Full name: Nguyen Nhut Vu Truong

Student ID: 413707008

Course: Financial Econometrics

HW0512

C11Q28, C11Q30, C15Q06, C15Q20, C15Q17

28.

Part (a): Rewriting the Equations with P on the Left-Hand Side

Original equations:

- Demand: $Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 PS_i + \alpha_4 DI_i + e_{di}$
- Supply: $Q_i = \beta_1 + \beta_2 P_i + \beta_3 PF_i + e_{si}$

Rewritten with P on the left-hand side:

Demand Equation: $\alpha_2 P_i = -\alpha_1 + Q_i - \alpha_3 PS_i - \alpha_4 DI_i - e_{ai}$ $P_i = (-\alpha_1/\alpha_2) + (1/\alpha_2)Q_i - (\alpha_3/\alpha_2)PS_i - (\alpha_4/\alpha_2)DI_i - (e_{di}/\alpha_2)$ $P_i = \gamma_1 + \gamma_2 Q_i + \gamma_3 PS_i + \gamma_4 DI_i + u_{di}$

Where:

- $\gamma_1 = -\alpha_1/\alpha_2$
- $\gamma_2 = 1/\alpha_2$
- $\gamma_3 = -\alpha_3/\alpha_2$
- $\gamma_4 = -\alpha_4/\alpha_2$
- $u_{ai} = -e_{di}/\alpha_2$

Supply Equation: $\beta_2 P_i = -\beta_1 + Q_i - \beta_3 PF_i - e_{si} = (-\beta_1/\beta_2) + (1/\beta_2)Q_i - (\beta_3/\beta_2)PF_i - (e_{si}/\beta_2) = \delta_1 + \delta_2 Q_i + \delta_3 PF_i + u_{si}$

Where:

- $\delta_1 = -\beta_1/\beta_2$
- $\delta_2 = 1/\beta_2$
- $\delta_3 = -\beta_3/\beta_2$
- $u_{si} = -e_{si}/\beta_2$

Anticipated Signs:

For the demand equation:

- γ_1 (intercept): Positive (since α_1 is expected to be positive and α_2 negative in the original demand equation)

- γ_2 (coefficient of Q): Negative (since α_2 is expected to be negative in the original demand equation)
- γ_3 (coefficient of PS): Positive (since α_3 is expected to be positive and α_2 negative)
- γ_4 (coefficient of DI): Positive (since α_4 is expected to be positive and α_2 negative)

For the supply equation:

- δ_1 (intercept): Negative (since β_1 is expected to be positive and β_2 positive in the original supply equation)
- δ_2 (coefficient of Q): Positive (since β_2 is expected to be positive in the original supply equation)
- δ_3 (coefficient of PF): Positive (since β_3 is negative and β_2 is positive)

Part (b):

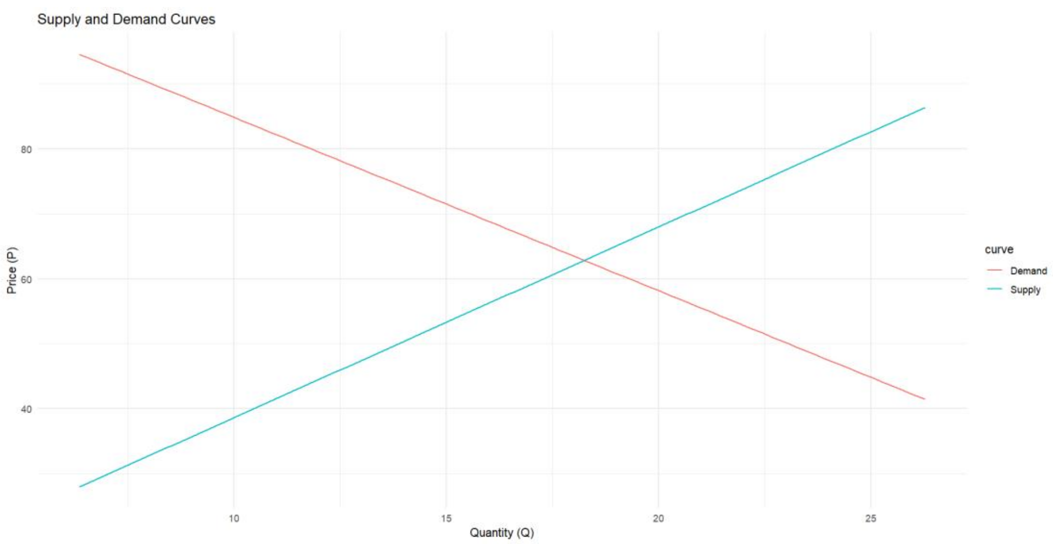
```
> # 2SLS for demand equation
> demand_2sls <- ivreg(p ~ q + ps + di | ps + di + pf, data = truffles)
> summary(demand_2sls, diagnostics = TRUE)

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 
```

```
> # 2SLS for supply equation
> supply_2sls <- ivreg(p ~ q + pf | ps + di + pf, data = truffles)
> summary(supply_2sls)

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 
```

