

4-5 Thursday

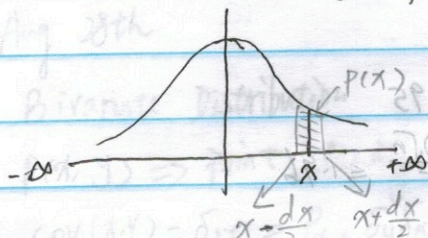
Baldy 115

Aug 26th

Univariate Case: X (income, age, ...)

$$X \sim N(\mu, \sigma^2) \quad -\infty \leq x \leq \infty$$

$p(x)$ prob. density of X



$$p(x)dx = \Pr\left(x - \frac{dx}{2} \leq x \leq x + \frac{dx}{2}\right)$$

$$\int p(x)dx = 1$$

$$\mu = \text{mean of } X = E(X) = \int_{-\infty}^{+\infty} x p(x) dx$$

$$\sigma^2 = \text{Variance of } X = V(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx$$

$E[(X - \mu)^2]$

n = sample size $x_1, x_2, \dots, x_n$; $\bar{x}$ = sample mean = $\frac{x_1 + \dots + x_n}{n} \neq \mu$

$\bar{x}$ is unbiased for μ .

sampling error: $\bar{x} - \mu$

average of $\bar{x} - \mu$ in all possible sample is zero.

$$T_1 \leq T_2$$

If $\Pr[T_1 \leq \mu \leq T_2] = 0.95$, we say (T_1, T_2) is a 95% CI for μ .

$\Rightarrow (T_1, T_2)$ include μ in 95% of the sample.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

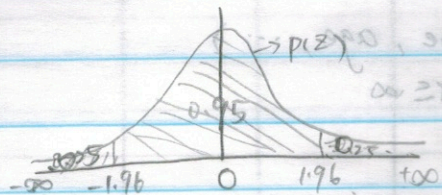
Testing of Hypothesis:

$H_0: \mu = \mu_0$ (specified)

$H_1: \mu \neq \mu_0$ (two sided alternative)

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \rightarrow \text{standardized normal}$$

$\therefore$ under $H_0: \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \sim N(0, 1)$



$$Pr[-1.96 \leq Z \leq 1.96] = 0.95$$

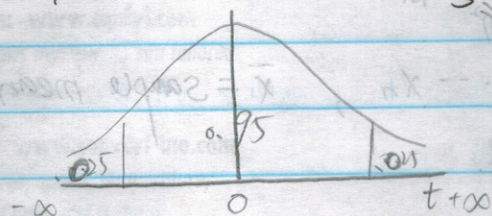
$$Pr[-1.96 \leq \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \leq 1.96] = 0.95$$

Accept H_0 if $-1.96 \leq \frac{(\bar{x} - \mu_0)\sqrt{n}}{\sigma} \leq 1.96$

Reject H_0 if otherwise

If σ^2 unknown, substitute S for σ

Then under $H_0: \frac{(\bar{x} - \mu_0)\sqrt{n}}{S} = t_{n-1}$ (t-variable with d.f. = n-1)



$$Pr[-t_{0.05} \leq t \leq t_{0.05}] = .95$$

$$Pr[-t_{0.05} \leq \frac{(\bar{x} - \mu_0)\sqrt{n}}{S} \leq t_{0.05}] = .95$$

Accept H_0 if $-t_{0.05} \leq \frac{(\bar{x} - \mu_0)\sqrt{n}}{S} \leq t_{0.05}$

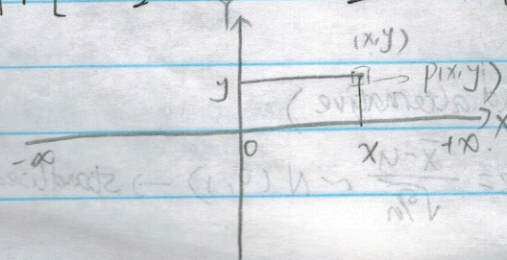
Reject H_0 if otherwise

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$x_1 - \bar{x} + x_2 - \bar{x} + \dots + x_n - \bar{x}$
loss one freedom

Bivariate Distribution: Y, X

$$Pr\left[x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2} \text{ and } y - \frac{dy}{2} \leq Y \leq y + \frac{dy}{2}\right]$$



$$p(x, y) dx dy$$

$$\begin{array}{cc} Y & X \\ \text{means} & \mu_y \quad \mu_x \\ \text{variance} & \sigma_y^2 \quad \sigma_x^2 \\ \text{cov}(X, Y) & \sigma_{xy} \end{array}$$

Aug 28th

Bivariate Distribution

$P(x, y) \Rightarrow$ joint density of Y & X .

$$\text{cov}(X, Y) = \sigma_{xy} = \rho_{xy} \sigma_y \sigma_x = E(X - \mu_x)(Y - \mu_y)$$

$$E(X) = \int_{R_x} \int_{R_y} x P(x, y) dx dy$$

Factorisation Rule:

$$P(x, y) = P_1(x) \cdot P_2(y|x)$$

marginal density of x

conditional density of y

conditional on $X=x$

$$E(X) = \int_{R_x} \int_{R_y} x P_1(x) \cdot P_2(y|x) dx dy$$

$$= \int_{R_x} x P_1(x) \left[\int_{R_y} P_2(y|x) dy \right] dx$$

$$= \int_{R_x} x P_1(x) dx$$

$$V(X) = E(X - \mu_x)^2 = \int_{R_x} \int_{R_y} (x - \mu_x)^2 P(x, y) dx dy = \int_{R_x} \int_{R_y} (x - \mu_x)^2 P_1(x) \cdot P_2(y|x) dx dy$$

$$= \int_{R_x} (x - \mu_x)^2 P_1(x) dx$$

$E(Y|x) =$ average value of Y when $x = x$

$$= \int_{R_y} y P_2(y|x) dy = \phi(x)$$

$\phi(x)$ is called the regression function of Y on X , $Y = \phi(x) + u$

where $E(u|x) = 0$

" distribution
at $bx + u$

two variable linear regression of Y on X

why $E(u|x) = 0$?

$$E(Y|x) = E(\phi(x)|x) + E(u|x)$$

$$\phi(x) = \phi(x) + \underbrace{E(u|x)}_0$$

$$Y = \underbrace{a+bx}_{Y^*} + u$$

$$Y = Y^* + u$$

Y^* - derived

Y - actual Y

$u = Y - Y^*$ error in realizing derived Y

Two implications of $E(u|x) = 0$

1035 B.1

(i) $E(u) = 0$ unconditional mean of $u \Rightarrow E(u) = E_x[E(u|x)]$

(ii) $\text{cov}(x, u) = 0$

$$\begin{aligned} \text{cov}(x, u) &= E(x, u) - E(x) \cdot E(u) = E(x, u) = E_x[E(xu|x)] \\ &= E_x[x E(u|x)] = 0 \end{aligned}$$

Suppose $\text{cov}(x, u) \neq 0$, then we may write $u = \theta x + v$ ($\theta \neq 0$)

when $E(u|x) = 0$, i.e. $E(u|x) = E[\theta x|x] + E(v|x) = \theta x + E(v|x) = 0$

Then $Y = a + bx + u = a + bx + \theta x + v = a + (b + \theta)x + v$ $E(v|x) = 0$

