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

- 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.

Review – LS Estimation with Linear Algebra

• OLS estimator: $\mathbf{b} = (\mathbf{X'X})^{-1}\,\mathbf{X'}\boldsymbol{y}$ (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.** $\text{Var}[a\,X + b] = a^2\,\text{Var}[X]$
- **Rule 5.** $\text{Var}[aX + bY + c] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\,\text{Cov}[X,Y]$

Review – Rules for Expectations: Linear Model

- Suppose we have the **CAPM DGP**:

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

where the error term, $\varepsilon_{i,t}$, has zero mean ($E[\varepsilon_{i,t}] = 0$), variance equal to $Var[\varepsilon_{i,t}]$ and unrelated to $r_{m,t} - r_f$.

Then, using rules of expectation, we derive $E[r_{i,t} - r_f]$ & $Var[r_{i,t} - r_f]$:

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

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

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

Also, by **Rule 5** & assuming $Cov[r_{m,t} - r_f, \varepsilon_{i,t}] = 0$):

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

Review – Rules for Expectations: Linear Model

Example: We compute $E[r_{i=IBM} - r_f]$ & $Var[r_{i=IBM} - r_f]$ for IBM, using OLS estimates for α_i & β_i & $Var[\varepsilon_{i,t}]$ and sample estimates for $E[r_{m,t} - r_f]$ & $Var[r_{m,t} - r_f]$.

Estimates:

b_1 (Intercept) = **-0.00579**, b_2 = **0.89577**, & Est. $Var[\varepsilon_{i,t}]$ = **0.003484**

Mean $[r_{m,t} - r_f]$ = **0.0056489**, & Estimated $Var[r_{m,t} - r_f]$ = **0.002148**

Then,

$E[r_i - r_f]$ = **-0.00579** + **0.89577** * **0.0056489** = **-0.000729** (**-0.0729%**)

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

= 0.89577^2 * **0.002148** + **0.003484** = .0052076

$\Rightarrow$ $SD[r_i - r_f]$ = sqrt(.0052076) = **0.07216** (7.22%)

OLS – Assumptions

• Typical OLS Assumptions

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

$\Rightarrow$ functional form known, but $\boldsymbol{\beta}$ is unknown.

(**2**) $E[\varepsilon_i] = 0$. $\Rightarrow$ expected value (mean) of the errors is 0.

(**3**) Explanatory variables $X_1, X_2, ..., X_k$, are given (& non random)

$\Rightarrow$ no correlation with ε ($Cov(\varepsilon_i, X_j) = 0$.)

(**4**) The k explanatory variables are independent.

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

(**6**) $Cov(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i \varepsilon_j] = 0$. (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 $f(\mathbf{X}, \boldsymbol{\theta}) = \mathbf{X}\,\boldsymbol{\beta}$ 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*.

We start with the DGP: $\boldsymbol{y} = \mathbf{X}\,\beta + \varepsilon$.

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

$$\Rightarrow E[\boldsymbol{y}|\mathbf{X}] = E[\mathbf{X}\,\beta|\mathbf{X}] + E[\varepsilon|\mathbf{X}] = \mathbf{X}\,\beta$$

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

Also, from (**A2**), it can be shown that $E[\boldsymbol{\varepsilon}\mathbf{X}] = 0 \quad \Rightarrow \boldsymbol{\varepsilon} \perp \mathbf{X}$.

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

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

LS – Assumptions with Linear Algebra Notation

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

That is, the (conditional) variance of the errors is given by:

