Problem Set 4

Program Evaluation and Causal Inference

Christian Birchler Fenqi Guo Mingrui Zhang Wenjie Tu Zunhan Zhang

Spring Semester 2021

Names are listed in alphabetical order

Instrumental Variables

1. Bias of the IV estimator

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

1(a)

$$\begin{aligned} Cov(y_i, z_i) &= Cov(\beta_0 + \beta_1 x_i + u_i, z_i) \\ &= \beta_1 Cov(x_i, z_i) + \underbrace{Cov(u_i, z_i)}_{0} \\ &= \beta_1 Cov(x_i, z_i) \end{aligned}$$

$$\beta_1 = \frac{Cov(y_i, z_i)}{Cov(x_i, z_i)}$$

where $Cov(y_i, z_i)$ can be obtained from reduced-form equation and $Cov(x_i, z_i)$ can be obtained from first-stage regression.

$$\begin{aligned} \hat{\beta}_{IV} &= \frac{\widehat{Cov(y_i}, z_i)}{\widehat{Cov(x_i}, z_i)} \\ &= \frac{\sum_{i=1}^n (z_i - \overline{z})(y_i - \overline{y})}{\sum_{i=1}^n (z_i - \overline{z})(x_i - \overline{x})} \\ &= \frac{\sum_{i=1}^n (z_i y_i - z_i \overline{y} - y_i \overline{z} + \overline{yz})}{\sum_{i=1}^n (z_i x_i - z_i \overline{x} - x_i \overline{z} + \overline{xz})} \\ &= \frac{\sum_{i=1}^n (z_i y_i - y_i \overline{z}) - \overline{y} \sum_{i=1}^n z_i + n\overline{yz}}{\sum_{i=1}^n (z_i x_i - x_i \overline{z}) - \overline{x} \sum_{i=1}^n z_i + n\overline{xz}} \\ &= \frac{\sum_{i=1}^n (z_i y_i - y_i \overline{z}) - n\overline{yz} + n\overline{yz}}{\sum_{i=1}^n (z_i x_i - x_i \overline{z}) - n\overline{xz} + n\overline{xz}} \\ &= \frac{\sum_{i=1}^n (z_i - \overline{z}) y_i}{\sum_{i=1}^n (z_i - \overline{z}) x_i} \end{aligned}$$

1(b)

$$\begin{aligned}
\hat{\beta}_{IV} &= \frac{\sum_{i=1}^{n}(z_i-\overline{z})y_i}{\sum_{i=1}^{n}(z_i-\overline{z})x_i} \\
&= \frac{\sum_{i=1}^{n}(z_i-\overline{z})(\beta_0+\beta_1 x_i+u_i)}{\sum_{i=1}^{n}(z_i-\overline{z})x_i} \\
&= \frac{\beta_0\sum_{i=1}^{n}(z_i-\overline{z})+\beta_1\sum_{i=1}^{n}(z_i-\overline{z})x_i+\sum_{i=1}^{n}(z_i-\overline{z})u_i}{\sum_{i=1}^{n}(z_i-\overline{z})x_i} \\
&= \beta_1+\underbrace{\frac{\sum_{i=1}^{n}(z_i-\overline{z})u_i}{\sum_{i=1}^{n}(z_i-\overline{z})x_i}}_{\text{bias}}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_{IV}\mid x_i,z_i\right] &= \beta_1+\mathbb{E}\left[\frac{\sum_{i=1}^{n}(z_i-\overline{z})u_i}{\sum_{i=1}^{n}(z_i-\overline{z})x_i}\middle| x_i,z_i\right] \\
&= \beta_1+\frac{\sum_{i=1}^{n}(z_i-\overline{z})}{\sum_{i=1}^{n}(z_i-\overline{z})x_i}\underbrace{\mathbb{E}\left[u_i\mid x_i,z_i\right]}_{\neq 0}
\end{aligned}$$

$\mathbb{E}\left[u_i\mid x_i,z_i\right]\neq 0$ since $u_i \not\perp x_i$.

