# ECON3150/4150: Introductory Econometrics - Exam Spring 2023 

1. ( $80 \%$ ) Suppose you have the following data on mother's smoking and children's birth weight. Descriptive statistics of your data stored in data frame df are as follows:

| \#\# | mean | SD | $\min$ | $\max$ | $N$ |
| :--- | ---: | ---: | ---: | ---: | ---: |
| \#\# bwght | 118.7080029 | 20.3588787 | 23.0 | 271 | 1387 |
| \#\# faminc | 29.0421774 | 18.7371174 | 0.5 | 65 | 1387 |
| \#\# motheduc | 12.9358327 | 2.3767284 | 2.0 | 18 | 1387 |
| \#\# parity | 1.6330209 | 0.8941882 | 1.0 | 6 | 1387 |
| \#\# male | 0.5212689 | 0.4997276 | 0.0 | 1 | 1387 |
| \#\# cigs | 2.0886806 | 5.9745789 | 0.0 | 50 | 1387 |
| \#\# cigtax | 19.5598414 | 7.7941839 | 2.0 | 38 | 1387 |

where

1. bwght birth weight, ounces ( 1 ounce $=28.35$ gr.)
2. faminc 1988 family income, $\$ 1000$ s
3. motheduc mother's yrs of educ
4. parity birth order of child
5. male $=1$ if male child
6. cigs cigs smked per day while preg
7. cigtax cig. tax in home state, 1988

You perform the following analysis:

```
reg1 = feols(bwght ~ cigs, dt)
reg2 = feols(log(bwght) ~ cigs, dt)
reg3 = feols(bwght ~ cigs + faminc, dt)
reg4 = feols(bwght ~ cigs + faminc + motheduc + parity + male, dt)
etable(reg1, reg2, reg3, reg4, signif.code = NA)
```

| \#\# |  |  | reg1 | reg2 |  | reg3 |  | reg4 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| \#\# | \# Dependent Var.: |  | bwght | $\log$ (bwght) |  | bwght |  | bwght |
| \#\# |  |  |  |  |  |  |  |  |
| \#\# | \# Constant | 119.8 | (0.575) | 4.77 (0.005) | 117.0 | (1.04) | 111.8 | (3.21) |
| \#\# | \# cigs | -0.514 | (0.088) | -0.004 (0.0008) | -0.464 | (0.089) | -0.473 | (0.090) |
| \#\# | \# faminc |  |  |  | 0.092 | (0.029) | 0.099 | (0.031) |
| \#\# | \# motheduc |  |  |  |  |  | 0.053 | (0.238) |
|  | \# parity |  |  |  |  |  | 1.65 | (0.601) |
|  | \# male |  |  |  |  |  | 3.15 | (1.07) |

```
##
## S.E. type Heterosk.-rob. Heteroske.-rob. Heterosk.-rob. Heterosk.-rob.
## Observations 1,387 1,387 1,387
## RMSE 20.118 0.18875 19.933
```

a. Interpret the estimates and their standard errors in the 1st column (reg1).

ANSWER HINT: Smoking one cigarette per day while pregnant is associated with a 0.5 ounce lower birth weight. this is precisely estimated and statistically significant at conventional levels. The standard error of 0.09 implies a $95 \%$ CI of about ( $0.34,0.69$ ).
b. What is the $t$-value corresponding to the null-hypothesis that the intercept equals 100 ?

ANSWER HINT: (119.8-100) / 0.575 = 34.43
c. Construct (briefly explain your steps) and interpret the 80 percent confidence interval for the estimate on cigs in 1.a.

ANSWER HINT: We need to find the critical level which puts $20 \%$ in the tails i.e. z for which $\mathrm{P}(\mathrm{Z}<\mathrm{z})=0.9$. the closed nr in the attached table is 0.9032 which implies $\mathrm{z}=1.3$. and the $80 \% \mathrm{CI}$ is therefore $(-0.514-1.3 * 0.088,-0.514-1.3 *$ $0.088)=(-0.6284,-0.3996)$.
d. Someone argues that cigarettes cannot have a beneficial effect on birth weight and suggests you do one-sided testing. What is the critical value for the null hypothesis that cigarettes do not affect birth weight at the $10 \%$ significance level?