$$Var[\boldsymbol{\varepsilon}|\mathbf{X}] = E[(\boldsymbol{\varepsilon} - E[\boldsymbol{\varepsilon}])\,(\boldsymbol{\varepsilon} - E[\boldsymbol{\varepsilon}])'|\mathbf{X}] \qquad (T\text{x}T) \text{ matrix}$$
$$= E[(\boldsymbol{\varepsilon} - \mathbf{0})\,(\boldsymbol{\varepsilon} - \mathbf{0})'|\mathbf{X}]$$
$$= E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}]$$
$$= \begin{bmatrix} E[\varepsilon_1^2|\mathbf{X}] & E[\varepsilon_2\,\varepsilon_1|\mathbf{X}] & \cdots & E[\varepsilon_T\,\varepsilon_1|\mathbf{X}] \\ E[\varepsilon_1\,\varepsilon_2|\mathbf{X}] & E[\varepsilon_2^2|\mathbf{X}] & \cdots & E[\varepsilon_T\,\varepsilon_2|\mathbf{X}] \\ \vdots & \vdots & \vdots & \vdots \\ E[\varepsilon_1\,\varepsilon_T|\mathbf{X}] & E[\varepsilon_2\,\varepsilon_T|\mathbf{X}] & \cdots & E[\varepsilon_T^2|\mathbf{X}] \end{bmatrix}$$
$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\,\mathbf{I}_T$$

LS – Assumptions with Linear Algebra Notation

• 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

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

LS – Assumptions with Linear Algebra Notation

• From Assumption (**A4**) $\Rightarrow$ 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.

$\Rightarrow$ **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):

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

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 $T \geq k$.

- OLS estimator: bjective function:

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

- 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? Yes, 2 **X'X** is a positive definite matrix.

OLS Estimation: Second Order Condition

$$\frac{\partial^2 S(x_i, \beta)}{\partial\beta\partial\beta'} = 2\boldsymbol{X}'\boldsymbol{X} = 2\begin{bmatrix} \Sigma_{i=1}^{T} x_{i1}^2 & \Sigma_{i=1}^{T} x_{i1}x_{i2} & \cdots & \Sigma_{i=1}^{T} x_{i1}x_{iK} \\ \Sigma_{i=1}^{T} x_{i2}x_{i1} & \Sigma_{i=1}^{T} x_{i2}^2 & \cdots & \Sigma_{i=1}^{T} x_{i2}x_{iK} \\ \cdots & \cdots & \cdots & \cdots \\ \Sigma_{i=1}^{T} x_{iK}x_{i1} & \Sigma_{i=1}^{T} x_{iK}x_{i2} & \cdots & \Sigma_{i=1}^{T} x_{iK}^2 \end{bmatrix}$$

If there were a single **b**, we would require this to be positive, which it would be: $2\,\boldsymbol{x}'\boldsymbol{x} = 2\sum_{i=1}^{T} x_i^2 > 0.$

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

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

OLS Estimation – Properties of b

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

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

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

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

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

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

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

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

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

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

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

$= \sigma^2\,(\mathbf{X'X})^{-1}$ $(k\text{x}k)$ 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.* $N(\mathbf{0}, \sigma^2\,\mathbf{I}_T)$,
we derive the distribution of **b**:

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

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

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

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

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

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

<u>Remark</u>: With (**A5**) we can do (exact) 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) $\mathbf{b} \xrightarrow{a} N(\boldsymbol{\beta}, \sigma^2 (\mathbf{X'X})^{-1})$ (**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 **β**, but we rely on an "approximate distribution."

OLS Estimation – Fitted Values and Residuals

• OLS estimates **β** with **b.** Now, we define *fitted values* as:

$\hat{\boldsymbol{y}} = \mathbf{X\,b}$ (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** $\Rightarrow \boldsymbol{\varepsilon} \perp \mathbf{X}$

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

Using $\boldsymbol{e}$, 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: $\boldsymbol{y} = \beta_1 + \beta_2 \boldsymbol{x} + \varepsilon$

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

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

$$\text{Var}[b_1|\mathbf{X}] = \sigma^2 \frac{\sum_i x_i^2}{T(\sum_i x_i^2 - T\bar{x}^2)} = \sigma^2 \frac{\sum_i x_i^2/T}{\sum_i (x_i - \bar{x})^2} > 0$$

$$\text{Var}[b_2|\mathbf{X}] = \sigma^2 \frac{1}{(\sum_i x_i^2 - T\bar{x}^2)} = \sigma^2 \frac{1}{\sum_i (x_i - \bar{x})^2} > 0$$

OLS Estimation – Var[b|X]

Example (continuation):