1(c)

$$\begin{aligned}
p\lim(\hat{\beta}_{IV}-\beta_1) &= \frac{p\lim\sum_{i=1}^{n}(z_i-\overline{z})u_i}{p\lim\sum_{i=1}^{n}(z_i-\overline{z})x_i} \\
&= \frac{p\lim\sum_{i=1}^{n}(z_i-\overline{z})(u_i-\overline{u})}{p\lim\sum_{i=1}^{n}(z_i-\overline{z})(x_i-\overline{x})} \\
&= \frac{p\lim\frac{1}{n}\sum_{i=1}^{n}(z_i-\overline{z})(u_i-\overline{u})}{p\lim\frac{1}{n}\sum_{i=1}^{n}(z_i-\overline{z})(x_i-\overline{x})} \\
&\approx \frac{Cov(z_i,u_i)}{Cov(z_i,x_i)}=0
\end{aligned}$$

By *Exogeneity Assumption*, $Cov(z_i,u_i)=0$ and by *Relevance Assumption*, $Cov(z_i,x_i)\neq 0$ Therefore, $\hat{\beta}_{IV}$ is a consistent estimator of β_1.

1(d)

In a small sample, $\hat{\beta}_{IV}$ is biased. But as the sample increases, β_{IV} will probability converge to the β_1. Therefore, in a large sample, IV estimator will be a consistent estimator of β_1 regardless of whether there exists an endogeneity problem.

2. Derivation of the Wald estimator

2(a)

From question 1, we know that

$$
\begin{aligned}
\delta^W = \hat{\beta}_1 &= \frac{Cov(y_i, z_i)}{Cov(d_i, z_i)} \\
&= \frac{\mathbb{E}[y_i|z_i = 1] - \mathbb{E}[y_i|z_i = 0]}{\mathbb{E}[d_i|z_i = 1] - \mathbb{E}[d_i|z_i = 0]} \\
&= \frac{\mathbb{E}[\beta_1 + \beta_1 d_i + u_i|z_i = 1] - \mathbb{E}[\beta_1 + \beta_1 d_i + u_i|z_i = 0]}{\mathbb{E}[d_i|z_i = 1) - \mathbb{E}(d_i|z_i = 0]} \\
&= \frac{\beta_1 \mathbb{E}[d_i|z_i = 1] + \mathbb{E}[u_i|z_i = 1] - \beta_1 \mathbb{E}[d_i|z_i = 0] - \mathbb{E}[u_i|z_i = 0]}{\mathbb{E}[d_i|z_i = 1) - \mathbb{E}(d_i|z_i = 0]} \\
&= \frac{\beta_1 \left(\mathbb{E}[d_i|z_i = 1] - \mathbb{E}[d_i|z_i = 0]\right) + \mathbb{E}[u_i|z_i = 1] - \mathbb{E}[u_i|z_i = 0]}{\mathbb{E}[d_i|z_i = 1] - \mathbb{E}[d_i|z_i = 0]} \\
&= \beta_1 + \frac{\mathbb{E}[u_i|z_i = 1] - \mathbb{E}[u_i|z_i = 0]}{\mathbb{E}[d_i|z_i = 1] - \mathbb{E}[d_i|z_i = 0]}
\end{aligned}
$$

2(b)

In order to identify β_1 using the instrument, we need

$$
\frac{\mathbb{E}[u_i|z_i = 1] - \mathbb{E}[u_i|z_i = 0]}{\mathbb{E}[d_i|z_i = 1] - \mathbb{E}[d_i|z_i = 0]} = 0 \iff \begin{cases} \mathbb{E}[u_i|z_i = 1] = \mathbb{E}[u_i|z_i = 0] & \text{Exclusion Assumption} \\ \mathbb{E}[d_i|z_i = 1] \neq \mathbb{E}[d_i|z_i = 0] & \text{Relevance Assumption} \end{cases}
$$

Assumptions

- SUTVA (Stable Unit Treatment Value Assumption): outcomes of the ith individual are independent of other individuals' outcome
- Exclusion restriction: $\mathbb{E}[y_i|z = 1, d] = \mathbb{E}[y_i|z = 0, d] \quad \forall i = 0, 1$
- Relevance assumption: $\mathbb{E}[d|z = 1] \neq \mathbb{E}[d|z = 0]$
- Monotonicity assumption: $d_i[z_i = 1] \geq d_i[z_i = 0] \quad \forall i$

Only relevance assumption can be tested empirically. The vadility of other assumptions must be assessed on a case-by-case basis.

