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 Hypothesis Testing

2.4.1 Two Types of Errors

We want to test H_0 against H_1 and there are two types of errors:

(1) Reject H_0 when H_0 is true;

(2) Reject H_1 when H_1 is true.

There is a trade-off between two types of errors: if $P(\text{Reject } H_0|H_0)$ decreases, then $P(\text{Reject } H_1|H_1)$ increases. But we fix the $P(\text{Reject } H_0|H_0)$ at a very small level, i.e., we reject H_0 when there is strong evidence against it.

2.4.2 Testing Steps

(1) Fix the $P(\text{Reject } H_0|H_0)$ at some level α.

(2) Define a test statistic and we have to know the distribution of test statistic under H_0.

(3) Define rejection region, $\{|t| > c\}$, where c is the critical value and

$$\alpha = P(|t| > c|H_0) = 1 - P(-c < t < c|H_0).$$

(4) Check the result and decide whether reject H_0 or not.

2.4.3 p-Value

Given the observed value of the test statistic, p-value is the smallest significance level α at which the null would be rejected. The smaller the p-value, the greatest the evidence against H_0.

2.5 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 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, 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 $\varepsilon > 0$,

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

Since

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

then

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

□

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)**.** $\text{Var}[U|X] = \sigma_U^2$.

From A1 and A5, we have

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

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

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

Therefore,

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

Besides,

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

and thus

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

3.4 Properties of OLS

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

Proof. We have

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

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

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

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

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

□

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

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

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

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

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

Proof. We have

$$\begin{aligned}
\sum_{i=1}^{n}\widehat{u}_i &= \sum_{i=1}^{n}(y_i-\widehat{y}_i) = \sum_{i=1}^{n}\left(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i\right) = \sum_{i=1}^{n}\left(y_i-\overline{y}+\widehat{\beta}_1\overline{x}-\widehat{\beta}_1x_i\right) \\
&= n\overline{y} - n\overline{y} + n\widehat{\beta}_1\overline{x} - n\widehat{\beta}_1\overline{x} = 0.
\end{aligned}$$

□

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

Proof. We have

$$\begin{aligned}
\sum_{i=1}^{n}\widehat{u}_i x_i &= \sum_{i=1}^{n}(y_i-\widehat{y}_i)\,x_i = \sum_{i=1}^{n}\left(y_i-\overline{y}+\widehat{\beta}_1\overline{x}-\widehat{\beta}_1x_i\right)x_i \\
&= \sum_{i=1}^{n}x_iy_i - n\overline{x}\overline{y} + \widehat{\beta}_1\left(n\overline{x}^2 - \sum_{i=1}^{n}x_i^2\right) = 0.
\end{aligned}$$

□

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

3.5 Interpretation

Table 3.1: Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-Level	y	x	$\Delta y = \beta_1 \Delta x$
Level-Log	y	$\ln x$	$\Delta y = \frac{\beta_1}{100} \% \Delta x$
Log-Level	$\ln y$	x	$\% \Delta y = 100 \beta_1 \Delta x$
Log-Log	$\ln y$	$\ln x$	$\% \Delta y = \beta_1 \% \Delta x$

Example 3.4 (Level-Log)**.** Suppose

$$hours = 33 + 45.1 \ln(wage),$$

then a 1% increase in *wage* increases the weekly hours worked by about 0.451.

Example 3.5 (Log-Level)**.** Suppose

$$\ln(wage) = 2.78 + 0.094 educ,$$

then one more year of education increases hourly wage by about 9.4%.

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

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

where $\mathbf{X}$ is $1 \times (k+1)$, β is $(k+1) \times 1$.

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

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)**.** $\mathrm{Var}[U|X_1, \cdots, X_k] = \sigma_U^2$.

From A1 and A5, we have

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

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 $\widehat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.

