HW0505 – Pinyo 312712017

11.1 Our aim is to estimate the parameters of the simultaneous equations model

$$y_1 = \alpha_1 y_2 + e_1$$
$$y_2 = \alpha_2 y_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$$

We assume that x_1 and x_2 are exogenous and uncorrelated with the error terms e_1 and e_2.

a. Solve the two structural equations for the reduced-form equation for y_2, that is, $y_2 = \pi_1 x_1 + \pi_2 x_2 + v_2$. Express the reduced-form parameters in terms of the **structural parameters** and the reduced-form error in terms of the structural parameters and e_1 and e_2. Show that y_2 is correlated with e_1.

Given the structural equations:

$$y_1 = \alpha_1 y_2 + e_1 \quad (1)$$
$$y_2 = \alpha_2 y_1 + \beta_1 x_1 + \beta_2 x_2 + e_2 \quad (2)$$

Substitute equation (1) into equation (2):

$$\begin{aligned} y_2 &= \alpha_2(\alpha_1 y_2 + e_1) + \beta_1 x_1 + \beta_2 x_2 + e_2 \\ &= \alpha_1\alpha_2 y_2 + \alpha_2 e_1 + \beta_1 x_1 + \beta_2 x_2 + e_2 \end{aligned}$$

Bring terms involving y_2 to one side:

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

Factor y_2:

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

Divide both sides:

$$y_2 = \frac{\beta_1}{1 - \alpha_1\alpha_2} x_1 + \frac{\beta_2}{1 - \alpha_1\alpha_2} x_2 + \frac{\alpha_2 e_1 + e_2}{1 - \alpha_1\alpha_2}$$

Define:

$$\pi_1 = \frac{\beta_1}{1 - \alpha_1\alpha_2}, \quad \pi_2 = \frac{\beta_2}{1 - \alpha_1\alpha_2}, \quad v_2 = \frac{\alpha_2 e_1 + e_2}{1 - \alpha_1\alpha_2}$$

Thus, the reduced-form equation is:

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

Correlation between y_2 and e_1:

Since v_2 includes e_1, and y_2 depends on v_2, y_2 is correlated with e_1. This implies that y_2 is endogenous in equation (1).

b. Which equation parameters are consistently estimated using OLS? Explain.

Explanation:

- **Equation (1):** $y_1 = \alpha_1 y_2 + e_1$

  Here, y_2 is endogenous (as shown above). OLS estimation would be inconsistent due to the correlation between y_2 and e_1.

- **Equation (2):** $y_2 = \alpha_2 y_1 + \beta_1 x_1 + \beta_2 x_2 + e_2$

  Similarly, y_1 is endogenous. OLS estimation would be inconsistent due to the correlation between y_1 and e_2.

Therefore, OLS does not provide consistent estimators for either structural equation due to endogeneity.

---

c. Identification of Parameters:

For identification in simultaneous equations:

- **Order Condition:** The number of excluded exogenous variables from an equation must be at least equal to the number of included endogenous variables.
- **Equation (1):** Excludes x_1 and x_2 (2 variables), includes y_2 (1 variable). Satisfies order condition.
- **Equation (2):** Excludes none of the exogenous variables, includes y_1 (1 variable). Does not satisfy order condition.

Thus, equation (1) is identified; equation (2) is not identified.

c. Which parameters are "identified," in the simultaneous equations sense? Explain your reasoning.

c. Identification of Parameters:

For identification in simultaneous equations:

- **Order Condition:** The number of excluded exogenous variables from an equation must be at least equal to the number of included endogenous variables.
- **Equation (1):** Excludes x_1 and x_2 (2 variables), includes y_2 (1 variable). Satisfies order condition.
- **Equation (2):** Excludes none of the exogenous variables, includes y_1 (1 variable). Does not satisfy order condition.

Thus, equation (1) is identified; equation (2) is not identified.

d. To estimate the parameters of the reduced-form equation for y_2 using the method of moments (MOM), which was introduced in Section 10.3, the two moment equations are

$$N^{-1}\sum x_{i1}(y_2 - \pi_1 x_{i1} - \pi_2 x_{i2}) = 0$$
$$N^{-1}\sum x_{i2}(y_2 - \pi_1 x_{i1} - \pi_2 x_{i2}) = 0$$

Explain why these two moment conditions are a valid basis for obtaining consistent estimators of the reduced-form parameters.