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

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

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

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

$$\text{RSS} = \sum_i e_i^2 = \boldsymbol{e}'\boldsymbol{e}$$

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:

$$s^2 = \frac{\text{RSS}}{(T-k)} = \frac{\sum_i e_i{}^2}{(T-k)}$$

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

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

OLS Estimation – Testing Only One Parameter

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

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*. We derive the distribution of the test under H_0, using assumption (**A5**) $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_T)$

We use b_k (OLS) to estimate β_k. From assumption (**A5**) we get

$$b_k|\mathbf{X} \sim N(\beta_k, v_{b,k}^2) \Rightarrow \text{Under } H_0\text{: } b_k|\mathbf{X} \sim N(\beta_k^0, s_{b,k}^2).$$

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

<u>Technical Note</u>: If (**A5**) does not hold, we rely on asymptotic distributions for the estimators & tests.

OLS Estimation – Testing Only One Parameter

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*($\hat{t}$) < α ⇒ 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 – Is IBM's Beta equal to 1?

Example: Using the 1-factor CAPM for IBM returns, we test if IBM's market β = 1, that is, if IBM bears the same market risk as the market. Using the *lm* function previous estimation:

```
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 %)
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                     # Adjust size (take one observation out )
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
```

OLS Estimation – Is IBM's Beta equal to 1?

Example (continuation):

```
> 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

b_ibm <- fit_ibm_capm$coefficients      # Extract from lm function OLS coefficients
SE_ibm <- sqrt(diag(vcov(fit_ibm_ff3)))# SE from fit_ibm (also a kx1 vector)
t_beta1 <- (b_ibm[2] - 1)/SE_ibm[2]     # t-stat for H0: Beta1 - 1
> t_beta1
```

[1] -1.934877 ⇒ $|\hat{t}_1 = -1.934877| < 1.96$ ⇒ Cannot reject H_0 at 5% level

```
p_val <- (1- pnorm(abs(t_beta1))) * 2    # pvalue for t_beta (adjusted b/c two sided test)
> p_val
```

[1] 0.0530 ⇒ cannot reject H_0: $\beta_{IBM} = 1$ at 5% level, but borderline!

OLS Estimation – Is IBM's Beta equal to 1?

Example (continuation):

Conclusions: Cannot reject H_0: $\beta_{IBM} = 1$ ⇒ IBM has a one-to-one risk relation with the market, but borderline test decision!

Note: You can get same results using linear algebra. From last class:

```
b <- solve(t(x)%*% x)%*% t(x)%*%y          # b = (X'X)^-1 X' y  (OLS regression)
SE <- sqrt(diag(vcov(fit_ibm_capm)))       # SE from fit_ibm (also a kx1 vector)
t_beta1 <- (b[2] - 1)/SE[2]                # t-stat for H0: Beta1 - 1

> t(b)
                Mkt_RF
[1,] -0.005791039 0.895773564

> t(SE_b)
                Mkt_RF
[1,] 0.002487 0.053867

> t_beta1
[1] -1.934877
```

OLS Estimation – Testing the CAPM

Example: Now, 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).

CAPM DGP:

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

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: $t_{\alpha_i} = \frac{\alpha_{i=IBM}}{\text{SE}[\alpha_{i=IBM}]}$

OLS Estimation – Testing the CAPM

Example (continuation):

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

$\Rightarrow |\hat{t}_\alpha| > 1.96$ $\Rightarrow$ 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 the cross-section of asset returns: IBM, Ford, Apple, Exxon, etc. Suppose we have N assets in the cross-section. Then, a test for the CAPM involves testing N α_i's:

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

$$H_0: \text{at least one } \alpha_i \neq 0.$$

• 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 do this join test. A common approach is a **two-step estimation**, popularly known as Fama-MacBeth (1973).

