Applied Econometrics

Derek Li

Contents

1 Introduction

Econometrics are quantitative methods of analyzing and interpreting economic data, which need to combine economic theory, math and statistics, data, and statistical or econometrics software.

We focus on estimating economic relationships, testing hypothesis involving economic behavior, and forecasting the behavior of economic variables.

1.1 Econometric Methodology

- Ask a question - statement of theory or hypothesis
- Specification of economic model
- Specification of econometric model
- Collection of Data
- Estimation of the econometric model
- Hypothesis testing
- Prediction or forecasting

1.2 Data Types

There are different data structures: cross-section, time series, repeated cross-section, and panel data.

1.2.1 Cross-Section

Cross-section consists of a sample of individuals, households, firms, countries, etc, taken at a given point in time. Observations are generally independent draws from the population. It is commonly indexed by i as x_i.

1.2.2 Time Series

Time series consists of observations on a variable or several variables over time. Observations are almost never independent of each other. It is commonly indexed by t as x_t.

1.2.3 Repeated Cross-Section

Repeated cross-section consists of two or more cross-sectional data in different points in time, and is different units in different periods. It is commonly index by it as x_{it}.

Example 1.1. Suppose two cross-sectional household surveys are taking in 1985 and 1990. In 1985, a random sample of households is surveyed for variables such as income. In 1990, a new random sample of households is taken using the same survey questions. This is a repeated cross-section data set.

1.2.4 Panel Data

Panel data consists of a time series for each cross-sectional unit in the data set. Observations are independent among units and dependent over time for each unit. It is commonly index by it as x_{it}.

2 Review of Statistics

Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$. Assume a random sample $\{x_1, \cdots, x_n\}$, i.e., identically and independently distributed (i.i.d.).

2.1 Estimator

Definition 2.1 (Statistic)**.** A statistic is a function of the data.

Definition 2.2 (Estimator)**.** An estimator is a statistic that is used to estimate the parameter of interest.

Take μ as an example, the proposed estimator is sample average

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

2.2 Sampling Distribution

We have

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$

i.e., $\overline{X}_n$ is an unbiased estimator of μ. Besides,

$$\mathrm{Var}\left[\overline{X}_n\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} x_i\right] \overset{\text{independet}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[x_i\right] = \frac{\sigma^2}{n}.$$

Definition 2.3 (Consistency)**.** Let W_n be an estimator of θ based on a sample $Y_1, \cdots, Y_n$. Then W_n is a consistent estimator of θ is for every $\varepsilon > 0$,

$$P(|W_n - \theta| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

As commonly stated, an estimator is called consistent when its sampling distribution becomes more and more concentrated around the parameters of interest as the sample size increases. Note that $\overline{X}_n$ is a consistent estimator of μ.

Theorem 2.1. If $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then

$$Y_1 + Y_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\mathrm{Cov}(Y_1, Y_2)),$$

By theorem, we have the sampling distribution

$$\overline{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}),$$

provided $X \sim \mathcal{N}(\mu, \sigma^2)$.

2.3 Confidence Interval

Theorem 2.2. If Y has $\mathbb{E}[Y]=\mu, \mathrm{Var}[Y]=\sigma^2$, then $Z=\frac{Y-\mu}{\sigma}$ is such that $\mathbb{E}[Z]=0, \mathrm{Var}[Z]=1$.

By theorem, we have

$$Z=\frac{\overline{X}_n-\mu}{\sqrt{\sigma^2/n}}\sim\mathcal{N}(0,1).$$

Therefore,

$$\begin{aligned}1-\alpha=P\left(-z_{\frac{\alpha}{2}}\leqslant Z\leqslant z_{\frac{\alpha}{2}}\right)&=P\left(-z_{\frac{\alpha}{2}}\leqslant\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\leqslant z_{\frac{\alpha}{2}}\right)\\&=P\left(\overline{X}_n-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\overline{X}_n+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right),\end{aligned}$$

i.e., $(1-\alpha)\%$ confidence interval is

$$\left[\overline{X}_n-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\overline{X}_n+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right].$$