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

$$\widehat{\Omega}_i = X_{i1} - \widehat{X}_{i1} = X_{i1} - \left(\widehat{\alpha}_0 + \widehat{\alpha}_1 X_{i2} + \cdots + \widehat{\alpha}_{k-1} X_{ik}\right).$$

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

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

then

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

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,

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

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

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2(\alpha_0 + \alpha_1 X_1 + V) + U = (\beta_0 + \beta_2\alpha_0) + (\beta_1 + \beta_2\alpha_1)X_1 + (U + \beta_2 V) \\ &:= \delta_0 + \delta_1 X_1 + \varepsilon. \end{aligned}$$

If $\mathbb{E}[\varepsilon|X_1] = 0$, then $\widetilde{\delta}_1$ for δ_1 is unbiased, i.e.,

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

then $\mathrm{Bias}\left(\widehat{\delta}_1\right) = \beta_2\alpha_1$. 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 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 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). $U|\mathbf{X} \sim \mathcal{N}(0, \sigma_U^2)$

With A6 we have

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

Property 5.1. Under A1 to A6,

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

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

5.1 Confidence Interval

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

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

5.2 t-Test

We want to test

$$H_0 : \beta_j = \beta_j^0 \text{ against } H_1 : \beta_j \neq \beta_j^0,$$

and we use the test statistic

$$T = \frac{\widehat{\beta}_j - \beta_j}{\text{SE}\left(\widehat{\beta}_j\right)}.$$

5.3 F-Test

Suppose the unrestricted model $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + U$. F-test is a joint test to test last q coefficients are zero, i.e., we want to test

$$H_0 : \beta_{k-q+1} = \cdots = \beta_k = 0 \text{ against } H_1 : \text{At least one of them is not 0.}$$

Under H_0, we have the restricted model $Y = \beta_0 + \beta_1 X_1 + \cdots \beta_{k-q} X_{k-q} + U$, and the test statistic is

$$F = \frac{(\text{SSR}_\text{r} - \text{SSR}_\text{ur})/q}{\text{SSR}_\text{ur}/(n-k-1)} \sim F_{(q,n-k-1)}$$

under H_0.

5.4 Example in STATA

Source	SS	df	MS
Model	SSE	k	$\frac{\text{SSE}}{k}$
Residual	SSR	$n-k-1$	$\frac{\text{SSR}}{n-k-1}$
Total	SST	$n-1$	$\frac{\text{SST}}{n-1}$

Number of obs $= n$
F(,) $= F$
Prob > F $= p$-value of the F-test
R $-$ squared $= 1 - \frac{\text{SSR}}{\text{SST}}$
Adj R-squared
Root MSE $= \sqrt{\widehat{\sigma}_U^2}$

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
X1	$\widehat{\beta}_1$	SE $\left(\widehat{\beta}_1\right)$	$\frac{\widehat{\beta}_j}{\text{SE}\left(\widehat{\beta}_j\right)}$	p-value of t-test	$\left[\widehat{\beta}_j \mp c \cdot \text{SE}\left(\widehat{\beta}_j\right)\right]$
X2	$\widehat{\beta}_2$	SE $\left(\widehat{\beta}_2\right)$			
_cons	$\widehat{\beta}_0$	SE $\left(\widehat{\beta}_0\right)$			

6 Multiple Regression: OLS Asymptotics

6.1 Consistency

Definition 6.1. $\widehat{\theta}_n$ is a consistent estimator of θ if

$$\forall \varepsilon > 0, P\left(\left|\widehat{\theta}_n - \theta\right| < \varepsilon\right) \to 1,$$

as $n \to \infty$, and we write $\widehat{\theta}_n \xrightarrow{P} \theta$.

Property 6.1. Let $\widehat{\theta}_n \xrightarrow{P} \theta, \widehat{\alpha}_n \xrightarrow{P} \alpha$, then

(1) $\widehat{\theta}_n + \widehat{\alpha}_n \xrightarrow{P} \theta + \alpha$;

(2) $\widehat{\theta}_n \cdot \widehat{\alpha}_n \xrightarrow{P} \theta \cdot \alpha$;

(3) $\frac{\widehat{\theta}_n}{\widehat{\alpha}_n} \xrightarrow{P} \frac{\theta}{\alpha}$, provided $\alpha \neq 0$.

