(2) Log (Wage)
Computer User (Dummy)	0.0292* (0.0150)	0.0256* (0.0150)
Years of Schooling	0.0545*** (0.00282)	0.0501*** (0.00291)
Experience	0.0292*** (0.00144)	0.0291*** (0.00144)
Experience Squared	-0.0495*** (0.00305)	-0.0498*** (0.00305)
Female (Dummy)	-0.128*** (0.0153)	-0.129*** (0.0154)
Married (Dummy)	0.101*** (0.0118)	0.0991*** (0.0118)
Female * Married (Dummy)	-0.101*** (0.0184)	-0.102*** (0.0183)
Observations	8,905	8,905
R^2	0.400	0.405
Occupation dummies	Yes	Yes
German and Math Score Dummies	No	Yes
Father Occupation Dummies	No	Yes

Standard errors in parentheses

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

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?

The results for the 1979 data is provided in Table 6. However the same model cannot be estimated for the 1985 data since the 1985 has missing values for father, german, math. This causes Stata to drop all observations. In Table 7, I have provided the results by excluding the german and math scores and the father occupational dummies.

Table 6: 1979 effects of tools used on log wages, with occupation controls

VARIABLES	(1) Log (Wage)	(2) Log (Wage)	(3) Log (Wage)	(4) Log (Wage)	(5) Log (Wage)
Calculator User (Dummy)	0.0350*** (0.0114)				
Years of Schooling	0.0503*** (0.00290)	0.0487*** (0.00290)	0.0492*** (0.00290)	0.0498*** (0.00290)	0.0499*** (0.00291)
Experience	0.0291*** (0.00144)	0.0290*** (0.00144)	0.0292*** (0.00144)	0.0290*** (0.00144)	0.0291*** (0.00144)
Experience Squared	-0.0496*** (0.00305)	-0.0495*** (0.00304)	-0.0500*** (0.00304)	-0.0496*** (0.00305)	-0.0497*** (0.00305)
Female (Dummy)	-0.130*** (0.0154)	-0.131*** (0.0153)	-0.131*** (0.0153)	-0.134*** (0.0154)	-0.130*** (0.0154)
Married (Dummy)	0.0989*** (0.0118)	0.0965*** (0.0118)	0.0965*** (0.0118)	0.0989*** (0.0118)	0.0990*** (0.0118)
Female * Married (Dummy)	-0.102*** (0.0183)	-0.0973*** (0.0183)	-0.0992*** (0.0183)	-0.101*** (0.0183)	-0.102*** (0.0183)
Telephone User (Dummy)		0.0722*** (0.0103)			
Pencil User (Dummy)			0.0600*** (0.0101)		
Sedantary Worker (Dummy)				0.0452*** (0.0108)	
Physical Tools User (Dummy)					-0.0347*** (0.0120)
Observations	8,905	8,905	8,905	8,905	8,905
R^2	0.405	0.408	0.407	0.406	0.405
Occupation dummies	Yes	Yes	Yes	Yes	Yes
German and Math Score Dummies	Yes	Yes	Yes	Yes	Yes
Father Occupation Dummies	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

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

Table 7: 1985 effects of tools used on log wages, with occupation controls

VARIABLES	(1) Log (Wage)	(2) Log (Wage)	(3) Log (Wage)	(4) Log (Wage)	(5) Log (Wage)
Calculator User (Dummy)	0.0604*** (0.00970)				
Years of Schooling	0.0395*** (0.00240)	0.0391*** (0.00240)	0.0393*** (0.00240)	0.0409*** (0.00239)	0.0410*** (0.00239)
Experience	0.0330*** (0.00145)	0.0328*** (0.00145)	0.0330*** (0.00145)	0.0331*** (0.00145)	0.0332*** (0.00145)
Experience Squared	-0.0540*** (0.00300)	-0.0540*** (0.00299)	-0.0541*** (0.00300)	-0.0543*** (0.00300)	-0.0545*** (0.00300)
Female (Dummy)	-0.0956*** (0.0148)	-0.100*** (0.0148)	-0.102*** (0.0148)	-0.0994*** (0.0148)	-0.0996*** (0.0148)
Married (Dummy)	0.104*** (0.0110)	0.105*** (0.0110)	0.104*** (0.0110)	0.107*** (0.0110)	0.106*** (0.0110)
Female * Married (Dummy)	-0.111*** (0.0173)	-0.110*** (0.0173)	-0.108*** (0.0173)	-0.114*** (0.0174)	-0.113*** (0.0174)
Telephone User (Dummy)		0.0697*** (0.00942)			
Pencil User (Dummy)			0.0673*** (0.00978)		
Sedantary Worker (Dummy)				0.0348*** (0.0113)	
Physical Tools User (Dummy)					-0.0238** (0.0112)
Observations	9,733	9,733	9,733	9,733	9,733
R^2	0.413	0.414	0.413	0.411	0.410
Occupation dummies	Yes	Yes	Yes	Yes	Yes
German and Math Score Dummies	No	No	No	No	No
Father Occupation Dummies	No	No	No	No	No

