Lecture 1

Regression Analysis & Extension

(Review & Added Issues)

Read
- Wooldridge* ch 4
- Verbeek* ch 2, 3, ch 4.10
- Green ch 2, 3, 4, 6

Linear Algebra of OLS Estimation

Classical Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + e_i \quad , \; i = 1, \dots n,$$

or
$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix} + \cdots + \beta_k \begin{pmatrix} x_{k1} \\ \vdots \\ x_{kn} \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

$$= \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}}_{n \times (k+1)} \underbrace{\begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}}_{(k+1)} + \underbrace{\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}}_{n \times 1}$$

or
$$y = X\beta + e$$

or $y_i = x_i \beta + e_i$ where $x_i = (1, x_{1i}, \cdots, x_{ki})$: $1 \times (k+1)$
.. row vector .. Verbeek

(or $y_i = x_i' \beta + e_i$ where $x_i = (1, x_{1i}, \cdots, x_{ki})'$: $(k+1) \times 1$)
... column vector .. Wooldridge
(this note, too)

we observe : y, X
we do not observe : β, e

Note - usually one includes an intercept

$x_{0i} = 1 \Rightarrow \beta_0$ is an intercept. (Thus, we drop x_{0i} and use $x_{1i} = 1$; k regressors)

- Multiple regression

$$\frac{\partial y_i}{\partial x_{ji}} = \beta_j \quad \text{other things being constant (partial effect)}$$

Def The model $y = X\beta + e$ satisfies ideal conditions if
a) the errors e_i are independently, identically distributed (i.i.d) and $N(0, \sigma^2)$
b) the regressors X are non-stochastic, has rank $k+1$ and $\lim_{n\to\infty} \frac{1}{n} X'X$ exists and is non-singular.

a) Assumptions about e (never residuals)

i) e_i normal $\forall i$ → non-normal, robust estimation

ii) $E(e_i) = 0$ $\forall i$ → non-zero mean, specification error (omitted variables)

iii) $Var(e_i) = \sigma^2$ $\forall i$ → heteroskedasticity, HAC error

iv) $Cov(e_i, e_s) = 0$ $\forall i \neq s$ → autocorrelation, Dynamic models

Note If X is treated as random, these are assumptions about $e|X$. ii)' $E(e_i|X) = 0$ iii)' $Var(e_i|X_i) = \sigma^2$ and so on.

b) Assumptions about X

i) non-stochastic, fixed → stochastic regressor

ii) $Rank(X) = k$ → multicollinearity

iii) 2nd moment exists → integrated regressor

$$\left(\frac{1}{n} X X'\right)_{sm} = \frac{1}{n} \sum_i^n X_{si} X_{mi} \text{ exists}$$

c) Assumption about both

i) $Cov(X_j, e) = 0$, $j = 1, \cdots, k$

⇒ X_j is "endogenous".

Note Wooldridge (p.50) discusses three sources of endogeneity

- Omitted variable $E(e_i | X_j) \neq 0$ ⇒ $Cov(X_j, e) \neq 0$
- measurement error $X_j = X_j^* + u_i$ (X_j^* true measures)
- simultaneity (big issue)

OLS Estimator

$$\text{Min } e'e = (y - X\beta)'(y - X\beta)$$
$$= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta$$

$$\frac{d(e'e)}{d\beta} = -X'y - X'y + 2X'X\beta \overset{\text{let}}{=} 0$$

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'y$$

Alternatively

$$\text{Min } e_i^2 = \Sigma(y_i - X_i\beta)'(y - X_i\beta) \quad \text{where } \underset{(1\times k)}{X_i} = (X_{1i}, \cdots, X_{ki})$$

$$\Rightarrow \hat{\beta} = (\Sigma X_i'X_i)^{-1}(\Sigma X_i'y_i)$$

Note Differentiation of a scalar w.r.t. a vector

Let z : scalar (1×1)

θ : vector $(p\times 1)$, $\theta = (\theta_1, \ldots, \theta_p)'$

$$\frac{dz}{d\theta} = \left(\frac{dz}{d\theta_1}, \cdots, \frac{dz}{d\theta_p}\right)'$$

i) If a and b are $p\times 1$ vector and

$z = a'b$ or $b'a$ is a scalar (1×1)

then $$\frac{\partial a'b}{\partial a} = b, \quad \frac{\partial a'b}{\partial b} = a$$

ii) If B is a symmetric matrix $(p\times p)$, and

c is a vector $(p\times 1)$, then $c'Bc$ is a scalar.

Then

$$\frac{\partial c'Bc}{\partial c} = 2Bc$$

Thus

$$\begin{cases} \dfrac{\partial\, y'X\beta}{\partial\beta} = \dfrac{\partial\,(X'y)'\beta}{\partial\beta} = X'y & \text{let } a = X'y,\ b=\beta \\ \dfrac{\partial\, \beta'X'y}{\partial\beta} = \dfrac{\partial\,\beta'(X'y)}{\partial\beta} = X'y & \text{let } a=\beta,\ b = X'y \\ \dfrac{\partial\, \beta'X'X\beta}{\partial\beta} = \dfrac{\partial\,\beta'(X'X)\beta}{\partial\beta} = 2(X'X)\beta & \text{let } c=\beta,\ B = X'X \end{cases}$$

Exercise ① Prove $\partial a'b/\partial a = b$, $\partial a'b/\partial b = a$

② Let $y = \alpha + e$. Find $\hat{\alpha}$ to minimize $e'e$.

