Financial econometrics

Chapter 4, Multiple Linear Regression

Overview I

Section 1, The multiple linear regression model

Objectives

- Define the (multiple) linear regression model.
- Make a distinction between the semi-parametric and parametric MLR model.
- Introduce the multiple linear Gaussian model.
- 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).
- E.g.

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

- This is called the multiple linear regression model.

Multiple linear regression model (cont'd)

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. A general formulation is given by

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t,2} + \cdots + \beta_k x_{t,K} + \varepsilon_t$$

where y is the dependent variable and $x_1, \ldots, x_K$ are K explanatory variables.

Notation: $x_{t,k}$ is the k-th explanatory variable for time t.

Notation

$$
\begin{aligned}
\underset{T\times 1}{\boldsymbol{y}} &= (y_1, y_2, \cdots, y_t, \cdots, y_T)^\top \\
\underset{T\times 1}{\boldsymbol{x}_k} &= (x_{1,k}, x_{2,k}, \cdots, x_{t,k}, \cdots, x_{T,k})^\top \\
\underset{T\times 1}{\boldsymbol{\varepsilon}} &= (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_t, \cdots, \varepsilon_T)^\top \\
\underset{T\times 1}{\boldsymbol{\beta}} &= (\beta_1, \beta_2, \cdots, \beta_t, \cdots, \beta_T)^\top
\end{aligned}
$$

Notation (cont'd)

$$\underset{T\times K}{\boldsymbol{X}} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k, \cdots, \boldsymbol{x}_K)$$

or equivalently

$$\underset{T\times K}{\boldsymbol{X}} = \begin{pmatrix} x_{1,1} & x_{1,1} & x_{1,3} & \cdots & x_{1,k} & \cdots & x_{1,K} \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,k} & \cdots & x_{2,K} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{t,1} & x_{t,2} & x_{t,3} & \cdots & x_{t,k} & \cdots & x_{t,K} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{T,1} & x_{T,2} & x_{T,3} & \cdots & x_{T,k} & \cdots & x_{T,K} \end{pmatrix}$$

Multiple linear regression model (cont'd)

Definition (Multiple linear regression model)

The multiple linear regression model can be written as

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

Multiple linear regression model (cont'd)

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

- Constant
- Time trend
- Dummies
- etc.

Therefore, $\boldsymbol{X}$ is generally a mixture of random variables and non-random variables

Multiple linear regression model (cont'd)

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

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

The matrix $\boldsymbol{X}$ becomes:

$$\underset{T\times(K+1)}{\boldsymbol{X}} = (\mathbf{1}, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_K)$$

or equivalently

$$\underset{T\times(K+1)}{\boldsymbol{X}} = \begin{pmatrix} 1 & x_{1,2} & x_{1,3} & \cdots & x_{1,k} & \cdots & x_{1,K} \\ 1 & x_{2,2} & x_{2,3} & \cdots & x_{2,k} & \cdots & x_{2,K} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{t,2} & x_{t,3} & \cdots & x_{t,k} & \cdots & x_{t,K} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{T,2} & x_{T,3} & \cdots & x_{T,k} & \cdots & x_{T,K} \end{pmatrix}$$

Example

Example

The CAPM for Intel Corp. can be written as

$$z_{\text{intel},t} = \beta_0 + \beta_1 z_{\text{market},t} + \varepsilon_t,$$

or equivalently

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

Question: what are $\boldsymbol{X}$, $\boldsymbol{y}$, ε , β?

Semi-parametric and semi-parametric specification

One key difference in the specification of the MLR:

- **Parametric model:** the distribution of the error terms is fully characterized, e.g.

$$\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}\right).$$

- **Semi-parametric model:** only a few moments of the error terms are specified, e.g.

$$\mathbb{E}\left[\boldsymbol{\varepsilon}\right] = \mathbf{0} \quad \text{and} \quad \mathbb{V}\left[\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\right] = \boldsymbol{\Omega}.$$

Parametric and semi-parametric specification (cont'd)

This difference does not matter for the derivation of the ordinary least square estimator. But this difference matters for (among others):

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

Parametric and semi-parametric specification (cont'd)

Definition (Semi-Parametric MLR)

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

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

where the error terms ε satisfies

$$\mathbb{E}\left[\varepsilon|\boldsymbol{X}\right] = \boldsymbol{0}$$
$$\mathbb{V}\left[\varepsilon|\boldsymbol{X}\right] = \sigma^2\boldsymbol{I}$$

and $\boldsymbol{I}$ is the identity matrix (of appropriate size).