2.4 Proves of Some Theorems and Results (optional)

Theorem 2.3 (Markov's Inequality)**.** If X is a nonnegative random variable and $a>0$, then

$$P(X\geqslant a)\leqslant\frac{\mathbb{E}[X]}{a}.$$

Proof. Since X is a nonnegative random variable, we have

$$\begin{aligned}\mathbb{E}[X]=\int_0^\infty xf(x)\mathrm{d}x&=\mathbb{E}[X]=\int_0^a xf(x)\mathrm{d}x+\int_a^\infty xf(x)\mathrm{d}x\\&\geqslant\int_a^\infty xf(x)\mathrm{d}x\geqslant\int_a^\infty af(x)\mathrm{d}x=a\int_a^\infty f(x)\mathrm{d}x=aP(X\geqslant a).\end{aligned}$$

Hence

$$P(X\geqslant a)\leqslant\frac{\mathbb{E}[X]}{a}.$$

□

Theorem 2.4 (Chebyshev Inequality)**.** For any $b>0$,

$$P(|X-\mathbb{E}[X]|\geqslant b)\leqslant\frac{\mathrm{Var}[X]}{b^2}.$$

Proof. By Markov's Inequality, we have

$$P((X-\mathbb{E}[X])^2\geqslant b^2)\leqslant\frac{\mathbb{E}[(X-\mathbb{E}[X])^2]}{b^2}=\frac{\mathrm{Var}[X]}{b^2}.$$

Therefore,

$$P(|X-\mathbb{E}[X]|\geqslant b)\leqslant\frac{\mathrm{Var}[X]}{b^2}.$$

□

Theorem 2.5 (Weak Law of Large Numbers)**.** Let $X_1,\cdots,X_n$ be a sequence of independent random variables with $\mathbb{E}[X_i]=\mu$. Let $\overline{X}_n=\frac{1}{n}\sum\limits_{i=1}^n X_i$, then

$$\forall\varepsilon>0,\lim_{n\to\infty}P\left(\left|\overline{X}_n-\mu\right|>\varepsilon\right)=0.$$

Proof. By Chebyshev Inequality, for all $\varepsilon > 0$,

$$P\left(\left|\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]\right| > \varepsilon\right) \leqslant \frac{\operatorname{Var}\left[\overline{X}\right]}{\varepsilon^2} \Leftrightarrow 0 \leqslant P(|\overline{X}_n - \mu| > \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon}.$$

Since

$$\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon} = 0,$$

then

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = 0.$$

$\square$

It follows that $\overline{X}_n$ is a consistent estimator of μ.

3 Simple Regression

3.1 Basic Property

Property 3.1. $\sum_{i=1}^{n}(x_i - \overline{x}) = 0.$

Proof. $\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - n\overline{x} = n\overline{x} - n\overline{x} = 0.$ □

Property 3.2. $\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i - \overline{x})y_i.$

Proof. On the one hand,

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i(y_i - \overline{y}) - \sum_{i=1}^{n} \overline{x}(y_i - \overline{y}) = \sum_{i=1}^{n} x_i(y_i - \overline{y}).$$

On the other hand,

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i - \overline{x})y_i - \sum_{i=1}^{n}(x_i - \overline{x})\overline{y} = \sum_{i=1}^{n}(x_i - \overline{x})y_i.$$

□

Corollary 3.1. $\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i(x_i - \overline{x}).$

Property 3.3. $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]].$

3.2 Econometric Model

Let (Y, X, U) be random variables with joint distribution s.t. $Y = g(X, U)$, where Y is dependent variable, X is explanatory (regressor/covariate) variable, and U is unobservable variable.

Assumption 1 (Linear in Parameters)**.** $Y = \beta_0 + \beta_1 X + U$.

Example 3.1. $y = \beta_0 + \beta_1 x^2 + u$, then $\frac{\partial y}{\partial x} = 2\beta_1 x$.

Example 3.2. $\ln y = \beta_0 + \beta_1 \ln x + u$, then $\frac{\partial \ln y}{\partial \ln x} = \beta_1 \approx \frac{\Delta y\%}{\Delta x\%}$.

