Lecture 3-c
Least Squares - Properties, Testing and Goodness of Fit

Brooks (4th edition): Chapters 3 & 4

 1

Review – LS Estimation with Linear Algebra

- Model (with linear algebra notation):

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

- Vectors will be column vectors: $\boldsymbol{y}$, $\boldsymbol{x}_j$, and $\boldsymbol{\varepsilon}$ are Tx1 vectors:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{y}' = [y_1\; y_2\; ....\; y_T]$$

$$\boldsymbol{x}_j = \begin{bmatrix} x_{j1} \\ \vdots \\ x_{jT} \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{x}_j' = [x_{j1}\; x_{j2}\; ....\; x_{jT}]$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{\varepsilon}' = [\varepsilon_1\; \varepsilon_2\; ....\; \varepsilon_T]$$

$\mathbf{X}$ is a Txk matrix. $\Rightarrow$ $\mathbf{X} = [\boldsymbol{x}_1\; \boldsymbol{x}_2\; ....\; \boldsymbol{x}_k]$

Review – LS Estimation with Linear Algebra

- Using linear algebra notation: $\boldsymbol{y} = \mathbf{X}\,\beta + \varepsilon$

$\mathbf{X}$ is a Txk matrix. $\Rightarrow$ $\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix}$

$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$ (a kx1 vector)

- The whole system (for all i) is:

$$y_1 = \beta_1 x_{11} + \beta_2\, x_{12} + \ldots + \beta_k x_{k1} + \varepsilon_1$$
$$y_2 = \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_k x_{k2} + \varepsilon_2$$
$$\ldots. \quad \ldots. \quad \ldots. \quad \ldots$$
$$y_T = \beta_1 x_{1T} + \beta_2 x_{2T} + \ldots + \beta_k x_{kT} + \varepsilon_T$$

Review – LS Estimation with Linear Algebra

- Assume *functional form*, $f(\mathbf{X}, \theta)$, is linear:

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

- LS Objective function: $S(x_i, \beta) = \Sigma_i\, \varepsilon_i^2 = \varepsilon'\varepsilon = (\boldsymbol{y} - \mathbf{X}\beta)'\,(\boldsymbol{y} - \mathbf{X}\beta)$
$= \boldsymbol{y}'\boldsymbol{y} - 2\,\beta'\mathbf{X}'\mathrm{y} + \beta'\mathbf{X}'\mathbf{X}\beta$

- First derivative w.r.t. β': $-2\,\mathbf{X}'\boldsymbol{y} + 2\,\mathbf{X}'\mathbf{X}\,\beta$ (a kx1 vector)

- F.o.c. (normal equations): $\mathbf{X}'\boldsymbol{y} - (\mathbf{X}'\mathbf{X})\,\mathbf{b} = \mathbf{0}$ $\Rightarrow (\mathbf{X}'\mathbf{X})\,\mathbf{b} = \mathbf{X}'\boldsymbol{y}$

- Assuming $(\mathbf{X}'\mathbf{X})$ is non-singular –i.e., invertible-, we solve for **b**:
$\Rightarrow \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{y}$ (a kx1 vector)

Note: **b** is called the **Ordinary Least Squares** (OLS) estimator.
(*Ordinary* = $f(\mathbf{X}, \theta)$ is linear.)

Review – LS Estimation with Linear Algebra

• OLS estimator: $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{y}$ (a kx1 vector)

• To derive **b**, we have made few assumptions:

1. Model is linear -i.e., $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
2. The explanatory variables are independent –i.e., rank(**X**) = k.

• To get properties for the OLS estimator, **b**, we need assumptions about **ε** (its mean and variance-covariance matrix) and about how **ε** relates to **X**.

Review – Rules for Expectations of a RV

• Let X denote a *discrete* RV with probability function $p(x)$, then the expected value of X, $E[X]$, is defined to be:

$$\text{E}[X] = \textstyle\sum_i \; x_i\, p(x_i)$$

and if X is *continuous* with probability density function $f(x)$:

$$\text{E}[X] = \int_{-\infty}^{\infty} x\, f(x) dx$$

• Rules:

- **Rule 1.** $\text{E}[c] = c$, where c is a constant.
- **Rule 2.** $\text{E}[c + d\,X] = c + d\,\text{E}[X]$, where c & d are constants.
- **Rule 3.** $\text{Var}[X] = \mu_2^0 = E[(X-\mu)^2] = E[X^2] - [E(X)]^2 = \mu_2 - \mu^2$
- **Rule 4.** $\text{Var}[a\,X + b] = a^2\,\text{Var}[X]$

Review – Rules for Expectations: Linear Model

- Suppose excess returns for asset i, $r_{i,t} - r_f$, are driven by the following linear model (DGP behind the CAPM):

$$r_{i,t} - r_f = \alpha_i + \beta_i (r_{m,t} - r_f) + \varepsilon_{i,t},$$

where

$r_{m,t} - r_f$ = excess return on the market portfolio at time t.