(Answer: $\hat{\alpha} = \bar{y}$)

Note Analogy principle (Wooldridge, p. 53)

{ "method of moments"
"generalized method of moments (GMM)" }

$$y = X\beta + e$$

multiply X' to both sides

$$X'y = X'X\beta + X'e$$

Take expectation

$$E(X'y) = E(X'X)\cdot\beta + \underbrace{E(X'e)}_{\to 0} \quad \text{since } Cov(X,e)=0$$

then

$$\beta = [E(X'X)]^{-1}\cdot E(X'y)$$

This holds in the population, thus it should hold in the sample (analogy principle)

$$\hat{\beta} = \underset{(k\times k)}{\left(\frac{1}{n}\sum X_i'X_i\right)^{-1}}\ \underset{(k\times 1)}{\left(\frac{1}{n}\sum X_i'y_i\right)}$$

Note (alternative notation): $\hat{\beta} = \left(\frac{1}{n}\sum X_i X_i'\right)^{-1}\frac{1}{n}\sum X_i y_i$ if X_i is $k\times 1$ vector

Simply, $X'e = 0 \Rightarrow X'(y - X\beta) = 0$

(orthogonal regression) $\Rightarrow X'y = X'X\beta \Rightarrow \hat{\beta} = (X'X)^{-1}X'y$

Exercise Find β s.t. $Z'e = 0$.

Remarks

① $X'\hat{e} = 0$ 'normal equation'

... always holds for any regressor! ($X'e = 0$ is an assumption.)

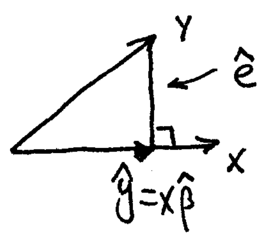
($\hat{\beta}$ is obtained in such a way that $X'\hat{e} = 0$ holds. why?)

② $\Sigma \hat{e}_i = 0$ if a constant term is included.

Note If a constant term is not included, this result does not hold. Also, R^2 can be negative or greater than 1 (depending on how R^2 is calculated.)

③ $\hat{y}_i$ = fitted value $= X_i\hat{\beta}$ $\Rightarrow$ $\frac{1}{n}\Sigma \hat{y}_i\hat{e}_i = 0$ always

$\hat{y}'\hat{e} = 0$. Fitted values and residuals are orthogonal.



$\hat{e} = Y - X\hat{\beta} = Y - \hat{Y}$ $\Rightarrow$ $\hat{e} \perp X$

$\Rightarrow$ $\hat{e} \perp \hat{y}$

; $\hat{y}$ is the projection onto the space spanned by X

Note Let $P = X(X'X)^{-1}X'$, $M = I - P = I - X(X'X)^{-1}X'$

Then $\hat{y} = Py$, $\hat{e} = My$

proof) $\hat{y} = X\hat{\beta} = X\cdot(X'X)^{-1}X'y = PY$

$\hat{e} = Y - \hat{Y} = Y - PY = (I-P)Y = MY$

Exercise a) Let $y_i = \alpha + \beta X_i + e_i$. Find two F.O.C's in $\min \Sigma e_i^2$.

b) Show : $MM = M$, $PP = P$, $MP = 0$, $MX = 0$, $PX = X$ (idempotent)

Properties of OLS estimator

① $E(\hat{\beta}) = \beta$; unbiased

proof) $E(\hat{\beta}) = E[(X'X)^{-1}X'y]$

$= E[(X'X)^{-1}X'(X\beta + e)]$

$= \underbrace{(X'X)^{-1}(X'X)}_{I} \cdot \beta + E(\underbrace{(\tfrac{1}{n}X'X)^{-1}}_{\text{ref } Q}(\tfrac{1}{n}X'e))$ ↓ alternatively $E(X'X)^{-1}X' \cdot \underset{0}{\underset{\|}{E(e)}}$

$= \beta + Q \cdot E(\tfrac{1}{n}X'e)$

$= \beta$ if $E(X'e) = 0$, which is true if $\mathrm{Cov}(X, e) = 0$

Thus, $\hat{\beta}$ is unbiased for β if $\mathrm{Cov}(X, e) = 0$ (no endogeneity!)

② $\mathrm{plim}\,\hat{\beta} = \beta$; consistent.

Note convergence in probability (Wooldridge, p. 36)

If $P[|X_N - a| > \varepsilon] \to 0$ as $n \to \infty$,

then $\mathrm{plim}\, X_N = a$. "weak convergence"

$$\hat{\beta} = \left(\tfrac{1}{n}\sum X_i'X_i\right)^{-1}\left(\tfrac{1}{n}\sum X_i' y_i\right) \quad \leftarrow y_i = X_i\beta + e_i$$

$$= \beta + \underbrace{\left(\tfrac{1}{n}\sum X_i'X_i\right)^{-1}}_{\downarrow\, Q}\underbrace{\left(\tfrac{1}{n}\sum X_i' e_i\right)}_{\downarrow\, 0}$$

since $\tfrac{1}{n}\sum X_i' e_i = E(X'e) = 0$

Thus $\mathrm{plim}\,\hat{\beta} = \beta$ $\Rightarrow$ We say, $\hat{\beta}$ is consistent.