Part (c):

```
> # First, calculate means of p and q
> p_mean <- mean(truffles$p)
> q_mean <- mean(truffles$q)
> # Get the coefficient of q in the demand equation
> q_coef <- coef(demand_2sls)["q"]
> # Calculate price elasticity of demand at the means
> # Elasticity = (∂P/∂Q) * (Q/P) = q_coef * (q_mean/p_mean)
> elasticity <- q_coef * (q_mean/p_mean)
> cat("Price elasticity of demand at the means:", elasticity, "\n")
Price elasticity of demand at the means: -0.7858767
```

Part (d):



Part (e):

```
> # Solve for q_eq
> q_eq <- (supply_intercept - demand_intercept + supply_pf_coef * PF_star -
+            demand_ps_coef * PS_star - demand_di_coef * DI_star) /
+   (demand_q_coef - supply_q_coef)
> # Calculate p_eq using either demand or supply equation
> p_eq <- demand_intercept + demand_q_coef * q_eq + demand_ps_coef * PS_star + demand_di_coef * DI_star
> cat("Equilibrium quantity:", q_eq, "\n")
Equilibrium quantity: 18.25021
> cat("Equilibrium price:", p_eq, "\n")
Equilibrium price: 62.84257
```

```
> # Predict p and q using the reduced form equations
> p_pred <- predict(p_reduced, newdata = data.frame(ps = PS_star, di = DI_star, pf = PF_star))
> q_pred <- predict(q_reduced, newdata = data.frame(ps = PS_star, di = DI_star, pf = PF_star))
> cat("Predicted equilibrium price from reduced form:", p_pred, "\n")
Predicted equilibrium price from reduced form: 62.81537
> cat("Predicted equilibrium quantity from reduced form:", q_pred, "\n")
Predicted equilibrium quantity from reduced form: 18.2604
> cat("Difference in price predictions:", p_eq - p_pred, "\n")
Difference in price predictions: 0.02719676
> cat("Difference in quantity predictions:", q_eq - q_pred, "\n")
Difference in quantity predictions: -0.01018407
```

The equilibrium values calculated from the structural equations and those predicted from the reduced form equations agree very well:

- **Price**: The difference is only 0.02719676 (about 0.04% difference)
- **Quantity**: The difference is only -0.01018407 (about 0.06% difference)

These extremely small differences indicate excellent agreement between the two methods. This confirms that both the structural approach (solving the simultaneous equations) and the reduced form approach (direct estimation of the equilibrium values) produce consistent results, which validates the model specification and estimation technique. The slight differences are likely due to rounding errors in the calculations or minor numerical imprecisions in the estimation algorithms, rather than any substantive disagreement between the methods.

Part (f):