$\varepsilon_{i,t}$ = idiosyncratic error term, with mean 0 & unrelated to $r_{m,t}$.

Then,

$\text{E}[r_i - r_f] = \text{E}[\alpha_i] + \beta_i \text{E}[r_{m,t} - r_f] + \text{E}[\varepsilon_{i,t}]$ (by **Rule 2**)

$\text{E}[r_i - r_f] = \alpha_i + \beta_i \text{E}[r_{m,t} - r_f] + \text{E}[\varepsilon_{i,t}]$ (by **Rule 1**)

$\text{E}[r_i - r_f] = \alpha_i + \beta_i \text{E}[r_{m,t} - r_f]$ -by mean 0 of $\varepsilon_{i,t}$

Also, by **Rule 4** (& assuming $\text{Cov}[r_{m,t} - r_f, \varepsilon_{i,t}] = 0$):

$$\text{Var}[r_i - r_f] = \beta_i^2 \text{Var}[r_{m,t} - r_f] + \text{Var}[\varepsilon_{i,t}]$$

OLS – Assumptions

- Typical OLS Assumptions

(1) DGP: $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i$, $i = 1, 2, \ldots, T$

⇒ functional form known, but β is unknown.

(2) $\text{E}[\varepsilon_i] = 0$. ⇒ expected value (mean) of the errors is 0.

(3) Explanatory variables $X_1, X_2, \ldots, X_k$, are given (& non random)

⇒ no correlation with ε ($\text{Cov}(\varepsilon_i, X_j) = 0$.)

(4) The k explanatory variables are independent.

(5) $\text{Var}[\varepsilon_i] = \text{E}[\varepsilon_i^2] = \sigma^2 < \infty$. (homoscedasticity = same variance)

(6) $\text{Cov}(\varepsilon_i, \varepsilon_j) = \text{E}[\varepsilon_i \varepsilon_j] = 0$. (no serial/cross correlation)

- These are the assumptions behind the *classical linear regression model* (CLM).

LS – Assumptions with Linear Algebra Notation

• We rewrite the assumptions using linear algebra. We condition on **X**, which allows **X** to be a random variable (though, once we condition, **X** becomes a matrix of numbers):

(**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (linear model) is correctly specified.

(**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$

(**A4**) **X** has full column rank –rank(**X**)= k–, where $T \geq k$.

• Assumption (**A1**) is called *correct specification*. We know how the data is generated. We call $\boldsymbol{y} = f(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ the Data Generating Process.

Note: The errors, $\boldsymbol{\varepsilon}$, are called *disturbances*. They are not something we add to $f(\mathbf{X}, \boldsymbol{\theta})$ because we don't know precisely $f(\mathbf{X}, \boldsymbol{\theta})$. No. The errors are part of the DGP.

LS – Assumptions with Linear Algebra Notation

• Assumption (**A2**) is called *regression*.

From Assumption (**A2**) we get:

(i) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0 \quad \Rightarrow E[\boldsymbol{y}\,|\mathbf{X}] = E[\mathbf{X}\,\boldsymbol{\beta}|\mathbf{X}] + E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{X}\,\boldsymbol{\beta}$

That is, the observed **y** will equal $E[\mathbf{y}|\mathbf{X}]$ + random variation.

(ii) Using rules of expectations (law of iterated expectations), we get:

(1) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0 \quad \Rightarrow E[\boldsymbol{\varepsilon}] = 0$

$\Rightarrow$ The conditional expectation = unconditional expectation

(2) $Cov(\boldsymbol{\varepsilon}, \mathbf{X}) = E[(\boldsymbol{\varepsilon} - 0)(\mathbf{X} - \mu_X)] = E[\boldsymbol{\varepsilon}\mathbf{X} - \boldsymbol{\varepsilon}\,\mu_X]$
$= E[\boldsymbol{\varepsilon}\mathbf{X}] - \mu_X\,E[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\mathbf{X}] = 0$

$\Rightarrow$ That is, $E[\boldsymbol{\varepsilon}\mathbf{X}] = 0 \quad \Rightarrow \boldsymbol{\varepsilon} \perp \mathbf{X}$.

There is no information about $\boldsymbol{\varepsilon}$ in **X** and vice-versa.

