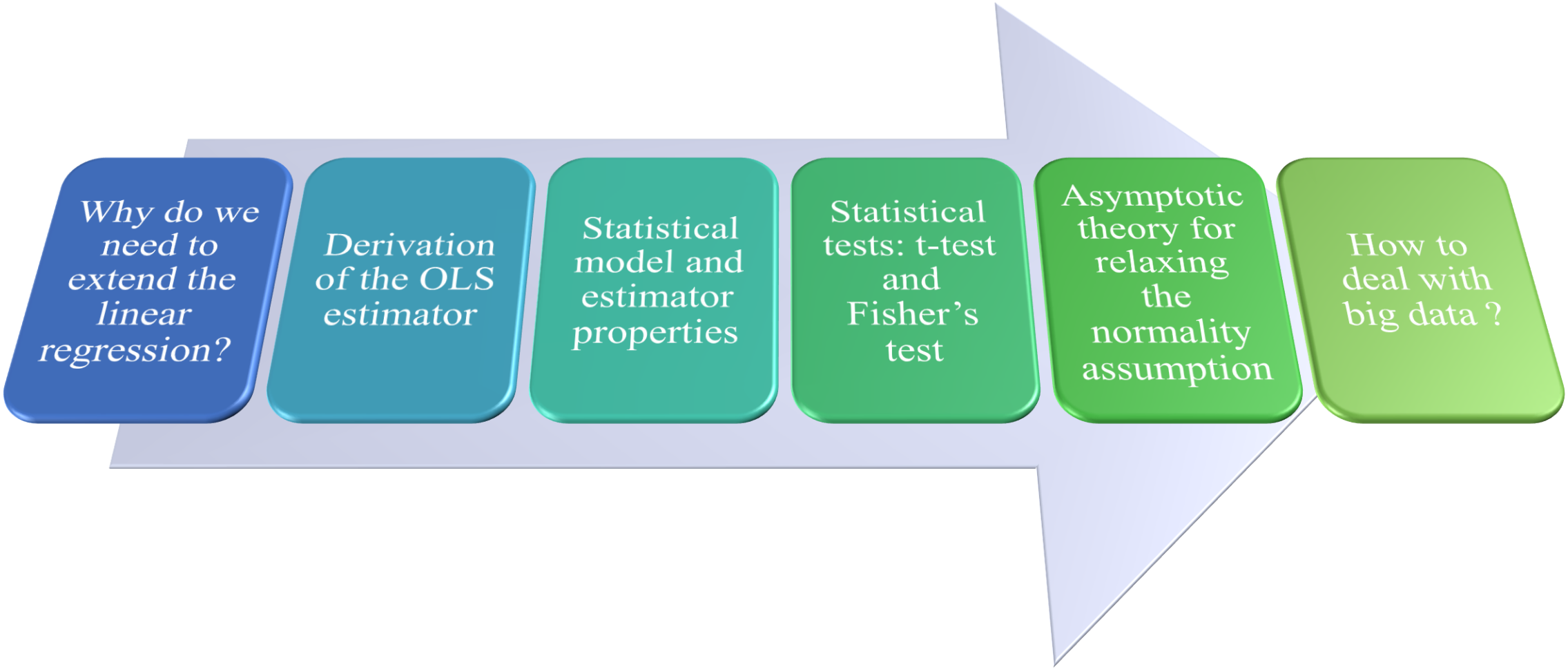
Introduction to Econometrics

Regression with multiple variables

ESSEC
BUSINESS SCHOOL

Outline of this course



Linear regression with multiple explanatory variables

Introduction

Why more variables ?

Source: Piff, Paul K., et al. "Having less, giving more: the influence of social class on prosocial behavior." *Journal of personality and social psychology* 99.5 (2010): 771.

Other studies on the topic

Study: 124 undergraduate students.

Dependent variable: How much they give in a dictator game (from 0 to 10).

Explanatory variable: In which social class they subjectively think they belong (from 1 to 10).

Simple linear regression

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t$$

$$\hat{\beta}_2 = -0.23$$

Statistical test:

Student test: $t = -2.52$ p-value: $p = 0.01$

Every time we claim the ladder of success by one unit, we give on average 0,23 token less

Is that enough ?
Did we prove that upper classes are more selfish ?

Multiple linear regression

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 z_{1t} + \beta_4 z_{2t} + \ldots + \epsilon_t$$

$$\hat{\beta}_2 = -0.22$$

Control variables:
Age, ethnicity, Religiosity

Statistical test:

Student test: $t = -2.10$ p-value: $p < 0.05$

Result remains valid when we control for many other variables (confounding variables)

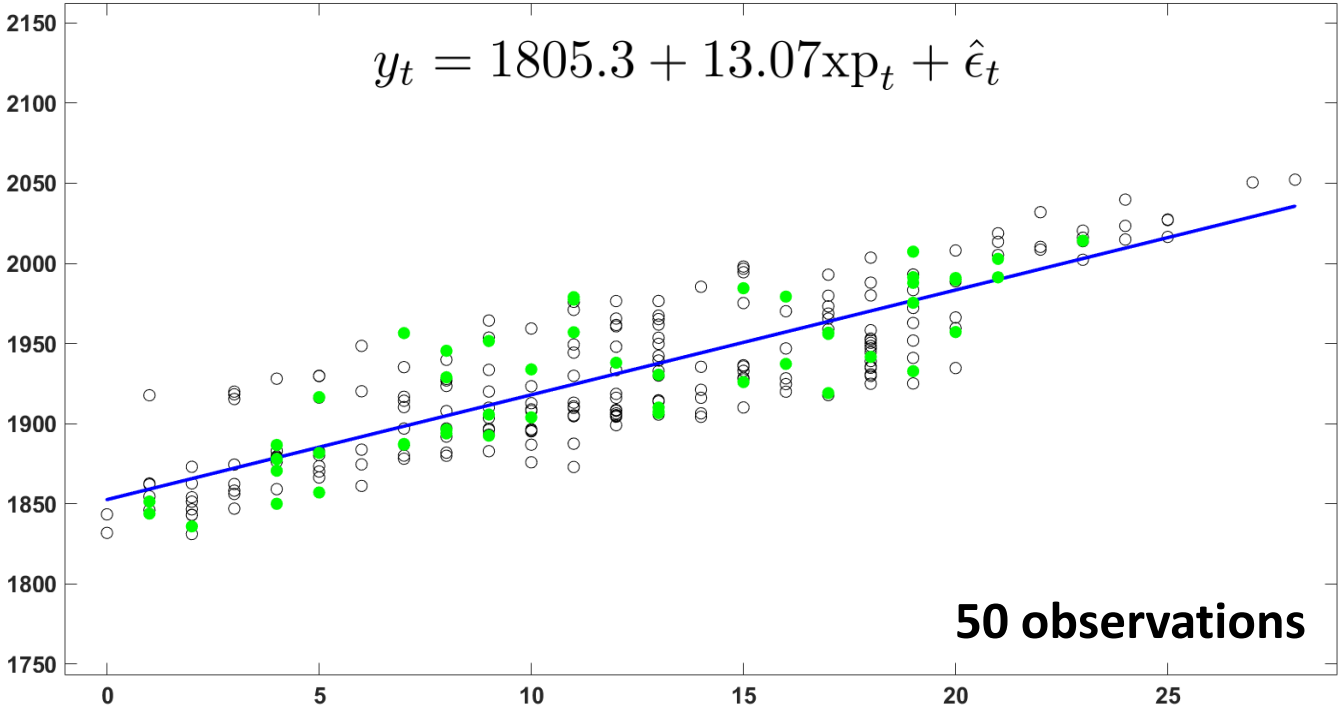
Why more variables ?

Salary-Experience example: Company of 250 employees.

Dependent variable: gross salary per month.

Explanatory variable: Experience per year.

True coefficient: $\beta_2 = 10$



Statistical test: $H_0 : \beta_2 = 10$ vs $H_1 : \beta_2 \neq 10$

t-stat: $t = 2.55$

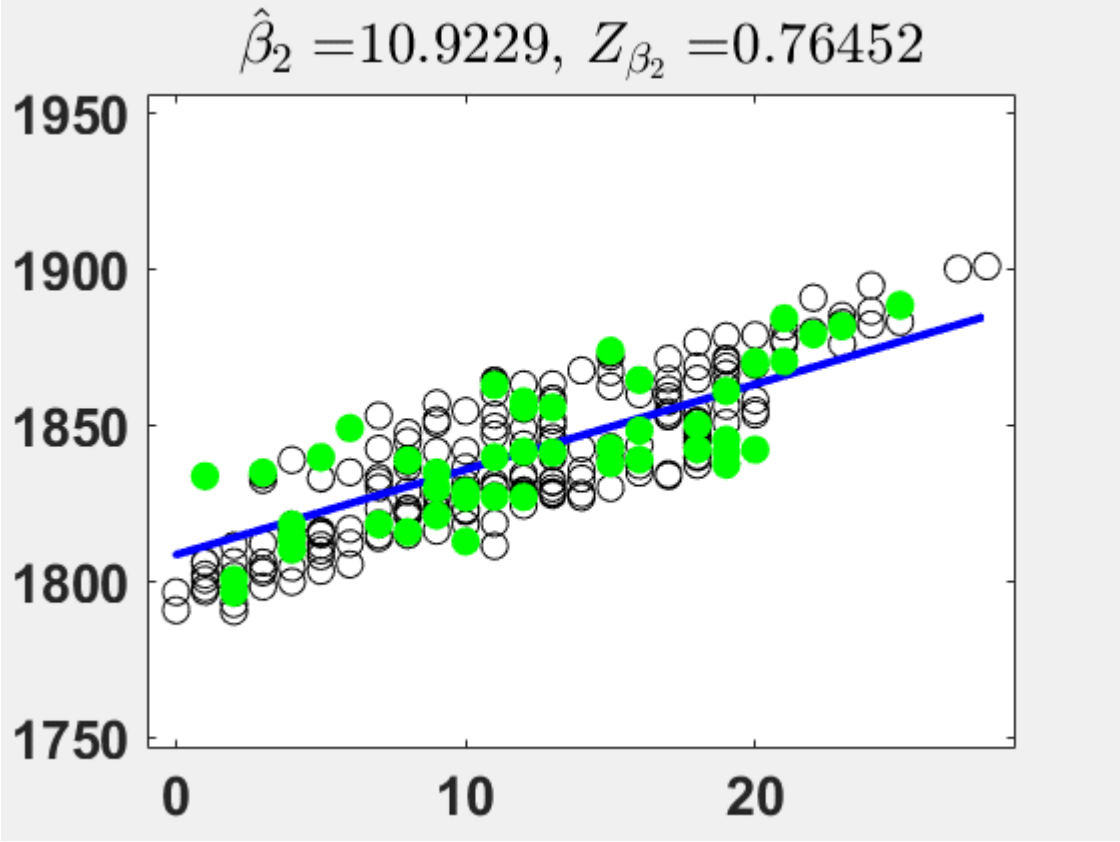
Confidence interval at 95% ($\alpha = 0.05$): [10.65,15.5]

We reject the Null.

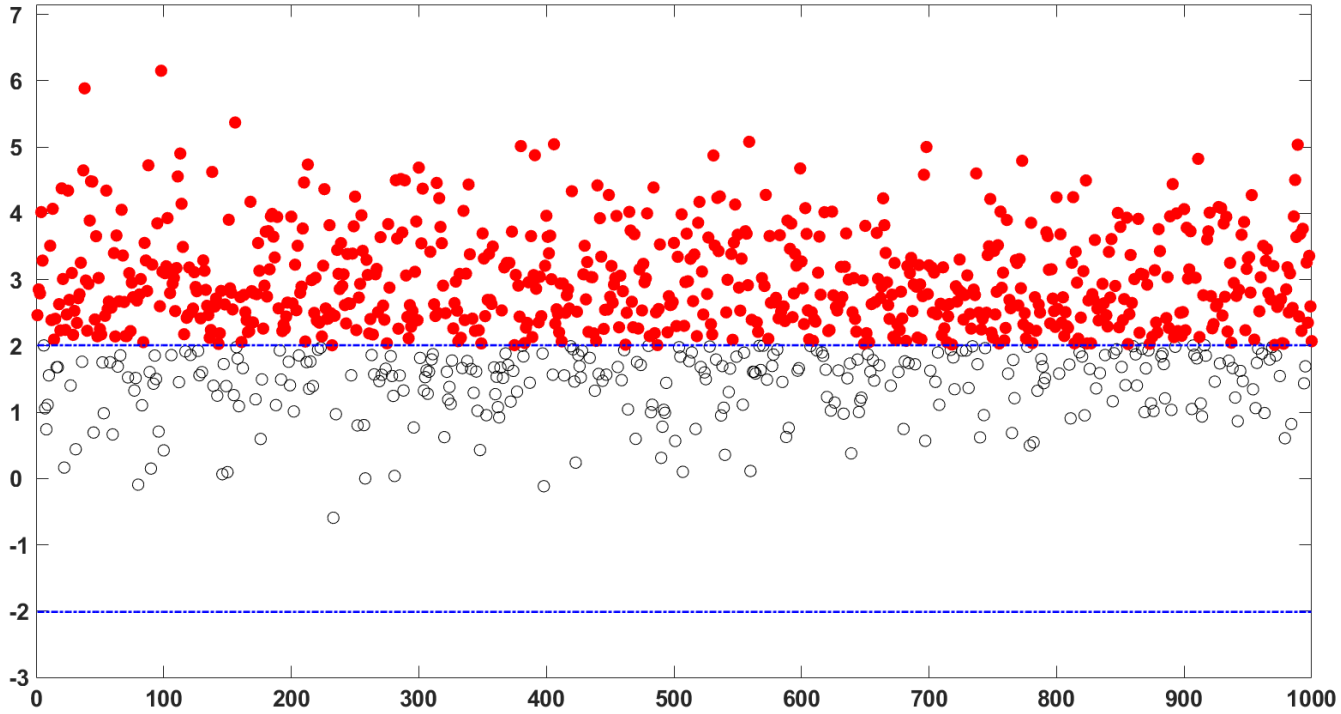
Why more variables ?

Salary-Experience example:

True coefficient: $\beta_2 = 10$



T-test rejected 696 out of 1000



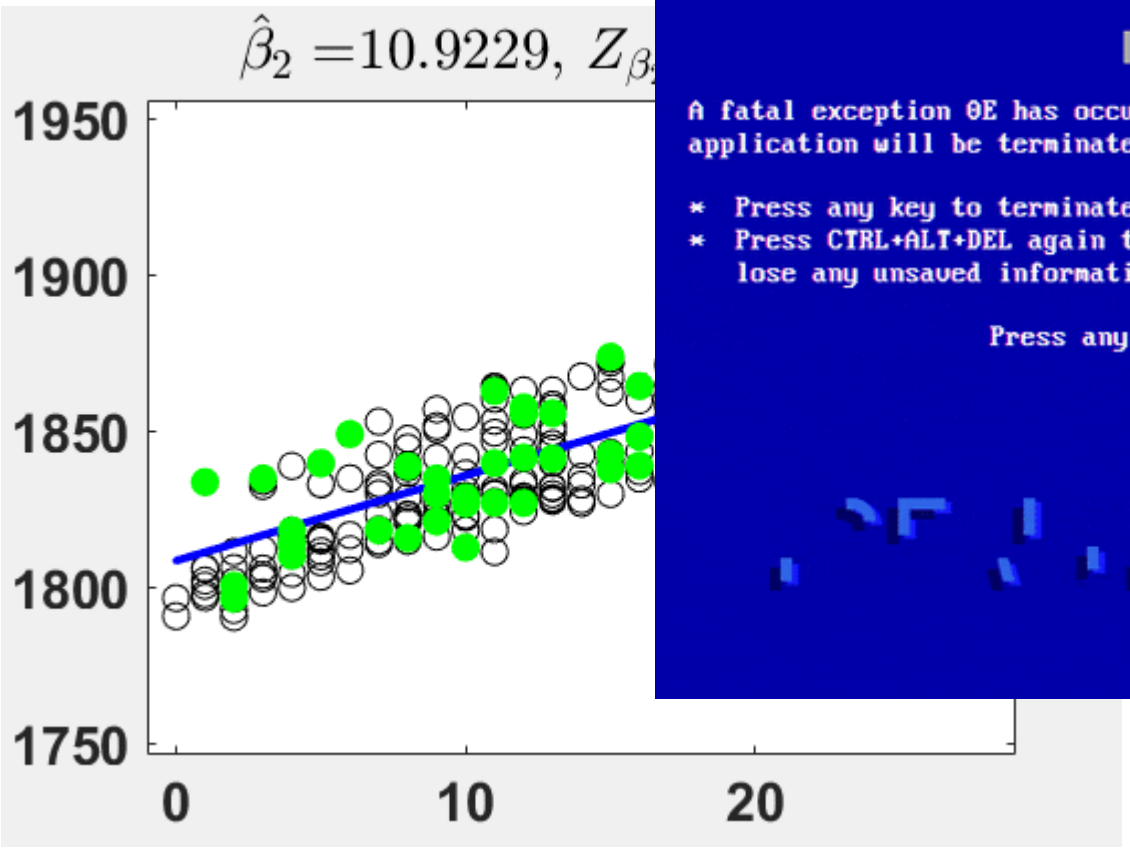
Typically overestimates the true coefficient