- ② is a neg eq ^{of Y on x} because $E(v|x) = 0$

$$Y = u_y + \varepsilon$$

$$X = u_x + v$$

$$u_y = a + b u_x$$

$$(Y - \varepsilon) = a + b(X - v)$$

$$Y = a + bx + \frac{\varepsilon - bv}{u}$$

$\text{COV}(V, s-bV) = 0$ required to make this a $\hat{y}$ eq

$$\text{COV}(V, s) - bV(V) = 0 \quad b = \frac{\text{COV}(V, s)}{V(V)} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$f = E(u^2)$$

choose a & b so as to $\min_{a,b} f$, $\frac{df}{da} = 0$ $\frac{df}{db} = 0$

$$\Rightarrow b = \frac{\sigma_{xy}}{\sigma_x^2}, \quad a = M_y - bM_x$$

$$\text{COV}(x, y) = \text{COV}(x, a + bx + u) = \text{COV}(x, a) + b \text{COV}(x, x) + \text{COV}(x, u)$$

$$= 0 + b \text{COV}(x, x) + 0 = b \sigma_x^2$$

Sep 2th

popⁿ level:

reg of Y on X :

$$Y = a + bx + u \quad (1)$$

$$\sigma(u|x) = 0 \quad \text{all } x$$

$$V(u|x) = \sigma^2 \quad \text{all } x$$

Note: (a) & in (1), the causation runs from X to Y , ($X = \text{Cause}$, $Y = \text{Effect}$)

In particular, Y cannot cause X (one-way)

(b) In a given app, we cannot have regression of Y on X and of X on Y . (not both)

hours worked H

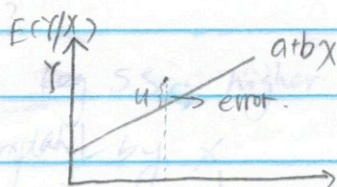
consumption C

Y income

$$C = \beta Y + \hat{u}$$

sampling level:

$$y = a + bx + u$$



estimate: a, b, σ^2

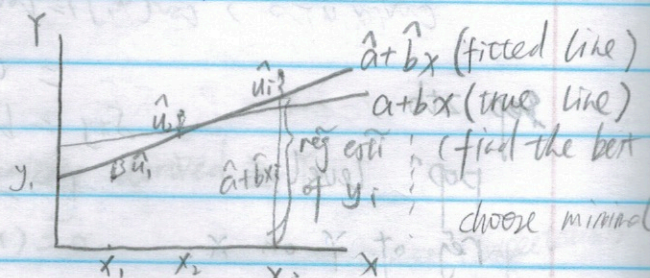
$Y = X$ estimating regression equation (model)

#1 y_1, x_1 $y_1 = a + bx_1 + u_1$

#2 y_2, x_2 $y_2 = a + bx_2 + u_2$

$\vdots$
#n y_n, x_n $y_n = a + bx_n + u_n$

mean $\bar{y}$ $\bar{x}$
variance $\frac{1}{n} \sum (y_i - \bar{y})^2$ $\frac{1}{n} \sum (x_i - \bar{x})^2$
cov $\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$



Choose a, b such that $f = \sum u_i^2 = \sum (y_i - (a + bx_i))^2$ is minimal
(Least Square method of estimating a & b)

$\frac{\partial f}{\partial a} = 0 \rightarrow \sum y_i - (a + b x_i) = 0$

$\frac{\partial f}{\partial b} = 0 \rightarrow \sum x_i (y_i - a - b x_i) = 0$

normal eq of LS

Let $\hat{a}, \hat{b}$ be the solutions $\rightarrow$ LS estimators of a & b ,
 u_i : cannot observe
 $\hat{u}_i$: observe can estimate of u_i respectively.

Define: $\hat{u}_i = y_i - \hat{a} - \hat{b}x_i$ as the i th LS residual

LS estimator of σ^2 is $s^2 = \frac{1}{n-2} \sum \hat{u}_i^2$

$\hat{a} = \bar{y} - \hat{b} \bar{x}$

$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

RSS - Residual Sum of Square

Algebraic Properties

(i) $\sum \hat{u}_i = 0 \Leftrightarrow \sum y_i - (\hat{a} + \hat{b}x_i) = 0 \Leftrightarrow \frac{\partial T}{\partial \hat{a}} = 0$

(ii) $\sum x_i \cdot \hat{u}_i = 0 \Leftrightarrow \sum x_i (y_i - \hat{a} - \hat{b}x_i) = 0 \Leftrightarrow \frac{\partial T}{\partial \hat{b}} = 0$

(iii) Decomposition of $TSS = \sum (y_i - \bar{y})^2$

fixed amount once sample is given

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum \hat{u}_i^2 + \sum (\hat{y}_i - \bar{y})^2 + 2 \sum \hat{u}_i (\hat{y}_i - \bar{y})$$

Where $\hat{y}_i = \hat{a} + \hat{b}x_i$ so that $\hat{u}_i = y_i - \hat{y}_i$

$$= \sum \hat{u}_i^2 + \sum (\hat{y}_i - \bar{y})^2$$

RSS

Reg SS

$$= (1 - r^2) TSS + r^2 TSS$$

unexplained Where $0 \leq r^2 \leq 1$ is the squared corr coeff between x & y
smaller RSS means larger r^2

Sep 4th

Decomposition of $TSS = \sum (y_i - \bar{y})^2$

$$\sum (y_i - \bar{y})^2 = \underbrace{\sum_{i=1}^n \hat{u}_i^2}_{RSS} + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{Reg SS}$$

How much of the variation Y (TSS) is caused by (explained by) the variation in X ?

This is indicated by Reg SS; higher the Reg SS, larger the fraction of TSS explained by X .

Explain: 100% of variation of Y is explained by the variation of X

Analyse of Variance Table. (ANOVA Table)

Source	df	SS	MSS = SS/df
Regression	k-1	$r^2 TSS$	$r^2 TSS / (k-1)$
Error/residual	n-k	$(1-r^2) TSS$	$(1-r^2) TSS / (n-k)$
Total	n-1	TSS	

Sampling Properties:

estimates: $\hat{a}, \hat{b}, \hat{S}^2$
 $\downarrow \quad \downarrow \quad \downarrow$
 a, b, σ^2

$$y_i = a + b x_i + u_i \quad E(u_i | x_i) = 0 \quad V(u_i | x_i) = \sigma^2$$

$$y_n = a + b x_n + u_n \quad E(u_n | x_n) = 0 \quad V(u_n | x_n) = \sigma^2$$

minimally, we must have $E(u_i | x_i) = 0$, we make a stronger

assumption $E(u_i | x_1, x_2, x_3, \dots, x_n) = 0$ (There is no externality in y_i)

~~$$E(u_i | x_i) = 0$$~~

→ The u 's are i.i.d, normal $(0, \sigma^2)$

Two sampling schemes, depending on whether we treat x 's

(i) fixed (general case) constants:

(ii) random

$\hat{b}$:

(i) $\hat{b}$ is unbiased for b

$$E(\hat{b} | x's) = b$$

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \frac{\sum w_i (a + b x_i + u_i)}{\sum w_i} \downarrow \frac{\sum (x_i - \bar{x})}{h^2}$$

$$\sum_{i=1}^n w_i (x_i - \bar{x}) = 0$$

$$= 0 \sum w_i + b \sum w_i x_i + \sum w_i u_i$$

$$= 0 + b \cdot 1 + \sum w_i u_i$$

$\hat{b} = b + \sum w_i u_i$ sampling error in $\hat{b}$ as an est of b

$$\therefore E(\hat{b} | x's) = b + E(\sum w_i u_i | x's)$$

$$= b + \sum w_i E(u_i | x's)$$

$$= b + 0$$

so unbiased

The unbiasedness result comes about only because of the "Orthogonality Condition"

u_i 's are orthogonal to x_i 's

u_i 's — independent of x_i 's

each u_i is indep of all x_i 's

Suppose x_i 's are random

$$E(\hat{b} | x's) = b, \text{ (even when } x_i \text{'s random are)}$$

i.e. this result doesn't involve x_i 's; therefore the unbiasedness result remained valid even when x_i 's are random.