Remarks:

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

$$\mathbb{E}[\varepsilon] = 0 \qquad \mathbb{V}[\varepsilon] = \sigma^2 \boldsymbol{I}$$

- If the conditional variance-covariance matrix of ε is not diagonal, i.e. if

$$\mathbb{V}[\varepsilon|\boldsymbol{X}] = \boldsymbol{\Omega}$$

  the model is called (multiple) **generalized regression model** (GLM).

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

$$\mathbb{E}\left[\boldsymbol{\varepsilon}|\boldsymbol{X}\right] = \mathbf{0} \qquad \mathbb{V}\left[\boldsymbol{\varepsilon}|\boldsymbol{X}\right] = \sigma^2\boldsymbol{I},$$

are equivalent to:

$$\mathbb{E}\left[\boldsymbol{y}|\boldsymbol{X}\right] = \boldsymbol{X}\boldsymbol{\beta} \qquad \mathbb{V}\left[\boldsymbol{y}|\boldsymbol{X}\right] = \sigma^2\boldsymbol{I}.$$

Parametric and semi-parametric specification (cont'd)

Definition (multiple linear Gaussian model)

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

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

where the error term ε is normally distributed

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

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

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

Remarks:

- The multiple linear Gaussian model is (by definition) a parametric model.
- If the matrix $\boldsymbol{X}$ is non stochastic (fixed), i.e. there are only fixed regressors, then the vector $\boldsymbol{y}$ has marginal normal distribution:

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

Section 2, Assumptions of the MLR

Assumptions of the MLR

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):

A1: Linearity

A2: Full rank condition (or identification)

A3: Exogeneity

A4: Spherical errors

A5: Data generation

A6: Normality

Assumptions 1, linearity

Definition (Linearity)

The model is linear with respect to the parameters $\beta_1, \ldots, \beta_K$ (i.e. $\boldsymbol{\beta}$).

Assumptions 1, linearity (cont'd)

Remarks:

- The models

$$y_t = \beta_0 + \beta_1 x_t + u_t$$
$$y_t = \beta_0 + \beta_1 \cos(x_t) + v_t$$
$$y_t = \beta_0 + \beta_1 \frac{1}{x_t} + \omega_t$$

  are all linear w.r.t. $\boldsymbol{\beta}$.

- The model

$$y_t = \beta_0 + \beta_1 x_t^{\beta_2} + \varepsilon_t$$

  is not linear w.r.t. $\boldsymbol{\beta}$.

- The model

$$y_t = x_t^{\beta} e^{\varepsilon_t}$$

  can turn to linear after appropriate transformation.

Assumptions 2, full column rank

Definition (Full column rank)

$\boldsymbol{X}$ is a $T \times K$ matrix with rank K.

Interpretation:

- There is no exact relationship among any of the independent variables in the model.
- The columns of $\boldsymbol{X}$ are linearly **independent**.
- A matrix $\boldsymbol{X}$ that is *not* full rank is also called **rank deficient**
- If the design matrix $\boldsymbol{X}$ some columns (or rows) can be obtained as a linear combination of the others: if this is the case, we say that there is a (multi)-collinearity problem.

! Remember that a (square) matrix that is rank deficient does not have an inverse.

Assumptions 2, full column rank (cont'd)

- **Perfect** multi-collinearity (that is one variable is linearly dependent from the others) is generally not difficult to spot and is signaled by most statistical software.
- Imperfect multi-collinearity is a more serious issue.

Definition (Imperfect multicollinearity)

Imperfect multicollinearity occurs when two or more explanatory variables in a statistical model are correlated with each other, but not perfectly. I.e. they are not linearly dependent but 'almost' (e.g. one variable has ¿0.9 correlation with another).

Assumptions 2, Example I

Example (Multicollinearity)

Suppose that we want to estimate the following model:

$$z_{\text{intel},t} = \beta_0 + \beta_1 z_{\text{market},t} + \beta_2 \left(z_{\text{market},t} \times 2\right) + \varepsilon_t$$

The identification condition *does not hold*, $z_{\text{market},t}$ *and* $z_{\text{market},t} \times 2$ *are perfectly collinear. It is impossible to estimate* $\boldsymbol{\beta}$*.*

Example (Full rank)

Suppose that we want to estimate the following model:

$$z_{\text{intel},t} = \beta_0 + \beta_1 z_{\text{market},t} + \beta_2 \left(z^2_{\text{market},t}\right) + \varepsilon_t$$

The identification condition *does hold*. No collinearity issues arise in estimating $\boldsymbol{\beta}$.

