Introduction to econometrics

III. Simple linear regression model

Content

Simple regression model

- No intercept.
- No need of matrix algebra.

$$Y_i = \beta X_i + \epsilon_i.$$

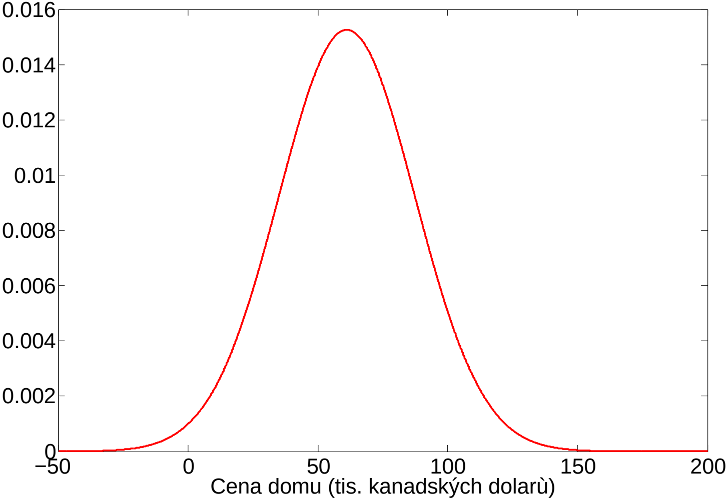
- Koop (2008) - chapter 3.

Content

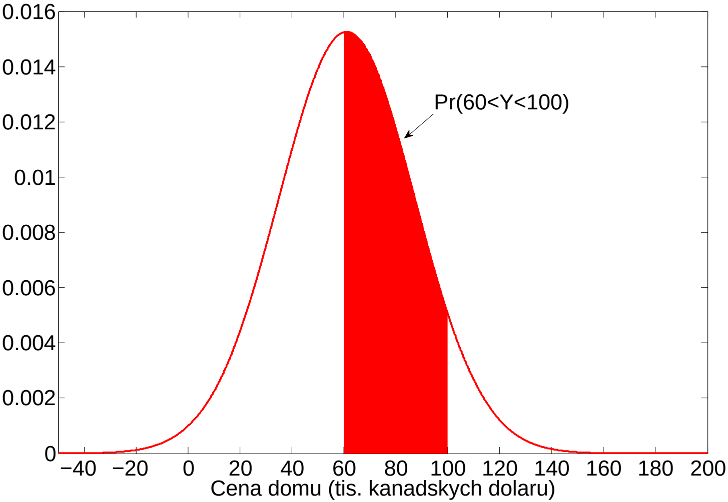
Randomness of dependent variable

- Uncertainty expressed by probability density function.
- House prices example – $N(61.153, 683.812)$.
- Koop (2008) - figure 3.1.

Normal p.d.f. of house price (lot size = 5000)



Normal p.d.f. of house price – density interval



Z-score

- $Y \sim N(\mu, \sigma^2)$ and new random variable:

$$Z = \frac{Y - \mu}{\sigma}.$$

- Expected value and variance of Z:

$$\begin{aligned} E(Z) &= E\left(\frac{Y-\mu}{\sigma}\right) = \frac{E(Y-\mu)}{\sigma} \\ &= \frac{E(Y)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0; \\ var(Z) &= var\left(\frac{Y-\mu}{\sigma}\right) = \frac{var(Y-\mu)}{\sigma^2} \\ &= \frac{var(Y)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$

- *Z-score*: $Z \sim N(0, 1)$.

Z-score – example

- Figure: $Y \sim N(61.153, 683.812)$.
- Shaded area: $\Pr(60 \leq Y \leq 100)$.

$$\begin{aligned}\Pr(60 \leq Y \leq 100) &= \Pr\left(\frac{60-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{100-\mu}{\sigma}\right) \\ &= \Pr\left(\frac{60-61.153}{\sqrt{683.812}} \leq \frac{Y-61.153}{\sqrt{683.812}} \leq \frac{100-61.153}{\sqrt{683.812}}\right) \\ &= \Pr\left(-0.04 \leq Z \leq 1.49\right).\end{aligned}$$

- Using statistical tables: $F(1.49) - F(-0.04) = 0.4479$.

Content

Classical assumption using dependent variable

1. $E(Y_i) = \beta X_i$.
2. $var(Y_i) = \sigma^2$.
3. $cov(Y_i, Y_j) = 0$ pro $i \neq j$.
4. Y_i is normally distributed.
5. X_i is fixed (not a random variable).

