Lecture 7: Confidence intervals

Economics 326 — Econometrics II

Vadim Marmer, UBC

Point estimation

- Our model:
  1. $Y_i = \beta_0 + \beta_1 X_i + U_i, \qquad i = 1, \dots, n.$
  2. $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$ for all i's.
  3. $\mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2$ for all i's.
  4. $\mathrm{E}\left[U_i U_j \mid \mathbf{X}\right] = 0$ for all $i \neq j$.
  5. U's are jointly normally distributed conditional on $\mathbf{X}$.
- The OLS estimator $\hat{\beta}_1$ is a **point estimator** of β_1.
- With probability **one**, we have that $\hat{\beta}_1 \neq \beta_1$.
- To construct interval estimators, we need to know the distribution of $\hat{\beta}_1$.

Normal distribution

- A normal rv is a continuous rv that can take on any value. The PDF of a normal rv X is

  $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ where}$$

  $$\mu = \mathrm{E}\left[X\right] \text{ and } \sigma^2 = \mathrm{Var}\left(X\right).$$

  We usually write $X \sim N(\mu, \sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$, then $a + bX \sim N(a + b\mu, b^2\sigma^2)$.

Standard normal distribution

- Standard normal rv has $\mu = 0$ and $\sigma^2 = 1$. Its PDF is $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$.
- Symmetric around zero (mean): if $Z \sim N(0,1)$, $P(Z > z) = P(Z < -z)$.
- Thin tails: $P(-1.96 \leq Z \leq 1.96) = 0.95$.
- If $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0,1)$.

Bivariate normal distribution

- X and Y have a bivariate normal distribution if their joint PDF is given by:

  $$f(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_X^2\sigma_Y^2}} \exp\left[-\frac{Q}{2(1-\rho^2)}\right],$$

  where $Q = \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}$,

  $\mu_X = \mathrm{E}\left[X\right], \mu_Y = \mathrm{E}\left[Y\right], \sigma_X^2 = \mathrm{Var}\left(X\right), \sigma_Y^2 = \mathrm{Var}\left(Y\right)$, and $\rho = \mathrm{Corr}(X,Y)$.

Properties of bivariate normal

If X and Y have a bivariate normal distribution:

- $a + bX + cY \sim N(\mu^*, (\sigma^*)^2)$, where

$$\mu^* = a + b\mu_X + c\mu_Y, \quad (\sigma^*)^2 = b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\rho\sigma_X\sigma_Y.$$

- $\text{Cov}\,(X, Y) = 0 \Longrightarrow X$ and Y are independent.
- Can be generalized to more than 2 variables (multivariate normal).

Normality of the OLS estimator

- Assume that U_i's are jointly normally distributed conditional on $\mathbf{X}$ (Assumption 5).
- Then $Y_i = \beta_0 + \beta_1 X_i + U_i$ are also jointly normally distributed conditional on $\mathbf{X}$.
- Since $\hat{\beta}_1 = \sum_{i=1}^n w_i Y_i$, where $w_i = \frac{X_i - \bar{X}}{\sum_{l=1}^n (X_l - \bar{X})^2}$ depend only on $\mathbf{X}$, $\hat{\beta}_1$ is also normally distributed conditional on $\mathbf{X}$.
- Conditionally on $\mathbf{X}$:

$$\hat{\beta}_1 \mid \mathbf{X} \sim N\left(\beta_1, \text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)\right),$$
$$\text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}.$$

Interval estimation problem

- We want to construct an **interval estimator** for β_1:
  - The interval estimator is called a **confidence interval** (CI).
  - A CI contains the **true** value β_1 **with some pre-specified probability** $1 - \alpha$, where α is a small probability of error.
  - For example, if $\alpha = 0.05$, then the random CI will contain β_1 with probability 0.95.
- $1 - \alpha$ is called the **coverage probability**.
- Confidence interval: $CI_{1-\alpha} = [LB_{1-\alpha}, UB_{1-\alpha}]$. The lower bound (LB) and upper bound (UB) should depend on the coverage probability $1 - \alpha$.
- The formal definition of CI: It is a **random interval** $CI_{1-\alpha}$ such that conditionally on $\mathbf{X}$,

$$P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) = 1 - \alpha.$$

  Note that the random element is $CI_{1-\alpha}$.
- Sometimes, a CI is defined as $P\left(\beta_1 \in CI_{1-\alpha}\right) \geq 1 - \alpha$.

