Le Thi Phuong Thao – 413707007
HW0512 (11.28, 11.30, 15.06)
11.28
a) original equations

$$\text{Demand: } Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 PS_i + \alpha_4 DI_i + e_{di} \qquad (11.11)$$

$$\text{Supply: } Q_i = \beta_1 + \beta_2 P_i + \beta_3 PF_i + e_{si} \qquad (11.12)$$

Rewrite

$$Q_i - \alpha_1 - \alpha_3 PS_i - \alpha_4 DI_i - e_{di} = \alpha_2 P_i$$
$$P_i = \frac{1}{\alpha_2} Q_i - \frac{\alpha_1}{\alpha_2} - \frac{\alpha_3}{\alpha_2} PS_i - \frac{\alpha_4}{\alpha_2} DI_i - \frac{1}{\alpha_2} e_{di}$$
$$Q_i - \beta_1 - \beta_3 PF_i - e_{si} = \beta_2 P_i$$
$$P_i = \frac{1}{\beta_2} Q_i - \frac{\beta_1}{\beta_2} - \frac{\beta_3}{\beta_2} PF_i - \frac{1}{\beta_2} e_{si}$$

Anticipated Signs of the Parameters for Demand Equation
$1/\alpha 2 < 0$ since $\alpha 2 < 0$
$-\alpha 3/\alpha 2 > 0$ since $\alpha 3 > 0$ and $\alpha 2 <0$
$-\alpha 4/\alpha 2 > 0$ since $\alpha 4 > 0$ and $\alpha 2 <0$
Anticipated Signs of the Parameters for Supply Equation
$1/\beta 2 > 0$ since $\beta 2 > 0$
$-\beta 3/\beta 2 > 0$ since $\beta 3 < 0$ and $\beta 2 > 0$

b) Answer

```
Call:
ivreg(formula = p ~ q + ps + di | ps + di + pf, data = truffles)

Residuals:
    Min      1Q  Median      3Q     Max 
-39.661  -6.781   2.410   8.320  20.251 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -11.428     13.592  -0.841  0.40810    
q             -2.671      1.175  -2.273  0.03154 *  
ps             3.461      1.116   3.103  0.00458 ** 
di            13.390      2.747   4.875 4.68e-05 ***

Diagnostic tests:
                 df1 df2 statistic  p-value    
Weak instruments   1  26     17.48 0.000291 ***
Wu-Hausman         1  25    120.03 4.92e-11 ***
Sargan             0  NA        NA       NA    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.17 on 26 degrees of freedom
Multiple R-Squared: 0.5567,	Adjusted R-squared: 0.5056 
Wald test: 17.37 on 3 and 26 DF,  p-value: 2.137e-06 

Call:
ivreg(formula = p ~ q + pf | ps + di + pf, data = truffles)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.7983 -2.3440 -0.6281  2.4350 11.1600 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -58.7982     5.8592  -10.04 1.32e-10 ***
q             2.9367     0.2158   13.61 1.32e-13 ***
pf            2.9585     0.1560   18.97  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.399 on 27 degrees of freedom
Multiple R-Squared: 0.9486,	Adjusted R-squared: 0.9448 
Wald test: 232.7 on 2 and 27 DF,  p-value: < 2.2e-16 
```

Signs are same as predicted

c) Answer

From the 2SLS results:

$$P = \gamma_0 + \gamma_1 Q + \gamma_2 PS + \gamma_3 DI + u$$

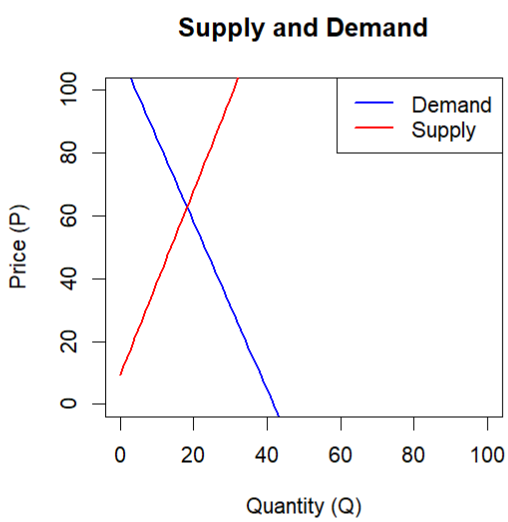
The **price elasticity of demand** at the means is:

$$E = \frac{\partial Q}{\partial P} \cdot \frac{\overline{P}}{\overline{Q}}$$

so $E = \frac{1}{\gamma_1} \cdot \frac{\overline{P}}{\overline{Q}}$

```
> # Part (c): Estimate price elasticity of demand at the means
> mean_p <- mean(truffles$p)
> mean_q <- mean(truffles$q)
> # Get the coefficient of q in the demand equation
> gamma1 <- coef(demand_2sls)["q"]
> # Calculate price elasticity of demand at the means
> elasticity <- (1 / gamma1) * (mean_p / mean_q)
> elasticity
        q
-1.272464
```

d) Answer



e) Answer

```
> cat("Equilibrium quantity:", q_eq, "\n")
Equilibrium quantity: 18.25021
> cat("Equilibrium price:", p_eq, "\n")
Equilibrium price: 62.84257
```

Difference in price predictions: 0.02719676
Difference in quantity predictions: -0.01018407

f) OLS results

```
Call:
lm(formula = p ~ q + ps + di, data = truffles)

Residuals:
     Min       1Q   Median       3Q      Max 
-25.0753  -2.7742  -0.4097   4.7079  17.4979 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -13.6195     9.0872  -1.499   0.1460    
q             0.1512     0.4988   0.303   0.7642    
ps            1.3607     0.5940   2.291   0.0303 *  
di           12.3582     1.8254   6.770 3.48e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.814 on 26 degrees of freedom
Multiple R-squared:  0.8013,	Adjusted R-squared:  0.7784 
F-statistic: 34.95 on 3 and 26 DF,  p-value: 2.842e-09
```

```
Call:
lm(formula = p ~ q + pf, data = truffles)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4721 -3.3287  0.1861  2.0785 10.7513 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -52.8763     5.0238  -10.53 4.68e-11 ***
q             2.6613     0.1712   15.54 5.42e-15 ***
pf            2.9217     0.1482   19.71  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.202 on 27 degrees of freedom
Multiple R-squared:  0.9531,	Adjusted R-squared:  0.9496 
F-statistic: 274.4 on 2 and 27 DF,  p-value: < 2.2e-16
```

Are the OLS signs "correct"?

Demand equation: the OLS slope on quantity q is positive (+0.15) even though theory (and 2SLS) says it should be negative. → Sign is wrong.

Supply equation: the OLS slope on q is positive—as theory predicts—so the sign is correct here.

Are the OLS coefficients statistically different from zero?

For demand, the q-slope is not significant ($p \approx 0.76$); the intercept is not significant either.

Only the shift variables ps and di pass the 5 % level.

For supply, all coefficients are highly significant ($p < 0.001$)

3 Comparison with part (b) (2SLS)

Feature	Demand	Supply
Direction & size of simultaneity bias	OLS pulls the slope toward +0 and even flips it to the wrong sign (downward-sloping demand becomes upward-sloping).	OLS slope remains positive but is **attenuated** (2.66 vs 2.94).
t-statistics	2SLS makes the q-slope significant; OLS does not.	Both methods yield highly significant slopes; bias mainly affects magnitude.
Economic interpretation	Ignoring endogeneity would lead one to conclude that higher quantities *raise* price—nonsense for a demand curve.	Direction is okay, but OLS understates the steepness of supply by roughly 9 %.

Quantity and price are determined simultaneously in the market => endogenity

OLS can give the bias results, we prefer 2SLS to OLS

11.30

a) Answer

```
Call:
lm(formula = i ~ p + plag + klag, data = klein)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.56562 -0.63169  0.03687  0.41542  1.49226 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.12579    5.46555   1.853 0.081374 .  
p            0.47964    0.09711   4.939 0.000125 ***
plag         0.33304    0.10086   3.302 0.004212 ** 
klag        -0.11179    0.02673  -4.183 0.000624 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.009 on 17 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9313,	Adjusted R-squared:  0.9192 
F-statistic: 76.88 on 3 and 17 DF,  p-value: 4.299e-10
```