Assumptions 2, Example II

Example (Imperfect multicollinearity)

Recall that $y = \log(1 + x) \approx x$ for $x \approx 0$. y is a non-linear transformation of x (indeed it involves the log that is non-linear), yet $y \approx x$. Look at the log-return section: by plotting x and y the relationship is pretty much linear around 0: the correlation is high.

Thus x and y and not linearly dependent (y can't be obtained from x as $a + bx$ where a, b are constants), however their correlation is very high: this is called imperfect collinearity.

Assumptions 2, Example III

Example (Questions)

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}, \boldsymbol{z} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

- $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is linearly dependent on $\boldsymbol{x}$?
- $\begin{pmatrix} 1 \\ -8 \end{pmatrix}$ is linearly dependent on $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$?
- $\begin{pmatrix} e^2 \\ -1 \end{pmatrix}$ is linearly dependent on $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$?

Example (Questions)

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \boldsymbol{z} = \begin{pmatrix} 10 \\ 14 \end{pmatrix},$$

Are linearly dependent?

If they are, you can write one vector as a linear combination of the others, e.g.

$$\boldsymbol{z} = a\boldsymbol{x} + b\boldsymbol{y}$$

Here

$$\boldsymbol{z} = 4a + 2b$$

But note that when taking them *two-by-two* they are linearly independent. In fact, you cannot obtain x as neither $a\boldsymbol{y}$ not $b\boldsymbol{z}$. Same for $\boldsymbol{y}$ and $\boldsymbol{z}$.

Linear independence I

Definition (Linear independence)

A set of vector is linearly independent if the only solution to

$$a_1 \boldsymbol{x}_1 + a_2 \boldsymbol{x}_2 + \cdots + a_K \boldsymbol{x}_K = \mathbf{0}$$

is

$$a_1 = a_2 = \cdots = a_K = 0$$

Linear independence II

Example

Form the previous example

$$2\boldsymbol{x} + \boldsymbol{y} - \frac{1}{2}\boldsymbol{z} = \boldsymbol{0}$$

We have that the null vector can be obtained as a linear combination where the coefficients are not-zero: thus $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are not linearly independent.

On the other hand, we claimed that e.g., $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent. In fact

$$a\boldsymbol{x} + b\boldsymbol{y} = \boldsymbol{0}$$

only if both $a = 0$ and $b = 0$.

Identification

Definition (Identification)

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

i. $\text{Rank}\,(\boldsymbol{X}) = K$

ii. The matrix $\boldsymbol{X}^\top \boldsymbol{X}$ is invertible

iii. $\boldsymbol{X}\beta_1 = \mathbf{0}$ implies $\boldsymbol{\beta} = \mathbf{0} \quad \forall \boldsymbol{\beta} \in \mathbb{R}^K$

iv. $\boldsymbol{X}\beta_1 = \boldsymbol{X}\beta_2$ implies $\beta_1 = \beta_2 \quad \forall\,(\beta_1, \beta_2) \in \mathbb{R}^{K \times K}$

Assumptions 3, exogeneity

Strict exogeneity of the regressors

The regressors are **exogenous** if:

$$\mathbb{E}\left[\varepsilon|\boldsymbol{X}\right] = \mathbf{0}$$

or equivalently

$$\mathbb{E}\left[\varepsilon_t|x_{s,k}\right] = 0$$

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

Assumptions 3, exogeneity (cont'd)

Remarks:

- 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).
- The explanatory variables are not predictors of the error terms.
- The strict exogeneity condition can be rewritten as:

$$\mathbb{E}\left[\boldsymbol{y}|\boldsymbol{X}\right] = \boldsymbol{X}\boldsymbol{\beta}$$

Assumptions 4, spherical errors

Spherical errors

The error terms are such that:

$$\mathbb{V}\left[\varepsilon_t|\boldsymbol{X}\right] = \mathbb{E}\left[\varepsilon_t^2|\boldsymbol{X}\right] = \sigma^2 \quad \forall t \in \{1, \ldots, T\}$$

and

$$\mathbb{C}\text{ov}\left(\varepsilon_t, \varepsilon_s|\boldsymbol{X}\right) = \mathbb{E}\left[\varepsilon_t\varepsilon_s|\boldsymbol{X}\right] = \mathbf{0} \quad \forall t \neq s$$