(4) $g\left(\widehat{\theta}_n\right) \xrightarrow{P} g(\theta)$, provided that g is continuous.

Theorem 6.1 (Law of Large Numbers)**.** Let $X_1, \cdots, X_n$ be i.i.d. with mean $\mathbb{E}[X] = \mu$, then

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X] = \mu.$$

Corollary 6.1. Let $X_1, \cdots, X_n$ be i.i.d. with mean $\mathbb{E}[X] = \mu$, then

$$\mathrm{Var}\left[\overline{X}\right] = \frac{\sigma^2}{n} \to 0,$$

as $n \to \infty$.

Theorem 6.2. Under A1 to A4, OLS is consistent, i.e.,

$$\widehat{\beta}_j \xrightarrow{P} \beta_j, j = 1, \cdots, k.$$

Proof of the simple case. We have

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i/n}{\sum_{i=1}^{n}(x_i - \overline{x})^2/n}.$$

We have

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})u_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mathbb{E}[x_i] + \mathbb{E}[x_i] - \overline{x})u_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mathbb{E}[x_i])u_i + \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[x_i] - \overline{x})u_i.$$

By A3, $\frac{1}{n}\sum_{i=1}^{n}(x_i - \mathbb{E}[x_i])u_i$ is i.i.d., and by LLN

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mathbb{E}[x_i])u_i \xrightarrow{P} \mathbb{E}[(X - \mathbb{E}[X])U] = \mathbb{E}[(X - \mathbb{E}[X])(U - \mathbb{E}[U])] = \mathrm{Cov}(X, U).$$

In fact,

$$\begin{aligned}\mathrm{Cov}(X,U) &= \mathbb{E}[(X-\mathbb{E}[X])U] = \mathbb{E}[XU]-\mathbb{E}[U]\mathbb{E}[X]\\ &= \mathbb{E}[\mathbb{E}[XU|X]] = \mathbb{E}[\mathbb{E}[X]\mathbb{E}[U|X]] = 0.\end{aligned}$$

Also,

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[x_i]-\overline{x})u_i = \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[x]-\overline{x})u_i \xrightarrow{P} 0.$$

Besides,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2 \xrightarrow{P} \mathbb{E}[X^2]-(\mathbb{E}[X])^2 = \mathrm{Var}[X].$$

Therefore,

$$\widehat{\beta}_1 \xrightarrow{P} \beta_1 + \frac{0}{\mathrm{Var}[X]} = \beta_1.$$

□

6.2 Asymptotic Distribution

Definition 6.2 (Asymptotic Distribution)**.** Let $\{Z_1.\cdots,Z_n,\cdots\}$ be a sequence of random variables s.t.

$$\forall z, P(Z_n \leqslant z) = F_{Z_n}(z) \to F_Z(z) = P(Z \leqslant z),$$

as $n \to \infty$. We say Z is an asymptotic distribution of Z_n, and write $Z_n \overset{a}{\sim} Z$.

Theorem 6.3 (Central Limit Theorem)**.** Let $\{X_1.\cdots,X_n\}$ be a random sample with mean μ and variance σ^2, then

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

Theorem 6.4. Under A1 to A5,

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

and

$$\frac{\widehat{\beta}_j-\beta_j}{\mathrm{SE}\left(\widehat{\beta}_j\right)} \overset{a}{\sim} \mathcal{N}(0,1).$$

With the theorem above, the usual CI, t-test and F-test are asymptotically valid.

