Econometrics I

Lecture 4

Understanding $\hat{\beta}$'s standard errors

Matías Cabello
matias.cabello@wiwi.uni-halle.de

October 27, 2025

Introduction

Tech Adoption: Linear vs Log-Linear Models

	Dependent variable:	
	Users	log(Users)
	Linear	Log-Linear
	(1)	(2)
Year	200.094***	$\boxed{0.241}$***
	(25.205)	(0.022)
Constant	-401,968.400***	$\boxed{-479.229}$***
	(50,787.630)	(44.104)
Observations	21	21
R2	0.768	$\boxed{0.864}$
Note:		*p<0.1; **p<0.05; ***p<0.01

$\hat{\beta}_2$ estimate

$\hat{\beta}_1$ estimate

R^2

Introduction


```
 Tech Adoption: Linear vs Log-Linear Models
=========================================================
                              Dependent variable:
                         ----------------------------
                             Users        log(Users)
                             Linear       Log-Linear
                              (1)            (2)
---------------------------------------------------------
Year                       200.094***     0.241***
                           (25.205)       (0.022)
Constant               -401,968.400***  -479.229***
                         (50,787.630)     (44.104)

---------------------------------------------------------
Observations                 21             21
R2                          0.768          0.864
=========================================================
Note:                  *p<0.1; **p<0.05; ***p<0.01
```

What is this?

Introduction

The information in parentheses corresponds to the respective **standard errors** of each estimated $\hat{\beta}_j$.

Define st. error of $\hat{\beta}_j \equiv s(\hat{\beta}_j) = \sqrt{\text{var}(\hat{\beta}_j)}$

Today's objective: understand what $s(\hat{\beta}_j)$ means and how to estimate it.

St. error $s(\hat{\beta})$ & sample size (n)

Reality (β) vs. estimate ($\hat{\beta}$)

Consider a data-generator process (a "population")

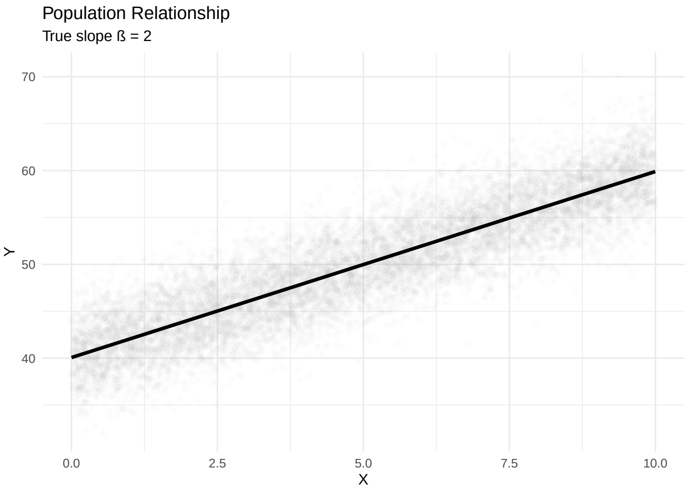
$$y = 40 + 2 \cdot x + u, \qquad \text{with random error } u \sim N(0, \sigma).$$

That is, the true parameter is $\beta = 2$.
In reality, we don't know the value of β. We only can guess the value through the estimation

$$\hat{\beta} = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

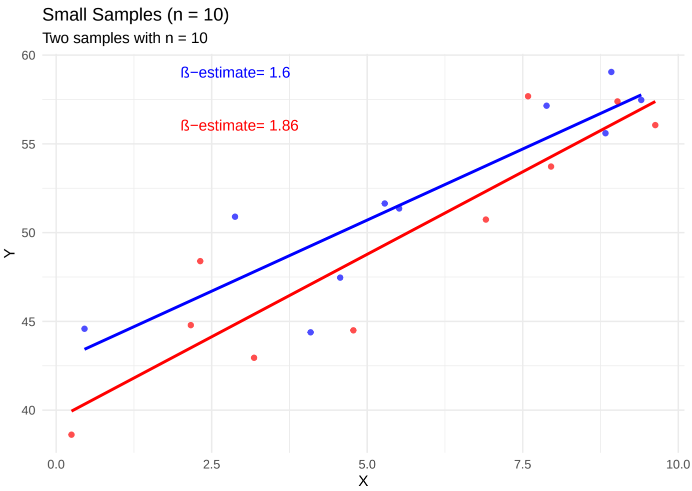
using n available data points.