The moment conditions are:

$$\frac{1}{N}\sum x_{i1}(y_{2i} - \pi_1 x_{i1} - \pi_2 x_{i2}) = 0$$
$$\frac{1}{N}\sum x_{i2}(y_{2i} - \pi_1 x_{i1} - \pi_2 x_{i2}) = 0$$

Since x_1 and x_2 are exogenous and uncorrelated with the error term v_2, these moment conditions are valid. They equate the sample moments to zero, allowing consistent estimation of π_1 and π_2.

e. Are the MOM estimators in part (d) the same as the OLS estimators? Form the sum of squared errors function for $y_2 = \pi_1 x_1 + \pi_2 x_2 + v_2$ and find the first derivatives. Set these to zero and show that they are equivalent to the two equations in part (d).

The sum of squared errors (SSE) function is:

$$SSE = \sum (y_{2i} - \pi_1 x_{i1} - \pi_2 x_{i2})^2$$

Taking derivatives with respect to π_1 and π_2, setting them to zero, yields the normal equations:

$$\sum x_{i1}(y_{2i} - \pi_1 x_{i1} - \pi_2 x_{i2}) = 0$$
$$\sum x_{i2}(y_{2i} - \pi_1 x_{i1} - \pi_2 x_{i2}) = 0$$

These are identical to the moment conditions in part (d), confirming that MOM and OLS estimators are the same in this context.

f. Using $\sum x_{i1}^2 = 1$, $\sum x_{i2}^2 = 1$, $\sum x_{i1}x_{i2} = 0$, $\sum x_{i1}y_{1i} = 2$, $\sum x_{i1}y_{2i} = 3$, $\sum x_{i2}y_{1i} = 3$, $\sum x_{i2}y_{2i} = 4$, and the two moment conditions in part (d) show that the MOM/OLS estimates of π_1 and π_2 are $\hat{\pi}_1 = 3$ and $\hat{\pi}_2 = 4$.

Given:

$$\sum x_{i1}^2 = 1, \quad \sum x_{i2}^2 = 1, \quad \sum x_{i1}x_{i2} = 0$$
$$\sum x_{i1}y_{2i} = 3, \quad \sum x_{i2}y_{2i} = 4$$

Using the normal equations:

$$\pi_1 \sum x_{i1}^2 + \pi_2 \sum x_{i1}x_{i2} = \sum x_{i1}y_{2i}$$
$$\pi_1 \sum x_{i1}x_{i2} + \pi_2 \sum x_{i2}^2 = \sum x_{i2}y_{2i}$$

Substitute the given values:

$$\pi_1(1) + \pi_2(0) = 3 \Rightarrow \pi_1 = 3$$
$$\pi_1(0) + \pi_2(1) = 4 \Rightarrow \pi_2 = 4$$

Thus, $\hat{\pi}_1 = 3$, $\hat{\pi}_2 = 4$.

g. The fitted value $\hat{y}_2 = \hat{\pi}_1 x_1 + \hat{\pi}_2 x_2$. Explain why we can use the moment condition $\sum \hat{y}_{i2}(y_{i1} - \alpha_1 y_{i2}) = 0$ as a valid basis for consistently estimating α_1. Obtain the IV estimate of α_1.

The fitted values from the reduced-form equation for y_2 are:

$$\hat{y}_{2i} = \hat{\pi}_1 x_{i1} + \hat{\pi}_2 x_{i2}$$

Since $\hat{y}_{2i}$ is a function of exogenous variables, it is uncorrelated with the error term e_1 in equation (1). Therefore, we can use the moment condition:

$$\sum \hat{y}_{2i}(y_{1i} - \alpha_1 y_{2i}) = 0$$

Solving for α_1:

$$\alpha_1 = \frac{\sum \hat{y}_{2i} y_{1i}}{\sum \hat{y}_{2i} y_{2i}}$$

This provides a consistent IV estimate of α_1.

h. Find the 2SLS estimate of α_1 by applying OLS to $y_1 = \alpha_1 \hat{y}_2 + e_1^*$. Compare your answer to that in part (g).