7 Multiple Regression: Further Issues

7.1 Beta Coefficients

Suppose $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_k x_k + \widehat{u}$, then

$$y_i - \overline{y} = \widehat{\beta}_1(x_1 - \overline{x}_1) + \cdots + \widehat{\beta}_k(x_k - \overline{x}_k) + \widehat{u}_i,$$

and thus

$$\frac{y - \overline{y}}{\widehat{\sigma}_y} = \frac{\widehat{\sigma}_1}{\widehat{\sigma}_y}\widehat{\beta}_1 \frac{x_1 - \overline{x}_1}{\widehat{\sigma}_1} + \cdots + \frac{\widehat{\sigma}_k}{\widehat{\sigma}_y}\widehat{\beta}_k \frac{x_k - \overline{x}_k}{\widehat{\sigma}_k} + \frac{\widehat{u}}{\widehat{\sigma}_y}.$$

Let $\widehat{b}_j = \frac{\widehat{\sigma}_j}{\widehat{\sigma}_y}\widehat{\beta}_j$ and $z_j = \frac{x_j - \overline{x}_j}{\widehat{\sigma}_j}$, then

$$z_y = \widehat{b}_1 z_1 + \cdots + \widehat{b}_k z_k + error,$$

where $\widehat{b}_j$ is the standardized coefficient or beta coefficient.

Beta coefficients have different interpretation: if x_j increases by one standard deviation, then $\widehat{y}$ changes by $\widehat{b}_j$ standard deviations.

7.2 Qualitative Information: Binary/Dummy Variables

Example 7.1. Consider the following simple model of hourly wage determination:

$$wage = \beta_0 + \beta_1 female + \beta_2 educ + u.$$

Because $female = 1$ when the person is female and $female = 0$ when the person is male, the parameter β_1 has the following interpretation: β_1 is the difference in hourly wage between females and males, given the same amount of education and the same error term u. In this example, we choose males to be the base group or benchmark group

8 Heteroskedasticity

If homoscedasticity assumption is dropped, the formula for SE $\left(\widehat{\beta}_j\right)$ is wrong, OLS is not guaranteed to be BLUE (but still unbiased and consistent), and the t-test and CI are invalid.

8.1 Heteroskedasticity-Robust SE

First, consider the simple regression, and we have

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

Under A1 to A4,

$$\text{Var}\left[\widehat{\beta}_1|x_i\right] = \frac{1}{SXX^2}\text{Var}\left[\sum_{i=1}^{n}(x_i - \overline{x})u_i\right] = \frac{1}{SXX^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\text{Var}[u_i|x_i] = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2\sigma_i^2}{SXX^2}.$$

Since

$$\text{Var}[u|x] = \mathbb{E}[u^2|x] - (\mathbb{E}[u|x])^2 \overset{\text{A2}}{=} \mathbb{E}[u^2|x],$$

then

$$\widehat{\text{Var}}\left[\widehat{\beta}_1|x_i\right] = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2\widehat{u}_i^2}{SXX^2}.$$

In multiple regression, $\widehat{\text{Var}}\left[\widehat{\beta}_j\right]$ is a consistent estimator of $\text{Var}\left[\widehat{\beta}_j\right]$ and thus

$$T = \frac{\widehat{\beta}_j - \beta_j}{\text{SE}\left(\widehat{\beta}_j\right)} \overset{a}{\sim} \mathcal{N}(0,1).$$

In STATA, using `reg y x, robust`.

8.2 Generalized Least Squares (GLS)

Assume that

$$\text{Var}[u|\mathbf{x}] = \sigma_u^2 h(\mathbf{x}),$$

where $h(\mathbf{x})$ is some function of the explanatory variables and must be positive. For example, $h(\mathbf{x}) = e^{\gamma_0 + \gamma_1 x_1 + \cdots + \gamma_k x_k}$. Therefore,

$$\sigma_u^2 h(\mathbf{x}) = \text{Var}[u|\mathbf{x}] \overset{\text{A2}}{=} \mathbb{E}[u^2|\mathbf{x}] \Rightarrow \sigma_u^2 = \frac{\mathbb{E}[u^2|\mathbf{x}]}{h(\mathbf{x})} = \mathbb{E}\left[\left(\frac{u}{\sqrt{h(\mathbf{x})}}\right)^2 \middle| \mathbf{x}\right] := \mathbb{E}[(u^*)^2|\mathbf{x}] = \text{Var}[u^*|\mathbf{x}].$$