Reality (β) vs. estimate ($\hat{\beta}$)

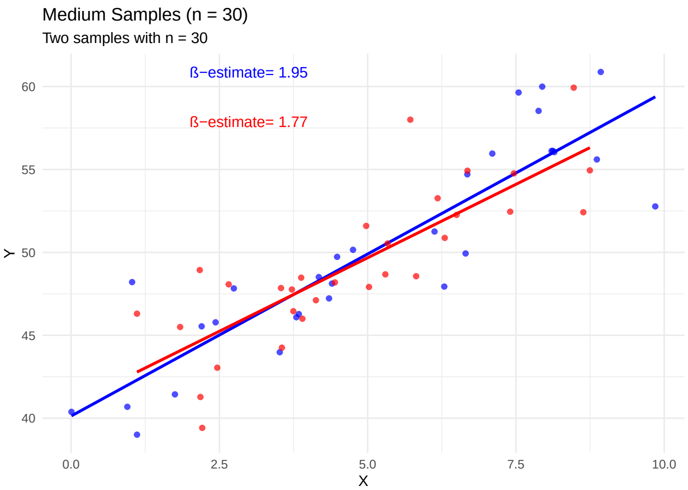


Will the sample size n affect the **precision** of our **estimate** $\hat{\beta}$?

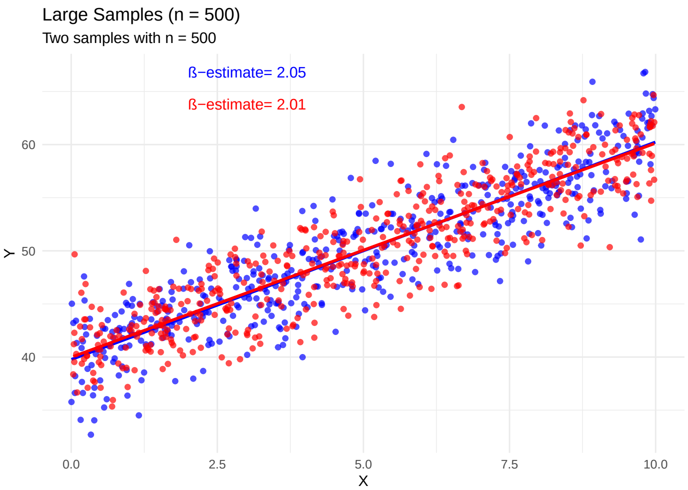
Sample size n vs. estimation reliability



Sample size n vs. estimation reliability



Sample size n vs. estimation reliability



Sample size n vs. estimation reliability

Takeaway:

- The larger the sample (the higher n) the more stable are the estimates of $\hat{\beta}_j$ – i.e., the lower is $s(\hat{\beta}_j) = \sqrt{\text{var}(\hat{\beta}_j)}$.

St. error $s(\hat{\beta}_j)$ & error variance (σ^2)

The importance of the error-variance σ^2

Let us play God and invent some data of the form

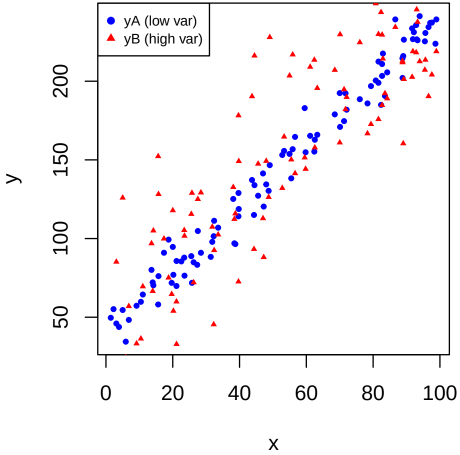
$$
\begin{aligned}
y_A &= 40 + 2 \cdot x + u_A, \; u_A \sim N(0, 10^2) \\
y_B &= 40 + 2 \cdot x + u_B, \; u_B \sim N(0, 40^2)
\end{aligned}
$$

Both populations are identical, except for the variances of the error terms u_A and u_B ($\sigma_B > \sigma_A$).

```
# Create dataset with 100 observations
n <- 100
x <- runif(n, 0, 100)  # Uniform distribution between
   0-100

# Generate y variables with different error variances
yA <- 40 + 2*x + rnorm(n, mean = 0, sd = 10)  # Low
   variance
yB <- 40 + 2*x + rnorm(n, mean = 0, sd = 40)  # High
   variance
```

The importance of the error-variance σ^2



Different error terms:

- Low-variance u_A vs. high-variance u_B.
- Both normally distributed

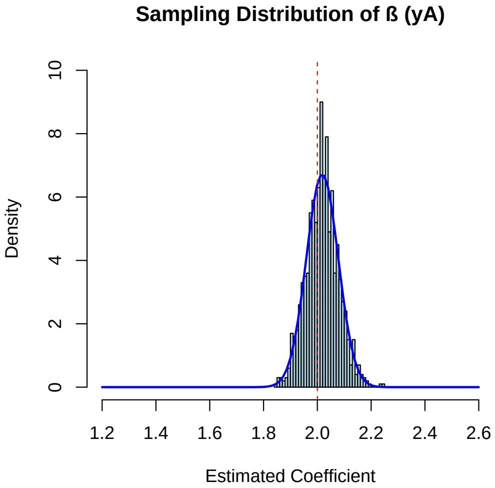
$$u_A \sim N(0, 10^2)$$
$$u_B \sim N(0, 40^2)$$

The importance of the error-variance σ^2

	Dependent variable:	
	yA (sigma = 10)	yB (sigma = 40)
x	2.015***	2.053***
	(0.032)	(0.132)
Constant	39.030***	39.310***
	(1.931)	(7.924)
Observations	100	100
R2	0.976	0.711
Note:	*p<0.1; **p<0.05; ***p<0.01	

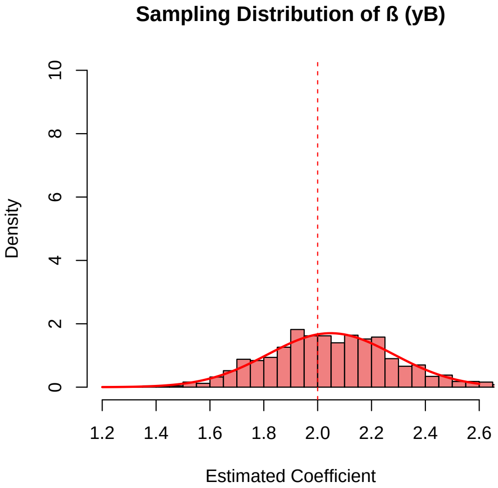
The importance of the error-variance σ^2

1000 different estimates $\hat{\beta}$ using $n = 30$ random samples of y_A



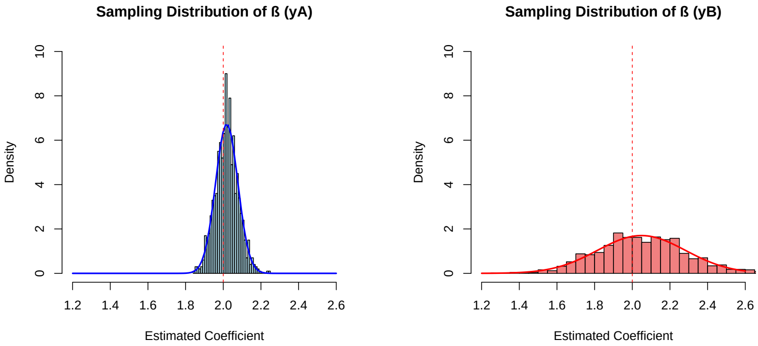
The importance of the error-variance σ^2

1000 different estimates $\hat{\beta}$ using $n = 30$ random samples of y_B



The importance of the error-variance σ^2

Comparison of Sampling Distributions



The importance of the error-variance σ^2

Takeaway:

- The smaller the error variance (the lower σ) the more stable are the estimates of $\hat{\beta}_j$ – i.e., the lower is $s(\hat{\beta}_j) = \sqrt{\text{var}(\hat{\beta}_j)}$.

The importance of the error-variance σ^2

Notice higher standard errors (in parentheses) of the coefficients of y_B:

	Dependent variable:	
	yA (sigma = 10)	yB (sigma = 40)
x	2.015*** (0.032)	2.053*** (0.132)
Constant	39.030*** (1.931)	39.310*** (7.924)
Observations	100	100
R2	0.976	0.711
Note:	*p<0.1; **p<0.05; ***p<0.01	

How is $s(\hat{\beta}_j)$ estimated?

In practice, we do not run thousands of estimates with different samples (like shown in the previous slides). Instead, we estimate the parameter vector $\hat{\beta} = \left[\hat{\beta}_1, \hat{\beta}_2, \dots \hat{\beta}_k\right]'$ just once.

Q: How do we then obtain an estimate of $s(\hat{\beta}_j) = \sqrt{\mathrm{var}(\hat{\beta}_j)}$?

