Advanced Econometrics

Chapter 2: The Classical Multiple Linear Regression Model

In Choi

Sogang University

Linear Regression Model

Reading: Chapter 2 of Greene

- Notation:
  y_i : dependent variable, regressand
  $x_{i1}, \cdots, x_{iK}$: independent variables, explanatory variables, control variables, regressors
  i : index.
- We want to explain y_i using $x_{i1}, \cdots, x_{iK}$. The multiple linear regression model takes the form

$$\begin{aligned} y_i &= \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + \varepsilon_i \;\; (i = 1, \cdots, n) \\ &= x_i' \beta + \varepsilon_i \end{aligned}$$

where

$$x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{pmatrix}, \; \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix} \text{ and } x_{i1} = 1 \text{ for all } i.$$

Linear Regression Model

- Equivalently, in matrix notation,

$$\begin{aligned} y &= \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_K\beta_K + \varepsilon \\ &= X\beta + \varepsilon \end{aligned}$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \ \mathbf{x}_l = \begin{pmatrix} x_{1l} \\ \vdots \\ x_{nl} \end{pmatrix} \quad (l = 1, \cdots, K)$$

Linear Regression Model

$$
\begin{aligned}
X &= (\mathbf{x}_1, ..., \mathbf{x}_K) \\
&= \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}
\end{aligned}
$$

and

$$
\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.
$$

Here ε_i is an error term.

Linear Regression Model

- *Loglinear model*:

$$\ln y = \beta_1 + \beta_2 \ln x_2 + \cdots + \beta_K \ln x_K + \varepsilon$$

$$\frac{\partial \ln y}{\partial \ln x_k} = \frac{\partial y / y}{\partial x_k / x_k} = \beta_k \text{ (constant elasticity)}$$

- *Semilog model:*

$$\ln y_t = X_t' \beta + \delta t + \varepsilon_t$$

Per period growth rate of y_t not explained by X_t is

$$\frac{d \ln y}{dt} = \delta$$

- *Polynomial equation*

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \cdots + \beta_p x^{p-1} + \varepsilon.$$

- *Interaction model*

$$y = \beta_1 + \beta_2 x + \beta_3 z + \beta_p xz + \varepsilon.$$

Example

Interaction model

$$D = \beta_1 + \beta_2 S + \beta_3 W + \beta_4 SW + \varepsilon$$

$$\begin{aligned} D &: \text{breaking distance} \\ W &: \text{road wetness} \\ S &: \text{speed} \end{aligned}$$

$$\frac{\partial D}{\partial S} = \beta_2 + \beta_4 W$$

If $\beta_4 > 0$, the marginal effect of higher speed on breaking distance is increased when the road is wetter.

Linear Regression Model

Example

Earnings and education

$$\text{earnings}_i = \beta_1 + \beta_2 \text{education}_i + \varepsilon_i$$

earnings_i : $i - th$ individual's hourly earning
education_i : $i - th$ individual's number of years in school
Any problems in the model?
Omitted variables such as job experience, job experience^2, gender, marital status, etc.

$$\begin{aligned} \text{earnings}_i &= \beta_1 + \beta_2 \text{education}_i + \beta_3 \text{job experience}_i + \beta_4 \text{job experience}_i^2 \\ &\quad + \beta_5 \text{gender}_i + \beta_6 \text{marital status}_i + \varepsilon_i \end{aligned}$$

Linear Regression Model

Example

Earnings and education (Continued)
For gender and marital status, use dummy variables. That is,

$$\text{gender}_i \begin{cases} = 1 \text{ if male} \\ = 0 \text{ if female} \end{cases}$$

$$\text{marital status}_i \begin{cases} = 1 \text{ if married} \\ = 0 \text{ if single} \end{cases}$$

See, for example, Ashenfelter and Krueger, American Economic Review, 1994, 73–85.