First Stage:

Regress y_2 on x_1 and x_2 to obtain $\hat{y}_2$.

Second Stage:

Regress y_1 on $\hat{y}_2$:

$$y_1 = \alpha_1 \hat{y}_2 + e_1^*$$

The estimator $\hat{\alpha}_1$ is:

$$\hat{\alpha}_1 = \frac{\sum \hat{y}_{2i} y_{1i}}{\sum \hat{y}_{2i}^2}$$

This 2SLS estimate may differ from the IV estimate in part (g) because the denominators are different: one uses $\sum \hat{y}_{2i} y_{2i}$, the other uses $\sum \hat{y}_{2i}^2$. However, both estimators are consistent for α_1.

11.16 Consider the following supply and demand model

$$\text{Demand: } Q_i = \alpha_1 + \alpha_2 P_i + e_{di}, \quad \text{Supply: } Q_i = \beta_1 + \beta_2 P_i + \beta_3 W_i + e_{si}$$

where Q is the quantity, P is the price, and W is the wage rate, which is assumed exogenous. Data on these variables are in Table 11.7.

TABLE 11.7 **Data for Exercise 11.16**

Q	P	W
4	2	2
6	4	3
9	3	1
3	5	1
8	8	3

a. Derive the algebraic form of the reduced-form equations, $Q = \theta_1 + \theta_2 W + v_2$ and $P = \pi_1 + \pi_2 W + v_1$, expressing the reduced-form parameters in terms of the structural parameters.

We solve the system of equations for Q and P in terms of the exogenous variable W.

From the **supply and demand equations**, equating both expressions for Q:

$$\alpha_1 + \alpha_2 P_i + e_{di} = \beta_1 + \beta_2 P_i + \beta_3 W_i + e_{si}$$

Solving for P_i:

$$(\alpha_2 - \beta_2) P_i = \beta_1 - \alpha_1 + \beta_3 W_i + e_{si} - e_{di}$$

$$P_i = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{\beta_3}{\alpha_2 - \beta_2} W_i + \frac{e_{si} - e_{di}}{\alpha_2 - \beta_2}$$

This gives the **reduced-form equation for price**:

$$P_i = \pi_1 + \pi_2 W_i + v_{1i}$$

Where:

- $\pi_1 = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2}$
- $\pi_2 = \frac{\beta_3}{\alpha_2 - \beta_2}$
- $v_1 = \frac{e_s - e_d}{\alpha_2 - \beta_2}$

Substitute this back into the **demand equation** to get Q in terms of W:

$$Q_i = \alpha_1 + \alpha_2 P_i + e_{di} = \alpha_1 + \alpha_2(\pi_1 + \pi_2 W_i + v_1) + e_{di}$$

$$Q_i = (\alpha_1 + \alpha_2 \pi_1) + \alpha_2 \pi_2 W_i + (\alpha_2 v_1 + e_{di})$$

This gives the **reduced-form equation for quantity**:

$$Q_i = \theta_1 + \theta_2 W_i + v_2$$

Where:

- $\theta_1 = \alpha_1 + \alpha_2 \pi_1$
- $\theta_2 = \alpha_2 \pi_2$
- $v_2 = \alpha_2 v_1 + e_d$

b. Which structural parameters can you solve for from the results in part (a)? Which equation is "identified"?

From reduced-form coefficients $\pi_1, \pi_2, \theta_1, \theta_2$, we want to identify the structural parameters.

Let's see:

From:

- $\pi_1 = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2}$
- $\pi_2 = \frac{\beta_3}{\alpha_2 - \beta_2}$

Then:

- $\alpha_1 + \alpha_2 \pi_1 = \theta_1$
- $\alpha_2 \pi_2 = \theta_2$

So:

- We can solve for $\alpha_2 = \frac{\theta_2}{\pi_2}$
- Plug this into $\theta_1 = \alpha_1 + \alpha_2 \pi_1$ to find α_1

Therefore, we **can identify the demand equation** (α_1, α_2), but **not** the supply parameters β_1, β_2, β_3 (they remain entangled in the reduced-form).

c. The estimated reduced-form equations are $Q = 5 + 0.5W$ and $P = 2.4 + 1W$. Solve for the identified structural parameters. This is the method of **indirect least squares**.

