Introduction to Financial Econometrics

Chapter 2: Multiple Linear Regression Model

Christophe Hurlin

Univ Orléans

May, 2019

1. Introduction

In finance, investors are interested in the trade-off between (expected) **returns** and **risk**.

Suppose that one would like to assess this trade-off in the case of a technology company (Intel corporation equity price).

Questions:

- How can one quantify and interpret such a relationship?
- Is there any evidence that Intel return amplifies/attenuates market risk?
- Is there any evidence that this stock outperforms/underperforms the market?

1. Introduction

- To quantify risk, one can proceed with statistical models (e.g., using financial theory), and especially the **multiple linear regression models**.
- Once the model is specified (step 1), the second step consists in estimating the (unknown) value of the model parameters.
- In the case of the multiple linear regression models, we generally consider the **Ordinary Least Squares (OLS)** estimator defined by

1. Introduction

The objectives of this chapter are the following:

1. Define the (multiple) linear regression model.
2. Introduce the ordinary least squares (OLS) estimator.
3. Consider the CAPM (Capital Asset Pricing Model) as an illustration.
4. Define the main statistical properties of the OLS estimator.

1. Introduction

The outline of this chapter is the following:

- **Section 2:** The multiple linear regression model
- **Section 3:** The ordinary least squares estimator
- **Section 4:** Statistical properties of the OLS estimator

The detail of the outline is the following:

1. Introduction

References

- Campbell, J.,Y., A.W. Lo and A.C. MacKinlay, The Econometrics of Financial Markets, Princeton University Press, 1997.
- Greene W. (2007), Econometric Analysis, sixth edition, Pearson - Prentice Hil
- Tsay, R., 2002, Analysis of Financial Time Series, Wiley Series

1. Introduction

Notations: In this chapter, I will (try to...) follow some conventions of notation.

Y	random variable
y	realization
$\mathbf{y}$	vector
$\mathbf{Y}$	matrix

Problem: this system of notations does not allow to discriminate between a vector (matrix) of random elements and a vector (matrix) of non-stochastic elements (realization).

Abadir and Magnus (2002), Notation in econometrics: a proposal for a standard, Econometrics Journal.

Section 2

The Multiple Linear Regression Model:

Specification and Assumptions

2. The multiple linear regression model

Definition (Linear regression model)

The **linear regression model** is used to study the relationship between a dependent variable and one explanatory variable. The generic form of the linear regression model is

$$y_t = \alpha + \beta x_t + \varepsilon_t, \qquad t = 1, \ldots, T$$

Notations

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

- α is called the **intercept**.
- β is the **slope** (parameter) of the regression.
- Both parameters are assumed to be fixed and unknown.

Notations (cont'd)

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

- y_t is the **dependent variable**, the **regressand**, or the **explained variable**.
- x_t is an **explanatory variable**, a **regressor** or a **covariate**.
- ε_t is the **error term** or **disturbance**.

IMPORTANT: do not use the term "residual"

Notations (cont'd)

The term ε_t is a **random disturbance**, so named because it "disturbs" an otherwise stable relationship. The disturbance arises for several reasons:

1. Primarily because we cannot hope to capture every influence on an economic variable in a model. The net effect (positive or negative) of these **omitted factors** is captured in the disturbance.

2. There are many **errors of measurement** on the variables used in the model.

Assumption

We assume that the error terms $\{\varepsilon_1, \ldots, \varepsilon_T\}$ are **independent and identically distributed (i.i.d.)** with

$$\mathbb{E}(\varepsilon_t) = 0 \qquad \forall t$$

$$\mathbb{V}(\varepsilon_t) = \sigma_\varepsilon^2 \qquad \forall t$$

where σ_ε^2 is the variance of error terms.

Note: The acronym i.i.d. means that all the random variables $\varepsilon_1, \ldots, \varepsilon_T$ have the same distribution and are independently distributed.

Sub-Section 2.1

The CAPM and the Linear Regression Model

2.1. The CAPM and the linear regression model

Objectives

1. Define the **Capital Asset Pricing Model** (CAPM).
2. Define the **systematic** and **idiosyncratic** risks.
3. To write the CAPM as a **linear regression model**.
4. To collect a **dataset** (sample) to evaluate a CAPM model.
5. To study the **descriptive statistics** of the data.

2.1. The CAPM and the linear regression model

Definition (Capital Asset Pricing Model)

The **Capital Asset Pricing Model** (CAPM) is an economic model that specifies what expected returns (and therefore prices) should be as a function of **systematic risk**.

Systematic vs. idiosyncratic risks

1. **Systematic risk** arises from market structure or dynamics which produce shocks or uncertainty faced by all agents in the market; such shocks could arise from government policy, international economic forces, etc.

2. **Idiosyncratic risk** is the risk to which only specific agents or industries are vulnerable.

3. The idiosyncratic risk can be reduced or eliminated through **diversification**; but since all market actors are vulnerable to systematic risk, it cannot be limited through diversification.

Source: Wikipedia

Remarks

- The CAPM is a model for pricing an individual security or portfolio
- The CAPM puts structure to Markowitz's (1952) mean-variance optimization theory.
- The CAPM assumes only one source of systematic risk: **market risk.**
- Investors are compensated for the market risk by a **risk premium.**
- Their compensation is proportional to the risk exposure.

Markowitz, H.M. (1952), Portfolio Selection, *The Journal of Finance*, 7(1), 77–91.

2.1. The CAPM and the linear regression model

Definition (security market line)

If the CAPM is true, then all securities should lie in the **security market line** (SML) which represents the expected rate of return of an individual security as a function of the systematic (market) risk, such that

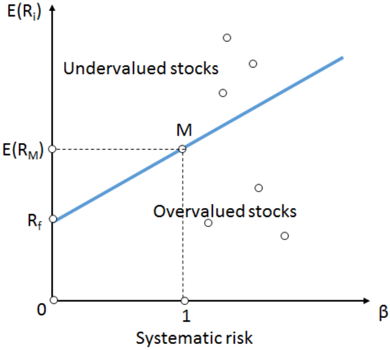
$$\mathbb{E}\left(R_i\right) = r_f + \beta_i \left(\mathbb{E}\left(R_m\right) - r_f\right)$$

Notations

- $\mathbb{E}\left(R_i\right)$ is the expected return of the asset i.
- $\mathbb{E}\left(R_m\right)$ is the expected return of the market portfolio.
- r_f is a risk-free rate (non-stochastic).
- β_i (**beta** of security i) represents the systematic (market) risk.

2.1. The CAPM and the linear regression model

Figure: Illustration of the security market line



Source: Wikipedia

Notations (cont'd)

- $\mathbb{E}(R_m) - r_f$ is the expected **excess return** of the market portfolio, also called the **market premium**.
- $\mathbb{E}(R_i) - r_f$ is the expected **excess return** of asset i, also called the **risk premium**.

2.1. The CAPM and the linear regression model

Definition (beta coefficient)

The **beta** parameter β_i represents the sensitivity of the expected excess asset returns to the expected excess market returns, with

$$\beta_i = \frac{\mathbb{C}ov\left(R_i, R_m\right)}{\mathbb{V}\left(R_m\right)}$$

Interpretation

$$\mathbb{E}(R_i) = r_f + \beta_i \left(\mathbb{E}(R_m) - r_f\right)$$

- If $\beta_i = 0$, asset i is not exposed to market risk. Thus, the investor is not compensated with higher return:

$$\mathbb{E}(R_i) = r_f$$

- If $\beta_i > 0$, asset i is exposed to market risk and $\mathbb{E}(R_i) > r_f$, provided that $\mathbb{E}(R_m) > r_f$.

- If $\beta_i = 1$, the expected return of asset i is equal to the expected market return

$$\mathbb{E}(R_i) = \mathbb{E}(R_m)$$

From the theoretical CAPM model to a linear regression model

Definition (the CAPM as a regression model)

The empirical CAPM model for an asset i at all time t can defined as

$$R_{i,t} - r_{f,t} = \alpha_i + \beta_i \left(R_{m,t} - r_{f,t} \right) + \varepsilon_{i,t}$$

where α_i is a constant term (intercept), β_i denotes the slope parameter and $\varepsilon_{i,t}$ is an error term with $\mathbb{E}\left(\varepsilon_{i,t}\right) = 0$ and $\mathbb{V}\left(\varepsilon_{i,t}\right) = \sigma^2$.

Note: if the intercept α_i is null, then we have

$$\mathbb{E}\left(R_{i,t}\right) = r_{f,t} + \beta_i \left(\mathbb{E}\left(R_{m,t}\right) - r_{f,t} \right)$$

Definition (excess return)

In the rest of the chapter, we denote the **excess return** by

$$z_{j,t} = R_{j,t} - r_{f,t}$$

for an asset j or the market portfolio. Remind that $z_{j,t}$ is a random variable since the return $R_{j,t}$ is stochastic.

Example (CAPM model for Intel Corp.)

Consider a CAPM model for the equity Intel Corp. (ticker: INTC) given by

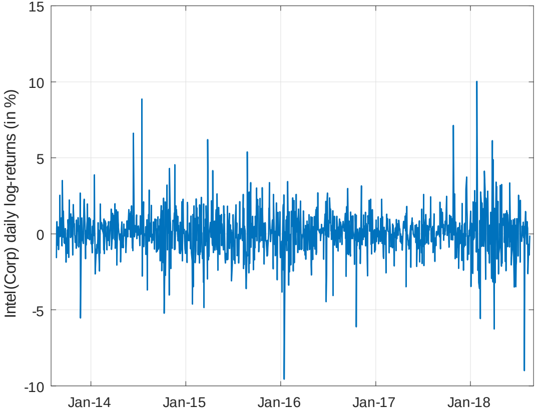
$$z_{\texttt{intel},t} = \alpha + \beta z_{\texttt{market},t} + \varepsilon_t$$

where $z_{\texttt{intel},t}$ (the dependent variable) is the **excess (log-) return of Intel**, $z_{\texttt{market},t}$ (the explanatory variable) is the **excess (log-) return of the market** and ε_t is an error term. We consider a sample of **daily** log(returns), based on the unadjusted closing prices for the equity INTEL Corp. from August 19, 2013 to August 17, 2018 (5 years).

2.1. The CAPM and the linear regression model

Figure: Closing daily prices for Intel Corp. (Aug 2013 - Aug 2018)

Figure: Daily log-returns for Intel Corp. (Aug 2013 - Aug 2018)



What is the Market Portfolio?

- It represents all wealth. We need to include not only all stocks, but all bonds, real estate, privately held capital, publicly held capital, and human capital in the world.
- Such a series does not exist: we have to use a **proxy**, typically a large portfolio of equities.
- In general, we consider the **SP500** index for the US market: this index is based on the market capitalizations of the 500 largest companies having common stock listed on the NYSE or NASDAQ.
- A measurement error is introduced: Roll's (1977) critique.

Roll R. (1977). A critique of the asset pricing theory's tests, *Journal of Financial Economics*, 4, 129-176.

Figure: Historical data available with Yahoo Finance (ticker: ^GSPC).