3. Self selection revisited

3(a)

$$
\begin{aligned}
D_i &= \mathbf{1}(Y_{1i} - Y_{0i} > 0) \\
&= \mathbf{1}(\beta_1 + u_{1i} - u_{0i} > 0)
\end{aligned}
$$

$$
\begin{aligned}
\Delta^{\text{ATE}} &= \mathbb{E}(Y_{1i} - Y_{0i}) \\
&= \mathbb{E}[(\beta_0 + \beta_1 + u_{1i}) - (\beta_0 + u_{0i})] \\
&= \mathbb{E}(\beta_1 + u_{1i} - u_{0i}) \\
&= \mathbb{E}(\beta_1) + \mathbb{E}(u_{1i}) + \mathbb{E}(u_{0i}) \\
&= \beta_1 > 0
\end{aligned}
$$

$$
\begin{aligned}
\Delta^{\text{ATT}} &= \mathbb{E}(Y_{1i} - Y_{0i} \mid D = 1) \\
&= \mathbb{E}(\beta_1 + u_{1i} - u_{0i} \mid D = 1) \\
&= \beta_1 + \mathbb{E}(u_{1i} - u_{0i} \mid D = 1) \\
&= \Delta^{\text{ATE}} + \mathbb{E}(u_{1i} - u_{0i} \mid D = 1) \\
\mathbb{E}(u_{1i} - u_{0i} \mid D = 1) &= \mathbb{E}(u_{1i} - u_{0i} \mid \beta_1 + u_{1i} - u_{0i} > 0) \\
&= \mathbb{E}(u_{1i} - u_{0i} \mid u_{1i} - u_{0i} > -\beta_1) > 0
\end{aligned}
$$

ATT is larger than ATE.

3(b)

$$
\begin{aligned}
\Delta^{\text{naive}} &= \mathbb{E}(Y_{1i} \mid D = 1) - \mathbb{E}(Y_{0i} \mid D = 0) \\
&= \mathbb{E}(Y_{1i} \mid D = 1) - \mathbb{E}(Y_{0i} \mid D = 0) - \mathbb{E}(Y_{0i} \mid D = 1) + \mathbb{E}(Y_{0i} \mid D = 1) \\
&= \underbrace{\mathbb{E}(Y_{1i} \mid D = 1) - \mathbb{E}(Y_{0i} \mid D = 1)}_{\Delta^{\text{ATT}}} + \underbrace{\mathbb{E}(Y_{0i} \mid D = 1) - \mathbb{E}(Y_{0i} \mid D = 0)}_{\text{selection bias}}
\end{aligned}
$$

$$
\begin{aligned}
\text{selection bias} &= \mathbb{E}(Y_{0i} \mid D = 1) - \mathbb{E}(Y_{0i} \mid D = 0) \\
&= \mathbb{E}(\beta_0 + u_{0i} \mid \beta_1 + u_{1i} - u_{0i} > 0) - \mathbb{E}(\beta_0 + u_{0i} \mid \beta_1 + u_{1i} - u_{0i} \leq 0) \\
&= \underbrace{\mathbb{E}(u_{0i} \mid u_{0i} < u_{1i} + \beta_1)}_{\text{negative}} - \underbrace{\mathbb{E}(u_{0i} \mid u_{0i} \geq u_{1i} + \beta_1)}_{\text{positive}} \\
&< 0
\end{aligned}
$$

If individuals can self-select themselves into the program, the naive estimator will be underestimated since the selection bias is negative ($\mathbb{E}(Y_{0i}|D = 1) < \mathbb{E}(Y_{0i}|D = 0)$)

3(c)

$$
\begin{cases}
D_{1i} = \mathbf{1}(Y_{1i} - Y_{0i} + Z_i > 0) & Z_i = 1 \\
D_{0i} = \mathbf{1}(Y_{1i} - Y_{0i} > 0) & Z_i = 0
\end{cases}
$$

$$
\begin{cases}
D_{1i} = \mathbf{1}(u_{1i} - u_{0i} > -\beta_1 - Z_i) & Z_i = 1 \\
D_{0i} = \mathbf{1}(u_{1i} - u_{0i} > -\beta_1) & Z_i = 0
\end{cases}
$$