Notes:

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

Assumptions 4, spherical errors (cont'd)

Comments:

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

Assumptions 4, spherical errors (cont'd)

Comments:

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

$$\underset{T\times T}{\mathbb{V}\left[\boldsymbol{\varepsilon}|\boldsymbol{X}\right]} = \underset{T\times T}{\mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\boldsymbol{X}\right]} =$$

$$\begin{pmatrix}
\mathbb{V}\left[\varepsilon_1|\boldsymbol{X}\right] & \mathbb{C}\text{ov}\left(\varepsilon_1\varepsilon_2|\boldsymbol{X}\right) & \cdots & \mathbb{C}\text{ov}\left(\varepsilon_1\varepsilon_t|\boldsymbol{X}\right) & \cdots & \mathbb{C}\text{ov}\left(\varepsilon_1\varepsilon_T|\boldsymbol{X}\right) \\
\mathbb{C}\text{ov}\left(\varepsilon_2\varepsilon_1|\boldsymbol{X}\right) & \mathbb{V}\left[\varepsilon_2|\boldsymbol{X}\right] & \cdots & \mathbb{C}\text{ov}\left(\varepsilon_2\varepsilon_t|\boldsymbol{X}\right) & \cdots & \mathbb{C}\text{ov}\left(\varepsilon_2\varepsilon_T|\boldsymbol{X}\right) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\mathbb{C}\text{ov}\left(\varepsilon_t\varepsilon_1|\boldsymbol{X}\right) & \mathbb{C}\text{ov}\left(\varepsilon_t\varepsilon_2|\boldsymbol{X}\right) & \cdots & \mathbb{V}\left[\varepsilon_t|\boldsymbol{X}\right] & \cdots & \mathbb{C}\text{ov}\left(\varepsilon_t\varepsilon_T|\boldsymbol{X}\right) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\mathbb{C}\text{ov}\left(\varepsilon_T\varepsilon_1|\boldsymbol{X}\right) & \mathbb{C}\text{ov}\left(\varepsilon_T\varepsilon_2|\boldsymbol{X}\right) & \cdots & \mathbb{C}\text{ov}\left(\varepsilon_T\varepsilon_t|\boldsymbol{X}\right) & \cdots & \mathbb{V}\left[\varepsilon_T|\boldsymbol{X}\right]
\end{pmatrix}$$

Assumptions 4, spherical errors (cont'd)

The assumptions of homoscedasticity and non-autocorrelation imply that:

$$\underset{T\times T}{\mathbb{V}\left[\varepsilon|\boldsymbol{X}\right]} = \mathbb{E}\underset{T\times T}{\left[\varepsilon\varepsilon^{\top}|\boldsymbol{X}\right]} =$$

$$\begin{pmatrix} \sigma^2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \cdots & \sigma^2 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \cdots & 0 & \cdots & \sigma^2 \end{pmatrix}$$

Assumptions 4, spherical errors (cont'd)

Notes:

- $\underset{T\times T}{\mathbb{V}\left[\boldsymbol{\varepsilon}|\boldsymbol{X}\right]} = \mathbb{E}\underset{T\times T}{\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top|\boldsymbol{X}\right]} = \sigma^2\boldsymbol{I}$
- homoscedasticity means the 'same variance' for all the error terms

$$\mathbb{V}\left[\varepsilon_1|\boldsymbol{X}\right] = \cdots = \mathbb{V}\left[\varepsilon_T|\boldsymbol{X}\right] = \sigma^2$$

- non-autocorrelation means 'no correlation' for two error terms at two different dates

$$\mathbb{C}\text{ov}\left(\varepsilon_t\varepsilon_s|\boldsymbol{X}\right) = 0 \quad \forall t \neq s$$

Assumption 5, data generating process

Data generation

The data in $(x_{t,1}, x_{t,2}, \ldots, x_{t,K})$ 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.

Assumption 5, data generating process (cont'd)

Comments:

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

Assumption 6, normality

Normality

The data disturbances are normally distributed

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

Question: How about independence or correlations?

Assumption 6, normality (cont'd)

Comments:

- Assumption 6 implies assumption 3 (exogeneity) and 4 (spherical errors):

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

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

- Normality is **not necessary** to obtain most of the results presented in the following, but practical for inference.

