Econometrics

Chapter 3: Simple Regression

In Choi

Sogang University

Introduction

- Regression is concerned with describing the relationship between one variable (explained variable, dependent variable, response variable, regressand) and one or more other variables (explanatory variables, independent variables, control variables, regressors).
- Notation:

y_i, y_t : dependent variable
x_{ji}, x_{jt} $(j = 1, ..., k)$: k independent variables.
i : index for cross-section data
t : index for time series data

Introduction

- Purposes of regression
  1. Examine whether or not $x_{1t}, ..., x_{kt}$ have significant effects on y_t.
  2. Predict the value of $y_{t+h} (h > 0)$ given $x_{1(t+h)}, ..., x_{k(t+h)}$.

Simple regression model

- We want to find the relationship

$$y_t = f(x_t)$$

for all t, where $f(\cdot)$ is a known function.

- In real world, it is difficult to find such an exact relationship due to:
  1. Randomness in human response
  2. Omitted independent variables not recognized
  3. Errors in measuring the dependent and independent variables.

For these unknown factors, we add an error term as

$$y_t = f(x_t) + u_t$$

where u_t is a random variable.

Simple regression model

- In simple linear regression, we assume that

$$f(x_t) = \mu + \beta x_t$$

where μ and β are constant (μ: intercept, β: slope). That is,

$$y_t = \mu + \beta x_t + u_t.$$

In this model, there is only one regressor.

- There can be various types of nonlinear regression. Examples are:

$$f(x_t) = \mu + \beta e^{\gamma x_t}$$

$$f(x_t) = \mu + \beta x_t \frac{1}{1 + e^{-\gamma(x_t - c)}} + \delta x_t, \gamma > 0$$

Simple regression model

- Linear regression: Linearity in parameters

$y_t = \mu + \beta \ln x_t + u_t$: Linear regression
$\ln y_t = \mu + \beta \ln x_t + u_t$: Linear regression
$y_t = \mu + \beta x_t^2 + u_t$: Linear regression

- Linear regression for log variables

$$\ln(y_t) = \mu + \beta \ln(x_t) + u_t$$

$$\beta = \frac{d\ln(y_t)}{d\ln(x_t)} = \frac{\frac{dy_t}{y_t}}{\frac{dx_t}{x_t}} : \text{elasticity}$$

Example

y_t : consumption, x_t : income $\Rightarrow$ β : income elasticity of consumption.

Ordinary least squares (OLS) estimation

- A problem for European researchers in late 18th century : Given data $(x_1, y_1), ..., (x_n, y_n)$, find $\hat{\mu}$ and $\hat{\beta}$ such that $\hat{y}_t = \hat{\mu} + \hat{\beta}x_t$ represents a line "close" to $(x_1, y_1), ..., (x_n, y_n)$. $\hat{\mu}$ and $\hat{\beta}$ are estimators of μ and β, respectively.

When $n = 2$, we may solve the equation:

$$y_1 = \mu + \beta x_1$$

$$y_2 = \mu + \beta x_2$$

When $n > 2$, we have too many equations for 2 unknowns μ and β. Legendre considered this problem in 1805 and devised the least squares method. See "The history of statistics : The measurement of uncertainty before 1900" by Stephen M. Stigler for details.

Ordinary least squares (OLS) estimation

- The OLS method obtain the estimation of μ and β by minimizing the sum:

$$Q(\mu, \beta) = \sum_{t=1}^{n} (y_t - \mu - \beta x_t)^2$$

with respect to μ and β.

Ordinary least squares (OLS) estimation

- First order conditions:

$$\frac{\partial Q}{\partial \mu} = 2\sum_{t=1}^{n}(y_t - \mu - \beta x_t)(-1) = 0$$

$$\frac{\partial Q}{\partial \beta} = 2\sum_{t=1}^{n}(y_t - \mu - \beta x_t)(-x_t) = 0.$$

These are called the normal equations. The solution of the equation are

$$\hat{\beta} = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

and

$$\hat{\mu} = \bar{y} - \hat{\beta}\bar{x}$$

where $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ and $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$. These are OLS estimators of β and μ, respectively.