Note A sufficient condition for "consistency" is

① $\mathrm{Bias} = E(\hat{\beta}) - \beta \to 0$ as $n \to \infty$; $E(\hat{\theta} - \theta)^2 = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}^2 \to 0$

② $\mathrm{Var}(\hat{\beta}) \to 0$ as $n \to \infty$

(2nd mean convergence, quadratic mean ", MSE convergence)

③ Asymptotic variance

$$Var(\hat\beta) = \begin{cases} \sigma^2 (X'X)^{-1} & \text{under homoskedasticity } Var(e_i)=\sigma^2 \\ (X'X)^{-1}\left(\Sigma \hat e_i^2 X_i' X_i\right)(X'X)^{-1} & \text{heteroskedasticity-consistent (robust) variance.} \end{cases}$$

<u>Note</u> Variance matrix; also called variance-covariance matrix

<u>Def</u> $Var(\hat\theta)$ or $Cov(\hat\theta)$ $\quad p\times p$

$$= E\Big[\underbrace{(\hat\theta - E(\hat\theta))}_{p\times 1}\ \underbrace{(\hat\theta - E(\hat\theta))'}_{1\times p}\Big]$$

- i, jth element

$$= E\left[(\hat\theta_i - E(\hat\theta_i))(\hat\theta_j - E(\hat\theta_j))\right] = cov(\hat\theta_i, \hat\theta_j)$$

- i-th diagonal element

$$= E\left[(\hat\theta_i - E(\hat\theta_i))^2\right] = Var(\hat\theta_i)$$

Back to $\hat\beta$.

$\hat\beta = \beta + (X'X)^{-1}X'e$ from the previous page

$E(\hat\beta) = \beta$

thus $\hat\beta - E(\hat\beta) = \beta + (X'X)^{-1}X'e - \beta = (X'X)^{-1}X'e$

$Var(\hat\beta)$ or $cov(\hat\beta)$

$$= E\left[(\hat\beta - E(\hat\beta))(\hat\beta - E(\hat\beta))'\right]$$

$$= E\left[(X'X)^{-1}\underbrace{X'e \cdot e'X}(X'X)^{-1}\right]$$

this term $= \begin{cases} X'\sigma^2 I X = \sigma^2(X'X) & \text{under } Var(e_i) = \sigma^2 \\ X'ee'X = Var(X'e), & \text{which is estimated as } \Sigma \hat e_i^2 X_i' X_i. \end{cases}$

(heteroskedasticity robust variance)

<u>Remark</u>

i) Is $\hat{\beta}$ consistent? yes

① $E(\hat{\beta}) - \beta = \text{bias} = 0$

② $Var(\hat{\beta}) = (X'X)^{-1}\sigma^2(X'X)(X'X)^{-1} = \sigma^2(X'X)^{-1}$ under homoskedasticity

$= \sigma^2 \frac{1}{n} \cdot \left(\frac{X'X}{n}\right)^{-1} \to 0$ as $n \to \infty$

Here, $\frac{X'X}{n} = \frac{1}{n}\Sigma X_i'X_i \to Q$ which is finite.

$\frac{\sigma^2}{n} \to 0$ as $n \to \infty$

ii) How about the expression $\sqrt{n}\hat{\beta}$?

$Var(\sqrt{n}\hat{\beta}) = n \cdot Var(\hat{\beta}) = n \cdot \frac{\sigma^2}{n} \cdot \left(\frac{X'X}{n}\right)^{-1} \to \sigma^2 Q$ finite

We denote

$$\sqrt{n}\hat{\beta} \to N(\beta, \underbrace{\sigma^2 Q}_{\text{finite}}) = N(\beta, \sigma^2 \underbrace{\left(\frac{X'X}{n}\right)^{-1}}_{\text{finite}})$$

we say

$\hat{\beta}$ is $\sqrt{n}$ - consistent.

... Most estimators are $\sqrt{n}$ - consistent.

In time series, unit root & cointegration,

we have n-consistent estimators.

; $Var(n\hat{\beta}) \to Q^*$ finite

$n\hat{\beta} \to N(\beta, Q^*)$ ... <u>super</u>-consistent

$var(\hat{\beta}) = \frac{1}{n^2} Q^*$

<u>Exercise</u> Show (assume $Var(e) = \sigma^2 I$)

a) $Var(R\hat{\beta}) = R \cdot var(\hat{\beta}) \cdot R' = \sigma^2 R(X'X)^{-1}R'$

b) $Var(\sqrt{n}\hat{\beta}) = n \cdot \sigma^2(X'X)^{-1} = \sigma^2\left(\frac{X'X}{n}\right)^{-1} = \sigma^2 Q$

Useful Facts

① Let λ be a $p\times 1$ vector. $\hat{\theta} = p\times 1$

$$Var(\lambda'\hat{\theta}) = \lambda' Var(\hat{\theta})\lambda \quad : 1\times 1$$

② Let R be a $m\times p$ matrix

$$Var(R\hat{\theta}) = \underset{(m\times p)}{R}\ \underset{(p\times p)}{Var(\hat{\theta})}\ \underset{(p\times m)}{R'} \quad ; m\times m$$

eg) $y = x_1\hat{\beta}_1 + x_2\hat{\beta}_2 + x_3\hat{\beta}_3 + \hat{e}$ where $var(\hat{\beta}) = \sigma^2(x'x)^{-1}$

i) $Var(\hat{\beta}_1 - 2\hat{\beta}_2 + 3\hat{\beta}_3)$

$$= Var\left[(1,-2,3)\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{pmatrix}\right] = (1,-2,3)\,\sigma^2(x'x)^{-1}\begin{pmatrix}1\\-2\\3\end{pmatrix}$$

ii) $$Var\begin{pmatrix}\hat{\beta}_1 - 2\hat{\beta}_2 + 3\hat{\beta}_3\\ \hat{\beta}_1 - \hat{\beta}_3\end{pmatrix} = \begin{pmatrix}1, & -2, & 3\\ 1, & 0, & -1\end{pmatrix}\sigma^2(x'x)^{-1}\begin{pmatrix}1 & 1\\ -2 & 0\\ 3 & -1\end{pmatrix}$$

Exercise (H/W)

1. Let $\tilde{\beta} = Cy$ where $C = (x'x)^{-1}x' + D$, $y = X\beta + e$, $Var(e) = \sigma^2 I$.

 i) Show that $DX = 0$ should hold to have $\tilde{\beta}$ as an unbiased estimator for β.