- Compact notation:

$$Y_i \sim N(\beta X_i, \sigma^2),$$

for $i = 1, \ldots, N$, where Y_i a Y_j are uncorrelated with one another if $i \neq j$.

Expected value of dependent variable

- $E(Y_i) = \beta X_i$: the linearity assumption
- Multiple regression model: $E(Y_i) = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$.
- We expect (in average) Y_i to lie on the regression line.

Alll observation have the same variance

- Violated in practice (house price example – prices of small and big houses; consumption example – households with different incomes).
- Homoskedasticity $\times$ heteroskedasticity.

Different observations are uncorrelated

- Remember:

$$corr(Y_i, Y_j) = \frac{cov(Y_i, Y_j)}{\sqrt{var(Y_i)var(Y_j)}}.$$

- $cov(Y_i, Y_j) = 0 \Leftrightarrow corr(Y_i, Y_j) = 0.$
- Not a problem for cross-sectional data.
- Time series – correlation over tima (e.g. unemployment).
- Autocorrelation.

Dependent variable normally distributed

- Hard to motivate.
- In many empirical applications – a reasonable assumption.
- Asymptotic theory to relax this assumption.

Fixed explanatory variables

- Experimental sciences – reasonable assumption.
- Economic research – usually not an experimental science.
- Asymptotic theory to relax this assumption (explanatory variables random but uncorrelated with the error term.

Classical assumptions using error term

1. $E(\epsilon_i) = 0$.
2. $var(\epsilon_i) = E(\epsilon_i^2) = \sigma^2$. Constant variability (homoskedasticity).
3. $cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. Errors are uncorrelated.
4. ϵ_i is normally distributed.
5. X_i is fixed (not random variable).

Violated classicalassumption.

- Problem $\rightarrow$ OLS results not correct.
- Testing using residuals (usually).
- Solution – asymptotic theory (large samples) or other estimation methods.

Content

Estimaor and estimate

- Estimator – a function of possible observations.
- Estimate – evaluating the estimator using data.
- Classical econometrics – unknown parameter is a scalar number $\times$ its estimate is a random variable.

OLS estimator

- No intercept:

$$Y_i = \beta X_i + \epsilon_i.$$

- Minimizing sum of squared resiuals:

$$SSR = \sum_{i=1}^{N} \widehat{\epsilon}_i^2 \to \min.$$

- Straightforward calculus problem⇒

$$\widehat{\beta} = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} X_i^2}.$$

- $\widehat{\beta}$ – normally distibuted (linear function in Y_i = normally distributed random variables).

Alternative expression for OLS estimator

- Replacing $Y_i \rightarrow$ true parameter + termi involving the explanatory variables and errors.
- Not useful for evaluating OLS estimator in practice.

$$\widehat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i (X_i \beta + \epsilon_i)}{\sum X_i^2}$$
$$= \beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}.$$

Expected value of the OLS estimator

- Unbiased estimator – „in average“ equals its true value.

$$E\left(\widehat{\beta}\right) = \beta.$$

$$\begin{aligned} E\left(\widehat{\beta}\right) &= E\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) = \beta + E\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\ &= \beta + \frac{1}{\sum X_i^2} E\left(\sum X_i \epsilon_i\right) = \beta + \frac{1}{\sum X_i^2} \sum X_i E(\epsilon_i) \\ &= \beta. \end{aligned}$$

Variance of the OLS estimator

- To quantify the variablitity of the random variable (to evaluate the effectivness of the estimator).

$$var\left(\widehat{\beta}\right) = \frac{\sigma^2}{\sum X_i^2}.$$

$$\begin{aligned} var\left(\widehat{\beta}\right) &= var\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) = var\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\ &= \left(\frac{1}{\sum X_i^2}\right)^2 var\left(\sum X_i \epsilon_i\right) = \left(\frac{1}{\sum X_i^2}\right)^2 \sum X_i^2 var(\epsilon_i) \\ &= \left(\frac{1}{\sum X_i^2}\right)^2 \sigma^2 \sum X_i^2 = \frac{\sigma^2}{\sum X_i^2}. \end{aligned}$$

Efficiency of the estimator

- Many unbiased estimator $\rightarrow$ choose the most efficient (the best), with the smallest variance.
- LInear regression model without intercept:

$$\tilde{\beta} = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i}.$$

- Unbiased wstimator, has a larger variance than OLS estimator.