S&P 500 (^GSPC)
SNP - SNP Real Time Price. Currency in USD

Add to watchlist

2,850.13 +9.44 (+0.33%)
At close: August 17 4:49PM EDT

Summary Chart Options Components Historical Data

Time Period: Aug 19, 2013 - Aug 19, 2018 Show: Historical Prices Frequency: Daily Apply

Currency in USD Download Data

Date	Open	High	Low	Close*	Adj Close**	Volume
Aug 17, 2018	2,838.32	2,855.63	2,833.73	2,850.13	2,850.13	3,024,100,000
Aug 16, 2018	2,831.44	2,850.49	2,831.44	2,840.69	2,840.69	3,219,880,000
Aug 15, 2018	2,827.95	2,827.95	2,802.49	2,818.37	2,818.37	3,645,070,000
Aug 14, 2018	2,827.88	2,843.11	2,826.58	2,839.96	2,839.96	2,976,970,000
Aug 13, 2018	2,835.46	2,843.40	2,819.88	2,821.93	2,821.93	3,158,450,000
Aug 10, 2018	2,838.90	2,842.20	2,825.81	2,833.28	2,833.28	3,256,040,000
Aug 09, 2018	2,857.19	2,862.48	2,851.98	2,853.58	2,853.58	3,047,050,000
Aug 08, 2018	2,856.79	2,862.44	2,853.09	2,857.70	2,857.70	2,972,200,000
Aug 07, 2018	2,855.92	2,863.43	2,855.92	2,858.45	2,858.45	3,162,770,000
Aug 06, 2018	2,840.29	2,853.29	2,835.98	2,850.40	2,850.40	2,874,540,000

2.1. The CAPM and the linear regression model

Figure: Closing daily prices for the S&P500 index (Aug 2013 - Aug 2018)

2.1. The CAPM and the linear regression model

Figure: Daily log-returns for Intel Corp. (Aug 2013 - Aug 2018)

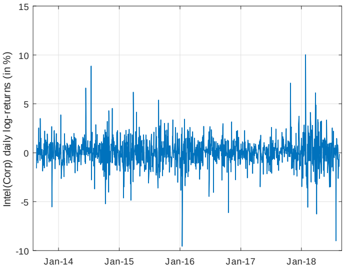
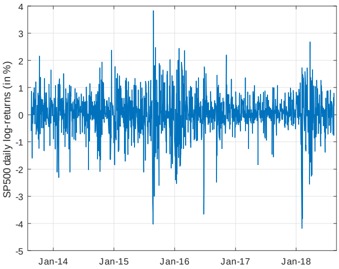


Figure: Daily log-returns for the S&P500 (Aug 2013 - Aug 2018)



2.1. The CAPM and the linear regression model

In order to estimate the parameters of the CAPM regression model, we need the **excess returns** (log-)returns.

Definition (excess returns)

The **excess** (log-)return is defined as difference between the return on the asset/portfolio (Intel or S&P500 in our case) and the (log-)return on a risk-free bond, denoted $r_{f,t}$.

$$z_{\texttt{intel},t} = R_{\texttt{intel},t} - r_{f,t}$$

$$z_{\texttt{market},t} = R_{\texttt{market},t} - r_{f,t}$$

Note: For the risk-free rate, we consider the **3 months treasury bill rate** for the US market.

Figure: Historical data available with Yahoo Finance (ticker: ^IRX).

13 WEEK TREASURY BILL (^IRX)
Chicago Options - Chicago Options Delayed Price. Currency in USD

Add to watchlist

1.995 -0.02 (-0.99%)
At close: August 17 2:59PM EDT

Summary Chart Conversations Options Components Historical Data

Time Period: Aug 19, 2017 - Aug 19, 2018 Show: Historical Prices Frequency: Daily Apply

Currency in USD Download Data

Date	Open	High	Low	Close*	Adj Close**	Volume
Aug 17, 2018	2.00	2.01	1.99	2.00	2.00	-
Aug 16, 2018	2.02	2.02	2.01	2.02	2.02	-
Aug 15, 2018	2.04	2.04	2.02	2.02	2.02	-
Aug 14, 2018	2.03	2.03	2.02	2.03	2.03	-
Aug 13, 2018	2.00	2.01	2.00	2.00	2.00	-
Aug 10, 2018	2.00	2.00	2.00	2.00	2.00	-
Aug 09, 2018	2.01	2.01	2.00	2.01	2.01	-
Aug 08, 2018	2.01	2.01	2.01	2.01	2.01	-
Aug 07, 2018	2.01	2.01	2.01	2.01	2.01	-
Aug 06, 2018	1.97	1.97	1.97	1.97	1.97	-

Remarks

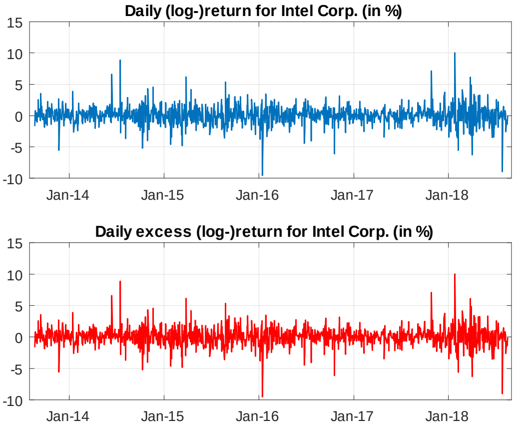
- In general, the T-Bill rate is quoted as an annualized interest rate.
- In order to compute the daily excess (log-) return, we have to **convert** this annual rate in a daily return.
- Denote by $r_{f,t} \equiv r_{f,t}^{[1]}$ the daily rate and $r_{f,t}^{[365]}$ the yearly interest rate, respectively. According to the simple interest formula (cf. Chapter 1), we have:

$$r_{f,t}^{[1]} = \frac{1}{365} r_{f,t}^{[365]}$$

Figure: Daily T-bill rate (Aug 2013 - Aug 2018)

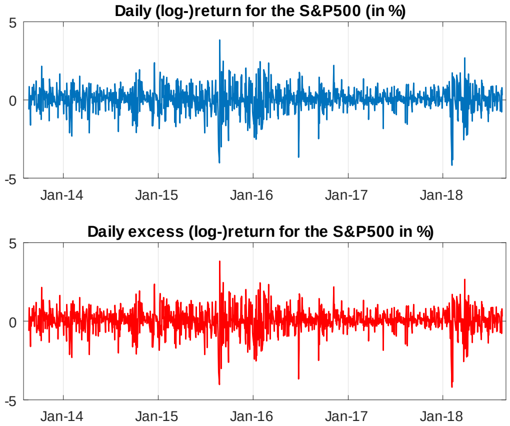
2.1. The CAPM and the linear regression model

Figure: Daily returns and excess returns for Intel Corp. (Aug 2013 - Aug 2018)



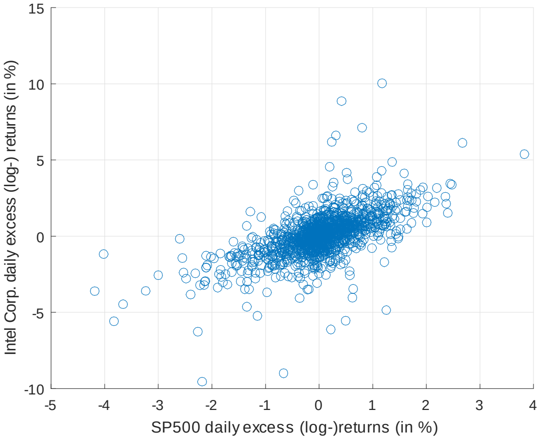
2.1. The CAPM and the linear regression model

Figure: Daily returns and excess returns for the S&P500 index (Aug 2013 - Aug 2018)



2.1. The CAPM and the linear regression model

Figure: Scatter plot of the excess returns of the S&P500 and Intel Corp.



2.1. The CAPM and the linear regression model

Table 1: Descriptive statistics for the daily excess returns (August 22, 2017 to August 17, 2018)

Daily excess return	Intel Corp.	S&P500
Mean	0.0011	0.0005
Median	0.0015	0.0009
Maximum	0.1002	0.0267
Minimum	-0.0898	-0.0418
Std. Dev.	0.0188	0.0078
Skewness	0.1942	-1.3194
Kurtosis	8.7062	8.9555
Jarque–Bera	340.74	442.00
p-value	0.0000	0.0000
Observations	250	250

Key Concepts

1. Capital Asset Pricing Model (CAPM)
2. Security Market Line (SML)
3. Systematic and idiosyncratic risk
4. Beta parameter
5. Market portfolio
6. Excess (log-) return
7. Descriptive statistics

Sub-Section 2.2

Specification of the Multiple Linear Regression Model

2.2. Specification of the multiple linear regression model

Objectives

1. Define the **(multiple) linear regression model**.
2. Make a distinction between the **semi-parametric** and **parametric** MLR model.
3. Introduce the multiple linear **Gaussian** model.
4. Introduce a **vectorial definition** of the MLR model.

Multiple linear regression model

- Other explanatory variables might explain variations of the excess (log-) return of Intel : macroeconomic variables (e.g., inflation), financial variables (e.g., Fama-French factors or price-to-dividend ratio), etc.
- For instance,

$$z_{\texttt{intel},t} = \beta_0 + \beta_1 z_{\texttt{market},t} + \beta_2 \text{inflation}_t + \varepsilon_t$$

- This is called the **multiple** linear regression model.

2.2. Specification of the multiple linear regression model

Definition (Multiple linear regression model)

The **multiple linear regression model** is used to study the (linear) relationship between a dependent variable and one or more independent variables, given by

$$y_t = x_{t,1}\beta_1 + x_{t,2}\beta_2 + \ldots + x_{t,K}\beta_K + \varepsilon_t$$

where y is the **dependent** (or explained) variable and $\mathbf{x}_1, .., \mathbf{x}_K$ are the **explanatory** (or independent) variables.

Notation

$$x_{t,k} = \text{value of the } k^{th} \text{ explanatory variable for time } t$$

$$x_{\text{time,variable}}$$

Notations (cont'd)

$$\underset{T\times 1}{\mathbf{y}} = \begin{pmatrix} y_1 \\ y_2 \\ .. \\ y_t \\ .. \\ y_T \end{pmatrix} \quad \underset{T\times 1}{\mathbf{x}_k} = \begin{pmatrix} x_{1,k} \\ x_{2,k} \\ .. \\ x_{t,k} \\ .. \\ x_{T,k} \end{pmatrix} \quad \underset{T\times 1}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ .. \\ \varepsilon_t \\ .. \\ \varepsilon_T \end{pmatrix} \quad \underset{K\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ .. \\ \beta_K \end{pmatrix}$$

Notations (cont'd)

$$\underset{T\times K}{\mathbf{X}} = \left(\mathbf{x}_1 : \mathbf{x}_2 : \ldots : \mathbf{x}_K \right)$$

or equivalently

$$\underset{T\times K}{\mathbf{X}} = \begin{pmatrix} x_{1,1} & x_{1,2} & .. & x_{1,k} & .. & x_{1,K} \\ x_{2,1} & x_{2,2} & .. & x_{2,k} & .. & x_{2,K} \\ .. & .. & .. & .. & .. & .. \\ x_{t,1} & x_{t,2} & .. & x_{t,k} & .. & x_{t,K} \\ .. & .. & .. & .. & .. & .. \\ x_{T,1} & x_{T,2} & .. & x_{T,k} & .. & x_{T,K} \end{pmatrix}$$

2.2. Specification of the multiple linear regression model

Definition (multiple linear regression model)

The **multiple linear regression model** can be written

$$\underset{T\times 1}{\mathbf{y}} = \underset{T\times K}{\mathbf{X}}\underset{K\times 1}{\boldsymbol{\beta}} + \underset{T\times 1}{\boldsymbol{\varepsilon}}$$

where K denotes the number of regressors (including the intercept).