LS – Assumptions with Linear Algebra Notation

• Assumption (**A3**) gives the model a constant variance for all errors and no relation between the errors at different measurements/times. That is, we have a diagonal variance-covariance matrix:

$$\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\, \mathbf{I}_T \qquad (T\text{x}T) \text{ matrix}$$

This assumption implies

(i) *homoscedasticity* $\Rightarrow \text{E}[\varepsilon_i^2|\mathbf{X}] = \sigma^2$ for all i.

(ii) *no serial/cross correlation* $\Rightarrow \text{E}[\varepsilon_i\, \varepsilon_j|\mathbf{X}] = 0$ for $i \neq j$.

It can be shown using the law of total variance that

$$\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \sigma^2\mathbf{I}_T \qquad \Rightarrow \text{Var}[\boldsymbol{\varepsilon}] = \sigma^2\, \mathbf{I}_T$$

LS – Assumptions with Linear Algebra Notation

Note: $\text{Var}[\boldsymbol{\varepsilon}|\mathbf{X}] = \text{E}[(\boldsymbol{\varepsilon} - \text{E}[\boldsymbol{\varepsilon}])\,(\boldsymbol{\varepsilon} - \text{E}[\boldsymbol{\varepsilon}])'|\mathbf{X}]$ $\qquad (T\text{x}T)$ matrix

$$= \text{E}[(\boldsymbol{\varepsilon} - \mathbf{0})\,(\boldsymbol{\varepsilon} - \mathbf{0})'|\mathbf{X}]$$

$$= \text{E}[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}]$$

$$= \begin{bmatrix} \text{E}[\varepsilon_1^2|\mathbf{X}] & \text{E}[\varepsilon_2\,\varepsilon_1|\mathbf{X}] & \cdots & \text{E}[\varepsilon_T\,\varepsilon_1|\mathbf{X}] \\ \text{E}[\varepsilon_1\,\varepsilon_2|\mathbf{X}] & \text{E}[\varepsilon_2^2|\mathbf{X}] & \cdots & \text{E}[\varepsilon_T\,\varepsilon_2|\mathbf{X}] \\ \vdots & \vdots & \vdots & \vdots \\ \text{E}[\varepsilon_1\,\varepsilon_T|\mathbf{X}] & \text{E}[\varepsilon_2\,\varepsilon_T|\mathbf{X}] & \cdots & \text{E}[\varepsilon_T^2|\mathbf{X}] \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\, \mathbf{I}_T$$

LS – Assumptions with Linear Algebra Notation

• From Assumption (**A4**) ⇒ the k independent variables in **X** are linearly independent. Then, the kxk matrix **X'X** will also have full rank –i.e., rank(**X'X**) = k.

⇒ **X'X** is invertible.

We need this result to solve a system of equations given by the 1st-order conditions of LS Estimation (normal equations):

$$\mathbf{X}'\boldsymbol{y} - \mathbf{b}\ \mathbf{X}'\mathbf{X} = 0$$

Note: To get asymptotic results we need more assumptions about **X**.

CLM: OLS – Summary

• *Classical linear regression model* (CLM) - Assumptions:

(**A1**) DGP: $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.

(**A2**) $E[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$

(**A3**) $Var[\boldsymbol{\varepsilon}|\mathbf{X}] = E[(\boldsymbol{\varepsilon} - \mathbf{0})\,(\boldsymbol{\varepsilon} - \mathbf{0})'|\mathbf{X}] = E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\,\mathbf{I}_T$

(**A4**) **X** has full column rank –rank(**X**) = k, where T ≥ k.

• OLS estimator: bjective function:

$$S(x_i, \boldsymbol{\beta}) = \Sigma_{i=1}^{n}\varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})'\,(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta})$$
$$= \boldsymbol{y}'\boldsymbol{y} - 2\,\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

• f.o.c: $\quad - 2\,\mathbf{X}'\boldsymbol{y} + 2\,\mathbf{X}'\mathbf{X}\,\mathbf{b} = 0$

Solving for **b** $\quad \Rightarrow \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{y} \qquad$ (kx1) vector

Q: Is **b** a minimum? Check second order condition!

OLS Estimation: Second Order Condition

$$\frac{\partial^2 S(x_i, \beta)}{\partial b \partial b'} = 2\boldsymbol{X'X} = 2\begin{bmatrix} \Sigma_{i=1}^n x_{i1}^2 & \Sigma_{i=1}^n x_{i1}x_{i2} & \cdots & \Sigma_{i=1}^n x_{i1}x_{iK} \\ \Sigma_{i=1}^n x_{i2}x_{i1} & \Sigma_{i=1}^n x_{i2}^2 & \cdots & \Sigma_{i=1}^n x_{i2}x_{iK} \\ \cdots & \cdots & \cdots & \cdots \\ \Sigma_{i=1}^n x_{iK}x_{i1} & \Sigma_{i=1}^n x_{iK}x_{i2} & \cdots & \Sigma_{i=1}^n x_{iK}^2 \end{bmatrix}$$

If there were a single **b**, we would require this to be positive, which it would be: $2\,\boldsymbol{x'x} = 2\sum_{i=1}^n x_i^2 > 0$.

