ECON-Online
Assignment - 8
Piyush Kulkarni
pxk161130

Q. 10.2

a) $Hours = \beta_1 + \beta_2 Wage + \beta_3 EDU + \beta_4 Age + \beta_5 Kids\,L6 + \beta_6 Kids\,618 + \beta_7 NWIFEINC + e$

Wage → +ve more wage more supply

EDU → could be +ve or -ve. How much education women have it dosen't mean they ~~mo~~ work more. They can work part time ~~too~~ if they are more qualified.

Age → Age ~~co~~ could be negative. More the age, less the number of hours.

Kids L6 → negative. ~~A~~ presence of young children will reduce the tendency ~~to~~ work

Kids 618 → +ve As children grow up, women can start working

NWifeINC → -ve more the househol income, less the tendency of women working.

b) The term wage could be endogenous. wages may be taking the effect of underlying or omitted variable → Ability. Ability could be corelated with education as well. & can be correlated with the error term.

wage → ~~core~~ correlated with error.
education → correlated with error
→ This will cause endogenity problem & OLS will fail.
Wage depends on experience also

c) wage & Expr & wage & $Expr^2$ could be correlated as ~~wage~~ expr increases, wage increases & after a certain point, it will start declining this shows wage & $expr^2$ is also correlated at the same point
Expr & $expr^2$ will be not correlated with the error term.
∴ they can be used as instrument variables.

d) Yes the equation is identified.
1 → endogenous variable → wage
2 → instrument variables → expr & $expr^2$
instrument var > endogenous variables

e) First estimate wage as
$$\hat{wage} = r_1 + r_2 Edu + r_3 Age + r_4 KidsL6 + r_5 Kids618 + r_6 NWifeINC + \theta_1 Exper + \theta_2 Exper^2 + error$$

~~seacond~~ second replace endogenous wage variable with regressed $\hat{wage}$ in first step. that is use two step least square.

Q.10.6

a) $y = \beta_1 + \beta_2 x + e$

$= 3 + x + e$

$\sigma_x^2 = 2, \quad \sigma_e^2 = 1$

$cor(x, e) = 0.9$

correlation bet^n x & e =

$r_{xe} = \dfrac{cor(x,e)}{\sqrt{var(x) \times var(e)}} = \dfrac{0.9}{\sqrt{2 \times 1}}$

$= \cancel{0.634}$

$= 0.6364$

b) Using R correlation = 0.65136

c) Plotted using R.

d)

Observations	β_1	error β_1	β_2	error β_2
10	2.775	0.3608	1.3722	0.1727
20	3.0169	0.2036	1.3876	0.1211
100	3.007	0.07872	1.4016	0.05336
500	3.01825	0.030410	1.4535	0.02367

for β_1 as the sample size increases, the value is moving closer ~~to~~ 3 (true value) but after 100 it ~~goes~~ moves away from 3
for β_2, as sample size is increasing, the value is moving away from true value.

this could be because of correlation between x & e.

e) z_1 & x have high correlation (0.62) & z_1 & e have practically zero correlation (-0.003) makes it a good instrument variable.
z_2 also can be used (0.28 & 0.027) correlation but z_1 is better choice than z_2 as it have higher correlation with x & low with e.

f)

Observations	β_1	error β_1	β_2	error β_2
10	2.7144	0.4277	1.0640	0.2526
20	3.0810	0.2500	1.0263	0.1966
100	2.9771	0.1051	0.9363	0.1132
500	3.03150	0.04512	0.99613	0.05044

As sample size increases, β_1 & β_2 both are converging towards true parameter.

g)

Observations	β_1	error β_1	β_2	error β_2
10	2.7144	0.4277	1.0640	0.2526
20	30810	0.2500	1.0263	0.1966
100	2.9771	0.1051	0.9363	0.1132
500	3.03150	~~0.996~~ 0.04512	0.99613	0.05044

for small sample size, estimates are far away from true parameters.
As sample size is increasing, estimats are moving closer to true value.
Z_1 gives better estimates compared with Z_2

h)

Observations	β_1	error β_1	β_2	error β_2
10	2.7144	0.4277	1.0640	0.2526
20	3.0810	0.2500	1.0263	0.1966
100	2.9771	0.1051	0.9363	0.1132
500	3.03150	~~0.99613~~ 0.04152	0.99613	0.05044

As sample size is increasing, estimates are converging towards true parameter.