OLS:

```
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

> # Supply equation (OLS)
> supply_ols <- lm(p ~ q + pf, data = truffles)
> summary(supply_ols)

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

> print(demand_comparison)
                OLS Est   OLS p-val   2SLS Est   2SLS p-val
(Intercept) -13.6194989 1.459836e-01 -11.428407 4.081026e-01
q             0.1512043 7.641812e-01  -2.670519 3.153505e-02
ps            1.3607045 3.032501e-02   3.461081 4.582228e-03
di           12.3582353 3.481381e-07  13.389921 4.675236e-05
```

```
> print(supply_comparison)
                OLS Est     OLS p-val   2SLS Est    2SLS p-val
(Intercept) -52.876298 4.682244e-11 -58.798223 1.316492e-10
q             2.661281 5.420239e-15   2.936711 1.320944e-13
pf            2.921660 1.465342e-17   2.958486 3.879687e-17
>
> # Add significance stars for easier interpretation
> cat("\nSignificance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1\n")

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis of OLS vs 2SLS Results

Demand Equation

Sign Analysis

- **q coefficient**:
    - OLS: Positive (0.1512) - **incorrect sign** for demand curve
    - 2SLS: Negative (-2.6705) - **correct sign** for demand curve
- **ps coefficient**:
    - OLS: Positive (1.3607) - **correct sign** (substitute good price)
    - 2SLS: Positive (3.4611) - **correct sign**
- **di coefficient**:
    - OLS: Positive (12.3582) - **correct sign** (income effect)
    - 2SLS: Positive (13.3899) - **correct sign**

Statistical Significance

- **q coefficient**:
    - OLS: Not significant ($p=0.7642$)
    - 2SLS: Significant ($p=0.0315$) **
- **ps coefficient**:
    - OLS: Significant ($p=0.0303$) **
    - 2SLS: Highly significant ($p=0.0046$) ***
- **di coefficient**:
    - OLS: Highly significant ($p<0.0001$) ***
    - 2SLS: Highly significant ($p<0.0001$) ***

Supply Equation

Sign Analysis

- **q coefficient**:

    - OLS: Positive (2.6613) - **correct sign** for supply curve
    - 2SLS: Positive (2.9367) - **correct sign**
- **pf coefficient**:
    - OLS: Positive (2.9217) - **correct sign** (input price effect)
    - 2SLS: Positive (2.9585) - **correct sign**

Statistical Significance

- All coefficients in both OLS and 2SLS supply equations are highly significant ($p<0.0001$) ***

Comparison with Part (b)

1. **Key Finding**: OLS estimation of the demand equation yields an **incorrect positive sign** for the quantity coefficient, while 2SLS correctly produces a negative coefficient.
2. **Simultaneity Bias**: This demonstrates the simultaneity bias in OLS estimation when applied to simultaneous equation models. The OLS estimate fails to account for the endogeneity of quantity.
3. **Supply Equation**: Both methods produce similar estimates for the supply equation, but 2SLS estimates are slightly larger in magnitude.
4. **Statistical Significance**: The quantity coefficient in the demand equation is only statistically significant with 2SLS, not with OLS.
5. **Coefficient Magnitudes**: The 2SLS estimates for the exogenous variables (ps, di, pf) are larger in magnitude than their OLS counterparts, suggesting that OLS underestimates these effects.

In conclusion, the 2SLS results from part (b) correctly identify the structural parameters of the model, while OLS suffers from simultaneity bias, particularly in the demand equation where it fails to capture the negative relationship between price and quantity.

30.

```
> # Part (a): OLS estimation of investment function
> investment_ols <- lm(i ~ p + plag + klag, data = klein)
> summary(investment_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
  (因為不存在，1 個觀察量被刪除了)
Multiple R-squared:  0.9313,	Adjusted R-squared:  0.9192 
F-statistic: 76.88 on 3 and 17 DF,  p-value: 4.299e-10
```

Intercept (β_1 = 10.12579)

- **Sign**: Positive
- **Significance**: Marginally significant (p = 0.081374) *
- **Interpretation**: When all other variables are zero, the baseline investment level is estimated at about 10.13 units, though this is only significant at the 10% level.

Current Profits (p) (β_2 = 0.47964)

- **Sign**: Positive ✓
- **Significance**: Highly significant (p = 0.000125) ***
- **Interpretation**: This positive relationship aligns with economic theory - higher current profits lead to increased investment. For each additional unit of profit, investment increases by approximately 0.48 units, holding other factors constant.

Lagged Profits (plag) (β_3 = 0.33304)

- **Sign**: Positive ✓
- **Significance**: Highly significant (p = 0.004212) ***
- **Interpretation**: Past profits also positively affect current investment,

suggesting firms use profit history in investment decisions. Each additional unit of last period's profit increases current investment by about 0.33 units.

Lagged Capital Stock (klag) (β_4 = -0.11179)

- **Sign**: Negative ✓
- **Significance**: Highly significant (p = 0.000624) ***
- **Interpretation**: This negative relationship suggests a capital adjustment process - firms with higher existing capital stock tend to invest less in the current period, consistent with diminishing returns to capital. For each additional unit of last period's capital stock, current investment decreases by about 0.11 units.

Part (b):