The matrix counterpart of a positive number is a *positive definite* (pd) matrix. We need **X'X** to be pd.

A square matrix (*m*x*m*) **A** "takes the sign" of the *quadratic form,* **z'A z**, where **z** is an *m*x1 vector. Then, **z'A z** is a scalar.

OLS Estimation: Second Order Condition

• A form is a polynomial expression in which each component term has a uniform degree. A quadratic form has a uniform 2^{nd} degree.

Examples:

$9\,x + 3\,y + 2\,z$ - 1st degree form.

$6\,x^2 + 2\,x\,y + 2\,y^2$ - 2nd degree (quadratic) form.

$d^2z = f_{xx}\,dx^2 + 2\,f_{xy}dx\,dy + f_{yy}\,dy^2$ (quadratic form)

A quadratic form can be written in matrix notation as **z'A z** , where **A** is (*m*x*m*) **A** and **z** is an *m*x1 vector. Then, **z'A z** is a scalar.

The above quadratic form can be written as:

$$q = [x \;\; y] * \begin{bmatrix} 6 & 1 \\ 1 & 2 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} = 6\,x^2 + 2\,x\,y + 2\,y^2$$

Once we have x & y, q is a number.

OLS Estimation: Second Order Condition

• Let q be a quadratic form. We say q is:
  Positive definite (pd) if q is invariably positive ($q > 0$)
  Positive semi-definite (psd) if q is invariably non-negative ($q \geq 0$)
  Negative semi-definite (nsd) if q is invariably non-positive ($q \leq 0$)
  Negative definite (nd) if q is invariably negative ($q < 0$)

Definition: Positive definite matrix

A matrix **A** is *positive definite* (pd) if $\mathbf{z}'\mathbf{A}\,\mathbf{z} > 0$ for *any* **z** (a kx1 vector).

For some matrices, it is easy to check. Let $\mathbf{A} = \mathbf{X}'\mathbf{X}$ (a kxk matrix).

Then, $\mathbf{z}'\mathbf{A}\,\mathbf{z} = \mathbf{z}'\mathbf{X}'\mathbf{X}\,\mathbf{z} = \boldsymbol{v}'\boldsymbol{v} = \sum_{i=1}^{T} v_i^2 > 0.$ ($\boldsymbol{v} = \mathbf{X}\mathbf{z}$ is Tx1)

$\Rightarrow$ **X'X** is pd $\Rightarrow$ **b** is a min!

Technical note 1: In general, we need eigenvalues of **A** to check this. If all the eigenvalues are positive, then **A** is pd.

Technical note 2: If **A** is pd, then $\mathbf{A}^{-1}$ is also pd.

OLS Estimation: Second Order Condition

• Loosely speaking, a matrix is positive definite if the diagonal elements are positive and the off-diagonal elements are not too large in absolute value relative to the diagonal elements.

Remark: This is an informal way of looking at a pd matrix, but, keep in mind for later, that the **diagonal elements are positive**.

OLS Estimation – Properties of b

- The OLS estimator of $\boldsymbol{\beta}$ in the CLM is

$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\mathbf{y}$ $\quad\Rightarrow$ **b** is a (linear) function of the data $(y_i,\ x_i)$.

$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\mathbf{y} = (\mathbf{X'X})^{-1}\,\mathbf{X'}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}$

$\Rightarrow \mathbf{b} - \boldsymbol{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}$

Under the typical assumptions, we can establish properties for **b**.

1) $\text{E}[\mathbf{b}|\mathbf{X}] = \text{E}[\boldsymbol{\beta}|\mathbf{X}] + \text{E}[(\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}|\mathbf{X}]$

$= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\,\text{E}[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\beta}$ (**b** is *unbiased*)

2) $\text{Var}[\mathbf{b}|\mathbf{X}] = \text{E}[(\mathbf{b} - \boldsymbol{\beta})\,(\mathbf{b} - \boldsymbol{\beta})'|\mathbf{X}] = \text{E}[(\mathbf{X'X})^{-1}\,\mathbf{X'}\,\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'\,\mathbf{X}(\mathbf{X'X})^{-1}|\mathbf{X}]$