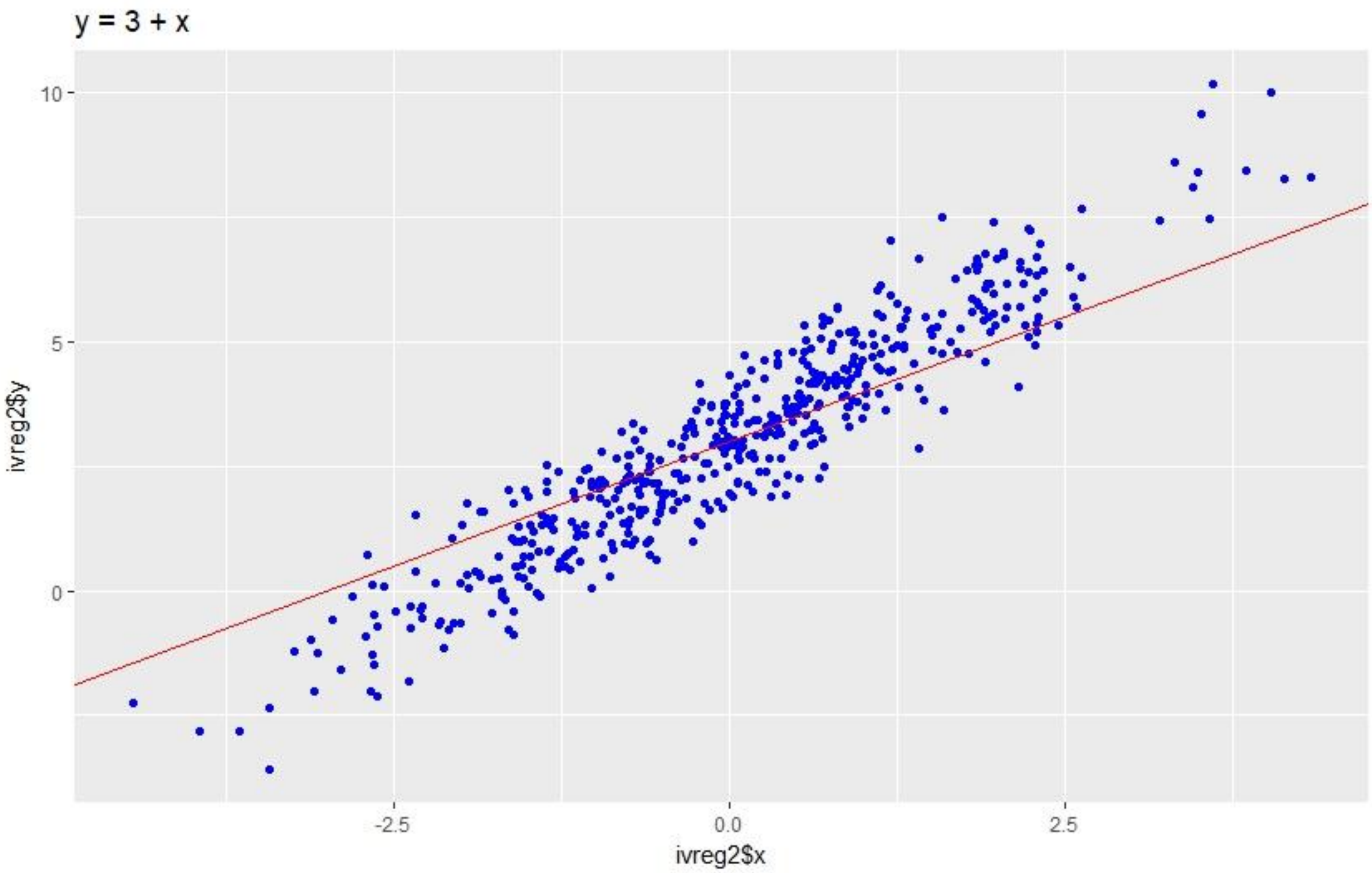
g) estimates

Observations	β_1	error β_1	β_2	error β_2
10	1.892	11.061	-2.950	51.705
20	3.2433	0.6975	0.1110	2.4712
100	2.99020	0.08865	1.13489	0.14697
500	3.0294	0.04236	1.06665	0.10136

for less no. of observations, estimates are way too far from true parameter
z_1 is better than z_2

h)