Current Profit (p, positive): Higher current profits lead to more net investment, reflecting that profits fund investments.
Lagged Profit (plag, positive): Past profits also boost current investment, showing a delayed investment response.
Lagged Capital Stock (klag, negative): Larger existing capital stock reduces current net investment, indicating firms slow investment when capital is already high.

b) Answer

p value > 0.1% so we cannot reject H0 that coeffificients are zero

```
> linearHypothesis(profit_rf, c("g = 0", "tx = 0", "w2 = 0", "time = 0", "elag = 0"))

Linear hypothesis test:
g = 0
tx = 0
w2 = 0
time = 0
elag = 0

Model 1: restricted model
Model 2: p ~ plag + klag + g + tx + w2 + time + elag

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     18 108.04                           
2     13  61.95  5    46.093 1.9345 0.1566
```

c) Answer

```
Call:
lm(formula = i ~ p + plag + klag + .resid, data = augmented_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.04645 -0.56030  0.06189  0.25348  1.36700 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 20.27821    4.70179   4.313 0.000536 ***
p            0.15022    0.10798   1.391 0.183222    
plag         0.61594    0.10147   6.070 1.62e-05 ***
klag        -0.15779    0.02252  -7.007 2.96e-06 ***
.resid       0.57451    0.14261   4.029 0.000972 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7331 on 16 degrees of freedom
Multiple R-squared:  0.9659,	Adjusted R-squared:  0.9574 
F-statistic: 113.4 on 4 and 16 DF,  p-value: 1.588e-11
```

V_hat (.resid) is significant at 1% level, so P is endogenous

d) Answer

```
Call:
ivreg(formula = i ~ p + plag + klag | plag + klag + g + tx +
    w2 + time + elag, data = klein)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2909 -0.8069  0.1423  0.8601  1.7956

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.27821    8.38325   2.419  0.02707 *
p            0.15022    0.19253   0.780  0.44598
plag         0.61594    0.18093   3.404  0.00338 **
klag        -0.15779    0.04015  -3.930  0.00108 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.307 on 17 degrees of freedom
Multiple R-Squared: 0.8849,     Adjusted R-squared: 0.8646
Wald test:  41.2 on 3 and 17 DF,  p-value: 5.148e-08
```

And OLS

```
Call:
lm(formula = i ~ p + plag + klag, data = klein)

Residuals:
     Min       1Q   Median       3Q      Max
-2.56562 -0.63169  0.03687  0.41542  1.49226

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.12579    5.46555   1.853 0.081374 .
p            0.47964    0.09711   4.939 0.000125 ***
plag         0.33304    0.10086   3.302 0.004212 **
klag        -0.11179    0.02673  -4.183 0.000624 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.009 on 17 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9313,    Adjusted R-squared:  0.9192
F-statistic: 76.88 on 3 and 17 DF,  p-value: 4.299e-10
```

Metric	OLS (`lm`)	2SLS / IV Regression (`ivreg`)
Intercept	10.13 (p = 0.081)	20.28 (p = 0.027)
p (coefficient)	0.48 (p = 0.0001) ***	0.15 (p = 0.446) (not significant)
plag	0.33 (p = 0.0042) **	0.62 (p = 0.0034) **
klag	-0.11 (p = 0.0006) ***	-0.16 (p = 0.0011) **
R-squared	0.9313	0.8849
Adj. R-squared	0.9192	0.8646
Residual Std. Err	1.009	1.307
F/Wald test p-value	4.30e-10	5.15e-08

OLS suggests p is significant, but 2SLS shows it is not, indicating that OLS may be biased due to endogeneity. 2SLS provides more reliable inference in the presence of endogenous regressors

e) Answer

Coefficients are the same, t-value / standard error is different. T-value of 2SLS in part d is larger

```
Call:
lm(formula = i ~ .fitted + plag + klag, data = augmented_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8778 -1.0029  0.3058  0.7275  2.1831 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 20.27821    9.97663   2.033  0.05802 . 
.fitted      0.15022    0.22913   0.656  0.52084   
plag         0.61594    0.21531   2.861  0.01083 * 
klag        -0.15779    0.04778  -3.302  0.00421 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.556 on 17 degrees of freedom
Multiple R-squared:  0.837,	Adjusted R-squared:  0.8082 
F-statistic: 29.09 on 3 and 17 DF,  p-value: 6.393e-07
```

f) Answer
H0: All instruments are exogenous (uncorrelated with the error term) ⇒ instruments are valid.
H1: At least one instrument is endogenous (correlated with the error) ⇒ instruments are invalid.
critical value = 9.487729
T*R^2 = 1.281519
If T*R^2>9.488, reject H0 so we cannot reject H0 => instruments are valid

