Lecture Notes: Econometrics I

Based on lectures by **Marko Milkota** in Autumn semester, 2024

Draft updated on December 2, 2024

These lecture notes were taken in the course *Econometrics I* taught by **Marko Milkota** at Graduate of International and Developoment Studies, Geneva as part of the International Economics program (Semester I, 2024).

Currently, these are just drafts of the lecture notes. There can be typos and mistakes anywhere. So, if you find anything that needs to be corrected or improved, please inform at `jingle.fu@graduateinstitute.ch`.

Contents

Lecture 1.

Statistical Inference

$$x : X \sim p(x|\theta)$$

θ is the parameter setting the shape of the distribution.

Definition 1.0.1. Point Estimator $\delta(X)$

A mapping δ from sample space of X to the parameter space Θ: $\delta : X \to \Theta$

Given X, what's the best Θ. $\delta(x)$ is an estimate.

1.1 Methods to get $\delta(X)$

1.1.1 LS Estimation

$$X_1 \sim p(x|\theta), \quad \mathbb{E}[X_1|\theta] = \theta, \quad X_1|\theta \sim N(0,1)$$

Point estimator $\hat{\theta}$ is the argument that minimizes the objective function

$$\hat{\theta}_{\text{LS}} = \arg\min_{\theta} \sum_{i=1}^{n} (x_i - \theta)^2$$

where we assume that $\theta = \mathbb{E}[x_i|\theta]$. Using the First Order Condition (FOC) to solve, we have

$$\frac{\partial(\cdot)}{\partial\theta} = \sum_{i=1}^{n} -2(x_i - \theta) = 0$$

We get $\hat{\theta}_{\text{LS}} = \frac{1}{n}\sum_{i=1}^{n} x_i$

1.1.2 Method of Moments

Find $\hat{\theta}$ such that

$$\mathbb{E}[X|\hat{\theta}_{\text{MM}}] = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Or, such that

$$\mathbb{E}[X^2|\theta]_{\theta=\hat{\theta}_{\text{MM}}} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$\mathbb{V}[X|\theta] = 1$, thus $\mathbb{E}[X^2|\theta] = \mathbb{V}[X|\theta] + (\mathbb{E}[X|\theta]^2) = 1 + \theta^2$.

$$1 + \hat{\theta}_{\text{MM}}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

We get

$$\hat{\theta}_{\text{MM}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - 1}$$

> **Note.** Choose $\hat{\theta}_{\text{MM}}$ s.t. $\mathbb{E}[h(X)|\theta]$ under $\theta = \hat{\theta}_{\text{MM}}$ is the mean of samples $\frac{1}{n}\sum_{i=1}^n x_i^2$.

1.1.3 Maximum Likelihood

$$\hat{\theta}_{\text{ML}} = \arg\max_{\theta} \mathcal{L}(\theta|x)$$

where

$$\mathcal{L}(\theta|x) = p(x|\theta)$$

is the PDF of the RV $X|\theta$.

> **Note.** Specify the whole distribution.

With X as i.i.d. distribution, we have

$$\begin{aligned}
\mathcal{L}(\theta|x) &= p(x|\theta) \\
&= \prod_{i=1}^n p(x_i|\theta) \\
&= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right\} \\
&\sim N(0,\sigma^2)
\end{aligned}$$

Then, let's define $\ell(\theta|x) = \log \mathcal{L}(\theta|x) = -\frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$

$$\hat{\theta}_{\text{ML}} = \frac{1}{n}\sum_{i=1}^n x_i = \hat{\theta}_{\text{MM}} = \hat{\theta}_{\text{LS}}$$

$$\begin{aligned}
\hat{\theta}_{\text{ML}} &= \arg\max_{\theta} -\frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2 \\
&= \arg\min_{\theta} \sum(x_i-\mu)^2
\end{aligned}$$

> **Definition 1.1.1.** $\delta(X)$ is unbiased if $\mathbb{E}[\delta(X)|\theta] = \theta$.

e.g. $\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i$ is unbiased because

$$\begin{aligned}
\mathbb{E}[\hat{\theta}|\theta] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i|\theta\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum X_i\right] \\
&= \frac{1}{n}\sum \mathbb{E}[X_i] \\
&= \theta
\end{aligned}$$

However, $\hat{\theta}_* = \frac{1}{n-1}\sum_{i=1}^n X$ is biased, because

$$\mathbb{E}[\hat{\theta}_*|\theta] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n X_i|\theta\right]$$

$$\begin{aligned} &= \frac{1}{n-1}\mathbb{E}\left[\sum X_i\right] \\ &= \frac{n}{n-1}\theta \\ &\neq \theta \end{aligned}$$

For the variance,

$$\mathbb{V}[\hat{\theta}|\theta] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i|\theta] = \frac{\sigma^2}{n}$$

and,

$$\mathbb{V}[\hat{\theta}_*|\theta] = \mathbb{V}\left[\frac{n}{n-1}\hat{\theta}\right] = \frac{n^2}{(n-1)^2}\mathbb{V}[\hat{\theta}] = \frac{n\sigma^2}{(n-1)^2}$$

To get the estimation of $\mathbb{E}$ and $\mathbb{V}$, we have to assume the mean and variance of the distribution. Under $X_i|\theta \sim N(0,1)$, we have $\hat{\theta} = \frac{1}{n}\sum X_i \sim N\left(0, \frac{1}{n}\right)$.

1.2 Asymptotic Properties

Definition 1.2.1. Point estimator $\delta(X)$ is consistent if $\delta(X) \xrightarrow{p} \theta$.

By Weak Law of Large Numbers (WLLN),

$$\hat{\theta} = \frac{1}{n}\sum X_i \xrightarrow{p} \mathbb{E}[X] = \theta$$

Now let's look at $\hat{\theta}_*$ again,

$$\hat{\theta}_* = \frac{1}{n-1}\sum X_i = \frac{n}{n-1}\frac{1}{n}\sum X_i = \frac{n}{n-1}\mathbb{E}[X] = \frac{n}{n-1}\hat{\theta} \xrightarrow{p} \theta$$

as $\lim\limits_{n\to\infty}\frac{n}{n-1} = 1$ and $\hat{\theta} \xrightarrow{p} \theta$. Slutsky's theorem tells us that we can form the limit of their product as the product of the limits.

Central Limit Theorem (CLT):

$$\begin{aligned} &\sqrt{n}\frac{\frac{1}{n}\sum X_i - \mathbb{E}[X_i]}{\sqrt{\mathbb{V}[X_i]}} \xrightarrow{p} N(0,1) \\ \Rightarrow &\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0,1) \text{ Set the variance to 1} \\ \Rightarrow &\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0,1) \\ \Rightarrow &\sqrt{n}(\hat{\theta} - \theta) \overset{\text{approx}}{\sim} N(0,1) \text{ in our finite sample of } n \\ \Rightarrow &(\hat{\theta} - \theta) \sim N\left(0, \frac{1}{n}\right) \\ \Rightarrow &\hat{\theta} \overset{\text{approx}}{\sim} N\left(\theta, \frac{1}{n}\right) \end{aligned}$$

RECALL: Statistical inference given x, what can we say?

1. Point inference, best guess for θ

2. Hypothesis testing, is θ larger than 1 or not?

3. Interval inference, give an interval where you are sure that θ lies in.

Lecture 2.

Problem Set 1

Solution.

For the Least Squares estimation, we need to solve the following problem:

$$\min \sum_{i=1}^{n} (x_i - \theta)^2$$

Denote $F = \sum\limits_{i=1}^{n} (x_i - \theta)^2$ and take the first order derivative with respect to θ, we have the FOC:

$$\begin{aligned} \frac{\partial F}{\partial \theta} &= \sum -2(x_i - \theta) = 0 \\ \Rightarrow \hat{\theta} &= \frac{1}{n} \sum_{i=1}^{n} x_i \end{aligned}$$

Solution.

The mean of $\hat{\theta}$ is its expectation, so we calculate:

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} x_i\right] \overset{1}{=} \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[x_i] \overset{2}{=} \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[u_i] = \frac{1}{n}\cdot n\theta = \theta$$

1. We are using the property of $\mathbb{E}$: $\mathbb{E}\left[\sum\limits_{i=1}^{n} x_i\right] = \sum\limits_{i=1}^{n} \mathbb{E}[x_i]$, no matter x_i are independent or not.
2. We are using the property of $\mathbb{E}$: $\mathbb{E}[\theta + u_i] = \mathbb{E}[u_i] + \theta$ if θ is a constant number.

Thus, $\hat{\theta}$ is unbiased and we make no other assumptions on pdf of $x_i|\theta$ and the sample $\{x_i\}_{i=1}^{n}$.

Solution.

$$\mathbb{V}[\hat{\theta}] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[x_i] = \frac{\sigma^2}{n}$$

By having this result, we make the following 2 assumptions:

1. For sample set $\{x_i\}_{i=1}^{n}$, we assume that for all samples x_i, they are mutually independent, which gives that $\mathbb{V}\left[\sum\limits_{i=1}^{n} u_i\right] = \sum\limits_{i=1}^{n} \mathbb{V}[u_i]$.
2. For the pdf of $x_i|\theta$, we assume that x_i are independently distributed, which means that $\mathbb{V}[x_i] = \mathbb{V}[u_i] = \sigma^2$

Lecture 3.

Hypothesis Testing

3.1 Some Basic Concepts

Definition 3.1.1. Null Hypothesis

The null hypothesis $\mathcal{H}_0$ is the set $\theta = \theta_0$ or $\beta \in \mathcal{B}_0$.

Or, we denote it as:

$$\mathcal{H}_0 : \theta \in \Theta_0$$

For econometrics, we usually set $\mathcal{H}_0 : \beta = 0$.

Definition 3.1.2. Alternative Hypothesis

The alternative hypothesis $\mathcal{H}_1$ is the set $\{\theta \in \Theta : \theta \neq \theta_0\}$ or $\{\beta \in \mathcal{B} : \beta \notin \mathcal{B}_0\}$.

Or, we denote it as:

$$\mathcal{H}_1 : \theta \in \Theta_1$$

For econometrics, we usually set $\mathcal{H}_1 : \beta \neq 0$. Often Θ_1 is the complement of Θ_0.

Note. Point estimator of $\mathbf{1}\{\theta \in \Theta_0\}$ (if $\theta \in \Theta_0$, the function equals 1).

	$\mathcal{H}_0$ is true	$\mathcal{H}_0$ is false
Accept $\mathcal{H}_0$	$\checkmark$: $1-\alpha$	$\times$: $1-\beta$
Reject $\mathcal{H}_0$	$\times$: α	$\checkmark$: β

α is the Type I error, $1-\beta$ is the Type II error.

Definition 3.1.3. A hypothesis test $\varphi \in \{0, 1\}$ is a rule that specifies when we reject and when we accept (do not reject) $\mathcal{H}_0$, with $\varphi = 0$ indicating rejection.

Definition 3.1.4. Power Function

$$\beta(\theta) = \mathbb{P}[\text{rejecting}|\theta \text{ is true}] = \mathbb{P}[\varphi = 0|\theta]$$

Definition 3.1.5. Size of a test

The size of a test is α if $\sup\limits_{\theta \in \Theta_0} \beta(\theta_0) = \alpha$, $\alpha \in (0, 1)$.