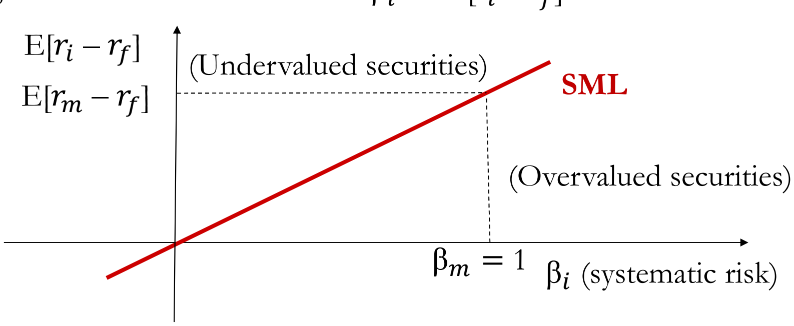
OLS Estimation – Testing the CAPM (CS)

According to the CAPM, in the cross-section of asset returns, assets with higher β_i should get, on average, higher compensation.

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

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 $E[r_i - r_f]$.



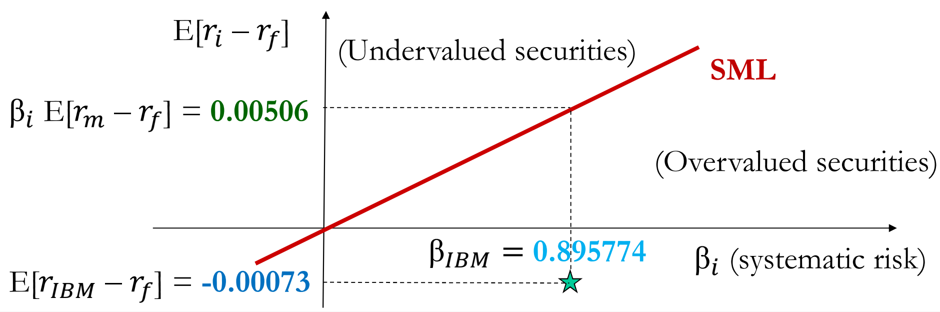
OLS Estimation – Testing the CAPM (CS)

Example (continuation):

IBM underperformed relative to what the CAPM expected by

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

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 (=-0.1295)!



OLS Estimation – Testing the CAPM (CS)

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

A linear cross-sectional 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 $\lambda \neq E\,[r_{m,t} - r_f] > 0$

<u>Note</u>: 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.

OLS Estimation – Testing the CAPM (CS)

• Fama-MacBeth (1973) proposed a well-known **two-step approach** to test the CAPM in the cross-section:

(1) Estimate β_i using the time series (T observations) for each asset i.
$r_{i,t}$ - $r_{f,t} = \alpha_i + \beta_i$ ($r_{M,t}$ - $r_{f,t}$) + $\varepsilon_{i,t}$, $t = 1, \ldots, T$ $\Rightarrow$ Get N b_i's.

(2) Using the N b_i's as regressors, estimate

$$(\bar{r}_i - \bar{r}_f) = \alpha + b_i\, \lambda + \varepsilon_i, \qquad i = 1, \ldots, N$$

where $(\bar{r}_i - \bar{r}_f)$ is the average excess return of asset i in our sample.

The usual execution of almost all 2-step procedures involves:
1) Since returns are estimated with a lot of noise, portfolios are used.
2) The estimation takes into account the possible change over time of beta coefficients, by estimating the β_i's every 5 or 10 years.

OLS Estimation – Testing the CAPM (CS)

Example: We test the CAPM, in the cross-section, using the 2-step Fama-MacBeth method. We use returns of 25 Fama-French portfolios (sorted by Size (ME) and Book-to-Market), downloaded, along the 3-Fama-French factors from Ken French's website.

```
FF_p_da <- read.csv("https://www.bauer.uh.edu/rsusmel/4397/FF_25_portfolios.csv", head=TRUE, sep=",")
FF_f_da <- read.csv("https://www.bauer.uh.edu/rsusmel/4397/FF_3_factors.csv", head=TRUE, sep=",")

# Extract variables from imported data
Mkt_RF_fm <- FF_f_da$Mkt_RF          # extract Market excess returns (in %)
HML_fm <- FF_f_da$HML                # extract HML returns (in %)
SMB_fm <- FF_f_da$SMB                # extract HML returns (in %)
RF_fm <- FF_f_da$RF                  # extract Risk-free rate (in %)
Y_p <- FF_p_da[,2:26] - RF_fm        # Compute excess returns of 25 portfolios

T <- length(HML_fm)                  # Number of observations (1926:July on)
x0 <- matrix(1,T,1)                  # Vector of ones, represents constant in X
```