15.06

(a) The OLS estimates differ very little for the 1987 and 1988 regressions. The differences are on the order of one standard error or less. The OLS models assume that all the population parameter values, including the intercept, for all individuals are identical. The underlying assumption is that there is no heterogeneity across individuals.

(b) In equation (XR15.6) we explicitly recognize the use of panel data, by adding both individual and time period identifying subscripts, *it*, to teach observable variable. Also the unobservable error components are specified, with one random error varying across individual and time, *eit*. This is the idiosyncratic error component. The other error component, *ui*, is time-invariant, and represents unobserved heterogeneity across individuals. It is assumed not to change over the period of the sample.

(c) All the fixed effects coefficient estimates seem substantially different. Constructing rough 95% interval estimates using the fixed effects model we find the intervals (-.0085, .1235), (-.0034, .001), (-.5777, -.0745) and (.0198, .1446) for *EXPER*, $EXPER^2$, *SOUTH* and *UNION* respectively. The OLS estimates for the coefficient of *EXPER* does not fall within the fixed effects interval estimate, giving some content to the statement that the estimates show some difference. The other OLS estimates do fall within the interval estimates, suggesting that they are not “significantly” different.

(d) The *F*-statistic has numerator degrees of freedom *N* - 1= 716-1=715, because there are that many pairs of hypotheses such as $\beta_{1i} = \beta_{1,i+1}$. The denominator degrees of freedom is *NT-N-(K-1)*=1432-716-4=712 . The number of individuals must be subtracted from the total because implicitly there are that many individual indicator variables in the model to account for individual differences even if they are not shown. Equivalently the within transformation uses *N* sample means in the calculations, each "costing" a degree of freedom. Here $K = 5$ is the number of parameters showing in equation (XR15.6). The 1% critical value using Statistical Table 5 is 1.0. Using computer software the critical value is 1.1904959. Using either we reject the null hypothesis that there are no individual differences because $F = 11.68$ is larger than the critical value (F.99,715,712)

(e) Following the within transformation the random error is $\widetilde{e_{it}} = e_{it} - \bar{e}_i$. As discussed in Exercise 15.10 these transformed random errors are serially correlated when the idiosyncratic errors are uncorrelated. By using cluster robust standard errors we are allowing for heteroskedasticity in *eit* across individuals and time, and/or serial correlation across time in the *eit* . In this example the ratios of the conventional to the robust standard errors are 1.0060976, 1.0, 0.50420842, and 0.85013624 for *EXPER*, *EXPER*2, *SOUTH* and *UNION* respectively. For the experience variables the robust standard errors are virtually identical to the conventional standard errors. For *SOUTH* the robust standard error is almost twice as large as the conventional one, and for *UNION* the robust error is 1.18 times as large as the conventional standard error.

(f) The random effects estimates for the coefficients of *EXPER* and *EXPER*2 are almost twice as large as the fixed effects estimates, 1.71 times and 1.92 times respectively. The random effects coefficient of *SOUTH* is 0.71 times the fixed effects estimate, and the random effects estimate for *UNION* is 1.25 times the fixed effects estimate.
The Hausman *t*-statistics are $= \frac{0.0575-0.0986}{[0.033\quad{}^2-0.022\quad{}^2]^{1/2}}$-1.67, 1.29, -0.77 and -1.06 respectively. Only the coefficient of *EXPER* shows any significant difference, and that at the 10% level. Thus the evidence of endogeneity is weak, and a strong argument against using the random effects estimator cannot be made based on these results. We conclude that the random effects estimator is appropriate in this example.
The results are based on the conventional fixed effects standard errors because the classic Hausman test is a contrast between the standard fixed effects estimator, which is consistent, and the random effects estimator, which is efficient under all its assumptions, including homoskedasticity and lack of serial correlation in the errors.