 ii) Show that

 $$Var(\tilde{\beta}) = \sigma^2(x'x)^{-1} + \sigma^2 D'D$$

2. Let $\tilde{e} = y - X\tilde{\beta} = (y - X\hat{\beta}) + X(\hat{\beta} - \tilde{\beta})$. Show that

 $$\tilde{e}'\tilde{e} = \hat{e}'\hat{e} + (\hat{\beta}-\tilde{\beta})'x'x(\hat{\beta}-\tilde{\beta}) \quad \text{where } \hat{e} = y - X\hat{\beta}$$

3. Show that $\hat{e}'\hat{e} = y'My = e'Me$
 where $M = I - X(X'X)^{-1}X'$

Useful Computational Note (orthogonal partitioned regression)

$y = X\beta + e$, Let $X = [X_1, X_2]$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$

$= X_1\beta_1 + X_2\beta_2 + e$

then

$$\hat\beta = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} = [(X_1, X_2)'(X_1, X_2)]^{-1}(X_1, X_2)'y$$

$$= \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

① $\hat\beta_1 = (X_1'M_2X_1)^{-1}X_1'M_2y$ (proof: see Greene, p. 26)

with $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$

$= (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'y$ or $(\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'\tilde{y}$

where $\tilde{X}_1 = M_2X_1$

= residual of the regression of X_1 on X_2

$= X_1 - \hat{X}_1$ with $\hat{X}_1 = X_2\hat\gamma$, for the regression

$X_1 = X_2\gamma + u$

Also,

$$\hat\beta_2 = (X_2'M_1X_2)^{-1}X_2'M_1y = (\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'y$$

$$= (\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'\tilde{y}$$

Here, $\tilde{y} = M_2y$ is the residual from the regression of y on X_2 only

$y = X_2c + u$

② Let $X_{1i} = 1$ (constant) $\Rightarrow X_1 = (1, \cdots, 1)' = i'$ (vector of ones)

$M_1 = I - i(i'i)^{-1}i'$

then $\tilde{X}_2 = M_1X_2 = X_2 - \bar{X}$, $\tilde{y} = M_1y = y - \bar{y}$

(demeaned variables; deviation from)

$\hat\beta_2$ = coeff. from the regression in demeaned variables.

Exercise H/W

$$\begin{cases} (A) & y = X_1\beta_1 + X_2\beta_2 + e \\ (B) & y = X_1\beta_1 + \tilde{X}_2\beta_2 + u, \quad \tilde{X}_2 = M_1X_2, \quad M_1 = I - X_1(X_1'X_1)^{-1}X_1' \\ (C) & y = X_1\beta_1 + \varepsilon \end{cases}$$

(a) show that the OLS estimator of β_2 from (A) is the same as the OLS estimator of β_2 from (B).

(b) show that the OLS estimator of β_1 from (B) is the same as the OLS estimator of β_1 from (C).

(Hint : $M_1X_1 = 0$)

(c) Suppose that (A) is a true model, but one uses Model (C) so that X_2 is omitted. Then show that the bias of the OLS estimator of β_1 in (C) is given as

$$E(\tilde{\beta}_1) - \beta_1 = \text{bias} = (X_1'X_1)^{-1}X_1'X_2\beta_2$$

$$= \gamma \cdot \beta_2 \quad \text{where } \gamma \text{ is the coefficient}$$

in the regression $X_1 = X_2\gamma + \text{error}$

Note this expression corresponds to eq (4.24) in the text (Wooldridge, ch 4, p.62) also p.67, eq (4.32)

Estimate of σ^2

$Var(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$ where $\underbrace{\boxed{\hat{\sigma}^2 = \frac{1}{n-k}\hat{e}'\hat{e}}}$, $\hat{e}'\hat{e} = SSE$

; $\hat{\sigma}^2$ is unbiased : $E(\hat{\sigma}^2) = \sigma^2$

Note $\dfrac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$. $E(\chi^2_{n-k}) = n-k$

thus

$$E\left[\frac{n-k}{\sigma^2}\hat{\sigma}^2\right] = \frac{n-k}{\sigma^2}E(\hat{\sigma}^2) = n-k \overset{\text{thus}}{\Rightarrow} E(\hat{\sigma}^2) = \sigma^2$$

: s^2 is often used instead of $\hat{\sigma}^2$. $s^2 = \frac{1}{n-k}\hat{e}'\hat{e}$

Testing Restrictions (Hypothesis)

- general form : $R\beta = r$ $\quad (\beta : k\times 1)$

  R = restriction matrix (selection matrix) : $m\times k$

  r = known vector : $m\times 1$

eg1) $\beta_2 = 0$, $\quad R = [0, 1, 0 \ldots 0]$, $r = 0$

eg2) $\beta_2 + \beta_3 = 1$, $\quad R = [0, 1, 1, \ldots 0]$, $r = 1$

eg3) $\beta_1 = \beta_2$ $\quad R = [1, -1, 0, \ldots 0]$, $r = 0$

$(\beta_1 - \beta_2 = 0)$

eg4) $\beta_2 + \beta_3 = 1$, $\beta_1 = \beta_2$ $\quad$ ($m = 2$ restrictions)

$$R = \begin{bmatrix} 0, & 1, & 1, & 0 \cdots & 0 \\ 1, & -1, & 0, & 0, \cdots & 0 \end{bmatrix} \qquad r = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Testing Hypothesis