Linear Regression Model

Example

Class attendance and test scores

$$
\begin{aligned}
\text{score}_i &= \beta_1 + \beta_2 \left(\text{fraction of lectures attended}\right)_i \\
&\quad + \beta_3 \left(\text{fraction of problem sets completed}\right)_i + \varepsilon_i
\end{aligned}
$$

See Romer, Journal of Economic Perspectives, 1993.

Linear Regression Model

Example

Threshold effects and categorical variables

$$income = \beta_1 + \beta_2 age + \delta_B B + \delta_M M + \delta_P P + \varepsilon$$

$$B \begin{cases} = 1, & \text{Bachelor's degree only} \\ = 0, & \text{otherwise} \end{cases}$$

$$M \begin{cases} = 1, & \text{Master's and Bachelor's degrees} \\ = 0, & \text{otherwise} \end{cases}$$

$$P \begin{cases} = 1, & \text{Ph.D., Master's and Bachelor's degrees} \\ = 0, & \text{otherwise} \end{cases}$$

Linear Regression Model

Example

Threshold effects and categorical variables

High school

$$E\left[Income|age, HS\right] = \beta_1 + \beta_2 age$$

Bachelor's

$$E\left[Income|age, B\right] = \beta_1 + \beta_2 age + \delta_B$$

Master's

$$E\left[Income|age, M\right] = \beta_1 + \beta_2 age + \delta_B + \delta_M$$

Ph.D.

$$E\left[Income|age, P\right] = \beta_1 + \beta_2 age + \delta_B + \delta_M + \delta_P$$

Example

Seasonal effects

Suppose y_t is a quarterly data

$$y_t = \beta_1 + \beta' X_t + \delta_1 D_{1t} + \delta_2 D_{2t} + \delta_3 D_{3t} + \varepsilon_t.$$

Here X_t does not contain 1.

$$D_{1t} \begin{cases} = 1, & \text{if spring} \\ = 0, & \text{otherwise} \end{cases}$$

$$D_{2t} \begin{cases} = 1, & \text{if summer} \\ = 0, & \text{otherwise} \end{cases}$$

$$D_{3t} \begin{cases} = 1, & \text{if fall} \\ = 0, & \text{otherwise} \end{cases}$$

Conditional expectation

Definition

For two continuous random variables, X and Y, we say that the conditional probability density function (pdf) of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

where $f(x, y)$ is the joint pdf of X and Y and $f_X(x)$ is the marginal pdf of X.

Conditional expectation

Remark

(i) $f_{Y|X}(y|x)$ *is a function of* x *and possibly a different pdf for each* x.
(ii) When we wish to describe the entire family of pdf, we use the phrase "the pdf of $Y \mid X$*".*
(iii) If X *and* Y *are independent,*

$$f_{Y|X}(y|x) = f_Y(y)$$

Conditional expectation

Definition

A conditional mean is the mean of the conditional distribution and is defined by

$$E\left(Y|X=x\right)=\begin{cases}\int_y yf_{Y|X}\left(y|x\right)dy & \text{if } y \text{ is continuous}\\ \sum_y yf_{Y|X}\left(y|x\right) & \text{if } y \text{ is discrete}\end{cases}$$

Remark

(i) Note that

$$E\left[Y|X=x\right]=E\left[Y\right]$$

if X and Y are independent.
(ii) $E(Y \mid X)$ is a random variable whose value depends on X.

Conditional expectation

- The Law of Iterated Expectations

$$E(X) = E\left[E(X|Y)\right].$$

This holds because

$$\begin{aligned} E\left(X\right) &= \int\int xf\left(x,y\right)dxdy \\ &= \int\left[\int xf\left(x \mid y\right)dx\right] f_Y\left(y\right)dy \\ &= \int \left[E(X \mid y)\right] f_Y\left(y\right)dy \\ &= E(X \mid Y). \end{aligned}$$

Conditional expectation

- $$E\left[g(Y)h(X)|X\right] = h(X)E\left[g(Y)|X\right]$$

  Trivially, $E\left[g(Y)h(X)|X = x\right] = h(x)E\left[g(Y)|X = x\right]$. But this holds for any $x \in R$. Thus, the stated result holds. See, e.g., Ash (1972) for a formal proof.