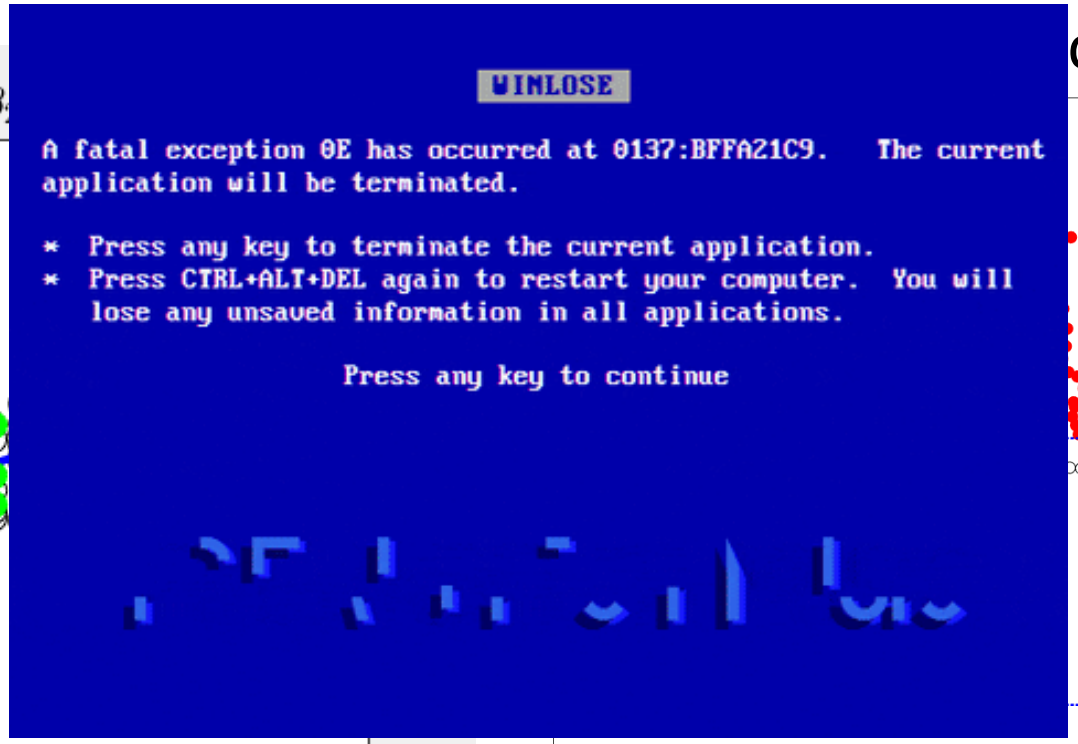
Reject too often the test statistics

Why more variables ?

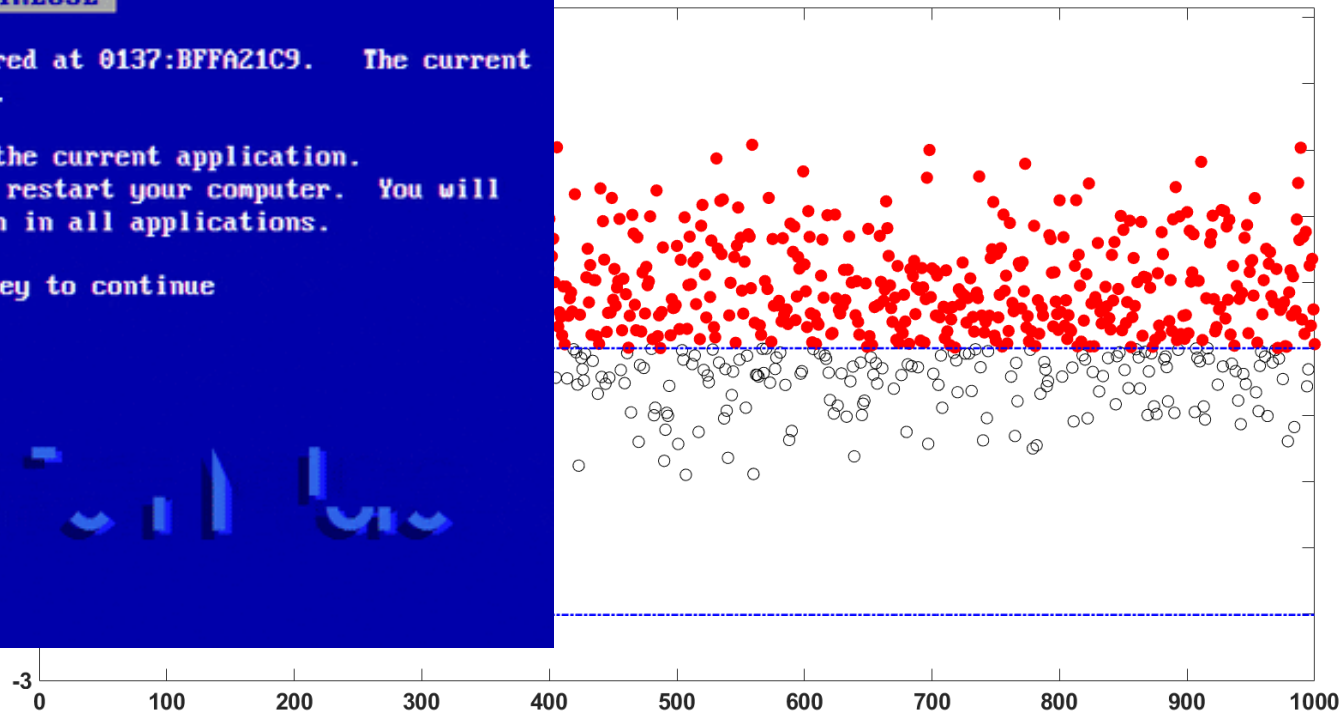
Salary-Experience example:

True coefficient: $\beta_2 = 10$





cted 696 out of 1000



Typically overestimates the true coefficient

Reject too often the test statistics

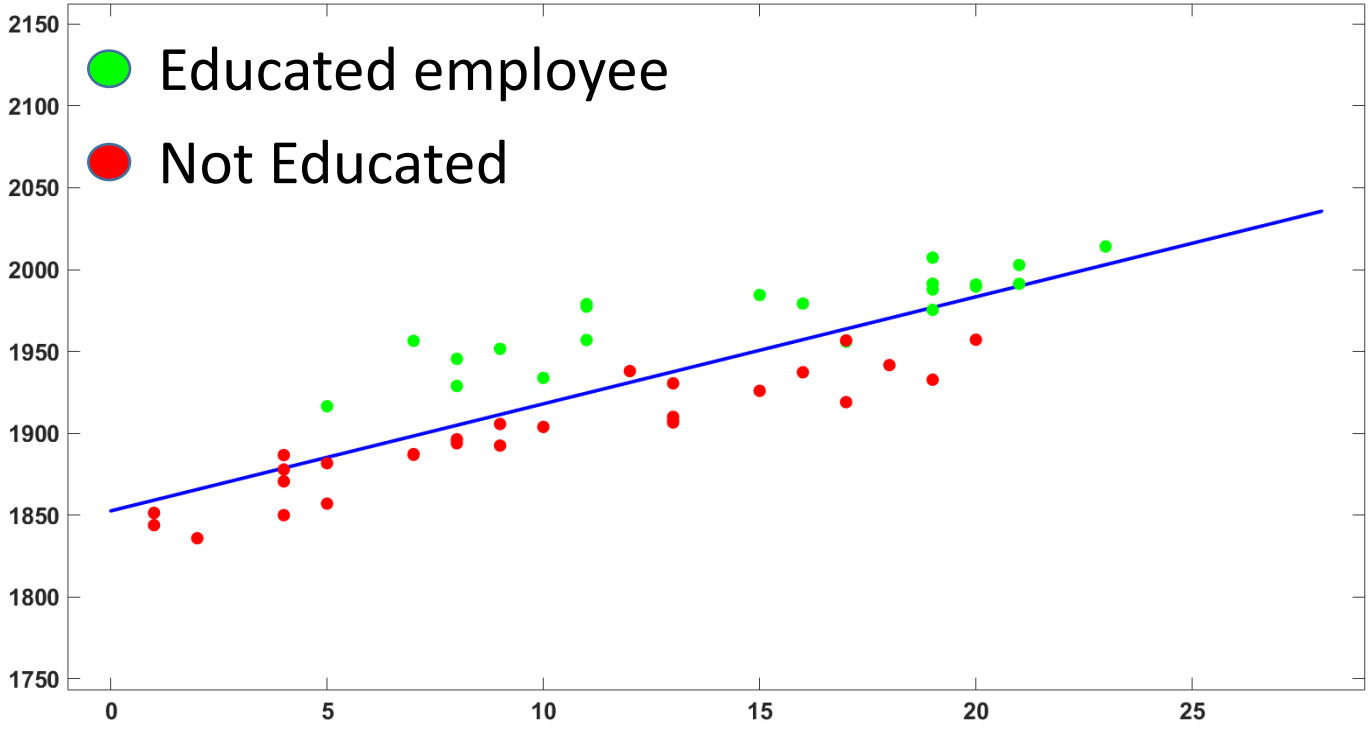
Why more variables ?

Biased estimates:

Strict exogeneity: Errors have zero expectations conditional on all the explanatory variables.

Additional expl. var. correlated with experience!

$$\text{True regression: } y_t = 1800 + 10\text{xp}_t + 100\text{educ}_t + \epsilon_t$$



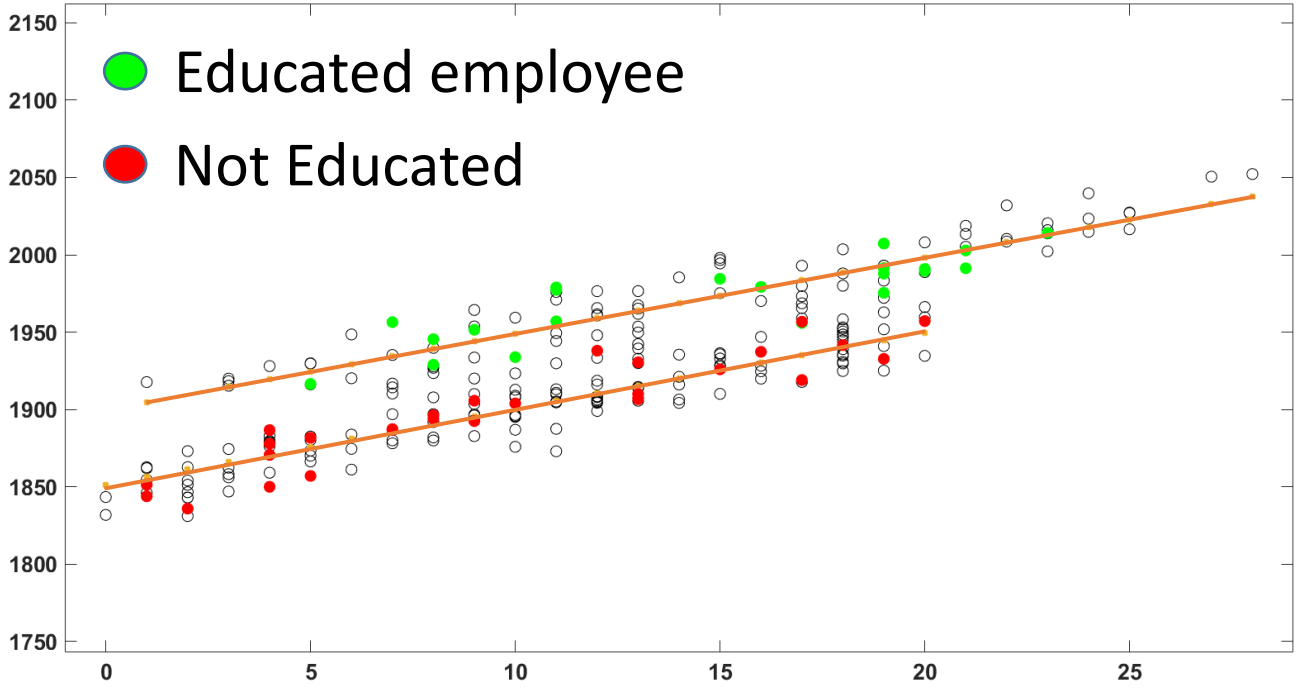
Education increases the wage
Education increases the experience

Omitted variable bias

Including Education

Omitted variable: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

$$y_t = 1802.84 + 9.83\text{xp}_t + 96.95\text{Educ}_t$$



Statistical test: $H_0 : \beta_2 = 10$ vs $H_1 : \beta_2 \neq 10$

t-stat: $t = -0.28$

We do not reject the Null.

Take Away

Omitted variable can create a bias.
Think about which variable could impact on X and y simultaneously

Linear regression

If true regression was:

For Educated employees: $y_t = \beta_1 + 30\text{xp}_t + \epsilon_t$

For other employees: $y_t = \beta_1 + 10\text{xp}_t + \epsilon_t$

→ True regression is not linear: $\begin{cases} y_t = \beta_1 + \beta_2 \text{xp}_t + \epsilon_t \text{ if } \text{Educ}_t = 0 \\ y_t = \beta_1 + \tilde{\beta}_2 \text{xp}_t + \epsilon_t \text{ if } \text{Educ}_t = 1 \end{cases}$

Including Education can make the regression linear:

→ $y_t = \beta_1 + \beta_2 \text{xp}_t + \underbrace{(\tilde{\beta}_2 - \beta_2)}_{\beta_3} \text{xp}_t \text{Educ}_t + \epsilon_t$

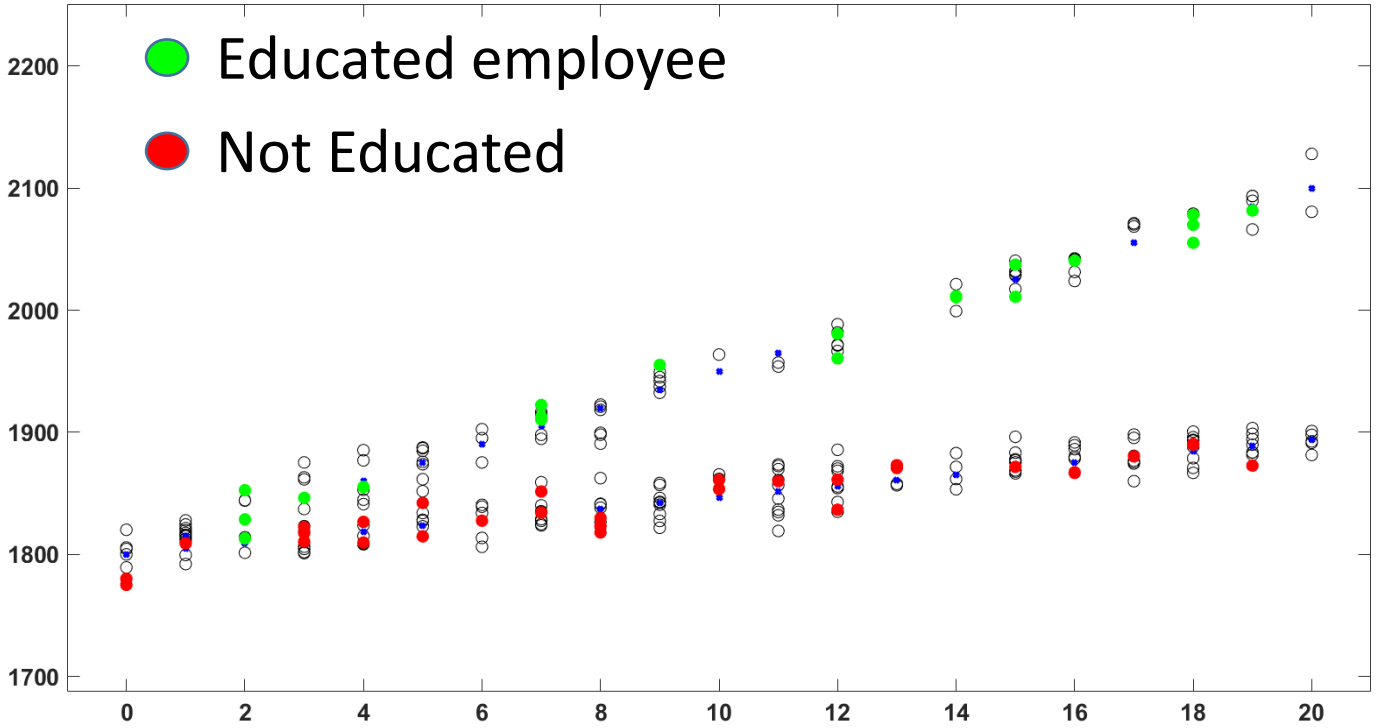
Including Education

For Educated employees: $y_t = \beta_1 + 30\text{xp}_t + \epsilon_t$

For other employees: $y_t = \beta_1 + 10\text{xp}_t + \epsilon_t$

Not linear w.r.t. Educ: $y_t = \beta_1 + \beta_2\text{xp}_t + \beta_3\text{xp}_t\text{Educ}_t + \epsilon_t$

$$y_t = 1800.01 + 9.36\text{xp}_t + 20.65\text{xp}_t\text{Educ}_t$$



Effect of Experience:

- 30.01 when $\text{Educ} = 1$
- 9.36 when $\text{Educ} = 0$

Dummy variables (equal to 0 or 1) can make linear a lot of complex regression.