More specifically, we have $E(\hat{b}) = b$

In general, if a conditional result doesn't involve conditioning variable, then the conditional result remain valid unconditionally as well, as when these variables allowed to vary.

When x_i 's are random, what we seek are unconditional results.

Sep 9th

$$E(\hat{\beta} | X's) = b$$

$$\therefore E(\hat{\beta}) = b$$

$$\begin{aligned} V(\hat{\beta} | X's) &= \text{Var}(b + \sum w_i u_i | X's) = \text{Var}(\sum w_i u_i) \\ &= \sum V(w_i u_i | X's) = \sum w_i^2 V(u_i | X's) \\ &= \sigma^2 \sum w_i^2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \end{aligned}$$

$$\frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \sigma_b^2, \text{ say}$$

(1) $\sigma_b^2 \downarrow$, when $\sum (x_i - \bar{x})^2 \uparrow \rightarrow$ choose X 's as far apart as possible. $\Rightarrow \text{eff}(\hat{\beta}) \uparrow$

(2) since σ_b^2 depends on the x 's, $V(\hat{\beta}) \neq V(\hat{\beta} | X's)$

$$V(\hat{\beta}) = \underbrace{E_x[V(\hat{\beta} | X's)]}_{\text{constant, given } X's} + \underbrace{V[E(\hat{\beta} | X)]}_0$$

$$\hat{\alpha} = a + \sum_{i=1}^n l_i u_i$$

$$E(\hat{\alpha} | X's) = a$$

$$V(\hat{\alpha} | X's) = V(\sum l_i u_i | X's) = E(\hat{\alpha})$$

$$= \sigma^2 \sum l_i^2$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right) \neq V(\hat{\alpha})$$

$$E(\hat{\beta}^2 | X's) = \sigma^2 \equiv E(S^2)$$

Gauss - Markoff Thm.

relates to LS est^s of coeffs.

$$\text{LS: } \hat{a}, \hat{b} \quad y_i = a + bx_i + u_i$$

$$\text{Alt: } \tilde{a}, \tilde{b}$$

"at least as efficient as"

$$\hat{a} \geq \tilde{a}, \text{ when } \tilde{a} \text{ is, like } \hat{a},$$

linear unbiased est of a .

$\hat{a}$ is linear in $y_1, \dots, y_n$

$\rightarrow \hat{a}$ is a linear est

$\rightarrow \hat{a}$ is unbiased for a .

$\hat{a}$ is BLUE of a ($\hat{a} \geq \tilde{a}$)

best linear unbiased estimator.

similarly, $\hat{b}$ is BLUE of b ($\hat{b} \geq \tilde{b}$)

The Thm actually makes stronger statements regarding the LS est

Testing of Hypothesis / Confidence Intervals

Ass: u_i 's are normally distributed each with mean zero, and variance σ^2 ; u_i 's are, as before, indep.

$$y_i = a + bx_i + u_i$$

$$y_n = a + bx_n + u_n$$

$$y_1, y_2, \dots, y_n \rightarrow N$$

with mean $y_i = a + bx_i$

$$\text{Var } y_i = \sigma^2$$

$$E(u_i | x_i) = 0$$

$$V(u_i | x_i) = \sigma^2$$

x_i 's fixed

$$\text{Cov}(u_i, u_j) = 0$$

$$u_i \sim N(0, \sigma^2)$$

Gauss - Markoff

set up

Consider $\hat{b}$

$\hat{b}$ is linear in $y_0, \dots, y_n$, which are normally distributed.
So, $\hat{b}$ is also normally distributed, with, as before, mean = b , var = σ_b^2

[Any linear function of normals is also normal]

$H_0: b = b_0$ (specified)

we have

$b \neq b_0$ (two-sided)

$$\hat{b} \sim N(b, \sigma_b^2)$$

$$\therefore Z = \frac{\hat{b} - b}{\sigma_b} \sim N(0, 1)$$

$$\therefore \text{under } H_0, Z = \frac{\hat{b} - b_0}{\sigma_b} \sim N(0, 1)$$

$$\sigma_b^2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$S_b^2 = \frac{S^2}{\sum (x_i - \bar{x})^2}$$

$$\therefore t = \frac{\hat{b} - b_0}{S_b} \sim t_{n-2}, \text{ a } t\text{-variable with d.f.} = n-2$$

Sep 11th

$H_0: b = b_0$ (specified)

$H_1: b \neq b_0$

$$t = \frac{\hat{b} - b_0}{S_b} \sim t_{n-2}, \text{ } t\text{-variable with } (n-2) \text{ d.f.}$$

Accept H_0 if $-t_{\alpha/2} \leq t \leq t_{\alpha/2}$, $\hat{b} - t_{\alpha/2} \cdot S_b \leq b_0 \leq \hat{b} + t_{\alpha/2} \cdot S_b$

Reject H_0 otherwise

Special case

$H_0: b = 0$ (i.e. $t_{10} = \rho = 0$)

$H_1: b \neq 0$ (i.e. $t_{10} = \rho \neq 0$)

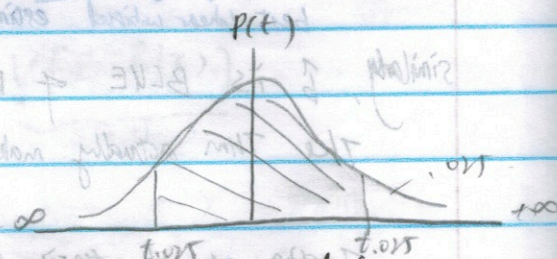
$$t = \frac{\hat{b} - 0}{S_b} \text{ (regression-based test)}$$

→ Goodness of Fit test

correlation-based test: $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$

$$\hat{b} = r \cdot \frac{s_y}{s_x} = \frac{\frac{1}{n} \sum (x_i y_i)}{\frac{1}{n} \sum (x_i^2)}$$

$$S_b^2 = \frac{S^2}{\sum (x_i - \bar{x})^2} = \frac{S^2}{n S_x^2}$$



ANOVA Table

Regression	$MS = SS/df$
Residual	$(1-r^2)TSS/n-k$

ANOVA Test is an F -test (Goodness of fit test)

$$F = \frac{a}{b} = \frac{r^2(n-2)}{1-r^2} \sim F_{1, n-2}$$

$H_0: b=0$ (slopes are zero)

Accept H_0 if $F \leq F_{0.05}$

$F = t^2$ (in two variable case) only

$$\left. \begin{array}{l} H_0: a=a_0 \\ H_1: a \neq a_0 \end{array} \right\} t = \frac{\hat{a} - a_0}{SE}$$