Assumptions

1. Full rank

- X is an $n \times K$ matrix with rank K.
    - The columns of X are linearly independent.
    - Obviously, we should have $n \geq K$.
    - If this assumption is violated, X contains redundant information.
    - We allow stochastic regressors.

- What if this assumption is violated?
  Suppose that

$$y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \varepsilon$$

and

$$X_1 = \alpha X_2.$$

Then, the ordinary least squares estimator does not exist. When X is of deficient rank, we say that there is a multicollinearity problem.

- $$E(\varepsilon|X) = \begin{bmatrix} E(\varepsilon_1|X) \\ \vdots \\ E(\varepsilon_n|X) \end{bmatrix} = 0.$$

  Whatever the value of X is, the mean of ε is zero. This assumption implies

  $$E(\varepsilon_i) = 0$$

  (this follows from the law of iterated expectation) and

  $$Cov(\varepsilon_i, x_i) = E\left[\varepsilon_i \left(x_i - E(x_i)'\right)\right] = 0 \text{ for all } i.$$

- In fact, ε_i is uncorrelated with any function of x_i.

$$
\begin{aligned}
Cov\left(\varepsilon_i, f\left(x_i\right)\right) &= E\left[\varepsilon_i \left\{f\left(x_i\right) - Ef\left(x_i\right)\right\}'\right] \\
&= EE\left[\varepsilon_i \left\{f\left(x_i\right) - Ef\left(x_i\right)\right\}' \mid x_i\right] \\
&= E\left[\left\{f\left(x_i\right) - Ef\left(x_i\right)\right\}' E\left(\varepsilon_i \mid x_i\right)\right] \\
&= 0.
\end{aligned}
$$

- $$\begin{cases} Var\left(\varepsilon_i|X\right) = \sigma^2 \text{ for all } i = 1, 2, \cdots, n \\ Cov\left(\varepsilon_i, \varepsilon_j|X\right) = 0 \text{ for all } i \neq j. \end{cases}$$

  These imply

  $$\begin{aligned} E\left(\varepsilon\varepsilon'|X\right) &= \begin{bmatrix} E\left(\varepsilon_1^2|X\right) & E\left(\varepsilon_1\varepsilon_2|X\right) & \cdots & E\left(\varepsilon_1\varepsilon_n|X\right) \\ & \ddots & & \\ & & & \\ & & & E\left(\varepsilon_n^2|X\right) \end{bmatrix} \\ &= \sigma^2 I. \end{aligned}$$

  The assumption of common variance for ε_i is called homoskedasticity.

Assumptions

- Alternatively, we may assume

1. X is an $n \times K$ nonstochastic matrix with rank K.
2. $\varepsilon_i \sim iid(0, \sigma^2)$.

- These assumptions are more restrictive.

Partial effects, elasticity and average partial effect

- The partial derivative of $E(y \mid X)$ with respect to x_j is called the partial effect of x_j on $E(y \mid X)$.
- The elasticity of $E(y \mid X)$ with respect to x_j, holding other variables fixed, is

$$\frac{\partial E(y \mid X)}{\partial x_j} \cdot \frac{x_j}{E(y \mid X)}$$

Example

Let

$$E(y \mid X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

$\frac{\partial E(y|X)}{\partial x_1} = \beta_1$ and $\frac{\partial E(y|X)}{\partial x_2} = \beta_2$.

Partial effects, elasticity and average partial effect

- Suppose that $E(y \mid X, q) = \mu(X, q)$, where q is not observed. Let $\theta_j(X, q) = \frac{\partial \mu(X,q)}{\partial x_j}$. The average partial effect (APE) of x_j is

$$E_q \theta_j(X, q).$$