Assumption 2 (Zero Conditional Mean)**.** $\mathbb{E}[U|X] = \mathbb{E}[U]$, and $\mathbb{E}[U] = 0$.

Example 3.3. Let Y be wage, X be training program, and U be ability. If training is assigned randomly, then X and U are fully independent. Nevertheless, if let X be education, then the education may influence the ability so that $\mathbb{E}[U|X = 0] \neq \mathbb{E}[U|X = 1]$, then A2 is violated.

From A1 and A2, we have

$$\begin{aligned}\mathbb{E}[Y|X] &\overset{\text{A1}}{=} \mathbb{E}[\beta_0 + \beta_1 X + U|X] = \beta_0 + \beta_1 \mathbb{E}[X|X] + \mathbb{E}[U|X] \\ &\overset{\text{A2}}{=} \beta_0 + \beta_1 X + \mathbb{E}[U] = \beta_0 + \beta_1 X.\end{aligned}$$

Assumption 3 (Random Sample)**.** $\{(x_i, y_i), i = 1, \cdots, n\}$ is i.i.d.

Assumption 4 (Sample Variation)**.** $\{x_1, \cdots, x_n\}$ are not all the same.

Assumption 5 (Homoscedasticity)**.** $\mathrm{Var}[U|X] = \sigma_U^2$.

From A1 and A5, we have

$$\mathrm{Var}[Y|X] \overset{\text{A1}}{=} \mathrm{Var}[\beta_0 + \beta_1 X + U|X] = \mathrm{Var}[U|X] \overset{\text{A5}}{=} \sigma_U^2.$$

3.3 Estimator: OLS

We want to solve

$$\min_{\widehat{\beta}_0, \widehat{\beta}_1} \sum_{i=1}^{n} Q\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2.$$

Let

$$\begin{aligned} \frac{\partial Q}{\partial \widehat{\beta}_0} &= -\sum_{i=1}^{n} 2\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0, \\ \frac{\partial Q}{\partial \widehat{\beta}_1} &= -\sum_{i=1}^{n} 2\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) x_i = 0. \end{aligned}$$

Therefore,

$$\widehat{\beta}_0 = \overline{y} - \beta_1 \overline{x}.$$

Besides,

$$\begin{aligned} \sum_{i=1}^{n} \left(x_i y_i - \widehat{\beta}_0 x_i - \widehat{\beta}_1 x_i^2\right) &= \sum_{i=1}^{n} (x_i y_i - x_i \overline{y} + \widehat{\beta}_1 x_i \overline{x} - \widehat{\beta}_1 x_i^2) \\ &= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \overline{y} + \sum_{i=1}^{n} \widehat{\beta}_1 x_i \overline{x} - \sum_{i=1}^{n} \widehat{\beta}_1 x_i^2 \\ &= \sum_{i=1}^{n} x_i y_i - n\overline{xy} + n\widehat{\beta}_1 \overline{x}^2 - \widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0, \end{aligned}$$

and thus

$$\widehat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum\limits_{i=1}^{n} (x_i y_i - x_i \overline{y})}{\sum\limits_{i=1}^{n} (x_i^2 - x_i \overline{x})} = \frac{\sum\limits_{i=1}^{n} x_i (y_i - \overline{y})}{\sum\limits_{i=1}^{n} x_i (x_i - \overline{x})} = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}.$$

3.4 Properties of OLS

Property 3.4. $\mathbb{E}\left[\widehat{\beta}_1\right] = \beta_1$.