Ordinary least squares (OLS) estimation

Because

$$\begin{aligned}\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y}) &= \sum_{t=1}^{n}x_ty_t-\bar{x}\sum_{t=1}^{n}y_t-\bar{y}\sum_{t=1}^{n}x_t+n\bar{x}\bar{y}\\ &= \sum_{t=1}^{n}x_ty_t-n\bar{x}\bar{y}-n\bar{x}\bar{y}+n\bar{x}\bar{y}\\ &= \sum_{t=1}^{n}x_ty_t-n\bar{x}\bar{y}\end{aligned}$$

and

$$\begin{aligned}\sum_{t=1}^{n}(x_t-\bar{x})^2 &= \sum_{t=1}^{n}x_t^2-2\bar{x}\sum_{t=1}^{n}x_t+n\bar{x}^2\\ &= \sum_{t=1}^{n}x_t^2-n\bar{x}^2,\end{aligned}$$

Ordinary least squares (OLS) estimation

we may write

$$\hat{\beta} = \frac{\sum_{t=1}^{n} x_t y_t - n\bar{x}\bar{y}}{\sum_{t=1}^{n} x_t^2 - n\bar{x}^2}.$$

Ordinary least squares (OLS) estimation

- Residuals

We write

$$\hat{y}_t = \hat{\mu} + \hat{\beta} x_t$$

and

$$\begin{aligned} \hat{u}_t &= y_t - \hat{\mu} - \hat{\beta} x_t \\ &= y_t - \hat{y}_t . \end{aligned}$$

We call $\hat{u}_t$ residuals. We deduce from the normal equations that

$$\sum_{t=1}^{n} \hat{u}_t = 0$$

$$\sum_{t=1}^{n} \hat{u}_t x_t = 0$$

Residuals $\hat{u}_t$ estimate the error terms u_t which are unobserved. However, not that $\sum_{t=1}^{n} u_t \neq 0$ and $\sum_{t=1}^{n} u_t x_t \neq 0$.

Ordinary least squares (OLS) estimation

- When $f(x) = \beta x_t$, (no intercept term), the OLS method minimizes

$$Q(\beta) = \sum_{t=1}^{n} (y_t - \beta x_t)^2$$

which gives the normal equation

$$\sum_{t=1}^{n} (y_t - \beta x_t)(-x_t) = 0$$

and

$$\hat{\beta} = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} x_t^2}$$

In this case, $\hat{u}_t = y_t - \hat{\beta} x_t$ satisfying the equation

$$\sum_{t=1}^{n} \hat{u}_t x_t = 0$$

($\sum_{t=1}^{n} \hat{u}_t = 0$ does not hold).

- To study the relationship between the firm performance measured by the average of past returns on equity and CEO compensation, consider the simple regression model

$$salary = \mu + \beta roe + u,$$

where *salary* =annual salary (1990, in thousands of US dollars) and *roe* =return on equity, 88-90 avg.

- Using the data set CEOSAL1.RAW[1], we obtain

$$\hat{salary} = 963.191 + 18.501 roe.$$

This result indicates that when the return on equity increases by one percent, the salary is predicted to change by about $18,500.

Is this regression result satisfactory? What are possible problems?

[1] CEOSAL1.RAW: A random sample of data reported in the May 6, 1991 issue of Businessweek.

- Is wage related to the education level? Of course it should be. To verify this empirically, consider the simple regression

$$wage = \mu + \beta edu + u,$$

where $wage$ = average hourly earnings and edu =years of education.

- Using the data set WAGE1.RAW[2], we obtain

$$wage = -0.90 + 0.54 edu.$$

  This result implies that one more year of education increases hourly wages by 54 cents an hour.

But there are many other variables that can explain one's wage. For example, tenure years with current employer, gender, race, marital status, number of dependents, industry, region, minimum wage law, etc., can be used.

[2] WAGE1.RAW: Data from the 1976 Current Population Survey, collected by Henry Farber.

Classical assumptions

Assumption 1: $\{x_t\}$ is a sequence of constants

Assumption 2 $u_t \sim iid(0, \sigma^2)$

- $E(u_t) = 0$
- $Var(u_t) = \sigma^2$: homoskedasticity
- u_t is independent of u_s $(t \neq s)$
- u_t have the same distribution
- Assumption 2 implies

$$Cov(u_t, u_s) = E(u_t u_s) = 0 \ (t \neq s).$$

  That is, u_t are not autocorrelated.
- $Var(u_t) = \sigma^2$ implies $Var(y_t) = \sigma^2$.

Properties of OLS estimation

- Unbiasedness
  1. $E(\hat{\mu}) = \mu$,
  2. $E(\hat{\beta}) = \beta$,
  3. $E(\hat{\sigma}^2) = \sigma^2$, where $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{t=1}^{n} \hat{u}_t{}^2$.