A: In the simple case of the univariate model ($k = 2$), it is estimated with the analytical solution

$$\boxed{\widehat{\mathrm{var}}(\hat{\beta}_2|x) = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2}} \qquad \text{with } \hat{\sigma}^2 = \frac{\sum_i^n \hat{u}_i^2}{n-2}$$

That is the formula that statistical software use to estimate the st. errors $\hat{s}(\beta_2|x) = \sqrt{\widehat{\mathrm{var}}(\hat{\beta}_2|x)}$ shown in parenthesis in the introduction.

Where does the $\widehat{\text{var}}(\hat{\beta}_2|x)$ formula come from?

Assume true model is $\boxed{y = \beta_1 + x\beta_2 + u}$. Then

$$\hat{\beta}_2 = \frac{\text{cov}(x, y)}{\text{var}(x)} = \frac{\text{cov}(x, \beta_1 + x\beta_2 + u)}{\text{var}(x)}$$

By the rules described in the appendix:

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\text{cov}(x, x\beta_2)}{\text{var}(x)} + \frac{\text{cov}(x, u)}{\text{var}(x)} \\
&= \beta_2 \frac{\text{cov}(x, x)}{\text{var}(x)} + \frac{\text{cov}(x, u)}{\text{var}(x)} \\
&= \beta_2 \frac{\text{var}(x)}{\text{var}(x)} + \frac{\text{cov}(x, u)}{\text{var}(x)} \\
&= \beta_2 + \frac{\text{cov}(x, u)}{\text{var}(x)} = \beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}$$

Where does the $\widehat{\text{var}}(\hat{\beta}_2|x)$ formula come from?

$$
\begin{aligned}
\text{var}(\hat{\beta}_2|x) &= \text{var}\left(\beta_2 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\middle| x\right) \\
&= \text{var}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\middle| x\right) \\
&= \frac{1}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}\,\text{var}\left(\sum_{i=1}^n (x_i - \bar{x})u_i\middle| x\right)
\end{aligned}
$$

Let's first expand $\text{var}\left(\sum_{i=1}^n (x_i - \bar{x})u_i | x\right)$:

Where does the $\widehat{\text{var}}(\hat{\beta}_2|x)$ formula come from?

Let's first expand $\text{var}\left(\sum_{i=1}^{n}(x_i - \bar{x})u_i|x\right)$ using the appendix rule 3:

$$\text{var}\left(\sum_{i=1}^{n}(x_i - \bar{x})u_i \middle| x\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \bar{x})(x_j - \bar{x})\,\text{cov}(u_i, u_j|x)$$

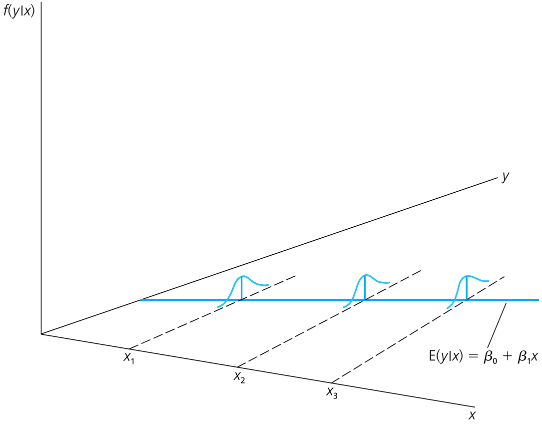
$$= \sum_{i=1}^{n}(x_i - \bar{x})^2\,\text{var}(u_i|x) + \sum_{i \neq j}(x_i - \bar{x})(x_j - \bar{x})\,\text{cov}(u_i, u_j|x)$$

We will now **assume**

- **Homoskedasticity**: $\text{var}(u_i|x) = \sigma^2$ for all i
- **No autocorrelation**: $\text{cov}(u_i, u_j|x) = 0$ for $i \neq j$

Homoskedasticity

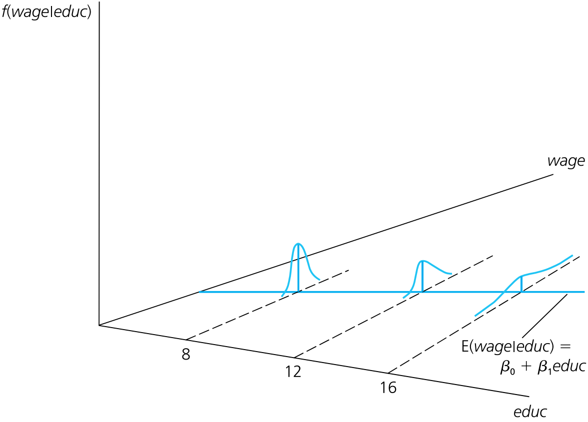
Homoskedastic errors: $\text{var}(u_i|x_i) = \sigma^2$ for all $i = 1 \ldots n$