In order to identify LATE, we need to determine the conditions for compliers. Compliers are those who are induced to switch treatment status as a result of the instrument. Therefore, D_{1i}>D_{0i}$ should hold for compliers.

$$
\text{LATE} = \mathbb{E}(Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i})
$$

3(d)

$$
D_{1i} > D_{0i} \implies \begin{cases} D_{1i} = 1 \\ D_{0i} = 0 \end{cases} \implies \begin{cases} u_{1i} - u_{0i} > -\beta_1 - Z_1 \\ u_{1i} - u_{0i} < -\beta_1 \end{cases} \implies -\beta_1 - Z_1 < u_{1i} + u_{0i} < -\beta_1
$$

$$
\begin{aligned}
\text{LATE} &= \mathbb{E}(Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}) \\
&= \beta_1 + \mathbb{E}(u_{1i} - u_{0i} \mid D_{1i} > D_{0i}) \\
&= \beta_1 + \underbrace{\mathbb{E}(u_{1i} - u_{0i} \mid -\beta_1 - Z_1 < u_{1i} + u_{0i} < -\beta_1)}_{<-\beta_1} \\
&< \beta_1 \equiv \text{ATE}
\end{aligned}
$$

4. Application: Angrist's (1990) study on military service

4(a)

The OLS estimate may be biased:

- There may be a self-selection bias from participants. People may self-select themselves into or not into the program.
- There may be a self-selection bias from experimenter. The military authority can select who can join the army and who cannot.

4(b)

	$Z = 0$	$Z = 1$
$D = 0$	5,928	1,875
$D = 1$	1,400	863

Due to monotonicity:

In the observed $Z = 0$ group, the individuals who received treatment ($D = 1$) must be always-takers.

$$p_A = \mathbb{E}(D_i \mid Z_i = 0) = \frac{\sum_i \mathbf{1}(D_i[Z_i = 0] = 1)}{\sum_i \mathbf{1}(Z_i = 0)} = \frac{1400}{5928 + 1400} = 0.191$$

In the observed $Z = 1$ group, the individuals who did not receive treatment ($D = 0$) must be never-takers.

$$p_N = 1 - \mathbb{E}(D_i \mid Z_i = 1) = \frac{\sum_i \mathbf{1}(D_i[Z_i = 1] = 0)}{\sum_i \mathbf{1}(Z_i = 1)} = \frac{1875}{1875 + 863} = 0.685$$

$$
\begin{aligned}
p_C &= \mathbb{E}(D_i \mid Z_i = 1) - \mathbb{E}(D_i \mid Z_i = 0) \\
&= 1 - 0.685 - 0.191 \\
&= 0.124
\end{aligned}
$$

Due to randomization:

The proportions of compliers, always-takers, and never-takers are the same between $Z = 0$ and $Z = 1$ group.

$$p_C = 1 - p_A - p_N = 0.124$$

Note:

- N denotes **never takers**
- C denotes **compliers**
- A denotes **always takers**

4(c)

	$Z = 0$	$Z = 1$
$D = 0$	$\widehat{\mathbb{E}(Y)} = 6.4472$	$\widehat{\mathbb{E}(Y)} = 6.4028$
$D = 1$	$\widehat{\mathbb{E}(Y)} = 6.4076$	$\widehat{\mathbb{E}(Y)} = 6.4289$

- Average potential outcome for always-takers $\mathbb{E}(Y_{1i} \mid D_i = 1, Z_i = 0) = 6.4076$
- Average potential outcome for never-takers $\mathbb{E}(Y_{0i} \mid D_i = 0, Z_i = 1) = 6.4028$

Average potential outcome for compliers:

In $Z = 0$ group:

$$\underbrace{\mathbb{E}(Y_0 \mid D = 0, Z = 0)}_{6.4472} = \frac{p_C}{p_N + p_C} \times \mathbb{E}(Y_{0i} \mid D_{1i} > D_{0i}) + \frac{p_N}{p_N + p_C} \times \mathbb{E}(Y_{0i} \mid D_{1i} = 0)$$

$$\mathbb{E}(Y_{0i} \mid D_{1i} > D_{0i}) = 6.6867$$

In $Z = 1$ group:

$$\underbrace{\mathbb{E}(Y_1 \mid D = 1, Z = 1)}_{6.4289} = \frac{p_C}{p_A + p_C} \times \mathbb{E}(Y_{1i} \mid D_{1i} > D_{0i}) + \frac{p_A}{p_A + p_C} \times \mathbb{E}(Y_{1i} \mid D_{1i} = 0)$$

$$\mathbb{E}(Y_{1i} \mid D_{1i} > D_{0i}) = 6.4616$$

- Average potential outcome for untreated compliers $\mathbb{E}(Y_0|Z = 0, C) = 6.6867$
- Average potential outcome for treated compliers $\mathbb{E}(Y_1|Z = 1, C) = 6.4616$

4(d)

$$\begin{aligned} \text{LATE} &= \mathbb{E}(Y_{1i} \mid D_{1i} > D_{0i}) - \mathbb{E}(Y_{0i} \mid D_{1i} > D_{0i}) \\ &= 6.4616 - 6.6867 \\ &= -0.2251 \end{aligned}$$

5. IV in action

5(a)

```r
# load relevant libraries
library(haven) # read dta file
library(lattice) # density plot
library(stargazer) # print summary statistics
library(ggplot2) # plot
library(AER) # iv regression
```

```r
d.mort <- read_dta('mortality.dta')
```

```r
demo <- subset(d.mort, select = c('before67dead', 'dist65_ageATend4emp', 'Zd_during'))
```

```r
subdata <- as.data.frame(
  subset(d.mort, select = c('before67dead', 'dist65_ageATend4emp', 'Zd_during'))
)
stargazer(subdata, header = F, title = 'Descriptive Statistics')
```

Table 3: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
before67dead	2,298	0.073	0.260	0	0	0	1
dist65_ageATend4emp	2,298	6.499	2.248	−5	5	8	11
Zd_during	2,298	0.479	0.500	0	0	1	1

5(b)

```r
model.ols1 <- lm(before67dead ~ dist65_ageATend4emp + czeit1yATage50 +
                   czeit2yATage50 + czeit5yATage50 + czeit10yATage50 +
                   czeit25yATage50 + I(czeit1yATage50^2) + I(czeit2yATage50^2) +
                   I(czeit5yATage50^2) + I(czeit10yATage50^2) +
                   I(czeit25yATage50^2) + as.factor(halfyearOFbirth) +
                   as.factor(nutsATage50), data=d.mort)

model.ols2 <- lm(before67dead ~ dist65_ageATend4emp, data=d.mort)

stargazer(model.ols1, model.ols2, keep.stat='n', header=F,
          keep='dist65_ageATend4emp', font.size='small',
          column.labels=c('Control', 'Non-control'), digits=4,
          title='Comparison between control and non-control')
```

5(c)

As we can see, the coefficient on the treatment slightly increases from column 1 (with control variables) to column 2 (without control variables).

Significance

- With control variables, *p-value* is smaller than 10

Table 4: Comparison between control and non-control

	Dependent variable:	
	before67dead	
	Control	Non-control
	(1)	(2)
dist65_ageATend4emp	0.0046*	0.0049**
	(0.0026)	(0.0024)
Observations	2,298	2,298
Note:	*p<0.1; **p<0.05; ***p<0.01	

- Without control variables, *p-value* is smaller than 5

We can reject the null hypothesis in both cases but we are more confident to reject $\beta_1 = 0$ with control variables.