ANSWER HINT: We rule out positive effects which means that we pick the critical level that puts $10 \%$ mass in the negative tail, ie the z for which $\mathrm{P}(\mathrm{Z}<-\mathrm{z})=0.1$. The attached table shows only probabilities for positive z but we know that the normal distribution is symmetric, so that $\mathrm{P}(\mathrm{Z}<-\mathrm{z})=1-\mathrm{P}(\mathrm{Z}<\mathrm{z})$ and we therefore have $\mathrm{P}(\mathrm{Z}<\mathrm{z})=1-.1=0.9$. The closest probability is 0.9032 which means that the critical level is -1.3 (ie we reject if the $t$ stat is more negative than -1.3).
e. Compute the R-squared of the regression in the first column.

ANSWER HINT: R-squared $=1$ - mean error / variance of the outcome $=1$ $(20.118 / 20.3589)^{\wedge} 2=0.02353$.
f. 1 ounce is about 28.35 grams. If you would measure birth weight in grams, how would this affect your estimates in the first column?

ANSWER HINT: Instead of estimating $y=b 0+b 1 \cdot$ cigs we estimate $y \cdot c=$ $b 0 \cdot c+b 1 \cdot c \cdot$ cigs which shows that all estimates are multiplied by $\mathrm{c}=28.35$.
g. Interpret the coefficients in the 2nd column (reg2).

ANSWER HINT: The coefficient on cigs implies that one cigarette per day during pregancy is associated with a $0.4 \%$ lower birth weight. the intercept is the estimate
of birth weight without smoking $($ cigs=0 $)$ which equals $\exp (4.77)=117.9$ ounces (ca 3343 grams).
h. 1 ounce is about 28.35 grams. If you would measure birth weight in grams, how would this affect your estimates in column 2?

ANSWER HINT: We estimate $\log (c \cdot y)=\log (c)+\log (y)=\log (c)+b 0+b 1$ cigs which shows that the coefficient on cigs is unaffected but that the intercept increases with $\log (28.35)$.
i. Can we give the estimate on cigs in column 3 a causal interpretation? Does the specification in column 4 that adds more regressors change your mind? Motivate your answer.

ANSWER HINT: Here we see that keeping family income constant a cigarette per day during pregnancy is associated with 0.464 ounce lower birth weight (very similar to column 1). We believe that it is causal if we can convincingly rule out omitted variables that may be correlated with maternal smoking and factors that contribute to birth weight and which are unaccounted for. Here one can think of alcohol consumption, nutrition, general health, sleep, stress, etc. none of which is kept constant in column 3. While column 4 adds some controls which may partially account for this (perhaps in particular maternal education), it therefore seems hard to be confident that the estimate is indeed causal.
j. The third column (reg3) adds family income to the specification. Use the omitted variable bias formula to compute the correlation between family income (faminc) and cigarette smoking (cigs).

ANSWER HINT: Compared to reg3, bwgt $=b 0+b 1 \cdot c i g s+b 2 \cdot f a m i n c$, the omitted variable bias formula for reg1 is $\widehat{b 1}=b 1+b 2 \cdot \operatorname{cov}($ cigs, faminc $) / v a r($ cigs $)$. We have $\widehat{b 1}=-0.514$ from reg $1, b 1=-0.464, b 2=0.092$ from reg 3 , and $\operatorname{var}(\operatorname{cigs})$ $=5.9746^{\wedge} 2=35.7$ from the descriptive statistics. This gives $\operatorname{cov}($ cigs, faminc $)=$ $(-0.514--0.464)^{*} 5.9746^{\wedge} 2 / 0.092=-19.4$. the correlation is therefore $-19.4 /$ $(18.7371 * 5.9746)=-0.1734$.
k. Perform an F-test (you will need to assume homoskedasticity) that tests the null hypothesis that the coefficients on motheduc, parity, and male are jointly zero in the final regression reg4.