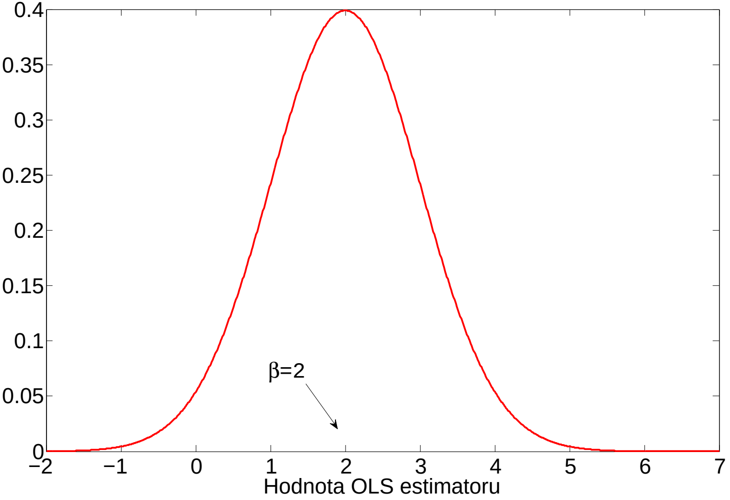
Distribution of the OLS estimator

- Under classical assumptions:

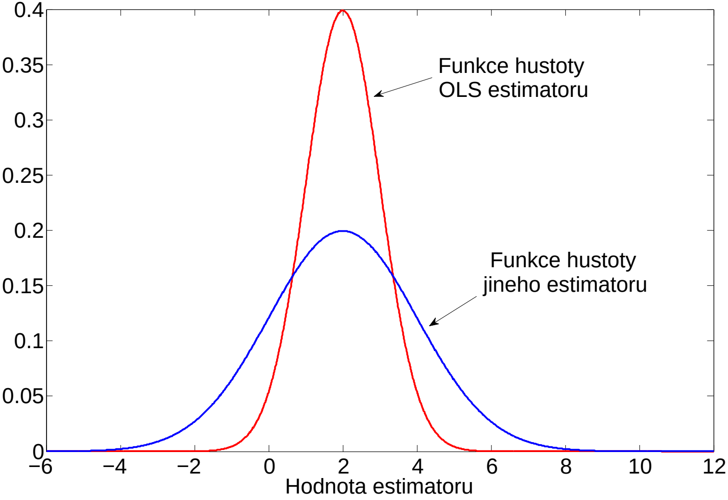
$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right).$$

- To derive confidence intervals and for hypothesis testing.
- Příklad: $\widehat{\beta} \sim N(2,1)$ a $\tilde{\beta} \sim N(2,4)$.
- Using statistical tables: e.g. $\Pr(1.0 \leq \widehat{\beta} \leq 3.0) = 0.68$ or $\Pr(0 \leq \widehat{\beta} \leq 1) = 0.14$.

The p.d.f. of the OLS estimator



OLS estimator and a less efficient estimator



Gauss-Markov theorem

- If the classical assumptions hold (not necessary normality): OLS estimator is BLUE (Best Linear Unbiased Estimator).

Carl Friedrich Gauss (1777–1855)

Andrej Markov (1856–1922)

Gauss-Markov theorem – unbiasness

- Simple LRM, no intercept, all assumptions hold.
- Linear estimator: $\beta^* = c_1 Y_1 + \ldots + c_N Y_N$ for constants $c_1, \ldots, c_N$.
- Unbiasness: $E(\beta^*) = \beta$.

$$\begin{aligned} E(\beta^*) &= E(c_1 Y_1 + \ldots + c_N Y_N) \\ &= c_1 E(Y_1) + \ldots + c_N E(Y_N) \\ &= c_1 \beta X_1 + \ldots + c_N \beta X_N \\ &= \beta \sum_{i=1}^{N} c_i X_i. \end{aligned}$$

- Unbiased estimator, β^* – it holds $\sum_{i=1}^{N} c_i X_i = 1$.