$H_0: a=a_0$ and $b=b_0$

H_1 : Negation of H_0
more efficient, take into account of the correlation of a & b

VS

(i) $H_0: a=a_0$
 $H_1: a \neq a_0$

(ii) $H_0: b=b_0$
 $H_1: b \neq b_0$

This t -test procedure remains valid when x 's are random also, because the distribution of t under H_0 (i.e. the null distribution of t) depends only on the degrees of freedom, and not in particular, on $x_1, \dots, x_n$

Let H_1 state that $b=b_1 > b_0$ (say) > 0

under H_1 , we have

$$t = \frac{\hat{b} - b_0 + b_1 - b_0}{S_b} = \frac{\hat{b} - b_0}{S_b} + \frac{b_1 - b_0}{S_b}$$

behaves like a t_{n-2} + a positive variable $Z = \frac{b_1 - b_0}{S_b}$

larger the S is, higher the power of the t-test

Power = Prob of Accepting H_1 , where H_1 is true
where H_0 is false

95% CI for b

T_1 (sample value)

T_2 (Sample value)

$T_1 < T_2$

If we can claim that $P[T_1 \leq \theta \leq T_2] = .95$, then we say that $[T_1, T_2]$ is a 95% CI for θ .

we know that, for true b ,

$$t = \frac{\hat{b} - b}{S_b} \sim t_{n-2}, \text{ df} = n-2$$

and hence that

$$Pr[-t_{.05} \leq t \leq t_{.05}] = .95$$

$$P[\hat{b} - t_{.05} S_b \leq b \leq \hat{b} + t_{.05} S_b] = .95$$

$$\therefore T_1 = \hat{b} - t_{.05} S_b$$

$$T_2 = \hat{b} + t_{.05} S_b$$

Sep 16th

Prediction / forecasting

$$\text{model: } y_i = a + b x_i + u_i$$

$$\hat{a}, \hat{b}; \hat{S}$$

$$y_n = a + b x_n + u_n$$

$$y_{n+1} = a + b x_{n+1} + u_{n+1}$$

True Problem

$$E(u_{n+1}) = 0$$

(i) estimate $E(y_{n+1} | x_{n+1}, x's, u's) = a + b x_{n+1} + \underbrace{E(u_{n+1} | x_{n+1}, x's, u's)}_0$

→ prediction problem

(ii) ~~est~~ estimate y_{n+1} itself (forecast)

(1) Est $y_{n+1}^* = \text{est}(a) + \text{est}(b x_{n+1})$

$$= \hat{a} + \hat{b} x_{n+1}$$

$$= y_{n+1}^p \text{ (say)}$$

The y_{n+1}^p is unbiased for y_{n+1}^* , under normality of $u's$ (e_p is intrinsic linear in $u's$),

Also, the prediction error is $e_p = y_{n+1}^p - y_{n+1}^*$ constant in $u's$,

$$= \hat{a} + \hat{b} x_{n+1} - (a + b x_{n+1})$$

$$= (\hat{a} - a) + (\hat{b} - b) x_{n+1}$$

clearly, $E(e_p | x_{n+1}, x's) = 0$ if e_p is linear in $u's$

if it is conditional on $u's$, then $\hat{a}, \hat{b}$ is given

(Assignment 11) $V(e_p | x_{n+1}, x's) = \sigma^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] = \sigma_p^2 \text{ (say)}$

e_p is also normal $e_p \sim N(0, \sigma_p^2)$

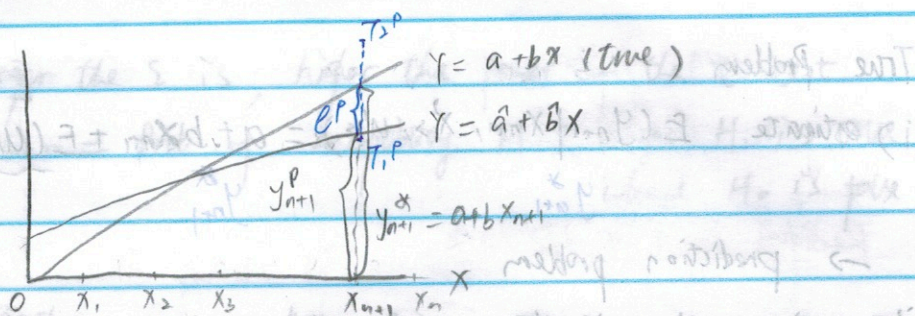
$$\therefore z = \frac{e_p}{\sigma_p} \sim N(0, 1)$$

$$t = \frac{e_p}{s_p} \sim t_{n-2}$$

$$s^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] = s_p^2$$

$$= \frac{y_{n+1}^p - y_{n+1}^*}{s_p}$$

$$\therefore T_1 = y_{n+1}^p - t_{0.05} s_p \quad T_2 = y_{n+1}^p + t_{0.05} s_p$$



(ii) Forecasting Problem:

estimate y_{n+1} itself

$$y_{n+1} = y_{n+1}^o + u_{n+1}$$

$$y_{n+1}^f = \text{est } y_{n+1} = \text{est } y_{n+1}^o + \text{est } u_{n+1}$$

$$= \text{est } y_{n+1}^o$$

$$= y_{n+1}^p$$

$$\text{est } y_{n+1} = y_{n+1}^f = y_{n+1}^p$$

In production, y_{n+1}^p estimates $(a + bx_{n+1})$

In forecasting, y_{n+1}^f estimates $[a + bx_{n+1} + u_{n+1}]$

But y_{n+1}^p is the same as y_{n+1}^f , because of this, we should expect that the y_{n+1}^f to perform worse than y_{n+1}^p .

Define the forecast error as $e_f = y_{n+1}^f - y_{n+1}$

$$= y_{n+1}^p - (a + bx_{n+1} + u_{n+1})$$

$$= -u_{n+1} + (y_{n+1}^p - y_{n+1}^o)$$

$$= -u_{n+1} + e_p$$

clearly, $E(e_f | x_{n+1}, x's) = 0$

$$V(e_f | x_{n+1}, x's) = V(u_{n+1} + e_p) = \frac{10}{1} \cdot 8$$

$$= V(u_{n+1}) + V(e_p) \quad (\because u_{n+1}, e_p \text{ are indep})$$

$$= \sigma^2 + \sigma_p^2 = \sigma_f^2 \text{ (say)}$$

Not unexpectedly, it is true that $\sigma_f^2 \geq \sigma_p^2$

Sep 18th

Prediction

Forecasting

$$y_{n+1} = a + b x_{n+1} + u_{n+1}$$

$$= y_{n+1}^* + u_{n+1}$$

Target

y_{n+1}^*

$$y_{n+1}^* + u_{n+1} = y_{n+1}$$

The performance of y_{n+1}^p is an est of y_{n+1}^* ($a + b x_n$) is naturally stronger than its performance as an est of $(y_{n+1}^* + u_{n+1} = y_{n+1})$

$$e_p = y_{n+1}^p - y_{n+1}^* = \text{predict error}$$

$$V(e_p) = \sigma_p^2$$

$$e_f = y_{n+1}^p - y_{n+1} = y_{n+1}^p - (a + b x_{n+1} + u_{n+1}) = y_{n+1}^p - y_{n+1}^* - u_{n+1}$$

