Full name: Nguyen Nhut Vu Truong

Student ID: 413707008

Course: Financial Econometrics

HW0505

C11Q01, C11Q16, C11Q17, C11Q28, C11Q30

1.

We have the simultaneous equations model:

$y_1 = \alpha_1 y_2 + e_1$

$y_2 = \alpha_2 y_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

Where x_1 and x_2 are exogenous and uncorrelated with error terms e_1 and e_2.

Part (a): Deriving the Reduced-Form Equation for y_2

To find the reduced-form equation for y_2, I need to substitute the first equation into the second.

First, I'll substitute the expression for y_1 into the second equation:

$y_2 = \alpha_2(\alpha_1 y_2 + e_1) + \beta_1 x_1 + \beta_2 x_2 + e_2$

Expanding:

$y_2 = \alpha_2\alpha_1 y_2 + \alpha_2 e_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

Rearranging to isolate y_2:

$y_2 - \alpha_2\alpha_1 y_2 = \alpha_2 e_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

$y_2(1 - \alpha_2\alpha_1) = \alpha_2 e_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

$y_2 = (\beta_1 x_1 + \beta_2 x_2 + \alpha_2 e_1 + e_2)/(1 - \alpha_2\alpha_1)$

This is the reduced-form equation for y_2, which we can write as:

$y_2 = \pi_1 x_1 + \pi_2 x_2 + v_2$

Where:

$\pi_1 = \beta_1/(1 - \alpha_2\alpha_1)$

$\pi_2 = \beta_2/(1 - \alpha_2\alpha_1)$

$v_2 = (\alpha_2 e_1 + e_2)/(1 - \alpha_2\alpha_1)$

Now, to show that y_2 is correlated with e_1, I'll examine the covariance:

$Cov(y_2, e_1) = Cov(\pi_1 x_1 + \pi_2 x_2 + v_2, e_1)$

Since x_1 and x_2 are exogenous and uncorrelated with e_1:

$Cov(y_2, e_1) = Cov(v_2, e_1)$

Substituting the expression for v_2:

$Cov(y_2, e_1) = Cov((\alpha_2 e_1 + e_2)/(1 - \alpha_2\alpha_1), e_1)$

Since e_2 is uncorrelated with e_1:

$Cov(y_2, e_1) = \alpha_2/(1 - \alpha_2\alpha_1) \times Cov(e_1, e_1) = \alpha_2/(1 - \alpha_2\alpha_1) \times Var(e_1)$

Since $Var(e_1) > 0$ and assuming $1 - \alpha_2\alpha_1 \neq 0$, we have $Cov(y_2, e_1) \neq 0$, which proves that y_2 is correlated with e_1.

Part (b): OLS Consistency

For OLS to consistently estimate parameters, the explanatory variables must be uncorrelated with the error term.

In the first equation: $y_1 = \alpha_1 y_2 + e_1$

y_2 is correlated with e_1 as shown in part (a)

Therefore, OLS estimation of α_1 will be inconsistent

In the second equation: $y_2 = \alpha_2 y_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

y_1 contains e_1 (from first equation), which affects y_2

y_1 is correlated with e_2 (can be shown similarly to part (a))

Therefore, OLS estimation of α_2 will be inconsistent

However, if we could somehow isolate the effects, β_1 and β_2 could be consistently estimated since x_1 and x_2 are exogenous

Part (c): Identification

In a system of M simultaneous equations, at least (M-1) variables must be absent from each equation for it to be identified.

Here, M = 2, so at least 1 variable must be absent from each equation.

For the first equation: $y_1 = \alpha_1 y_2 + e_1$

Both exogenous variables x_1 and x_2 are absent

Therefore, this equation is identified (over-identified since 2 > M-1)

For the second equation: $y_2 = \alpha_2 y_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

No exogenous variables are absent

Therefore, this equation is not identified (under-identified)

In summary: The first equation's parameter α_1 is identified and can be consistently estimated using 2SLS. The second equation's parameters α_2, β_1, and β_2 are not identified and cannot be consistently estimated using any method.