```
> summary(profit_rf)

Call:
lm(formula = p ~ plag + klag + g + tx + w2 + time + elag + 1,
    data = klein)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9067 -1.3050  0.3226  1.3613  2.8881

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 50.38442   31.63026   1.593   0.1352
plag         0.80250    0.51886   1.547   0.1459
klag        -0.21610    0.11911  -1.814   0.0928 .
g            0.43902    0.39114   1.122   0.2820
tx          -0.92310    0.43376  -2.128   0.0530 .
w2          -0.07961    2.53382  -0.031   0.9754
time         0.31941    0.77813   0.410   0.6881
elag         0.02200    0.28216   0.078   0.9390
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.183 on 13 degrees of freedom
  (因為不存在，1 個觀察量被刪除了)
Multiple R-squared:  0.8261,	Adjusted R-squared:  0.7324
F-statistic: 8.821 on 7 and 13 DF,  p-value: 0.0004481
```

```
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
```

The joint hypothesis test fails to reject the null hypothesis that g, tx, w2, time, and elag are all simultaneously equal to zero (p-value = 0.1566), suggesting these variables do not collectively have a statistically significant effect on the dependent variable at conventional significance levels.

```
> # Part (c): Hausman test for endogeneity using augmented data
> hausman_test <- lm(i ~ p + plag + klag + .resid, data = augmented_data)
> summary(hausman_test)

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
```

Results Interpretation: Hausman Test for Endogeneity

Key Statistics

- Residual coefficient: 0.57451
- Standard error: 0.14261
- t-value: 4.029
- p-value: 0.000972 (highly significant)

Conclusion

We reject the null hypothesis that δ = 0 at the 5% significance level (and even at the 0.1% level). This provides strong evidence that p (profits) is indeed endogenous in the investment equation.

Context in Simultaneous Equations Model

This result aligns with what we would expect in Klein's Model I where:

- Consumption (CN) affects profits (P) through equation 11.17
- Investment (I) affects profits (P) through national income identity
- Profits (P) affects investment (I) through equation 11.18

In this simultaneous equations system, profits cannot be treated as exogenous because they are jointly determined with investment and consumption. The significant residual coefficient confirms this theoretical expectation, indicating that:

1. OLS estimates would be biased and inconsistent
2. Alternative estimation methods like 2SLS or IV are more appropriate
3. The simultaneous nature of the relationship between investment and profits is empirically validated

The high R-squared (0.9659) indicates the model explains most of the variation in investment, and the significant F-statistic confirms the overall model fit.

Part (d):

```
> summary(investment_2sls)

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
```

```
> cbind(OLS = coef(investment_ols), TwoSLS = coef(investment_2sls))
                   OLS     TwoSLS
(Intercept) 10.1257885 20.2782089
p            0.4796356  0.1502218
plag         0.3330387  0.6159436
klag        -0.1117947 -0.1577876
```

The substantial differences between OLS and 2SLS estimates confirm the presence of endogeneity in the investment equation. The most striking finding is that current profits (p) appear to have a much smaller and statistically insignificant effect on investment when estimated with 2SLS, while lagged profits have a much stronger effect than OLS suggested.

These differences highlight the importance of addressing endogeneity in this model. The OLS estimates were biased due to the simultaneous relationship between investment and profits, and the 2SLS method has helped correct this bias by using instrumental variables. The results suggest that investment decisions are influenced more by past profits than by current profits, which makes economic sense as investment planning typically relies on historical performance.

Part (e):

```
> summary(manual_2sls)