$$= e_p - u_{n+1}$$

$$V(e_f) = \sigma_f^2 = \sigma_p^2 + \sigma^2$$

$$V(e_p) < V(e_f)$$

95% CI for y_{n+1}

$$e_t \sim N(0, \sigma_f^2)$$

$$Z = \frac{e_t}{\sigma_f} \sim N(0, 1)$$

$$t = \frac{e_t}{S_f} \sim t_{n-2}$$

$$S_f^2 = S_p^2 + S^2$$

$$\therefore y_{n+1}^L = t_{\alpha/2} S_f, \quad y_{n+1}^U = t_{\alpha/2} S_f$$

$\therefore$ 95% CL for y_{n+1} is

$$T_1^L = \hat{a} + \hat{b}x_{n+1} - t_{\alpha/2} S_f, \quad T_2^U = \hat{a} + \hat{b}x_{n+1} + t_{\alpha/2} S_f$$

$$\text{Length of the CL for } y_{n+1} = 2t_{\alpha/2} S_f$$

$$y_{n+1}^* = \hat{a} + \hat{b}x_{n+1} \quad (T_2^U - T_1^L) = 2t_{\alpha/2} S_f$$

$$y_{n+1} = a + bx_{n+1} + u_{n+1}$$

known

known

$(y_{n+1}, x_{n+1}) \rightarrow$ new observation

H_0 : no structure break

If this observation can be assumed to have confront the same pop ~~from~~ ^{as} the sample came, then it must be added on to the sample and the regression should be re-estimated.

Let a_{n+1}, b_{n+1} denote LS est of a, b (respectively), and

Let $a_n = \hat{a}, b_n = \hat{b}$. Then $a_{n+1} \geq a_n$ means at least as

$$\Rightarrow V(a_{n+1}) \leq V(a_n)$$

$$V(b_{n+1}) \leq V(b_n)$$

Accept H_0 if $T_1^L \leq y_{n+1} \leq T_2^U$ (structure break test)

If H_0 is accepted, add the $(n+1)^{\text{th}}$ obs to the sample and obtain a_{n+1}, b_{n+1} .

Sep 23rd

Maximum Likelihood Estimation (MLE)

$$y_i = a + bx_i + u_i$$

$$y_n = a + bx_n + u_n$$

Must specify the distri of u_i 's

Assume: $u_i \sim NID(0, \sigma^2)$

(1) obtain the likelihood of the sample

$$L(a, b, \sigma^2 | \text{data}) = P(y_1, y_2, \dots, y_n)$$

$$= P(u_1, \dots, u_n) \times |J|$$

where J is the Jacobian of the transformation from u_i 's into y_i 's
and $||$ denotes absolute value of $\det J$

$$J = \frac{\partial(u_1, u_2, \dots, u_n)}{\partial(y_1, y_2, \dots, y_n)} = \begin{bmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \dots & \frac{\partial u_1}{\partial y_n} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \dots & \frac{\partial u_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_n}{\partial y_1} & \frac{\partial u_n}{\partial y_2} & \dots & \frac{\partial u_n}{\partial y_n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

$$L = P(u_1, \dots, u_n) = \prod_{i=1}^n P(u_i)$$

$$P(u_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{u_i^2}{2\sigma^2}}$$

$$L = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \cdot e^{-\sum \frac{u_i^2}{2\sigma^2}} = L(a, b, \sigma^2 | \text{data})$$

to be maximized w.r.t a, b, σ^2

(2) The maximizing value of a, b, σ^2 are called the m.l.e's of a, b, σ^2 (respectively)

(3) Maximization proceeds as follows

Step 1: Set $L(a, b, \sigma^2 | \text{data}) = \log L(a, b, \sigma^2 | \text{data})$

Set $\frac{\partial L}{\partial \sigma^2} = 0$

$$L(a, b, \sigma^2 | \text{data}) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum u_i^2$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4} \sum (y_i - a - bx_i)^2 = 0$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - a - bx_i)^2 = 0$$

$$\sigma^2 = \frac{1}{n} \sum (y_i - a - bx_i)^2$$

which is the conditional m.l.e of σ^2 given a & b

$$L^* = -\frac{n}{2} \log \left[\frac{1}{n} \sum (y_i - a - bx_i)^2 \right] - \frac{1}{2\sigma^2} \sum (y_i - a - bx_i)^2$$

constant $n/2$

Next, we need to max L^* w.r.t a & b .

$$(max) L^* = \text{constant} - \frac{n}{2} \log \left[\frac{1}{n} \sum (y_i - a - bx_i)^2 \right]$$

$= \min (y_i - a - bx_i)^2$ the same as LS

$\therefore$ m.l.e of $a = \tilde{a}$, say $= \hat{a}$

m.l.e of $b = \tilde{b}$, say $= \hat{b}$.

Also, then the m.l.e of σ^2 is

$$\hat{S}^2 = \hat{\sigma}^2 \text{ at } a = \hat{a}, b = \hat{b} = \frac{1}{n} \sum (y_i - \hat{a} - \hat{b}x_i)^2 = \frac{1}{n} \text{RSS}$$

Under normality of u 's, it is ^{divided} that

ML for the coefficient = LS for the coefficient

However, for σ^2 , $ML \neq LS$

$$\therefore \frac{\text{RSS}}{n} = \hat{S}^2 \neq S^2 = \frac{\text{RSS}}{n-2}$$

Note; S^2 is unbiased for σ^2 (whether or not x 's are fixed or

$\therefore \hat{S}^2$ is biased for σ^2 random)

$$E(\hat{S}^2) = E\left[\frac{\text{RSS}}{n}\right] = E\left[\frac{(n-2)S^2}{n}\right] = \frac{n-2}{n} E[S^2] = \frac{n-2}{n} \sigma^2$$

if in large samples, as $n \rightarrow \infty$, it becomes unbiased

so $\hat{S}^2$ is asymptotically unbiased for σ^2

Covariance and variance of a, b are the same as in LS

(when in complicated situation, LS can not work)

In general, ML gives asymptotically efficient estimates, which, though are biased in small samples (biases $\rightarrow 0$ as $n \rightarrow \infty$)

Sep 15th

constant term or no constant term

Two types of misspecification

(1) True: $y = a + bx + u$ ($a \neq 0$)

False: $y = bx + u$ ($a = 0$) is used ^{missp of} omission of a relevant

$$\text{From false: } \hat{b}_0 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (a + bx_i + u_i)}{\sum x_i^2} = \frac{a \sum x_i}{\sum x_i^2} + b + \frac{\sum x_i u_i}{\sum x_i^2}$$

$$E(\hat{b}_0 | x\text{'s}) = a \frac{\sum x_i}{\sum x_i^2} + b \neq b, \text{ unless } \sum x_i = 0$$

$\therefore \hat{b}_0$ is biased for b

$\hat{b}_0$ is biased

$\hat{b}$ is unbiased

$$V(\hat{b}_0 | x's) = \frac{\sigma^2}{\sum x_i^2} = \sigma_{b_0}^2$$

(2) True: $y = bx + u$

False: $y = a + bx + u$

is used

misspecification of commission - adding an irrelevant represents variable

True

$$\hat{b}_0 = \frac{\sum x_i u_i}{\sum x_i^2}$$

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \text{ an estimate of } b$$

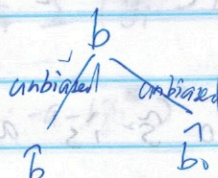
$E(\hat{b} | x's) = b$
unbiased

$$= \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(bx_i + u_i)}{\sum (x_i - \bar{x})^2} = b + \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$$