Section 3,
The OLS estimator in MLR

OLS estimator

Consider the MLR model

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

or equivalently, for every t

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

Objective: find an estimator of $\boldsymbol{\beta}$ and σ^2 under assumptions A1-A5.

OLS estimator (cont'd)

Three equivalent approaches

1. Minimize the sum of squared residuals (SSR).
2. Use a geometric rationale/interpretation.
3. Solve the minimization problem with matrix notation.

Minimize the sum of squared residuals

As for the simple linear regression, we have

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

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

Minimize the sum of squared residuals (cont'd)

OLS and multiple linear regression model

In the MLR model $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta} + \varepsilon_t$, with $\boldsymbol{x}_t = (x_{t,1}, \dots, x_{t,K})^\top$, the OLS estimator $\hat{\boldsymbol{\beta}}$ is the solution of

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

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

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^\top\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{x}_t y_t\right)$$

Geometric interpretation

- The ordinary least squares estimation methods consist in determining the adjusted vector, $\hat{\boldsymbol{y}}$, which is the closest to $\boldsymbol{y}$ (in a certain space...) such that the squared norm between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ by is minimized.
- Finding $\hat{\boldsymbol{y}}$ by is equivalent to find an estimator of $\boldsymbol{\beta}$.

Geometric interpretation

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

Geometric interpretation (cont'd)

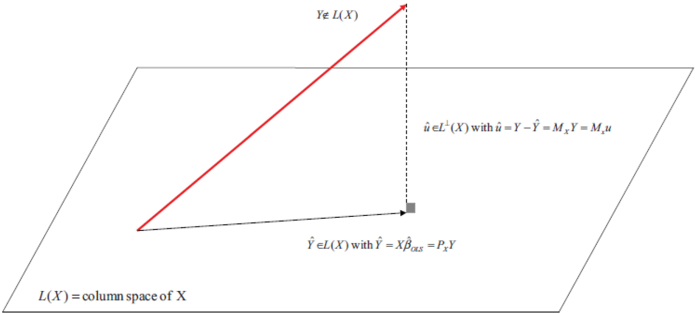

Figure: OLS as a projection.

OLS in matrix notation

OLS and multiple linear regression model

For the MLR model $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$ the OLS estimator $\hat{\boldsymbol{\beta}}$ is the solution of the minimization problem

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \hat{\varepsilon}^{\top}\hat{\varepsilon} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X}\beta)^{\top} (\boldsymbol{y} - \boldsymbol{X}\beta)$$

The **OLS estimator** of β is

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}^{\top}\boldsymbol{y}\right)$$

Derivation of the OLS estimator I

In the multiple regresiion context, the RSS would be minimized w.r.t. all the elements in $\boldsymbol{\beta}$. The RSS is the relevant loss (L) and would be given in matrix notation by

$$L = \hat{\varepsilon}^\top \hat{\varepsilon} = \hat{\varepsilon}_1^2 + \cdots + \hat{\varepsilon}_T^2 = \sum \hat{\varepsilon}_t^2.$$

Denoting the vector of estimated parameters as $\hat{\boldsymbol{\beta}}$, it is possible to write

$$L = \hat{\varepsilon}^\top \hat{\varepsilon} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^\top \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) = \boldsymbol{y}^\top \boldsymbol{y} - 2\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{y} + \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

In order to find the parameter values that minimize loss we differentiate L w.r.t. $\hat{\boldsymbol{\beta}}$ and set it to zero:

$$\frac{\partial L}{\partial \hat{\boldsymbol{\beta}}} = -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \mathbf{0}.$$

Derivation of the OLS estimator II

Rearranging the above gives

$$
\begin{aligned}
2\boldsymbol{X}^\top \boldsymbol{y} &= 2\boldsymbol{X}^\top \boldsymbol{X}\hat{\beta} \\
\boldsymbol{X}^\top \boldsymbol{y} &= \boldsymbol{X}^\top \boldsymbol{X}\hat{\beta},
\end{aligned}
$$

and pre-multiplying both sides by the inverse of $\boldsymbol{X}^\top \boldsymbol{X}$ gives

$$
\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}.
$$

Thus the vector of OLS coefficient stimates for a set of k parameters is given by:

$$
\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k\right)^\top = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}.
$$

Estimate of the variance of the errors

The estimation of the variance of the errors is analogous to the SLR:

- Previously we had
$$s^2 = \frac{1}{T-2}\sum \hat{\varepsilon}_t^2$$
- Under MLR
$$s^2 = \frac{1}{T-K}\sum \hat{\varepsilon}_t^2 = \frac{\hat{\varepsilon}^\top \hat{\varepsilon}}{T-K}$$
- s^2 is unbiased for σ^2.