Remark

More generally, the matrix $\mathbf{X}$ may as well contain **stochastic** and **non stochastic** elements such as:

- Constant;
- Time trend;
- Dummy variables (for specific episodes in time);
- etc.

Therefore, $\mathbf{X}$ is generally a mixture of fixed and random variables.

Remark: If the model includes a **constant term** (intercept), then we have

$$y_t = \underbrace{1 \times \beta_1}_{\text{intercept}} + x_{t,2}\beta_2 + .. + x_{t,K}\beta_K + \varepsilon_t$$

The matrix $\mathbf{X}$ becomes

$$\underset{T \times K}{\mathbf{X}} = (\mathbf{e} : \mathbf{x}_2 : .. : \mathbf{x}_K) = \begin{pmatrix} \mathbf{1} & x_{1,2} & .. & x_{1,k} & .. & x_{1,K} \\ \mathbf{1} & x_{2,2} & .. & x_{2,k} & .. & x_{2,K} \\ .. & .. & .. & .. & .. & .. \\ \mathbf{1} & x_{t,2} & .. & x_{t,k} & .. & x_{t,K} \\ .. & .. & .. & .. & .. & .. \\ \mathbf{1} & x_{T,2} & .. & x_{T,k} & .. & x_{T,K} \end{pmatrix}$$

where $\mathbf{e}$ is a unit $T \times 1$ vector

Example (CAPM model for Intel Corp.)

The CAPM model for Intel Corp. can be written as

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \varepsilon_t$$

or equivalently

$$\underset{T\times 1}{y} = \underset{T\times 2}{\mathbf{X}}\underset{2\times 1}{\boldsymbol{\beta}} + \underset{T\times 1}{\boldsymbol{\varepsilon}}$$

with

$$\underset{T\times 2}{\mathbf{X}} = \begin{pmatrix} 1 & z_{\texttt{market},1} \\ .. & .. \\ .. & .. \\ 1 & z_{\texttt{market},T} \end{pmatrix} \quad \underset{T\times 1}{\mathbf{y}} = \begin{pmatrix} z_{\texttt{intel},1} \\ .. \\ .. \\ z_{\texttt{intel},T} \end{pmatrix} \quad \underset{T\times 1}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ .. \\ .. \\ \varepsilon_T \end{pmatrix} \quad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

One key difference for the specification of the MLRM:

Parametric/semi-parametric specification

Parametric model: the distribution of the error terms is fully characterized, e.g. $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega})$

Semi-Parametric specification: only a few moments of the error terms are specified, e.g. $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathbb{V}(\boldsymbol{\varepsilon}) = \mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\right) = \mathbf{\Omega}$.

2.2. Specification of the multiple linear regression model

This **difference** does not matter for the derivation of the ordinary least square estimator

But this difference matters for (among others):

1. The characterization of the statistical properties of the OLS estimator (e.g., efficiency);
2. The choice of alternative estimators (e.g., the maximum likelihood estimator, etc.).

Definition (Semi-parametric multiple linear regression model)

The **semi-parametric multiple linear regression model** is defined by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where the error term $\boldsymbol{\varepsilon}$ satisfies

$$\mathbb{E}\left(\boldsymbol{\varepsilon}\middle|\,\mathbf{X}\right) = \underset{T\times 1}{\mathbf{0}}$$

$$\mathbb{V}\left(\boldsymbol{\varepsilon}\middle|\,\mathbf{X}\right) = \sigma^2 \underset{T\times T}{\mathbf{I}_T}$$

and $\mathbf{I}_T$ is the identity matrix of order T.

Remarks

1. If the matrix X is non stochastic (fixed), i.e. there are only fixed regressors, then the conditions on the error term $\boldsymbol{\varepsilon}$ read:

$$\mathbb{E}\left(\boldsymbol{\varepsilon}\right)=\mathbf{0}$$

$$\mathbb{V}\left(\boldsymbol{\varepsilon}\right)=\sigma^2\mathbf{I}_T$$

2. If the (conditional) variance covariance matrix of $\boldsymbol{\varepsilon}$ is not diagonal, i.e. if

$$\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)=\boldsymbol{\Omega}$$

the model is called the **Multiple Generalized Linear Regression Model**

Remarks (cont'd)

The two conditions on the error term $\boldsymbol{\varepsilon}$

$$\mathbb{E}\left(\boldsymbol{\varepsilon} \mid \mathbf{X}\right) = \mathbf{0}_{T \times 1}$$

$$\mathbb{V}\left(\boldsymbol{\varepsilon} \mid \mathbf{X}\right) = \sigma^2 \mathbf{I}_T$$

are equivalent to

$$\mathbb{E}\left(\mathbf{y} \mid \mathbf{X}\right) = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbb{V}\left(\mathbf{y} \mid \mathbf{X}\right) = \sigma^2 \mathbf{I}_T$$

2.2. Specification of the multiple linear regression model

Definition (The multiple linear Gaussian model)

The **(parametric) multiple linear Gaussian model** is defined by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where the error term $\boldsymbol{\varepsilon}$ is normally distributed

$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_T\right)$$

As a consequence, the vector $\mathbf{y}$ has a conditional normal distribution with

$$\mathbf{y} | \mathbf{X} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T\right)$$

Remarks

1. The **multiple linear Gaussian model** is (by definition) a parametric model.

2. If the matrix X is non stochastic (fixed), i.e. there are only fixed regressors, then the vector $\mathbf{y}$ has **marginal** normal distribution:

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T\right)$$

Key Concepts

1. The multiple linear regression model.
2. Semi-parametric multiple linear regression model
3. Parametric multiple linear regression model
4. Multiple linear Gaussian model

Sub-Section 2.3

Assumptions on the Multiple Linear Regression Model

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

2.3. Assumptions on the multiple linear regression model

Definition (Assumption 1: Linearity)

The model is **linear** with respect to the parameters $\beta_1, .., \beta_K$.

Remarks

- The model specifies a linear relationship between the dependent variable and the regressors. For instance, the models

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

$$y_t = \beta_0 + \beta_1 \cos(x_t) + v_t$$

$$y_t = \beta_0 + \beta_1 \times \frac{1}{x_t} + \omega_t$$

are all linear with respect to (w.r.t.) $\boldsymbol{\beta}$.

- In contrast, the model $y_t = \beta_0 + \beta_1 x_t^{\beta_2} + \varepsilon_t$ is non linear w.r.t. $\boldsymbol{\beta}$.

- The model can be linear after some transformations. Starting from $y_t = A x_t^{\beta} \exp(\varepsilon_t)$, one has a **log-linear** specification:

$$\ln(y_t) = \ln(A) + \beta \ln(x_t) + \varepsilon_t$$

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

2.3. Assumptions on the multiple linear regression model

Definition (Assumption 2: Full column rank)

$\mathbf{X}$ is an $T \times K$ matrix with rank K.

Interpretation

1. There is no exact relationship among any of the independent variables in the model.
2. The columns of $\mathbf{X}$ are linearly independent.

Remarks

1. Perfect multi-collinearity is generally not difficult to spot and is signalled by most statistical software.
2. Imperfect multi-collinearity is a more serious issue.

Example (identification)

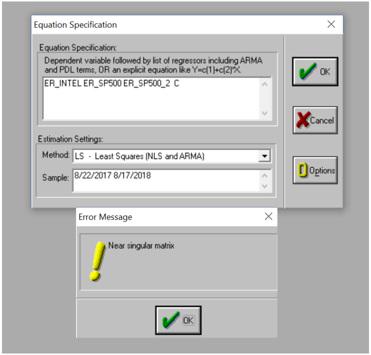
Suppose that we want to estimate the following model:

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \beta_3 \left(z_{\texttt{market},t} \times 2 \right) + \varepsilon_t,$$

The identification condition does not hold since the variables $z_{\texttt{market},t}$ and $z_{\texttt{market},t} \times 2$ are perfectly collinear. It is impossible to estimate β_2 and β_3.

2.3. Assumptions on the multiple linear regression model

Figure: Example of perfect multi-collinearity. Source: Eviews 3



Definition (Identification)

The multiple linear regression model is **identifiable** if and only if one the following equivalent assertions holds:

(i) $\operatorname{rank}(\mathbf{X}) = K$

(ii) The matrix $\mathbf{X}^{\top}\mathbf{X}$ is invertible

(iii) $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2 \Longrightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \quad \forall (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \in \mathbb{R}^K \times \mathbb{R}^K$

(iv) $\mathbf{X}\boldsymbol{\beta} = 0 \Longrightarrow \boldsymbol{\beta} = 0 \quad \forall \boldsymbol{\beta} \in \mathbb{R}^K$

(v) $\ker(\mathbf{X}) = \{0\}$

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

2.3. Assumptions on the multiple linear regression model

Definition (Assumption 3: Strict exogeneity of the regressors)

The regressors are **exogenous** if:

$$\mathbb{E}\left(\varepsilon\middle|\,\mathbf{X}\right) = \mathbf{0}_{T\times 1}$$

or equivalently

$$\mathbb{E}\left(\varepsilon_t\middle|\,x_{s,k}\right) = 0$$

for any explanatory variable $k \in \{1, \ldots, T\}$ and any time $(t, s) \in \{1, \ldots, T\}$.

2.3. Assumptions on the multiple linear regression model

Comments

1. The expected value of the error term at time t is not a function of the explanatory variables observed at any observation (including the t^{th} observation).
2. The explanatory variables are not predictors of the error terms.
3. The strict exogeneity condition can be rewritten as:

$$\mathbb{E}\left(\mathbf{y} \mid \mathbf{X}\right) = \mathbf{X}\beta$$

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

Definition (Assumption 4: Spherical disturbances)

The error terms are such that:

$$\mathbb{V}\left(\varepsilon_t \middle| \mathbf{X}\right) = \mathbb{E}\left(\varepsilon_t^2 \middle| \mathbf{X}\right) = \sigma^2 \quad \text{for all time } t \in \{1, ..T\}$$

and

$$\mathbb{C}ov\left(\varepsilon_t, \varepsilon_s \middle| \mathbf{X}\right) = \mathbb{E}\left(\varepsilon_t \times \varepsilon_s \middle| \mathbf{X}\right) = 0 \quad \text{for all } t \neq s$$

Notes:

1. The condition of constant variances is called **homoscedasticity**.
2. The uncorrelatedness across observations is called **non-autocorrelation.**

2.3. Assumptions on the multiple linear regression model

Comments

1. Spherical disturbances = **homoscedasticity** + **non-autocorrelation**
2. If the errors are not spherical, we call them nonspherical disturbances.
3. The assumption of homoscedasticity is a strong one: this is the exception rather than the rule!