$H_0 : R\beta = r$ $\qquad H_a : R\beta \neq r$

(1) Single Restriction $(m = 1)$

Let $s_R^2 = s^2 R (X'X)^{-1} R' = Var(R\hat\beta - r) = Var(R\hat\beta)$

then $\dfrac{R\hat\beta - r}{s_R} \sim t_{n-k}$

Note It's given as $\dfrac{R\hat\beta - r}{\sqrt{Var(R\hat\beta - r)}} = \dfrac{R\hat\beta - r}{\sqrt{\hat\sigma^2 R (X'X)^{-1} R'}}$

eg1) $H_0 : \beta_1 = 2$ $\qquad H_a : \beta_1 \neq 2$

$R = [1, 0 \ldots 0]$, $r = 2$

$R\hat\beta - r = \hat\beta_1 - 2$

$Var(R\hat\beta - r) = \sqrt{Var(\hat\beta_1)} = s(\hat\beta_1)$: std. error of $\hat\beta_1$

$t = \dfrac{\hat\beta_1 - 2}{s(\hat\beta_1)}$ $\quad$ where $Var(\hat\beta_1)$ is the (1,1) element of $s^2(X'X)^{-1}$

... 1st diagonal term

eg2) $H_0: \beta_1 - \beta_2 = 3$ $\quad H_a: \beta_1 - \beta_2 \neq 3$

$$R = [1, -1, 0 \cdots 0] \qquad r = 2$$

$$R\hat{\beta} - r = \hat{\beta}_1 - \hat{\beta}_2 - 3$$

$$Var(R\hat{\beta} - r) = S_R^2 = s^2 R(X'X)^{-1}R'$$

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 3}{\sqrt{S_R^2}}$$

Note $S_R^2 = s^2 R(X'X)^{-1}R' = R \begin{bmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) \\ Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) \end{bmatrix} R'$

if $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$

$$= Var(\hat{\beta}_1 - \hat{\beta}_2)$$

$$= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2\,Cov(\hat{\beta}_1, \hat{\beta}_2)$$

Thus,

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 3}{\sqrt{Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)}}$$

which is the same as the above.

Exercise (H/W) show that (refer to any book)

$$\frac{R\hat{\beta} - r}{S_R} = \frac{N(0,1)}{\sqrt{\frac{1}{n-k}\chi^2_{n-k}}} = t_{n-k}$$

Note t-dist is defined as

$$t = \frac{z}{\sqrt{\chi^2_{df}/df}}$$

Note let $\sigma_R^2 = \sigma^2 R(X'X)^{-1}R'$

then

$$\frac{S_R}{\sigma_R} \sim \sqrt{\frac{1}{n-k}\chi^2_{n-k}}$$

(2) Joint restriction $(m \geq 1)$

(i) $F = \dfrac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/m}{s^2} \sim F_{m,n-k}$

(ii) $\text{Wald} = (R\hat{\beta}-r)'[s^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r) \sim \chi^2_m$

Note $\begin{cases} \text{Wald} = (\hat{\theta}-\theta_0)' \, Var(\hat{\theta})^{-1} (\hat{\theta}-\theta_0) \sim \chi^2_p \\ \quad : \text{quadratic form} \\ \text{Wald} = (R\hat{\beta}-r)'(R\hat{V}R')^{-1}(R\hat{\beta}-r) \end{cases}$

where $\hat{V} = Var(\hat{\beta}) = \begin{cases} s^2(X'X)^{-1} & \text{under homoskedasticity} \\ \text{Robust variance} & \text{(p.7)} \end{cases}$

Note If $\hat{\theta} \sim N(\theta, \Sigma)$, then

$W = (\hat{\theta}-\theta)'\Sigma^{-1}(\hat{\theta}-\theta) \sim \chi^2_{\dim \Sigma}$

(iii) LM (Lagrange multiplier) test

$LM = nR^2 \sim \chi^2_m$

where R^2 is the usual R^2 from the regression of the restricted residual on regressors.

ex) $y = X_1\beta_1 + X_2\beta_2 + e$

$H_0: \beta_2 = 0$

Restricted model $\Rightarrow y = X_1\beta_1 + u \Rightarrow \hat{u}$ restricted residual

Run the regression

$\hat{u} = X_1 b_1 + X_2 b_2 + \text{error}$

and obtain R^2.

$LM = nR^2 \sim \chi^2_m$

(iv) LR test $= 2(\log L_u - \log L_R) \sim \chi^2_m$ ... later

Exercise H/W

$$F = \frac{(R\hat\beta - r)'[R(X'X)^{-1}R']^{-1}(R\hat\beta - r)/m}{s^2} \quad (p.13)$$

$$= \frac{(R_u^2 - R_R^2)/m}{(1-R_u^2)/(n-k)}$$

where $R_u^2 = R^2$ from unrestricted model
$R_R^2 = R^2$ from restricted model

$$= \frac{(RSS_R - RSS_u)/m}{RSS_u/(n-k)}$$

where RSS_R = RSS (residual sum of squares) from the restricted model
$= \tilde e' \tilde e$

RSS_u = RRS from the unrestricted model
$= \hat e' \hat e$

$$= \frac{(\tilde\beta - \hat\beta)'(X'X)(\tilde\beta - \hat\beta)/m}{s^2}$$

where $\hat\beta$ is unrestricted estimator
$\tilde\beta$ = restricted estimator
$= \hat\beta + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat\beta)$

Hint $\tilde\beta - \hat\beta = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\underbrace{(r - R\hat\beta)}$

where $r - R\hat\beta = R\beta - R\hat\beta$ since $R\beta = r$
$= R(\beta - \hat\beta) = -R(X'X)^{-1}X'e$

Hint $\tilde e'\tilde e - \hat e'\hat e = (\tilde\beta - \beta)'X'X(\tilde\beta - \beta)$: p.9

Hint $$\frac{RSS_R - RSS_u}{RSS_u} = \frac{\frac{RSS_R}{TSS} - \frac{RSS_u}{TSS}}{\frac{RSS_u}{TSS}} = \frac{(1-R_R^2) - (1-R_u^2)}{(1-R_u^2)}$$