Symmetric CIs

- One approach to constructing CIs is to consider a **symmetric** interval around the estimator $\hat{\beta}_1$:

$$CI_{1-\alpha} = \left[\hat{\beta}_1 - c_{1-\alpha}, \hat{\beta}_1 + c_{1-\alpha}\right]$$

- The problem is choosing $c_{1-\alpha}$ such that $P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) = 1 - \alpha$.
- In choosing $c_{1-\alpha}$ we will be relying on the fact that given our assumptions and conditionally on $\mathbf{X}$:

$$\hat{\beta}_1 \mid \mathbf{X} \sim N\left(\beta_1, \text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)\right),$$
$$\text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}.$$

- Note that conditionally on $\mathbf{X}$:

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}} \sim N(0,1).$$

Standard normal quantiles

- Let $Z \sim N(0,1)$. The τ-th **quantile** (percentile) of the standard normal distribution is z_τ such that

$$P(Z \leq z_\tau) = \tau.$$

- **Median**: $\tau = 0.5$ and $z_{0.5} = 0$. ($P(Z \leq 0) = 0.5$).
- If $\tau = 0.975$ then $z_{0.975} = 1.96$. Due to symmetry, if $\tau = 0.025$ then $z_{0.025} = -1.96$.

σ^2 is known (infeasible CIs)

- **Suppose** (for a moment) that σ^2 is known, and we can compute exactly the variance of $\hat{\beta}_1$:

$$\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}.$$

- Consider the following CI:

$$\begin{aligned} CI_{1-\alpha} = \Big[&\hat{\beta}_1 - z_{1-\alpha/2}\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}, \\ &\hat{\beta}_1 + z_{1-\alpha/2}\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}\Big]. \end{aligned}$$

- For example, if $1 - \alpha = 0.95 \iff \alpha = 0.05 \iff z_{1-\alpha/2} = z_{0.975} = 1.96$, and

$$\begin{aligned} CI_{0.95} = \Big[&\hat{\beta}_1 - 1.96\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}, \\ &\hat{\beta}_1 + 1.96\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}\Big]. \end{aligned}$$

Infeasible CI validity (σ^2 known)

- We need to show that

$$P\left(\beta_1 \in CI_{1-\alpha} \mid \mathbf{X}\right) = 1 - \alpha.$$

- Next, let $\sigma_{\hat{\beta}_1} = \sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}$. Then:

$$\begin{aligned} &\hat{\beta}_1 - z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \\ &\iff -z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \leq \beta_1 - \hat{\beta}_1 \leq z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \\ &\iff -z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \leq \hat{\beta}_1 - \beta_1 \leq z_{1-\alpha/2}\,\sigma_{\hat{\beta}_1} \\ &\iff -z_{1-\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \leq z_{1-\alpha/2} \end{aligned}$$

Infeasible CI validity (σ^2 known)

- We have that

$$\begin{aligned} &\beta_1 \in CI_{1-\alpha} \\ &\iff -z_{1-\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)}} \leq z_{1-\alpha/2}. \end{aligned}$$

- Let $Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\mathrm{Var}(\hat{\beta}_1 | \mathbf{X})}} \sim N(0,1)$ conditionally on $\mathbf{X}$.

$$
\begin{aligned}
&P\left(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2} \mid \mathbf{X}\right) \\
&= P\left(z_{\alpha/2} \le Z \le z_{1-\alpha/2} \mid \mathbf{X}\right) \\
&= 1 - \alpha/2 - \alpha/2 = 1 - \alpha.
\end{aligned}
$$

Feasible CIs (σ^2 unknown)

- Since σ^2 is unknown, we must estimate it from the data:

$$
s^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{U}_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2.
$$

- We can replace σ^2 by s^2; however, the result does not have a normal distribution anymore:

$$
\begin{aligned}
&\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right)}} \sim t_{n-2}, \\
&\text{where } \widehat{\mathrm{Var}}\left(\hat{\beta}_1\right) = \frac{s^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.
\end{aligned}
$$

  Here t_{n-2} denotes the t-distribution with $n-2$ degrees of freedom.

- The degrees of freedom depend on
  - the sample size (n),
  - and the number of parameters one has to estimate to compute s^2 (two in this case, β_0 and β_1).