Comments

Let us consider the (conditional) variance covariance matrix of the error terms:

$$\underbrace{\mathbb{V}\left(\boldsymbol{\varepsilon}\middle|\,\mathbf{X}\right)}_{T\times T} = \underbrace{\mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\middle|\,\mathbf{X}\right)}_{T\times T} =$$

$$\begin{pmatrix}
\mathbb{E}\left(\varepsilon_1^2\middle|\,\mathbf{X}\right) & \mathbb{E}\left(\varepsilon_1\varepsilon_2\middle|\,\mathbf{X}\right) & .. & \mathbb{E}\left(\varepsilon_1\varepsilon_t\middle|\,\mathbf{X}\right) & .. & \mathbb{E}\left(\varepsilon_1\varepsilon_T\middle|\,\mathbf{X}\right) \\
\mathbb{E}\left(\varepsilon_2\varepsilon_1\middle|\,\mathbf{X}\right) & \mathbb{E}\left(\varepsilon_2^2\middle|\,\mathbf{X}\right) & .. & \mathbb{E}\left(\varepsilon_2\varepsilon_t\middle|\,\mathbf{X}\right) & .. & \mathbb{E}\left(\varepsilon_2\varepsilon_T\middle|\,\mathbf{X}\right) \\
.. & & .. & .. & .. & .. \\
\mathbb{E}\left(\varepsilon_s\varepsilon_1\middle|\,\mathbf{X}\right) & & .. & \mathbb{E}\left(\varepsilon_s\varepsilon_t\middle|\,\mathbf{X}\right) & .. & \mathbb{E}\left(\varepsilon_s\varepsilon_T\middle|\,\mathbf{X}\right) \\
.. & & .. & .. & .. & .. \\
\mathbb{E}\left(\varepsilon_T\varepsilon_1\middle|\,\mathbf{X}\right) & & .. & \mathbb{E}\left(\varepsilon_T\varepsilon_t\middle|\,\mathbf{X}\right) & .. & \mathbb{E}\left(\varepsilon_T^2\middle|\,\mathbf{X}\right)
\end{pmatrix}$$

Comments

Let us consider the (conditional) variance covariance matrix of the error terms:

$$\underbrace{\mathbb{V}\left(\boldsymbol{\varepsilon}\middle|\mathbf{X}\right)}_{T\times T} = \underbrace{\mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\middle|\mathbf{X}\right)}_{T\times T} =$$

$$\begin{pmatrix} \mathbb{V}\left(\varepsilon_1\middle|\mathbf{X}\right) & \mathbb{C}ov\left(\varepsilon_1\varepsilon_2\middle|\mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_1\varepsilon_t\middle|\mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_1\varepsilon_T\middle|\mathbf{X}\right) \\ \mathbb{C}ov\left(\varepsilon_2\varepsilon_1\middle|\mathbf{X}\right) & \mathbb{V}\left(\varepsilon_2\middle|\mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_2\varepsilon_t\middle|\mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_2\varepsilon_T\middle|\mathbf{X}\right) \\ .. & & .. & .. & .. & .. \\ \mathbb{C}ov\left(\varepsilon_s\varepsilon_1\middle|\mathbf{X}\right) & & .. & \mathbb{C}ov\left(\varepsilon_s\varepsilon_t\middle|\mathbf{X}\right) & .. & \mathbb{C}ov\left(\varepsilon_s\varepsilon_T\middle|\mathbf{X}\right) \\ .. & & .. & .. & .. & .. \\ \mathbb{C}ov\left(\varepsilon_T\varepsilon_1\middle|\mathbf{X}\right) & & .. & \mathbb{C}ov\left(\varepsilon_T\varepsilon_t\middle|\mathbf{X}\right) & .. & \mathbb{V}\left(\varepsilon_T\middle|\mathbf{X}\right) \end{pmatrix}$$

Comments

The two assumptions (homoscedasticity and nonautocorrelation) imply that:

$$\underbrace{\mathbb{V}\left(\left.\boldsymbol{\varepsilon}\right|\mathbf{X}\right)}_{T\times T} = \underbrace{\mathbb{E}\left(\left.\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\right|\mathbf{X}\right)}_{T\times T} = \sigma^2\ \mathbf{I}_T$$

$$= \begin{pmatrix} \sigma^2 & 0 & .. & 0 & .. & 0 \\ 0 & \sigma^2 & .. & 0 & .. & 0 \\ .. & & .. & .. & .. & .. \\ .. & & .. & .. & .. & 0 \\ .. & & .. & .. & .. & .. \\ 0 & & .. & 0 & .. & \sigma^2 \end{pmatrix}$$

Comments

$$\underbrace{\mathbb{V}\left(\boldsymbol{\varepsilon}\middle|\mathbf{X}\right)}_{T\times T}=\underbrace{\mathbb{E}\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\middle|\mathbf{X}\right)}_{T\times T}=\sigma^2\ \mathbf{I}_T$$

1. **homoscedasticity** means the "same variance" for all the error terms

$$\mathbb{V}\left(\varepsilon_1\middle|\mathbf{X}\right)=\ldots=\mathbb{V}\left(\varepsilon_T\middle|\mathbf{X}\right)=\sigma^2$$

2. **non-autocorrelation** means "no correlation" for two error terms at two different dates

$$\mathbb{C}\left(\varepsilon_s\varepsilon_t\middle|\mathbf{X}\right)=0 \quad \text{if } s\neq t$$

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

Definition (Assumption 5: Data generation)

The data in $(x_{t,1}\ x_{t,2}\ ...x_{t,K})$ may be any mixture of **constants** and **random variables**.

Example (non-stochastic terms)

Some examples of non-stochastic terms used as regressors: a constant term (intercept), a time trend, or some dummy variables (in some particular cases).

Comments

- The fact that the columns of $\mathbf{X}$ are stochastic (or not) has an impact on the asymptotic properties.
- If the explanatory variables are randomly distributed, additional assumptions regarding $(x_{t,1}, \ldots, x_{t,K})$ are required. This is a statement about how the sample is drawn.
- In the sequel, we assume that $(x_{t,1} \; x_{t,2} \; ...x_{t,K})$ are **independently and identically distributed (i.i.d)** for $t = 1, .., T$.

2.3. Assumptions on the multiple linear regression model

The classical linear regression model consists of a set of assumptions that describes how the data set is produced by a data generating process (DGP)

Assumption 1: Linearity

Assumption 2: Full rank condition or identification

Assumption 3: Exogeneity

Assumption 4: Spherical error terms

Assumption 5: Data generation

Assumption 6: Normal distribution

Definition (Assumption 6: Normal distribution)

The disturbances are **normally** distributed.

$$\boldsymbol{\varepsilon} | \mathbf{X} \sim \mathcal{N}\left(\mathbf{0}_{T\times 1}, \sigma^2 \mathbf{I}_T\right)$$

Comments

1. Normality is **not necessary** to obtain most of the results presented below.
2. Assumption 6 implies assumptions 3 (exogeneity) and 4 (spherical disturbances).

$$\boldsymbol{\varepsilon} | \mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \ \mathbf{I}_T\right)$$

$$\mathbb{E}\left(\boldsymbol{\varepsilon} | \mathbf{X}\right) = \mathbf{0} \qquad \mathbb{V}\left(\boldsymbol{\varepsilon} | \mathbf{X}\right) = \sigma^2 \ \mathbf{I}_T$$

Summary

The main assumptions of the multiple linear regression model

A1: linearity	The model is linear with β
A2: identification	$\mathbf{X}$ is an $T \times K$ matrix with rank K
A3: exogeneity	$\mathbb{E}\left(\varepsilon \mid \mathbf{X}\right) = \mathbf{0}_{T\times 1}$
A4: spherical error terms	$\mathbb{V}\left(\varepsilon \mid \mathbf{X}\right) = \sigma^2 \mathbf{I}_T$
A5: data generation	$\mathbf{X}$ may be fixed or random
A6: normal distribution	$\varepsilon \mid \mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_T\right)$

2.3. Assumptions on the multiple linear regression model

Key Concepts

1. Assumptions of the multiple linear regression model
2. Linearity (A1)
3. Identification (A2)
4. Exogeneity (A3)
5. Spherical error terms (A4)
6. Data generation (A5)
7. Normal distribution (A6)

Section 3

The Ordinary Least Squares (OLS) Estimator

Introduction

1. The simple linear regression **model** assumes that the following specification is true in the **population**:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \qquad t = 1, \ldots, T$$

where other unobserved factors determining y_t are captured by the error term ε_t.

2. Consider a **sample** $\{X_t, Y_t\}_{i=1}^{T}$ of random variables and only one realization $\{x_t, y_t\}_{i=1}^{T}$ of this sample (your data set).

3. How to **estimate** the parameters β_1 and β_2?

4. A solution here is to use the **ordinary least squares estimator (OLS)**.

Sub-Section 3.1

Intuition of the OLS Estimator

Objectives

1. Define the **Ordinary Least Squares (OLS)** estimator.
2. Define the **Sum of Squared Residuals (SSR)** or **Residual Sum of Squares (RSS)**.
3. Define the notions of **predicted values** and **residuals**.
4. Define the **variance of the error terms**.
5. Define the **Standard Error (SE) of the regression**.

Intuition

Let us consider the following linear regression model :

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

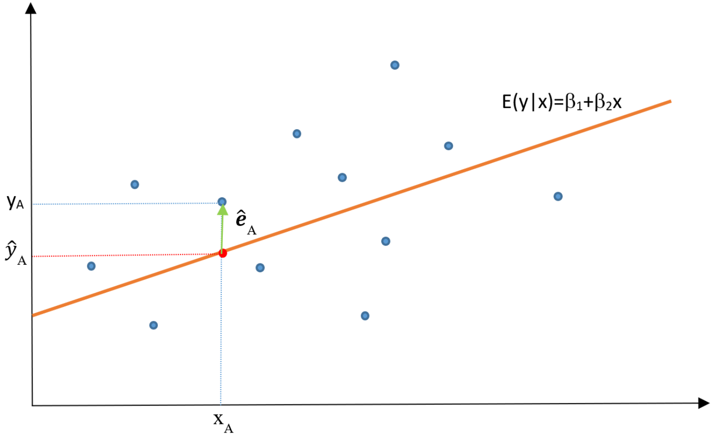
The general idea of the OLS consists in minimizing the "distance" between the points (x_t, y_t) and the regression line defined by

$$\widehat{y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 x_t$$

or the points $(x_i, \widehat{y}_i)$ for all $t = 1, .., T$.

3.1. Intuition of the OLS estimator

Figure: Intuition of the OLS estimator



3.1. Intuition of the OLS estimator

Definition (Sum of Squared Residuals)

Estimates of β_1 and β_2 are chosen by minimizing the **Sum of Squared Residuals (SSR),** or **Residual Sum of Squares (RSS)**

$$SSR = \sum_{t=1}^{T} \widehat{\varepsilon}_t^2$$

Note: The SSR corresponds to the sum of squares of the vertical distances between the actual y values and the predicted values of y, i.e.

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \sum_{t=1}^{T} \left(y_t - \widehat{\beta}_1 - \widehat{\beta}_2 x_t \right)^2$$

3.1. Intuition of the OLS estimator

Definition (OLS - simple linear regression model)

In the **simple linear regression** model $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, the **OLS estimators** $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are the solutions of the minimization problem

$$\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = \underset{(\beta_1,\beta_2)}{\arg\min} \sum_{t=1}^{T} \left(y_t - \beta_1 - \beta_2 x_t\right)^2$$

Definition (OLS estimators)

In the simple linear regression model $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, the **OLS estimators** $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are defined by

$$\widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2 \overline{x}$$

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{T} \left(x_t - \overline{x}_T \right) \left(y_t - \overline{y}_T \right)}{\sum_{t=1}^{T} \left(x_t - \overline{x}_T \right)^2}$$

where $\overline{y}_T = T^{-1} \sum_{t=1}^{T} y_t$ and $\overline{x}_T = T^{-1} \sum_{t=1}^{T} x_t$ respectively denote the sample mean of the dependent variable y and the regressor x.