Generic form:

$$\varphi(x; \alpha) = \mathbf{1}\{T(x) < c_\alpha\}$$

3.2 T-test

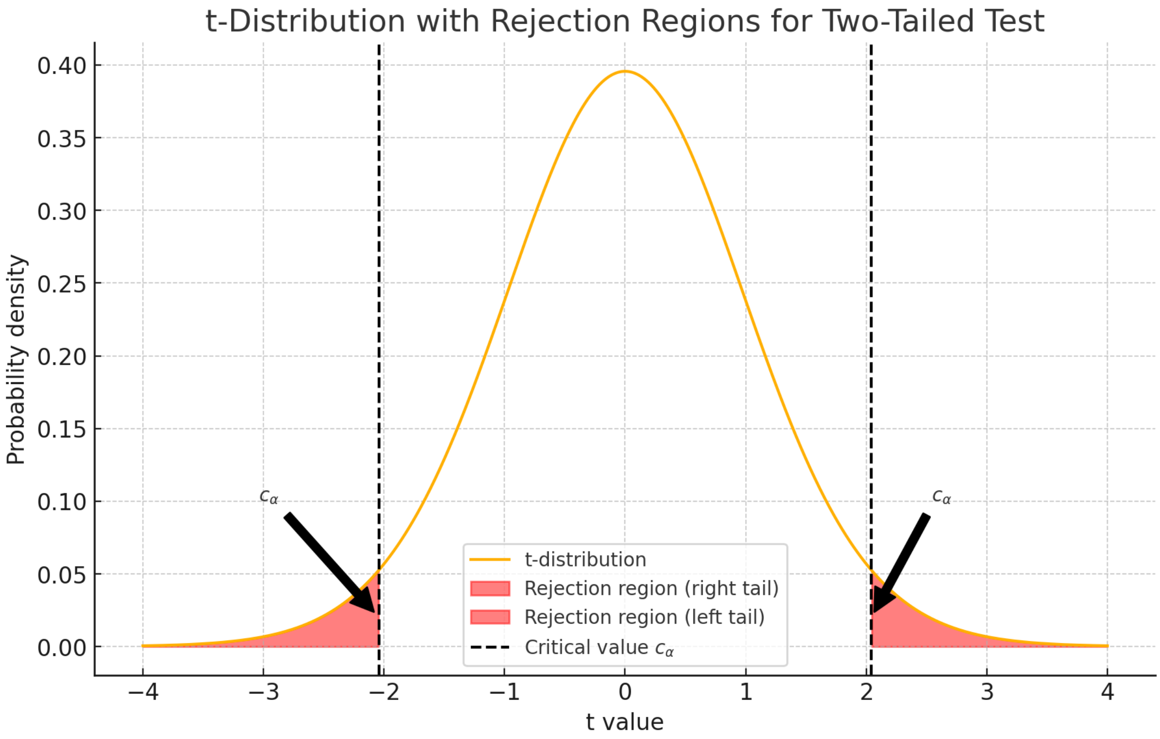
Definition 3.2.1. Suppose $\hat{\theta}|\theta \sim N(\theta, v^2)$, and we are testing a point hypothesis $\mathcal{H}_0 : \theta = \theta_0$.

Under the alternative $\mathcal{H}_1 : \theta \neq \theta_0$, the two-sided t-test is

$$\varphi_t(x) = \mathbf{1}\left\{\left|\frac{\hat{\theta}-\theta_0}{v}\right| < c\right\}$$

Note. The test-statistic $T(X) = \left|\frac{\hat{\theta}-\theta_0}{v}\right|$ is a function of data X because our estimator θ, $\hat{\theta}$, is a function of X.

Example 1. Let $X|\theta \sim N(\theta, v^2)$, under $\mathcal{H}_0 : \hat{\theta} \sim N(\theta_0, v^2) \to \frac{\hat{\theta}-\theta_0}{v} \sim N(0,1)$



$$\begin{aligned}
\beta(\theta) &= \mathbb{P}[\text{rejecting}|\theta \text{ is true}] \\
&= \mathbb{P}[\varphi = 0|\theta] \\
&= \mathbb{P}[T(X) > c_\alpha|\theta] \\
&= \mathbb{P}\left[\left|\frac{\hat{\theta}-\theta_0}{v}\right| > c_\alpha|\theta\right] \\
&= 1 - \mathbb{P}\left[\left|\frac{\hat{\theta}-\theta_0}{v}\right| \leq c_\alpha|\theta\right] \\
&= 1 - \mathbb{P}\left[-c_\alpha \leq \frac{\hat{\theta}-\theta_0}{v} \sim N(0,1) \leq c_\alpha|\theta\right] \\
&= 1 - \left(\mathbb{P}\left[\frac{\hat{\theta}-\theta_0}{v} \leq c_\alpha|\theta\right] - \mathbb{P}\left[\frac{\hat{\theta}-\theta_0}{v} \leq -c_\alpha|\theta\right]\right) \\
&= 1 - [\Phi(c_\alpha) - \Phi(-c_\alpha)] \\
&= 1 - [\Phi(c_\alpha) - (1 - \Phi(c_\alpha))] \\
&= 2 - 2\Phi(c_\alpha)
\end{aligned}$$

$$= \alpha$$

Under $\alpha = 0.05$, we get $c_\alpha = 1.64$, $\alpha = 0.1$, we get $c_\alpha = 1.96$.

To compute the power of this test, we need to think about what happens if $\mathcal{H}_0$ is false. Assuming that $\tilde{\theta}$ is the true value of θ.

$$\begin{aligned}\beta(\tilde{\theta}) &= \mathbb{P}[\text{rejecting}|\tilde{\theta} \text{ is true}] \\ &= \mathbb{P}[\varphi = 0|\tilde{\theta}] \\ &= \mathbb{P}[T(X) > c_\alpha|\tilde{\theta}] \\ &= \mathbb{P}\left[\left|\frac{\hat{\theta}-\theta_0}{v}\right| > c_\alpha|\tilde{\theta}\right]\end{aligned}$$

To find this, under $\tilde{\theta}$ is true, $\tilde{\theta} \sim N(\tilde{\theta}, v^2)$, $\hat{\theta}-\tilde{\theta} \sim N(0, v^2)$, $\hat{\theta}-\theta_0 \sim N(\tilde{\theta}-\theta_0, v^2)$, $\frac{\hat{\theta}-\theta_0}{v} \sim N(\tilde{\theta}-\theta_0, 1)$, $\frac{\hat{\theta}-\theta_0}{v} - (\tilde{\theta}-\theta_0) \sim N(0,1)$

So,

$$\begin{aligned}\beta(\tilde{\theta}) &= \mathbb{P}\left[\left|\frac{\hat{\theta}-\theta_0}{v}\right| > c_\alpha|\tilde{\theta}\right] \\ &= 1 - \mathbb{P}\left[\left|\frac{\hat{\theta}-\theta_0}{v}\right| \leq c_\alpha|\theta\right] \\ &= 1 - \mathbb{P}\left[-c_\alpha \leq \frac{\hat{\theta}-\theta_0}{v} \leq c_\alpha|\theta\right] \\ &= 1 - \mathbb{P}[-c_\alpha - (\tilde{\theta}-\theta_0) \leq z \sim N(0,1) \leq c_\alpha - (\tilde{\theta}-\theta_0)] \\ &= 1 - (\Phi[c_\alpha - (\tilde{\theta}-\theta_0)] - \Phi[-c_\alpha - (\tilde{\theta}-\theta_0)])\end{aligned}$$

The higher the probability of wrongly accepting (or failing to reject) $\mathcal{H}_0$. It is common to be rather conservative (i.e. erring on the side of not rejecting $\mathcal{H}_0$) and report test results for sizes of 10%, 5% and 1%.

3.3 Likelihood Ratio Test

Definition 3.3.1.

$$\varphi_{\text{LR}}(x) = \mathbf{1}\left\{\frac{\sup_{\theta\in\Theta_1} p(x|\theta)}{\sup_{\theta\in\Theta_0} p(x|\theta)} < c_\alpha\right\}, T_{\text{LR}}(X) = \frac{\sup_{\theta\in\Theta_1} p(x|\theta)}{\sup_{\theta\in\Theta_0} p(x|\theta)}.$$

So, if there are points in Θ_0 for which observed x is more likely than points in Θ_1, the ratio is small — the test is likely to accept $\mathcal{H}_0$.

Under $\mathcal{H}_0 : \theta = \theta_0$ and the alternative $\mathcal{H}_1 : \theta \neq \theta_0$,

$$\varphi_{\text{LR}}(x) = \mathbf{1}\left\{\frac{p(x|\hat{\theta}_{\text{ML}})}{p(x|\theta_0)} < c\right\}$$

Example 2. $\{x_i\}_{i=1}^n$, $x_i|\theta \sim N(\theta, 1)$,

$$p(x|\theta) = \prod_i p(x_i|\theta) = (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\theta)^2\right\}$$

if $n = 1$. As $x = \hat{\theta}$, $p(x|\hat{\theta}) = (2\pi)^{-\frac{1}{2}}$

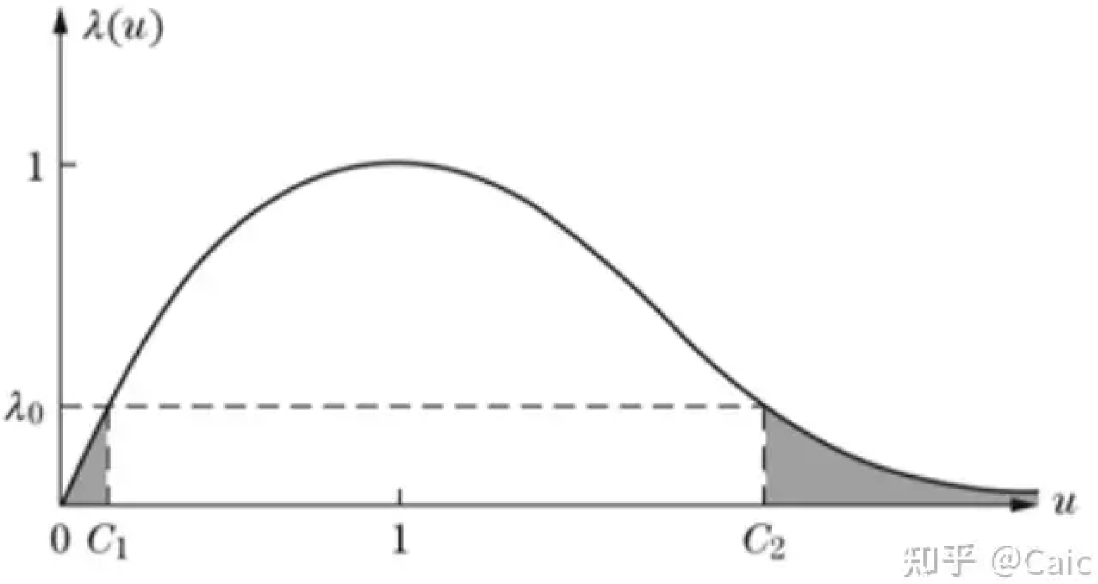
$$T(x) = \frac{p(x|\hat{\theta}_{\mathrm{ML}})}{p(x|\theta_0)} = \frac{(2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\hat{\theta})^2\right\}}{(2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\theta_0)^2\right\}} = \exp\left\{\frac{1}{2}(x-\theta_0)^2\right\}$$

as $x = \hat{\theta}$.

$$\varphi_{\mathrm{LR}} = \mathbf{1}\left\{\exp\left\{\frac{1}{2}(x-\theta_0)^2\right\} < c_\alpha\right\}$$

$$\begin{aligned}\alpha &= \mathbb{P}\{\text{Reject}|\theta_0 \text{ is actually true}\}\\ &= \mathbb{P}\left\{\exp\left\{\frac{1}{2}(x-\theta_0)^2\right\} \geq c_\alpha|\theta_0\right\}\\ &= \mathbb{P}\{(x-\theta_0)^2 \geq 2\log c_\alpha|\theta_0\}\\ &= 1-\mathbb{P}\left\{(x-\theta_0)^2 < 2\log c_\alpha|\theta_0\right\}\end{aligned}$$

If $x \sim N(\theta_0, 1)$, $(x-\theta_0)^2 \sim \chi_1^2$



Note. **Uniformly most powerful test** Highest probability of rejection, of wrong acceptance.

$$\varphi(x) = \mathbf{1}\{T(x) < c_\alpha\}$$
$$\alpha = \mathbb{P}\{T(x) > c_\alpha|\theta_0\}(\text{reject rule})$$

3.3.1 Numerical Hypothesis Testing

Algorithm 1: Numerical Hypothesis Testing

```
Input  : Distribution $N(\theta_0, 1)$, sample size $M$, significance level $\alpha$
Output: Decision to accept or reject $H_0$
for $m = 1$ to $M$ do
    Draw $x^m \sim N(\theta_0, 1)$;
    Compute $T(x^m)$;
end
Sort $\{T(x^m)\}_{m=1}^M$ in ascending order;
Set $c_\alpha$ to the $100(1-\alpha)$ quantile of $\{T(x^m)\}_{m=1}^M$;
Compute $T(x)$ for your observed realization $x$;
if $T(x) \leq c_\alpha$ then
    Accept $H_0$;
else
    Reject $H_0$;
end
```

Small p-value means $\mathcal{H}_0$ is likely to be rejected, larger p-value means it's likely to be true.

3.4 Coverage Sets

3.4.1 Frequentist Confidence Sets

A confidence set $C(X) \subseteq \Theta$ is a (random) set that should cover the true θ with a prespecified probability:

$$\inf_{\theta \in \Theta} \mathbb{P}[\theta \in C(X) | \theta] = 1 - \alpha$$

$$C(x) = \{\theta_0 \in \Theta : \varphi(x; \theta_0) = 1\}$$

contains all the values of θ_0 that we would accept.