Proof. We have

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2} = \frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})y_i}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2} = \frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(\beta_0+\beta_1x_i+u_i)}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2} \\
&= \frac{1}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\left[\sum_{i=1}^{n}(x_i-\overline{x})\beta_0+\sum_{i=1}^{n}(x_i-\overline{x})\beta_1x_i+\sum_{i=1}^{n}(x_i-\overline{x})u_i\right] \\
&= \beta_1+\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})u_i}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}.
\end{aligned}
$$

By A4, $\sum\limits_{i=1}^{n}(x_i-\overline{x})^2 \neq 0$, and thus,

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta}_1|x_1,\cdots,x_n\right] &= \mathbb{E}\left[\beta_1+\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})u_i}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\middle|x_1,\cdots,x_n\right] \\
&= \beta_1+\mathbb{E}\left[\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})u_i}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\middle|x_1,\cdots,x_n\right] \\
&= \beta_1+\frac{1}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\sum_{i=1}^{n}(x_i-\overline{x})\mathbb{E}[u_i|x_1,\cdots,x_n] \\
&\overset{\text{A3}}{=} \beta_1+\frac{1}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\sum_{i=1}^{n}(x_i-\overline{x})\mathbb{E}[u_i|x_i] \\
&\overset{\text{A2}}{=} \beta_1+\frac{1}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\sum_{i=1}^{n}(x_i-\overline{x})\cdot 0 = \beta_1.
\end{aligned}
$$

Therefore, $\mathbb{E}\left[\widehat{\beta}_1\right] = \mathbb{E}\left[\mathbb{E}\left[\widehat{\beta}_1|x_1,\cdots,x_n\right]\right] = \mathbb{E}\left[\beta_1\right] = \beta_1$. As a consequence, OLS is unbiased. □

Property 3.5. $\mathrm{Var}\left[\widehat{\beta}_1|x_1,\cdots,x_n\right] = \frac{\sigma_u^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$

Proof. We have

$$\begin{aligned}
\operatorname{Var}\left[\widehat{\beta}_1\right] &= \operatorname{Var}\left[\beta_1 + \frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})u_i}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\middle| x_1,\cdots,x_n\right] \\
&= \frac{1}{\left(\sum\limits_{i=1}^{n}(x_i-\overline{x})^2\right)^2}\operatorname{Var}\left[\sum_{i=1}^{n}(x_i-\overline{x})u_i\middle| x_1,\cdots,x_n\right] \\
&\overset{\text{A3}}{=} \frac{1}{\left(\sum\limits_{i=1}^{n}(x_i-\overline{x})^2\right)^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\operatorname{Var}[u_i|x_1,\cdots,x_n] \\
&\overset{\text{A3}}{=} \frac{1}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}\operatorname{Var}[u_i|x_i] \overset{\text{A5}}{=} \frac{\sigma_u^2}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}.
\end{aligned}$$

□

Notice that larger $\sum\limits_{i=1}^{n}(x_i-\overline{x})^2$ implies smaller $\operatorname{Var}\left[\widehat{\beta}_1\right]$.

Theorem 3.1 (Gauss-Markov Theorem)**.** Under A1 to A5, OLS is the best linear unbiased estimator.

Definition 3.1 (Standard Error)**.** Let $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i, \widehat{u}_i = y_i - \widehat{y}_i, \widehat{\sigma}_U^2 = \frac{1}{n-2}\sum\limits_{i=1}^{n}\widehat{u}_i^2$, define

$$\operatorname{SE}\left(\widehat{\beta}_1\right) = \sqrt{\widehat{\operatorname{Var}}\left[\widehat{\beta}_1\right]} = \sqrt{\frac{\widehat{\sigma}_U^2}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}}.$$

Property 3.6. $\sum\limits_{i=1}^{n}\widehat{u}_i = 0$.

Proof. We have

$$\begin{aligned}
\sum_{i=1}^{n}\widehat{u}_i &= \sum_{i=1}^{n}(y_i-\widehat{y}_i) = \sum_{i=1}^{n}\left(y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_i\right) = \sum_{i=1}^{n}\left(y_i-\overline{y}+\widehat{\beta}_1\overline{x}-\widehat{\beta}_1 x_i\right) \\
&= n\overline{y} - n\overline{y} + n\widehat{\beta}_1\overline{x} - n\widehat{\beta}_1\overline{x} = 0.
\end{aligned}$$

□

Property 3.7. $\sum\limits_{i=1}^{n}\widehat{u}_i x_i = 0$.