Gauss-Markov theroem – effectivness

$$\begin{aligned} var\left(\beta^*\right) &= var\left(c_1 Y_1 + \ldots + c_N Y_N\right) \\ &= c_1^2 var\left(Y_1\right) + \ldots + c_N^2 var\left(Y_N\right) \\ &= c_1^2 \sigma^2 + \ldots + c_N^2 \sigma^2 \\ &= \sigma^2 \sum_{i=1}^{N} c_i^2. \end{aligned}$$

- Looking for $c_1, \ldots, c_N$ to minimize $\sigma^2 \sum_{i=1}^{N} c_i^2$ subject to constraint $\sum_{i=1}^{N} c_i X_i = 1$.
- Standard calculus problem.

Gauss-Markov theorem – solution

- Solution: $c_j = \dfrac{X_j}{\sum_{i=1}^N X_i^2}$, pro $j = 1, \ldots, N$.
- Plugging this into expression for β^*:

$$\begin{aligned} \beta^* &= c_1 Y_1 + \ldots + c_N Y_N \\ &= \frac{X_1 Y_1}{\sum X_i^2} + \ldots + \frac{X_N Y_N}{\sum X_i^2} \\ &= \frac{\sum X_i Y_i}{\sum X_i^2} = \widehat{\beta}. \end{aligned}$$

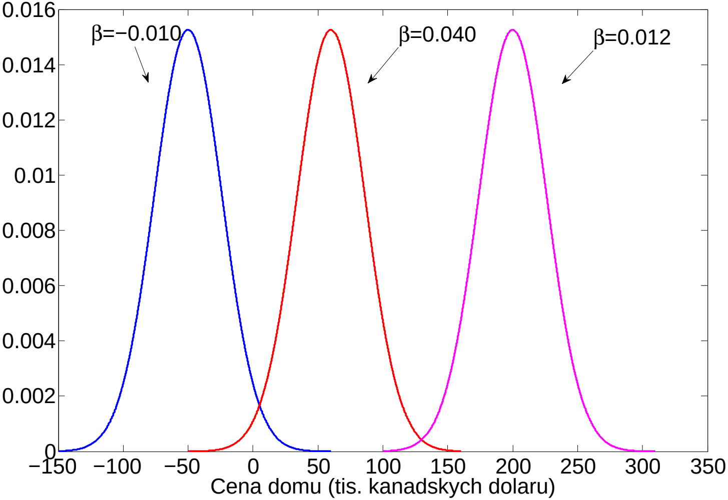
- Conclusion: linear unbiased estimator with minimum variance is the OLS estimator.

Content

Motivation

- Maximal likelihood (ML).
- Probability density function for Y should be in accordance with observed data.
- Example – house prices (for different parameter values β).

Motivation – figure



Likelihood function funkce

- Likelihood function: joint probability density function for all observation.
- Under classical assumption: $Y_1, \ldots, Y_N$ are normally distributed, uncorrelated random variables (mean βX_i and variance σ^2).
- P.D.F. for each random variable: $p(Y_i|X_i, \beta)$ (conditional densities).
- Joint p.d.f for all observations:

$$p\left(Y_1, \ldots, Y_N\right) = \prod_{i=1}^{N} p\left(Y_i|X_i, \beta\right).$$

ML estimate

- We know X_i and Y_i $\Rightarrow$ likelihood function depends on parameters – $L(\beta)$.
- ML estimate = selecting β that maximizes $L(\beta)$ (in case of LRM – without numerical optimalization).

$$p\left(Y_i|X_i\beta\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(Y_i - \beta X_i\right)^2\right].$$

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(Y_i - \beta X_i\right)^2\right] \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(Y_i - \beta X_i\right)^2\right]. \end{aligned}$$

Maximizing likelihood function.

- Working with loga-likelihood, $l(\beta)$ (maximum is not changed).

$$\begin{aligned} l(\beta) &= \ln\left[L(\beta)\right] \\ &= \ln\left\{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i-\beta X_i)^2\right]\right\} \\ &= \ln\left\{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\right\}-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i-\beta X_i)^2. \end{aligned}$$

- The value of β that maximizes $l(\beta)$ using calculus $\rightarrow$ one nedd to minimize $\sum_{i=1}^{N}(Y_i-\beta X_i)^2 \rightarrow$ OLS estimator.