Consider $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \neq \theta_0$, $\varphi_\alpha(x) = \mathbf{1}\{T(x) < c_{\alpha;\theta_0}\}$.

Example 3. T-test: $T(x) = \left|\frac{\hat{\theta}-\theta_0}{v}\right| \Rightarrow \varphi\{x; \theta_0, \alpha\} = \mathbf{1}\{T(x) < c\}$

$$\begin{aligned} C(x) &= \{\theta_0 \in \Theta, \varphi\{x; \theta_0, \alpha\} = 1\} \\ &= \left\{\theta_0 \in \Theta, \left|\frac{\hat{\theta} - \theta_0}{v}\right| < c\right\} \\ &= \left\{\theta_0 \in \Theta, -c < \frac{\hat{\theta} - \theta_0}{v} < c\right\} \\ &= \{\theta_0 \in \Theta, -cv + \hat{\theta} < \theta_0 < cv + \hat{\theta}\} \end{aligned}$$

Example 4. LR-test: $\varphi(x; \theta_0, \alpha) = \mathbf{1}\{(x - \theta_0)^2 < \tilde{c_\alpha}\}$

$$\begin{aligned} C(x) &= \{\theta_0 \in \Theta, \varphi\{x; \theta_0, \alpha\} = 1\} \\ &= \{\theta_0 \in \Theta, (x - \theta_0)^2 < \tilde{c_\alpha}\} \\ &= \left\{\theta_0 \in \Theta, -\sqrt{\tilde{c_\alpha}} < x - \theta_0 < \sqrt{\tilde{c_\alpha}}\right\} \\ &= \left\{\theta_0 \in \Theta, x - \sqrt{\tilde{c_\alpha}} < \theta_0 < x + \sqrt{\tilde{c_\alpha}}\right\} \end{aligned}$$

3.4.2 Numerical Confidence Set Construction

Algorithm 2: Numerical Confidence Set Construction

```
Data: Choose a grid $\mathcal{T}$ of values for $\theta_0$
for each $\theta_0 \in \mathcal{T}$ do
    for $m = 1$ to $M$ do
        Draw $x^m \sim N(\theta_0, 1)$ ;                // Distribution of $X$ under $H_0 : \theta = \theta_0$
        Compute $T(x^m, \theta_0)$;
    end
    Get the critical value $c_\alpha(\theta_0)$ as the $(1-\alpha)$th quantile of $\{T(x^m, \theta_0)\}_{m=1}^{M}$;
    Compute $T(x; \theta_0)$ for observed $x$;
    if $T(x; \theta_0) \leq c_\alpha(\theta_0)$ then
        $\theta_0 \in C(x)$;
    else
        $\theta_0 \notin C(x)$;
    end
end
```

Lecture 4.

Problem Set 2

Solution.

Step 1: Write the pdf of observations $x_i|\theta$

Since $x_i = \theta + u_i$, and we assume $u_i \sim \mathcal{N}(0, \sigma^2)$, then $x_i \sim \mathcal{N}(\theta, \sigma^2)$, and we have:

$$p(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right\}.$$

Step 2: Define Likelihood Function

We have assumed that observations in the sample are independent. Thus,

$$L_n(\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

Log-linearize the function, and we define the log-likelihood function:

$$\ell_n(\theta) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2$$

Step 3: Define the Likelihood Estimation problem and find the $\hat{\theta}$

For maximum likelihood estimation, we need to solve the following problem:

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\max_{\theta\in\Theta} \ell_n(\theta)$$

So, we need to maximize:

$$\ell_n(\theta) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2$$

Take the derivative of $\ell_n(\theta)$ with respect to θ, and set it to zero for maximization,

$$\begin{aligned}\frac{\partial \ell_n(\theta)}{\partial\theta} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta)\\ &= \frac{1}{\sigma^2}(\sum_{i=1}^{n}x_i - n\theta)\\ &= 0\end{aligned}$$

Thus, we have

$$\hat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

Solution.

Step 1: Find the likelihood function

For $\mathcal{H}_0 : \theta = \theta_0$, the likelihood function is:

$$L_n(\theta_0) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta_0)^2\right\}$$

For $\mathcal{H}_1 : \theta \neq \theta_0$, the maximum likelihood estimator is $\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i = \hat{\theta}$. The likelihood function is:

$$L_n(\theta_1) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \hat{\theta})^2 \right\}$$

Step 2: Define the likelihood ratio test and c_α

The likelihood ratio and the test is firstly defined as follows(we'll simplify to another version later):

$$\varphi_{LR}(x) = \mathbf{1} \{ LR_n < c \}$$

$$\begin{aligned} LR_n &= \frac{L_n(\theta_1)}{L_n(\theta_0)} \\ &= \exp \left\{ \frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \theta_0)^2 - \sum_{i=1}^{n} (x_i - \hat{\theta})^2 \right] \right\} \end{aligned}$$

Denote

$$D = \sum_{i=1}^{n} (x_i - \theta_0)^2 - \sum_{i=1}^{n} (x_i - \hat{\theta})^2$$

Using the identity:

$$\sum_{i=1}^{n} (x_i - \theta_0)^2 = \sum_{i=1}^{n} (x_i - \hat{\theta} + \hat{\theta} - \theta_0)^2 = \sum_{i=1}^{n} (x_i - \hat{\theta})^2 + n(\hat{\theta} - \theta_0)^2$$

Thus,

$$\begin{aligned} D &= \sum_{i=1}^{n} (x_i - \theta_0)^2 - \sum_{i=1}^{n} (x_i - \hat{\theta})^2 \\ &= \sum_{i=1}^{n} (x_i - \hat{\theta})^2 + n(\hat{\theta} - \theta_0)^2 - \sum_{i=1}^{n} (x_i - \hat{\theta})^2 \\ &= n(\hat{\theta} - \theta_0)^2 \end{aligned}$$

So, the likelihood ratio LR_n is:

$$LR_n = \exp \left\{ \frac{n}{2\sigma^2} (\hat{\theta} - \theta_0)^2 \right\}$$

Then, we simplify the expression and define the test statistic $T(x)$ as below:

$$T(x) = 2 \log (LR_n) = \frac{n}{\sigma^2} (\hat{\theta} - \theta_0)^2$$

And our LR test would be:

$$\varphi_{LR}(x) = \mathbf{1} \left\{ T(x) = \frac{n}{\sigma^2} (\hat{\theta} - \theta_0)^2 < c' \right\}$$

where $c' = 2 \log(c)$. To get a size α test, we find c' so as to set the Type I error to α, which is:

$$\mathbb{P} [T(x) \geq c' | \mathcal{H}_0] = \alpha$$

we can denote that $c' = c_\alpha$.

Step 3: Determine the distribution of $T(x)$ under $\mathcal{H}_0$ and find the value of c_α

Under $\mathcal{H}_0$, $\hat{\theta} \sim \mathcal{N}(\theta_0, \frac{\sigma^2}{n})$, because:

$$\mathbb{E}[\hat{\theta}] = \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^{n} x_i \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[x_i] = \frac{1}{n} \cdot n\theta_0 = \theta_0$$

$$\mathbb{V}[\hat{\theta}] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[x_i] = \frac{1}{n}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Then, standardizing $\hat{\theta}$, we'll have:

$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0,1)$$

Using the hint, we know that

$$Z^2 = \left(\frac{\hat{\theta} - \theta_0}{\sqrt{\sigma^2/n}}\right)^2 = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 \sim \chi_1^2$$

Therefore, under $\mathcal{H}_0$,

$$T(x) = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 = Z^2 \sim \chi_1^2$$

Given $\alpha = 0.05$,

$$c_\alpha = \chi^2_{1,0.95} \approx 3.8415$$

Step 4: Set the decision rule

- Reject $\mathcal{H}_0$: $T(x) > c_\alpha = 3.8415$
- Do not reject $\mathcal{H}_0$: $T(x) \leq c_\alpha = 3.8415$

Solution.

We have $\sigma^2 = 6$, $n = 4$, $x_1 = 178$, $x_2 = 161$,$x_3 = 168$, $x_4 = 172$, $\theta_0 = 175$, so $\hat{\theta} = 169.75$.

Put this data back into our $T(x)$ and LR test, we have:

$$T(x) = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 = \frac{4}{6}(169.75 - 175)^2 = 18.735 > 3.8415$$

We reject $\mathcal{H}_0$.

Solution.

Numerical approximation of c_α: 3.6266

Analyticalc_α from chi-squared distribution: 3.8415

Difference between numerical and analytical c_α: 0.2148, which is about 5.6% of the analytical c_α, so our approximation is not very close to the true value c_α.

I expect the estimated approximation to get closer to the real analytical value of c_α as M is larger.

Since the $T(x)$ we get is 18.735 which is greatly larger than 3.84 and 3.62, which is our numerical result, the conclusion from previous exercise doesn't change, we still reject $\mathcal{H}_0$.

Solution.

Based on our previous LR test, we have:

$$\begin{aligned}
\varphi_{LR}(x) &= \mathbf{1}\left\{T(x) = \frac{n}{\sigma^2}(\hat{\theta} - \theta_0)^2 < c_\alpha\right\} \\
&= \mathbf{1}\left\{(\hat{\theta} - \theta_0)^2 < \frac{c_\alpha \sigma^2}{n}\right\} \\
&= \mathbf{1}\left\{-\sqrt{\frac{c_\alpha \sigma^2}{n}} < (\hat{\theta} - \theta_0) < \sqrt{\frac{c_\alpha \sigma^2}{n}}\right\} \\
&= \mathbf{1}\left\{\hat{\theta} - \sqrt{\frac{c_\alpha \sigma^2}{n}} < \theta_0 < \hat{\theta} + \sqrt{\frac{c_\alpha \sigma^2}{n}}\right\}
\end{aligned}$$

Thus, we can define $C(X)$ as:

$$C(X) = \left[\hat{\theta} - \sqrt{\frac{c_\alpha \sigma^2}{n}}, \hat{\theta} + \sqrt{\frac{c_\alpha \sigma^2}{n}}\right]$$

Apply our previous data: $\sigma^2 = 6$, $n = 4$, $x_1 = 178$, $x_2 = 161$,$x_3 = 168$, $x_4 = 172$, $\theta_0 = 175$, $\hat{\theta} = 169.75$, and $c_\alpha = 3.8415$, we have:

$$C(X) = [169.75 - 2.4, 169.75 + 2.4] = [167.35, 172.15]$$

$\theta_0 = 175$ is not in this interval.

Because we rejected $\mathcal{H}_0 : \theta = 175$, it's consistent that 175 is not within the 95% confidence interval.

Lecture 5.

Least Squares Estimation of the Linear Regression Model

5.1 Finite Sample Properties

$$y_i = x_i'\beta + u$$
$$Y = X\beta + U$$

where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}, \quad X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}_{n\times k}, \quad U = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}_{n\times 1}$$

Assumption 5.1.1. (Independent Sampling). Observations $z_i = \{y_t, x_i\}_{i=1}^n$ are independent across i.

Assumption 5.1.2. (Full rank). The matrix $X'X = \sum x_i x_i'$ is of full rank.

Assumption 5.1.3. (Conditional Independence). $\mathbb{E}[u_i|x_i] = 0$.
$\mathbb{E}[y_i] = \mathbb{E}[x_i'\beta + u_i|x_i] = \mathbb{E}[x_i'\beta|x_i] + \mathbb{E}[u_i|x_i] = x_i'\beta$

Assumption 5.1.4. (Homoskedasticity). $\mathbb{V}[u_i|x_i] = \sigma^2$ for all i.
$\mathbb{V}[y_i] = \mathbb{V}[x_i'\beta + u_i|x_i] = \sigma^2$

The OLS estimator:

$$\hat{\beta}_{\text{OLS}} = \arg\min_{\beta\in\mathbb{R}^k} \sum u_i^2 = \arg\min_{\beta\in\mathbb{R}^k} \sum_{i=1}^n (y_i - x_i'\beta)^2 = \arg\min_{\beta\in\mathbb{R}^k} (Y - X\beta)'(Y - X\beta)$$

$$\begin{aligned} \frac{\partial (Y - X\beta)'(Y - X\beta)}{\partial \beta} &= X'(Y - X\beta) = 0 \\ &\Rightarrow X'Y - X'X\beta = 0 \\ &\Rightarrow (X'X)^{-1}X'Y = \hat{\beta} \end{aligned}$$

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = P_X Y$$

where $P_X = X(X'X)^{-1}X'$ is the projection matrix.

$$y = x\hat{\beta} + \hat{u} \to y = \hat{y} + \hat{u}$$

thus,

$$\hat{U} = Y - \hat{Y} = Y - P_X Y = (I - P_X)Y = M_X Y$$

where M_X is another projection matrix.

$$Y = P_X Y + M_X Y = (P_X + M_X)Y.$$

In another sense:

$$\hat{U} = Y - \hat{Y} = X\beta + U - X(X'X)^{-1}X'(X\beta + U) = (I_n - X(X'X)^{-1}X')U = (I_n - P_X)U = M_X U.$$

P_X and M_X are idempotent: $P_X = P_X'$ and $P_X P_X = P_X$, and are orthogonal to each other: $P_X M_X = M_X P_X = 0$.