Linear regression with multiple explanatory variables
Mathematical framework

OLS estimator

Advanced linear regression

6Blue1Brown (1')

- Linear regression: $y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + \epsilon_t.$
- Matrix form: $y_t = x_t'\beta + \epsilon_t,$

$$y = X\beta + \epsilon.$$

Notation : $x_t = \underset{(K \times 1)}{\begin{pmatrix} 1 \\ x_{t,2} \\ \ldots \\ x_{t,K} \end{pmatrix}}$, $\beta = \underset{(K \times 1)}{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_K \end{pmatrix}}$, $X = \underset{(T \times K)}{\begin{pmatrix} x_1' \\ x_2' \\ \ldots \\ x_T' \end{pmatrix}} = \begin{pmatrix} 1 & \ldots & x_{1,K} \\ 1 & \ldots & x_{2,K} \\ \ldots & & \\ 1 & \ldots & x_{T,K} \end{pmatrix}$, $y = \underset{(T \times 1)}{\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_T \end{pmatrix}}$, $\epsilon = \underset{(T \times 1)}{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \ldots \\ \epsilon_T \end{pmatrix}}$

Advanced linear regression

- Linear regression: $y = X\beta + \epsilon.$

OLS Criterion: $\hat{\beta} = \text{Argmin}_{\beta}\epsilon'\epsilon = \sum_{t=1}^{T}\epsilon_t^2$

Analytical solution: $\hat{\beta} = (X'X)^{-1}X'y$

Reminder

$$\frac{d(a'\beta)}{d\beta} = a$$

$$\frac{d(\beta'A\beta)}{d\beta} = 2A\beta$$

Proof:

Advanced linear regression

Linear regression: $y = X\beta + \epsilon.$

OLS estimator: $\hat{\beta} = (X'X)^{-1}X'y$

No regressor: $y_t = \beta_1 + \epsilon_t \Rightarrow \hat{\beta}_1 = \bar{y}$

One regressor: $y_t = \beta_1 + \beta_2 x_{t,1} + \epsilon_t \Rightarrow$

$$\begin{cases} \hat{\beta}_2 = \dfrac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \\ \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}. \end{cases}$$

General estimator leads to the same estimators

Fitting criterion

Linear regression: $y = X\beta + \epsilon.$

Coefficient of determination: $R^2 = 1 - \frac{\sum_{t=1}^{T} \hat{\epsilon}_t^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$

- Coefficient of determination between 0 and 1.

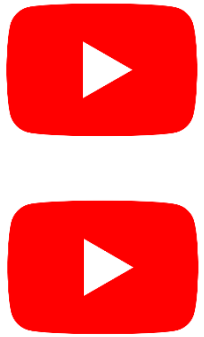
By how much do we improve the fit compared to the sample average ?

- **Fitted values:** $\hat{y} = X\hat{\beta}$

Coefficient of determination is still equal to the squared empirical correlation

$$R^2 = \rho^2_{\hat{y}y}$$

Coefficient of determination is not equal anymore to the squared empirical correlation with one particular explanatory variable.



Linear regression with multiple explanatory variables

Statistical framework

OLS properties

Statistical model

Linear regression: $y = X\beta + \epsilon.$

Assumptions for the generalized regression:

1. **Linear regression:** The variables are linearly related.
2. **No Collinearity:** Rank of matrix X is K.
3. **Strict exogeneity:** Errors have zero expectations conditional on all the explanatory variables.
4. **White noise:** No linear dependence between the error terms.
5. **Homoskedasticity:** The variance of the error term is constant.
6. **Distribution:** Error term is normally distributed.

➡ Collinearity assumption is slightly different.

Statistical model

Linear regression: $y = X\beta + \epsilon.$

Assumptions for the simple regression: $y_t = \beta_1 + \beta_2 x_{t,1} + \epsilon_t$

Collinearity: The explanatory variable is not constant.

If does not hold: $\hat{\beta}_2 = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{0}$

Assumptions for the generalized regression: $y = X\beta + \epsilon.$

Collinearity: Rank of matrix X is K.

OLS estimator: $\hat{\beta} = (X'X)^{-1}X'y$

Reminder

$$(X'X)^{-1} = \frac{\text{Adjugate}(X'X)}{|X'X|}$$

If does not hold: $|X'X| = 0$

$$\Rightarrow \hat{\beta} = \frac{\text{Adjugate}(X'X)}{0}X'y$$

Statistical model

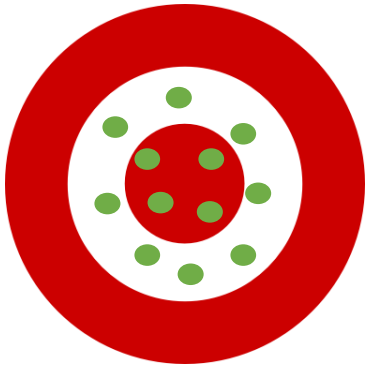
Linear regression: $y = X\beta + \epsilon.$

OLS estimator: $\hat{\beta} = (X'X)^{-1}X'y$

Useful for OLS properties: $\hat{\beta} = (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t y_t)$

- **Unbiasedness** (assumptions 1 to 3)
- OLS estimator is **Best Linear Unbiased Estimator** (assumptions 1 to 5)
- OLS estimator is **Best Unbiased Estimator** (assumptions 1 to 6)

Unbiasedness



On average, estimates = true coefficient

$$E(\hat{\beta}|X) = \beta$$

Proof:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'(X\beta + \epsilon), \\ &= \beta + (X'X)^{-1}X'\epsilon. \end{aligned}$$

$$\begin{aligned} E(\hat{\beta}|X) &= \beta + E((X'X)^{-1}X'\epsilon|X), \\ &= \beta. \end{aligned}$$

Statistical model

Linear regression: $y = X\beta + \epsilon.$

OLS estimator:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'\epsilon, \\ &= \beta + (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t)\end{aligned}$$

Consistency

Asymptotically, estimator = coefficient.

Proof (simplified):

Law of large number:

$$\frac{1}{T}(\sum_{t=1}^{T} x_t \epsilon) \quad \rightarrow \quad E(x_t \epsilon_t) = 0 \text{ Strict exogeneity}$$

Variance of the estimator

Linear regression: $y = X\beta + \epsilon.$

OLS estimator: $$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'\epsilon.\end{aligned}$$

- **Assumptions 4 and 5: Spherical assumptions**

White noise: No linear dependence between the error terms: $\text{Cov}(\epsilon_t, \epsilon_j|X) = 0$

Homoskedasticity: The variance of the error term is constant: $V(\epsilon_t|X) = \sigma^2$

Matrix developments:

$$\begin{aligned}\text{V}(\epsilon|X) &= E(\epsilon\epsilon'|X) - E(\epsilon|X)E(\epsilon|X)', \\ &= E(\epsilon\epsilon'|X)\end{aligned}$$

$$E(\epsilon\epsilon'|X) = \begin{pmatrix} E(\epsilon_1^2|X) & E(\epsilon_1\epsilon_2|X) & \dots & E(\epsilon_1\epsilon_T|X) \\ E(\epsilon_2\epsilon_1|X) & E(\epsilon_2^2|X) & \dots & E(\epsilon_2\epsilon_T|X) \\ & & \dots & \\ E(\epsilon_T\epsilon_1|X) & E(\epsilon_T\epsilon_2|X) & \dots & E(\epsilon_T^2|X) \end{pmatrix} = \sigma^2 I_T$$

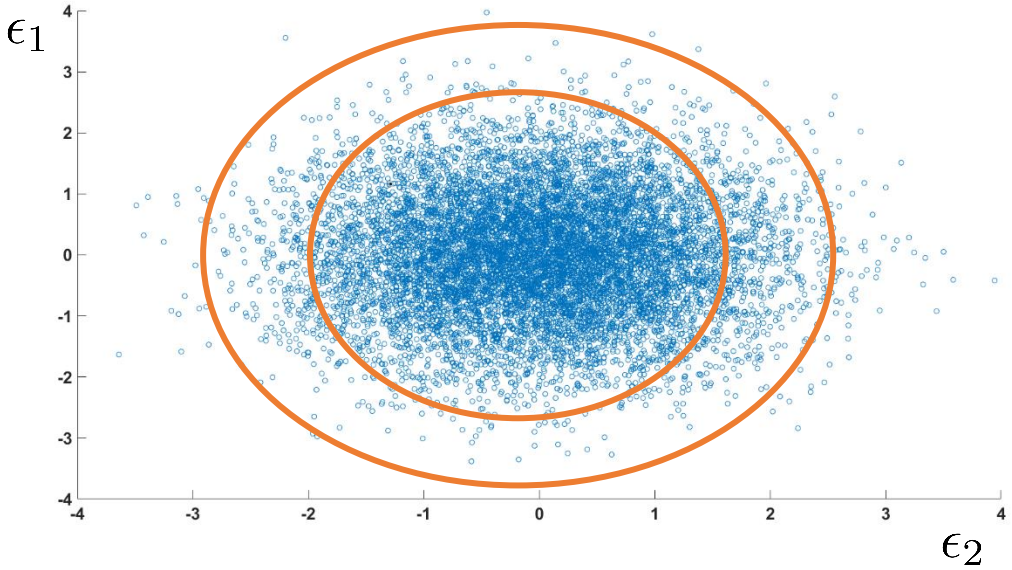
Why spherical ?

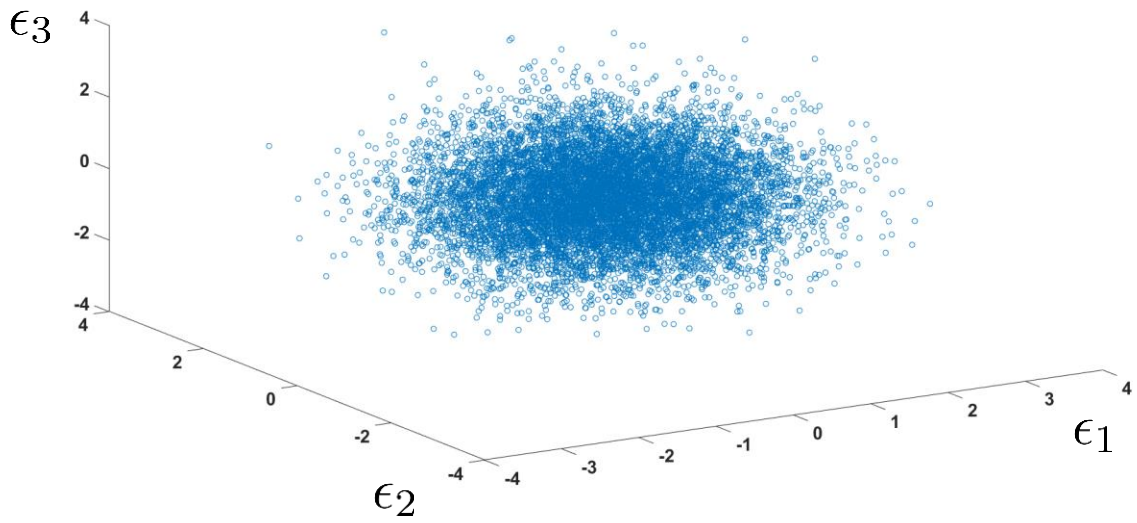
Linear regression: $y = X\beta + \epsilon.$

White noise: No linear dependence between the error terms: $\mathrm{Cov}(\epsilon_t, \epsilon_j | X) = 0$

Homoskedasticity: The variance of the error term is constant: $V(\epsilon_t^2 | X) = \sigma^2$

➡ When error terms are bell-shaped





Ball of dimension T

Variance of the estimator

Linear regression: $y = X\beta + \epsilon.$

Assumptions 3 to 5: $E(\epsilon\epsilon'|X) = \sigma^2 I_T$

OLS estimator:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= \beta + (X'X)^{-1}X'\epsilon.\end{aligned}$$

Estimator variance:

$$\begin{aligned}\mathrm{V}(\hat{\beta}|X) &= \mathrm{V}(\beta + (X'X)^{-1}X'\epsilon|X), \\ &= \mathrm{V}((X'X)^{-1}X'\epsilon|X), \\ &= \sigma^2(X'X)^{-1}\end{aligned}$$

6. Distribution: Error term is normally distributed.

Estimator is a linear function of the error terms → **Normally distributed!**

$$\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1})$$

Multivariate Normal distribution

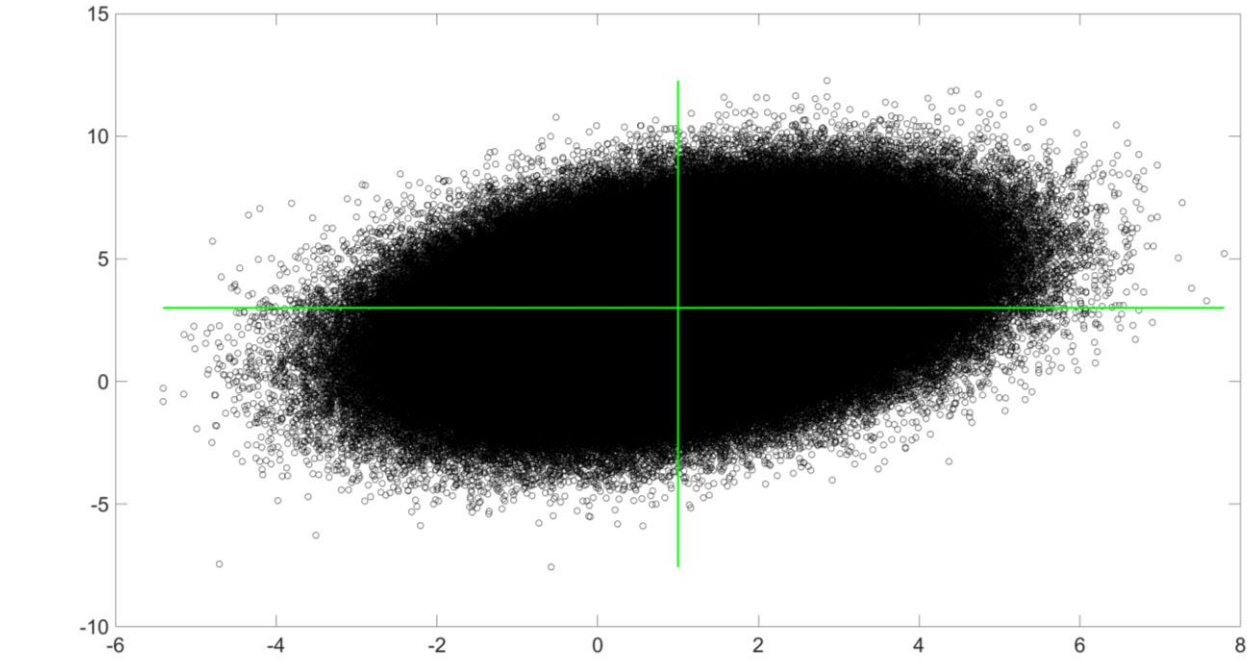
Joint distribution:

$$\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1})$$

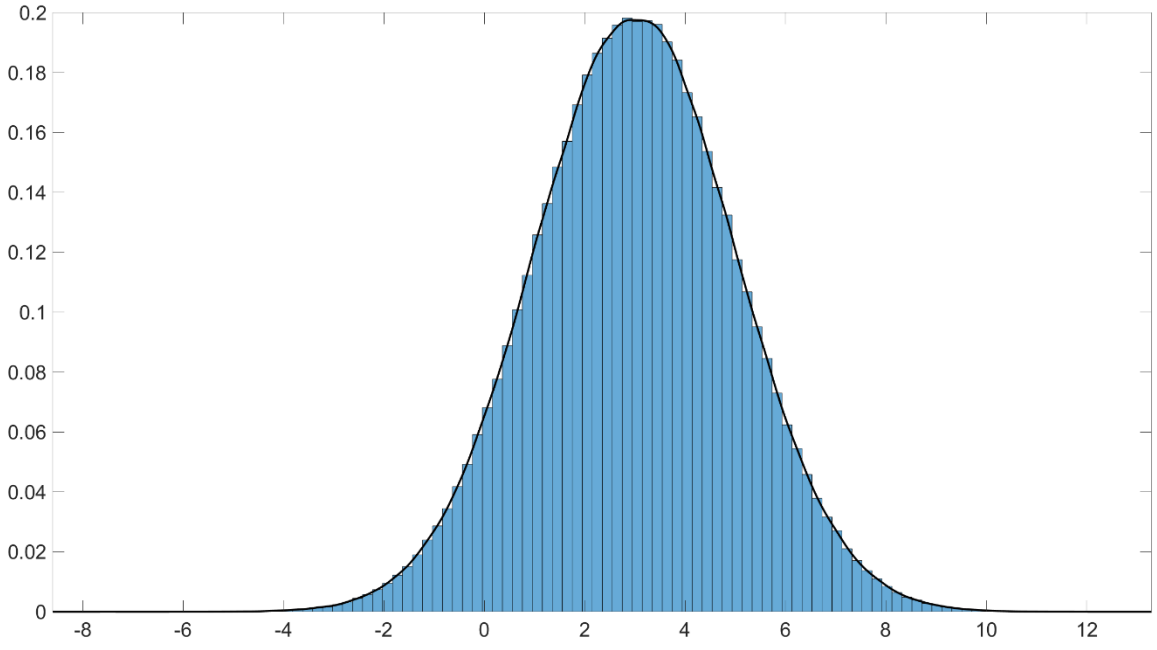
Marginal distribution:

$$\hat{\beta}_2|X \sim N(\beta_2, \sigma^2((X'X)^{-1})_{22})$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \right)$$



$$\hat{\beta}_2 \sim N(3, 4)$$



Linear regression with multiple explanatory variables

Statistical framework

t-test and Fisher's test

Statistical inference

Linear regression: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

Hypothesis tests:

Two-sided test: $H_0 : \beta_2 = 10$ vs $H_1 : \beta_2 \neq 10$

One-sided test: $H_0 : \beta_2 = 10$ vs $H_1 : \beta_2 > 10$ or $H_0 : \beta_2 = 10$ vs $H_1 : \beta_2 < 10$

Notation:

$$\left((X'X)^{-1}\right)_{22} = \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{pmatrix}$$

Joint distribution:

$$\hat{\beta}|X \quad \sim \quad N(\beta, \sigma^2(X'X)^{-1})$$

Marginal distribution:

$$\hat{\beta}_2|X \quad \sim \quad N(\beta_2, \sigma^2((X'X)^{-1})_{22})$$

Under the Null: $$\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\sigma^2((X'X)^{-1})_{22}}} \quad \sim \quad N(0,1)$$

Statistical inference

Distribution: $\frac{\hat{\beta}_2 - \beta_2}{\sqrt{\sigma^2((X'X)^{-1})_{22}}} \sim N(0,1)$

Linear regression: $y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + \epsilon_t.$

Variance estimation: $\hat{\sigma}^2$?

Intuition: $\hat{\sigma}^2 = \sum_{t=1}^{T} \frac{\hat{\epsilon}_t^2}{T-K}$ ✓

Properties:

Unbiased estimator: $E(\hat{\sigma}^2) = \sigma^2$

Consistent estimator: $\hat{\sigma}^2 \to \sigma^2$

Other popular estimator: $\hat{\sigma}^2 = \sum_{t=1}^{T} \frac{\hat{\epsilon}_t^2}{T} = \frac{SSR}{T}$

→ Biased but consistent estimator.

Linear restriction

Linear regression: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

Hypothesis test: $H_0 : \beta_2 = \beta_3$ vs $H_1 : \beta_2 \neq \beta_3$

First idea: a t-test

Under the Null: $$\frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{\sigma^2((X'X)^{-1})_{22}}} \sim N(0,1)$$

Problem: $\beta_3 \neq \hat{\beta}_3$

Estimates can over-(under-)estimate the coefficient

Restriction on the model:

$$y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t \Rightarrow y_t = \beta_1 + \beta_2 (\text{xp}_t + \text{educ}_t) + \epsilon_t$$

We need a test for comparing models

Multiple hypotheses ?

Linear regression: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

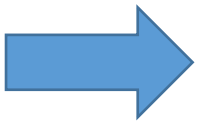
True regression: $y_t = 1800 + 10\text{xp}_t + 100\text{Educ}_t$

Hypothesis test: $H_0 : \beta_2 = 10 \text{ and } \beta_3 = 100 \text{ vs } H_1 : \beta_2 \neq 10 \text{ or/and } \beta_3 \neq 100$

First idea: Two t-tests

Under the Null: $\dfrac{\hat{\beta}_2 - 10}{\sqrt{\sigma^2((X'X)^{-1})_{22}}} \sim N(0,1)$ and $\dfrac{\hat{\beta}_3 - 100}{\sqrt{\sigma^2((X'X)^{-1})_{33}}} \sim N(0,1)$

First problem: Prob rejecting the Null $= \alpha$ Prob rejecting the Null $= \alpha$



Under the Null, reject the Null with Probability: 2α

Multiple testing problem

Second problem: Test does not take correlation into account!

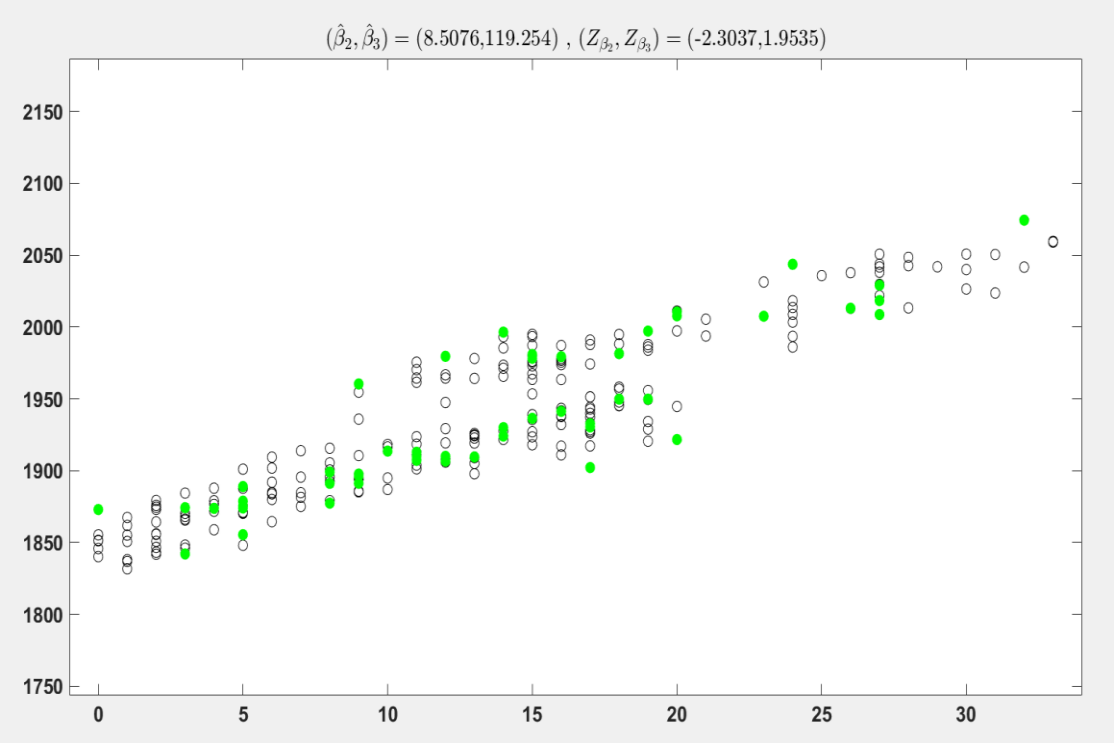
Multiple hypotheses ?

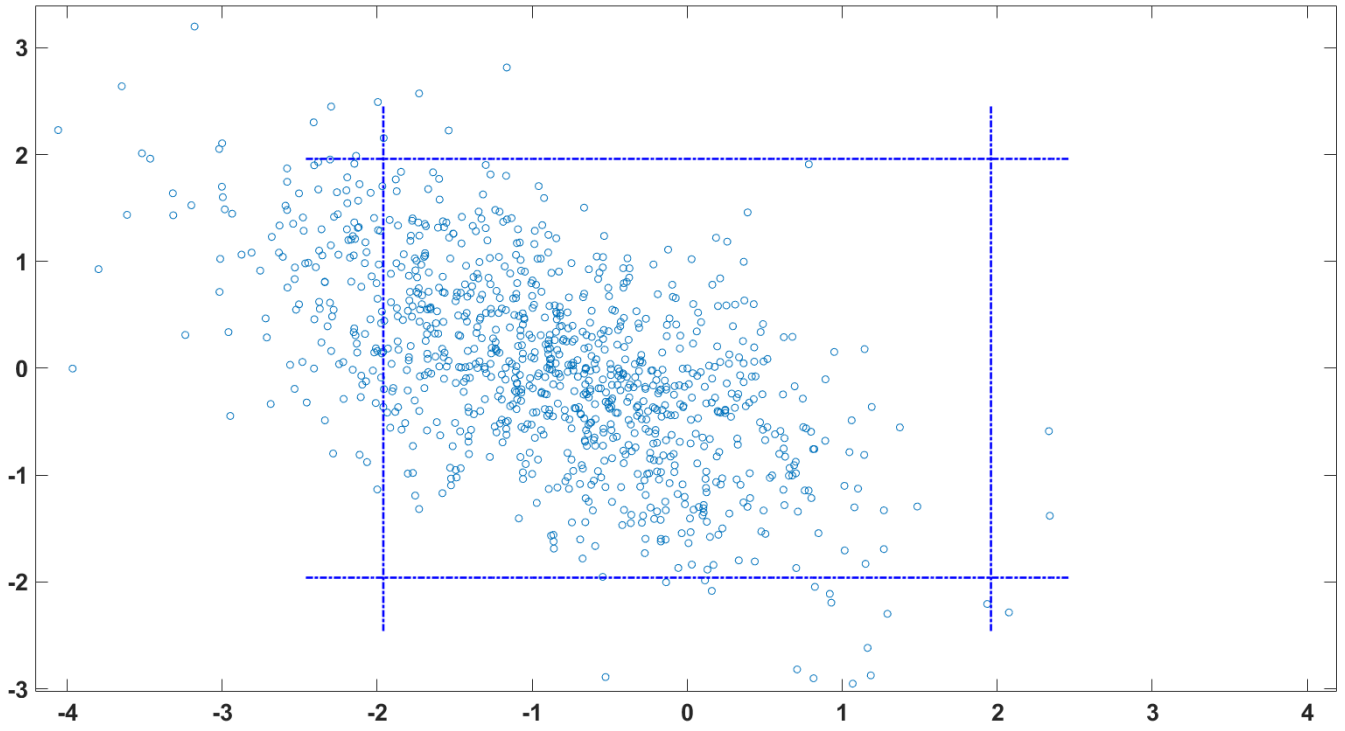
Linear regression: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

True regression: $y_t = 1800 + 10\text{xp}_t + 100\text{Educ}_t$

Hypothesis test: $H_0 : \beta_2 = 10 \text{ and } \beta_3 = 100 \text{ vs } H_1 : \beta_2 \neq 10 \text{ or/and } \beta_3 \neq 100$

Under the Null: $\dfrac{\hat{\beta}_2}{\sqrt{\sigma^2((X'X)^{-1})_{22}}} \sim N(0,1)$ and $\dfrac{\hat{\beta}_3}{\sqrt{\sigma^2((X'X)^{-1})_{33}}} \sim N(0,1)$





If β_2 is over-estimated, β_3 is likely to be under-estimated.

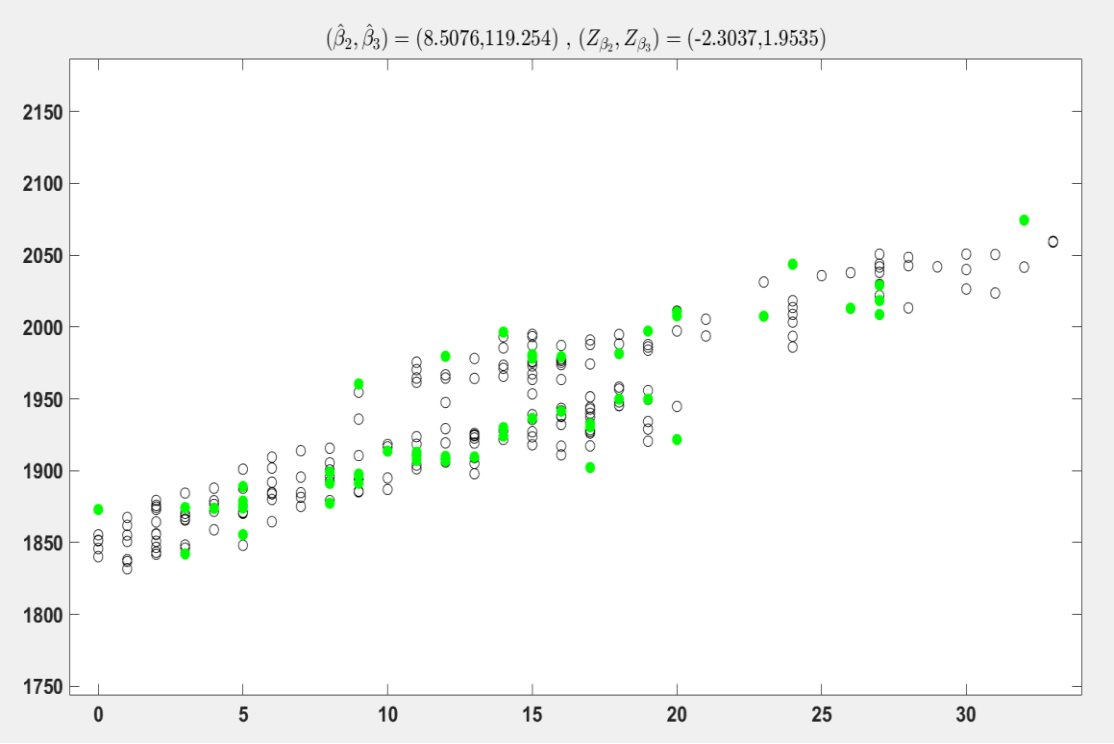
Multiple hypotheses ?