Content

Confidence interval for the parameter

- $\widehat{\beta}$ is normally distributed with the mean β and a given variance (explained later).
- To construct *Z-skóru*:

$$Z = \frac{\widehat{\beta} - E(\widehat{\beta})}{\sqrt{var(\widehat{\beta})}} = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}}.$$

- $Z \sim N(0,1) \Rightarrow$ using statistical tables for standardized normal distribution.
- For example:

$$\Pr\left[-1.96 \leq Z \leq 1.96\right] = 0.95.$$

Deriving confidence interval

- 95% confidence interval $\rightarrow$ to put β in the middle:

$$\Pr\left[-1.96 \leq \frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}} \leq 1.96\right]$$

$$= \Pr\left[-1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}} \leq \widehat{\beta}-\beta \leq 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}\right]$$

$$= \Pr\left[\widehat{\beta}-1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}} \leq \beta \leq \widehat{\beta}+1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}\right]$$

$$= 0.95.$$

Alternative expressions

- 95% confidence interval:

$$\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}} \leq \beta \leq \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}$$

- Or:

$$\widehat{\beta} \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}$$

$$\left[\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}, \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}\right].$$

- „95 %“: *confidence level*.
- Other confidence levels can be handled beginning with a different probability (e.g. 90 %) ⇒ other „critical values“ (e.g. 1.64).
- Practical problem: unknown error variance, σ^2.

General steps on hypothesis testing

1. Specify a null hypothesos, H_0 and an alternative hypothesis H_1.
2. Specify test statistic.
3. Figure out the distribution of the test statistic, assuming H_0 is true.
4. Choose a level of significance.
5. Use steps 3 and 4 to obtain critical value.
6. Calculate your test statistics from step 2 and compare with the critical value from step 5:
   - Reject H_0 if the absolut value of the test statistic is greater than the critical value (else do not reject H_0.

Hypothesis test about a parameter

1. $H_0 : \beta = \beta_0$, where β_0 is known (usually $\beta_0 = 0$). Alternative hypothesis is usually $H_1 : \beta \neq \beta_0$ (one-sided hypothesis possible = other critical value).
2. $Z = \dfrac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}}$.
3. $Z = \dfrac{\widehat{\beta} - \beta_0}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}} \sim N(0,1)$.
4. Common choice 5 % (i.e. 0.05).
5. Since $Z \sim N(0,1)$ and $\Pr[-1.96 \leq Z \leq 1.96] = 0.95$, the critical value is 1.96.
   - Two-sided test (like two-sided confidence interval) → critical value = 97.5% (0.975) quantile of standardized normal distribution $(1 - 0.05/2)$.
6. In our case: reject H_0 if $|Z| > 1.96$.

Error variance estimator

- Commonly used estimator for σ^2 is reffered as s^2 (sample variance).
- OLS residuals: $\widehat{\epsilon}_i = Y_i - \widehat{\beta} X_i$; OLS estimator σ^2: $s^2 = \frac{\sum \widehat{\epsilon}_i^2}{N-1}$.
- May be shown that $E(s^2) = \sigma^2$.
- Intuition: $E\left(\epsilon_i^2\right) = \sigma^2 \Rightarrow \epsilon_i^2$ might be a good estimator for σ^2.
- Use all errors $\rightarrow$ sample average of squared errors to estimate σ^2: $\frac{\sum \epsilon_i^2}{N}$.
- Substitute unobserved ϵ_i: $\frac{\sum \widehat{\epsilon}_i^2}{N}$ (biased estimator for σ^2 – ML estimator).
- Unbiased OLS estimator: N replaced by $N-1$ (degrees of freedom)
- Multiple regression with k explanatory variables and with the intercept term:

$$s^2 = \frac{\sum \widehat{\epsilon}_i^2}{N-k-1} = \frac{SSR}{N-k-1}.$$

Modifications when error variance is unknown

- Construct Z-score:

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{s^2}{\sum X_i^2}}} \sim t_{N-1},$$

- Hypothesis testing and test statistics:

$$t = \frac{\widehat{\beta} - \beta_0}{\sqrt{\frac{s^2}{\sum X_i^2}}} \sim t_{N-1}.$$

- t_{N-1} is Student t-distribution with $N-1$ degrees of freedom.
- Statistical tables for t-distribution or p-value for the test (reject H_0 if p-value is less than significance level).