Exercise Let

X_i	Y_i
1	3
2	5
3	4

$n=3$

(a) Find $\hat{\beta} = \begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}$ where $y_i = a + bX_i + e_i$

$$\begin{pmatrix}3\\5\\4\end{pmatrix} = \begin{pmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{pmatrix}\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} + \begin{pmatrix}\hat{e}_1\\ \hat{e}_2\\ \hat{e}_3\end{pmatrix}$$

or $y = X\hat{\beta} + \hat{e}$

$$\hat{\beta} = (X'X)^{-1}X'y = \left[\begin{pmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{pmatrix}\right]^{-1}\begin{pmatrix}1 & 1 & 1\\ 1 & 2 & 3\end{pmatrix}\begin{pmatrix}3\\5\\4\end{pmatrix}$$

$$= \underbrace{\begin{pmatrix}2.33 & -1.0\\ -1.0 & 0.5\end{pmatrix}}_{(X'X)^{-1}}\underbrace{\begin{pmatrix}12\\ 25\end{pmatrix}}_{X'y} = \begin{pmatrix}3\\ 0.5\end{pmatrix} \Rightarrow \hat{a} = 3,\ \hat{b} = 0.5$$

(b) Residuals $\hat{e}_i = y_i - \hat{y}_i$ where $\hat{y}_i = X\hat{\beta} = \begin{pmatrix}1 & 1\\ 1 & 2\\ 1 & 3\end{pmatrix}\begin{pmatrix}3\\ 0.5\end{pmatrix} = \begin{pmatrix}3.5\\ 4.0\\ 4.5\end{pmatrix}$

$$\hat{e}_i = y_i - \hat{y}_i = \begin{pmatrix}3\\5\\4\end{pmatrix} - \begin{pmatrix}3.5\\ 4.0\\ 4.5\end{pmatrix} = \begin{pmatrix}-0.5\\ 1\\ 0.5\end{pmatrix} = \begin{pmatrix}\hat{e}_1\\ \hat{e}_2\\ \hat{e}_3\end{pmatrix}$$

$$SSR = \Sigma \hat{e}_i^2 = \hat{e}'\hat{e} = (-0.5, 1, 0.5)\begin{pmatrix}-0.5\\ 1\\ 0.5\end{pmatrix} = 1.5$$

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\hat{e}'\hat{e} = \frac{1}{3-1-1}(1.5) = 1.5$$

(c) $Var(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = 1.5\begin{pmatrix}2.33 & -1.0\\ -1.0 & 0.5\end{pmatrix} = \begin{pmatrix}3.5 & -1.5\\ -1.5 & 0.75\end{pmatrix}$ under homosked.

thus $var(\hat{a}) = 3.5$, $var(\hat{b}) = 0.75$, $cov(\hat{a}, \hat{b}) = -1.5$

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{s(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{\sqrt{Var(\hat{\beta}_2)}} = \frac{\hat{b}}{\sqrt{Var(\hat{b})}} = \frac{0.5}{\sqrt{0.75}} = 0.577$$

(d) Robust std. error (heteroskedasticity consistent std. error)

$$\mathrm{Var}(\hat\beta) = (X'X)^{-1} \left(\sum \hat e_i^2 X_i' X_i \right) (X'X)^{-1}$$

where X_i = ith row of X with $X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$

$$\sum_1^n \hat e_i^2 X_i' X_i = \hat e_1^2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1\ 1) + \hat e_2^2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} (1,\ 2) + \hat e_3 \begin{pmatrix} 1 \\ 3 \end{pmatrix} (1\ 3)$$

$$= .25 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + 1 \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} + 0.25 \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1.5 & 3 \\ 3 & 6.25 \end{pmatrix}$$

Thus

$$\mathrm{Var}(\hat\beta) = \begin{pmatrix} 2.33 & -1.0 \\ -1.0 & 0.5 \end{pmatrix} \begin{pmatrix} 1.5 & 3 \\ 3 & 6.25 \end{pmatrix} \begin{pmatrix} 2.33 & -1.0 \\ -1.0 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.667 & -0.25 \\ -0.25 & 0.125 \end{pmatrix}$$

(e) Test $H_0: a = 2,\ b = 1$ $\quad (\beta_1 = 2,\ \beta_2 = 1)$

$$H_0: R\beta = r \Rightarrow H_0: \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$(R \quad \beta \quad = r)$$

i) Wald test

$$W = (R\hat\beta - r)' [R \hat V R']^{-1} (R\hat\beta - r)$$

where $\hat V = \begin{cases} \hat\sigma^2 (X'X)^{-1} & \text{under homoskedasticity} \\ (X'X)^{-1} (\sum \hat e_i^2 X_i' X_i)(X'X)^{-1} & \text{heteroskedasticity consistent variance} \end{cases}$

$$R\hat\beta - r = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 0.5 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -0.5 \end{pmatrix}$$

$$R\hat V R' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.667 & -0.25 \\ -0.25 & 0.125 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}' = \begin{pmatrix} 0.667 & -0.25 \\ -0.25 & 0.125 \end{pmatrix}$$

$$(R\hat V R')^{-1} = \begin{pmatrix} 6 & 12 \\ 12 & 32 \end{pmatrix}$$

thus

$$\text{Wald} = (1,\ -0.5) \begin{pmatrix} 6 & 12 \\ 12 & 32 \end{pmatrix} \begin{pmatrix} 1 \\ -0.5 \end{pmatrix} = 2$$

$\chi_2^2 = 5.99$ with $\alpha = 5\%$ level. We <u>cannot</u> reject H_0.