Observations	β_1	error β_1	β_2	error β_2
10	2.7114	0.4337	1.0491	0.2549
20	3.0852	0.2555	1.0026	0.1989
100	2.98076	0.09974	0.99206	0.09316
500	3.03113	0.04458	1.00899	0.04496




```
> # -----------------------calculating error
>
> e <- ivreg2$y- 3 - ivreg2$x
>
> # --- Correlation in x and e
>
> cor(ivreg2$x, e)
[1] 0.65136
>
> line <- abline(3,1)
Error in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...) :
  plot.new has not been called yet
>
> # --- Scatterplot ---
>
```

```
> library(ggplot2)
>
> ggplot(data = ivreg2, aes(x = ivreg2$x, y = ivreg2$y)) +
+         geom_point(color = 'blue') +
+             geom_abline(intercept = 3, slope = 1, color = "red") + ggtitle('y = 3 + x')
> # ------ OLS Regression with different observations -------
> data_10 <- ivreg2[1:10,]
> model_10 <- lm(y ~ x, data = data_10)
> summary(model_10)

Call:
lm(formula = y ~ x, data = data_10)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6450 -0.6888 -0.2390  0.4484  1.9556

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.7775     0.3608   7.698 5.76e-05 ***
x             1.3722     0.1727   7.945 4.59e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.136 on 8 degrees of freedom
Multiple R-squared:  0.8875,	Adjusted R-squared:  0.8735
F-statistic: 63.12 on 1 and 8 DF,  p-value: 4.589e-05

> data_20 <- ivreg2[1:20,]
> model_20 <- lm(y ~ x, data = data_20)
> summary(model_20)
```

```
Call:
lm(formula = y ~ x, data = data_20)

Residuals:
    Min       1Q   Median       3Q      Max
-1.83171 -0.52577  0.08304  0.45379  1.75205

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0169     0.2036   14.81 1.59e-11 ***
x             1.3876     0.1211   11.46 1.05e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9056 on 18 degrees of freedom
Multiple R-squared:  0.8795,	Adjusted R-squared:  0.8728
F-statistic: 131.4 on 1 and 18 DF,  p-value: 1.053e-09

> data_100 <- ivreg2[1:100,]
> model_100 <- lm(y ~ x, data = data_100)
> summary(model_100)

Call:
lm(formula = y ~ x, data = data_100)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1199 -0.5289  0.0271  0.5255  1.7940

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.00783    0.07872   38.21   <2e-16 ***
```

```
x           1.40164    0.05330   26.30   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7864 on 98 degrees of freedom
Multiple R-squared:  0.8759,	Adjusted R-squared:  0.8746
F-statistic: 691.5 on 1 and 98 DF,  p-value: < 2.2e-16

> model_500 <- lm(y ~ x , data = ivreg2)
> summary(model_500)

Call:
lm(formula = y ~ x, data = ivreg2)

Residuals:
    Min      1Q   Median      3Q     Max
-2.20345 -0.51588 -0.01086  0.52412  2.26606

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.01825    0.03410    88.5   <2e-16 ***
x            1.45352    0.02367    61.4   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7624 on 498 degrees of freedom
Multiple R-squared:  0.8833,	Adjusted R-squared:  0.8831
F-statistic:  3770 on 1 and 498 DF,  p-value: < 2.2e-16

> # ---------correlation between x, z1, z2 ,e-------
> dat <- cbind(ivreg2, e)
> cor(dat)
```

```
          x         y          z1          z2            e
x  1.0000000 0.9398447  0.620821104  0.28948601  0.651359982
y  0.9398447 1.0000000  0.399870154  0.19965601  0.871374239
z1 0.6208211 0.3998702  1.000000000 -0.01530765 -0.003447192
z2 0.2894860 0.1996560 -0.015307651  1.00000000  0.027708992
e  0.6513600 0.8713742 -0.003447192  0.02770899  1.000000000