The total sum of squares (SST) is given by:

$$\sum_{i=1}^{n} y_i^2 = Y'Y$$

It measures the variability in y_i across observations i.

We can decompose it into the explained sum of squares (SSE) and the residual sum of squares (SSR)

$$\begin{aligned}
\text{SST} = Y'Y &= (P_X Y + M_X Y)'(P_X Y + M_X Y) \\
&= Y'P_X P_X Y + Y'P_X M_X Y + Y'M_X P_X Y + Y'M_X M_X Y \\
&= Y'P_X P_X Y + Y'M_X M_X Y \\
&= \hat{Y}'\hat{Y} + \hat{U}'\hat{U} \\
&= \text{SSE} + \text{SSR}
\end{aligned}$$

Based on that, we get the R^2-statistic as a measure of how well X accounts for the variation in Y in the linear regression model:

$$R^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y} = 1 - \frac{\hat{U}'\hat{U}}{Y'Y} \in [0,1] \left(\frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}} \right)$$

Look at **Assumption 5.1.3** again, it always gives product have intercept $x_{i1} = 1$.

$$\begin{aligned}
\mathbb{E}[\hat{\beta}|X] &= \mathbb{E}[(X'X)^{-1}X'Y|X] \\
&= \mathbb{E}[(X'X)^{-1}X'(X\beta + U)|X] \\
&= \mathbb{E}[\beta|X] + \mathbb{E}[(X'X)^{-1}X'U|X] \\
&= \beta + (X'X)^{-1}X'\mathbb{E}[U|X] \\
&= \beta \\
\Rightarrow \mathbb{E}[\hat{\beta}] &= \mathbb{E}[\mathbb{E}[\hat{\beta}|X]] = \beta
\end{aligned}$$

For **Assumption 5.1.4**, the conditional variance of $\hat{\beta}_{OLS}$ is given by:

$$\begin{aligned}
\mathbb{V}[\hat{\beta}|X] &= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'U)((X'X)^{-1}X'U)'|X] \\
&= \mathbb{E}[(X'X)^{-1}X'UU'X(X'X)^{-1}|X] \\
&= (X'X)^{-1}X'\mathbb{E}[UU'|X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

> **Note.**
>
> $$UU' = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

$$\mathbb{E}[UU'|X] = \begin{bmatrix} \mathbb{E}[u_1^2|X] & \mathbb{E}[u_1u_2|X] & \cdots & \mathbb{E}[u_1u_n|X] \\ \mathbb{E}[u_2u_1|X] & \mathbb{E}[u_2^2|X] & \cdots & \mathbb{E}[u_2u_n|X] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[u_nu_1|X] & \mathbb{E}[u_nu_2|X] & \cdots & \mathbb{E}[u_n^2|X] \end{bmatrix}$$

By **Assumption 5.1.3** and **Assumption 5.1.4**, we have: $\mathbb{E}[u_i|X] = 0$ and $\mathbb{E}[u_i^2|X] = \sigma^2$. Furthermore, $\mathbb{E}[u_iu_j|X] = 0$ for $i \neq j$.

$$\mathbb{E}[UU'|X] = \sigma^2 I_n$$

By LIE again, the unconditional variance of $\hat{\beta}_{OLS}$ is given by:

$$\begin{aligned} \mathbb{V}[\hat{\beta}] &= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= \mathbb{E}[\mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X]] \\ &= \mathbb{E}[\mathbb{V}[\hat{\beta}|X]] \\ &= \mathbb{E}[\sigma^2(X'X)^{-1}] \\ &= \sigma^2\mathbb{E}[(X'X)^{-1}] \end{aligned}$$

Theorem 5.1.1 (Gauss-Markov Theorem)**.**

If **Assumption 5.1.1** to **Assumption 5.1.4** hold, then the OLS estimator $\hat{\beta}_{OLS}$ is the best linear unbiased estimator (**BLUE**) of β.

Note.

- **Best** means that the OLS estimator has the smallest variance among all linear unbiased estimators. $\mathbb{V}[\hat{\beta}_{OLS}] \leq \mathbb{V}[\hat{\beta}]$ for all linear unbiased estimators $\hat{\beta}$.
- **Linear** means that the estimator is a linear function of the dependent variable. $\hat{\beta} = c + dY$.
- **Unbiased** means that the expected value of the estimator is equal to the true value of the parameter. $\mathbb{E}[\hat{\beta}] = \beta$.

If, we know $\hat{\beta}$, $\mathbb{E}[\hat{\beta}]$ and $\mathbb{V}[\hat{\beta}]$. To find the unconditional expectation of $\hat{\beta}_{OLS}$, we could only use asymptotic properties of the OLS estimator. We can show that the OLS estimator is consistent and asymptotically normal.

$$\begin{aligned} \hat{\beta}_{OLS} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + U) \\ &= \beta + (X'X)^{-1}X'U \\ &= \beta + \left(\sum x_ix_i'\right)^{-1}\left(\sum x_iu_i\right) \\ &= \beta + \left(\frac{1}{n}\sum x_ix_i'\right)^{-1}\frac{1}{n}\left(\sum x_iu_i\right) \\ &\xrightarrow{p} \beta + \mathbb{E}[x_ix_i']^{-1}\mathbb{E}[x_iu_i] \end{aligned}$$

Note. By WLLN, we know that $\frac{1}{n}\sum z_i \xrightarrow{p} \mathbb{E}[z_i]$. So, $\frac{1}{n}\sum x_iu_i \xrightarrow{p} \mathbb{E}[x_iu_i] = Q$, and $\left[\frac{1}{n}\sum x_ix_i'\right]^{-1} \xrightarrow{p} \{\mathbb{E}[x_ix_i']\}^{-1} \xrightarrow{p} \mathbb{E}[x_ix_i']^{-1} = Q^{-1}$.

So, we have:

$$\hat{\beta}_{OLS} - \beta \xrightarrow{p} \mathbb{E}[x_ix_i']^{-1}\mathbb{E}[x_iu_i]$$

$$
\begin{aligned}
&= \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[\mathbb{E}[x_i u_i|x_i]] \\
&= \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i \mathbb{E}[u_i|x_i]] \\
&= \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i \cdot 0] \\
&= 0
\end{aligned}
$$

> **Note.** By the Central Limit Theorem, we know that:
>
> $$
> \begin{aligned}
> \sqrt{n}(\hat{\beta}_{OLS} - \beta) &= (X'X)^{-1}X'U \\
> &= \underbrace{\left(\frac{1}{n}\sum x_i x_i'\right)^{-1}}_{\mathbb{E}[x_i x_i']^{-1}} \underbrace{\sqrt{n}\ \frac{1}{n}\sum x_i u_i}_{\xrightarrow{d}\mathcal{N}(0,\mathbb{V}[x_i u_i])} \\
> &\xrightarrow{p} \mathbb{E}[x_i x_i']^{-1}\mathcal{N}(0,\mathbb{V}[x_i u_i])
> \end{aligned}
> $$

Thus, we have:

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_{OLS} - \beta) &\xrightarrow{d} \mathcal{N}(0, \mathbb{E}[x_i x_i']^{-1}\mathbb{V}[x_i u_i]\mathbb{E}[x_i x_i']^{-1'}) \\
\mathbb{V}[x_i u_i] &= \mathcal{N}(0, \mathbb{E}\left[(x_i u_i - \mathbb{E}[x_i u_i])(x_i u_i - \mathbb{E}[x_i u_i])'\right]) \\
&= \mathcal{N}(0, \mathbb{E}[x_i u_i u_i' x_i] - \mathbb{E}[x_i u_i]\mathbb{E}[x_i u_i]') \\
&= \mathcal{N}(0, \mathbb{E}[x_i u_i u_i' x_i']) \\
&= \mathcal{N}(0, \mathbb{E}[\mathbb{E}[u_i^2|x_i]x_i x_i']) \\
&= \mathcal{N}(0, \mathbb{E}[\sigma^2 x_i x_i']) \\
&= \sigma^2\mathbb{E}[x_i x_i'] \\
&= \sigma^2 Q \\
\Rightarrow \sqrt{n}(\hat{\beta}_{OLS} - \beta) &\xrightarrow{d} \mathcal{N}(0, \mathbb{E}[x_i x_i']^{-1}\sigma^2\mathbb{E}[x_i x_i']\mathbb{E}[x_i x_i']^{-1}) = \mathcal{N}(0, \sigma^2 Q^{-1})
\end{aligned}
$$

Then, we could say that:

- For finite n, $\sqrt{n}(\hat{\beta}_{OLS} - \beta) \overset{approx}{\sim} \mathcal{N}(0, \sigma^2 Q^{-1})$;
- $\hat{\beta} \overset{approx}{\sim} \mathcal{N}(\beta, \frac{\sigma^2}{n}Q^{-1})$;
- Replace unknown σ^2 and Q^{-1} by $\hat{\sigma}^2$ and $\hat{Q}^{-1}$ to get the t-distribution. We would have:

$$\hat{\beta} \overset{approx}{\sim} \mathcal{N}\left(\beta, \frac{\hat{\sigma}^2}{n}\hat{Q}^{-1}\right).$$

5.2 Hypothesis Testing

As $\hat{\beta} \overset{approx}{\sim} \mathcal{N}\left(\beta, \frac{\hat{\sigma}^2}{n}\hat{Q}^{-1}\right)$, we know that:

$$\hat{\beta}_j \overset{approx}{\sim} \mathcal{N}(\beta_j, \frac{\hat{\sigma}^2}{n}[\hat{Q}^{-1}]_{jj})$$

for a single parameter $\beta_j \in \beta$ where $[\hat{Q}^{-1}]_{jj}$ is the j-th diagonal element of $\hat{Q}^{-1}$.

This enables us to test a point hypothesis $\mathcal{H}_0 : \beta_j = \beta_{j,0}$ against the alternative $\mathcal{H}_1 : \beta_j \neq \beta_{j,0}$ using the t-test:

$$\varphi_t(x) = \mathbf{1}\left\{T_x < c\right\}, \text{with } T_t = \left|\frac{\hat{\beta}_{j,0} - \beta_j}{\hat{\sigma}_{\beta_{j,0}}}\right|,$$

where $\hat{\sigma}_{\beta_{j,0}} = \sqrt{\frac{\hat{\sigma}^2}{n}\hat{Q}^{-1}_{jj}}$.

Because the distribution of $\hat{\beta}_j$ is not exact, but only asymptotically valid, so too does the resulting test-statistic only asymptotically converge to a standard Normal distribution:

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{\frac{\hat{\sigma}^2}{n}[\hat{Q}^{-1}]_{jj}}} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\hat{\beta}_j \overset{approx}{\sim} \mathcal{N}\left(\beta_j, \underbrace{\frac{\sigma^2}{n}\left(\frac{1}{n}\sum x_i x_i'\right)^{-1}_{jj}}_{V_j}\right)$$

$$t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{V_j}} \overset{approx}{\sim} \mathcal{N}(0,1)$$

Definition 5.2.1 (Wald Test).

The asymptotic distribution of the Wald-test-statistic, T_W, follows from asymptotic Normality of $\hat{\beta}$, the Delta Method and the fact that $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_{dim(X)}$ for $X \sim N(\mu, \Sigma)$.

Using $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0,V)$ and the Delta method, we get

$$\sqrt{n}\left(g(\hat{\beta}) - g(\beta_0)\right) \xrightarrow{d} G \cdot N(0,V) = N(0, GVG'), \quad \text{with} \quad G = \left.\frac{\partial g(\beta)}{\partial \beta}\right|_{\beta=\beta_0}.$$

Therefore,

$$\sqrt{n}\left(g(\hat{\beta}) - g(\beta_0)\right)' [GVG']^{-1} \sqrt{n}\left(g(\hat{\beta}) - g(\beta_0)\right) \xrightarrow{d} \chi^2_m.$$

Under $\mathcal{H}_0 : g(\beta_0) = 0$. Also, because we do not know β_0, we replace G with $G(\hat{\beta})$, as $\hat{\beta}$ is our estimator of β_0.

