Econometrics - 1

Lecture 2 (Jan 15, 2024)

Simple Linear Regression (model) → SLRM

Components

y → dependendent variable

x → explanatory variable (to explain systematic relation w/ y)

u → $\boxed{\text{random}}$, unobserved factors that may have helped in explaining y.

u is a random variable, then by def^n → $u = \{u_1, u_2, \ldots, u_N\}$

$p_1, p_2 \ldots, p_N$

$p_1 + p_2 + p_3 + \ldots + p_N = 1$

Specification

→ Except for x everything else is unobserved.
- by def^n. of a simple linear regression model

$$\underbrace{y}_{\text{dependent}} = \beta_0 + \beta_1 \underbrace{x}_{\text{explanatory variable}} + \underbrace{u}_{\text{random error.}} \rightarrow \underbrace{1}_{\text{SLRM}}$$

Note: u : unobserved variable is likely to be a "vector"

E.g. $\quad wage_p = \beta_0 + \beta_1 \, training_p + u$

u: $\boxed{educ_p, \; exper_p, \; ability_p}$ $\boxed{\text{random variable vector}}$

Intercept parameter: $\beta_0 \Rightarrow$ value of y when $x = 0$ and $u = 0$

Slope parameter: $\beta_1 \Rightarrow \frac{\Delta y}{\Delta x}$ when all factors in u are held constant

$$y = \beta_0 + \beta_1 x + u \quad \text{--- (1)}$$

(wage) (educ)

2 preliminary issues:

1. Linearity b/w y and x.

$\Rightarrow$ for every 1 unit change in X, the model infers/predicts the same, β_1 units of change in y, for all levels of x.

$\boxed{\text{UNREALISTIC.}}$

2. Causality

Can we say w/ confidence that x affects y?

What if I write:

$$x = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} y - \frac{u}{\beta_1}$$

$$\boxed{x = \gamma_0 + \gamma_1 y + \eta} \quad \text{--- (2)}$$

Question is which among the following should we trust?

$$y = \beta_0 + \beta_1 x + u$$

$$x = \gamma_0 + \gamma_1 y + \eta.$$

To understand further, we need a structural understanding of the process and (social) context/system.

role of economic reasoning/

theory/

Intuition.

u is likely a random "vector"?

wages, educ.

$$y = \beta_0 + \beta_1 x + \underbrace{u}$$

the observed variable.

composite variable (experience, training, etc.)

$$u = f(u_1 + u_2 + u_3)$$

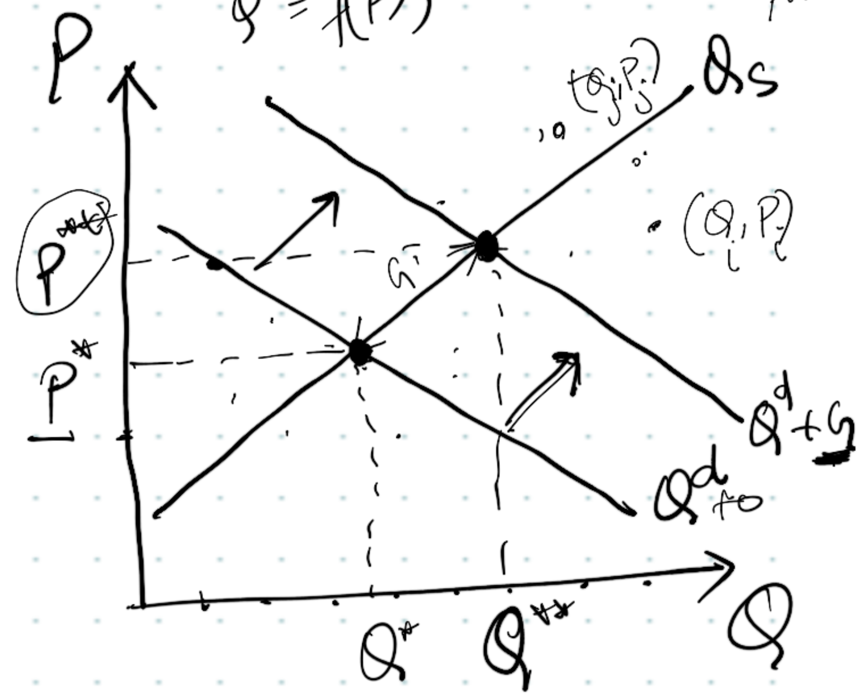
Economic Model of Demand and Supply.

$Q = f(P)$ Qs a fn of price

① $Q_d = a - bP$

② $Q_s = -c + dP$

③ $Q_s = Q_d + G$



structural model of a market

Parameters : a, b, c, d (unknown) to be estimated

Exogenous : G

Endogenous : Q_d, Q_s, P

4 parameters, 3 eqⁿs

$-c + dP = a - bP + G$

$\Rightarrow P^* = \frac{1}{d+b}[a + c + G]$ variable

$$-c + dP = a - bP + G$$

$$\Rightarrow \quad P = \frac{1}{d+b}\left[a + c + G\right]$$

$$\Rightarrow \quad P = \frac{a}{d+b} + \frac{c}{d+b} + \frac{1}{d+b} G$$

reduced form eq^n:

$$\boxed{P = \frac{a+c}{d+b} + \frac{1}{d+b} G}$$

$$y = \beta_0 + \beta_1 x + u$$

$$P = A_1 + A_2 G + u$$

$$\frac{a+c}{d+b} = A_1 \qquad = \frac{1}{d+b}$$

$$a + c = \frac{A_1}{A_2} \qquad d + b = \frac{1}{A_2}$$

$$P = \beta_0 + \beta_1 G + u \qquad \frac{\partial P}{\partial G} = \beta_1$$

Theory predicts: $\beta_1 \geq 0$ ← Test this using real-world data

$\underbrace{\hspace{3cm}}$ Hypothesis testing.

Data structures

* Cross-sectional data:

i	P_i	G_i
1	-	-
2	-	-
3	-	-
⋮	⋮	⋮
100 = N	-	-

$$P_i = \beta_0 + \beta_1 G_i + u_i$$

* Time-series

t	P_t	G_t
1	-	-
2	-	-
⋮	⋮	⋮
T	-	-

$$P_t = \beta_0 + \beta_1 G_t + u_t$$

- Panel or longitudinal data

i	t	P_{it}	G_{it}
1	1		
1	2		
1	3		
⋮	⋮		
1	10		
2	1		
2	2		
⋮	⋮		
2	10		
⋮	⋮		
N	1		
N	2		
⋮	⋮		
N	10		

$$P_{it} = \beta_0 + \beta_1 G_{it} + u_{it}$$

Concept of Ceteris Paribus

all else held constant

Assumptions

causality

$$y = \beta_0 + \beta_1 x + u$$

Aside : $y = f(x)$ → plot a graph.

$\boxed{Q_d = a - bP}$ → $Q = f(P)$