For original equation $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$, we divide both side by $\sqrt{h(\mathbf{x})}$ and have

$$\frac{y}{\sqrt{h(\mathbf{x})}} = \beta_0\frac{1}{\sqrt{h(\mathbf{x})}} + \beta_1\frac{x_1}{\sqrt{h(\mathbf{x})}} + \cdots + \beta_k\frac{x_k}{\sqrt{h(\mathbf{x})}} + \frac{u}{\sqrt{h(\mathbf{x})}},$$

or

$$y^* = \beta_0 x_0^* + \beta_1 x_1^* + \cdots + \beta_k x^* + u^*$$

with $\text{Var}[u^*|\mathbf{x}] = \sigma_u^2$. Hence, generalized least squares (GLS) estimators are BLUE. Note that in most case we do not know $h(\mathbf{x})$ and thus we use heteroskedasticity-robust.

8.3 Testing for Heteroskedasticity

We want to test

$$H_0 : \text{Var}[u|\mathbf{x}] = \sigma_u^2 \text{ against } H_1 : \text{Var}[u|\mathbf{x}] \neq \sigma_u^2.$$

We regress y on $\mathbf{x}$ and have

$$\widehat{u}_i = y_i - \mathbf{x}_i\widehat{\beta},$$

and then estimate the regression

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + \text{Interaction terms} + v.$$

Then, we can do the F-test on $H_0 : \delta_1 = \delta_2 = \cdots = 0$.

9 Instrumental Variables

If $\text{Cov}(X,U) \neq 0$, OLS is inconsistent. Some reasons cause $\text{Cov}(X,U) \neq 0$: omitted variable bias, measurement error, simultaneity, and/or sample section. We can solve this problem by instrumental variables.

Definition 9.1 (Instrument Variable)**.** If Z is observed and is s.t. $\text{Cov}(Z,U) = 0$ (valid instrumental variable) and $\text{Cov}(Z,X) \neq 0$ (relevant instrumental variable), then we call Z an instrument variable.

Example 9.1. Let $Y = wage, X = education, U = unboserved\ ability$. $Z = SIN$ is not IV since $\text{Cov}(Z,X) = 0$. $Z = IQ$ is not IV since $\text{Cov}(Z,U) \neq 0$. Note that good proxies are bad IV. $Z = tutition\ subsidies$ could be IV, since $\text{Cov}(Z,X) \neq 0$, and $\text{Cov}(Z,U)$ may be 0 depends on how subsidies are allocated.

Consider simple regression with $\text{Cov}(X,U) \neq 0$, i.e., dropping A2 (zero conditional mean).

Assumption 2'. $\text{Cov}(Z,U) = 0, \text{Cov}(Z,X) \neq 0$.

Assumption 3'. $\{(y_i, x_i, z_i), i = 1, \cdots, n\}$ is i.i.d.

Assumption 5'. $\text{Var}[U|Z] = \sigma_U^2$,

Property 9.1. $\beta_1 = \frac{\text{Cov}(Y,Z)}{\text{Cov}(X,Z)}$ under A1, A2' and A3'.