More general hypotheses $\mathcal{H}_0 : g(\beta) = 0$ vs. $\mathcal{H}_1 : g(\beta) \neq 0$ for some function $g : \mathbb{R}^k \to \mathbb{R}^m$ (i.e. $m \leq k$ restrictions) can be tested using the Wald test. It uses the following statistic:

$$T_W = ng\left(\hat{\beta}_{OLS}\right)'\left[G\left(\hat{\beta}_{OLS}\right)\hat{V}G\left(\hat{\beta}_{OLS}\right)'\right]^{-1} g\left(\hat{\beta}_{OLS}\right) \xrightarrow{d} \chi^2_m,$$

where $\hat{V} = \hat{\sigma}^2\hat{Q}^{-1}$ and where $G\left(\hat{\beta}_{OLS}\right) = \partial g(\beta)/\left.\partial\beta\right|_{\beta=\hat{\beta}_{OLS}}$ is the $m \times k$ matrix of derivatives of g with respect to β evaluated at $\hat{\beta}_{OLS}$. The short derivation in the Appendix illustrates that the Wald test-statistic is based on the idea that if $\mathcal{H}_0$ is true, then the difference between $g\left(\hat{\beta}_{OLS}\right)$ and $g(\beta) = 0$ should be small. Suppose we are interested in testing $\mathcal{H}_0 : \{\beta_2 + \beta_3 = 5, \beta_4 = 0\}$ under a five-dimensional vector β. Then we would take

$$g(\beta) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}\beta - \begin{bmatrix} 5 \\ 0 \end{bmatrix}, \quad \text{with} \quad G(\beta) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

If $g(\beta) = 0$ is s.t. it tests only $\beta_j = \beta_{j,0}$ for a single β_j , then the Wald test is equivalent to the t-test: $\varphi_W = \varphi_t$.

Theorem 5.2.1 (Delta Method).

$X \xrightarrow{d} \mathcal{N}(\mu, \sigma^2)$, and $g : \mathbb{R}^k \to \mathbb{R}^q$ is a differentiable function. Then, $g(X) \xrightarrow{d} \mathcal{N}(g(\mu), \sigma^2(g'(\mu))^2)$.

Let $\beta \in \mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}^q$ be a differentiable function. If $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \xi$, then

$$\sqrt{n}(g(\hat{\beta}) - g(\beta)) \xrightarrow{d} G'\xi,$$

where $G = G(\beta) = \frac{\partial}{\partial \beta} g(\beta)'$.
In particular, if $\xi \sim \mathcal{N}(0, V)$, then

$$\sqrt{n}(g(\hat{\beta}) - g(\beta)) \xrightarrow{d} \mathcal{N}(0, G'VG).$$

By previous results, if we have $V = \sigma^2 Q^{-1}$, then

$$\sqrt{n}(g(\hat{\beta}) - g(\beta)) \xrightarrow{d} \mathcal{N}(0, G'\sigma^2 Q^{-1} G)$$

where $G(u) = \frac{\partial}{\partial u} g(u)'$ and $G = G(\beta)$.

5.3 Violations of Ideal Conditions

First of all, note that while unbiasedness requires the conditional independence assumption 3 to hold, both consistency and asymptotic Normality go through even under the weaker exogeneity assumption $\mathbb{E}[u_i x_i] = 0$.[1]

5.3.1 Singular $X'X$

If $X'X$ is not of full rank, then the OLS estimator is not even defined. There are two reasons that lead to this case.

5.3.2 Heteroskedasticity

Suppose we replace the Assumption 5.1.4 with the weaker assumption that $\mathbb{V}[u_i|x_i] = \sigma_i^2$ for all i. Then, the OLS estimator is still unbiased, but the variance of the OLS estimator is now given by:

$$\begin{aligned}
\mathbb{V}[x_i u_i] &= \mathbb{E}[(x_i u_i - \mathbb{E}[x_i u_i])\,(x_i u_i - \mathbb{E}[x_i u_i])'] \\
&= \mathbb{E}[x_i u_i u_i' x_i'] - \mathbb{E}[x_i u_i]\mathbb{E}[x_i u_i]' \\
&= \mathbb{E}[\mathbb{E}[u_i^2|x_i] x_i x_i'] \\
&= \mathbb{E}[x_i x_i' u_i^2] \\
&= \mathbb{E}[x_i x_i' \sigma_i^2] \\
\Rightarrow \sqrt{n}(\hat{\beta}_{OLS} - \beta) &\xrightarrow{d} \mathcal{N}(0, Q^{-1}\mathbb{E}[x_i x_i' u_i^2] Q^{-1})
\end{aligned}$$

The asymptotic variance can again be estimated by replacing $\mathbb{E}[x_i x_i' u_i^2]$ with its sample analogue: as a consistent estimator: $\frac{1}{n}\sum_{i=1}^n x_i x_i' \hat{u}_i^2$.
Note that if the variances $\{\sigma_i^2\}_{i=1}^n$ were known, we could transform the heteroskedastic model into a homoskedastic one by writing the regression as:

$$\frac{y_i}{\sigma_i} = \frac{x_i'}{\sigma_i}\beta + \frac{u_i}{\sigma_i}.$$

In this model, observations are weighted by the inverses of their standard deviations and, as a result, less noisy observations are given more weight as they are more informative about the relation between X and Y. Let $\mathbb{V}[U|X] = \Sigma = diag(\sigma_1^2, \cdots, \sigma_n^2)$. We can then write the regression in matrix form as: $\Sigma^{-\frac{1}{2}} Y = \Sigma^{-\frac{1}{2}} X\beta + \Sigma^{-\frac{1}{2}} U$, with $\mathbb{V}[\Sigma^{-\frac{1}{2}} U|X] = I$.

[1] This is because it's implied by the conditional independence assumption.

The OLS estimator in this weighted regression model is referred to as the Generalized Least Squares (GLS) estimator. It is given by:

$$\hat{\beta}_{GLS} = \left(\left(\Sigma^{-\frac{1}{2}}X\right)'\left(\Sigma^{-\frac{1}{2}}X\right)\right)\left(\Sigma^{-\frac{1}{2}}X\right)'\left(\Sigma^{-\frac{1}{2}}Y\right) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y.$$

Under otherwise the same conditions as for OLS, this estimator is unbiased[2] and consistent[3]. and has variance:

$$\mathbb{V}[\hat{\beta}_{GLS}] = \mathbb{E}\left[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}UU'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\right] = \mathbb{E}\left[(X'\Sigma^{-1}X)^{-1}\right].$$

5.3.3 Endogeneity

Omitted Variables

Suppose the true model is given by:

$$y_i = x_i'\beta + z_i'\delta + \varepsilon_i \text{ , with } \mathbb{E}[x_i\varepsilon_i] = 0,$$

i.e. exogeneity holds in this true model, whereas the researcher estimates

$$y_i = x_i'\gamma + u_i.$$

Notice that we have written the coefficient as γ rather than β and the error as u rather than ε. Goldberger (1991) introduced the catchy labels long regression and short regression to emphasize the distinction. Typically, $\beta \neq \gamma$,except in special cases. To see this, we calculate

$$\begin{aligned}\gamma &= \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[XY\right] \\ &= \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[X\left(X'\beta + Z'\delta + \varepsilon\right)\right] \\ &= \beta + \left(\mathbb{E}\left[XX'\right]\right)^{-1}\mathbb{E}\left[XZ'\right]\delta \\ &= \beta + \Gamma\delta\end{aligned}$$

where $\Gamma = Q^{-1}Q_{XZ}$ is the coefficient matrix from a projection of Z on X.

Observe that $\gamma = \beta + \Gamma\delta \neq \beta$ unless $\Gamma = 0$ or $\delta = 0$. Thus the short and long regressions have different coefficients. They are the same only under one of two conditions. First, if the projection of Z on X yields a set of zero coefficients (they are uncorrelated), or second, if the coefficient on Z in is zero. The difference $\Gamma\delta$ between γ and β is known as omitted variable bias. It is the consequence of omission of a relevant correlated variable.

[2] $\mathbb{E}\left[\hat{\beta}_{GLS}|X\right] = \mathbb{E}\left[\left(X'\Sigma^{-1}X\right)^{-1}(X\beta + U)|X\right] = \beta + \mathbb{E}\left[\left(X'\Sigma^{-1}X\right)^{-1}U|X\right] = \beta.$

[3] $\hat{\beta}_{GLS} - \beta = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}U \xrightarrow{p} \frac{1}{n}\mathbb{E}\left[\frac{x_i x_i'}{\sigma_i^2}\right]\frac{1}{n}\mathbb{E}\left[\frac{x_i u_i}{\sigma_i^2}\right] \xrightarrow{p} 0.$

Lecture 6.

Likelihood-Based Inference

In previous lectures we have discussed the least squares estimation of the linear regression model. In this lecture we will discuss the likelihood-based inference.

	ch2	ch3(LRM)
LS	$\mathbb{E}[y_i\|\theta]$	$\mathbb{E}[y_i\|x_i,\beta]$
ML	$p(y_i\|\theta)$	$p(y_i\|x_i,\beta)$

6.1 ML for LRM

$$y_i = x_i'\beta + u_i$$

For LS, we assume $\mathbb{E}[u_i|x_i]=0$, which gives $\mathbb{E}[y_i|x_i]=x_i'\beta$. For ML, we assume $u_i \sim N(0,\sigma^2)$, which gives $y_i|x_i \sim N(x_i'\beta,\sigma^2)$.

$$p(y_i|x_i) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right\}$$

We use θ to represent parameters (β,σ^2).

$$\begin{aligned}\Rightarrow \mathcal{L}(\theta|y_i,x) &= p(y|x_i,\theta)\\ &= \prod_{i=1}^n p(y_i|x_i,\theta)\\ &= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right\}\\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i - x_i'\beta)^2\right\}\end{aligned}$$

Then, we can get the log-likelihood function:

$$\begin{aligned}\ell(\theta|y_i,x) &= \log \mathcal{L}(\theta|y_i,x)\\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\\ \Rightarrow \hat{\theta}_{ML} = (\hat{\beta}_{ML},\hat{\sigma}^2_{ML}) &= \arg\max_{\theta=(\beta,\sigma^2)} \ell(\theta|y_i,x)\end{aligned}$$

Take the first order condition (FOC) of β and σ^2:

$$\begin{aligned}\beta :& \frac{1}{\sigma^2}X'(Y-X\beta) = 0\\ \Rightarrow \hat{\beta}_{ML} &= (X'X)^{-1}X'Y\\ \sigma^2 :& -\frac{n}{2\sigma^2} + \frac{2}{(2\sigma^2)^2}(Y-X\beta)'(Y-X\beta) = 0\\ \Rightarrow \hat{\sigma}^2_{ML} &= \frac{1}{n}(Y-X\beta)'(Y-X\beta) = \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 = \hat{\sigma}^2_{LS}\end{aligned}$$

As we know that:

$$\hat{\beta} = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'U = \beta + \left(\frac{1}{n}\sum x_i x_i'\right)^{-1}\sum x_i u_i$$

$$\Rightarrow \hat{\beta}|X \sim \mathcal{N}(\beta, V)$$

where

$$\begin{aligned} V &= \mathbb{V}\left[(X'X)^{-1}X'U|X\right] \\ &= (X'X)^{-1}\mathbb{V}[X'U|X](X'X)^{-1} \\ &= (X'X)^{-1}X'\mathbb{V}[U|X]X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

We define the **Score Function** as:

Definition 6.1.1 (Score Function).

$$\begin{aligned} S(\theta) &= \frac{\partial \ell(\theta|y_i, x)}{\partial \theta} \\ &= \begin{bmatrix} \frac{\partial \ell(\theta|y_i,x)}{\partial \beta} \\ \frac{\partial \ell(\theta|y_i,x)}{\partial \sigma^2} \end{bmatrix} \end{aligned}$$

As we know that:

$$\begin{aligned} \frac{\partial \ell(\theta|y_i, x)}{\partial \beta} &= \frac{1}{\sigma^2}X'(Y - X\beta) \\ \frac{\partial \ell(\theta|y_i, x)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(Y - X\beta)'(Y - X\beta) \end{aligned}$$