$= (\mathbf{X'X})^{-1}\,\mathbf{X'}\,\text{E}[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}]\,\mathbf{X}(\mathbf{X'X})^{-1}$

$= (\mathbf{X'X})^{-1}\,\mathbf{X'}\,\{\sigma^2\,\mathbf{I}_T\}\,\mathbf{X}(\mathbf{X'X})^{-1} = \sigma^2\,(\mathbf{X'X})^{-1}\,\mathbf{X'X}\,(\mathbf{X'X})^{-1}$

$= \sigma^2\,(\mathbf{X'X})^{-1}$ $\quad(k\text{x}k)$ matrix

OLS Estimation – Properties of b

3) **Gauss-Markov Theorem**: **b** is BLUE (*Best Linear Unbiased Estimator*). No other linear & unbiased estimator has a lower variance.

4) If we also assume: (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim$ *i.i.d.* $\text{N}(\mathbf{0}, \sigma^2\,\mathbf{I}_T)$,

we derive the distribution of **b**:

$\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}$ $\quad\Rightarrow$ **b** is a linear combination of normal variates

$\Rightarrow \mathbf{b}|\mathbf{X} \sim$ *i.i.d.* $\text{N}(\boldsymbol{\beta}, \sigma^2\,(\mathbf{X'X})^{-1})$

$\text{SD}[\mathbf{b}|\mathbf{X}] = \text{sqrt(diagonal elements of } \sigma^2\,(\mathbf{X'X})^{-1})$

<u>Note</u>: The marginal distribution of a multivariate normal is also normal, then

$\text{b}_k|\mathbf{X} \sim \text{N}(\beta_k, v_{b,k}^2)$

$\text{Std Dev}\,[\text{b}_k|\mathbf{X}] = \text{sqrt}\{[\sigma^2(\mathbf{X'X})^{-1}]_{kk}\} = v_{b,k}$

<u>Remark</u>: With (**A5**) we can do tests of hypothesis.

OLS Estimation – Properties of b

5) If (**A5**) is not assumed, we still can obtain a (limiting) distribution for **b**. Under additional assumptions –mainly, the matrix **X'X** does not explode as T becomes large–, as $T \rightarrow \infty$,

(i) $\mathbf{b} \xrightarrow{p} \boldsymbol{\beta}$ (**b** is consistent)

(ii) $\mathbf{b} \xrightarrow{a} N(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$ (**b** is asymptotically normal)

• Properties (1)-(4) are called *finite* (or *small)* sample properties, they hold for every sample size.

• Properties (5.i) and (5.ii) are called *asymptotic* properties, they only hold when T is large (actually, as T tends to ∞).

Property (5.ii) is very important: When the errors are not normally distributed we still can do testing about $\boldsymbol{\beta}$, but we rely on an "approximate distribution."

OLS Estimation – Fitted Values and Residuals

• OLS estimates $\boldsymbol{\beta}$ with **b.** Now, we define *fitted values* as:

$\hat{\boldsymbol{y}} = \mathbf{X}\,\mathbf{b}$ (what we expect $\boldsymbol{y}$ to be, given observed **X**)

Now we define the estimated error, $\boldsymbol{e}$:

$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$

$\boldsymbol{e}$ represents the unexplained part of **y**, what the regression cannot explain. They are usually called *residuals*.

Note that **e** is uncorrelated (orthogonal) with **X** $\Rightarrow \boldsymbol{\varepsilon} \perp \mathbf{X}$

$\boldsymbol{e} = \boldsymbol{y} - \mathbf{X}\mathbf{b} \quad \Rightarrow \mathbf{X}'\boldsymbol{e} = \mathbf{X}'(\boldsymbol{y} - \mathbf{X}\mathbf{b}) = \mathbf{X}'\boldsymbol{y} - \mathbf{X}'\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\boldsymbol{y} = \mathbf{0}$

Using **e**, we can define a measure of unexplained variation:

Residual Sum of Squares (RSS) = $\boldsymbol{e}'\boldsymbol{e} = \sum_i {e_i}^2$

OLS Estimation – Var[b | X]

We use the variance to measure precision of estimates. For OLS:

$\text{Var}[\mathbf{b}|\mathbf{X}] = \sigma^2 (\mathbf{X'X})^{-1}$

Example: One explanatory variable model.