$E(\hat{b} | x's) = b \rightarrow$ unbiased

$$V(\hat{b} | x's) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \sigma_b^2 \text{ (say)}$$

$$= \frac{\sigma^2}{\sum x_i^2 - nx\bar{x}^2}$$



$$\therefore V(\hat{b} | x's) > V(\hat{b}_0 | x's)$$

$$\hat{b}_0 > \hat{b}$$

$\hat{b}_0$ and $\hat{b}$ are both unbiased

— Three variable case

$$Y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$\text{Variance: } S_1^2 = \frac{1}{n} \sum (y - \bar{y})^2$$

$$S_2^2 = \frac{1}{n} \sum (x_2 - \bar{x}_2)^2$$

$$S_3^2 = \frac{1}{n} \sum (x_3 - \bar{x}_3)^2$$

	constant	x_2	x_3
y_1	✓	✓	✓
y_2	✓	✓	✓
$\vdots$			
y_n	✓	✓	✓

mean, $\bar{y}, \bar{x}_2, \bar{x}_3$

number

$$\text{Cov: } S_{12} = \frac{1}{n} \sum (y - \bar{y})(x_2 - \bar{x}_2) = r_{12} S_1 S_2$$

$$S_{p2} = \frac{1}{n} \sum (y - \bar{y})(x_3 - \bar{x}_3) = r_{13} S_1 S_3$$

$$S_{23} = \frac{1}{n} \sum (x_2 - \bar{x}_2)(x_3 - \bar{x}_3) = r_{23} S_2 S_3$$

$$E(b_1 | x_1) = \frac{\sum x_i (b x_i + u_i)}{\sum x_i^2} = b + \frac{\sum x_i u_i}{\sum x_i^2} = b$$

$$LS: \min_{\beta_1, \beta_2, \beta_3} f = \sum u_i^2 = \sum (y_i - \beta_1 - \beta_2 x_{i2} - \beta_3 x_{i3})^2$$

$$\frac{\partial f}{\partial \beta_1} = 0 \Rightarrow \beta_1 = \bar{y} - \beta_2 \bar{x}_2 - \beta_3 \bar{x}_3$$

$$\frac{\partial f}{\partial \beta_2} = 0 \Rightarrow \beta_2 \cdot \sum (x_{i2} - \bar{x}_2)^2 + \beta_3 \sum (x_{i3} - \bar{x}_3)(x_{i2} - \bar{x}_2) = \sum (x_{i2} - \bar{x}_2)(y_i - \bar{y}) \quad (1)$$

$$\frac{\partial f}{\partial \beta_3} = 0 \Rightarrow \beta_2 \sum (x_{i2} - \bar{x}_2)(x_{i3} - \bar{x}_3) + \beta_3 \sum (x_{i3} - \bar{x}_3)^2 = \sum (x_{i3} - \bar{x}_3)(y_i - \bar{y}) \quad (2)$$

put cov relation into (1), (2)

$$\hat{\beta}_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \times \frac{s_1}{s_2}$$

$$\hat{\beta}_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \times \frac{s_1}{s_3}$$

comments on $\hat{\beta}_2$: suppose we regress y on x_2 only,

$$b_2 = \frac{\sum (x_{i2} - \bar{x}_2)(y_i - \bar{y})}{\sum (x_{i2} - \bar{x}_2)^2} \quad \text{Vs } \hat{\beta}_2$$

$$= r_{12} \frac{s_1}{s_2} \quad \text{Vs } \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \cdot \frac{s_1}{s_2}$$

(i) $\text{sign } \hat{\beta}_2 \neq \text{sign } b_2$ in general

when this is 0, we can drop x_2
 $\text{sign}(r_{12} - r_{13}r_{23}) \neq \text{sign}(r_{12})$

(ii) If $r_{23} = 0$, it does not necessarily follow that we can drop x_3 from the regression.

(iii) we can drop x_2 when $r_{12} = 0$ and $r_{23} = 0$

(iv) If $r_{23} = 0$, $\hat{\beta}_2 = b_2$, $\hat{\beta}_3 = b_3$ (when b_3 is obtained by regression y on x_3 only)

(V) If $r_{23}^2 \approx 1$, we have the problem of near-perfect

Feb

Sep 30th

(Vi) $r_{23}^2 = 1$, $\hat{\beta}_2, \hat{\beta}_3$ will both blow up and so will their SE's, with result that the respective t-value ($\hat{\beta}_2/SE, \hat{\beta}_3/SE$) will be very small

→ multicollinearity problem.

$r_{23}^2 = 1 \Leftrightarrow X_2$ and X_3 are exact-linear function of each other

$$X_2 = h + gX_3 \text{ (say)}$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

$$= \beta_1 + \beta_2 (h + gX_3) + \beta_3 X_3 + u$$

$$= \beta_1 + \beta_2 h + (\beta_2 g + \beta_3) X_3 + u$$

$$= \text{const} + CX_3 + u$$

Ex: $Unemp = \beta_1 + \beta_2 GNP + \beta_3 r + u$

need to drop one, right

$\Rightarrow$ exact connection between $\hat{\beta}_2$ and $\hat{\beta}_3$

$$Y = C + b_2 X_2 - A$$

$$Y = C + \beta_2 X_2 + \beta_3 X_3 - B$$

(B) is likely to be more correct than (A) because (A) is a special case of (B) (which holds when $\beta_3 = 0$) $\rightarrow$ B encompasses (A). If it is known that (B) is correct, then (A) is false.

one normal eqⁿ for $\hat{\beta}_1$ & $\hat{\beta}_3$ is $\sum (x_1 - \bar{x}_1) = 0$ pos

$$\hat{\beta}_2 \sum (x_2 - \bar{x}_2) + \hat{\beta}_3 \sum (x_3 - \bar{x}_3)(x_2 - \bar{x}_2) = \sum (x_2 - \bar{x}_2)(y - \bar{y})$$

$$\hat{\beta}_2 + \hat{\beta}_3 \frac{\sum (x_2 - \bar{x}_2)(x_3 - \bar{x}_3)}{\sum (x_2 - \bar{x}_2)^2} = \frac{\sum (x_2 - \bar{x}_2)(y - \bar{y})}{\sum (x_2 - \bar{x}_2)^2} = \hat{b}_2$$

(connection.) $\hat{\beta}_2 + \hat{\beta}_3 \hat{d}_2 = \hat{b}_2$

$\hat{d}$ is the LS est of d in $x_3 = c + d x_2$, which is the "auxiliary" regression of x_3 on x_2 .

$$\Delta Y = 0 + \hat{\beta}_2 \Delta X_2 + \hat{\beta}_3 \Delta X_3$$

x_3 is controlled
 x_3 is accounted for
 x_3 is allowed for

Direct effect of $\Delta x_2 = 1$ on Y , when $\Delta x_3 = 0$, i.e. when x_3 is held constant

Y	x_2	x_3
#1	✓	5
#2	✓	5
⋮	⋮	⋮
#n	✓	5

Run regression like this cannot get $\hat{\beta}_2$, because it becomes two-variable regression.