ANSWER HINT: We can use the R-squared formulation of the F-test. the R2 in reg3 equals $1-(20.046 / 20.3589)^{\wedge} 2=0.0305$, and in reg4 : $1-(19.933 /$ $20.3589)^{\wedge} 2=0.0414$. the F-statistic is therefore: $((0.0414-0.0305) / 3) /((1-$ $0.0414) /(1387-3))=5.245706$ and follows an $F(3,1384)$ distribution under the null.

1. You want to test whether cigarettes have the same effect on birthweight for kids with mothers who have completed high school (motheduc $>=12$ ) compared to children with mothers who do not have a high school diploma (motheduc $<12$ ). What is the specification the regression that you will estimate and the null hypothesis you
will be testing?
ANSWER HINT: Add a dummy I(motheduc $>=12$ ) and an interaction of the dummy and cigs: $\mathrm{I}(\text { motheduc }>=12)^{*}$ cigs to your regression. the null hypothesis involves a zero coefficient on the interaction.

A friend suggests to use instrumental variable estimation rather than OLS, and proposes to use the cigarette tax in the home state as the instrumental variable. She also provides the following OLS regression results:

```
reg5 = feols(cigtax ~ cigs + faminc + motheduc + parity + male, dt)
reg6 = feols(cigs ~ cigtax + faminc + motheduc + parity + male, dt)
reg7 = feols(bwght ~ cigtax + faminc + motheduc + parity + male, dt)
etable(reg5, reg6, reg7, signif.code = NA)
```


m . Compute and interpret the IV estimate of the effect of cigarettes on birth weight.
ANSWER HINT: IV $=$ reduced form $/$ first stage $=0.106 / 0.031=3.42$ smoking one cigarette per day while pregnant is estimated to increase birth weight with 3.42 ounces. this has the unexpected sign.
n. Do you think this instrument satifies the exclusion restriction? Motivate your answer.

ANSWER HINT: No. Cigarette taxes vary across states but the regression does not control for differences across states that may correlate with birth weight (like things noted in answer hint 1.i. above). The instrument is therefore probably not independent from the error term in the outcome (2nd stage) equation.
o. Is the instrument relevant? Motivate your answer.

ANSWER HINT: No. The instrument is not statistically significant in the firststage reg6: it has a $t=0.106 / 0.070=1.51$ or an F of $1.5^{\wedge} 2=2.3$.
2. (20\%) In 1980, Kentucky raised its cap on weekly earnings that were covered by worker's disability compensation program. We want to know if this new policy caused workers to spend more time unemployed. The cap increase did not affect low-earnings workers, but did affect high-earnings workers. You have data on the following:

1. durat duration of benefits
2. afchnge $=1$ if after change in benefits
3. highearn $=1$ if high earner
4. male $=1$ if male
5. married $=1$ if married
6. $\mathrm{ky}=1$ for Kentucky
7. $\mathrm{mi}=1$ for Michigan
8. ldurat $\log$ (durat)
9. afhigh afchnge * highearn
10. head $=1$ if head injury
11. neck $=1$ if neck injury
12. upextr $=1$ if upper extremities injury
13. trunk $=1$ if trunk injury
14. lowback $=1$ if lower back injury
15. lowextr $=1$ if lower extremities injury
16. occdis $=1$ if occupational disease
17. manuf $=1$ if manufacturing industry
18. construc $=1$ if construction industry