(**A1'**) DGP: $\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{x} + \boldsymbol{\varepsilon}$

$$\text{Var}[\mathbf{b}|\mathbf{X}] = \sigma^2 (\mathbf{X'X})^{-1} = \sigma^2 \begin{bmatrix} \sum_i 1 & \sum_i 1x_i \\ \sum_i 1x_i & \sum_i x_i^2 \end{bmatrix}^{-1} = \sigma^2 \begin{bmatrix} T & T\bar{x} \\ T\bar{x} & \sum_i x_i^2 \end{bmatrix}^{-1}$$

$$= \sigma^2 \frac{1}{T(\sum_i x_i^2 - T\bar{x}^2)} \begin{bmatrix} \sum_i x_i^2 & -T\bar{x} \\ -T\bar{x} & T \end{bmatrix}$$

$$\text{Var}[b_1|\mathbf{X}] = \sigma^2 \frac{\sum_i x_i^2}{T(\sum_i x_i^2 - T\bar{x}^2)} = \sigma^2 \frac{\sum_i x_i^2/T}{\sum_i (x_i - \bar{x})^2}$$

$$\text{Var}[b_2|\mathbf{X}] = \sigma^2 \frac{1}{(\sum_i x_i^2 - T\bar{x}^2)} = \sigma^2 \frac{1}{\sum_i (x_i - \bar{x})^2}$$

OLS Estimation – Var[b | X]

Example (continuation):

$$\text{Var}[b_1|\mathbf{X}] = \sigma^2 \frac{\sum_i x_i^2}{T(\sum_i x_i^2 - T\bar{x}^2)} = \sigma^2 \frac{\sum_i x_i^2/T}{\sum_i (x_i - \bar{x})^2}$$

$$\text{Var}[b_2|\mathbf{X}] = \sigma^2 \frac{1}{(\sum_i x_i^2 - T\bar{x}^2)} = \sigma^2 \frac{1}{\sum_i (x_i - \bar{x})^2}$$

$$\text{Covar}[b_1, b_2|\mathbf{X}] = \sigma^2 \frac{-\bar{x}}{\sum_i (x_i - \bar{x})^2}$$

- In general, we do not know σ^2. It needs to be estimated. We estimate σ^2 using the residual sum of squares (RSS):

$\text{RSS} = \sum_i {e_i}^2$

The natural estimator of σ^2 is $\hat{\sigma}^2 = \text{RSS}/T$. Given the LLN, this is a consistent estimator of σ^2. However, this not unbiased.

OLS Estimation – Var[b|X]

• The unbiased estimator of σ^2 is s^2:

$$s^2 = \text{RSS}/(T-k) = \sum_i e_i^{\,2}/(T-k) = \boldsymbol{e}'\boldsymbol{e}/(T-k)$$

To get E[s^2], we use a property of a RV with a χ^2_υ distribution:

$$\text{E}[\chi^2_\upsilon] = \upsilon$$

We know $(T-k)\, s^2/\sigma^2 \sim \chi^2_{T-k}$.

$\Rightarrow$ $\text{E}[\boldsymbol{e}'\boldsymbol{e}/\sigma^2|\mathbf{X}] = (T-k)$

$\Rightarrow$ $\text{E}[\boldsymbol{e}'\boldsymbol{e}/(T-k)|\mathbf{X}] = \text{E}[s^2|\mathbf{X}] = \sigma^2$

Note: $(T-k)$ is referred as a *degrees of freedom* correction.

• Then, the estimator of $\text{Var}[\mathbf{b}|\mathbf{X}] = s^2(\mathbf{X}'\mathbf{X})^{-1}$ (a kxk matrix)

$\Rightarrow$ $\text{SE}[\text{b}_k|\mathbf{X}] = \text{sqrt}\{[s^2(\mathbf{X}'\mathbf{X})^{-1}]_{kk}\} = s_{b,k}$

OLS Estimation – Testing Only One Parameter

• We are interested in testing a hypothesis about one parameter in our linear model: $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

1. Set H_0 and H_1 (about only one parameter): H_0: $\beta_k = \beta_k^0$
H_1: $\beta_k \neq \beta_k^0$

2. Appropriate T(X): *t-statistic*. To derive the distribution of the test under H_0, we will rely on assumption (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim \text{N}(\mathbf{0}, \sigma^2\mathbf{I}_T)$ (otherwise, results are only asymptotic).

Let b_k = OLS estimator of β_k $\text{SE}[\text{b}_k|\mathbf{X}] = \text{sqrt}\{[s^2(\mathbf{X}'\mathbf{X})^{-1}]_{kk}\} = s_{b,k}$

From assumption (**A5**), we know that

$$\text{b}_k\,|\mathbf{X} \sim \text{N}(\beta_k, v^2_{b,k}) \Rightarrow \text{Under } H_0\text{: } \text{b}_k\,|\mathbf{X} \sim \text{N}(\beta_k^0, s^2_{b,k}).$$

$$\Rightarrow \text{Under } H_0\text{: } t_k = (\text{b}_k - \beta_k^0)/s_{b,k}\,|\mathbf{X} \sim t_{T-k}$$

OLS Estimation – Testing Only One Parameter

- We measure distance in standard error units:

$$t_k = \frac{b_k - \beta_k^0}{s_{b,k}}$$

Note: t_k is an example of the *Wald* (normalized) *distance measure*. Most tests statistics in econometrics will use this measure.