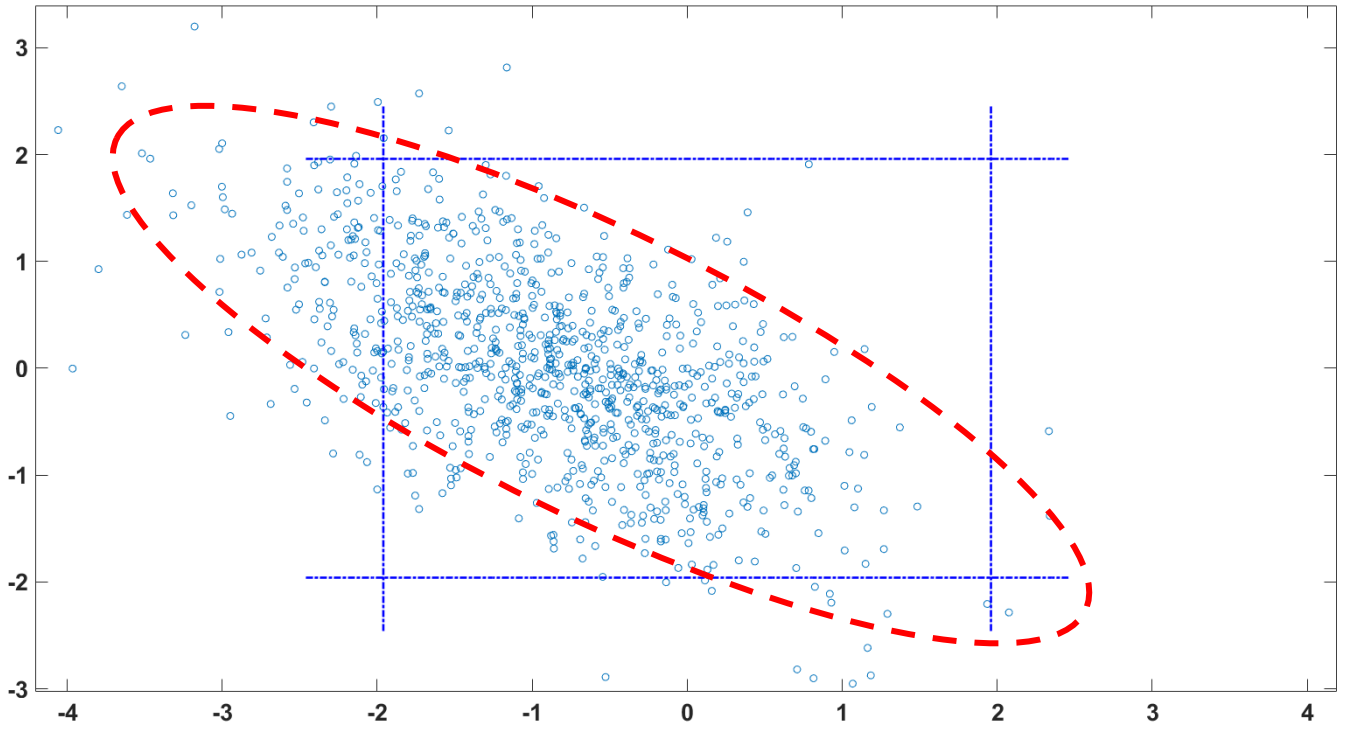
Linear regression: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

True regression: $y_t = 1800 + 10\text{xp}_t + 100\text{Educ}_t$

Hypothesis test: $H_0 : \beta_2 = 10 \text{ and } \beta_3 = 100 \text{ vs } H_1 : \beta_2 \neq 10 \text{ or/and } \beta_3 \neq 100$

We need a test which corrects for joint over-(under-) estimations





Multiple hypotheses ?

Linear regression: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

1. We need a test which corrects for joint over-(under-) estimations
2. We need a test for comparing models

F-test or Fisher-test

Test for models with linear restrictions:

Restricted model: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$ with linear restrictions on β.

Unrestricted model: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

Example: $H_0 : \beta_2 = 10$ and $\beta_3 = 100$ vs $H_1 : \beta_2 \neq 10$ or/and $\beta_3 \neq 100$

Restricted model: $y_t = \beta_1 + 10\text{xp}_t + 100\text{educ}_t + \epsilon_t$

Unrestricted model: $y_t = \beta_1 + \beta_2 \text{xp}_t + \beta_3 \text{educ}_t + \epsilon_t$

} # number constraint: $K^* = 2$

Estimation of the restricted model: $\underbrace{y_t - 10\text{xp}_t - 100\text{educ}_t}_{\tilde{y}_t} = \beta_1 + \epsilon_t$

Linear constraints

Notation: $\tilde{y}_t = y_t - 10\text{xp}_t - 100\text{educ}_t$

Example: $H_0 : \beta_2 = 10$ and $\beta_3 = 100$ vs $H_1 : \beta_2 \neq 10$ or/and $\beta_3 \neq 100$

Restricted model:

$$y_t = \beta_1 + 10\text{xp}_t + 100\text{educ}_t + \epsilon_t$$

Unrestricted model:

$$y_t = \beta_1 + \beta_2\text{xp}_t + \beta_3\text{educ}_t + \epsilon_t$$

Criterion for finding the estimates

$$\text{SSR}_R = \min_{\beta_1}[\sum_{t=1}^{T}\epsilon_t^2],$$

$$= \sum_{t=1}^{T}(\tilde{y}_t - 10\text{xp}_t - 100\text{educ}_t)^2$$

$$\text{SSR}_U = \min_{\beta_1,\beta_2,\beta_3}[\sum_{t=1}^{T}\epsilon_t^2].$$

Property: $SSR_R \geq SSR_U$ ⇒ Equality occurs when the constraints are exact.

Intuition: When $SSR_R - SSR_U \approx 0$ then do not reject the Null.

⇒ $SSR_R - SSR_U$ is sensitive to the scale of y.

Appropriate test: $\frac{SSR_R - SSR_U}{SSR_U}$ to delete the scaling problem.

Linear constraints

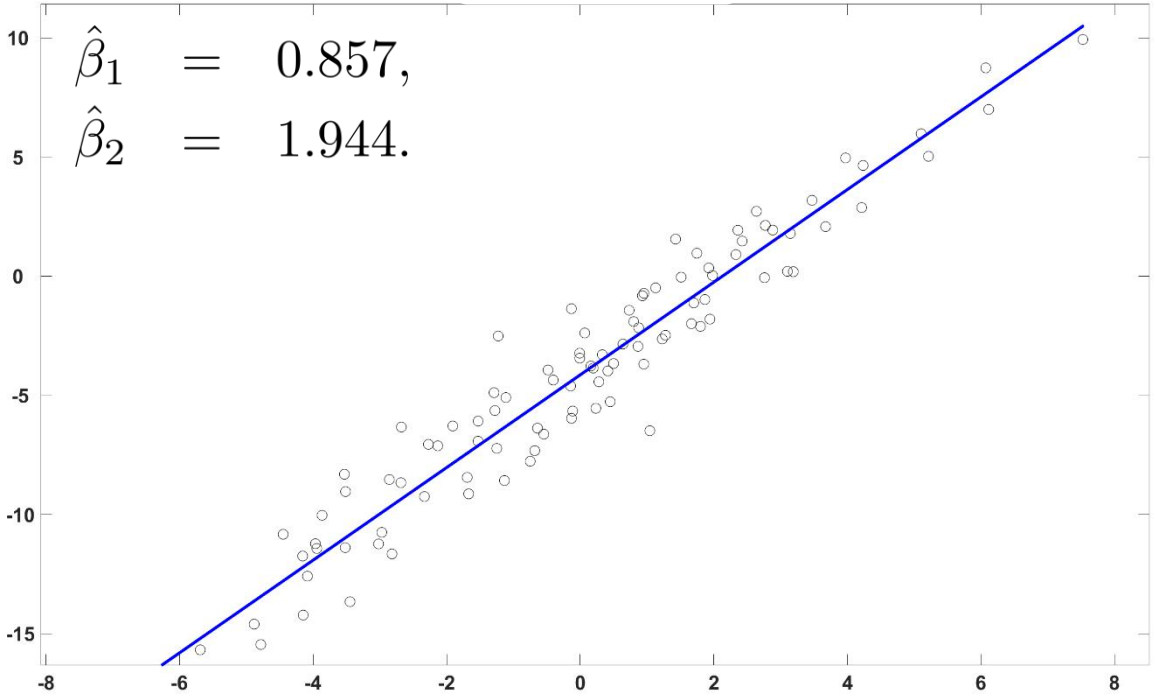
$SSR_R - SSR_U$ is sensitive to the scale of y.

Linear model: $y = X\beta + \epsilon,$

scale of k

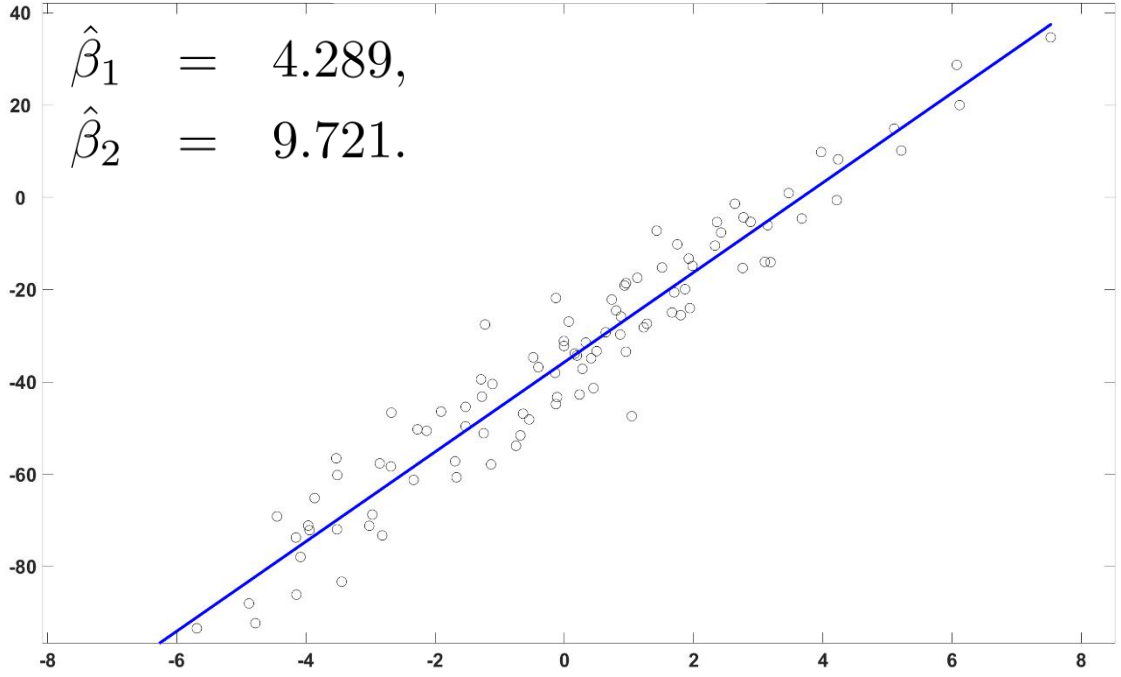
$$\begin{aligned} ky &= k(X\beta) + (k\epsilon), \\ ky &= X\tilde{\beta} + \tilde{\epsilon}. \end{aligned}$$

$\text{SSR} = 211.09$



$k = 5$

$\text{SSR} = 5277.25 = k^2 211.09$



Fisher distribution

$K^* =$ Number of constraints.

F-test or Fisher-test: $\frac{SSR_R - SSR_U}{SSR_U}\frac{T-K}{K^*} \sim F(K^*, T-K)$

Example: $H_0 : \beta_2 = 10$ and $\beta_3 = 100$ vs $H_1 : \beta_2 \neq 10$ or/and $\beta_3 \neq 100$

Restricted model:

$y_t = \beta_1 + 10\text{xp}_t + 100\text{educ}_t + \epsilon_t$

$SSR_R = \sum_{t=1}^{T}(y_t - \bar{\bar{y}}_t - 10\text{xp}_t - 100\text{educ}_t)^2$

$SSR_R = 234.8$

Unrestricted model:

$y_t = \beta_1 + \beta_2\text{xp}_t + \beta_3\text{educ}_t + \epsilon_t$

$\text{SSR}_U = \min_{\beta_1,\beta_2,\beta_3}[\sum_{t=1}^{T}\epsilon_t^2] = 210.4$

Other needed quantities:

Sample size: $T = 484$ #parameters: $K = 3$ #constraints: $K^* = 2$

Fisher statistics: $\frac{SSR_R - SSR_U}{SSR_U}\frac{T-K}{K^*} = \frac{234.8-210.4}{210.4}\frac{484-3}{2} = 27.89$

Choose the quantile in the Fisher distribution: $P[X \leq Q_{1-\alpha}] = 1-\alpha$

Reject the Null hypothesis if $27.89 > Q_{1-\alpha}$

Fisher distribution

F-test or Fisher-test: $\frac{SSR_R - SSR_U}{SSR_U} \frac{T-K}{K^*} \sim F(K^*, T-K)$

$$H_0 : \beta_2 = 10 \text{ and } \beta_3 = 100 \text{ vs } H_1 : \beta_2 \neq 10 \text{ or/and } \beta_3 \neq 100$$

Sample size: $T = 484$ #parameters: $K = 3$ #constraints: $K^* = 2$

Fisher statistics: $\frac{SSR_R - SSR_U}{SSR_U} \frac{T-K}{K^*} = \frac{234.8-210.4}{210.4} \frac{484-3}{2} = 27.89$

Intuition about rejecting the Null:

For large T, the F-statistic is proportional to a chi-square realization:

Large T: $K^* F_{\text{test}} \sim \chi^2(K^*)$

Expectation of a chi-square: $E(X) = K^*$ ➡ $0.5 \leq P[X \leq E(X)] \leq 0.69$ (empirically)

➡ If F-statistics below 1, we do not reject the Null hypothesis.

Linear regression with multiple explanatory variables

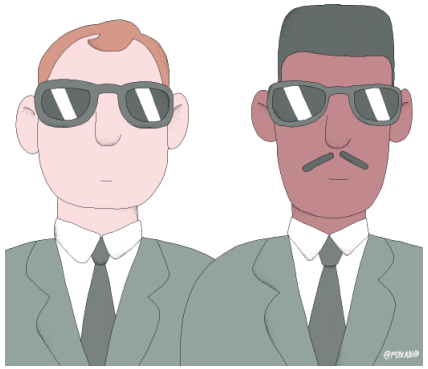
Testing and relaxing the Normality assumption

Statistical model

Linear regression: $y = X\beta + \epsilon.$

Assumptions for the generalized regression:

1. **Linear regression:** The variables are linearly related.
2. **No Collinearity:** Rank of matrix X is K.
3. **Strict exogeneity:** Errors have zero expectations conditional on all the explanatory variables.
4. **White noise:** No linear dependence between the error terms.
5. **Homoskedasticity:** The variance of the error term is constant.



Relaxing the regression assumptions

Linear regression: $y = X\beta + \epsilon.$

Relaxing assumptions is not a free lunch:

➡ **Assumptions are replaced by more flexible (but more technical) assumptions**

Relaxing assumptions requires an asymptotic framework

Asymptotic theory relies on the two statistics pillars (LLN and CLT)

➡ **Average of random variables tends to a constant (LLN – under conditions)**

➡ **Average of random variables behaves like Normal distribution (CLT – under conditions)**

OLS Estimator:

$$\hat{\beta} = (X'X)(X'y),$$

$$= \beta + \underbrace{(\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}}\underbrace{(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t)}$$

Under new assumptions: LLN and CLT are applicable!