Properties of OLS estimation

- Proof of 2: Let

$$Q = \sum_{t=1}^{n} (x_t - \bar{x})^2 \text{ and } w_t = \frac{x_t - \bar{x}}{Q}.$$

Write

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{t=1}^{n} (x_t - \bar{x}) y_t}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \\
&= \sum_{t=1}^{n} \left(\frac{x_t - \bar{x}}{Q} \right) y_t \\
&= \sum_{t=1}^{n} w_t y_t \\
&= \sum_{t=1}^{n} w_t (\mu + \beta x_t + u_t) \\
&= \mu \sum_{t=1}^{n} w_t + \beta \sum_{t=1}^{n} w_t x_t + \sum_{t=1}^{n} w_t u_t .
\end{aligned}$$

Properties of OLS estimation

- But

$$\begin{aligned}\sum_{t=1}^{n} w_t &= \frac{1}{Q}\sum_{t=1}^{n}(x_t - \bar{x}) \\ &= \frac{1}{Q}(\sum_{t=1}^{n} x_t - n\bar{x}) = 0\end{aligned}$$

and

$$\begin{aligned}\sum_{t=1}^{n} w_t x_t &= \frac{1}{Q}\sum_{t=1}^{n}(x_t - \bar{x})x_t \\ &= \frac{1}{Q}\sum_{t=1}^{n}(x_t - \bar{x})^2 = 1\end{aligned}$$

which give

$$\hat{\beta} = \beta + \sum_{t=1}^{n} w_t u_t.$$

Properties of OLS estimation

- Therefore,

$$\begin{aligned} E(\hat{\beta}) &= \beta + E(\sum_{t=1}^{n} w_t u_t) \\ &= \beta + \sum_{t=1}^{n} w_t E(u_t) \\ &= \beta. \end{aligned}$$

Recall that

$$E(a_1 x_1 + ... + a_n x_n) = a_1 E(x_1) + ... + a_n E(x_n)$$

where a_i are constants and x_i are r.v.'s.

Properties of OLS estimation

Proof of 1: Write

$$\begin{aligned}\hat{\mu} &= \bar{y} - \hat{\beta}\bar{x} \\ &= \mu + \beta\bar{x} + \bar{u} - \hat{\beta}\bar{x} \\ &= \mu + \bar{u} - (\hat{\beta} - \beta)\bar{x}\end{aligned}$$

where $\bar{u} = \frac{1}{n}\sum_{t=1}^{n} u_t$. Thus,

$$E(\hat{\mu}) = \mu + E(\bar{u}) - \bar{x}E(\hat{\beta} - \beta) = \mu.$$

Properties of OLS estimation

- Variance

1. $$Var(\hat{\mu}) = \frac{\sigma^2}{n} \frac{\sum_{t=1}^{n} x_t^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2}.$$

2. $$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2}.$$

Proof of 2: Recall that

$$\hat{\beta} = \beta + \sum_{t=1}^{n} w_t u_t$$

and $E(\hat{\beta}) = \beta$. Therefore,

$$\begin{aligned}
Var(\hat{\beta}) &= E(\hat{\beta} - E(\hat{\beta}))^2 \\
&= E(\hat{\beta} - \beta)^2 \\
&= E(\sum_{t=1}^{n} w_t u_t)^2 \\
&= E(\sum_{t=1}^{n} w_t^2 u_t^2 + \sum_{t=1}^{n} \sum_{s=1 \ (t \neq s)}^{n} w_t w_s u_t u_s) \\
&= \sum_{t=1}^{n} w_t^2 E(u_t^2) + \sum_{t=1}^{n} \sum_{s=1 \ (t \neq s)}^{n} w_t w_s E(u_t u_s) \\
&= \frac{\sigma^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2}.
\end{aligned}$$

Properties of OLS estimation

Proof of 1: Recall that

$$\hat{\mu} - \mu = \bar{u} - (\hat{\beta} - \beta)\bar{x}$$

and that

$$E(\hat{\mu}) = \mu.$$

Thus,

$$\begin{aligned} Var(\hat{\mu}) &= E(\hat{\mu} - E(\hat{\mu}))^2 = E(\hat{\mu} - \mu)^2 = E(\bar{u} - (\hat{\beta} - \beta)\bar{x})^2 \\ &= E(\bar{u}^2) + \bar{x}^2 E(\hat{\beta} - \beta)^2 - 2\bar{x}E(\bar{u}(\hat{\beta} - \beta) \end{aligned}$$