Proof. We have $\mathbb{E}[Y] = \beta_0 + \beta_1 \mathbb{E}[X] + \mathbb{E}[U]$ and thus

$$\mathbb{E}[Y]\mathbb{E}[Z] = \beta_0 \mathbb{E}[Z] + \beta_1 \mathbb{E}[X]\mathbb{E}[Z] + \mathbb{E}[U]\mathbb{E}[Z].$$

Besides, since $YZ = \beta_0 Z + \beta_1 XZ + UZ$, then

$$\mathbb{E}[YZ] = \beta_0 \mathbb{E}[Z] + \beta_1 \mathbb{E}[XZ] + \mathbb{E}[UZ].$$

Hence,

$$\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z] = \beta_1(\mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z]) + \mathbb{E}[UZ] - \mathbb{E}[U]\mathbb{E}[Z],$$

i.e.,

$$\text{Cov}(Y,Z) = \beta_1 \text{Cov}(X,Z) + \text{Cov}(U,Z) \Rightarrow \beta_1 = \frac{\text{Cov}(Y,Z)}{\text{Cov}(X,Z)}.$$

□

The estimator is

$$\widehat{\beta}_1^{\text{IV}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})(x_i - \overline{x})}.$$

Property 9.2. Under A1, A2' and A3',

$$\widehat{\beta}_1^{\text{IV}} \xrightarrow{P} \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,X)} = \beta_1,$$

i.e., $\widehat{\beta}_1^{\text{IV}}$ is consistent.

Property 9.3. $\widehat{\beta}_1^{\text{IV}}$ is biased.

Property 9.4. Under A1, A2', A3' and A5',

$$\widehat{\mathrm{Var}}\left[\widehat{\beta}_1^{\mathrm{IV}}\right] = \frac{\widehat{\sigma}_U^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2 R_{XZ}^2},$$

where R_{XZ}^2 is the R^2 obtained by regression X on Z.

Property 9.5. Under A1, A2', A3' and A5',

$$T = \frac{\widehat{\beta}_1^{\mathrm{IV}} - \beta_1}{\mathrm{SE}\left(\widehat{\beta}_1^{\mathrm{IV}}\right)} \overset{a}{\sim} \mathcal{N}(0,1).$$

In STATA, using `ivreg y (x = z), (robust)`.

9.1 Comparison of OLS and IV

- If $\mathrm{Cov}(X,U) = 0$ and $\mathrm{Cov}(Z,U) = 0$, then $\widehat{\beta}^{\mathrm{OLS}} \xrightarrow{P} \beta$ and $\widehat{\beta}^{\mathrm{IV}} \xrightarrow{P} \beta$. Also under homoscedasticity,

  $$\widehat{\mathrm{Var}}\left[\widehat{\beta}^{\mathrm{OLS}}\right] = \frac{\widehat{\sigma}_u^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \leqslant \frac{\widehat{\sigma}_u^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2 R_{XZ}^2} = \widehat{\mathrm{Var}}\left[\widehat{\beta}^{\mathrm{IV}}\right].$$

  Therefore, we use OLS.

- If $\mathrm{Cov}(X,U) \neq 0$ and $\mathrm{Cov}(Z,U) = 0$, then $\widehat{\beta}^{\mathrm{OLS}}$ is inconsistent to β and $\widehat{\beta}^{\mathrm{IV}} \xrightarrow{P} \beta$. Though $\widehat{\mathrm{Var}}\left[\widehat{\beta}^{\mathrm{OLS}}\right] \leqslant \widehat{\mathrm{Var}}\left[\widehat{\beta}^{\mathrm{IV}}\right]$, we use IV.

- If $\mathrm{Cov}(X,U) \neq 0$ and $\mathrm{Cov}(Z,U) \neq 0$, then

  $$\widehat{\beta}^{\mathrm{OLS}} \xrightarrow{P} \beta + \frac{\mathrm{Cov}(X,U)}{\mathrm{Var}[X]}, \widehat{\beta}^{\mathrm{IV}} \xrightarrow{P} \beta + \frac{\mathrm{Cov}(Z,U)}{\mathrm{Cov}(Z,X)}.$$

  In general ,it is unclear which one is better. But if $\mathrm{Cov}(Z,X)$ is very small, even if $|\mathrm{Cov}(Z,U)| < |\mathrm{Cov}(X,U)|$, we still have

  $$\left|\frac{\mathrm{Cov}(X,U)}{\mathrm{Var}[X]}\right| < \left|\frac{\mathrm{Cov}(Z,U)}{\mathrm{Cov}(Z,X)}\right|.$$