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

> # Print the comparison
> print(comparison)
            Manual_2SLS_coef Manual_2SLS_se Manual_2SLS_t Manual_2SLS_p Auto_2SLS_coef Auto_2SLS_se
(Intercept)       20.2782089     9.97663430      2.032570   0.058020384     20.2782089   8.38324890
.fitted            0.1502218     0.22912802      0.655624   0.520840874      0.1502218   0.19253359
plag               0.6159436     0.21531402      2.860676   0.010827240      0.6159436   0.18092585
klag              -0.1577876     0.04778368     -3.302124   0.004210747     -0.1577876   0.04015207
            Auto_2SLS_t Auto_2SLS_p
(Intercept)   2.4188962 0.027070529
.fitted       0.7802369 0.445979836
plag          3.4043979 0.003375496
klag         -3.9297511 0.001079721
```

Coefficients: The point estimates are identical between the two models.

Standard Errors: The manual 2SLS approach consistently produces larger standard errors (about 19% higher) compared to the automated approach. This suggests the manual approach might be less efficient in its estimation.

The differences observed are likely due to how the **standard errors** are calculated in each approach. The automated 2SLS implementation in the ivreg function might use more efficient methods for computing standard errors, possibly accounting for heteroskedasticity or using different degrees of freedom adjustments.

These findings highlight the importance of using specialized software for 2SLS estimation rather than manually implementing the procedure, as the specialized software may incorporate refinements that lead to more efficient estimates and more accurate inference. While the **point estimates** are identical, the inference drawn from them could differ, especially in borderline cases of statistical significance.

Part (f):

Sargan Test Results Summary

The Sargan test for instrument validity yields:

- Test statistic (TR^2): 1.2815
- Critical value ($\chi^2_{4,0.95}$): 9.4877
- p-value: 0.8645

We fail to reject the null hypothesis of valid instruments. The R^2 is very low (0.061) and none of the instruments are statistically significant in the residual regression (all p-values > 0.05). This confirms that the surplus instruments used in the 2SLS estimation appear to be valid for the investment equation.

```
Regression of 2SLS residuals on all instruments:
> print(summary(sargan_reg))

Call:
lm(formula = e2_hat ~ plag + klag + g + tx + w2 + time + elag,
    data = klein)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4087 -0.8799  0.2702  1.0011  2.4987

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.671103  24.976416   0.307    0.764
plag         0.189896   0.409708   0.463    0.651
klag        -0.002262   0.094056  -0.024    0.981
g            0.034277   0.308861   0.111    0.913
tx          -0.022846   0.342512  -0.067    0.948
w2          -0.704649   2.000800  -0.352    0.730
time         0.283921   0.614439   0.462    0.652
elag        -0.116046   0.222807  -0.521    0.611