Relaxing the regression assumptions

Linear regression: $y = X\beta + \epsilon.$

OLS Estimator: $\hat{\beta} = \beta + (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t)$

We define: $z_t = x_t \epsilon_t$

For cross-sectional data:

(i.e. $E(x_t x_t')$ and $E(z_t z_t')$)

$\{x_t\}$ and $\{z_t\}$ are i.i.d. processes with the second moments that exist

For time series data:

$\{x_t\}$ is a stationary and ergodic process

$\{z_t\}$ is a stationary and ergodic process that satisfies a Central limit theorem

Under these new assumptions: LLN and CLT are applicable!

$$(\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1} \to E(x_t x_t')^{-1} \text{ (LLN)}$$

$$(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t) \to N(0, \frac{V(x_t \epsilon_t)}{T}) \text{ (CLT)}$$

Implication of stationarity

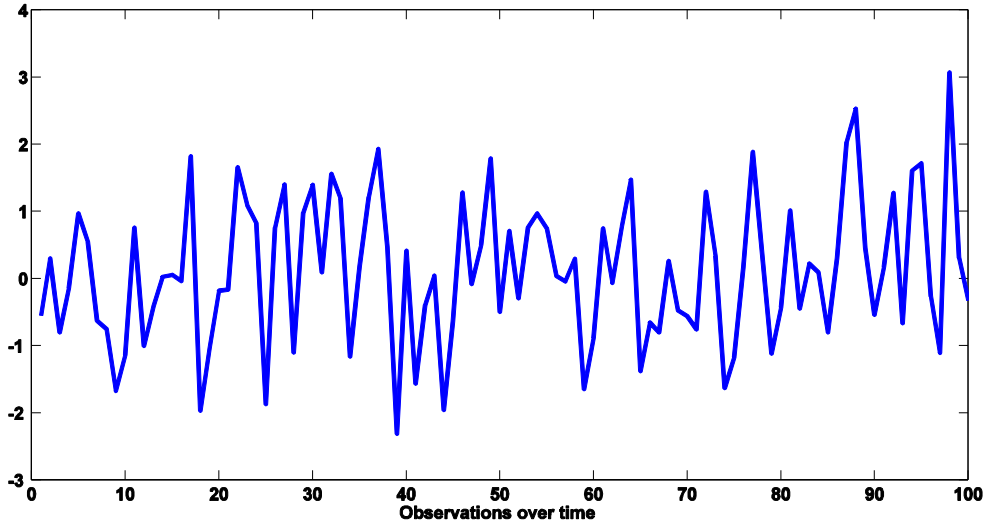
Linear regression: $y = X\beta + \epsilon.$

For time series data:

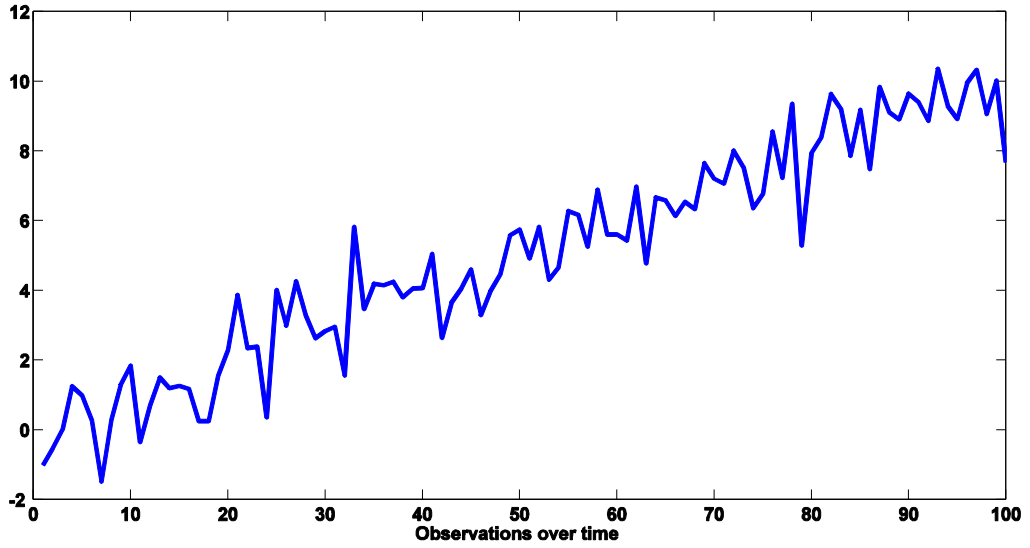
Asymptotic theory requires stationary processes.

Stationarity: The distribution and the dependence do not change over time.

$x_t = \epsilon_t$ with $\epsilon_t \sim N(0,1)$



$x_t = \beta_0 + \beta_1 t + \epsilon_t$ with $\epsilon_t \sim N(0,1)$



Take Away

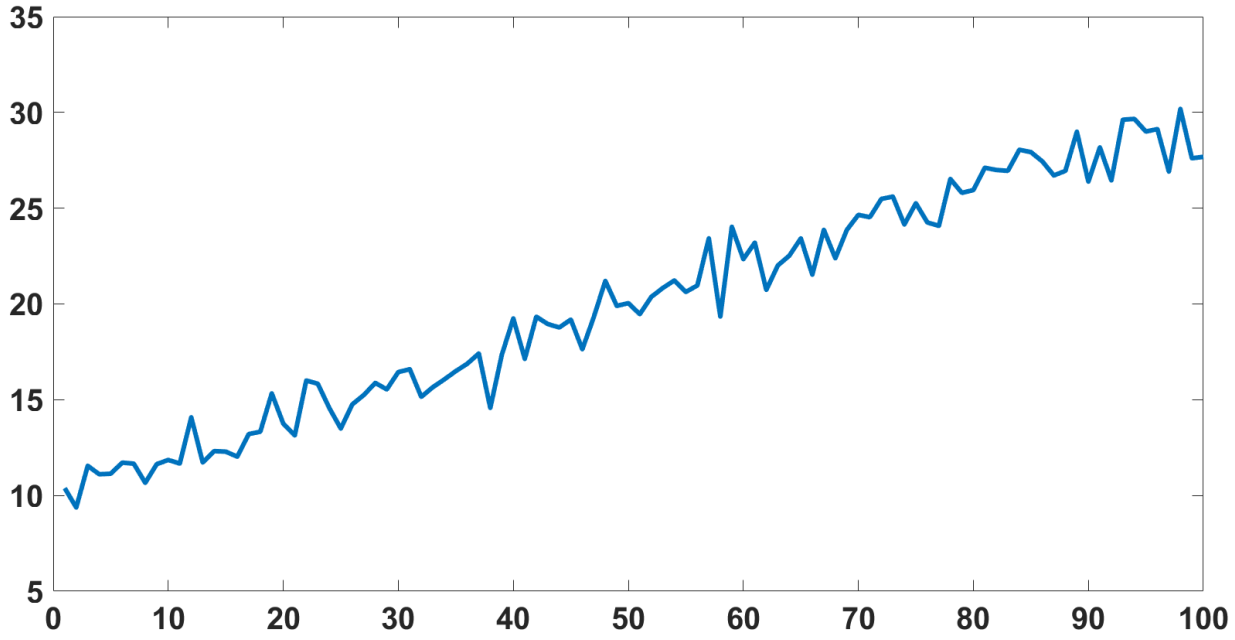
Stationarity implies that
expl. var. do not present trends

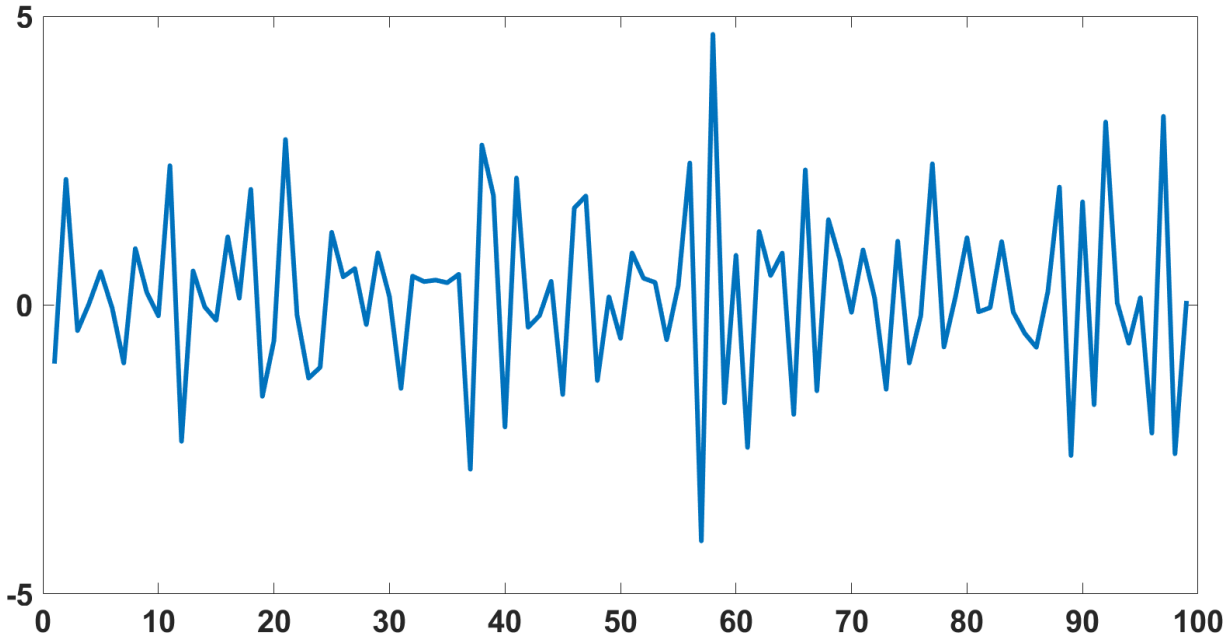
Implication of stationarity

Dealing with stationarity:

Linear trend: First-difference:

$$x_t = \beta_0 + \beta_1 t + \epsilon_t \text{ with } \epsilon_t \sim N(0,1) \quad \Rightarrow \quad \Delta x_t = \beta_1 + \epsilon_t \text{ with } \epsilon_t \sim N(0,2)$$



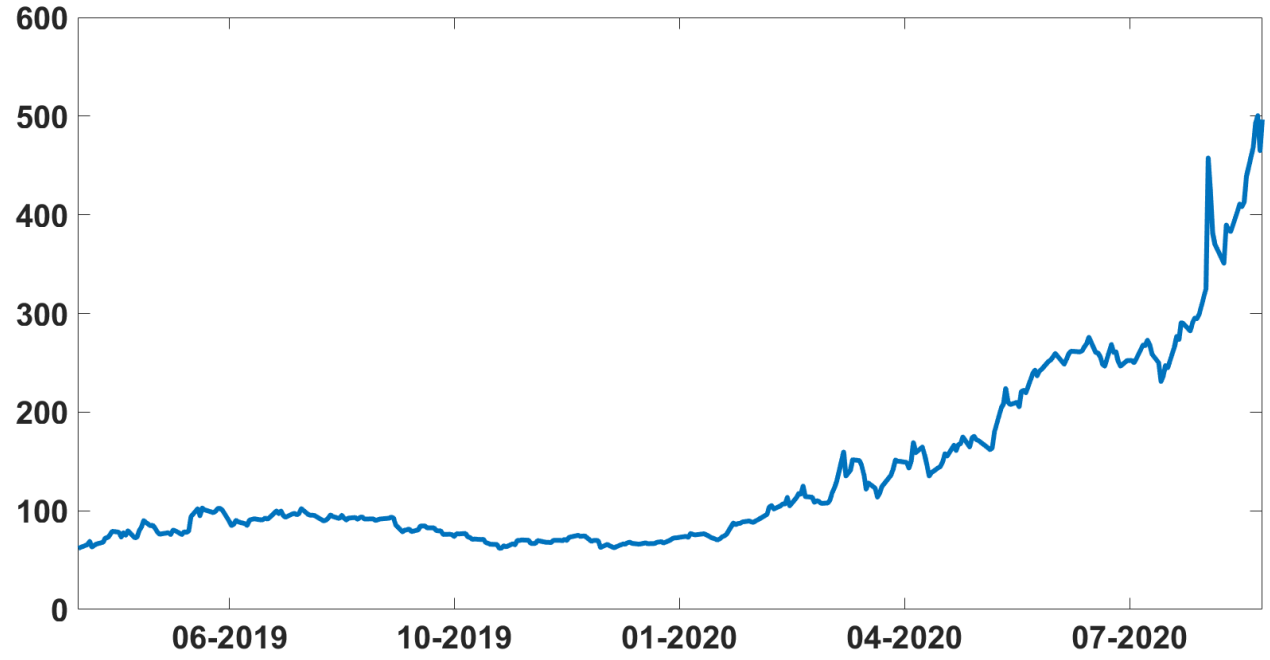


Implication of stationarity

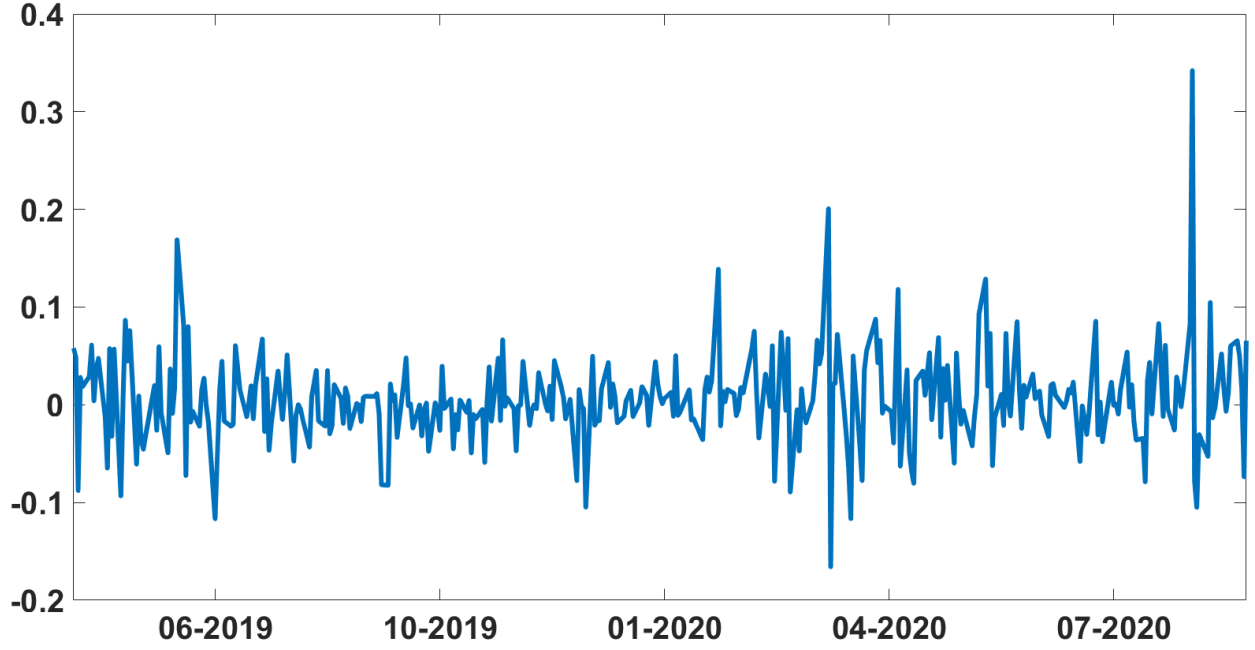
Dealing with stationarity:

Exponential trend: First-difference of logarithm: $\ln y_t - \ln y_{t-1} \approx \frac{(y_t - y_{t-1})}{y_{t-1}}$

Zoom asset prices



Zoom asset returns



Distribution

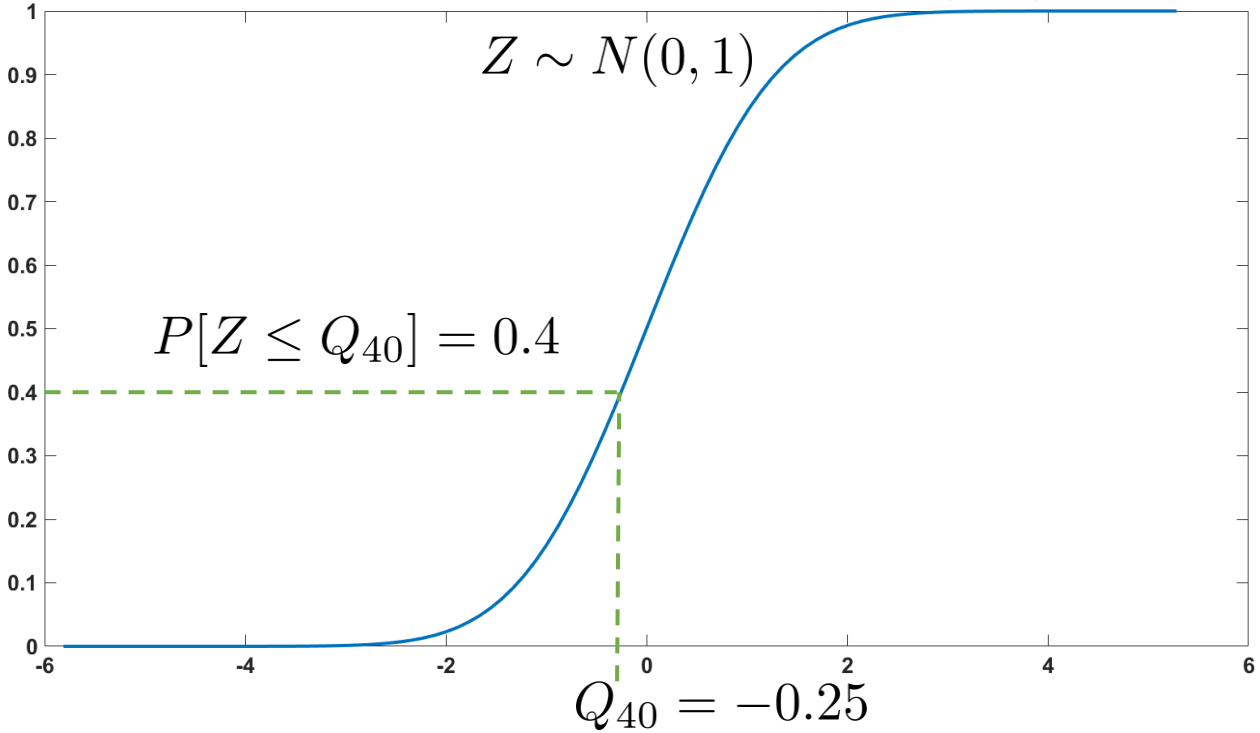
Linear regression: $y = X\beta + \epsilon.$

Testing the distribution: Error term is normally distributed.