Feasible CIs (σ^2 unknown)

- Let $t_{df,\tau}$ be the τ-th quantile of the t-distribution with the number of degrees of freedom df: If $T \sim t_{df}$ then

$$
P\left(T \le t_{df,\tau}\right) = \tau.
$$

- Similarly to the normal distribution, the t-distribution is centered at zero and is symmetric around zero: $t_{n-2,1-\alpha/2} = -t_{n-2,\alpha/2}$.
- We can now construct a feasible confidence interval with $1-\alpha$ coverage as:

$$
\begin{aligned}
CI_{1-\alpha} = &\left[\hat{\beta}_1 - t_{n-2,1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right)},\right. \\
&\left.\hat{\beta}_1 + t_{n-2,1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right)}\right], \\
\text{where } &\widehat{\mathrm{Var}}\left(\hat{\beta}_1\right) = \frac{s^2}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}.
\end{aligned}
$$

Example: Data

- Data: `rental` from the `wooldridge` R package. 64 US cities in 1990.
  - `rent`: average monthly rent ($)
  - `avginc`: per capita income ($)
- Model: $\mathrm{Rent}_i = \beta_0 + \beta_1 \mathrm{AvgInc}_i + U_i$.

```r
library(wooldridge)
data("rental")
rental90 <- subset(rental, y90 == 1)
head(rental90[, c("city", "rent", "avginc")])
```

```
   city rent avginc
2     1  342  19568
4     2  496  31885
6     3  351  21202
8     4  588  29044
10    5  925  56307
12    6  630  35103
```

Example: OLS regression

```r
reg <- lm(rent ~ avginc, data = rental90)
summary(reg)
```

```

Call:
lm(formula = rent ~ avginc, data = rental90)

Residuals:
   Min     1Q Median     3Q    Max 
-94.67 -47.27 -13.68  25.65 228.46 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.488e+02  3.210e+01   4.635 1.89e-05 ***
avginc      1.158e-02  1.308e-03   8.851 1.34e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 66.56 on 62 degrees of freedom
Multiple R-squared:  0.5582,    Adjusted R-squared:  0.5511 
F-statistic: 78.34 on 1 and 62 DF,  p-value: 1.341e-12
```

Example: Extracting key values

```r
# Estimated slope and its standard error
beta1_hat <- coef(reg)["avginc"]
se_beta1  <- summary(reg)$coefficients["avginc", "Std. Error"]
cat("beta_1_hat =", round(beta1_hat, 4), "\n")
```

```
beta_1_hat = 0.0116 
```

```r
cat("SE(beta_1_hat) =", round(se_beta1, 4), "\n")
```

```
SE(beta_1_hat) = 0.0013 
```

```r
# Degrees of freedom: n - 2
n  <- nrow(rental90)
df <- n - 2
cat("n =", n, ",  df =", df, "\n")
```

```
n = 64 ,  df = 62 
```

Example: 95% confidence interval

```r
# Critical value
t_95 <- qt(0.975, df)
cat("t_{62, 0.975} =", round(t_95, 3), "\n")
```

```
t_{62, 0.975} = 1.999 
```

```r
# 95% CI: beta_1_hat +/- t * SE
CI_95 <- c(beta1_hat - t_95 * se_beta1,
           beta1_hat + t_95 * se_beta1)
round(CI_95, 4)
```

```
avginc avginc
0.0090 0.0142
```

```r
# Check with confint()
confint(reg, "avginc", level = 0.95)
```

```
             2.5 %     97.5 %
avginc 0.008964625 0.01419539
```

Example: 90% confidence interval

```r
# Critical value
t_90 <- qt(0.95, df)
cat("t_{62, 0.95} =", round(t_90, 3), "\n")
```

```
t_{62, 0.95} = 1.67
```

```r
# 90% CI: beta_1_hat +/- t * SE
CI_90 <- c(beta1_hat - t_90 * se_beta1,
           beta1_hat + t_90 * se_beta1)
round(CI_90, 4)
```

```
avginc avginc
0.0094 0.0138
```

```r
# Check with confint()
confint(reg, "avginc", level = 0.90)
```

```
               5 %       95 %
avginc 0.009395296 0.01376472
```

The effect of estimating σ^2

- The t-distribution has heavier tails than the normal.
- $t_{df,1-\alpha/2} > z_{1-\alpha/2}$, but as df increases $t_{df,1-\alpha/2} \to z_{1-\alpha/2}$.
- When the sample size n is large, $t_{n-2,1-\alpha/2}$ can be replaced with $z_{1-\alpha/2}$.

Interpretation of confidence intervals

- The confidence interval $CI_{1-\alpha}$ is a function of the **sample** $\{(Y_i, X_i) : i = 1, \ldots, n\}$, and therefore is **random**. This allows us to talk about the probability of $CI_{1-\alpha}$ containing the true value of β_1.
- Once the confidence interval is computed given the data, we have its **one realization**. The realization of $CI_{1-\alpha}$ (the computed confidence interval) is not random, and it does not make sense anymore to talk about the probability that it includes the true β_1.
- **Once the confidence interval is computed, it either contains the true value β_1 or it does not.**