3. Compute t_k, $\hat{t}$, using b_k, β_k^0, s, and $(\mathbf{X'X})^{-1}$. Get *p-value*$(\hat{t})$.

4. Rule: Set an α level. If *p-value*$(\hat{t}) < \alpha$ ⇒ Reject H_0: $\beta_k = \beta_k^0$
Alternatively, if $|\hat{t}| > t_{T-k,1-\alpha/2}$ ⇒ Reject H_0: $\beta_k = \beta_k^0$.

OLS Estimation – Testing Only One Parameter

- Special case: H_0: $\beta_k = 0$
H_1: $\beta_k \neq 0$.

Then,

$$t_k = \frac{b_k}{\text{sqrt}\{[s^2(\mathbf{X'X})^{-1}]_{kk}\}} = \frac{b_k}{\text{SE}[b_k]} \qquad \Rightarrow t_k \sim t_{T-k}.$$

This special case of t_k is called the ***t-value*** or *t-ratio* (also refer as the "**t-stats**"). That is, the t-value is the ratio of the estimated coefficient and its SE.

- The t-value is routinely reported in all regression packages. In the lm() function, it is reported in the third column of numbers.

- Usually, α = 5%, then if $|t_k| >$ **1.96** ≈ 2, we say the coefficient b_k is "*significant*."

OLS Estimation – Testing the CAPM

Example: We test the CAPM for IBM using the time-series.

CAPM: $E[r_{i,t} - r_f] = \beta_i\, E[(r_{m,t} - r_f)]$.

According to the CAPM, equilibrium expected excess returns are only determined by expected excess market returns –i.e., the CAPM is a one factor model (no constant or extra factors besides the market).

A linear data generating process (DGP) consistent with the CAPM is:

$$r_{i,t} - r_f = \alpha_i + \beta_i\,(r_{m,t} - r_f) + \varepsilon_{i,t}, \quad i = 1, ..., N \;\&\; t = 1, \ldots, T$$

Thus, we test the CAPM by testing H_0 (CAPM holds): $\alpha_{i=IBM} = 0$

H_1 (CAPM rejected): $\alpha_{i=IBM} \neq 0$

```
SFX_da <-
read.csv("http://www.bauer.uh.edu/rsusmel/4397/Stocks_FX_1973.csv",head=TRUE,sep=",")
x_ibm <- SFX_da$IBM                   # Extract IBM price data
x_Mkt_RF <- SFX_da$Mkt_RF             # Extract Market excess returns (in %)
```