(d) These moment conditions arise from the assumptions that the x's are exogenous. It follows that

$$E(x_{i1}v_{i1}\mid \mathbf{x}) = E(x_{i2}v_{i2}\mid \mathbf{x}) = 0$$

From part (a), the reduced form equation for y_2 is

$$y_2 = \frac{\beta_1}{(1-\alpha_1\alpha_2)}x_1 + \frac{\beta_2}{(1-\alpha_1\alpha_2)}x_2 + \frac{e_2+\alpha_2 e_1}{(1-\alpha_1\alpha_2)} = \pi_1 x_1 + \pi_2 x_2 + v_2$$

The reduced form error is uncorrelated with the x's because

$$E\left[x_{ik}\left(\frac{e_2+\alpha_2 e_1}{(1-\alpha_1\alpha_2)}\right)\middle|\,\mathbf{x}\right] = E\left[\frac{1}{(1-\alpha_1\alpha_2)}x_{ik}e_2\middle|\,\mathbf{x}\right] + E\left[\frac{\alpha_2}{(1-\alpha_1\alpha_2)}x_{ik}e_1\middle|\,\mathbf{x}\right] = 0+0$$

(e) The sum of squares function, omitting the subscript i for convenience, is $S(\pi_1,\pi_2\mid \mathbf{y},\mathbf{x}) = \sum(y_2-\pi_1x_1-\pi_2x_2)^2$. The first derivatives are

$$\frac{\partial S(\pi_1,\pi_2\mid \mathbf{y},\mathbf{x})}{\partial \pi_1} = 2\sum(y_2-\pi_1x_1-\pi_2x_2)x_1 = 0$$

$$\frac{\partial S(\pi_1,\pi_2\mid \mathbf{y},\mathbf{x})}{\partial \pi_2} = 2\sum(y_2-\pi_1x_1-\pi_2x_2)x_2 = 0$$

Divide these equations by 2, and multiply the moment equations by N to see that they are equivalent.

(f) The moment conditions are

$$N^{-1}\sum x_{i1}(y_2-\pi_1x_{i1}-\pi_2x_{i2}) = 0$$

$$N^{-1}\sum x_{i2}(y_2-\pi_1x_{i1}-\pi_2x_{i2}) = 0$$

Multiplying these out we have

$$\sum x_{i1}y_{i2} - \pi_1\sum x_{i1}^2 - \pi_2\sum x_{i1}x_{i2} = 0$$

$$\sum x_{i2}y_{i2} - \pi_1\sum x_{i1}x_{i2} - \pi_2\sum x_{i2}^2 = 0$$

Inserting the given values, we have

$$3-\hat{\pi}_1 = 0 \Rightarrow \hat{\pi}_1 = 3$$

$$4-\hat{\pi}_2 = 0 \Rightarrow \hat{\pi}_2 = 4$$

(g) The first structural equation is $y_1 = \alpha_1 y_2 + e_1$, so that

$$E\left[(\pi_1x_1+\pi_2x_2)e_1\mid \mathbf{x}\right] = E\left[(\pi_1x_1+\pi_2x_2)(y_1-\alpha_1y_2)\mid \mathbf{x}\right] = 0$$

The empirical analog of this moment condition is

$$N^{-1}\sum(\pi_1 x_{i1}+\pi_2 x_{i2})(y_{i1}-\alpha_1 y_{i2})=0$$

If we knew π_1 and π_2 we could solve this moment condition for an estimator of α_1. While we do not know these parameters we can consistently estimate them from the reduced form equations. In large samples the consistent estimators converge to the true parameter values,

$$\operatorname{plim}\hat{\pi}_1=\pi_1 \text{ and } \operatorname{plim}\hat{\pi}_2=\pi_2$$