ii) F-test

$$F = \frac{(R\hat\beta - r)'[R\hat V R']^{-1}(R\hat\beta - r)/m}{s^2}$$ if heteroskedasticity-consistent variance $\hat V$ is used.

$$= \frac{\text{Wald}/m}{s^2} = \frac{2/2}{1.5} = \frac{1}{1.5} \sim F_{2,1}$$

$(m=2,\ n-k-1=1)$

F crit. value = 199.5 (Greene, p.956)

we cannot reject H_0.

iii) LM-test

$y_i = a + bX_i + e_i \qquad H_0: a=2,\ b=1$

restricted model : $y_i = 2 + 1X_i + u_i$ (no parameter to estimate!)

restricted residual

$$\hat u_i = y_i - 2 - X_i = \begin{pmatrix}3\\5\\4\end{pmatrix} - 2\begin{pmatrix}1\\1\\1\end{pmatrix} - 1\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}0\\1\\-1\end{pmatrix}$$

Run OLS $\hat u_i$ on $(1, X_i)$ and obtain R^2.

$\widehat{\hat u_i} = 1 - 0.5X_i$ (estimation result)

$R^2 = 0.25$

thus

$LM = nR^2 = 3 \times 0.25 = 0.75 \sim \chi^2_2 = 5.99$

Thus we cannot reject H_0.

Exercise (H/W)

Consider

X_i	Y_i
1	5
2	3
3	4

$y_i = a + bX_i + e_i,\quad \beta = \begin{pmatrix}a\\b\end{pmatrix}$

Hint $(X'X)^{-1} = \begin{pmatrix}2.33 & -1\\ -1 & 0.5\end{pmatrix}$

$\sum \hat e_i^2 x_i x_i' = \begin{pmatrix}1.5 & 3\\ 3 & 6.5\end{pmatrix}$

a) Find $\hat a, \hat b, \hat e_i, \hat\sigma^2$

b) Find $Var(\hat\beta)$ using 2 methods

c) Find t-stat for $b=1$

d) Find Wald, F-stat and LM stat for $H_0: a=2,\ b=1$ using the heteroskedasticity consistent variance estimator.

three sources of endogeneity

$Cov(X_i, e_i) \neq 0$ or $E(e_i | X_i) \neq 0$

i) Omitted variables

True : $y = X_1\beta_1 + X_2\beta_2 + e$

estimate : $y = X_1\beta_1 + u \Rightarrow \tilde{\beta}_1$ is estimated.

$E(\tilde{\beta}_1) \neq \beta_1 \Rightarrow$ Bias $= E(\tilde{\beta}_1) - \beta_1 = \gamma \cdot \beta_2$

where γ is from $X_1 = X_2\gamma + error$

* direction of bias

positive if γ and β_2 are both positive or negative

... $\tilde{\beta}_1$ is bigger than its true value

negative if their signs alternate

* Omitted variables can be unobservable (say, q)

no problem if $E(q) = 0$, but the variance of the estimator can be affected. Using robust variance mitigates the heteroskedasticity problem (see Wooldridge p. 69)

ii) measurement error

a) m.e. in regressors

say, $X_i = X_i^* + v_i$, X_i^* = true value

then $plim\ \hat{\beta} = \beta \cdot \dfrac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma^2} < \beta$ (attenuation bias)

b) m.e. in the dependent variable

say, $y_i = y_i^* + w_i$

then $y_i = X_i\beta + (e_i + w_i)$ ... Not a problem

but, std. error of $\hat{\beta}$ can be higher.

iii) Time endogeneity : simultaneity
... Lecture 2 IV estimation.

HAC std. error (Newey-West std. error)

"HAC" : "Heteroskedasticity Autocorrelation consistent" std err
(Verbeek, p.110, ch4.10)

$$y_t = X_t \beta + e_t \quad \text{.. time series}$$

$$Var(\hat{\beta}) = (X'X)^{-1} T \hat{V}^* (X'X)^{-1}$$

where $\hat{V}^* = \frac{1}{T} \sum \hat{e}_t^2 X_t' X_t + \frac{1}{T} \sum_{j=1}^{m-1} w_j \sum_{s=j+1}^{T} \hat{e}_s \hat{e}_{s-j} (X_s' X_{s-j} + X_{s-j}' X_s)$

where $w_j = 1 - \frac{j}{m}$: Bartlett weights (kernel)

Interpretation of Regression coefficients

(Verbeek ch3)

i) linear model

$$y_i = X_{1i} \beta_1 + X_{2i} \beta_2 + e_i$$

$$\beta_j = \frac{\partial E(y_i | X_i)}{\partial X_{ji}} \quad \text{: partial effect}$$

ii) quadratic form

$$y_i = X_{1i} \beta_1 + X_{2i} \beta_2 + (X_{2i}^2) \beta_3 + e_i$$

$$\frac{\partial E(y_i | X_i)}{\partial X_{2i}} = \beta_2 + 2\beta_3 \cdot X_{2i} \quad \text{... depends on } X_{2i}$$

iii) dummy variable

$$y_i = X_{1i}\beta_1 + X_{2i}\beta_2 + e_i$$

where $X_{2i} = D_i = \begin{cases} 1 & \text{for group 1 (say, female)} \\ 0 & \text{for group 2 (male)} \end{cases}$

$$\beta_2 = E(y_i | X_{1i}, X_{2i} = 1) - E(y_i | X_{1i}, X_{2i} = 0)$$

$= $ difference

eg) $\widehat{Wage}_i = 2 + 2.5\, Edu_i \underbrace{- 1.5}_{}\, Female$

\$1.50 lower wage for female workers, on average

iv) log-linear model

$$\log y_i = X_i\beta + e_i$$

$\beta = $ % change (or "times")

eg1) $\log\widehat{wage} = +1.5 + \underbrace{0.12}_{}\, Edu$

One more year of education gives 12% increase in salary (or 0.12 times higher salary)

eg2) $\log\widehat{wage}$ per hour $= 1.0 + \underbrace{0.15}_{}\, Edu \underbrace{- 0.25}_{}\, Female$

$\hat{\sigma} = 0.45$

15% increase per year of edu.