9.2 Weak IV

When $\mathrm{Cov}(Z,X)$ is small, we say Z is weak. There are two problems from weak IV:

- $\widehat{\mathrm{Var}}\left[\widehat{\beta}^{\mathrm{IV}}\right] = \frac{\widehat{\sigma}_u^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2 R_{XZ}^2}$ becomes larger since R_{XZ}^2 is smaller;

- $\frac{\widehat{\beta}^{\mathrm{IV}}-\beta}{\mathrm{SE}\left(\widehat{\beta}^{\mathrm{IV}}\right)} \overset{a}{\nsim} \mathcal{N}(0,1)$ and thus usual CI, t and F test are invalid.

9.2.1 Detection of Weak IV

We regress X on Z and get the F statistics. If $F < 10$, then Z is a weak IV; otherwise, Z is a strong IV.

9.3 IV Estimation of the Multiple Regression Model

Consider the structural equation

$$Y = \beta_0 + \beta_1 X + \beta_2 Z_1 + \beta_3 Z_2 + \cdots + \beta_k Z_{k-1} + U,$$

where X is endogenous regressor ($\text{Cov}(X, U) \neq 0$), $Z_1, \cdots, Z_{k-1}$ are exogenous regressor ($\text{Cov}(Z_j, U) = 0$), and $Z_k, \cdots, Z_q$ are instrumental variables ($\text{Cov}(Z_j, U) = 0$). We say $Z = (Z_1, \cdots, Z_{k-1}, Z_k, \cdots, Z_q)$ is exogenous.

We call

$$X = \pi_0 + \pi_1 Z_1 + \cdots + \pi_q Z_q + V = Z\pi + V$$

as reduced form equation, where $\text{Cov}(Z_j, V) = 0$.

In this case we have

$$\begin{aligned} 0 \neq \text{Cov}(X, U) &= \text{Cov}(Z\pi + V, U) = \text{Cov}(\pi_0 + \pi_1 Z_1 + \cdots + \pi_q Z_q + V, U) \\ &= \pi_1 \text{Cov}(Z_1, U) + \cdots + \pi_q \text{Cov}(Z_q, U) + \text{Cov}(V, U) \\ &= \text{Cov}(V, U). \end{aligned}$$

In STATA, using `y (x = zk zk+1 ... zq) z1 ... zk-1, robust`.

9.4 Two Stage Least Squares (2SLS)

In first stage, we regress X on all Z's $(Z_1, \cdots, Z_q)$ and get

$$\widehat{X}_i = Z_i \widehat{\pi}.$$

In second stage, we regress Y on $\widehat{X}_i$ and all exogenous regressors $Z_1, \cdots, Z_{k-1}$ (no IV included).

Mathematically,

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \beta_2 Z_1 + \cdots + \beta_k Z_{k-1} + U \\ &= \beta_0 + \beta_1(\pi_0 + \pi_1 Z_1 + \cdots + \pi_{k-1} Z_{k-1} + \pi_k Z_k + \cdots + \pi_q Z_q + V) + \beta_2 Z_1 + \cdots + \beta_k Z_{k-1} + U \\ &= (\beta_0 + \beta_1 \pi_0) + (\beta_1 \pi_1 + \beta_2) Z_1 + \cdots + (\beta_1 \pi_{k-1} + \beta_k) Z_{k-1} + \beta_1 \pi_k Z_k + \cdots + \beta_1 \pi_q Z_q + (U + \beta_1 V) \\ &:= \alpha_0 + \alpha_1 Z_1 + \cdots + \alpha_q Z_q + \varepsilon. \end{aligned}$$

Since $\alpha_k = \beta_1 \pi_k, \cdots, \alpha_q = \beta_1 \pi_q$ and $\alpha_j, \pi_j, j = k, \cdots, q$ can be estimated, then we can find $\widehat{\beta}_1$. For $\alpha_1 = \beta_1 \pi_1 + \beta_2, \cdots, \alpha_{k-1} = \beta_1 \pi_{k-1} + \beta_k$, we can also find $\widehat{\beta}_i, i = 2, \cdots, k$.