Standard errors in parentheses

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

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

From the evidence from Table 6 and Table 7 it is clear that our current identification strategy is unable to yield a causal inference of the role of computer usage on wages since if we were to accept that we must also accept the rather absurd conclusion that using writing instruments such as pencils also lead to a higher wage. This conundrum is clarified by an understanding that high wage earners are also likely to be higher users of writing instruments, of computers and telephones. Clearly the causality of ability and skill to wages is not captured by our identification strategy here. Presumably there may be some impact of the usage of computers on wages earned but a causal inference may only be able to be made with random assignment of persons to computer usage and computer non-usage, and with a solid identification strategy backing it. With this study however, we are unable to make any specific causal conclusions of the role played by computer usage on increasing wages.

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 whether 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 data 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).

The instructions specified in the question have been executed in the following Stata code.

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

2b. Run a regression of the log 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 8 presents the results of this basic regression. If one were to take the results in Table 8 at face value (as suggested by the question), one would conclude that both Implied contract (mico) and Public policy (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 8 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 employment).

The estimates in Table 8 may be both biased and have a higher variance. The bias may arise because of an incorrect specification that ignores potential endogeneity in the identification. Potential missing variable bias may be corrected with an appropriate quasi experimental setup where the researcher exploits the change in policy across states at various points of time. Additionally, state level effects, year level effects and time trends are not controlled for that may lead to higher standard errors of the estimates. It is possible that different states have different starting dispositions that might affect the level of temporary workers.

The instructions specified in the question have been executed in the following Stata code.

```
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 8:

Basic Regression	
(1)	
VARIABLES	ln(temporary employment)
Implied contract	0.4210*** (0.1333)
Public policy	0.5706*** (0.1350)
Good faith	-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 9 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 9, we observe that the standard error values are higher than those in Table 8. The higher standard error values in Table 9 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 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.

Table 9: Regression Results

Regression with Robust Cluster SE	
(1)	
VARIABLES	ln(temporary employment)
Implied contract	0.4210 (0.3491)
Public policy	0.5706 (0.3877)
Good faith	-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	

The instructions specified in the question have been executed in the following Stata code.

```
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 10 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 10: Regression with Year and State Controls

	(1)	(2)
	Log(Temporary Employment)	Log(Temporary Employment)
Implied contract	-0.117 (0.369)	0.812*** (0.125)
Public policy	0.192 (0.380)	0.808*** (0.138)
Good faith	-0.298 (0.611)	0.545** (0.199)
Constant	7.917*** (0.257)	6.651*** (0.125)
Observations	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

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

The instructions specified in the question have been executed in the following Stata code.

```

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 specific 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 11 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 11) 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.

The instructions specified in the question have been executed in the following Stata code.

```
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 11: 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 (0.099)	0.137** (0.062)
Public policy	0.126 (0.094)	-0.023 (0.058)
Good faith	0.100 (0.113)	-0.079 (0.093)
Constant	6.596*** (0.099)	6.117*** (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 12 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.

The instructions specified in the question have been executed in the following Stata code.

```
gen lnannemp = ln(anneemp)
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 12: 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. However from the data it does not seem like the above are the reasons. I suppose, therefore that the reason they author chose the region by year dummies so as to give up fewer degrees of freedom and improve the significance of his results given the sample size of 850.

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. [Autor \(2003\)](#) has very cleverly exploited the time variation in policy enforcement to lay out a strong causal inference of implied contract on the creation of temporary jobs. While it may be hard to make the claim that column 7 of table 5 in [Autor \(2003\)](#) is devoid of all bias, the results are definitely believable and robust having been controlled for all variable one can imagine.

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 13 replicates the results in column (2) of table 7 in Autor (2003). The reason the author demonstrates the effect of lagged and leading policy variables is to strengthen the claim for the causality being argued for. Table 7 clearly suggests that only lagged policy variables affect the outcome, and this makes his case hard to argue with.

The instructions specified in the question have been executed in the following Stata code.

```
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 13: 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)
	(0.058)
Law change _{t+2}	-0.017 (0.080)
Law change _{t+1}	-0.001
Law change _{t0}	0.108 (0.052)
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.