QQ plot: Quantiles versus quantiles plot

Quantiles

CDF of standard Normal(0,1)

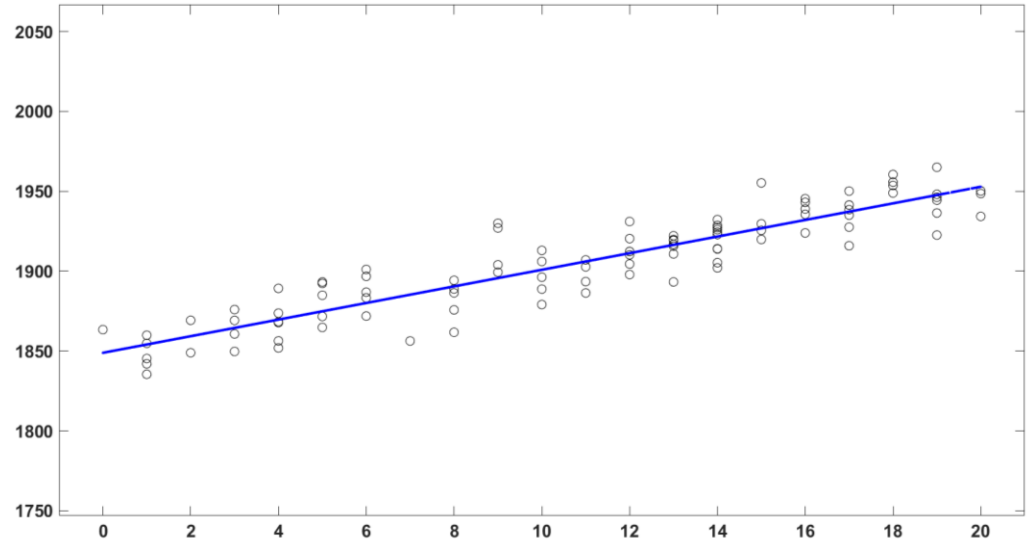


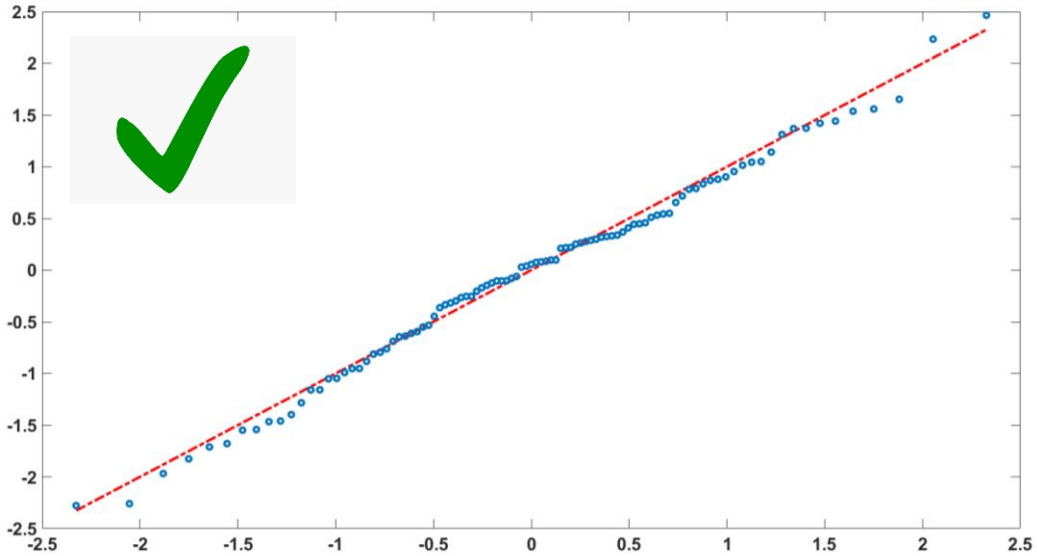
1. Normalize the residuals: $\frac{\hat{\epsilon}_t}{\hat{\sigma}}$
2. Sort the residuals by ascending order
3. Create a list containing the corresponding quantiles of the standard normal distribution
$$Q_{1/T}, Q_{2/T}, \ldots, Q_{(T-1)/T}$$
4. Plot the quantiles of the Normal distribution w.r.t. the sorted normalized residuals.

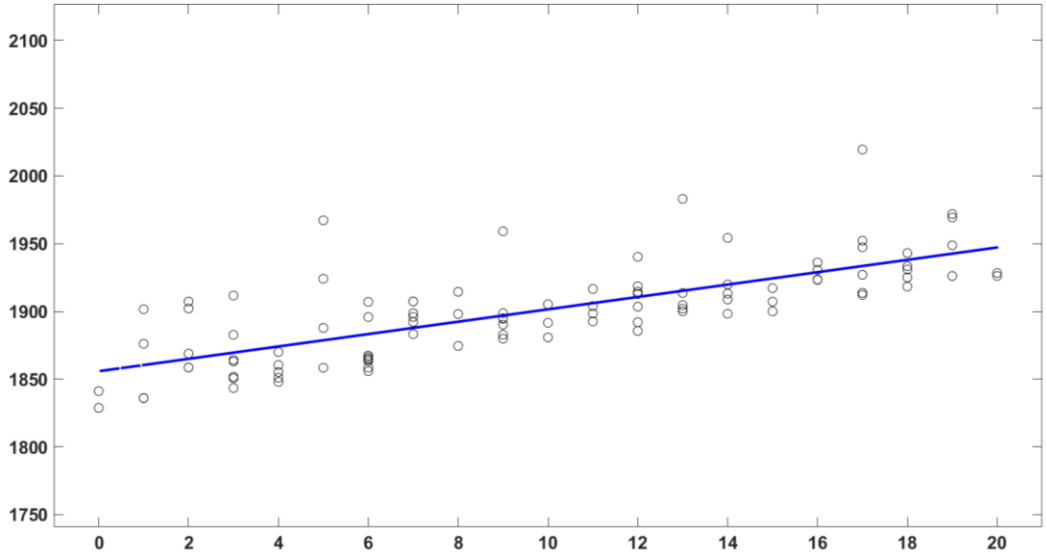
Distribution

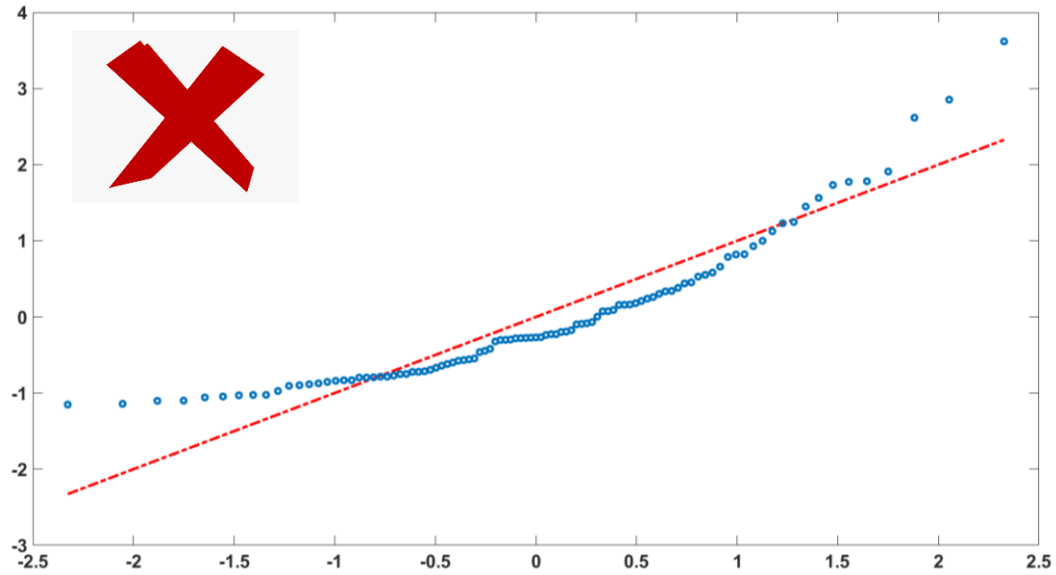
Linear regression: $y = X\beta + \epsilon.$

QQ plot for checking the Normality assumption









Distribution

Linear regression: $y = X\beta + \epsilon.$

Relaxing the assumption: Error term is normally distributed.

OLS estimator: $\hat{\beta} = \beta + (\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t)$

Application of the law of large numbers

$(\frac{1}{T}\sum_{t=1}^{T} x_t x_t')^{-1} = (\frac{X'X}{T})^{-1} \to_p E(x_t x_t')^{-1} = \text{Constant}$

Application of a Central limit theorem

If $x_t \epsilon_t$ is i.i.d. $\forall t$ then
$\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t \to_d N(0, \frac{S}{T})$

For large T, $\hat{\beta}|X \sim N(\beta, (X'X)^{-1}\sigma^2)$

Social dilemma

Linear regression with multiple explanatory variables

Too many variables

Big data

Linear regression: $y = X\beta + \epsilon.$

Big data: Very large set of data

In statistical words: Number of variables K larger than the sample size T

Example: **Asset pricing models:**

$$y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + \epsilon_t$$

Dependent variable: excess-return of a financial asset

Expl. variables: Risk factors (market, SMB, HML, …)

Harvey, Campbell R., Yan Liu, and Heqing Zhu. "… and the cross-section of expected returns."
The Review of Financial Studies 29.1 (2016): 5-68. (paper)

435 risk factors have been proposed so far…

Issues with Big data

Linear regression: $y = X\beta + \epsilon.$

Example: **Asset pricing models:**

$$y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + \epsilon_t$$

Dependent variable: excess-return of a financial asset

Expl. variables: 435 Risk factors (market, SMB, HML, …)

To estimate the model: you need at least 435 excess-returns

Monthly returns → **More than 36 years of data**

Yearly returns → **More than 435 years of data**

Problem:

Unlikely that you observe all the data over the last 435 years

Parameters may have changed over the last 435 years… Many crises since then

Issues with Big data

Linear regression: $y = X\beta + \epsilon.$

Example: **Asset pricing models:**

$$y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + \epsilon_t$$

Dependent variable: excess-return of a financial asset

You cannot use all the factors at the same time with the OLS approach!

Search for the best model by using F-test:

	x_1	(x_1, x_2)	(x_1, x_2, x_3)
Number Models:	2	$2^2 = 4$	$2^3 = 8$

Problem:

The number of models exponentially grows with the number of expl. var.

$$2^{435} \approx 10^{131} \quad >>> \quad 10^{21}$$

Issues with Big data

Linear regression: $y = X\beta + \epsilon.$

Regression assumption:

No Collinearity: Rank of matrix X is K.

Problems:

1. OLS estimator cannot be used when $T < K$
2. OLS estimator does not shrink irrelevant parameters

Hot research topic called penalized regression

→ Many estimators exist!

Most popular: Ridge and Lasso estimators

Penalized estimator

Linear regression: $y = X\beta + \epsilon.$

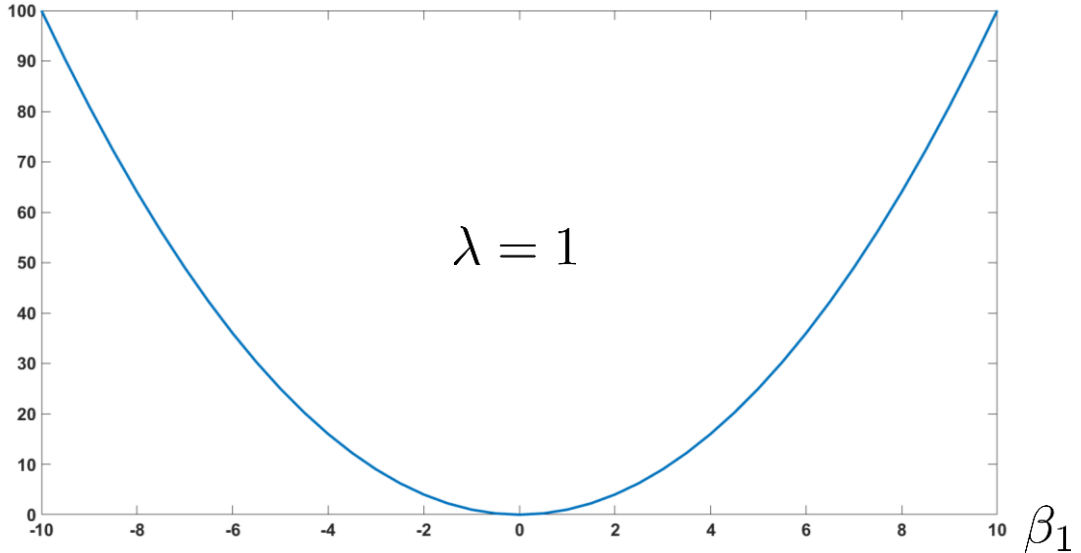
OLS Criterion: $\hat{\beta} = \text{Argmin}_\beta \epsilon'\epsilon = \sum_{t=1}^{T} \epsilon_t^2$ ➡ **Solution:** $\hat{\beta} = (X'X)^{-1}X'y$

Penalized Criterion: $\hat{\beta} = \text{Argmin}_\beta \epsilon'\epsilon + p(\beta)$ **Penalty:** Convex with $p(0) = 0$

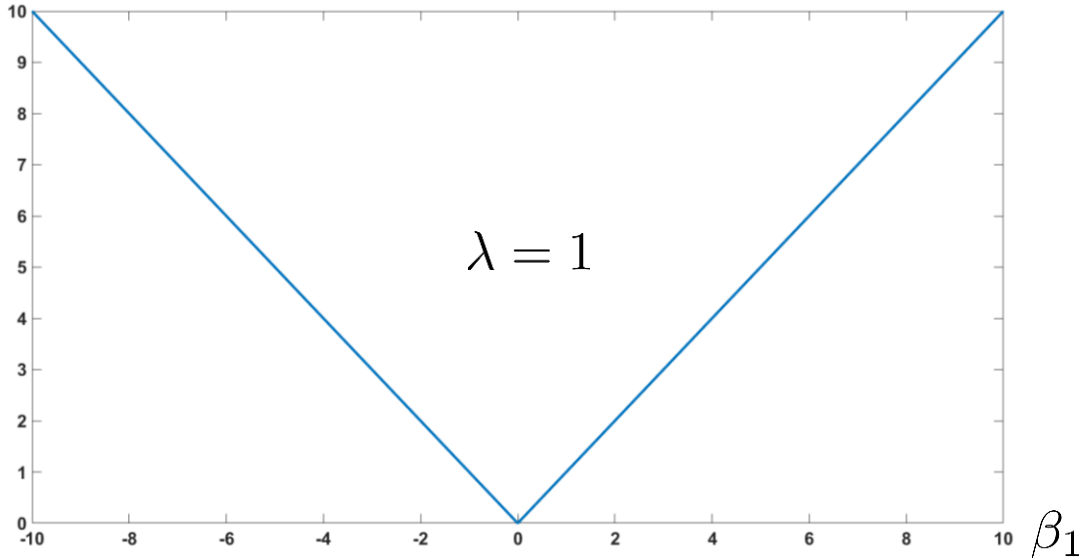
➡ **A lot of research on the choice of the penalty function!**

Most popular

Ridge penalty: $p(\beta) = \lambda \sum_{i=1}^{K} \beta_i^2$



Lasso penalty: $p(\beta) = \lambda \sum_{i=1}^{K} |\beta_i|$



Ridge estimator

Linear regression: $y = X\beta + \epsilon.$

Penalized Criterion: $\hat{\beta}_R = \text{Argmin}_\beta \epsilon'\epsilon + \lambda \sum_{i=1}^{K} \beta_i^2$

Solution: $\hat{\beta}_R = (X'X + \lambda I_K)^{-1} X'y$

Note: $(X'X + \lambda I) = (UDU' + \lambda UU') = U(D + \lambda I)U'$ → **Always invertible!**

Proof:

$$\begin{aligned}
\text{Pen. SSR}(\beta) &= \sum_{t=1}^{T} \epsilon_t^2 + \lambda \sum_{i=1}^{K} \beta_i^2, \\
&= \epsilon'\epsilon + \lambda\beta'\beta, \\
&= (y - X\beta)'(y - X\beta) + \lambda\beta'\beta, \\
&= y'y + \beta'(X'X + \lambda I_K)\beta - 2y'X\beta, \\
\frac{d\text{Pen. SSR}(\beta)}{d\beta} &= 2(X'X + \lambda I_K)\beta - 2X'y \quad (= 0) \\
\hat{\beta}_R &= (X'X + \lambda I_K)^{-1} X'y.
\end{aligned}$$