CAPM example I

Example

Estimate the parameters $\beta_1, \beta_2, \sigma^2$ in the CAPM model

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

We consider a sample of $T = 250$ observations (1 year) from Jan. 3, 2020 to Dec. 29, 2020. We observe

$$\sum z_{\text{market},t} = 14.069 \qquad \sum z^2_{\text{market},t} = 1202.6$$

$$\sum z_{\text{intel},t} = -18.702 \qquad \sum z_{\text{market},t} z_{\text{intel},t} = 1393.6$$

Question: Compute the OLS extimates of the parameters $\hat{\beta}$ and σ^2 with the matrix OLS solution.

CAPM example II

The MLR model writes as

$$\underset{250\times 1}{\boldsymbol{y}} = \underset{250\times 2}{\boldsymbol{X}}\underset{2\times 1}{\boldsymbol{\beta}} + \underset{250\times 1}{\boldsymbol{\varepsilon}}$$

Where:

- $\boldsymbol{y} = (z_{\text{intel},1}, z_{\text{intel},2}, \ldots, z_{\text{intel},250})$
- $\underset{250\times 2}{\boldsymbol{X}} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_{\text{market},1} & z_{\text{market},2} & \cdots & z_{\text{market},250} \end{pmatrix}^\top$
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{250})^\top$
- $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$

CAPM example III

The OLS estimator for $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top y$$

with

$$\underset{2\times 2}{\boldsymbol{X}^\top \boldsymbol{X}} = \begin{pmatrix} T & \sum z_{\text{market},t} \\ \sum z_{\text{market},t} & \sum z^2_{\text{market},t} \end{pmatrix} = \begin{pmatrix} 250 & 14.069 \\ 14.096 & 1202.6 \end{pmatrix}$$

$$\underset{2\times 1}{\boldsymbol{X}^\top \boldsymbol{y}} = \begin{pmatrix} \sum z_{\text{intel},t} \\ \sum z_{\text{intel},t} z_{\text{market},t} \end{pmatrix} = \begin{pmatrix} -18.702 \\ 1393.6 \end{pmatrix}$$

CAPM example IV

The OLS solution is:

$$\underset{2\times 1}{\hat{\boldsymbol{\beta}}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$$= \begin{pmatrix} 250 & 14.069 \\ 14.096 & 1202.6 \end{pmatrix}^{-1} \begin{pmatrix} -18.702 \\ 1393.6 \end{pmatrix} = \begin{pmatrix} -0.1401 \\ 1.1605 \end{pmatrix}$$

$$= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

The estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{T-K} \varepsilon^\top \varepsilon = \frac{1}{T-K} SSR$$

In this example we have $SSR = 1283.9$, then

$$\hat{\sigma}^2 = \frac{1283.9}{248} = 5.1769$$

CAPM example V

and the S.E. of the regression (RMSE) is $\sqrt{\hat{\sigma}^2} = 2.2753$.
Lastly, for variance-covariance matrix:

$$\mathbb{V}\left[\hat{\boldsymbol{\beta}}\right] = \hat{\sigma}^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} = 5.1769 \times \begin{pmatrix} 0.00400 & -0.00004 \\ -0.00004 & 0.00083 \end{pmatrix}$$

$$= \begin{pmatrix} 0.0207 & -0.0002 \\ -0.0002 & 0.0043 \end{pmatrix}$$

Code example

Example

Write a code that computes the OLS estimates based on the above example.

The solution is quite simple:

```
% ... ... ...
% ... ... ...
% ... code to import the data and subtract Rf

X       = [ones(250,1),zm];
B       = pinv(X'*X)*X'*zi;
Y       = X*B;
res     = zi-Y;
s2      = sum(res.^2)/(T-2);
VCov    = s2*pinv(X'*X);
```

Section 4,
Statistical Properties of the OLS estimator

Finite sample properties

Definition (Finite sample properties and finite sample distribution)

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

Unbiasedness

Definition

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

$$\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$$

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

Unbiasedness (cont'd)

Proof:

$$
\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\beta}}\right] &= \mathbb{E}\left[\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)\right] \\
&= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}^\top \boldsymbol{X}\right) \boldsymbol{\beta} + \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \mathbb{E}\left[\boldsymbol{\varepsilon}\right] \\
&= \boldsymbol{I}\boldsymbol{\beta} + \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\varepsilon} | \boldsymbol{x}\right]\right] \\
&= \boldsymbol{\beta} + 0 = \boldsymbol{\beta}
\end{aligned}
$$