| \#\# | mean | SD | min | max | N |
| :---: | :---: | :---: | :---: | :---: | :---: |
| \#\# durat | 9.92220280 | 24.4975417 | 0.250000 | 182.000000 | 7150 |
| \#\# afchnge | 0.47328671 | 0.4993208 | 0.000000 | 1.000000 | 7150 |
| \#\# highearn | 0.39888112 | 0.4897025 | 0.000000 | 1.000000 | 7150 |
| \#\# male | 0.78062798 | 0.4138501 | 0.000000 | 1.000000 | 7134 |
| \#\# married | 0.69225157 | 0.4615955 | 0.000000 | 1.000000 | 6853 |
| \#\# hosp | 0.26209790 | 0.4398064 | 0.000000 | 1.000000 | 7150 |
| \#\# indust | 2.29249123 | 0.8767738 | 1.000000 | 3.000000 | 7125 |
| \#\# injtype | 4.45090909 | 1.5169241 | 1.000000 | 8.000000 | 7150 |
| \#\# age | 34.70584943 | 12.5902519 | 12.000000 | 98.000000 | 7146 |
| \#\# prewage | 329.72848184 | 182.7989394 | 81.780602 | 1583.099976 | 7150 |
| \#\# totmed | 1714.42189230 | 27853. 3720862 | 0.000000 | 2323376.500000 | 7150 |
| \#\# injdes | 4384.61034965 | 1332.2382953 | 1007.000000 | 9052.000000 | 7150 |
| \#\# benefit | 162.92344510 | 61.4193557 | 14.869200 | 742.220886 | 7150 |
| \#\# ky | 0.78685315 | 0.4095592 | 0.000000 | 1.000000 | 7150 |
| \#\# mi | 0.21314685 | 0.4095592 | 0.000000 | 1.000000 | 7150 |
| \#\# ldurat | 1.33271222 | 1.3085423 | -1.386294 | 5.204007 | 7150 |
| \#\# afhigh | 0.19300699 | 0.3946861 | 0.000000 | 1.000000 | 7150 |
| \#\# lprewage | 5.65397907 | 0.5359215 | 4.404040 | 7.367140 | 7150 |

```
## lage 3.48346979 0.3548254 2.484907 4.584968 7146
## ltotmed 5.92723504 1.7438252 0.000000 14.6585327150
## head 0.03636364 0.1872064 0.000000 1.000000 7150
## neck 0.01706294 0.1295150 0.000000 1.0000007150
## upextr 0.29524476 0.4561846 0.000000 1.000000 7150
## trunk 0.11412587 0.3179863 0.000000 1.0000007150
## lowback 0.26181818 0.4396549 0.000000 1.0000007150
## lowextr 0.23160839 0.4218896 0.000000 1.0000007150
## occdis 0.01076923 0.1032218 0.000000 1.0000007150
## manuf 0.28084211 0.4494421 0.000000 1.0000007125
## construc 0.14582456 0.3529550 0.000000 1.0000007125
## highlpre 2.48413655 3.0532814 0.000000 7.3671407150
injury[ky==1, mean(log(durat)), by = c("highearn", "afchnge")]
```

| \#\# | highearn | afchnge | V1 |
| :---: | :---: | :---: | :---: |
| \#\# | <int> | <int> | <num> |
| \#\# 1: | 1 | 1 | 1.580352 |
| \#\# 2: | 0 | 1 | 1.133273 |
| \#\# 3: | 1 | 0 | 1.382094 |
| \#\# 4: | 0 |  | 1.125615 |

You decide to estimate the impact on the logarithm of benefit duration.
a. Use the data in the output above to compute the difference-in-differences estimate.

ANSWER HINT: (1.580352-1.382094) - (1.133273-1.125615) $=0.1906$. So the policy led to a $19 \%$ increase in the duration of benefits.
b. Explain in the context of this application what you need to assume to give the estimate in 2.a a causal interpretation.

ANSWER HINT: The common trend assumption must hold. In this context it means that without the increase in the cap the high-earnings workers would have experienced the same change in the $\log$ (duration) of benefits as the low earnings workers.
c. Use the variables in the dataset above to write out the regression equation you would estimate to recover the answer in 2.a.

ANSWER HINT: ldurat $=\mathrm{b} 0+\mathrm{b} 1 *$ afhigh $+\mathrm{b} 2 *$ highearn $+\mathrm{b} 3 *$ afchnge + residual where b1 is the difference-in-differences estimate.

## The Cumulative Standard Normal Distribution Function, $\mathbf{P}(\mathrm{Z}<=\mathrm{z})$

```
## Second decimal value of z:
```



So for example, $\mathrm{P}(\mathrm{Z}<=0.22)=0.5871$