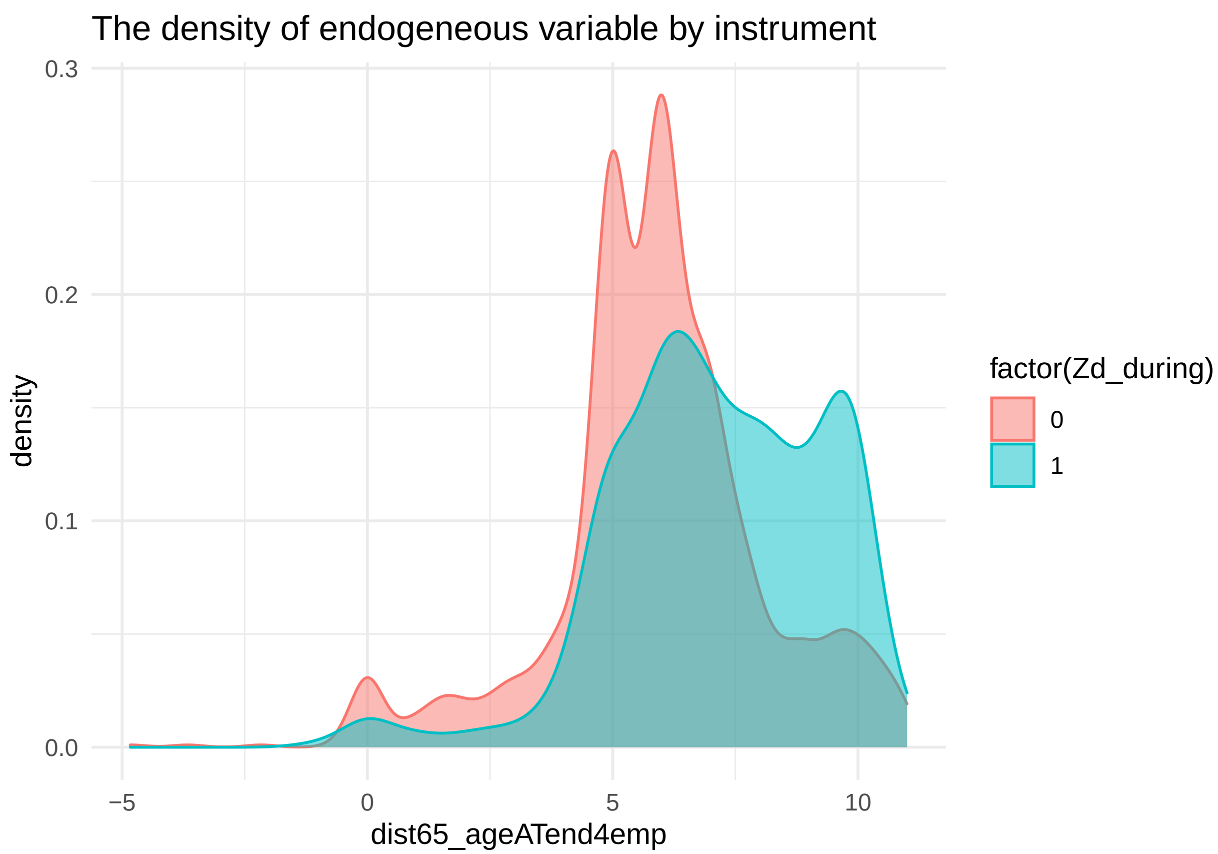
5(d)

Omitted-variable bias

Health status. If people are in a bad physical condition, they are more likely to spend less years in their early retirement or even die before retirement. Therefore, the estimator for β_1 is biased upwards and we expect a positive bias.

5(e)

```r
ggplot(d.mort, aes(x = dist65_ageATend4emp)) +
  geom_density(aes(group = factor(Zd_during),
                   color = factor(Zd_during),
                   fill = factor(Zd_during)), alpha = 0.5) +
  theme_minimal() + ggtitle('The density of endogeneous variable by instrument')
```



5(f)

```
# first stage regression
iv.1st.stage <- ivreg(dist65_ageATend4emp ~ Zd_during*as.factor(halfyearOFbirth) +
                        czeit1yATage50 + czeit2yATage50 + czeit5yATage50 +
                        czeit10yATage50 + czeit25yATage50 + I(czeit1yATage50^2) +
                        I(czeit2yATage50^2) + I(czeit5yATage50^2) + I(czeit10yATage50^2) +
                        I(czeit25yATage50^2) + as.factor(halfyearOFbirth) +
                        as.factor(nutsATage50), data=d.mort)

stargazer(iv.1st.stage, keep='Zd_during', keep.stat='n', header=F,
          font.size='small', title='First stage regression', no.space=T)
```

5(g)

```
# second stage regression
iv.2nd.stage <- lm(before67dead ~ predict(iv.1st.stage) + czeit1yATage50 +
                     czeit2yATage50 + czeit5yATage50 + czeit10yATage50 +
                     czeit25yATage50 + I(czeit1yATage50^2) + I(czeit2yATage50^2) +
                     I(czeit5yATage50^2) + I(czeit10yATage50^2) +
                     I(czeit25yATage50^2) + as.factor(halfyearOFbirth) +
```