Variance of the OLS estimator

Definition

Variance of the OLS estimator Under the assumption A4 (spherical errors) the variance-covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$ is:

$$\mathbb{V}\left[\hat{\boldsymbol{\beta}}\right] = \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1},$$

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

Variance of the OLS estimator (cont'd) I

Proof:
Remember that for a non-vector random variable x, with mean μ_x

$$\mathbb{V}[x] = \mathbb{E}\left[(x - \mu_x)^2\right]$$

If $\boldsymbol{x}$ is a vector-random variable, this is similar, but in place of the square you have $\mathbb{E}\left[(\boldsymbol{x} - \boldsymbol{\mu_x})(\boldsymbol{x} - \boldsymbol{\mu_x})^\top\right]$, and the variance $\mathbb{V}[\boldsymbol{x}]$ is actually a variance-covariance matrix: in position (i, i) of $\mathbb{V}[\boldsymbol{x}]$ you read the value of the variance of x_i (the i-th element of $\boldsymbol{x}$), in position (i, j) you read the covariance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, $\mathbb{C}ov(x_i, x_j)$.

Variance of the OLS estimator (cont'd) II

As above, the variance-covariance for $\hat{\boldsymbol{\beta}}$ is then:

$$\mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^{\top}\right].$$

Given the OLS solution for $\hat{\boldsymbol{\beta}}$ we can state that

$$\hat{\boldsymbol{\beta}}=\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\left(\boldsymbol{X}\beta+\varepsilon\right).$$

Expanding the parenthesis one gets

$$\hat{\boldsymbol{\beta}}=\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{X}\beta+\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\varepsilon$$
$$\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}+\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\varepsilon.$$

Variance of the OLS estimator (cont'd) III

Thus, we express the variance of $\hat{\boldsymbol{\beta}}$ as

$$
\begin{aligned}
&\mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^{\top}\right] \\
=&\mathbb{E}\left[\left(\boldsymbol{\beta}+\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{\varepsilon}-\boldsymbol{\beta}\right)\left(\boldsymbol{\beta}+\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{\varepsilon}-\boldsymbol{\beta}\right)^{\top}\right] \\
=&\mathbb{E}\left[\left(\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{\varepsilon}\right)\left(\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{\varepsilon}\right)^{\top}\right] \\
=&\mathbb{E}\left[\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\right] \\
=&\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\mathbb{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\top}\right]\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1} \\
=&\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\sigma^{2}\boldsymbol{I}\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1} \\
=&\sigma^{2}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{X}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1} \\
=&\sigma^{2}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}
\end{aligned}
$$

Variance of the OLS estimator (cont'd) IV

Therefore we have:

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

Remarks:

- $\sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}$ is a variance-covariance matrix of the coefficients.
  Question: how do we estimate it?
- The diagonal gives the estimated variances of the coefficients:

$$\mathrm{se}\left(\hat{\beta}_k\right) = \sqrt{\left[\hat{\sigma}^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\right]_{kk}} = \sqrt{\hat{\mathbb{V}}\left[\hat{\beta}_k\right]}$$

- The off-diagonal terms give the estimated covariance between the parameter estimates

Variance of the OLS estimator (cont'd)

Remark:

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

$$\mathbb{V}\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] = \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}.$$

The unconditional variance covariance matrix is equal to

$$\mathbb{V}\left[\hat{\boldsymbol{\beta}}\right] = \sigma^2 \mathbb{E}_{\boldsymbol{X}}\left[\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\right].$$

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

Finite sample distribution of the estimators

Theorem (Finite sample distribution of $\hat{\boldsymbol{\beta}}$ and σ^2)

Under the assumption A6 (normality), the estimators $\hat{\boldsymbol{\beta}}$ and σ^2 have finite sample distributions given by:

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

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

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

Finite sample distribution of the estimators (cont'd)

Note: as a result,

$$\mathbb{E}\left[\hat{\sigma}^2\right] = \sigma^2, \qquad \mathbb{V}\left[\hat{\sigma}^2\right] = \frac{2\sigma^4}{T-K}.$$

Proof, first part

- The first part is quite trivial. In $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ the stochastic component is entirely embedded in $\boldsymbol{\varepsilon} \sim \mathcal{N}$, thus the distribution of $\boldsymbol{y}$ is Normal as well. Similarly, $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ is normal as well. The mean and var-cov matrix of such distribution have already been computed and shown to be respectively $\boldsymbol{\beta}$ and $\sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}$.
- **Note:** Without A6 the mean and variance of the sampling distribution of $\hat{\boldsymbol{\beta}}$ are known, but now, under A6, we are able to characterize its distribution.