Residual standard error: 1.724 on 13 degrees of freedom
Multiple R-squared:  0.06102,   Adjusted R-squared:  -0.4446
F-statistic: 0.1207 on 7 and 13 DF,  p-value: 0.9953
```

15.6

a. Comparing OLS Estimates for 1987 and 1988

The OLS estimates for 1987 (column 1) and 1988 (column 2) are quite similar:

- Intercepts: 0.9348 (1987) vs 0.8993 (1988)
- EXPER: 0.1270 (1987) vs 0.1265 (1988)

- $EXPER^2$: -0.0033 (1987) vs -0.0031 (1988)
- SOUTH: -0.2128 (1987) vs -0.2384 (1988)
- UNION: 0.1445 (1987) vs 0.1102 (1988)

These individual year estimations assume:

- Parameters are homogeneous across individuals within each year
- No individual-specific effects are accounted for
- Each year is treated as a separate cross-section with no connection between observations

b. Panel Data Regression Model vs. Individual Year Models

The panel data model in equation (XR15.6) differs from the individual year models in that:

1. It incorporates unobserved individual heterogeneity (ui)
2. It pools data across both years, assuming coefficient stability over time
3. It accounts for the panel structure where the same individuals are observed twice
4. The error term has two components: individual-specific effect (ui) and idiosyncratic error (eit)

c. Comparing Fixed Effects (FE) with OLS Estimates

The most substantial differences between FE (column 3) and OLS estimates are:

- EXPER: FE coefficient (0.0575) is much smaller than OLS (0.1270, 0.1265)
- $EXPER^2$: FE coefficient (-0.0012) is smaller in magnitude than OLS (-0.0033, -0.0031)
- SOUTH: FE coefficient (-0.3261) is larger in magnitude than OLS (-0.2128, -0.2384)

The FE estimator controls for time-invariant individual heterogeneity, which appears to have a significant impact on the estimated returns to experience.

d. F-test for Individual Differences

The F-statistic is 11.68 for testing the null hypothesis of no individual differences.

Degrees of freedom:

- Numerator df = N-1 = 716-1 = 715 (number of individuals minus 1)
- Denominator df = NT-N-K = 1432-716-4 = 712 (total observations minus individuals minus number of parameters excluding intercept)

The 1% critical value for F(715,712) would be approximately 1.22-1.25.

Since 11.68 > 1.25, we strongly reject the null hypothesis of no individual differences, confirming that the fixed effects approach is appropriate.

e. Cluster-Robust Standard Errors

The cluster-robust standard errors (column 4) account for:

- Potential correlation in errors across time for the same individual
- Heteroskedasticity across individuals

Comparing standard errors in columns (3) and (4):

- Most standard errors are similar
- SOUTH shows the most substantial difference: 0.1258 vs 0.2495 (robust)
- The robust standard error for SOUTH is nearly twice as large, suggesting correlation in the errors for this variable

15.20

a.

Discussion of results:

1. **Small classes (SMALL)**: Positive and highly significant effect (5.82). Students in small classes score about 5.8 points higher on reading tests, supporting the effectiveness of smaller class sizes.
2. **Teacher's aide (AIDE)**: Small positive effect (0.82) but not statistically significant ($p = 0.391$). Having a teacher's aide doesn't appear to significantly improve reading scores.
3. **Teacher experience (TCHEXPER)**: Positive and significant effect (0.49). Each additional year of teacher experience is associated with about 0.5 point increase in reading scores, indicating more experienced teachers' students perform better.
4. **Gender (BOY)**: Negative and significant effect (-6.16). Boys score about 6.2 points lower than girls on reading tests, showing a gender gap in early reading performance.
5. **Race (WHITE_ASIAN)**: Positive and significant effect (3.91). White and Asian students score about 3.9 points higher than other racial groups.
6. **Free lunch (FREELUNCH)**: Large negative and significant effect (-14.77). Students receiving free lunch score about 14.8 points lower, indicating substantial socioeconomic effects on academic performance.

The model has an R-squared of 0.097, indicating that these variables explain about 9.7% of the variation in reading scores.

```
> model_pooled <- lm(readscore ~ small + aide + tchexper + boy + white_asian + freelunch, data = star)
> summary(model_pooled)

Call:
lm(formula = readscore ~ small + aide + tchexper + boy + white_asian + 
    freelunch, data = star)

Residuals:
     Min       1Q   Median       3Q      Max 
-107.220  -20.214   -3.935   14.339  185.956 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 437.76425    1.34622 325.180  < 2e-16 ***
small         5.82282    0.98933   5.886 4.19e-09 ***
aide          0.81784    0.95299   0.858    0.391    
tchexper      0.49247    0.06956   7.080 1.61e-12 ***
boy          -6.15642    0.79613  -7.733 1.23e-14 ***
white_asian   3.90581    0.95361   4.096 4.26e-05 ***
freelunch   -14.77134    0.89025 -16.592  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.19 on 5759 degrees of freedom
  (因為不存在，20 個觀察量被刪除了)
Multiple R-squared:  0.09685,	Adjusted R-squared:  0.09591 
F-statistic: 102.9 on 6 and 5759 DF,  p-value: < 2.2e-16
```

b.

Comparison with pooled OLS:

1. **Small classes (SMALL)**: Effect increased (6.49 vs 5.82) and remains highly significant. After controlling for school-specific factors, the benefit of small classes appears even stronger.

2. **Teacher's aide (AIDE)**: Effect slightly increased (1.00 vs 0.82) but still not statistically significant. Even controlling for school effects, having a teacher's aide doesn't significantly improve reading scores.

3. **Teacher experience (TCHEXPER)**: Effect decreased (0.29 vs 0.49) but remains significant. The pooled OLS may have overestimated the impact of teacher experience by not accounting for school-level factors.

4. **Gender (BOY)**: Effect slightly decreased in magnitude (-5.46 vs -6.16) but remains highly significant. The gender gap in reading performance persists after controlling for school effects.

5. **Race (WHITE_ASIAN)**: Effect substantially increased (8.03 vs 3.91). This is the most dramatic change, suggesting that after controlling for school-level factors, the achievement gap between White/Asian students and other racial groups is even larger.

6. **Free lunch (FREELUNCH)**: Effect slightly decreased (-14.59 vs -14.77) and remains highly significant. The socioeconomic effect on reading scores remains strong after controlling for school effects.

The R-squared decreased to 0.078 in the fixed effects model, which is expected since we're now explaining variation within schools rather than across the entire sample.