We take the Hessians of β and σ^2:

$$\begin{aligned} \mathcal{H}(\theta) &= \frac{\partial^2 \ell(\theta|y_i, x)}{\partial \theta \partial \theta'} = \frac{\partial S(\theta)}{\partial \theta'} \\ &= \begin{bmatrix} \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \beta \partial \beta} & \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix} \\ \frac{\partial^2 \ell(\theta|y_i, x)}{\partial \beta \partial \beta} &= -\frac{1}{\sigma^2}X'X \\ \frac{\partial^2 \ell(\theta|y_i, x)}{\partial \sigma^2 \partial \sigma^2} &= \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(Y - X\beta)'(Y - X\beta) \\ \frac{\partial^2 \ell(\theta|y_i, x)}{\partial \beta \partial \sigma^2} &= \frac{1}{(\sigma^2)^2}X'(Y - X\beta) \end{aligned}$$

Then, we can get the **Information Matrix** as:

Definition 6.1.2 (Information Matrix).

$$\begin{aligned} I(\theta) &= \mathbb{E}[s(\theta)s(\theta)'] = -\mathbb{E}\left[\frac{\partial^2 \ell(\theta|y_i, x)}{\partial \theta \partial \theta'}\right] \\ &= -\mathbb{E}\left[\begin{bmatrix} \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \beta \partial \beta} & \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 \ell(\theta|y_i,x)}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix}\right] \end{aligned}$$

$$
\begin{aligned}
\mathbb{E}[s(\beta)s(\beta)'] &= -\mathbb{E}\left[\frac{1}{\sigma^2}X'(Y-X\beta)\left[\frac{1}{\sigma^2}X'(Y-X\beta)\right]'\right] \\
&= \mathbb{E}\left[\frac{1}{\sigma^4}X'UU'X\right] \\
&= \frac{1}{\sigma^4}X'\mathbb{E}[UU']X \\
&= \frac{1}{\sigma^4}X'\sigma^2 IX \\
&= \frac{1}{\sigma^2}X'X \\
&= \mathbb{E}[-\mathcal{H}(\beta)]
\end{aligned}
$$

Then, we could have the **Cramer-Rao Lower Bound**:

Definition 6.1.3 (Cramer-Rao Lower Bound).

Let $\tilde{\theta}$ be an unbiased estimator of θ, then:

$$
\begin{aligned}
\mathbb{V}[\tilde{\theta}|X] &\geq I^{-1}(\theta) \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

Take a model $\ell(\theta|y)$, and

$$
\begin{aligned}
\hat{\theta} &= \underset{\theta}{\operatorname{argmax}}\,\ell(\theta|y) \\
\bar{\theta} &= \underset{\theta, g(\theta)=0}{\operatorname{argmax}}\,\ell(\theta|y)
\end{aligned}
$$

and $\sqrt{n}(\hat{\theta}_0 - \theta) \xrightarrow{d} \mathcal{N}(0, V)$. We want to test: $\mathcal{H}_0 : g(\theta) = 0$. Previously, we have three tests:

- t-test: $t = \frac{\hat{\theta}-\theta_0}{\sqrt{\frac{1}{n}\hat{V}}} \xrightarrow{d} \mathcal{N}(0,1)$
- Wald Test: $W = ng(\hat{\theta})'[g(\hat{\theta})\hat{V}g(\hat{\theta})']^{-1}g(\hat{\theta}) \sim \chi^2_k$
- LR Test: $LR = -2(\ell(\hat{\theta}_0) - \ell(\bar{\theta})) \sim \chi^2_k$
- LM Test: $LM = S(\bar{\theta})'\hat{I}(\bar{\theta})^{-1}S(\bar{\theta}) \sim \chi^2_k$

Lecture 7.

Likelihood-Based Inference(2)

7.1 Binary Choice: Logit Model & Probit Model

Suppose $y_i \in \{0,1\}$, the common approach is still:

$$\begin{aligned} y_i &= x_i'\beta + u_i \\ \mathbb{E}[u_i|x_i] &= 0 \end{aligned}$$

But, the linear regression is not attractive, because $\mathbb{E}[y_i|x_i] = \mathbb{P}[y_i = 1|x_i]$ is bounded between 0 and 1, while the linear regression is unbounded.

We define a new regression model:

$$\begin{aligned} y_i^* &= x_i'\beta + u_i \\ u_i|x_i &\sim \mathcal{N}(0,1) \end{aligned}$$

and assume we have: $y_i = \mathbf{1}\{y_i^* \geq 0\}$. If utility is positive $y_i = 1$, if it is negative, $y_i = 0$.

Then, we have the probability of $y_i = 1$ as:

$$\mathbb{P}[y_i = 1] = \mathbb{P}[x_i'\beta + u_i \geq 0] = \mathbb{P}[u_i \geq -x_i'\beta] = 1 - \Phi(-x_i'\beta) = \Phi(x_i'\beta)$$

and $\mathbb{P}[y_i = 0] = 1 - \Phi(x_i'\beta) = \Phi(-x_i'\beta)$, where Φ is the CDF of a standard normal RV.

Hence, y_i had the PDF:

$$p(y_i|x_i,\beta) = \begin{cases} \Phi(x_i'\beta) & y_i = 1 \\ \Phi(-x_i'\beta) & y_i = 0 \end{cases} = \Phi(x_i'\beta)^{y_i}\Phi(-x_i'\beta)^{1-y_i}$$

which is the Bernoulli distribution with probability of success $\Phi(x_i'\beta)$.

This leads to our likelihood function:

$$\begin{aligned} \mathcal{L}(\beta|Y,X) &= \prod_{i=1}^{n} p(y_i|x_i,\beta) \\ &= \prod_{i=1}^{n} \Phi(x_i'\beta)^{y_i}\Phi(-x_i'\beta)^{1-y_i} \end{aligned}$$

Then, we can get the log-likelihood function:

$$\begin{aligned} \ell(\beta|Y,X) &= \log \mathcal{L}(\beta|Y,X) \\ &= \sum_{i=1}^{n} \{y_i \log \Phi(x_i'\beta) + (1-y_i)\log \Phi(-x_i'\beta)\} \end{aligned}$$

where we have the estimator defines as: $\hat{\beta} = \arg\max_\beta \ell(\beta|Y,X)$.

Then, we can have:

$$\hat{y}_i = \mathbb{E}[y_u|x_i'\beta] = \Phi(x_i'\hat{\beta}) \Rightarrow R^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y}.$$

The partial effect of X on Y is:

$$\delta = \mathbb{E}[y_i|x_i = x_2, \beta] - \mathbb{E}[y_i|x_i = x_1, \beta]$$

$$= \Phi(x_2'\beta) - \Phi(x_1'\beta)$$

approximately,

$$\frac{\partial \mathbb{E}[y_i|x_i\beta]}{\partial x_i} = \frac{\partial \Phi(x_i'\beta)}{\partial x_i} = \phi(x_i'\beta)\beta.$$

which is:

$$\Delta \mathbb{E}[y_i|x_i\beta] = \phi(x_i'\beta)\beta\Delta x_i.$$

Since $\Phi(\cdot)$ is a strictly increasing function, the sign of β_j reveals the sign of the partial effect of x_j, but the size of β_j is not interpretable. Only the relative sizes of two coefficients β_k and β_l have (qualitative) meaning. Nevertheless, we can test for the partial effect of x_j being zero by testing $\mathcal{H}_0 : \beta_j = 0$, because the former is zero iff $\beta_j = 0$.

7.2 Censored Outcomes: Tobit Model

Suppose y_i is censored at 0, i.e., $y_i \geq 0$. We can deal with this by assuming:

$$\begin{aligned} y_i^* &= x_i'\beta + u_i \\ u_i|x_i &\sim \mathcal{N}(0, \sigma^2) \\ y_i &= y_i^*\mathbf{1}\{y_i^* \geq 0\} \end{aligned}$$

In this case, the probability of observing $y_i = 0$ is:

$$\mathbb{P}[y_i = 0] = \mathbb{P}[y_i^* < 0] = \mathbb{P}[x_i'\beta + u_i < 0] = \Phi\left(-\frac{x_i'\beta}{\sigma}\right).$$

To get the PDF $p(y_i)$ for $y_i > 0$, we derive the CDF:

$$\mathbb{P}[y_i < y] = \mathbb{P}[y_i^* < y] = \mathbb{P}\left[\frac{u_i}{\sigma} < \frac{y_i - x_i'\beta}{\sigma}\right] = \Phi\left(\frac{y_i - x_i'\beta}{\sigma}\right),$$

which gives that

$$p(y) = \frac{\partial \mathbb{P}[y_i < y]}{\partial y} = \frac{1}{\sigma}\phi\left(\frac{y - x_i'\beta}{\sigma}\right)$$

for $y > 0$.

Hence, the PDF observations y_i is:

$$p(y_i|x_i, \beta, \sigma) = \begin{cases} \Phi\left(-\frac{x_i'\beta}{\sigma}\right) & y_i = 0 \\ \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) & y_i > 0 \end{cases}$$

Then, we can get the likelihood function:

$$\begin{aligned} \mathcal{L}(\beta, \sigma|Y, X) &= \prod_{i=1}^{n} p(y_i|x_i, \beta, \sigma) \\ &= \prod_{i=1}^{n} \left\{\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right\}^{1\{y_i=0\}} \left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}^{1\{y_i>0\}} \end{aligned}$$

Then, we can get the log-likelihood function:

$$\begin{aligned} \ell(\beta, \sigma|Y, X) &= \log \mathcal{L}(\beta, \sigma|Y, X) \\ &= \sum_{i=1}^{n} \left\{\mathbf{1}\{y_i = 0\} \log \Phi\left(-\frac{x_i'\beta}{\sigma}\right) + \mathbf{1}\{y_i > 0\} \log\left(\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right)\right\} \end{aligned}$$

Let $\mathcal{G} = \{i, y_i = 0\}$, we can wirte the log-likelihood function as:

$$\ell(\beta, \sigma | Y, X) = \sum_{i \in \mathcal{G}} \left\{ \log \Phi \left(-\frac{x_i' \beta}{\sigma} \right) \right\} + \sum_{i \notin \mathcal{G}} \left\{ \log \left(\frac{1}{\sigma} \phi \left(\frac{y_i - x_i' \beta}{\sigma} \right) \right) \right\}$$

Our estimator is defined as:

$$\hat{\theta}_{ML} = (\hat{\beta}_{ML}, \hat{\sigma}_{ML}) = \arg \max_{\theta = (\beta, \sigma)} \ell(\theta | Y, X).$$

To compute partial effects, we need to derive the (conditional) expectation $\mathbb{E}[y_i | x_i]$. Using the result that for $Z \sim N(0,1)$, $\mathbb{E}[Z | Z > -c] = \phi(c)/\Phi(c)$ (inverse Mills ratio), we get

$$\mathbb{E}[y_i | y_i > 0] = \mathbb{E}[x_i' \beta + u_i | u_i > -x_i' \beta] = x_i' \beta + \sigma \phi \left(\frac{x_i' \beta}{\sigma} \right) / \Phi \left(\frac{x_i' \beta}{\sigma} \right).$$

> **Note.** The inverse Mills ratio is the ratio of the probability density function to the complementary cumulative distribution function of a distribution. Its use is often motivated by the following property of the truncated normal distribution. If X is a random variable having a normal distribution with mean μ and variance σ^2, then
>
> $$\mathbb{E}[\, X \mid X > \alpha \,] = \mu + \sigma \frac{\phi\left(\frac{\alpha - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\alpha - \mu}{\sigma}\right)},$$
> $$\mathbb{E}[\, X \mid X < \alpha \,] = \mu - \sigma \frac{\phi\left(\frac{\alpha - \mu}{\sigma}\right)}{\Phi\left(\frac{\alpha - \mu}{\sigma}\right)}$$
>
> where α is a constant, ϕ denotes the standard normal density function, and Φ is the standard normal cumulative distribution function.
>
> The two fractions are the inverse Mills ratios.