In a sense, having consistent estimators of parameters is "just as good as" knowing the parameter values. Replacing the unknowns by their estimates in the empirical moment condition we have

$$\sum(\hat{\pi}_1 x_{i1}+\hat{\pi}_2 x_{i2})(y_{i1}-\alpha_1 y_{i2})=\sum\hat{y}_{i2}(y_{i1}-\alpha_1 y_{i2})=0$$

So that

$$\sum\hat{y}_{i2}y_{i1}-\alpha_1\sum\hat{y}_{i2}y_{i2}=0\Rightarrow\hat{\alpha}_{1,IV}=\frac{\sum\hat{y}_{i2}y_{i1}}{\sum\hat{y}_{i2}y_{i2}}$$

Inserting the values, we find

$$\hat{\alpha}_{1,IV}=\frac{\sum\hat{y}_{i2}y_{i1}}{\sum\hat{y}_{i2}y_{i2}}=\frac{\sum(\hat{\pi}_1 x_{i1}+\hat{\pi}_2 x_{i2})y_{i1}}{\sum(\hat{\pi}_1 x_{i1}+\hat{\pi}_2 x_{i2})y_{i2}}=\frac{\hat{\pi}_1\sum x_{i1}y_{i1}+\hat{\pi}_2\sum x_{i2}y_{i1}}{\hat{\pi}_1\sum x_{i1}y_{i2}+\hat{\pi}_2\sum x_{i2}y_{i2}}=\frac{18}{25}$$

(h) The least squares estimator of the simple regression model with no intercept is given in Exercise 2.4. Applying that result here, and substituting $\hat{y}_2$ for *x* and y_1 for *y*, we have

$$\hat{\alpha}_{1,2SLS}=\frac{\sum\hat{y}_{i2}y_{i1}}{\sum\hat{y}_{i2}^2}$$

To show that the equations are equivalent, recall that $\hat{v}_2=y_2-\hat{y}_2\Rightarrow\hat{y}_2=y_2-\hat{v}_2$ and therefore

$$\sum\hat{y}_{i2}^2=\sum\hat{y}_{i2}(y_2-\hat{v}_2)=\sum\hat{y}_{i2}y_2-\sum\hat{y}_{i2}\hat{v}_2=\sum\hat{y}_{i2}y_2$$

The term

$$\sum\hat{y}_{i2}\hat{v}_{i2}=\sum(\hat{\pi}_1 x_{i1}+\hat{\pi}_2 x_{i2})\hat{v}_{i2}=\hat{\pi}_1\sum x_{i1}\hat{v}_{i2}+\hat{\pi}_2\sum x_{i2}\hat{v}_{i2}=0$$

because $\sum x_{i1}\hat{v}_{i2}=0$ and $\sum x_{i2}\hat{v}_{i2}=0$. This is a fundamental property of OLS that is illustrated in Exercises 2.3(f) and 2.4(g).

16.

Part (a): Deriving Reduced-Form Equations

Given model:

- Demand: Qi = α1 + **α2**Pi + edi
- Supply: Qi = β1 + β2Pi + β3Wi + esi

Setting demand equal to supply in equilibrium: α1 + α2Pi + edi = β1 + β2Pi + β3Wi + esi

Solving for Pi: Pi(α2 - β2) = β1 - α1 + β3Wi + esi - edi Pi = (β1 - α1 + β3Wi + esi -

$edi)/(\alpha 2 - \beta 2)$

This gives the reduced-form equation for P: $Pi = \pi 1 + \pi 2Wi + v1i$

Where:

- $\pi 1 = (\beta 1 - \alpha 1)/(\alpha 2 - \beta 2)$
- **$\pi 2 = \beta 3/(\alpha 2 - \beta 2)$**
- $v1i = (esi - edi)/(\alpha 2 - \beta 2)$