Remark

The estimator for the slope parameter $\widehat{\beta}_2$ can also be expressed as

$$\widehat{\beta}_2 = \frac{cov\left(x_t, y_t\right)}{var\left(x_t\right)}$$

where $cov\left(x_t, y_t\right)$ is the **empirical (or sample) covariance** of y_t and x_t, and $var\left(x_t\right)$ denotes the **empirical (or sample) variance** of x_t:

$$cov\left(x_t, y_t\right) = \frac{1}{T-1}\sum_{t=1}^{T}\left(x_t - \overline{x}_T\right)\left(y_t - \overline{y}_T\right)$$

$$var\left(x_t\right) = \frac{1}{T-1}\sum_{t=1}^{T}\left(x_t - \overline{x}_T\right)^2$$

Example (CAPM)

We want to estimate the intercept and the slope parameter in the CAPM model

$$z_{\texttt{intel},t} = \alpha + \beta z_{\texttt{market},t} + \varepsilon_t$$

where $z_{\texttt{intel},t}$ is the excess (log-) return for Intel Corp and $z_{\texttt{market},t}$ is the excess (log-) return for the S&P500. We consider a sample of **250 observations** from August 22, 2017 to August 17, 2018 (1 year) for which we get

$$\sum_{t=1}^{T} \left(z_{\texttt{market},t} - \overline{z}_{\text{market}} \right) \left(z_{\texttt{intel},t} - \overline{z}_{\text{intel}} \right) = 0.023990$$

$$\sum_{t=1}^{T} \left(z_{\texttt{market},t} - \overline{z}_{\text{market}} \right)^2 = 0.015526$$

$$\sum_{t=1}^{T} z_{\texttt{intel},t} = 0.2889 \quad \sum_{t=1}^{T} z_{\texttt{market},t} = 0.1498$$

Question: compute the OLS estimates of the parameters α and β.

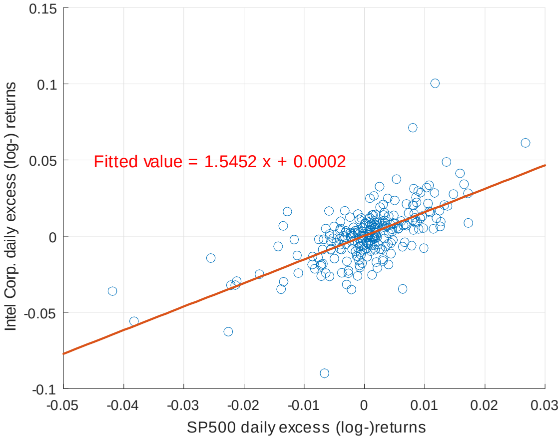
Solution

In the CAPM model, the OLS estimators $\widehat{\beta}$ and $\widehat{\alpha}$ are respectively defined by

$$\widehat{\beta} = \frac{\sum_{t=1}^{T}\left(z_{\texttt{market},t} - \overline{z}_{\text{market}}\right)\left(z_{\texttt{intel},t} - \overline{z}_{\text{intel}}\right)}{\sum_{t=1}^{T}\left(z_{\texttt{market},t} - \overline{z}_{\text{market}}\right)^2} = \frac{0.023990}{0.015526} = 1.5442$$

$$\widehat{\alpha} = \overline{z}_{\text{intel}} - \widehat{\beta}\,\overline{z}_{\text{market}} = \frac{0.2889}{250} - 1.5442 \times \frac{0.1498}{250} = 2.3032e^{-04} \simeq 0.0002$$

Figure: Regression line and fitted values



3.1. Intuition of the OLS estimator

Definition (Fitted value)

The **predicted (or fitted) value** of y_t is:

$$\widehat{y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 x_t$$

Note: The sample mean of the fitted values is equal to the sample mean of the observations

$$\overline{\widehat{y}}_T = \frac{1}{T} \sum_{t=1}^{T} \widehat{y}_t = \overline{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t$$

3.1. Intuition of the OLS estimator

Definition (residual)

The **residual** at time t is defined as:

$$\widehat{\varepsilon}_t = y_t - \widehat{\beta}_1 - \widehat{\beta}_2 x_t$$

with a sample mean equal to zero by definition:

$$\bar{\widehat{\varepsilon}}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t = 0$$

Remark

It is necessary to estimate **the variance σ_{ε}^{2} of the errors terms**

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

$$\varepsilon_t \quad \text{i.i.d. with } \mathbb{E}\left(\varepsilon_t\right) = 0 \quad \mathbb{V}\left(\varepsilon_t\right) = \sigma_{\varepsilon}^{2} \quad \forall t$$

So we have **3 parameters** to estimate

1. The intercept β_1.
2. The slope parameter β_2.
3. The variance σ_{ε}^{2} of the error terms.

Definition (standard error of the regression)

An estimator of the **variance of the error terms** σ_{ε}^2 is defined by

$$\widehat{\sigma}_{\varepsilon}^2 = \frac{1}{T-K}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 = \frac{SSR}{T-K}$$

where SSR denotes the sum of the squared residuals and K is the number of regressors (including the constant).

Note: The quantity $\widehat{\sigma}_{\varepsilon}$ is also called the **Standard Error (S.E.) of the regression** or the **Root Mean Squared Error (RMSE)**.

3.1. Intuition of the OLS estimator

Example (CAPM)

Write a Matlab script to estimate by OLS the intercept and the slope parameter in the CAPM model

$$z_{\texttt{intel},t} = \alpha + \beta z_{\texttt{market},t} + \varepsilon_t$$

where $z_{\texttt{intel},t}$ is the excess (log-) return for Intel Corp and $z_{\texttt{market},t}$ is the excess (log-) return for the S&P500. For that, consider a sample of 250 observations from August 22, 2017 to August 17, 2018 (1 year).

Note: the data are available within the file `Data_CAPM_returns.xlsx`.

3.1. Intuition of the OLS estimator

Figure: Matlab code for estimating a linear regression model

```
%====================================================
% PURPOSE: Data for CAPM model
% Course Name "Financial Econometrics", EDHEC Business School
% Chapter 1, Section 1. Multiple Linear Regression Models
%-------------------------
% Author: Christophe Hurlin
% Version: August 2018
%====================================================

clear , clc , close all

%=======================
%== Data importation ===
%=======================
[dataset,date]=xlsread('Data_CAPM_returns.xlsx');   % Importation of the data
date=datetime(date(2:end,1));                       % Transformation of the date
r_Intel_ex=dataset(:,1);                            % Excess daily (log-) returns for Intel
r_SP500_ex=dataset(:,2);                            % Excess daily (log-) returns for S&P500

%=================================================
%== Linear regression model and OLS estimator ===
%=================================================
reg=fitlm(r_SP500_ex,r_Intel_ex);
disp(reg)
```

Figure: CAPM model for Intel Corp (Aug 2017 - Aug 2018): Matlab function `fitlm`

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	0.00022961	0.00090928	0.25252	0.80085
x1	1.5452	0.11505	13.431	2.8468e-31

```
Number of observations: 250, Error degrees of freedom: 248
Root Mean Squared Error: 0.0143
R-squared: 0.421,  Adjusted R-Squared 0.419
F-statistic vs. constant model: 180, p-value = 2.85e-31
```

3.1. Intuition of the OLS estimator

Figure: CAPM model for Intel Corp (Aug 2017 - Aug 2018): Matlab function `fitlm`

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                   Estimate         SE        tStat       pValue
                  __________    __________    ______    __________

    (Intercept)   0.00022961    0.00090928    0.25252      0.80085
    x1                1.5452       0.11505     13.431   2.8468e-31