Source: Wooldridge's *Introductory Econometrics*

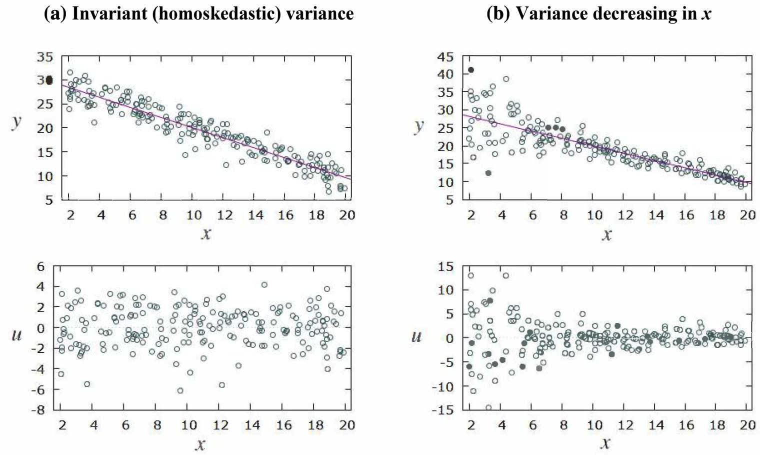
Homoskedasticity

Heteroskedastic errors: $\mathrm{var}(u_i|x_i) \neq \mathrm{var}(u_j|x_j)$ for some $i \neq j$



Source: Wooldridge's *Introductory Econometrics*

Homoskedasticity



Homoskedastic assumption $\mathbb{E}(u_i^2|x_i) = \sigma^2$ violated in (b).

Homoskedasticity

Thus:

$$\text{var}\left(\sum_{i=1}^{n}(x_i-\bar{x})u_i \,\middle|\, x\right) = \sum_{i=1}^{n}(x_i-\bar{x})^2\sigma^2 + 0$$
$$= \sigma^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Now completing the derivation:

$$\text{var}(\hat{\beta}_2|x) = \frac{1}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}\text{var}\left(\sum_{i=1}^{n}(x_i-\bar{x})u_i \,\middle|\, x\right)$$
$$= \frac{1}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}\cdot\sigma^2\sum_{i=1}^{n}(x_i-\bar{x})^2$$
$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

The final formula (if $k = 2$)

Thus, if we assume

- **Model = Population**: $y_i = \beta_1 + \beta_2 + u_i$
- **Homoskedasticity**: $\text{var}(u_i|x) = \sigma^2$ for all i
- **No autocorrelation**: $\text{cov}(u_i, u_j|x) = 0$ for $i \neq j$

The variance is

$$\boxed{\text{var}(\hat{\beta}_2|x) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}}$$

Note that σ^2 is also a population parameter that needs to be estimated.

We can use the SSR for that:

$$\boxed{\hat{\sigma}^2 = \frac{\sum_i^n \hat{u}_i^2}{n-2}}$$

The final formula (if $k = 2$)

Notice that $\sum_{i=1}^{n}(x_i - \bar{x})^2$ grows with n. This becomes explicit when substituting for $\text{var}(x) = \sum(x_i - \bar{x})^2/n$:

$$\text{var}(\hat{\beta}_2|x) = \frac{\sigma^2}{n\ \text{var}(x_1)}$$

As we saw before, **precision** will be higher (the variance lower):

- with more data ($\uparrow n$)
- with lower error variance ($\downarrow \sigma^2$)

Takeaways

Key takeaways

You should now know:

- Difference between sample and population
- What standard errors represent
- How $\text{var}(\hat{\beta}_2|x)$ depends on n and σ
- Basic variance and covariance rules
- Homo- vs. heteroskedasticity
- How to derive $\text{var}(\hat{\beta}_2|x)$ in the univariate case

Appendix

Variance and covariance rules

Basic Variance Rules:

1. $\text{Var}(a) = 0 \quad \text{(constant } a)$
2. $\text{Var}(aX + b) = a^2\text{Var}(X)$
3. $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$

Covariance Rules:

4. $\text{Cov}(X, X) = \text{Var}(X)$
5. $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
6. $\text{Cov}(X, a) = 0 \quad \text{(constant } a)$
7. $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$
8. $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$