OLS Estimation – Testing the CAPM

Example (continuation):

```
x_RF <- SFX_da$RF                     # Extract risk free rate (in %)
T <- length(x_ibm)                    # Sample size
 lr_ibm <- log(x_ibm[-1]/x_ibm[-T])   #  Log returns for IBM (lost one observation)
 Mkt_RF <- x_Mkt_RF[-1]/100           # Adjust size (take one observation out )
RF <- x_RF[-1]/100
ibm_x <- lr_ibm – RF                  # Define excess returns for IBM

fit_ibm_capm <- lm(ibm_x ~ Mkt_RF)    # OLS estimation with lm package in R
> summary(fit_ibm_capm)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.005791   0.002487  -2.329   0.0202 *
xMkt_RF      0.895774   0.053867  16.629   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Q: Is intercept ($\alpha_{i=IBM}$) equal to 0? Check t-value: $t_{\alpha_i} = \dfrac{\alpha_{i=IBM}}{SE[\alpha_{i=IBM}]}$

OLS Estimation – Testing the CAPM

Example (continuation):

We use the t-value: $\hat{t}_\alpha = \frac{\alpha_{i=IBM}}{SE[\alpha_{i=IBM}]} = \frac{-0.005791}{0.002487} = -2.329$

$\Rightarrow |\hat{t}_\alpha| > 1.96$ $\Rightarrow$ Reject H_0 (CAPM) at 5% level

Conclusion: The CAPM is rejected for IBM at the 5% level.
Note: You can also reject H_0 by looking at the *p-value* of intercept.

Interpretation: Given that the intercept is significant (& negative), IBM underperformed relative to what the CAPM expected:

- $r_{IBM,t} - r_f$: mean(ibm_x) = **-0.00073141**
- $r_{IBM,t} - r_f$ (CAPM): β_i * mean(Mkt_RF) = **0.895774** * 0.0056489 = **0.0050601**
- Ex-post difference: **-0.00073141** - **0.0050601** = **-0.00579151** ($\approx \alpha_{IBM}$)

OLS Estimation – Testing the CAPM: Remark

• We tested (& rejected) the CAPM for one asset only, IBM. But, the CAPM should apply to all assets. Suppose we have N assets. Then, a test for the CAPM involves testing N α_i's:

H_0: $\alpha_1 = \alpha_2 = \ldots. = \alpha_N = 0$

H_0: at least one $\alpha_i \neq 0$.

• This test is a joint test. It requires a simultaneous estimation of N CAPM equations. Usually, since returns are estimated with a lot of noise, portfolios are used. Also, the estimation usually takes into account the possible change over time of beta coefficients.

• There are different ways to approach this simultaneous estimation, a common approach is a two-step estimation, popularly known as Fama-MacBeth (1973).

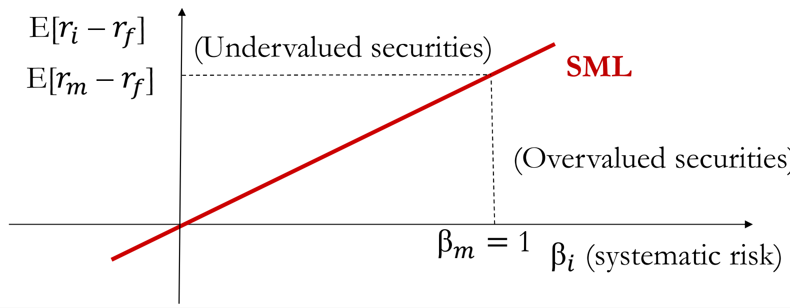
OLS Estimation – Testing the CAPM (CS)

The CAPM also tells a cross-section story for asset returns: Assets with higher β_i should get, on average, higher compensation.

CAPM (cross-section): $\quad E[r_i - r_f] = \beta_i \lambda$

where λ, in equilibrium, is the market excess return (or factor return).

If we have β_i's for N assets, we can estimate the *security market line* (SML), where we show the effect of β_i on $E[r_i - r_f]$.



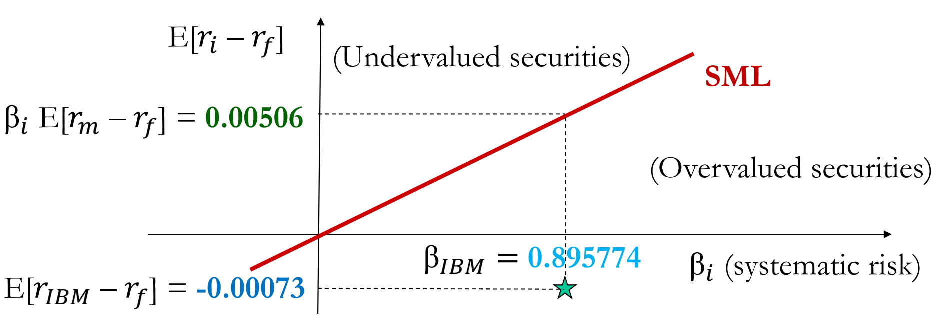
OLS Estimation – Testing the CAPM (CS)

Example (continuation):

IBM underperformed relative to what the CAPM expected by

$$\alpha_{i=IBM} = \textbf{-0.005791}$$

Then, according to the CAPM, IBM has been overvalued. The average, negative, performance (**-0.00073**) is the performance of a much safer asset, with a small, negative β!



OLS Estimation – Testing the CAPM (CS)

Q: Which assets pay a higher return? The SML answers this question: Assets with the higher exposure to market risk –i.e., higher β_i.

A linear DGP consistent with the CAPM is:

$$(r_i - r_f) = \alpha + \beta_i \lambda + \varepsilon_i, \qquad i = 1, ..., N$$

Testing implication of the SML for the cross-section of stock returns:

H_0 (CAPM holds in the CS): $\alpha = 0$ & $\lambda = E[r_{m,t} - r_f] > 0$

H_1 (CAPM rejected in the CS): $\alpha \neq 0$ and/or $\lambda \neq E[r_{m,t} - r_f] > 0$

Note: Fama and French (1992, 1993) estimated variations of the DGP with more factors. They found that β was weakly significant or not significant ("*Beta is dead*") in explaining the C-S of stock returns. The debate about β & what (& how many) factors to include continues.