We're given:

- $\hat{Q}_i = 5 + 0.5W_i \Rightarrow \theta_1 = 5, \theta_2 = 0.5$
- $\hat{P}_i = 2.4 + 1W_i \Rightarrow \pi_1 = 2.4, \pi_2 = 1$

Then:

$$\alpha_2 = \frac{\theta_2}{\pi_2} = \frac{0.5}{1} = 0.5$$

$$\alpha_1 = \theta_1 - \alpha_2 \pi_1 = 5 - (0.5)(2.4) = 5 - 1.2 = 3.8$$

d. Obtain the fitted values from the reduced-form equation for P, and apply 2SLS to obtain estimates of the demand equation.

We want to estimate:

$$Q_i = \alpha_1 + \alpha_2 P_i + e_{di}$$

But since P_i is endogenous, we use **instrumental variables** via **2SLS**.

Step 1 (First Stage): Regress P_i on exogenous variable W_i

Given:

$$\hat{P}_i = 2.4 + 1 \cdot W_i$$

Step 2 (Second Stage): Regress Q_i on $\hat{P}_i$

Let's compute $\hat{Q}_i$ as a function of $\hat{P}_i$:

From:

- $Q_i = \alpha_1 + \alpha_2 \hat{P}_i + \text{error}$

Using:

W	P̂ = 2.4 + 1*W	Q
2	4.4	42
3	5.4	64
1	3.4	93
1	3.4	35
3	5.4	88

Now regress Q on $\hat{P}$:

This is simple OLS on:

- y = Q: [42, 64, 93, 35, 88]
- x = P̂: [4.4, 5.4, 3.4, 3.4, 5.4]

We can use OLS formulas to get estimates of α_1 and α_2.

Let me compute it:

Mean of Q = $(42 + 64 + 93 + 35 + 88)/5 = 64.4$

Mean of P̂ = $(4.4 + 5.4 + 3.4 + 3.4 + 5.4)/5 = 4.4$

Now calculate slope (α_2):

$$\alpha_2 = \frac{\sum(\hat{P}_i - \bar{\hat{P}})(Q_i - \bar{Q})}{\sum(\hat{P}_i - \bar{\hat{P}})^2}$$

We get:

$\hat{P}$	Q	$\hat{P}$−mean	Q−mean	Product	($\hat{P}$−mean)^2
4.4	42	0.0	-22.4	0.0	0.0
5.4	64	1.0	-0.4	-0.4	1.0
3.4	93	-1.0	28.6	-28.6	1.0
3.4	35	-1.0	-29.4	29.4	1.0
5.4	88	1.0	23.6	23.6	1.0

Sum of product = 24.0
Sum of ($\hat{P}$−mean)^2 = 4.0

$$\alpha_2 = \frac{24.0}{4.0} = 6.0$$

$$\alpha_1 = \bar{Q} - \alpha_2 \cdot \bar{\hat{P}} = 64.4 - 6.0 * 4.4 = 64.4 - 26.4 = 38.0$$

11.17 Example 11.3 introduces Klein's Model I.

a. Do we have an adequate number of IVs to estimate each equation? Check the necessary condition for the identification of each equation. The necessary condition for identification is that in a system of M equations at least $M - 1$ variables must be omitted from each equation.

The **necessary condition for identification** is that in a system of **M equations**, at least **M − 1 variables** must be omitted from each equation.

- **M** = **8** endogenous variables → Each equation must omit at least **7** endogenous variables.