Number of observations: 250, Error degrees of freedom: 248
Root Mean Squared Error: 0.0143
R-squared: 0.421,  Adjusted R-Squared 0.419
F-statistic vs. constant model: 180, p-value = 2.85e-31
```

Key Concepts

1. Ordinary least squares estimator.
2. Sum of Squared Residuals (SSR) or Residual Sum of Squares (RSS).
3. Predicted (or fitted) value and residuals.
4. Variance of the error terms.
5. Standard Error of the regression.

Sub-Section 3.2

Definition of the OLS Estimator

3.2. Definition of the OLS estimator

Objectives

1. Define the OLS estimator in a **multiple** linear regression model.
2. Write the **vectorial formula** for the OLS estimator.
3. **Minimize** the sum of squared residuals.
4. Solve the same minimization problem with **matrix notation**.
5. Give a **geometrical interpretation** of the OLS estimator.

3.2. Definition of the OLS estimator

Now consider the **multiple** linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

or

$$y_t = \sum_{k=1}^{K} \beta_k x_{t,k} + \varepsilon_t$$

Objective: Find an estimator (and an estimate) of $\beta_1, \beta_2, .., \beta_K$ and σ^2 under the assumptions A1-A5.

3.2. Definition of the OLS estimator

OLS and multiple linear regression model

Different (equivalent) methods:

1. Minimize the sum of squared residuals (SSR).
2. Solve the minimization problem with matrix notation.
3. Geometrical interpretation.

1. Minimize the sum of squared residuals (SSR)

As for the simple linear regression, we have

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \varepsilon_t^2 = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \left(y_i - \sum_{k=1}^{K} \beta_k x_{t,k} \right)^2$$

One can derive the first order conditions with respect to β_k for $k = 1, .., K$ and solve a system of K equations with K unknowns.

3.2. Definition of the OLS estimator

Definition (OLS and multiple linear regression model)

In the **multiple** linear regression model $y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \varepsilon_t$, with $\mathbf{x}_t = (x_{t,1}, .., x_{t,K})^\top$, the OLS estimator $\widehat{\boldsymbol{\beta}}$ is the solution of

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} \sum_{t=1}^{T} \left(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} \right)^2$$

The **OLS estimators** of $\boldsymbol{\beta}$ is:

$$\underset{(K,1)}{\widehat{\boldsymbol{\beta}}} = \left(\sum_{t=1}^{T} \underset{(K,1)}{\mathbf{x}_t} \underset{(1,K)}{\mathbf{x}_t^\top} \right)^{-1} \left(\sum_{t=1}^{T} \underset{(K,1)}{\mathbf{x}_t} \underset{(1,1)}{y_t} \right)$$

3.2. Definition of the OLS estimator

2. Using matrix notations

Definition (OLS and multiple linear regression model)

the **multiple** linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, The OLS estimator $\widehat{\boldsymbol{\beta}}$ is the solution of the minimization problem

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min}\ \boldsymbol{\varepsilon}^{\top}\boldsymbol{\varepsilon} = \underset{\boldsymbol{\beta}}{\arg\min}\ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\top}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

The **OLS estimators** of $\boldsymbol{\beta}$ is:

$$\underset{(K,1)}{\widehat{\boldsymbol{\beta}}} = \left(\underset{(K,T)}{\mathbf{X}^{\top}}\underset{(T,K)}{\mathbf{X}}\right)^{-1}\left(\underset{(K,T)}{\mathbf{X}^{\top}}\underset{(T,1)}{\mathbf{y}}\right)$$

3. Geometric interpretation

1. The ordinary least squares estimation methods consists in determining the adjusted vector, $\widehat{\mathbf{y}}$, which is the closest to $\mathbf{y}$ (in a certain space...) such that the squared norm between $\mathbf{y}$ and $\widehat{\mathbf{y}}$ is minimized.

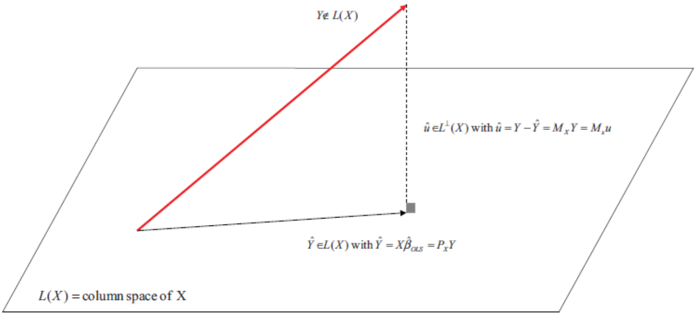
2. Finding $\widehat{\mathbf{y}}$ is equivalent to find an estimator of $\boldsymbol{\beta}$.

Definition (Geometric interpretation)

The adjusted vector, $\widehat{\mathbf{y}}$, is the (orthogonal) projection of $\mathbf{y}$ onto the column space of $\mathbf{X}$. The fitted error terms, $\widehat{\boldsymbol{\varepsilon}}$, is the projection of $\mathbf{y}$ onto the orthogonal space engendered by the column space of $\mathbf{X}$. The vectors $\widehat{\mathbf{y}}$ and $\widehat{\boldsymbol{\varepsilon}}$ are orthogonal.

3.2. Definition of the OLS estimator

3. Geometric interpretation



3. Geometric interpretation

Definition (Projection matrices)

The vectors $\widehat{\mathbf{y}}$ and $\widehat{\boldsymbol{\varepsilon}}$ are defined to be:

$$\widehat{\mathbf{y}} = \mathbf{P} \times \mathbf{y}$$

$$\widehat{\boldsymbol{\varepsilon}} = \mathbf{M} \times \mathbf{y}$$

where $\mathbf{P}$ and $\mathbf{M}$ denote the two following **projection matrices**:

$$\mathbf{P} = \mathbf{X} \left(\mathbf{X}^{\top} \mathbf{X} \right)^{-1} \mathbf{X}^{\top}$$

$$\mathbf{M} = \mathbf{I}_T - \mathbf{P} = \mathbf{I}_T - \mathbf{X} \left(\mathbf{X}^{\top} \mathbf{X} \right)^{-1} \mathbf{X}^{\top}$$

Remarks

Suppose that there is a **constant term** in the model.

1. The least squares residuals sum to zero:

$$\sum_{t=1}^{T} \widehat{\varepsilon}_t = 0$$

2. The regression hyperplane passes through the point of means of the data $(\overline{\mathbf{x}}_T, \overline{y}_T)$.

3. The mean of the fitted values of y equals the mean of the actual values of y:

$$\overline{\widehat{y}}_T = \overline{y}_T$$

3.2. Definition of the OLS estimator

Definition (Unbiased variance estimator)

An **unbiased estimator of** σ^2 is given by

$$\widehat{\sigma}^2 = \frac{1}{T-K}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 \equiv \frac{SSR}{T-K}$$

Note: The estimator $\widehat{\sigma}^2$ can also be written as

$$\widehat{\sigma}^2 = \frac{1}{T-K}\sum_{t=1}^{T}\left(y_t - \mathbf{x}_t^\top\widehat{\boldsymbol{\beta}}\right)^2$$

$$\widehat{\sigma}^2 = \frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^\top\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{T-K}$$

3.2. Definition of the OLS estimator

Key Concepts

1. Matrix notation for the OLS estimator.
2. Minimize the Sum of Squared Residuals.
3. Geometric interpretation of the OLS estimator.
4. Projection matrix.

Sub-Section 3.3

Applications to the CAPM model

3.3. Applications: CAPM model

Objectives

1. Apply the OLS estimator to the **CAPM model**.
2. Run an OLS estimation with **Excel**.
3. Interpret the OLS **estimation outputs** of various software

3.3. Applications: CAPM model

Example (CAPM)

We want to estimate the parameters β_1, β_2 and σ^2 in the CAPM model

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \varepsilon_t$$

$$\varepsilon_t \ \text{ i.i.d. } \left(0, \sigma^2\right)$$

For that we consider a sample of 250 observations from August 22, 2017 to August 17, 2018 (1 year), for which we observe

$$\sum_{t=1}^{250} z_{\texttt{market},t} = 0.1498 \qquad \sum_{t=1}^{250} z^2_{\texttt{market},t} = 0.0156$$

$$\sum_{t=1}^{250} z_{\texttt{intel},t} = 0.2889 \qquad \sum_{t=1}^{250} z_{\texttt{intel},t} \times z_{\texttt{market},t} = 0.0242$$

Question: Compute the OLS estimates of the parameters β_1, β_2 and σ^2 with the vectorial formula.

Note: the data are available within the file `Data_CAPM_returns.xlsx`.

3.3. Applications: CAPM model

Solution

The linear regression model can be written as

$$\underset{(250,1)}{\mathbf{y}} = \underset{(250,2)}{\mathbf{X}} \underset{(2,1)}{\boldsymbol{\beta}} + \underset{(250,1)}{\boldsymbol{\varepsilon}}$$

with $T = 250$ and $K = 2$

$$\underset{(250\times1)}{\mathbf{y}} = \begin{pmatrix} z_{\texttt{intel},1} \\ z_{\texttt{intel},2} \\ .. \\ z_{\texttt{intel},t} \\ .. \\ z_{\texttt{intel},250} \end{pmatrix} \quad \underset{250\times2}{\mathbf{X}} = \left(\mathbf{e} : \mathbf{z}_{\text{market}}\right) = \begin{pmatrix} 1 & z_{\texttt{market},1} \\ 1 & z_{\texttt{market},2} \\ .. & .. \\ 1 & z_{\texttt{market},t} \\ .. & .. \\ 1 & z_{\texttt{market},250} \end{pmatrix}$$

Solution

The linear regression model can be written as

$$\underset{(250,1)}{\mathbf{y}} = \underset{(250,2)}{\mathbf{X}} \underset{(2,1)}{\boldsymbol{\beta}} + \underset{(250,1)}{\boldsymbol{\varepsilon}}$$

with $T = 250$ and $K = 2$

$$\underset{(250\times 1)}{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ .. \\ \varepsilon_t \\ .. \\ \varepsilon_{250} \end{pmatrix} \quad \underset{(2\times 1)}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Solution (cont'd)

The OLS estimators of $\boldsymbol{\beta}$ is:

$$\underset{(2,1)}{\widehat{\boldsymbol{\beta}}} = \left(\underset{(2,250)}{\mathbf{X}^{\top}} \underset{(250,2)}{\mathbf{X}} \right)^{-1} \left(\underset{(2,250)}{\mathbf{X}^{\top}} \underset{(250,1)}{\mathbf{y}} \right)$$

with

$$\underset{(2,2)}{\mathbf{X}^{\top}\mathbf{X}} = \begin{pmatrix} T & \sum_{t=1}^{250} z_{\texttt{market},t} \\ \sum_{t=1}^{250} z_{\texttt{market},t} & \sum_{t=1}^{250} z_{\texttt{market},t}^{2} \end{pmatrix} = \begin{pmatrix} 250 & 0.1498 \\ 0.1498 & 0.0156 \end{pmatrix}$$

$$\underset{(2,1)}{\mathbf{X}^{\top}\mathbf{y}} = \begin{pmatrix} \sum_{t=1}^{250} z_{\texttt{intel},t} \\ \sum_{t=1}^{250} z_{\texttt{intel},t} \times z_{\texttt{market},t} \end{pmatrix} = \begin{pmatrix} 0.2889 \\ 0.0242 \end{pmatrix}$$

Solution (cont'd)

The OLS estimators of $\boldsymbol{\beta}$ is equal to:

$$\begin{aligned}
\underset{(2,1)}{\widehat{\boldsymbol{\beta}}} &= \left(\underset{(2,250)}{\mathbf{X}^{\top}} \underset{(250,2)}{\mathbf{X}} \right)^{-1} \left(\underset{(2,250)}{\mathbf{X}^{\top}} \underset{(250,1)}{\mathbf{y}} \right) \\
&= \begin{pmatrix} 250 & 0.1498 \\ 0.1498 & 0.0156 \end{pmatrix}^{-1} \begin{pmatrix} 0.2889 \\ 0.0242 \end{pmatrix} \\
&= \begin{pmatrix} 0.0002 \\ 1.5452 \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix}
\end{aligned}$$

3.3. Applications: CAPM model

Solution (cont'd)

The estimator of σ^2 is given by

$$\widehat{\sigma}^2 = \frac{1}{T-K}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 = \frac{SSR}{T-K}$$

Here, we have

$$SSR = 0.0510 \qquad T = 250 \qquad K = 2$$

$$T - K = 248 \text{ (number of degrees of freedom)}$$

$$\widehat{\sigma}^2 = \frac{0.0510}{248} = 2.0551e^{-04}$$

$$\text{S.E. of regression} = \text{RMSE} = \sqrt{\widehat{\sigma}^2} = 0.0143$$

3.3. Applications: CAPM model

Figure: CAPM model for Intel Corp (Aug 2017 - Aug 2018): Matlab function `fitlm`

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                    Estimate         SE          tStat       pValue
                   __________    __________    _______    __________

    (Intercept)    0.00022961    0.00090928    0.25252       0.80085
    x1                 1.5452       0.11505     13.431    2.8468e-31


Number of observations: 250, Error degrees of freedom: 248
Root Mean Squared Error: 0.0143
R-squared: 0.421,  Adjusted R-Squared 0.419
F-statistic vs. constant model: 180, p-value = 2.85e-31
```

Example (CAPM)

Question: Write a **Matlab** code to estimate by OLS the parameters β_1, β_2 and σ^2 in the CAPM model

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \varepsilon_t$$

$$\varepsilon_t \text{ i.i.d. } \left(0, \sigma^2\right)$$

by using the **vectorial formula** for the OLS estimators.

Note: the data are available within the file `Data_CAPM_returns.xlsx`.

Figure: Matlab code for estimating a linear regression model

```
clear , clc , close all

%======================
%== Data importation ===
%======================
[dataset,date]=xlsread('Data_CAPM_returns.xlsx');   % Importation of the data
date=datetime(date(2:end,1));                       % Transformation of the date
r_Intel_ex=dataset(:,1);                            % Excess daily (log-) returns for Intel
r_SP500_ex=dataset(:,2);                            % Excess daily (log-) returns for S&P500

%=================
%== Regression ===
%=================
T=length(r_Intel_ex);        % Sample size (T=250)
X=[ones(T,1) r_SP500_ex];    % Matrix X(250,2)
Y=r_Intel_ex;                % Vector Y(250,1)

disp(' ')
disp(X'*X)
disp(X'*Y)
```

3.3. Applications: CAPM model

Figure: Matlab code for estimating a linear regression model

```
beta=pinv(X'*X)*X'*Y;          % OLS estimates
disp(' OLS estimates')
disp(beta)

residuals=Y-X*beta;            % Residuals
SSR=sum(residuals.^2);         % Sum of squared residuals
sigma2=SSR/(T-size(X,2));      % Estimator of the variance

disp(' SSR')
disp(SSR)

disp('S.E. of the regression')
disp(sqrt(sigma2))
```

3.3. Applications: CAPM model

Figure: Estimation result with Matlab

```
 250.0000    0.1498
   0.1498    0.0156

   0.2889
   0.0242

OLS estimates
   0.0002
   1.5452

SSR
   0.0510

S.E. of the regression
   0.0143
```

3.3. Applications: CAPM model

Example (CAPM)

Question: Use **Excel** to compute the OLS estimates of the parameters β_1, β_2 and σ^2 in the CAPM model

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \varepsilon_t$$

$$\varepsilon_t \ \text{ i.i.d. } \left(0, \sigma^2\right)$$

Note: the data are available within the file `Data_CAPM_returns.xlsx`.

How to run an OLS estimation with Excel?

You can use the data analysis tools in the **Analysis ToolPak**.