Good estimator

Question: what is a 'good' estimator of β?

- The question is to know if there this estimator is preferred to other unbiased estimators, i.e.

$$\mathbb{V}\left[\beta_{OLS}\right] < \mathbb{V}\left[\beta_{other}\right] ?$$

- This answer this question one has to use Cramer-Rao Bound (CRB) and study the efficiency of the estimator.
- The computation of the CRB is based on likelihood theory and as such requires assumptions about the distribution of ε, otherwise is just impossible to compute it.

Efficiency I

Theorem (Efficiency of the Gaussian MLR model)

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

$$\mathbb{V}\left[\hat{\beta}\right] = CRB$$

Remark: The CRB expresses a lower bound on the variance of unbiased estimators of a fixed though unknown parameter, stating that the variance of *any* such estimator is at least as high the CRB. An unbiased estimator which achieves this lower bound is said to be efficient.

Remark: As a consequence, there is no other estimator with lower variance than the CRB. In this view, β_{OLS} is the best choice as it reaches the CRB and no other unbiased estimator can have smaller variance.

Efficiency and BLUE

Problem:

- In a semi-parametric model (with no assumption on the distribution of ε), it is impossible to compute the CRB and to show the efficiency of the OLS estimator.
- 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 β is the best linear unbiased estimator (BLUE) of whether $\boldsymbol{X}$ is stochastic or nonstochastic.

Summary

Property	Assumption
$\hat{\beta}$ is unbiased	A3: Exogeneity
$\mathbb{V}\left[\hat{\beta}\right] = \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}$	A4: Sph. Errors
σ^2 is unbiased	A3 and A4
$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\right)$	A6: Normality
$\frac{\hat{\sigma}^2}{\sigma^2}(T-K) \sim \chi^2_{(T-K)}$	A6: Normality

Summary (cont'd)

Property	Assumptions
$\hat{\beta}$ is BLUE	$+$A3, & $+$A4
$\hat{\beta}$ is efficient and BLUE	$+$ A3, $+$ A4 & $+$A6 (Normality)

(In addition to, A1, A2, A5)

Section 5, Asymptotic Properties

Asymptotic Properties

Question: what is the behavior of the random variable $\hat{\beta}$ when the sample size 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**...

Consistency

Definition (Consistency)

An estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ if

$$\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}.$$

Limit distribution of the OLS

Theorem

Under assumption A1-A5, the OLS estimator $\hat{\beta}$ is **asymptoptically normally** *distributed*

$$\sqrt{N}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{Q}^{-1}\right)$$

where

$$\boldsymbol{Q} = \text{plim}\,\frac{1}{N}\boldsymbol{X}^\top\boldsymbol{X} = \mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{x}_i^\top \boldsymbol{x}_i\right]$$

Equivalently, asymptotically:

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{N}\boldsymbol{Q}^{-1}\right)$$

Remark:
The asymptotic variance is consistently estimated by the estimated variance matrix

$$\hat{\mathbb{V}}\left[\hat{\boldsymbol{\beta}}\right] = s^2\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}$$

where s^2 is consistent for σ^2.
For example:

$$s^2 = \frac{1}{N-K}\hat{\boldsymbol{\varepsilon}}^\top\hat{\boldsymbol{\varepsilon}}$$

or

$$s^2 = \frac{1}{N}\hat{\boldsymbol{\varepsilon}}^\top\hat{\boldsymbol{\varepsilon}}$$

as asymptotically the correction for the number of estimated parameters is irrelevant.

Section 6, References

Slides

Disclaimer:

- Some slides from Christophe Hurlin's (University of Orleans), financial econometrics course (2019), available online.
- Some slides original (made ad-hoc for this course).
- Some slides from Colin Cameron's (University of California, Davis) lecture notes.
- Some slides based on (Hayashi, 2011, Ch. 1).

Bibliography I

[Hay11] F. Hayashi. *Econometrics*. Princeton University Press, 2011. ISBN: 9781400823833. URL: https://books.google.dk/books?id=QyIW8WUIyzcC.