Proof. We have

$$\begin{aligned}
\sum_{i=1}^{n}\widehat{u}_i x_i &= \sum_{i=1}^{n}(y_i-\widehat{y}_i)\,x_i = \sum_{i=1}^{n}\left(y_i-\overline{y}+\widehat{\beta}_1\overline{x}-\widehat{\beta}_1 x_i\right)x_i \\
&= \sum_{i=1}^{n}x_i y_i - n\overline{x}\overline{y} + \widehat{\beta}_1\left(n\overline{x}^2 - \sum_{i=1}^{n}x_i^2\right) = 0.
\end{aligned}$$

□

Definition 3.2. $R^2 = 1 - \frac{\text{SSR}}{\text{SST}}$, where $\text{SSR} = \sum\limits_{i=1}^{n}\widehat{u}_i^2, SST = \sum\limits_{i=1}^{n}(y_i-\overline{y})^2$.

4 Multiple Regression: Estimation

4.1 Econometric Model

Assumption 1 (Linear in Parameters)**.** $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$. We can write Y as

$$Y = \begin{pmatrix} 1 & X_1 & \cdots & X_k \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = X\beta + U,$$

where X is $1 \times (k+1), \beta$ is $(k+1) \times 1$.

Assumption 2 (Zero Conditional Mean)**.** $\mathbb{E}[U|X] = 0$.

Assumption 3 (Random Sample)**.** $\{(y_i, x_{1i}, \cdots, x_{ki}), i = 1, \cdots, n\}$ is i.i.d.

Assumption 4 (No Perfect Collinearity)**.** There is no exact linear relationship among the explanatory variables.

Example 4.1. Let $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U, X_3 = X_1 + X_2$. Then

$$Y = \beta_0 + (\beta_1 + \beta_3) X_1 + (\beta_2 + \beta_3) X_2 + U,$$

which violates No Perfect Collinearity.

Example 4.2. Let $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 (X_1 \cdot X_2) + U$, then

$$\frac{\partial Y}{\partial X_1} = \beta_1 + 2\beta_2 X_1 + \beta_3 X_2,$$

and thus the model does not violate No Perfect Collinearity.

Assumption 5 (Homoscedasticity)**.** $\text{Var}[U|X_1, \cdots, X_k] = \sigma_U^2$.

From A1 and A5, we have

$$\text{Var}[Y|X] = \sigma_U^2.$$

4.2 Estimator: OLS

We want to solve

$$\min_{\widehat{\beta}_0, \cdots, \widehat{\beta}_k} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{1i} - \cdots - \widehat{\beta}_k x_{ki} \right)^2.$$

With F.O.C., we have $\widehat{\beta} = (X^T X)^{-1} X^T Y$.

4.3 Properties of OLS

Property 4.1. R^2 increases when the model includes more explanatory variables.

Property 4.2 (Partialling Out)**.** Regress X_1 on $X_2, \cdots, X_k$:

$$X_1 = \alpha_0 + \alpha_1 X_2 + \cdots + \alpha_{k-1} X_k + \Omega.$$

Thus

$$\widehat{\Omega}_i = X_{1i} - \widehat{X}_{1i} = X_{1i} - \left(\widehat{\alpha}_0 + \widehat{\alpha}_1 X_{2i} + \cdots + \widehat{\alpha}_{k-1} X_{ki}\right).$$

Regress Y_i on $\widehat{\Omega}_i$:

$$Y_i = \gamma_0 + \gamma_1 \widehat{\Omega}_i + V_i,$$

then

$$\widehat{\gamma} = \frac{\sum_{i=1}^{n} \widehat{\Omega}_i Y_i}{\sum_{i=1}^{n} \widehat{\Omega}_i^2} = \widehat{\beta}_1.$$

Therefore, $\widehat{\beta}_1$ measures the effect of X_1 on Y after $X_2, \cdots, X_k$ have been partialled out.

Property 4.3. Under A1 to A4, OLS is unbiased, i.e., $\mathbb{E}\left[\widehat{\beta}_j\right] = \beta_j, j = 0, \cdots, k.$

Property 4.4. Under A1 to A5,

$$\operatorname{Var}\left[\widehat{\beta}_j\right] = \frac{\sigma_U^2}{\left[\sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2\right](1 - R_j^2)},$$

for $j = 1, \cdots, k$, where R_j^2 is the R^2 of the regression of X_j on all other X's.