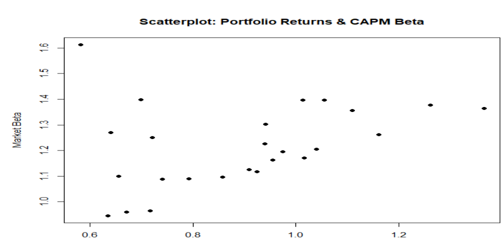
OLS Estimation – Testing the CAPM (CS)

Example (continuation):

```
x <- cbind(x0, Mkt_RF_fm)                    # Matrix X (Tx2)
k <- ncol(Y_p)

## First Pass (CAPM Betas)
Allbs = NULL                                 # Initialize empty (a space to put betas)
for (i in seq(1,k,1)){
  y <- Y_p[,i]      # select Y (portfolio)
  b <- solve(t(x)%*% x)%*% t(x)%*%y          # OLS regression = (X'X)^(-1) X'y
  Allbs  =cbind(Allbs,b)                     # accumulate b as rows
}

beta_ret <- cbind(colMeans(Y_p),t(Allbs))  # Mean portfolio returns along alpha & beta estimates
cor(beta_ret[,1], beta_ret[,3])              # Correlation of mean portfolio return & beta

> cor(beta_ret[,1], beta_ret[,3])
[1] 0.3326008

plot(beta_ret[,1], beta_ret[,3], main="Scatterplot: Portfolio Returns & CAPM Beta",
    xlab="Mean Portfolio Returns ", ylab="Market Beta", pch=19)
```

OLS Estimation – Testing the CAPM (CS)

Example (continuation):




```
## Second Pass (CAPM-SML only)
fit_fm_capm_25 <- lm(beta_ret[,1] ~ beta_ret[,3])
> summary(fit_fm_capm_25)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.3728     0.3113   1.198    0.243
beta_ret[, 3]  0.4289     0.2536   1.691    0.104
```

⇒ Not significant: Beta plays no role!

Conclusion: CAPM's beta does not seem to be useful to explain expected returns.

OLS Estimation – Testing the CAPM (CS)

- Fama and French (1992, 1993) estimated variations of the DGP with more factors. They found that β was weakly significant or not significant, even with the wrong sign, in explaining the C-S of stock returns, which created a big splash in the literature ("**Beta is dead**").

- Other researchers dispute the "Beta is dead" finding, criticizing the selection of estimation period, construction of portfolios, number of factors, statistical problems like measurement error and incorrect SE, etc.

- The debate about β & what (& how many) factors to include in the DGP continues.

OLS Estimation – Testing Multi-factor Models

- Fama-French (1992, 1993) generalized Fama-MacBeth two-step approach to test β_i in multi-factor models in the cross-section. In their 3-factor model:

(1) First pass

Using the time series (T observations), run a regression with the 3 Fama-French factors (Market, SMB, HML) to estimate 3 β_i's for each asset $i = 1, \ldots, N$.

$r_{i,t} - r_{f,t} = \alpha_i + \beta_{1,i}\,(r_{M,t} - r_{f,t}) + \beta_{2,i}\, SMB_t + \beta_{3,i}\, HML_t + \varepsilon_{i,t}, t = 1, \ldots, T$

$\Rightarrow$ Get N $\mathbf{b}_i = [b_{1,i}, b_{2,i}, b_{3,i}]$.

(2) Second pass

Using the N b_i's as regressors, estimate

$(\bar{r}_i - \bar{r}_f) = \alpha + b_{1,i}\,\lambda_1 + b_{2,i}\,\lambda_2 + b_{3,i}\,\lambda_3 + \varepsilon_i, \qquad i = 1, \ldots, N$

where $(\bar{r}_i - \bar{r}_f)$ is the average excess return of asset i in our sample.