Table 5: First stage regression

	Dependent variable:
	dist65_ageATend4emp
Zd_during	0.664*
	(0.345)
Zd_during:as.factor(halfyearOFbirth)-50	−0.448
	(0.492)
Zd_during:as.factor(halfyearOFbirth)-49	−0.070
	(0.505)
Zd_during:as.factor(halfyearOFbirth)-48	0.179
	(0.506)
Zd_during:as.factor(halfyearOFbirth)-47	−0.280
	(0.492)
Zd_during:as.factor(halfyearOFbirth)-46	0.424
	(0.471)
Zd_during:as.factor(halfyearOFbirth)-45	0.253
	(0.509)
Zd_during:as.factor(halfyearOFbirth)-44	0.525
	(0.504)
Zd_during:as.factor(halfyearOFbirth)-43	0.558
	(0.495)
Zd_during:as.factor(halfyearOFbirth)-42	0.150
	(0.460)
Zd_during:as.factor(halfyearOFbirth)-41	1.229***
	(0.450)
Zd_during:as.factor(halfyearOFbirth)-40	0.681
	(0.451)
Zd_during:as.factor(halfyearOFbirth)-39	0.608
	(0.463)
Zd_during:as.factor(halfyearOFbirth)-38	0.524
	(0.479)
Zd_during:as.factor(halfyearOFbirth)-37	0.345
	(0.459)
Observations	2,298
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
                     as.factor(nutsATage50), data=d.mort)

# iv regression
model.iv <- ivreg(before67dead ~ dist65_ageATend4emp + czeit1yATage50 +
                     czeit2yATage50 + czeit5yATage50 + czeit10yATage50 +
                     czeit25yATage50 + I(czeit1yATage50^2) + I(czeit2yATage50^2) +
                    I(czeit5yATage50^2) + I(czeit10yATage50^2) + I(czeit25yATage50^2) +
                    as.factor(halfyearOFbirth) + as.factor(nutsATage50) |
                    Zd_during*as.factor(halfyearOFbirth) + czeit1yATage50 +
                    czeit2yATage50 + czeit5yATage50 + czeit10yATage50 +
                    czeit25yATage50 + I(czeit1yATage50^2) + I(czeit2yATage50^2) +
                    I(czeit5yATage50^2) + I(czeit10yATage50^2) + I(czeit25yATage50^2) +
                    as.factor(halfyearOFbirth) + as.factor(nutsATage50), data=d.mort)

stargazer(iv.2nd.stage, model.iv, font.size='small', header=F,
          keep.stat=c('n', 'f'), title='Comparsion between 2SLS and ivreg',
          keep=c('iv.1st.stage', 'dist65_ageATend4emp'), digits=4)
```

Table 6: Comparsion between 2SLS and ivreg

	Dependent variable:	
	before67dead	
	OLS	*instrumental variable*
	(1)	(2)
predict(iv.1st.stage)	−0.0143	
	(0.0109)	
dist65_ageATend4emp		−0.0143
		(0.0110)
Observations	2,298	2,298
F Statistic	1.3992* (df = 32; 2265)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

As we can see, **2SLS** and **ivreg** yield exactly the same estimate but with different standard errors.

5(h)

```r
stargazer(model.ols1, model.iv, font.size='small', header=F,
          keep.stat=c('n', 'f'), keep='dist65_ageATend4emp',
          title='Comparison between OLS and 2SLS results', digits=4)
```

As expected, from column(1) to column(2), we see a decrease in the coefficient on *dist65_ageATend4emp*, which verifies our statement in 5(d) - a positive bias in the OLS estimator.

Table 7: Comparison between OLS and 2SLS results

	Dependent variable:	
	before67dead	
	OLS	*instrumental variable*
	(1)	(2)
dist65_ageATend4emp	0.0046*	−0.0143
	(0.0026)	(0.0110)
Observations	2,298	2,298
F Statistic	1.4431* (df = 32; 2265)	
Note:		*p<0.1; **p<0.05; ***p<0.01