1. This is not a standard part of Excel's instillation. To install the ToolPak on your computer, select "Tools" from the menu bar and look for the "Data Analysis" option.

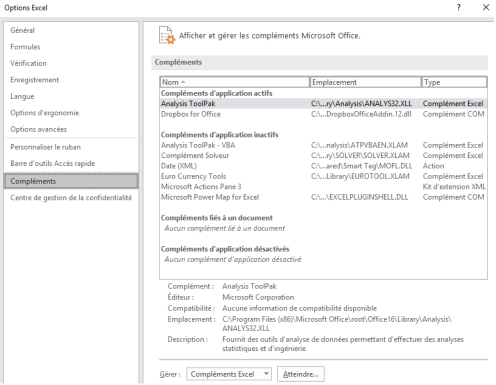
2. If you do not see Data Analysis, select Add-ins from the Tools menu. Check the box for the Analysis ToolPak and click on OK to install them.

3. Once you install the Analysis ToolPak, it will continue to load each time you launch Excel.

Note. For the French version of Excel: Fichier => Options => Compléments => Gérer "compléments Excel"

3.3. Applications: CAPM model

Figure: Installation of the Analysis ToolPak in Excel (French version)

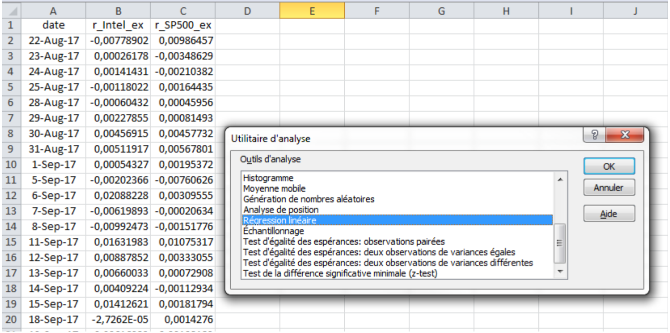


To run a regression with Excel (cont'd):

1. Select **Data Analysis** from the Tools menu, which opens the Data Analysis window.
2. Scroll through the window, select **Regression** from the available options.
3. Place the cursor in the box for **Input** Y **range** and then click and drag over cells for the Y data.
4. Place the cursor in the box for **Input** X **range** and click and drag over cells for the X data.
5. Select the radio button for Output range and click on any empty cell; this is where Excel will place the results.

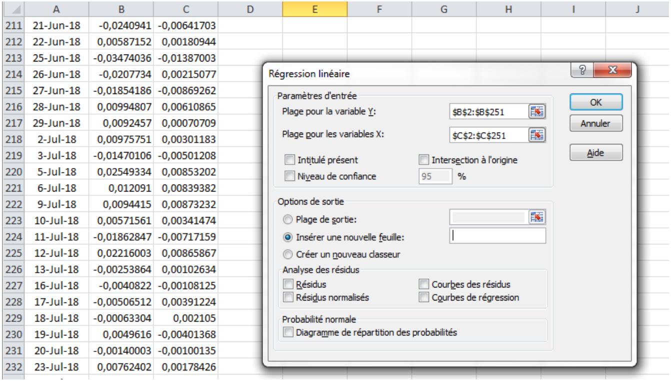
Note. for the French version: Données => Utilitaires d'analyse => Régression linéaire

Figure: Run an OLS regression (step 1)



3.3. Applications: CAPM model

Figure: Run an OLS regression (step 2)



3.3. Applications: CAPM model

Figure: Output from the Excel estimation Tool Pack

RAPPORT DÉTAILLÉ

Statistiques de la régression	
Coefficient de détermination multiple	0,6489074
Coefficient de détermination R^2	0,42108081
Coefficient de détermination R^2	0,41874646
Erreur-type	0,01433559
Observations	250

ANALYSE DE VARIANCE

	Degré de liberté	*Somme des carrés*	*Moyenne des carrés*	*F*	*Valeur critique de F*
Régression	1	0,03707066	0,03707066	180,384488	2,8468E-31
Résidus	248	0,05096626	0,00020551		
Total	249	0,08803692			

	Coefficients	*Erreur-type*	*Statistique t*	*Probabilité*	*Limite inférieure pour seuil de confiance = 95%*	*Limite supérieure pour seuil de confiance = 95%*	*Limite inférieure pour seuil de confiance = 95,0%*	*Limite supérieure pour seuil de confiance = 95,0%*
Constante	0,00022961	0,00090928	0,25251848	0,8008495	-0,00156128	0,0020205	-0,00156128	0,0020205
Variable X 1	1,54515835	0,1150465	13,4307293	2,8468E-31	1,31856557	1,77175113	1,31856557	1,77175113

Example (extended CAPM)

We want to estimate by OLS the parameters of the extended CAPM model given by

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \beta_3 \text{inflation}_t + \varepsilon_t$$

$$\varepsilon_t \quad \text{i.i.d.} \quad \left(0, \sigma^2\right)$$

where $z_{\texttt{intel},t}$ is the excess (log-) return for Intel Corp and $z_{\texttt{market},t}$ is the excess (log-) return for the S&P500. For that we consider a sample of 120 observations. **Question:** Write a Matlab code to compute the OLS estimates of the parameters β_1, β_2, β_3 and σ^2 with the vectorial formula.

Note: the data are available in the file `Data_CAPM_extended.xlsx`.

3.3. Applications: CAPM model

Figure: Matlab code for estimating a linear regression model

```
clear , clc , close all

%======================
%== Data importation ===
%======================
dataset=xlsread('Data_CAPM_extended.xlsx');      % Importation of the data
r_Intel=dataset(:,1);                            % Excess daily (log-) returns for Intel
r_SP500=dataset(:,2);                            % Excess daily (log-) returns for S&P500
inflation=dataset(:,3);                          % Inflation rate

%=================
%== Regression ===
%=================
T=length(r_Intel);                 % Sample size T
X=[ones(T,1) r_SP500 inflation];   % Matrix X
Y=r_Intel;                         % Vector Y

disp(' ')
disp(X'*X)
disp(X'*Y)
```

Figure: Matlab code for estimating a linear regression model

```
beta=pinv(X'*X)*X'*Y;           % OLS estimates
disp(' OLS estimates')
disp(beta)

residuals=Y-X*beta;             % Residuals
SSR=sum(residuals.^2);          % Sum of squared residuals
sigma2=SSR/(T-size(X,2));       % Estimator of the variance

disp(' SSR')
disp(SSR)

disp('S.E. of the regression')
disp(sqrt(sigma2))
```

Figure: Estimation result with Matlab

```
 120.0000    0.0871    0.2330
   0.0871    0.2652    0.0007
   0.2330    0.0007    0.0027

   0.3344
   0.4176
  -0.0027

OLS estimates
   0.0052
   1.5777
  -1.8285

SSR
   0.7764

S.E. of the regression
   0.0815
```

3.3. Applications: CAPM model

Figure: Estimation results obtained with the Matlab function `fitlm`

```
Linear regression model:
    y ~ 1 + x1 + x2

Estimated Coefficients:
                   Estimate        SE        tStat       pValue
                   _________    _________    _______    __________

    (Intercept)    0.0051922    0.0081507    0.63703       0.52535
    x1                1.5777      0.15825     9.9698    2.5606e-17
    x2               -1.8285        1.719    -1.0637       0.28965


Number of observations: 120, Error degrees of freedom: 117
Root Mean Squared Error: 0.0815
R-squared: 0.461,  Adjusted R-Squared 0.452
F-statistic vs. constant model: 50.1, p-value = 1.94e-16
```

Section 4

Statistical Properties of the OLS Estimator

4. Statistical properties of the OLS estimator

Objectives

The objectives of this section are the following**:**

1. Compute the two **first moments** of the (unknown) finite sample distribution of the OLS estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$.
2. Determine the **finite sample distribution** of the OLS estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}$ under assumption A6.
3. Determine the **asymptotic properties** of the OLS estimators.
4. Determine if the OLS estimators are "good": **efficient estimator** versus **BLUE**.

4. Statistical properties of the OLS estimator

In order to study the statistical properties of the OLS estimator, we have to distinguish:

1. The **finite sample properties**
2. The large sample or **asymptotic properties**

Sub-Section 4.1

Finite Sample Properties

4.1. Finite sample properties

Definition (Finite sample properties and finite sample distribution)

The finite sample properties of an estimator $\widehat{\boldsymbol{\beta}}$ correspond to the properties of its **finite sample** distribution (or exact distribution) defined for any sample size $T \in \mathbb{N}$.

Definition (Unbiased estimator)

Under the assumption **A3 (strict exogeneity)**, the OLS estimator $\widehat{\boldsymbol{\beta}}$ is **unbiased**:

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}_0$$

where $\boldsymbol{\beta}_0$ denotes the true value of the vector of parameters. This result holds whether or not the matrix $\mathbf{X}$ is considered as random.

Definition (Variance of the OLS estimator, non-stochastic regressors)

Under the assumption **A4 (spherical error terms)**, the variance covariance matrix of the OLS estimator $\widehat{\boldsymbol{\beta}}$ is

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

where $\mathbf{X}$ is non-stochastic.

Remark

If the matrix $\mathbf{X}$ is **stochastic**, the conditional variance covariance matrix of the OLS estimator $\widehat{\boldsymbol{\beta}}$ is

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}\middle|\mathbf{X}\right)=\sigma^2\ \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}$$

The unconditional variance covariance matrix is equal to

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}\right)=\sigma^2\ \mathbb{E}_X\left(\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\right)$$

where $\mathbb{E}_X$ denotes the expectation with respect to the distribution of $\mathbf{X}$.

Question

How to **estimate** the variance covariance matrix of the OLS estimator?

$$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

Definition (estimator for the variance of errors)

Under the assumption **A3 (exogeneity)**, the estimator $\widehat{\sigma}^2$ defined by

$$\widehat{\sigma}^2 = \frac{1}{T-K} \sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^\top \widehat{\boldsymbol{\varepsilon}}}{T-K}$$

is unbiased

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \sigma^2$$

Note: this result holds whether or not the matrix $\mathbf{X}$ is considered as random.

Definition (Variance estimator)

An **unbiased estimator** of the variance covariance matrix of the OLS estimator is

$$\widehat{\mathbb{V}}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right) = \widehat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$$

where $\widehat{\sigma}^2 = (T-K)^{-1} \widehat{\boldsymbol{\varepsilon}}^\top \widehat{\boldsymbol{\varepsilon}}$ is an unbiased estimator of σ^2. This result holds whether $\mathbf{X}$ is stochastic or non stochastic.