$$Y = \beta_1 + \beta_2 x_2 + \beta_3 \cdot 5 + u$$

constant

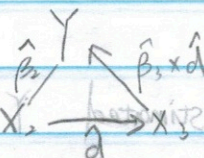
$$= (\beta_1 + 5\beta_3) + \beta_2 x_2 + u$$

estimate form:

$$Y = \sqrt{+} + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

$$= \sqrt{+} + \hat{\beta}_2 x_2 + \hat{\beta}_3 (c + \hat{d} x_2)$$

$$= \sqrt{+} + (\hat{\beta}_2 + \hat{\beta}_3 \hat{d}) x_2$$



total effect for $\Delta x_2 = 1$ is $\hat{\beta}_2 + \hat{\beta}_3 \hat{d}$

Exp,

$$\log Q = \alpha_0 + \alpha_1 \log K + \alpha_2 \log L + U$$

use: $\log Q = c_0 + c_1 \log L$

$$\log K = c + d \log L$$

$$\hat{c}_1 = \hat{a}_2 + \frac{\hat{a}_1 d}{\hat{a}_2} \Rightarrow \hat{c}_1 > \hat{a}_2$$

Oct 2nd

parent's edu

$$E = \beta_1 + \beta_2 S + \beta_3 S_p + U$$

$$\text{cov}(Y, X_2) = 0 \quad r_{12} = 0$$

$$\text{cov}(X_2, X_3) \neq 0 \quad r_{23} \neq 0$$

$$\hat{\beta}_2 = \frac{r_{12} - r_{13}r_{23}}{1 - r_{23}^2} \cdot \frac{S_1}{S_2}$$

even $r_{12} = 0$, $\beta_3 \neq 0$

$$\hat{\beta}_3 = \frac{r_{13} - r_{12}r_{23}}{1 - r_{23}^2} \cdot \frac{S_1}{S_3}$$

cannot drop X_3

$$E = \beta_1 + \beta_2 S + \beta_3 \text{Age} + U$$

$$r_{23} = 0$$

$$r_{12} \neq 0$$

$$\hat{\beta}_2 = r_{12} \frac{S_1}{S_2} = \hat{b}_2$$

Y on X_2 only

$$\hat{\beta}_3 = r_{13} \frac{S_1}{S_3} = \hat{b}_3$$

Y on X_3 only

$$E = \beta_1 + \beta_2 S + \beta_3 \text{church} + U$$

$$r_{23} = 0$$

$$r_{12} = 0$$

in this case, can drop X_3

When in two cases, $\hat{b}_2$ is called, total reg

$\beta_2, \beta_3 \rightarrow$ partial regression coefficient

Estimated $Y = c + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3$

estimate $\hat{\beta}_2$ from $\hat{b}_2$, controlling for X_3 .

1) regress Y on X_3 :

$$Y = a_y + f \cdot X_3 + \text{erro} \rightarrow (\hat{a}_y, \hat{f}) \rightarrow S_i = Y_i - \hat{a}_y - \hat{f}X_{i3}, \text{ which is the}$$

LS residual, free from the linear influence from $X_3 \rightarrow$ partial value of Y

(2) Regress X_2 on X_3 .

$X_2 = c + gX_3 + \text{error} \rightarrow (\hat{c}, \hat{g}) \rightarrow y_i = X_{i2} - \hat{c} - \hat{g}X_{i3}$ which is the LS residual free from the linear influence from $X_3 \rightarrow$ purified value of X_{i2}

(3) Regress purified Y on purified X_2 .

$\varepsilon_i = \theta y_i + \text{error}$ $\mu = m(\varepsilon_i) + \theta m(y_i)$ mean of residual is 0, ~~it has~~ constant here.

$$\text{Then } \theta = \frac{\sum y_i \varepsilon_i}{\sum \varepsilon_i^2} = \hat{\beta}_2 = \frac{\sum \varepsilon_i}{\sum y_i} = \frac{1}{n} \sum \varepsilon_i = \frac{1}{n} \sum y_i^2$$

Note: $\text{corr}(Y, X_2) = r_{12}$ total corr coefficient between Y & X_2

$$\sum \varepsilon_i^2 = TSS_y (1 - r_{12}^2)$$

$$\sum y_i^2 = TSS_{X_2} (1 - r_{23}^2)$$

$$\sum (X_2 - \bar{X}_2)^2$$

General Linear Regression

Multiple Regression

$$\text{Pop}^n: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + u$$

$$(i) E(u|X's) = 0 \Rightarrow E(Y|X's) = \text{linear in } X_1, X_2, \dots, X_n$$

$$(ii) V(u|X's) = \sigma^2$$

We take $X_1 \equiv 1$ to accommodate for a constant term, which then is β_0 .

Oct 7th

Assumption: All first and second moments of $Y, X_1, \dots, X_k$ exist
 $\rightarrow$ means, variances, covariances all exist

Sampling

	Y	$X_1=1$	X_2	$\dots$	X_k
#1	y_1	1	x_{12}	x_{1k}	$\Rightarrow y_1 = \beta_1 + \beta_2 x_{12} + \dots + \beta_k x_{1k}$
#2	y_2	1	x_{22}	x_{2k}	$\Rightarrow y_2 = \beta_1 + \beta_2 x_{22} + \dots + \beta_k x_{2k}$
$\vdots$					
#n	y_n	1	x_{n2}	x_{nk}	$\Rightarrow y_n = \beta_1 + \beta_2 x_{n2} + \dots + \beta_k x_{nk}$

define $\beta_{(k \times 1)} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$ $y_{(n \times 1)} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ $u_{(n \times 1)} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ $X_{(n \times k)} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$

$$y_{(n \times 1)} = X_{(n \times k)} \cdot \beta_{(k \times 1)} + u_{(n \times 1)}$$

The GM model is

$$y = X\beta + u$$

$E(u_i | \text{all } x\text{'s}) = E(u_i | X) = 0 \Rightarrow E(u | X) = 0$ orthogonal

$E(u_i | X) = 0$

$E(u_n | X) = 0$

(2) X may be fixed or stochastic, if stochastic

(3) $r(X) = k < n$, if X stochastic, $r(X) = k$

(4) The u 's are iid $(0, \sigma^2)$ $[V(u_i | X) = \sigma^2]$

$\rightarrow$ Vector (or matrix) of expectation = Expectation of the Vector (Matrix)

	$u_1, u_2, \dots, u_n$
u_1	$\sigma_1^2, \sigma_{12}, \dots, \sigma_{1n}$
u_2	$\sigma_{21}, \sigma_{22}, \dots, \sigma_{2n}$
$\vdots$	
u_n	$\sigma_{n1}, \sigma_{n2}, \dots, \sigma_{nn}$

$V(u | X) \stackrel{\text{def}}{=} E(uu' | X)$ u 's are homoscedastic

$$E(uu') = \begin{pmatrix} E(u_1^2) & E(u_1 u_2) & \dots & E(u_1 u_n) \\ E(u_2 u_1) & E(u_2^2) & \dots & E(u_2 u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_n u_1) & E(u_n u_2) & \dots & E(u_n^2) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

$= E(u_n)E(u_1)$

$$\frac{\partial L}{\partial \beta} = -X' y \quad \frac{\partial L}{\partial \beta} = -X' X \beta$$