Properties of OLS estimation

But

$$
\begin{aligned}
E(\bar{u}^2) &= \frac{1}{n^2} E(\sum_{t=1}^{n} u_t)^2 \\
&= \frac{1}{n^2} [\sum_{t=1}^{n} E(u_t^2) + \sum_{t=1}^{n} \sum_{s=1 \ (t \neq s)}^{n} E(u_t u_s)] \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

and

$$
\begin{aligned}
E(\bar{u}(\hat{\beta} - \beta)) &= E\left(\frac{\sum_{t=1}^{n} u_t}{n}\right) (\sum_{t=1}^{n} w_t u_t) \\
&= E(\frac{1}{n} \sum_{t=1}^{n} w_t u_t^2 + \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1 \ (t \neq s)}^{n} w_t u_t u_s) \\
&= \sigma^2 \frac{1}{n} \sum_{t=1}^{n} w_t + \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1 \ (t \neq s)}^{n} w_t E(u_t u_s)
\end{aligned}
$$

Properties of OLS estimation

Thus,

$$
\begin{aligned}
Var(\hat{\mu}) &= \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \\
&= \frac{\sigma^2}{n}\frac{\sum_{t=1}^{n} x_t^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}.
\end{aligned}
$$

- Interpretations
  1. For any sample size, the pdfs of $\hat{\mu}$ and $\hat{\beta}$ are centered around μ and β respectively.
  2. As sample size n grows, $Var(\hat{\mu})$ and $Var(\hat{\beta})$ decreases. This implies $\hat{\mu}$ gets closer to μ as sample size increase.
  3. Higher variation in x decreases the variance of $\hat{\beta}$.

- Under the classical assumptions, the OLS estimator is the best, linear, unbiased estimator (BLUE). That is, it is the most preferred estimation in the class of linear unbiased estimators.
- Linear: linear combination of y_t,

$$\hat{\beta} = \sum_{t=1}^{n} w_t y_t, \; w_t = \frac{x_t - \bar{x}}{\sum_{t=1}^{n}(x_t - \bar{x})^2},$$

$$\begin{aligned}\hat{\mu} &= \bar{y} - \hat{\beta}\bar{x} \\ &= \frac{1}{n}\sum_{t=1}^{n} y_t - \bar{x}\sum_{t=1}^{n} w_t y_t \\ &= \sum_{t=1}^{n}\left(\frac{1}{n} - \bar{x}w_t\right)y_t.\end{aligned}$$

Gauss-Markov theorem

- Unbiased:

$$E(\hat{\mu}) = \mu, E(\hat{\beta}) = \beta.$$

- Best: Minimum variance (most accurate)

Example

Consider the model

$$y_t = \beta x_t + u_t$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2}$$

$$\hat{\beta}_2 = \frac{\sum_{t=1}^n y_t}{\sum_{t=1}^n x_t}$$

Gauss-Markov theorem

Example

(continued) Both are linear and unbiased estimator. $\hat{\beta}_2$ is unbiased since

$$\hat{\beta}_2 = \frac{\beta \sum_{t=1}^{n} x_t + \sum_{t=1}^{n} u_t}{\sum_{t=1}^{n} x_t} = \beta + \frac{\sum_{t=1}^{n} u_t}{\sum_{t=1}^{n} x_t}$$

$$\Rightarrow E(\hat{\beta}_2) = \beta + E\left(\frac{\sum_{t=1}^{n} u_t}{\sum_{t=1}^{n} x_t}\right) = \beta + \frac{\sum_{t=1}^{n} E(u_t)}{\sum_{t=1}^{n} x_t} = \beta$$

But $\hat{\beta}_1$ is preferred because it is the OLS estimator.

Coefficient of determination

- Is the variation in the dependent variable largely explained by the variation in the independent variable?

Write

$$\begin{aligned} y_t &= \hat{\mu} + \hat{\beta}x_t + \hat{u}_t \\ &= \hat{y}_t + \hat{u}_t \end{aligned}$$

$$\Rightarrow y_t - \bar{y} = \hat{y}_t - \bar{y} + \hat{u}_t$$

where

$y_t - \bar{y}$: variation in the y_t around its mean

$\hat{y}_t - \bar{y}$: variation in $\hat{y}_t$ (part of y_t explained by x_t) around its mean ($\bar{\hat{y}} = \frac{1}{n}\sum_{t=1}^{n}\hat{y}_t = \hat{\mu} + \hat{\beta}\frac{1}{n}\sum_{t=1}^{n}x_t = \hat{\mu} + \hat{\beta}\bar{x} = \bar{y}$).