Then, we obtain

$$\begin{aligned}
\mathbb{E}[y_i] &= \mathbb{E}[y_i | y_i \geq 0] \mathbb{P}[y_i \geq 0] \\
&= \mathbb{E}[y_i | y_i = 0] \mathbb{P}[y_i = 0] + \mathbb{E}[y_i | y_i > 0] \mathbb{P}[y_i > 0] \\
&= \mathbb{E}[y_i | y_i > 0] (1 - \mathbb{P}[y_i = 0]) \\
&= \mathbb{E}[y_i^* | y_i^* > 0] \left(1 - \Phi \left(\frac{-x_i' \beta}{\sigma} \right) \right) \\
&= \mathbb{E}[x_i' \beta + u_i | u_i + x_i' \beta > 0] \left(1 - \Phi \left(\frac{-x_i' \beta}{\sigma} \right) \right) \\
&= \mathbb{E}[x_i' \beta + u_i | u_i > -x_i' \beta] \left(1 - \Phi \left(\frac{-x_i' \beta}{\sigma} \right) \right) \\
&= x_i' \beta + \sigma \frac{\phi(x_i' \beta / \sigma)}{\Phi(x_i' \beta / \sigma)} \left(1 - \Phi \left(\frac{-x_i' \beta}{\sigma} \right) \right) \\
&= \Phi \left(\frac{x_i' \beta}{\sigma} \right) \left(x_i' \beta + \sigma \frac{\phi \left(\frac{x_i' \beta}{\sigma} \right)}{\Phi \left(\frac{x_i' \beta}{\sigma} \right)} \right) \\
&= \Phi \left(\frac{x_i' \beta}{\sigma} \right) x_i' \beta + \sigma \phi \left(\frac{x_i' \beta}{\sigma} \right).
\end{aligned}$$

Lecture 8.

Review Sesison

Lecture 9.

Topics in Econometrics

9.1 Numerical Estimation

For previous estomation methods, we have:

$$\hat{\theta} = \arg\min_{\theta} \mathcal{Q}_n(\theta; Y_n).$$

Take the first-order condition to higher orders, we have:

$$\begin{aligned}
\mathcal{Q}^{(1)}(\theta) = \frac{\partial \mathcal{Q}(\theta)}{\partial \theta} &= \begin{bmatrix} \frac{\partial \mathcal{Q}(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \mathcal{Q}(\theta)}{\partial \theta_k} \end{bmatrix} \\
\Rightarrow \mathcal{Q}^{(n)}(\theta^{m+1}) &= \mathcal{Q}^{(1)}(\theta^m) + \mathcal{Q}^{(2)}(\theta^m)(\theta^{m+1} - \theta^m) + \frac{1}{2}\mathcal{Q}^{(3)}(\theta^m)(\theta^{m+1} - \theta^m)^2 + \cdots \\
&\approx \mathcal{Q}^{(1)}(\theta^m) + \mathcal{Q}^{(2)}(\theta^m)(\theta^{m+1} - \theta^m) \\
&= 0. \\
\Rightarrow \theta^{m+1} &= \theta^m - \left[\mathcal{Q}^{(2)}(\theta^m)\right]^{-1} \mathcal{Q}^{(1)}(\theta^m).
\end{aligned}$$

Note (Newton-Raphson Method).

The Newton-Raphson method is a root-finding algorithm that uses the first few terms of the Taylor series of a function $f(x)$ in the vicinity of a starting point x_0 to find the root of the function. The method is based on the idea that a continuous and differentiable function can be approximated by a straight line tangent to it. The method is iterative and converges quadratically to the root.

Algorithm 3: Newton-Raphson Method

```
Input: Initialize $\theta_0$, tolerence level $\varepsilon > 0$
for $m = 1$ to $M$ do
    Given $\theta^m$, compute $\mathcal{Q}^{(2)}(\theta^m, Y^n)^{-1}$ and $\mathcal{Q}^{(1)}(\theta^m, Y^n)$;
    Set $\theta^{m+1} = \theta^m - \mathcal{Q}^{(2)}(\theta^m, Y^n)^{-1} \mathcal{Q}^{(1)}(\theta^m, Y^n)$;
    if $\|\theta^{m+1} - \theta^m\| < \varepsilon$ then
        $\hat{\theta} = \theta^{m+1}$;
    else
        Proceed to the next iteration;
    end
end
```

In some cases, we are not able to find the joint expectation of β and Σ, but may know identically.

Example 5.

$$\hat{\theta} = \arg\min_{\theta} \mathcal{Q}(\theta; Y_n)$$

cannot be obtained analytically.
As $\theta = [\theta_1, \theta_2]'$, we know $\hat{\theta}_1 \mid \theta_2$ and $\hat{\theta}_2 \mid \theta_1$

Suppose we have the following model:

$$y_i = x_i'\beta + u_i, \quad u_i|x_i \sim \mathcal{N}(0, \sigma_i^2).$$

Then, the likelihood function is:

$$\begin{aligned}
\hat{\theta}\left(\hat{\beta}, \hat{\Sigma}\right) &= \arg\max_{\theta} p\left(y \mid \beta, \Sigma\right) \\
\Rightarrow p\left(y \mid \beta, \Sigma\right) &= \prod_{i=1}^{n} p\left(y_i \mid \beta, \Sigma\right) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}\left(y_i - x_i'\beta\right)^2\right\} \\
&= (2\pi)^{-\frac{n}{2}} \prod_{i=1}^{n} \sigma_i^{-1} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left(y_i - x_i'\beta\right)^2\right\} \\
&= (2\pi)^{-\frac{n}{2}} \left(\prod_{i=1}^{n} \sigma_i^{-1}\right) \exp\left\{-\frac{1}{2}\left(Y - X\beta\right)' \Sigma^{-1}\left(Y - X\beta\right)\right\} \\
&= (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(Y - X\beta\right)' \Sigma^{-1}\left(Y - X\beta\right)\right\}. \\
\Rightarrow \hat{\beta} \mid \Sigma &= \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}Y, \\
\hat{\Sigma} \mid \beta &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'\hat{\beta}\right)^2 = diag(Y - X\beta)(Y - X\beta)'.
\end{aligned}$$

WE can obtain the estimator $\hat{\theta}$ by iterating between $\hat{\beta} \mid \hat{\Sigma}$ and $\hat{\Sigma} \mid \hat{\beta}$.

Note (Meng and Rubin (1993) Algorithm).

Algorithm 4: Meng and Rubin (1993) Algorithm

```
Input: Initialize $\Sigma_0$(i.e. $= I$), tolerence level $\varepsilon > 0$
for $m = 1$ to $M$ given $\Sigma^m$ do
    Compute $\beta^{m+1} = \hat{\beta} \mid \hat{\Sigma}^m$ ;
    Compute $\Sigma^{m+1} = \hat{\Sigma} \mid \hat{\beta}^{m+1}$;
    if $\left\|\theta^{m+1} - \theta^m\right\| < \varepsilon$ then
        $\hat{\theta} = \theta^{m+1}$;
    else
        Proceed to the next iteration;
    end
end
```

9.2 Bootstrapping

For some point estimator $\hat{\theta}$, we only have the aympototic distribution of $\hat{\theta}$, but not the finite-sample distribution.

Example 6. Suppose we have the following model:

$$y_i = x_i'\beta + u_i, \quad u_i|x_i \sim \mathcal{N}(0, \sigma^2).$$

Then, the estimator $\hat{\beta}$ is:

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y.$$

We know that:

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \left(X'X\right)^{-1}\right).$$

However, we do not know the finite-sample distribution of $\hat{\beta}$.

But, we could use the bootstrapping method to estimate the finite-sample distribution of $\hat{\beta}$.

Note (Bootstrapping Method).

Algorithm 5: Bootstrapping Method

```
Input: Sample $Y^n = \{y_1, \cdots, y_n\}$, number of bootstrap samples $B$
for $m = 1$ to $M$ do
    Generate a bootstrap sample of $n_B$ observations by sampling with replacement from $\{z_i\}_{i=1}^n$;
    Compute the bootstrap estimator $\hat{\theta}^m$ using $\{z_i^m\}_{i=1}^{n_B}$;
end
The set $\{\hat{\theta}_m\}_{m=1}^M$ approximates the finite-sample distribution of $\hat{\theta} \mid \theta$ for sample size $n_B$;
Compute the bootstrap standard error $\hat{\sigma}_{\hat{\theta}} = \sqrt{\frac{1}{B}\sum_{m=1}^B \left(\hat{\theta}^m - \bar{\hat{\theta}}\right)^2}$;
```

9.3 Extremum Estimation

9.3.1 Standard Asymptotics

We have a general form of question:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{Q}_n(\theta; Y^n).$$

Problem.

- Consistency: $\hat{\theta} \xrightarrow{p} \theta_0$?
- Distribution: $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0, V) \to \hat{\theta}_0 \overset{approx}{\to} \mathcal{N}\left(0, \frac{1}{n}\hat{V}\right)$?

Example 7 (Probit Model).

$$\hat{\beta} = \arg\min_{\beta} -\frac{1}{n}\sum_{i=1}^n \left\{y_i \log\Phi\left(x_i'\beta\right) + (1-y_i)\log\left(-\Phi\left(x_i'\beta\right)\right)\right\} = \ell(\beta \mid y).$$

Proposition 9.3.1 (Consistency).

Assumption 9.3.1.

- Θ is compact.
- $\mathcal{Q}_n(\theta, Y^n)$ converges uniformly in probability to $\mathcal{Q}(\theta)$ uniformly in θ; i.e.
$$\forall \varepsilon > 0, \mathbb{P}\left[\sup_{\theta\in\Theta}\left|\mathcal{Q}_n(\theta, Y^n) - \mathcal{Q}(\theta)\right| < \varepsilon\right] \to 1;$$
- $\mathcal{Q}(\theta)$ is continuous in Θ;
- $\mathcal{Q}(\theta)$ is uniquely minimized by θ_0, i.e. $\mathcal{Q}(\theta) > \mathcal{Q}(\theta_0) \quad \forall\theta \in \Theta, \theta \neq \theta_0$.

Then, $\hat{\theta} \xrightarrow{p} \theta_0$.

For our original model:

$$\{y_i\}_n^{i=1} \quad \text{with} \quad \mathbb{E}[y_i] = 0.$$

we have:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} (y_i - \theta)^2$$
$$\Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathbb{E}[y_i] = 0.$$

In this case, we have:

- Θ is compact, take $\Theta = [-c, c]$ for some large c.
- $\mathscr{Q}_n(\theta, Y^n) = \sum_{i=1}^{n} (y_i - \theta)^2 \xrightarrow{p} \underbrace{\mathbb{E}\left[(y_{i-\theta})^2\right]}_{Q|\theta\rangle}$ by LLN;
- $Q\,|\theta\rangle = \mathbb{E}[y_i^2] - 2\theta\mathbb{E}[y_i] + \theta^2 = \mathbb{V}[y_i] + (\mathbb{E}[y_i] - \theta)^2$ is continuous.
- $Q\,|\theta\rangle$ is uniquely minimized by $\theta_0 = \mathbb{E}[y_i]$.
- $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathbb{E}[y_i] = 0$.

Example 8 (Nonlinear Least Squares(NLS)).
Consider the nonlinear least squares (NLS) estimation of the regression model:

$$y_i = (x_i'\beta)^3 + u_i, \quad \mathbb{E}[u_i \mid x_i] = 0.$$

Define $\mathscr{B} = \left\{\beta \in \mathbb{R}^k : \|\beta\| \le c\right\}$ for some $c > 0$ large and let

$$\hat{\beta} = \arg\min_{\beta \in \mathscr{B}} Q_n(\beta, Y^n) = \arg\min_{\beta \in \mathcal{B}} \frac{1}{2n}\sum_{i=1}^{n}\left(y_i - (x_i'\beta)^3\right)^2.$$

To show $\hat{\beta} \xrightarrow{p} \beta_0$, we show uniform convergence in probability of Q_n:

- $\mathscr{B}$ is compact in $\mathbb{R}^k$.
- $\mathscr{Q}_n(\beta) = \frac{1}{n}\sum \frac{1}{2}\left(y_i - (x_i'\beta)^3\right)^2 \xrightarrow{p} \mathbb{E}\left[\frac{1}{2}\left(y_i - (x_i'\beta)^3\right)^2\right] = \mathscr{Q}(\beta)$; [a]
- We know $\mathbb{E}\left[(y_i - h(x_i))^2\right]$ is uniquely minimized at $h(x_i) = \mathbb{E}[y_i \mid x_i] = (x_i'\beta_0)^3$. Thus, $Q_n(\beta) = \frac{1}{2}\mathbb{E}\left[\left(y_i - (x_i'\beta)^3\right)^2\right]$ is uniqely minimized at $\beta = \beta_0$.
- $Q(\theta)$ is continuous

[a] $Q_n(\beta) = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} m((x_i, y_i), \beta)$ converges uniformly in probability to $Q(\beta) = \frac{1}{2}\mathbb{E}[m(x;\theta)] = \mathbb{E}\left[\left(y_i - (x_i'\beta)^3\right)^2\right]$ because $m((x_i, y_i), \beta) = \left(y_i - (x_i'\beta)^3\right)^2$ satisfies the conditions for the ULLN; the first three are obvious, and for the fourth it is sufficient to assume $\mathbb{E}\left[\|u_i\|^2\right] < \infty$ and $\mathbb{E}\left[\|x_i\|^6\right] < \infty$, along with $\|\beta\| \le c$:

$$\mathbb{E}\left[\sup_{\beta \in \mathscr{A}} \|m((x_i, y_i), \beta)\|\right] \leqslant \mathbb{E}\left[|y_i|^2\right] + \sup_{\beta \in \mathscr{A}} 2\mathbb{E}\left[|y_i|\,|x_i|^3\,\|\beta\|^3\right] + \sup_{\beta \in \mathscr{A}}\left[|x_i|^6\,\|\beta\|^6\right]$$

$$< \infty.$$

In general, there are three ways to show that θ_0 is the unique minimizer of $Q(\theta)$. First, one can write out $Q(\theta)$ to see it explicitly by looking at FOCs (and SOCs) as in the first example above. Second, one can use the conditional-expectation-argument as in the second example above. Third, one can show that $Q(\tilde{\theta}) - Q(\theta_0) > 0 \quad \forall \tilde{\theta} \neq \theta_0$.