Content

Motivation

- What happens in large (infinite) samples?
- No assumpotion about what the p.d.f. of the errors is.
- X_i is *independent and identically distributed – i.i.d.*, independent of error terms $\Rightarrow E(X_i) = \mu_X$ and $var(X_i) = \sigma_X^2$.
- Use OLS estimator:

$$\widehat{\beta} = \beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}.$$

Consistency of OLS estimator – introduction

- $\text{plim}(\widehat{\beta}) = \beta$.

$$\begin{aligned}
\text{plim}\left(\widehat{\beta}\right) &= \text{plim}\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\
&= \beta + \text{plim}\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \text{ by Slutsky's theorem} \\
&= \beta + \text{plim}\left(\frac{\frac{1}{N}\sum X_i \epsilon_i}{\frac{1}{N}\sum X_i^2}\right) \\
&= \beta + \frac{\text{plim}\left(\frac{1}{N}\sum X_i \epsilon_i\right)}{\text{plim}\left(\frac{1}{N}\sum X_i^2\right)} \text{ by Slutsky's theorem.}
\end{aligned}$$

Consistency of OLS estimator (continued)

- *Law of large numbers* to figure out the plim $\left(\frac{1}{N}\sum X_i\epsilon_i\right)$.
- *„Average converge to expected values"*.
- Errors and explanatory variables are assumed to be independent of one another:

$$\text{plim}\left(\frac{1}{N}\sum X_i\epsilon_i\right) = E(X_i\epsilon_i) = 0$$

- Law of large numbers $\rightarrow$ plim $\left(\frac{1}{N}\sum X_i^2\right) = E(X_i^2)$.
- From definition of the variance, $var(X_i) = E(X_i^2) - [E(X_i)]^2$, we can write $E(X_i^2) = var(X_i) + [E(X_i)]^2 = \sigma_X^2 + \mu_X^2 \Rightarrow$

$$\text{plim}\left(\widehat{\beta}\right) = \beta + \frac{0}{\sigma_X^2 + \mu_X^2} = \beta.$$

Asymptotic normality – introduction

- As $N \to \infty$ we have:

$$\sqrt{N}(\widehat{\beta} - \beta) \sim N\left(0, \frac{\sigma^2}{\sigma_X^2 + \mu_X^2}\right).$$

- Equation can be written:

$$\sqrt{N}(\widehat{\beta} - \beta) = \sqrt{N}\frac{\sum X_i\epsilon_i}{\sum X_i^2} = \sqrt{N}\frac{\frac{1}{N}\sum X_i\epsilon_i}{\frac{1}{N}\sum X_i^2}.$$

- *Central limit theorem* as $N \to \infty$:

$$\sqrt{N}\frac{1}{N}\sum X_i\epsilon_i \sim N\left(0, var(X_i\epsilon_i)\right).$$

Asymptotic normality (continued)

$$\begin{aligned} var(X_i\epsilon_i) &= E(X_i^2\epsilon_i^2) - [E(X_i\epsilon_i)]^2 \\ &= E(X_i^2)E(\epsilon_i^2) - [E(X_i)E(\epsilon_i)]^2 \\ &= (\sigma_X^2 + \mu_X^2)\sigma^2 - [\mu_X 0]^2 \\ &= (\sigma_X^2 + \mu_X^2)\sigma^2. \end{aligned}$$

- We showed:

$$\text{plim}\left(\frac{1}{N}\sum X_i^2\right) = (\sigma_X^2 + \mu_X^2).$$

Asymptotic normality (end)

- Using Cramer's theorem we can combine the result form the central limit theorem with the forms for $var\left(X_i\epsilon_i\right)$ and plim $\left(\frac{1}{N}\sum X_i^2\right)$:

$$\sqrt{N}(\widehat{\beta}-\beta)\xrightarrow{N} N\left(0,\frac{(\sigma_X^2+\mu_X^2)\sigma^2}{(\sigma_X^2+\mu_X^2)^2}\right).$$

- Cancelling out the common factor in the variance:

$$\sqrt{N}(\widehat{\beta}-\beta)\xrightarrow{N} N\left(0,\frac{\sigma^2}{(\sigma_X^2+\mu_X^2)}\right).$$

Using the asymptotic results in practice

- As $N \to \infty$, $\sqrt{N}(\widehat{\beta} - \beta)$ converges to a normal distribution.
- Using the properties of the expected value and variance operators:

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{N(\sigma_X^2 + \mu_X^2)}\right).$$

- Problem: $(\sigma_X^2 + \mu_X^2)$ unknown.
- It holds $\text{plim}\left(\frac{1}{N}\sum X_i^2\right) = \sigma_X^2 + \mu_X^2 \Rightarrow \left(\frac{1}{N}\sum X_i^2\right)$ is a consistent estimator of $\sigma_X^2 + \mu_X^2$.
- Main result:

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right).$$

- All the derivations and results for finite samples hold (approximately)!