$\hat{u}_t$: variation in y_t not explained by x_t.

Coefficient of determination

- Consider

$$\sum_{t=1}^{n}(y_t - \bar{y})^2 = \sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{n} \hat{u}_t^2$$

where
$\sum_{t=1}^{n}(y_t - \bar{y})^2$: Total sum of squares (TSS)
$\sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2$: Explain sum of squares (ESS)
$\sum_{t=1}^{n} \hat{u}_t^2$: Residual sum of squares (RSS)

Coefficient of determination

- Coefficient of determination (R^2)

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

1. $0 \leq R^2 \leq 1$
2. The closer to one R^2 is , the better x_t explains y_t.

Assume: $u_t \sim iidN(0, \sigma^2)$. Then

1. $$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n} \frac{\sum_{t=1}^n x_t^2}{\sum_{t=1}^n (x_t - \bar{x})^2}\right).$$

2. $$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{t=1}^n (x_t - \bar{x})^2}\right).$$

Note that

$$z_t \sim iidN(\mu, \omega^2) \text{ or } z_t - u \sim iidN(0, \omega^2)$$

$$\Rightarrow \sum_{t=1}^n a_t z_t \sim N\left(\mu \sum_{t=1}^n a_t, \omega^2 \sum_{t=1}^n a_t^2\right)$$

That is, a linear combination of iid normal variables has a normal distribution.

Proof of 2. Recall that

$$\hat{\beta} = \beta + \sum_{t=1}^{n} w_t u_t$$

Thus,

$$\hat{\beta} - \beta \sim N\left(0, \sigma^2 \sum_{t=1}^{n} w_t^2\right) = N\left(0, \frac{\sigma^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}\right)$$

or

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}\right).$$

(Note that when $Z \sim N(\mu, \omega^2)$, $aZ \sim N(a\mu, a^2\omega^2)$).

Proof of 1. We have

$$
\begin{aligned}
\hat{\mu} - \mu &= \bar{u} - (\hat{\beta} - \beta)\bar{x} \\
&= \sum_{t=1}^{n} \left(\frac{1}{n} - \bar{x} w_t\right) u_t \\
&\sim N\left(0, \sigma^2 \sum_{t=1}^{n} \left(\frac{1}{n} - \bar{x} w_t\right)^2\right) \\
&= N\left(0, \frac{\sigma^2}{n} \frac{\sum_{t=1}^{n} x_t^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2}\right)
\end{aligned}
$$

as required.

- T-test

Using the above results, we have

$$\frac{\hat{\mu}-\mu}{\sqrt{\frac{\sigma^2}{n}\frac{\sum_{t=1}^{n}x_t^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2}}} \sim N(0,1)$$

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum_{t=1}^{n}(x_t-\bar{x})^2}}} \sim N(0,1)$$

Hypothesis testing and CI

Small sample approach

- Because σ^2 is not known, replace it with $\frac{1}{n-2}\sum \hat{u}_t^2$. Then,

$$t_\mu = \frac{\hat{\mu} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n} \frac{\sum_{t=1}^n x_t^2}{\sum_{t=1}^n (x_t - \bar{x})^2}}} \sim t(n-2)$$

$$t_\beta = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^n (x_t - \bar{x})^2}}} \sim t(n-2)$$

When we test the hypothesis,

$$H_0 : \beta = \beta^o \quad \text{versus} \quad H_1 : \beta \neq \beta^o$$

Compute t_{β^o} and reject when $|t_{\beta^o}| > t_{\alpha/2}(n-2)$, where $t_{\alpha/2}(n-2)$ is the percentil for the tail probablity $\alpha/2$ from Student's t-distribution with d.f. $n-2$..

Example

$n = 10$

$$\hat{y}_t = \underset{(0.11)}{0.24} + \underset{(0.03)}{0.97} x_t$$

The number in parenthesis are estimators of $\sqrt{Var(\hat{\alpha})}$ and $\sqrt{Var(\hat{\beta})}$ using $\hat{\sigma}^2$. Consider testing

$$H_0 : \beta = 1 \quad \text{versus} \quad H_1 : \beta \neq 1$$

at the 5% level. The t-test is computed as

$$t_1 = \frac{0.97 - 1}{0.03} = -1.0$$

Because

$$|-1.0| < t_{0.25}(8) = 2.306$$

do not reject the null.