> # ------- IV reg  Z1--------
>
> library(AER)
> ivreg_10 <- ivreg( y ~ x | z1 , data =data_10)
> summary(ivreg_10)

Call:
ivreg(formula = y ~ x | z1, data = data_10)

Residuals:
     Min       1Q   Median       3Q      Max
-2.63964 -0.67660 -0.09229  1.06554  1.56448

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.7144     0.4277   6.346 0.000222 ***
x             1.0640     0.2526   4.211 0.002951 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.343 on 8 degrees of freedom
Multiple R-Squared: 0.8427,       Adjusted R-squared: 0.8231
Wald test: 17.73 on 1 and 8 DF,  p-value: 0.002951

> ivreg_20 <- ivreg( y ~ x | z1 , data =data_20)
> summary(ivreg_20)
```

```
Call:
ivreg(formula = y ~ x | z1, data = data_20)

Residuals:
     Min       1Q   Median       3Q      Max
-3.13540 -0.43910  0.09362  0.65356  1.77434

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0810     0.2500  12.323 3.29e-10 ***
x             1.0263     0.1966   5.219 5.79e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.107 on 18 degrees of freedom
Multiple R-Squared: 0.8199,	Adjusted R-squared: 0.8099
Wald test: 27.24 on 1 and 18 DF,  p-value: 5.79e-05

> ivreg_100 <- ivreg( y ~ x | z1 , data =data_100)
> summary(ivreg_100)

Call:
ivreg(formula = y ~ x | z1, data = data_100)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3405 -0.8786  0.0898  0.7411  2.1542

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.9771     0.1051  28.320  < 2e-16 ***
```

```
x            0.9363     0.1132   8.268 6.77e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.049 on 98 degrees of freedom
Multiple R-Squared: 0.7793,       Adjusted R-squared: 0.7771
Wald test: 68.36 on 1 and 98 DF,  p-value: 6.775e-13

> model <- ivreg( y ~ x | z1 , data =ivreg2)
> summary(model)

Call:
ivreg(formula = y ~ x | z1, data = ivreg2)

Residuals:
    Min       1Q   Median       3Q      Max
-3.18955 -0.72094  0.01315  0.62964  3.54344

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.03150    0.04512   67.18   <2e-16 ***
x           0.99613    0.05044   19.75   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.008 on 498 degrees of freedom
Multiple R-Squared: 0.7958,       Adjusted R-squared: 0.7954
Wald test:   390 on 1 and 498 DF,  p-value: < 2.2e-16

> # ------------IV Reg Z2--------------
>
>
```

```
> library(AER)
> ivregz2_10 <- ivreg( y ~ x | z2 , data =data_10)
> summary(ivregz2_10)

Call:
ivreg(formula = y ~ x | z2, data = data_10)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.593  -5.249  -1.467   4.665  18.598 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.892     11.061   0.171    0.868
x             -2.950     51.705  -0.057    0.956

Residual standard error: 10.11 on 8 degrees of freedom
Multiple R-Squared: -7.919,	Adjusted R-squared: -9.034 
Wald test: 0.003256 on 1 and 8 DF,  p-value: 0.9559 

> ivregz2_20 <- ivreg( y ~ x | z2 , data =data_20)
> summary(ivregz2_20)

Call:
ivreg(formula = y ~ x | z2, data = data_20)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.4388 -1.3000  0.2736  1.4712  4.5745 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)   3.2433     0.6975   4.650 0.000199 ***
x            0.1110     2.4712   0.045 0.964670
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.426 on 18 degrees of freedom
Multiple R-Squared: 0.1351,       Adjusted R-squared: 0.08702
Wald test: 0.002017 on 1 and 18 DF,  p-value: 0.9647

> ivregz2_100 <- ivreg( y ~ x | z2 , data =data_100)