25% lower salary for female workers

"If $u \sim N(0, \sigma^2)$, $E\exp(u) = \exp(\frac{1}{2}\sigma^2)$"

Note $\hat{y}_i = \exp(X_i\hat{\beta}) \cdot \exp(\frac{1}{2}\hat{\sigma}^2) \simeq \exp(X_i\hat{\beta})$ but not precise.

eg2) A female worker with 16 yrs of edu.

$$\widehat{wage} = \exp(1.0 + 0.15 \times 16 - 0.25 \times 1) \cdot \exp(\tfrac{1}{2} 0.45^2)$$
$$= \exp(3.15) \cdot \exp(\tfrac{1}{2} \times 0.45^2) = \$25.82/\text{hour}$$

v) linear-log

$$y_i = X_{1i}\beta_1 + (\log X_{2i})\cdot\underbrace{\beta_2}_{} + e_i$$

How much y_i will increase when X_{2i} increase by 100%.

vi) log-log form

$$\log y_i = (\log X_i)\underbrace{\beta_1}_{} + e_i$$

$$\frac{\%\text{ of } y}{\%\text{ of } X} = \frac{\Delta Y/Y}{\Delta X/X} = \text{elasticity}$$

Model Selections

① Testing restrictions

eg) should we include X_2?

$$y = X_1\beta_1 + X_2\beta_2 + e$$

Test $H_0: \beta_2 = 0$ $\quad$ $H_a: \beta_2 \neq 0$

if H_0 is not rejected, drop X_2.

(t-test, F-test, LM test, LR test (later), Wald, ..)

② Adj R^2 (higher)

$$\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = R^2 - \underbrace{\frac{k-1}{n-k}(1-R^2)}_{}$$

penalty term as k increases.

③ AIC or BIC (lower)

$$AIC = \log \frac{1}{n}\sum \hat{e}_i^2 + 2k/n$$

$$BIC = \log \frac{1}{n}\sum \hat{e}_i^2 + \frac{k}{n}(\log n)$$

④ Ramsey's "RESET" test (Regression equation specification error test)

$$y_i = X_i\beta + e_i$$

H_0: The model is correctly specified

H_a: H_0 is not true

Testing procedure

i) Estimate $\hat{y}_i = X_i\hat{\beta}$

ii) Add $\hat{y}_i^2$, $\hat{y}_i^3$ and test the significance

$$y_i = X_i\beta + \alpha_2 \hat{y}_i^2 + \alpha_3 \hat{y}_i^3 + \text{error}$$

Test: $H_0: \alpha_2 = \alpha_3 = 0$ (F-test, Wald ..)

If H_0 is rejected, the model is mis-specified.

⇒ But, the question is why?

(nonlinearity? omitted variables? incorrect function form? ..)

Testing Non-nested Models

(A) $y_i = X_i\beta + e_i$ which model is correct?

(B) $y_i = z_i\gamma + u_i$

i) Non-nested F-test

To check the validity of Model (A), let $z_i = (z_{1i}, z_{2i})$ where z_{1i} is included in X_i, but z_{2i} is not.

Then test $H_0: \delta = 0$

$$y_i = X_i\beta + z_{2i}\delta + \text{error}$$

vice versa.

ii) J-test

$$y = (1-\delta) X_i \beta + \delta \underbrace{Z_i \hat{\gamma}}_{\to \hat{y}_i} + u_i$$

where $\hat{\gamma}$ is from Model (B), and $\hat{y}_i$ is , tw.

Then test $\delta = 0$

$$y = X_i \beta^* + \delta \hat{y}_i + u_i$$

where $\beta^* = (1-\delta)\beta$

... see Verbeek p.61 (also Greene, p.155)

cox test.

Testing for a Structural Break

(CHOW test)

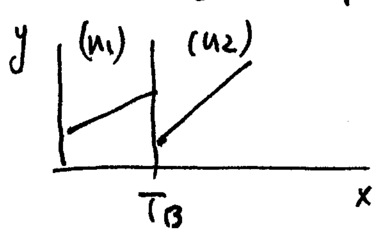
(1) $y_1 = X_1 \beta_1 + u_1$ $\quad n_1$ obs

(2) $y_2 = X_2 \beta_2 + u_2$ $\quad n_2$ obs

$\Big) \; n_1 + n_2 = n$

: same equations, different observations

eg1) Two different periods



different intercept & slope

eg2) Two different groups

$$\begin{cases} wage_1 = X\beta_1 + u_1 & : \text{female} \\ wage_2 = X\beta_2 + u_2 & : \text{male} \end{cases}$$

$X = (\text{edu}, \text{age} ..)$

Consider Stacked regression

(3) $y = X\beta + e$ or $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\beta + e$

with the restriction

$\beta_1 = \beta_2 = \beta$

$$F = \frac{[RSS_R - (RSS_1 + RSS_2)]/k}{(RSS_1 + RSS_2)/(n-2k)} \sim F_{k,\, n-2k}$$

where RSS_R is from (3)

RSS_1 & RSS_2 are from (1) or (2).

<u>Note</u> Restricted model is (3).

unrestricted RSS = $RSS_1 + RSS_2$

\# of restriction = k.

<u>Exercises</u> H/W (Empirical exercises)

- Verbeek ch2

  Ex 2.2 (p.46–47), Ex 2.3 (p.48) Asset pricing

- Verbeek ch3

  Ex 3.2 (p.77)