This result is particularly important as it allows to compute the **Standard Error (SE)** associated to the estimators $\widehat{\beta}_k$

$$\underset{(K,K)}{\widehat{\mathbb{V}}\left(\widehat{\boldsymbol{\beta}}\right)} = \begin{pmatrix} \widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_1\right) & \widehat{\mathbb{C}}ov\left(\widehat{\beta}_1, \widehat{\beta}_k\right) & \widehat{\mathbb{C}}ov\left(\widehat{\beta}_1, \widehat{\beta}_K\right) \\ \widehat{\mathbb{C}}ov\left(\widehat{\beta}_k, \widehat{\beta}_1\right) & \widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_k\right) & \\ \widehat{\mathbb{C}}ov\left(\widehat{\beta}_K, \widehat{\beta}_1\right) & \widehat{\mathbb{C}}ov\left(\widehat{\beta}_K, \widehat{\beta}_k\right) & \widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_K\right) \end{pmatrix}$$

$$\mathrm{SE}_{\widehat{\beta}_k} = \sqrt{\widehat{\mathbb{V}}\left(\widehat{\beta}_k\right)} \quad \forall k = 1, \ldots, K$$

Figure: Estimation results obtained with the Matlab function `fitlm`

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                   Estimate         SE         tStat       pValue
                   __________    __________    _______    __________

    (Intercept)    0.00022961    0.00090928    0.25252       0.80085
    x1                 1.5452       0.11505     13.431    2.8468e-31

Number of observations: 250, Error degrees of freedom: 248
Root Mean Squared Error: 0.0143
R-squared: 0.421,  Adjusted R-Squared 0.419
F-statistic vs. constant model: 180, p-value = 2.85e-31
```

Theorem (Linear gaussian regression model)

Under the assumption **A6 (normality)***, the estimators* $\widehat{\boldsymbol{\beta}}$ *and* $\widehat{\sigma}^2$ *have a finite sample distribution given by:*

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}_0, \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)$$

$$\frac{\widehat{\sigma}^2}{\sigma^2}(T-K) \sim \chi^2(T-K)$$

Moreover, $\widehat{\boldsymbol{\beta}}$ *and* $\widehat{\sigma}^2$ *are independent. This result holds whether or not the matrix* $\mathbf{X}$ *is considered as random. In this last case, the distribution of* $\widehat{\boldsymbol{\beta}}$ *is conditional to* $\mathbf{X}$*.*

Question: OLS estimator = "good" estimator of β?

- The question is to know if there this estimator is **preferred** to other unbiased estimators?

$$\mathbb{V}\left(\widehat{\beta}_{OLS}\right) \lesseqgtr \mathbb{V}\left(\widehat{\beta}_{other}\right)$$

- In general, to answer to this question we use the **FDCR or Cramer-Rao bound** and study the **efficiency** of the estimator.

- **Problem:** the computation of the FDCR bound requires an assumption on the distribution of ε

Theorem (Efficiency - Gaussian model)

Under the assumption **A6 (normality)**, *the OLS estimator* $\widehat{\beta}$ *is* **efficient.** *Its variance reaches the FDCR or Cramer-Rao bound:*

$$\mathbb{V}\left(\widehat{\beta}\right) = FDCR\ bound = I_T^{-1}\left(\beta_0\right)$$

This result holds whether or not the matrix $\mathbf{X}$ *is considered as random.*

Problem

1. In a **semi-parametric** model (with no assumption on the distribution of ε), it is impossible to compute the FDCR bound and to show the **efficiency** of the OLS estimator.

2. The solution consists in introducing the concept of best linear unbiased estimator (**BLUE**): the Gauss-Markov theorem.

Theorem (Gauss-Markov theorem)

In the linear regression model under assumptions A1-A5, the least squares estimator $\widehat{\beta}$ is the best linear unbiased estimator **(BLUE)** *of β_0 whether* $\mathbf{X}$ *is stochastic or nonstochastic.*

Comment

The estimator $\widehat{\beta}_k$ for $k = 1, ..K$ is the BLUE of β_k

Best = smallest variance

Linear (in $\mathbf{y}$ or y_i) : $\widehat{\beta}_k = \sum_{i=1}^{T} \omega_{ki} y_i$

Unbiased: $\mathbb{E}\left(\widehat{\beta}_k\right) = \beta_k$

Estimator: $\widehat{\beta}_k = f(y_1, .., y_T)$

4.1. Finite sample properties

Summary

Properties	Assumptions required
$\widehat{\boldsymbol{\beta}}$ is unbiased: $\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}_0$	A3: Exogeneity
$\mathbb{V}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}$	A4: Spherical disturbances
$\widehat{\sigma}^2$ is unbiased: $\mathbb{E}\left(\widehat{\sigma}^2\right) = \sigma^2$	A3 and A4
$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\right)$	A6: Normality
$\frac{\widehat{\sigma}^2}{\sigma^2}\left(T-K\right) \sim \chi^2\left(T-K\right)$	A6: Normality

Summary (cont'd)

Properties	Assumptions required
$\widehat{\beta}$ is **efficient** and **BUE**	A6: Normality
$\widehat{\beta}$ is the **BLUE**	A3,A4

BUE: Best Unbiased Estimator

Sub-Section 4.2

Asymptotic Properties

Question: what is the behavior of the random variable $\widehat{\beta}$ when the sample size T tends to infinity?

Definition (Asymptotic theory)

Asymptotic or **large sample theory** consists in the study of the distribution of the estimator when the sample size is sufficiently large.

The asymptotic theory is fundamentally based on the notion of **convergence**...

Theorem (Consistency)

Under assumptions A1-A5, the OLS estimators $\widehat{\boldsymbol{\beta}}$ *and* $\widehat{\sigma}^2$ *are* (**weakly) consistent**

$$\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$$

$$\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$$

Theorem (Asymptotic distribution)

Under assumptions A1-A5, the OLS estimator $\widehat{\beta}$ is **asymptotically normally** *distributed*

$$\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\overset{d}{\to}\mathcal{N}\left(\mathbf{0},\sigma^2\mathbf{Q}^{-1}\right)$$

where

$$\mathbf{Q}=plim\frac{1}{T}\mathbf{X}^{\top}\mathbf{X}=\mathbb{E}_X\left(\mathbf{x}_i^{\top}\mathbf{x}_i\right)$$

or equivalently

$$\widehat{\beta}\overset{asy}{\approx}\mathcal{N}\left(\beta_0,\frac{\sigma^2}{T}\mathbf{Q}^{-1}\right)$$

4.2. Asymptotic properties

Definition (Estimator of the asymptotic variance matrix)

A consistent estimator of the **asymptotic** variance covariance matrix is given by:

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\beta}}\right)=\widehat{\sigma}^2\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}$$

where $\widehat{\sigma}^2$ is consistent estimator of σ^2 :

$$\widehat{\sigma}^2=\frac{1}{T-K}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2=\frac{\widehat{\boldsymbol{\varepsilon}}^{\top}\widehat{\boldsymbol{\varepsilon}}}{T-K}$$

Asymptotic variance covariance matrix

Even **without** the normality assumption A6, the OLS estimator $\widehat{\boldsymbol{\beta}}$ has a normal distribution as soon T is sufficiently large

$$\widehat{\boldsymbol{\beta}} \overset{asy}{\approx} \mathcal{N}\left(\boldsymbol{\beta}_0, \frac{\sigma^2}{T}\mathbf{Q}^{-1}\right)$$

The estimator of its (asymptotic) variance covariance matrix is always defined as

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\sigma}^2\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}$$

Sub-Section 4.3

Applications to the CAPM model

4.3. Applications: CAPM model

Example (CAPM)

Consider the CAPM model

$$z_{\texttt{intel},t} = \beta_1 + \beta_2 z_{\texttt{market},t} + \varepsilon_t$$

$$\varepsilon_t \quad \text{i.i.d.} \quad \left(0, \sigma^2\right)$$

and a sample of 250 observations from August 22, 2017 to August 17, 2018f. **Question:** Compute the standard errors of the OLS estimator, knowing that

$$SSR = \sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = 0.0510$$

$$\begin{pmatrix} T & \sum_{t=1}^{T} z_{\texttt{market},t} \\ \sum_{t=1}^{T} z_{\texttt{market},t} & \sum_{t=1}^{T} z_{\texttt{market},t}^2 \end{pmatrix} = \begin{pmatrix} 250 & 0.1498 \\ 0.1498 & 0.0156 \end{pmatrix}$$

Note: the data are available within the file `Data_CAPM_returns.xlsx`.

Solution

The linear regression model can be written as

$$\underset{(250,1)}{\mathbf{y}} = \underset{(250,2)}{\mathbf{X}}\underset{(2,1)}{\boldsymbol{\beta}} + \underset{(250,1)}{\boldsymbol{\varepsilon}}$$

with $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, $T = 250$ an,d $K = 2$

$$\underset{(250\times 1)}{\mathbf{y}} = \begin{pmatrix} z_{\texttt{intel},1} \\ z_{\texttt{intel},2} \\ .. \\ z_{\texttt{intel},t} \\ .. \\ z_{\texttt{intel},250} \end{pmatrix} \quad \underset{250\times 2}{\mathbf{X}} = (\mathbf{e} : \mathbf{z}_{\text{market}}) = \begin{pmatrix} 1 & z_{\texttt{market},1} \\ 1 & z_{\texttt{market},2} \\ .. & .. \\ 1 & z_{\texttt{market},t} \\ .. & .. \\ 1 & z_{\texttt{market},250} \end{pmatrix}$$

Solution (cont'd)

The estimator of σ^2 is given by:

$$\widehat{\sigma}^2 = \frac{1}{T-K}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 = \frac{SSR}{T-K} = \frac{0.0510}{250-2} = 2.0551e^{-04}$$

$$\text{S.E. of regression} = \text{RMSE} = \sqrt{\widehat{\sigma}^2} = 0.0143$$

Solution (cont'd)

As there is no information on the distribution of ε, we consider the asymptotic properties of the OLS estimator. A consistent estimator of the **asymptotic** variance covariance matrix is given by:

$$\begin{aligned}
\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}\right) &= \widehat{\sigma}^2 \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \\
&= 2.0551e^{-04} \times \begin{pmatrix} 250 & 0.1498 \\ 0.1498 & 0.0156 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 0.0083e^{-04} & -0.0793e^{-04} \\ -0.0793e^{-04} & 132.3570e^{-04} \end{pmatrix}
\end{aligned}$$

Solution (cont'd)

So we have:

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\boldsymbol{\beta}}\right) = \begin{pmatrix} \widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_1\right) & \widehat{\mathbb{C}}ov_{asy}\left(\widehat{\beta}_1, \widehat{\beta}_2\right) \\ \widehat{\mathbb{C}}ov_{asy}\left(\widehat{\beta}_1, \widehat{\beta}_2\right) & \widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_2\right) \end{pmatrix}$$

$$= \begin{pmatrix} 0.0083e^{-04} & -0.0793e^{-04} \\ -0.0793e^{-04} & 132.3570e^{-04} \end{pmatrix}$$

and the standard errors are equal to

$$\text{SE}_{\widehat{\beta}_1} = \sqrt{\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_1\right)} = \sqrt{0.0083e^{-04}} = 0.0009$$

$$\text{SE}_{\widehat{\beta}_2} = \sqrt{\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_2\right)} = \sqrt{132.3570e^{-04}} = 0.1150$$

4.3. Applications: CAPM model

Figure: Estimation results obtained with the Matlab function `fitlm`

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                   Estimate        SE          tStat       pValue
                  __________    __________    _______    __________

    (Intercept)   0.00022961    0.00090928    0.25252       0.80085
    x1                1.5452       0.11505     13.431    2.8468e-31


Number of observations: 250, Error degrees of freedom: 248
Root Mean Squared Error: 0.0143
R-squared: 0.421,  Adjusted R-Squared 0.419
F-statistic vs. constant model: 180, p-value = 2.85e-31
```

4. Statistical properties of the OLS estimator

Key Concepts

1. The OLS estimator is unbiased under assumption A3 (exogeneity)
2. Mean and variance of the OLS estimator under assumptions A3-A4 (exogeneity - spherical disturbances)
3. BLUE estimator and efficient estimator (FDCR bound)
4. The OLS estimator is weakly consistent
5. The OLS estimator is asymptotically normally distributed
6. Asymptotic variance covariance matrix
7. Estimator of the asymptotic variance

End of Chapter 2

Christophe Hurlin