- Confidence interval (CI)

Because

$$P(-t_{\alpha/2}(n-2) < t_\mu < t_{\alpha/2}(n-2)) = 1 - \alpha,$$

the $100 \cdot (1-\alpha)\%$ CI for μ is

$$\hat{\mu} - t_{\alpha/2}(n-2)SE(\hat{\mu}) < \mu < \hat{\mu} + t_{\alpha/2}(n-2)SE(\hat{\mu})$$

where

$$SE(\hat{\mu}) = \sqrt{\frac{\hat{\sigma}^2}{n} \frac{\sum_{t=1}^n x_t^2}{\sum_{t=1}^n (x_t - \bar{x})^2}}.$$

$SE(\hat{\mu})$, called standard error of $\hat{\mu}$, is an estimator of $\sqrt{Var(\hat{\mu})}$.

Hypothesis testing and CI

Small sample approach

- In the same manner, the $100 \cdot (1-\alpha)\%$ CI for β is

$$\hat{\beta} - t_{\beta/2}(n-2)SE(\hat{\beta}) < \beta < \hat{\beta} + t_{\beta/2}(n-2)SE(\hat{\beta})$$

where

$$SE(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^{n}(x_t - \bar{x})^2}}.$$

$SE(\hat{\beta})$ is an estimator of $\sqrt{Var(\hat{\beta})}$.

- The confidence level is chosen to be 90%, 95% or 99%. If it is too large, the interval is too wide to be useful. If it is too small, the interval is too narrow to be informative.

Example

(continued):

$$\begin{array}{cccc} \hat{y}_t = & 0.24 & + & 0.97x_t \\ & (0.11) & & (0.03) \end{array}$$

The 95% CI for β is:

$$(0.97 - 0.03 \cdot 2.306, 0.97 + 0.03 \cdot 2.306) = (0.90, 1.04).$$

Hypothesis testing and CI

Large sample approach

- Assume: $u_t \sim iid(0, \sigma^2)$ (No normality assumption). Then,

$$t_\mu = \frac{\hat{\mu} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n} \frac{\sum_{t=1}^n x_t^2}{\sum_{t=1}^n (x_t - \bar{x})^2}}} \simeq N(0,1)$$

and

$$t_\beta = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{t=1}^n (x_t - \bar{x})^2}}} \simeq N(0,1).$$

Thus, use standard normal distribution instead of t-distribution for hypothesis testing and CIs.

Example

(continued):

$$\hat{y}_t = 0.24 + 0.97x_t$$
$$(0.11) \qquad (0.03)$$

$$H_0 : \beta = 1 \quad \text{versus} \quad H_1 : \beta \neq 1$$

$t_\beta = -1$. Because $|-1| \leq z_{0.025} = 1.96$, do not reject the null. The 95% CI for β is

$$(0.97 - 0.03 \cdot 1.96, 0.97 + 0.03 \cdot 1.96)$$

$$H_0 : \beta = 1 \quad \text{versus} \quad H_1 : \beta < 1$$

Because $t_\beta > -z_{0.05} = -1.64$, do not reject the null.

Forecasting

- y_{n+h} can be forecasted by
$$\hat{y}_{n+h} = \hat{\mu} + \hat{\beta} x_{n+h}$$
when x_{n+h} is known.
- Error in forecasting y_{n+h} by $\hat{y}_{n+h}$ is
$$\begin{aligned} \hat{y}_{n+h} - y_{n+h} &= \hat{\mu} + \hat{\beta} x_{n+h} - \mu - \beta x_{n+h} - u_{n+h} \\ &= (\hat{\mu} - \mu) + (\hat{\beta} - \beta) x_{n+h} - u_{n+h}. \end{aligned}$$
This is called the the forecast error.
- Properties of the forecast error
  1. $$E(\hat{y}_{n+h} - y_{n+h}) = 0$$
  2. $$Var(\hat{y}_{n+h}) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+h} - \bar{x})^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \right).$$

Forecasting

- A practical problem is that x_{n+h} is not known. A regression model often used to circumvent this problem is the *predictive regression model* that is defined by

$$y_t = \mu + \beta x_{t-h} + u_t, \ (t = h + 1, ..., n).$$

In this model, the regressor is known when future values of y are predicted.

Example

y : return of S&P 500, x : divident yield, earning price ratio, etc.