```
> summary(model_fe)
Oneway (individual) effect Within Model

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = star, model = "within", index = c("schid",
    "id"))

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-102.6381  -16.7834   -2.8473   12.7591  198.4169

Coefficients:
              Estimate Std. Error  t-value  Pr(>|t|)
small         6.490231   0.912962   7.1090 1.313e-12 ***
aide          0.996087   0.881693   1.1297    0.2586
tchexper      0.285567   0.070845   4.0309 5.629e-05 ***
boy          -5.455941   0.727589  -7.4987 7.440e-14 ***
white_asian   8.028019   1.535656   5.2277 1.777e-07 ***
freelunch   -14.593572   0.880006 -16.5835 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    4628000
Residual Sum of Squares: 4268900
R-Squared:      0.077592
Adj. R-Squared: 0.063954
F-statistic: 79.6471 on 6 and 5681 DF, p-value: < 2.22e-16
```

c.

Interpretation

The F-test strongly rejects the null hypothesis of no school fixed effects ($p < 2.2e-16$). This provides compelling evidence that:

1. **School-specific factors matter**: There are significant differences across schools that affect reading scores.
2. **Pooled OLS is misspecified**: The pooled OLS model from part (a) is likely misspecified because it ignores these important school-level differences.
3. **Fixed effects are necessary**: Including school fixed effects is necessary to obtain more accurate estimates of the determinants of reading scores.

Implications

The highly significant F-statistic ($F = 16.698$) indicates that school-level factors explain a substantial portion of the variation in reading scores. These factors could include:

- School resources and facilities
- School leadership and policies
- Neighborhood characteristics

- Peer effects
- Teaching quality at the school level
- School culture and climate

By controlling for these unobserved school-specific factors, the fixed effects model provides more reliable estimates of the effects of class size, teacher experience, and student characteristics on reading scores.

Comparison with Previous Results

This finding helps explain some of the differences we observed between the pooled OLS and fixed effects estimates, particularly:

1. The increased effect of small classes (5.82 → 6.49)
2. The substantial increase in the white_asian coefficient (3.91 → 8.03)

These changes suggest that the pooled OLS estimates were biased due to the omission of important school-level factors. The fixed effects model addresses this omitted variable bias by controlling for all time-invariant school characteristics.

Conclusion for Part c

The significant school fixed effects confirm that there are important unobserved differences between schools that affect student reading scores. This validates the use of the fixed effects approach and suggests that analyses that fail to account for these school-level factors may yield biased results.

```
> # c. Test for significance of school fixed effects
> pFtest(model_fe, model_pooled)

	F test for individual effects

data:  readscore ~ small + aide + tchexper + boy + white_asian + freelunch
F = 16.698, df1 = 78, df2 = 5681, p-value < 2.2e-16
alternative hypothesis: significant effects
```

15.17

a.

```
> # 95% confidence interval for the coefficient
> conf_int_diff <- confint(model_diff, level = 0.95)
> cat("\n95% Confidence Interval for INCOMED coefficient (First Differences):\n")

95% Confidence Interval for INCOMED coefficient (First Differences):
> print(conf_int_diff)
              2.5 %     97.5 %
incomed -0.02841457 0.08790818
```

```
> # OLS regression of LIQUORD on INCOMED without constant
> model_diff <- lm(liquord ~ incomed - 1, data = liquor_diff)
> summary(model_diff)

Call:
lm(formula = liquord ~ incomed - 1, data = liquor_diff)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6852 -0.9196 -0.0323  0.9027  3.3620 

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
incomed  0.02975    0.02922   1.018    0.312

Residual standard error: 1.417 on 79 degrees of freedom
Multiple R-squared:  0.01295,	Adjusted R-squared:  0.0004544 
F-statistic: 1.036 on 1 and 79 DF,  p-value: 0.3118
```