In STATA, using `ivreg 2sls y (x = ...)`.

9.5 Endogeneity Test

We know that OLS is BLUE, and unless there is a compelling reason, we should use OLS. If we have a suspicious OLS estimator due to potential endogeneity and a plausible 2SLS estimator, we can test if $\text{Cov}(X, U) = 0$.

We cannot use OLS to test this since the sample covariance between x_i and $\widehat{u}_i = y_i - x_i \widehat{\beta}^{\text{OLS}}$ is zero by construction $\sum_{i=1}^{n} \widehat{u}_i x_i = 0$. But we can use a valid instrument to test.

Suppose $Y = \beta_0 + \beta_1 X + U$ and $X = \pi_0 + \pi_1 Z + V$ with $\text{Cov}(Z, U) = \text{Cov}(Z, V) = 0$, then $\text{Cov}(X, U) = 0$ iff $\text{Cov}(V, U) = 0$, which is what we want to test. Let

$$U = \delta V + e,$$

where $\text{Cov}(V, e) = 0, \mathbb{E}[e] = 0$. U and V are uncorrelated iff $\delta = 0$.

Therefore, we have

$$Y = \beta_0 + \beta_1 X + \delta V + e,$$

which is called control function and is numerically equivalent to 2SLS.

We take endogeneity test following three steps:

- Step 1: Regress X on Z (OLS) and get $\widehat{V} = X - Z\widehat{\pi}$.
- Step 2: Regress Y on X and $\widehat{V}$ (OLS).
- Step 3: Test $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$ (t test).

9.6 Over-Identification Test

Note that control function depends on $\text{Cov}(Z, U) = 0$ but we cannot use IV estimator to test it. In the data, the covariance between z_i and $\widehat{u}_i = y_i - x_i \widehat{\beta}^{\text{IV}}$ is zero by construction $\sum_{i=1}^{n} \widehat{u}_i z_i = 0$.

Suppose $Y = \beta_0 + \beta_1 X + U$ and $X = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + V$ with

$$\text{Cov}(Z_1, U) = \text{Cov}(Z_2, U) = \text{Cov}(Z_1, V) = \text{Cov}(Z_2, V) = 0.$$

If we run 2SLS using Z_1 or Z_2 only, we have $\widetilde{\beta}_1^{\text{IV}}$ and $\breve{\beta}_1^{\text{IV}}$, respectively. Under the assumption that both Z_1 and Z_2 are valid IVs, we should have $\widetilde{\beta}_1^{\text{IV}} \approx \breve{\beta}_1^{\text{IV}}$. If they are different, then either Z_1 or Z_2 or both are invalid.

We take over-identification test following three steps:

- Step 1: Estimate structural equation using all IVs and get $\widehat{U}$ (2SLS).
- Step 2: Regress $\widehat{U}$ on all exogenous variables (OLS) and get R^2.
- Step 3: Test $H_0 : \text{Cov}(Z_1, U) = \text{Cov}(Z_2, U) = 0$ against H_1 : At least one of covariance is not zero. Under $H_0, nR^2 \overset{a}{\sim} \chi^2_{(q)}$ where q = Number of IVs − Number of endogenous regressors and n = Number of observations.

We cannot test if we have one IV. If we have more IVs than endogenous regressors, we can run the over-identification test. If we reject H_0, we only know either Z_1 or Z_2 or both are invalid but we do not know which one is invalid. If we do not reject H_0, it does not guarantee that both IVs are valid because they could both be invalid and deliver deliver similar estimates, $\widetilde{\beta}_1^{\text{IV}} \approx \breve{\beta}_1^{\text{IV}}$.