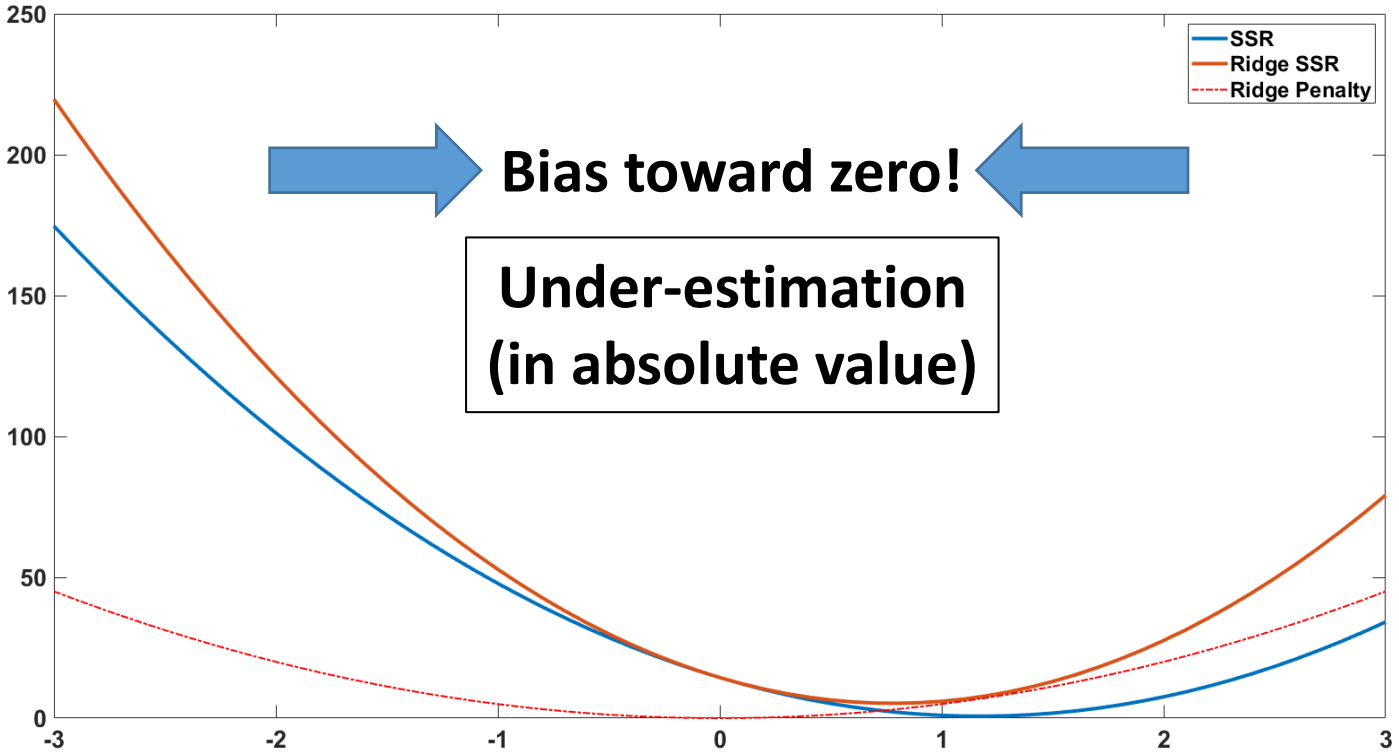
Biased estimator

Linear regression: $y = X\beta + \epsilon.$

Ridge estimator: $\hat{\beta}_R = (X'X + \lambda I_K)^{-1} X'y$

$= (X'X + \lambda I_K)^{-1} X'X\beta + (X'X + \lambda I_K)^{-1} X'\epsilon$

Bias: $E(\hat{\beta}_R|X) = (X'X + \lambda I_K)^{-1} X'X\beta \neq \beta$



Biased estimator

Linear regression: $y = X\beta + \epsilon.$

Ridge estimator: $\hat{\beta}_R = (X'X + \lambda I_K)^{-1} X'y$

$$= (\frac{1}{T}\sum_{t=1}^{T} x_t x_t' + \frac{\lambda}{T} I_K)^{-1} (\frac{1}{T}\sum_{t=1}^{T} x_t y_t)$$

Consistent: $\frac{\lambda}{T} \to 0$ → **The penalty is dominated by the sample information**

The ridge estimator converges to the OLS estimator as T increases

Efficient: **The ridge estimator has a smaller variance than the OLS estimator**

Biased estimator

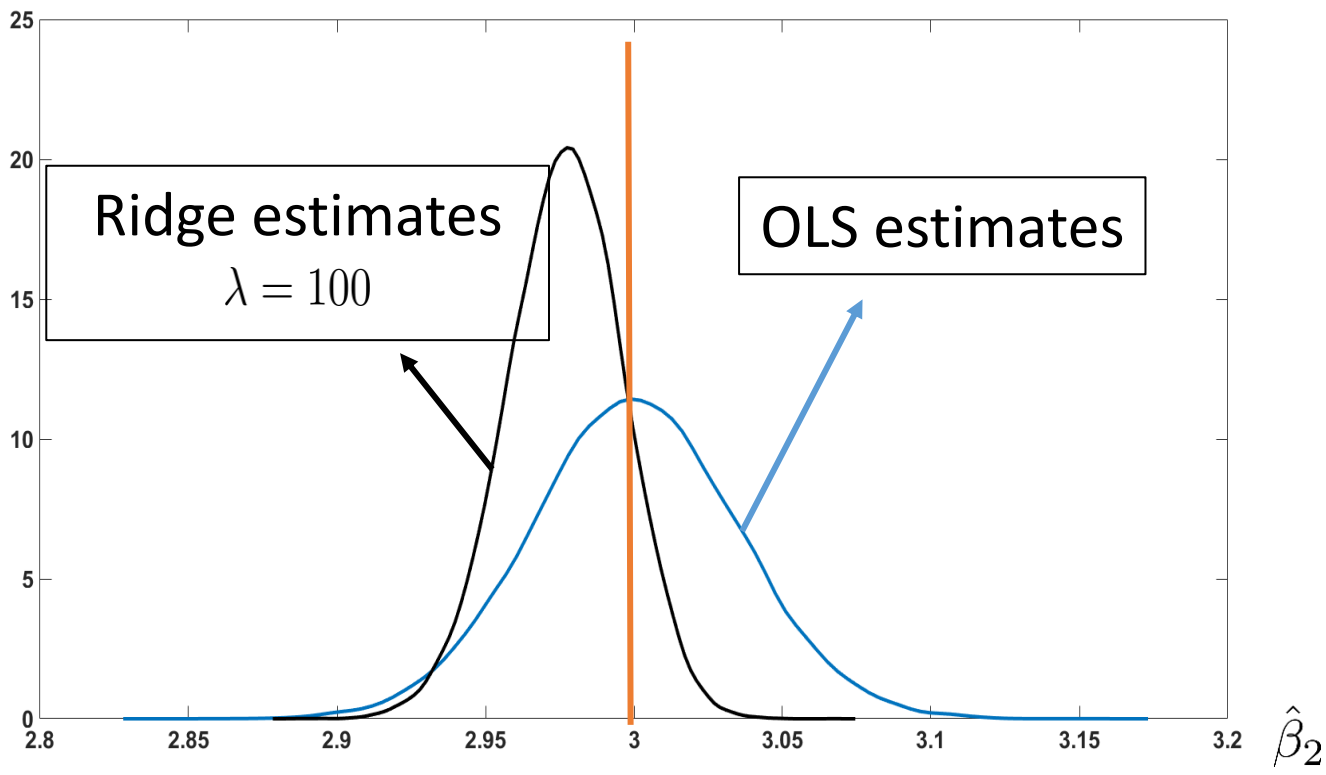
Linear regression: $y = X\beta + \epsilon.$

Ridge estimator: $\hat{\beta}_R = (X'X + \lambda I_K)^{-1} X'y$

Efficient: **The ridge estimator has a smaller variance than the OLS estimator.**

$$y_t = \underbrace{\beta_1}_{=1} + \underbrace{\beta_2}_{=3} \underbrace{x_t}_{\sim U[0,10]} + \underbrace{\epsilon_t}_{\sim N(0,1)}$$

50 000 replications with T=100



Summary

Linear regression: $y = X\beta + \epsilon.$

Ridge estimator: $\hat{\beta}_R = (X'X + \lambda I_K)^{-1} X'y$

Advantage of the Ridge:

1. **Works with any sample size and with collinear variables.**
2. **Consistent estimator.**

Drawback

1. **Biased estimator**
2. **Does not shrink parameters to zero**

Lasso estimator shrinks parameters to zero

Revolution in linear regression!

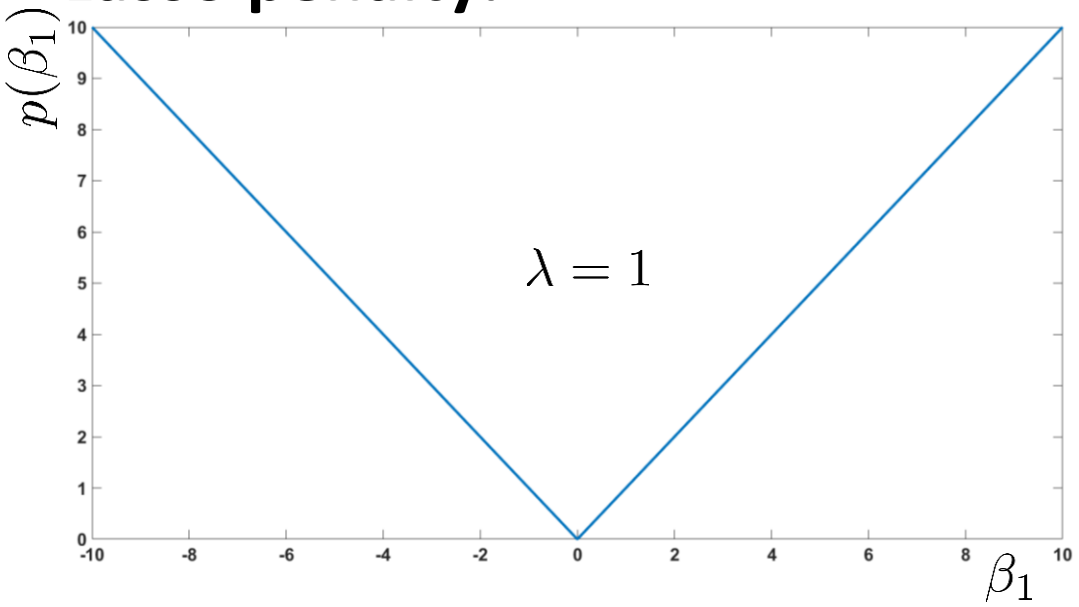
Lasso estimator

Linear regression: $y = X\beta + \epsilon.$

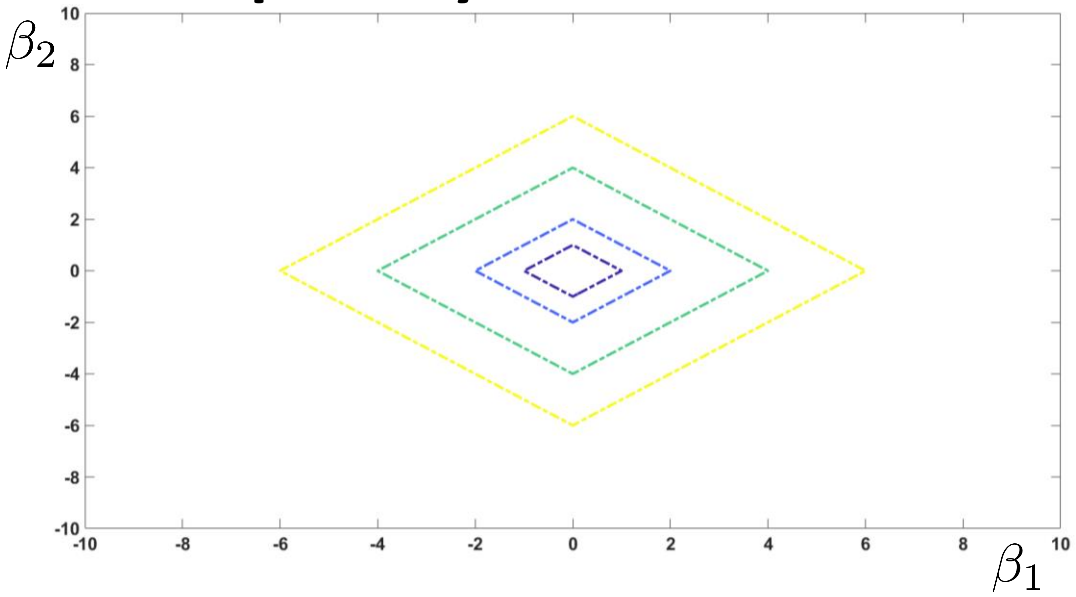
Penalized Criterion: $\hat{\beta}_L = \text{Argmin}_\beta \epsilon'\epsilon + \lambda \sum_{i=1}^{K} |\beta_i|$

Solution: Unique but not analytical…

Lasso penalty: $p(\beta) = \lambda|\beta_1|$



Lasso penalty: $p(\beta) = \lambda|\beta_1| + \lambda|\beta_2|$



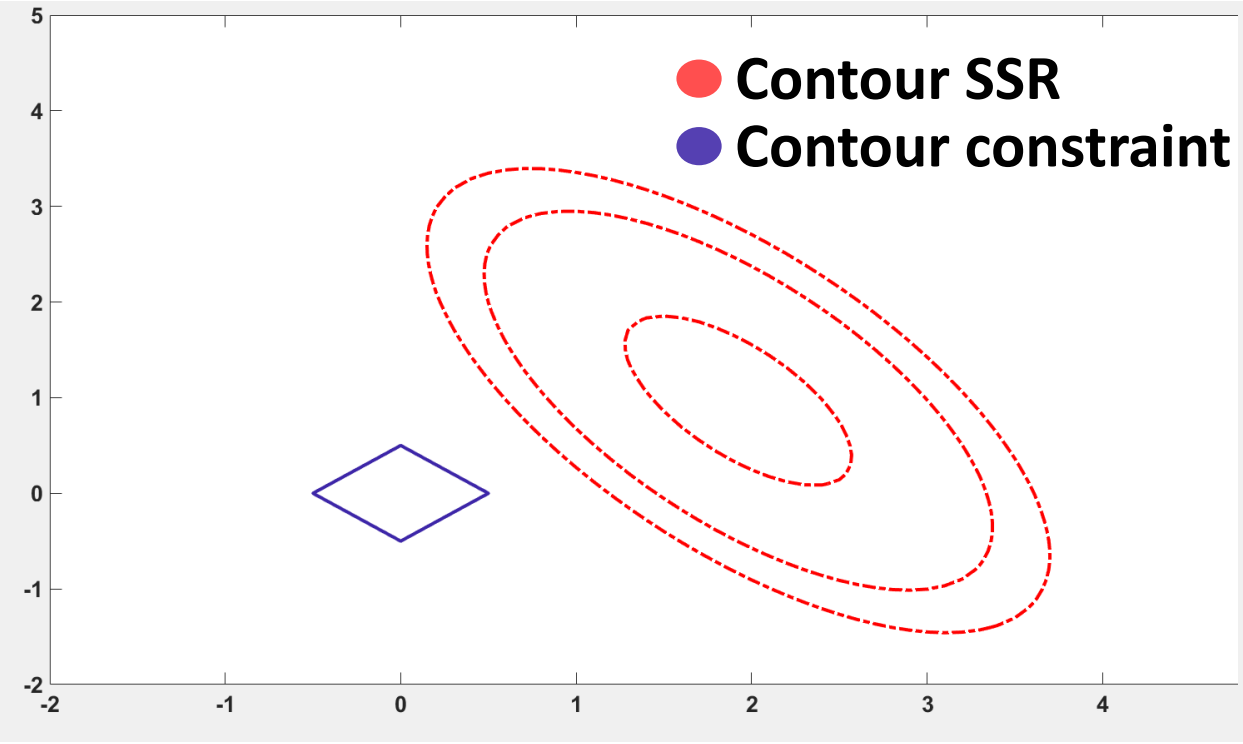
- **Biased but consistent estimator**

Lasso estimator

Linear regression: $y = X\beta + \epsilon.$

Why does the Lasso estimator shrink and Ridge does not ?

Lasso



Ridge