Proposition 9.3.2 (Asymptotic Normality).

In addition to the conditions in 9.3.1, we assume:

Assumption 9.3.2.

- $\theta_0 \in int(\Theta)$;
- $\sqrt{n}\mathcal{Q}_n^{(1)}(\theta_0, Y^n) \xrightarrow{d} \mathcal{N}(0, M)$.
- $\mathcal{Q}_n(\theta, Y^n) \in \mathcal{C}^2$ w.r.t. $\theta \quad \forall Y^n$. Also, $\exists H$ s.t. $\mathcal{Q}_n^{(2)}(\theta_0, Y^n) \xrightarrow{p} H \quad \forall \theta_n \xrightarrow{p} \theta_0$. Then,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, H^{-1}MH^{-1}).$$

Example 9.

Let's look at the NLS example again. We have:

- $\beta_0 \in int(\mathscr{B})$ for large c.
- By CLT, we have:

$$\begin{aligned}
\mathcal{Q}_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - (x_i'\beta)^3\right)^2 \\
\sqrt{n}\mathcal{Q}_n^{(1)}(\beta_0, Y^n) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(y_i - (x_i'\beta_0)^3\right) 3\left(x_i'\beta_0\right)^2 x_i \\
&\xrightarrow{d} \mathbb{E}\left[9u_i^2(x_i'\beta_0)^4 x_i x_i'\right] \\
&\equiv \mathcal{N}(0, M).
\end{aligned}$$

- $\mathcal{Q}_n(\beta) \in \mathcal{C}^2$ w.r.t. β.

$$\begin{aligned}
\mathcal{Q}_n^{(2)}(\beta_0, Y^n) &= \frac{1}{n}\sum_{i=1}^{n} -\left(y_i - (x_i'\beta_0)^3\right)\left[6(x_i'\beta_0)^4 x_i' x_i\right] + 9\left(x_i'\beta_0\right)^4 x_i x_i' \\
&\xrightarrow{d} \mathbb{E}\left[9(x_i'\beta_0)^4 x_i x_i'\right] \\
&\equiv H.
\end{aligned}$$

As we know $M = \mathbb{E}\left[9u_i^2(x_i'\beta_0)^4 x_i x_i'\right]$ and $H = \mathbb{E}\left[9(x_i'\beta_0)^4 x_i x_i'\right]$, we have:

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \underbrace{H^{-1}MH^{-1}}_{V}).$$

where

$$V = \left(\mathbb{E}\left[9(x_i'\beta_0)^4 x_i x_i'\right]\right)^{-1} \mathbb{E}\left[9u_i^2(x_i'\beta_0)^4 x_i x_i'\right] \left(\mathbb{E}\left[9(x_i'\beta_0)^4 x_i x_i'\right]\right)^{-1}$$

$$\begin{aligned}
\Rightarrow \hat{V} &= \frac{1}{9}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^4 x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^4 x_i x_i'\hat{u}_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^4 x_i x_i'\right)^{-1} \\
\overset{\mathbb{E}[u_i^2|x_i]=0}{\Longrightarrow} \hat{V} &= \frac{1}{9}\left(\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^4 x_i x_i'\right)^{-1}\sigma^2\mathbb{E}\left[(x_i'\beta)^4 x_i x_i'\right]\left(\frac{1}{n}\sum_{i=1}^{n}(x_i'\hat{\beta})^4 x_i x_i'\right)^{-1} \\
&= \frac{\hat{\sigma}^2}{9}\left(\mathbb{E}\left[(x_i'\beta)^4 x_i x_i'\right]\right)^{-1} \\
&= \frac{\hat{\sigma}^2}{9}\left[\frac{1}{n}\sum(x_i'\beta)^4 x_i x_i'\right]^{-1}
\end{aligned}$$

Appendix for Extremum Estimation

Lecture 10.

Cross-Sectional Topics

10.1 Recall

For sample $X : \{x_i\}_{i=1}^n$, we draw $p(x;\theta)$ giving us :

- point estimator $\hat{\theta}$:
  - $p_n(\hat{\theta})$ in finite samples
  - $p(\hat{\theta})$ in Asymptotic samples, where $\operatorname*{plim}_{n\to\infty}(\hat{\theta})$ and $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathcal{N}(0,V)$.
- Hypothesis testing: $\mathcal{H}_0 : \theta = \theta_0 \to g(\theta) : \theta - \theta_0 = 0$.
- CI construction:

10.2 Parameter Transformation

Example 10.

LRM: $y_i = x_i'\beta + u_i$, interest in: $\beta : \tilde{x}'\beta = \mathbb{E}[y_i \mid x_i = \tilde{x}]$.

Probit:

$$y_i^* = x_i'\beta + u_i$$
$$y_i = \mathbf{1}\{y_i^* > 0\}.$$

interest (mostly) in $\mathbb{E}[y_i \mid x_i = \tilde{x}] = \mathbb{E}[y_i = 1 \mid x_i = \tilde{x}] = \Phi(\tilde{x}_i'\beta)$.

1. Point estimator
2. Hypothesis testing & CI construction:
   - We know how to test: $\mathcal{H}_0 : g(\theta) = 0$, apply to $\mathcal{H}_0 : f(\theta) = f_0$, rewrite $\mathcal{H}_0 : g(\theta) = f(\theta) - f_0 = 0$.
   - We know that $CI : \{\mathcal{H}_0 : f(\theta) = f_0$ is accepted$\}$. But finding this set can be very hard.

     Example 11.

     $$\begin{aligned} T_w &= ng(\hat{\theta})'[g(\hat{\theta})g(\hat{\theta})']^{-1}g(\hat{\theta}) < c \\ &= n\left[f(\hat{\theta}) - f_0\right]'[GVG]^{-1}\left[f(\hat{\theta}) - f_0\right] < c \\ &\xrightarrow{d} \chi^2(r), \end{aligned}$$

     where $r = \text{rank}[g(\theta)]$

   - It's easier if $f(\theta)$ is a scalar. Try to find the distribution of $f(\hat{\theta})$.

- Analytically, in finite samples:

  If $\hat{\beta} \mid X \sim \mathcal{N}\left(\beta_0, \underbrace{\sigma^2(X'X)^{-1}}_{V}\right)$, then,

$$\underbrace{\tilde{x}'\hat{\beta}}_{\hat{\delta}} \mid X \sim \mathcal{N}\left(\tilde{x}'\beta_0, \tilde{x}'V\tilde{x}\right)$$
$$\hat{\delta} \sim \mathcal{N}(\delta_0, V_\delta)$$
$$\Rightarrow T_t(x) = \left|\frac{\delta - \delta_0}{\sqrt{V_\delta}}\right|$$

- Analytically, in asymptotic properties:

  Example 12. $\Phi(\tilde{x}'\hat{\beta})$

  * No finite sample distribution
  * Asymptotic distribution based on Delta Method:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &\xrightarrow{d} \mathcal{N}(0, V) \\ \Rightarrow \sqrt{n}\left(g(\hat{\theta}) - g(\theta)\right) &\xrightarrow{d} \mathcal{N}(0, GVG') \\ \Rightarrow \sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} \mathcal{N}(0, V_\beta) \\ \Rightarrow \sqrt{n}\left(\Phi(\tilde{x}'\hat{\beta}) - \Phi(\tilde{x}'\beta_0)\right) &\xrightarrow{d} \mathcal{N}\left(0, \phi(\tilde{x}')\tilde{x}'V\tilde{x}\phi(\tilde{x}'\beta)\right) \\ \Rightarrow \Phi(\tilde{x}'\hat{\beta}) &\xrightarrow{d} \mathcal{N}\left(\Phi(\tilde{x}'\beta_0), \frac{\phi^2(\tilde{x}'\beta_0)\tilde{x}'V\tilde{x}}{n}\right) \end{aligned}$$

- Bootstrapping: to get distribution of $f(\hat{\theta})$, take $\left\{\hat{\theta}^m\right\}_{i=1}^{M}$ as the estimator applied to data $\{x_i^m\}_{i=1}^{n_B}$. Then, we get the distribution of $f(\hat{\theta})$: take $\left\{f(\hat{\theta^m})\right\}_{i=1}^{M}$.

10.3 Instrumental Variables

Background: $y_i = x_i'\beta + u_i$, with $\mathbb{E}[x_i u_i] \neq 0$, meaning x_i is endogenous. β is consistent if $\mathbb{E}[x_i u_i] = 0$.

Example 13.

So, we want to find a way to estimate β consistently when x_i is endogenous.

$$\begin{aligned} y_i &= x_i'\beta + u_i \\ x_i &= z_i'\gamma + e_i \end{aligned}$$

1. z_i is exogenous to error term u_i: $\mathbb{E}[z_i u_i] = 0$.
2. z_i is relevant to regressor x_i: $\mathbb{E}[z_i v_i] \neq 0$

Then, we have the 2SLS method:

1. Estimate $\hat{\gamma}$ from $x_i = z_i'\gamma + e_i$. $\hat{\gamma} = (Z'Z)^{-1}Z'X$ and

$$\begin{aligned} \hat{x}_i &= z_i'\hat{\gamma} \\ \hat{X} &= Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X \\ &= Z\gamma + Z(Z'Z)^{-1}Z'e \\ &= X + Z(Z'Z)^{-1}Z'e \end{aligned}$$

2. Estimate $\hat{\beta}$ from $y_i = \hat{x}_i'\beta + u_i^*$. This gives us

$$\begin{aligned}
\hat{\beta}_{2SLS} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left((P_Z X)'P_Z X\right)^{-1}(P_Z X)'Y \\
&= \left(X'P_Z'P_Z X\right)^{-1}X'P_Z Y \\
&= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= \beta + (\cdots)\underbrace{Z'U}_{V}
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_{IV} &= \left(\sum_{i=1}^{n} z_i x_i'\right)^{-1}\sum_{i=1}^{n} z_i y_i \\
\sqrt{n}(\hat{\beta}_{IV} - \beta) &\xrightarrow{d} \mathcal{N}(0, V_{IV})
\end{aligned}$$

- V_{IV} is not easy to find.
- $CI : \{\mathcal{H}_0 : \beta = \beta_0$ is accepted$\}$.

Lecture 11.

Appendix

.Recommended Resources

Books

[1] James H. Stock and Mark W. Watson. *Introduction to Econometrics*. 4th ed. New York: Pearson, 2003

[2] Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach*. 7th ed. Cengage Learning, 2020

[3] Bruce E. Hansen. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2022

[4] Fumio Hayashi. *Econometrics*. Princeton, New Jersey: Princeton University Press, 2000

[5] Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, Massachusetts: The MIT Press, 2010

[6] Joshua Chan et al. *Bayesian Econometric Methods*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press, 2019

[7] Badi H. Baltagi. *Econometric Analysis of Panel Data*. 6th ed. Cham, Switzerland: Springer, 2021

[8] James D. Hamilton. *Time Series Analysis*. Princeton, New Jersey: Princeton University Press, 1994. ISBN: 9780691042893

Others

[9] Chen N Yang and Robert L Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. In: *Physical Review* 96.1 (Oct. 1, 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191