4.3.1 Omitted Variable Bias

Suppose the true model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U,$$

with $\mathbb{E}[U|X_1, X_2] = 0$. Now ignore X_2 and instead consider $Y = \beta_0 + \beta_1 X_1 + U$.

Suppose $X_2 = \alpha_0 + \alpha_1 X_1 + V, \mathbb{E}[V|X_1] = 0$, and thus

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2(\alpha_0 + \alpha_1 X_1 + V) + U = (\beta_0 + \beta_2\alpha_0) + (\beta_1 + \beta_2\alpha_1)X_1 + (\beta_2 V + U) \\ &:= \widetilde{\beta}_0 + \widetilde{\beta}_1 X_1 + \varepsilon. \end{aligned}$$

If $\mathbb{E}[\varepsilon|X_1] = 0$, then $\widetilde{\beta}_1$ for $\widetilde{\beta}_1$ is unbiased, i.e.,

$$\mathbb{E}\left[\widehat{\delta}_1\right] = \delta_1 = \beta_1 + \beta_2\alpha_1,$$

then $\text{Bias}\left(\widehat{\delta}_1\right) = \beta_2\alpha_1$. Because the bias in this case arises from omitting x_2, we call $\beta_2\alpha_1$ the omitted variable bias.

4.3.2 Including New Variables

Suppose we have $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$, and we consider whether X_{k+1} should be included in the model.

We consider these cases:

- If $\beta_{k+1} = 0$, then we should not include X_{k+1}.
- If $\beta_{k+1} \neq 0$ and X_{k+1} is uncorrelated with all other X's:

  - There is no omitted variable bias problem.

  - $\Delta R_j^2 = 0$ and σ_U^2 decreases, and thus $\mathrm{Var}\left[\widehat{\beta}_j\right]$ decreases.

  Hence we should include X_{k+1}.
- If $\beta_{k+1} \neq 0$ and X_{k+1} is correlated with other X's:

  - Excluding X_{k+1} leads to omitted variable bias.

  - R_j^2 increases and σ_U^2 decreases, and thus the change in $\mathrm{Var}\left[\widehat{\beta}_j\right]$ is unclear.

  We should still include X_{k+1} to avoid omitted variable bias.

5 Multiple Regression: Inference

Assumption 6 (Normality)**.** $U|X \sim \mathcal{N}(0, \sigma_U^2)$

With A6 we have

$$Y|X \sim \mathcal{N}(X\beta, \sigma_U^2).$$

Property 5.1. Under A1 to A6,

$$\widehat{\beta}_j \sim \mathcal{N}\left(\beta_j, \frac{\sigma_U^2}{\left[\sum\limits_{i=1}^{n}(x_{ij} - \overline{x}_j)^2\right](1 - R_j^2)}\right),$$

for $j = 1, \cdots, k$.

Corollary 5.1. We have

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\text{Var}\left[\widehat{\beta}_j\right]}} \sim \mathcal{N}(0, 1).$$

Corollary 5.2. We have

$$T = \frac{\widehat{\beta}_j - \beta_j}{\text{SE}\left(\widehat{\beta}_j\right)} \sim t_{(n-k-1)},$$

where $\text{SE}\left(\widehat{\beta}_j\right) = \sqrt{\frac{\widehat{\sigma}_U^2}{\left[\sum\limits_{i=1}^{n}(x_{ij}-\overline{x}_j)^2\right](1-R_j^2)}}, \widehat{\sigma}_U^2 = \frac{1}{n-k-1}\sum\limits_{i=1}^{n}\left(\widehat{u}_i\right)^2 = \frac{1}{n-k-1}\sum\limits_{i=1}^{n}\left(y_i - \mathbf{x}_i\widehat{\beta}\right)^2.$

5.1 Confidence Interval

The $(1-\alpha)\%$ confidence interval for β_j is

$$\left[\widehat{\beta}_j \pm c \cdot \text{SE}\left(\widehat{\beta}_j\right)\right].$$