Let's examine each of the three structural equations:

1. **Consumption equation (CNt):**
   Endogenous RHS variables: **W1t, Pt**
   Omitted endogenous variables: CNt, It, W1t, Pt, Et, Et−1 → at least 6, possibly more if definitions are included
   ✅ **Satisfies necessary condition**
2. **Investment equation (It):**
   Endogenous RHS variables: **Pt**
   Omitted: CNt, W1t, Et, It, etc. → ≥7
   ✅ **Satisfies necessary condition**
3. **Wage equation (W1t):**
   Endogenous RHS variables: **Et**
   Omitted: CNt, It, Pt, etc. → ≥7
   ✅ **Satisfies necessary condition**

Conclusion: All equations satisfy the necessary condition for identification.

b. An equivalent identification condition is that the number of excluded exogenous variables from the equation must be at least as large as the number of included right-hand side endogenous variables. Check that this condition is satisfied for each equation.

This condition states:

 # excluded exogenous variables ≥ # included endogenous RHS variables

Check each equation:

1. **Consumption equation**:

   Endogenous RHS: **W1t, Pt** → 2 variables

   Excluded exogenous variables: **TIMEt, Gt, Kt−1, TXt** → at least 4

   ✅ **Identified**

2. **Investment equation**:

   Endogenous RHS: **Pt** → 1 variable

   Excluded exogenous variables: **W1t, CNt, Gt, W2t, Et, Et−1, TIMEt** → many

   ✅ **Identified**

3. **Wage equation**:

   Endogenous RHS: **Et** → 1 variable

   Excluded exogenous variables: **Pt, CNt, It, TXt, Kt−1** → several

   ✅ **Identified**

Conclusion: All equations satisfy the exclusion restriction condition for identification.

c. Write down in econometric notation the first-stage equation, the reduced form, for W_{1t}, wages of workers earned in the private sector. Call the parameters $\pi_1, \pi_2, \ldots$

We want to express **W1t** as a function of **only exogenous and predetermined variables**.

Let's define the reduced-form for **W1t**:

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

Here:

- The **π's** are reduced-form coefficients.
- All variables on the RHS are **exogenous or predetermined**, satisfying IV conditions.

d. Describe the two regression steps of 2SLS estimation of the consumption function. This is not a question about a computer software command.

Structural equation (consumption function):

$$CN_t = \alpha_1 + \alpha_2(W_{1t} + W_{2t}) + \alpha_3 P_t + \alpha_4 P_{t-1} + e_{1t}$$

Endogenous RHS variables: **W1t, Pt**

Exogenous/instrumental variables: **W2t, P_{t-1}, G_t, K_{t-1}, TX_t, TIME_t, E_{t-1}**

Step 1: First-stage regressions

Regress each endogenous RHS variable on all exogenous and predetermined variables:

- Estimate:

$$\hat{W}_{1t} = \gamma_1 + \gamma_2 W_{2t} + \gamma_3 G_t + \gamma_4 TX_t + \gamma_5 TIME_t + \gamma_6 P_{t-1} + \gamma_7 K_{t-1} + \gamma_8 E_{t-1}$$

$$\hat{P}_t = \delta_1 + \delta_2 W_{2t} + \delta_3 G_t + \delta_4 TX_t + \delta_5 TIME_t + \delta_6 P_{t-1} + \delta_7 K_{t-1} + \delta_8 E_{t-1}$$

Step 2: Second-stage regression

Use the **predicted values** from Step 1:

$$CN_t = \alpha_1 + \alpha_2(\hat{W}_{1t} + W_{2t}) + \alpha_3 \hat{P}_t + \alpha_4 P_{t-1} + \text{error}$$

Estimate this regression using OLS to get 2SLS estimates of the α coefficients.

e. Does following the steps in part (d) produce regression results that are identical to the 2SLS estimates provided by software specifically designed for 2SLS estimation? In particular, will the *t*-values be the same?

- The **point estimates** of the coefficients (α's) will be the same in both manual 2SLS and software 2SLS.
- However, the **standard errors and t-values may differ** unless proper correction for generated regressors is made.

Software like **Stata, R, or Python's statsmodels** accounts for the fact that the regressors in the second stage are **predicted values**, and adjusts standard errors accordingly.

✅ **Estimates will be the same**

⚠️ **t-values may differ** due to standard error adjustments