Substituting this expression for Pi into the demand equation: $Qi = \alpha 1 + \alpha 2[(\beta 1 - \alpha 1 + \beta 3Wi + esi - edi)/(\alpha 2 - \beta 2)] + edi$

After simplification: $Qi = (\alpha 2\beta 1 - \alpha 1\beta 2)/(\alpha 2 - \beta 2) + [\alpha 2\beta 3/(\alpha 2 - \beta 2)]Wi + (\alpha 2esi - \beta 2edi)/(\alpha 2 - \beta 2)$

This gives the reduced-form equation for Q: $Qi = \theta 1 + \theta 2Wi + v2i$

Where:

- $\theta 1 = (\alpha 2\beta 1 - \alpha 1\beta 2)/(\alpha 2 - \beta 2)$
- **$\theta 2 = \alpha 2\beta 3/(\alpha 2 - \beta 2)$**
- $v2i = (\alpha 2esi - \beta 2edi)/(\alpha 2 - \beta 2)$

Part (b): Identifying Structural Parameters

From the reduced-form parameters:

- We can identify $\alpha 2 = \theta 2/\pi 2 = (\alpha 2\beta 3/(\alpha 2-\beta 2))/(\beta 3/(\alpha 2-\beta 2)) = \alpha 2$

The demand equation is identified because:

- The wage rate W appears in the supply equation but not in the demand equation
- This satisfies the order condition for identification (at least M-1 variables absent)

The supply equation is not identified because:

- There are no exogenous variables that appear in the demand equation but not in the supply equation

Part (c): Indirect Least Squares

Given estimated reduced-form equations:

- $\hat{Q} = 5 + 0.5W$
- $\hat{P} = 2.4 + 1W$

We can identify:

- $\alpha 2 = \theta 2/\pi 2 = 0.5/1 = 0.5$

For $\alpha 1$ (intercept in demand equation):

- $\alpha 1 = 5 - 0.5(2.4) = 5 - 1.2 = 3.8$

Therefore, the identified demand equation is: $Q = 3.8 + 0.5\hat{P} + e_d$

Part (d)

Step 1: Calculate fitted values from the reduced-form equation for P

Using the estimated reduced-form equation for P: $\hat{P} = 2.4 + 1W$

For each observation in the data:

Observation	W	$\hat{P} = 2.4 + 1W$
1	2	$2.4 + 1(2) = 4.4$
2	3	$2.4 + 1(3) = 5.4$
3	1	$2.4 + 1(1) = 3.4$
4	1	$2.4 + 1(1) = 3.4$
5	3	$2.4 + 1(3) = 5.4$

Step 2: Apply 2SLS to estimate the demand equation

The demand equation is: $Q = \alpha_1 + \alpha_2 P + e_d$

For 2SLS, we replace P with $\hat{P}$ in the second stage: $Q = \alpha_1 + \alpha_2\hat{P} + e_d$

Creating a table with the original Q and the fitted $\hat{P}$:

Observation	Q	$\hat{P}$
1	4	4.4
2	6	5.4
3	9	3.4
4	3	3.4
5	8	5.4

Running the regression of Q on $\hat{P}$ gives us: $Q = 3.8 + 0.5\hat{P} + e_d$

Verification using the data: To verify this result, we can calculate the predicted Q values using our 2SLS equation:

Observation	$\hat{P}$	$\hat{Q} = 3.8 + 0.5\hat{P}$	Actual Q	Residual
1	4.4	3.8 + 0.5(4.4) = 6.0	4	-2.0
2	5.4	3.8 + 0.5(5.4) = 6.5	6	-0.5
3	3.4	3.8 + 0.5(3.4) = 5.5	9	3.5
4	3.4	3.8 + 0.5(3.4) = 5.5	3	-2.5
5	5.4	3.8 + 0.5(5.4) = 6.5	8	1.5

Comparison with OLS: For comparison, if we had incorrectly used OLS directly (regressing Q on P), we would get: $Q = 4.75 + 0.28P$

This differs from our 2SLS estimate because OLS doesn't account for the endogeneity between P and Q in the simultaneous equations system.