> summary(ivregz2_100)

Call:
ivreg(formula = y ~ x | z2, data = data_100)

Residuals:
     Min        1Q    Median        3Q       Max
-2.672103 -0.632638 -0.007235  0.667518  1.778049

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.99020    0.08865  33.729  < 2e-16 ***
x           1.13489    0.14697   7.722  9.9e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8812 on 98 degrees of freedom
Multiple R-Squared: 0.8441,       Adjusted R-squared: 0.8426
Wald test: 59.63 on 1 and 98 DF,  p-value: 9.904e-12

> modelz2 <- ivreg( y ~ x | z2 , data =ivreg2)
> summary(modelz2)
```

```
Call:
ivreg(formula = y ~ x | z2, data = ivreg2)


Residuals:
     Min       1Q   Median       3Q      Max
-2.945541 -0.687447 -0.007559  0.586303  3.291229


Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.02946    0.04236   71.51   <2e-16 ***
x            1.06665    0.10136   10.52   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.945 on 498 degrees of freedom
Multiple R-Squared: 0.8207,	Adjusted R-squared: 0.8204
Wald test: 110.7 on 1 and 498 DF,  p-value: < 2.2e-16


> # ------------IV Reg z1 + Z2--------------
>
>
> library(AER)
> ivregz1z2_10 <- ivreg( y ~ x | z1 + z2 , data =data_10)
> summary(ivregz1z2_10)


Call:
ivreg(formula = y ~ x | z1 + z2, data = data_10)


Residuals:
    Min      1Q  Median      3Q     Max
-2.6876 -0.6756 -0.0865  1.1177  1.5746
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.7114     0.4337   6.252 0.000245 ***
x             1.0491     0.2549   4.116 0.003362 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.361 on 8 degrees of freedom
Multiple R-Squared: 0.8383,	Adjusted R-squared: 0.8181
Wald test: 16.94 on 1 and 8 DF,  p-value: 0.003362

> ivregz1z2_20 <- ivreg( y ~ x | z1 + z2 , data =data_20)
> summary(ivregz1z2_20)

Call:
ivreg(formula = y ~ x | z1 + z2, data = data_20)

Residuals:
     Min       1Q   Median       3Q      Max
-3.22102 -0.41855  0.08397  0.63692  1.78915

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0852     0.2555  12.074 4.57e-10 ***
x             1.0026     0.1987   5.045 8.42e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.132 on 18 degrees of freedom
Multiple R-Squared: 0.8118,	Adjusted R-squared: 0.8013
Wald test: 25.45 on 1 and 18 DF,  p-value: 8.42e-05
```

```
> ivregz1z2_100 <- ivreg( y ~ x | z1 + z2 , data =data_100)
> summary(ivregz1z2_100)

Call:
ivreg(formula = y ~ x | z1 + z2, data = data_100)

Residuals:
    Min       1Q   Median       3Q      Max
-3.15279 -0.82933  0.05788  0.70203  1.95563

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.98076    0.09974   29.88   <2e-16 ***
x            0.99206    0.09316   10.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9955 on 98 degrees of freedom
Multiple R-Squared: 0.8011,	Adjusted R-squared: 0.799
Wald test: 113.4 on 1 and 98 DF,  p-value: < 2.2e-16

> modelz1z2 <- ivreg( y ~ x | z1 + z2 , data =ivreg2)
> summary(modelz1z2)

Call:
ivreg(formula = y ~ x | z1 + z2, data = ivreg2)

Residuals:
     Min        1Q    Median        3Q       Max
-3.145058 -0.721230  0.008719  0.622161  3.497457
```

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.03113    0.04458   67.99   <2e-16 ***
x            1.00899    0.04490   22.47   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9964 on 498 degrees of freedom
Multiple R-Squared: 0.8007,       Adjusted R-squared: 0.8003
Wald test:   505 on 1 and 498 DF,  p-value: < 2.2e-16
```