Conclusion:

The 2SLS estimate of the demand equation is: $Q = 3.8 + 0.5\hat{P} + e_d$

This matches our result from the indirect least squares method in part (c), confirming that both methods yield the same estimates for the identified equation when applied correctly.

17.

(a) There are M = 8 equations requiring 7 omitted variables in each equation. There is a total of 16 variables in the system. The consumption equation includes 6 variables and omits 10. The necessary condition is satisfied. The investment equation includes 5 variables and omits 11. The necessary condition is satisfied. The private sector wage equation includes 5 variables and omits 11. The necessary condition is satisfied.

(b) The consumption equation has 2 RHS endogenous variables and excludes 5 exogenous variables. The investment and private wage equations have 1 RHS endogenous variable and omit 5 exogenous variables.

(c) $$W_{1t} = \pi_1 + \pi_2 G_t + \pi_3 W_{2t} + \pi_4 TX_t + \pi_5 TIME_t + \pi_6 P_{t-1} + \pi_7 K_{t-1} + \pi_8 E_{t-1} + v$$

(d) Obtain fitted values $\hat{W}_{1t}$ from the estimated reduced form equation in part (c) and similarly obtain $\hat{P}_t$. Create $W_t^* = \hat{W}_{1t} + W_{2t}$. Regress CN_t on W_t^*, $\hat{P}_t$ and P_{t-1} plus a constant by OLS.

(e) The coefficient estimates will be the same. The t-values will not be because the standard errors in part (d) are not correct 2SLS standard errors.

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):

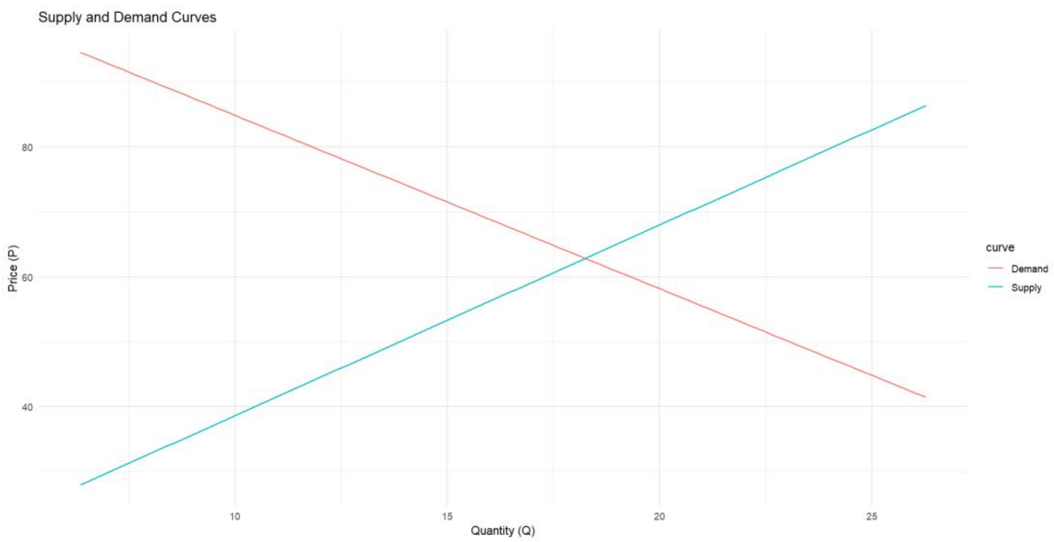
```
> # 2SLS for demand equation
> demand_2sls <- ivreg(p ~ q + ps + di | ps + di + pf, data = truffles)
> summary(demand_2sls, diagnostics = TRUE)

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 $\delta = 0$ 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
```