LS estimation

$$\min_{\beta} [f = \sum u_i^2 = u'u = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta]$$

$$\frac{\partial f}{\partial \beta_1} = 0$$

$$\frac{\partial L}{\partial \beta} = -2X'y + 2X'X\beta = 0$$

$$\frac{\partial \mathcal{L}}{\partial \beta_n} = 0$$

$$\underset{(K \times K)}{X'X} \beta = \underset{(K \times 1)}{X'y}$$

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

$$\begin{aligned} &\equiv x_1' \beta + u_1 \\ &\equiv x_2' \beta + u_2 \\ &\vdots \\ &\equiv x_n' \beta + u_n \end{aligned} \Rightarrow \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

This system, known as normal eqⁿ, consists of k eqⁿ in k unknowns: $\beta_1, \beta_2, \dots, \beta_k$, and provide a solution, say $\hat{\beta}$, which is unique because $r(X'X) = k \Rightarrow X'X$ - non-singular.

$$\hat{\beta} = (X'X)^{-1}X'Y = AY, \Rightarrow \hat{\beta} \text{ is linear in } Y$$

Oct 9th

The solution is $\hat{\beta} = (X'X)^{-1}X'y$, which is the est of β , i.e. writing $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$, $\hat{\beta}_i$ is the LS est of β_i , $i=1, \dots, k$

if A is not square, then A^{-1} is not unique, $AA^{-1}A = A$
 $r(A^{-1}) \geq r(A)$

Define $\hat{u} = y - x\beta = y - \hat{y}$ as the LS residual vector

finally, the LS est of σ^2 is $S^2 = \frac{1}{n-k} \hat{u}'\hat{u} = \frac{1}{n-k} RSS$

$$(\sum u_i^2 = \hat{u}'\hat{u}, \text{ also } \hat{u}_i = y_i - \hat{y}_i = y_i - x_i\beta,$$
 (β, s, u) (β, σ^2, μ)

$$N^T X^T (X'X)^{-1} X' (X'X)^{-1} X' = (N^T + g^T x) X^T (X'X)^{-1} X' = N^T X^T (X'X)^{-1} X' = \hat{g}^T - A \hat{g} \quad (19)$$

$$E((A-K)^T x) = \frac{1}{n} \text{tr}(K^T A X^T X + I_n) = \sigma^2 \text{tr}(N)$$

variation $(y_i - \bar{y})^2$
variance $\frac{1}{n}(y_i - \bar{y})^2$

Some Algebraic Results.

Note first that, from the normal eqⁿ, we have $X'X\beta = X'y$

i.e. $X'(y - X\beta) = 0$

i.e. $X'u = 0$

i.e. column of X are orthogonal to $\hat{u}$. In particular,

$1'u = 0$, when $1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is the first column of X , or $\sum \hat{u}_i = 0$

Decomposition of TSS = $\sum_{i=1}^n (y_i - \bar{y})^2$ fix number

$$= \sum (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$$

$$= \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 + 2 \sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

$$= \sum \hat{u}_i^2 + \sum (\hat{y}_i - \bar{y})^2$$

R^2 = multiple corr coefficient

$$= (R^2)TSS + R^2TSS$$

The explanatory (regression) variables explain 60% of variation in y

ANOVA Table

Source	d.f.	SS	MSS
Regression	$k-1$	R^2TSS	$R^2TSS/k-1$
Residual	$n-k$	$(1-R^2)TSS$	$(1-R^2)TSS/n-k$
Total	$n-1$	TSS	-

$$F = \frac{a}{b} \rightarrow \text{goodness of fit test}$$

Sampling Properties.

As regards $u_1, u_2, \dots, u_n$, the only assumption we currently make is that u_i 's are iid, $E(u|x) = 0$ orthogonal condition.

(i) $\hat{\beta}$:
$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'u$$

$$= \beta + (X'X)^{-1}X'u \text{ sampling err}$$

given x , constant matrix

$$E(\hat{\beta}|x) = E(\beta|x) + E((x'x)^{-1}x'u|x) \\ = \beta + (x'x)^{-1}x'E(u|x) = 0 \quad (\text{orthogonal condition})$$

i.e. $\hat{\beta}$ is unbiased for β

i.e. $\hat{\beta}_i$ is β_i , $i=1, \dots, k$

$$(2) \hat{u} = y - x\hat{\beta} = y - x(x'x)^{-1}x'y = (I_n - x(x'x)^{-1}x')y = Ny$$

$$= N(x\beta + u) = Nx\beta + Nu$$

$$Nx = x - x(x'x)^{-1}x'x = 0$$

$$MN = M - M^2 = 0$$

$$NM = M - M^2 = 0$$

Oct 14th

M, N

(i) M & N are symmetric

(ii) M & N are idempotent ($M^2=M, N^2=N$, use $Mx=x$)

(iii) $MN=NM=0$ (use $Mx=x$) orthogonal

(iv) $V(M)$ = sum of rows of M, which are 1 's
similarly for N

$$\text{so } \hat{u} = N(x\beta + u) = Nx\beta + Nu = Nu$$

$$E(\hat{u}|x) = NE(u|x) = 0$$

$\hat{u}$ is unbiased for u , in the specific sense that $E[\hat{u}-u|x]=0$

$$E(s^2|x) = \sigma^2 \quad s^2 = \frac{RSS}{n-k} = \frac{\sum \hat{u}_i^2}{n-k} = \frac{\hat{u}'\hat{u}}{n-k}$$

$$\therefore (n-k)s^2 = \hat{u}'\hat{u} = (Nu)'(Nu) = u'N'Nu = u'NNU = duu \\ = \text{tr } u'Nu = \text{tr } Nu u' \quad (\text{tr } AB = \text{tr } BA)$$

$$E((n-k)s^2|x) = \text{tr } NE(uu'|x) = \text{tr } N\sigma^2 I_n = \sigma^2 \text{tr } N$$

$$\begin{aligned} \text{tr } N &= \text{tr} [I_n - X(X'X)^{-1}X'] = \text{tr } I_n - \text{tr} \left[\frac{X}{A} \frac{(X'X)^{-1}}{B} \frac{X'}{C} \right] \\ &= n - \text{tr} [(X'X)^{-1}X'X] = n - \text{tr } I_K = n - K \\ &\quad \text{tr}(AB) = \text{tr}(BA) \end{aligned}$$

$$\therefore F(S^2|X) = \sigma^2$$

Variance, Covariance:

$$\begin{aligned} V(\hat{\beta}|X) &= E[\hat{\beta} - E(\hat{\beta}|X)] [\hat{\beta} - E(\hat{\beta}|X)]' \\ &= E \left[(X'X)^{-1}X'u \cdot u'X(X'X)^{-1} \middle| X \right] \\ &\quad \underset{E[u \cdot u'|X]}{=} \\ &= (X'X)^{-1}X' \sigma^2 I_n X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned}$$

Oct 20th

$$V(\hat{\beta}|X) = \sigma^2 (X'X)^{-1} = \sigma^2 B, \text{ say } B = MN \neq NM \quad (i)$$

$$V(\hat{\beta}_i|x) = \sigma^2 B_{ii} \quad (ii)$$

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \text{cov}(\hat{\beta}_j, \hat{\beta}_i) = \sigma^2 B_{ij} = \sigma^2 B_{ji} \quad (iii)$$

$$V(\hat{u}|X) = V(Nu|X) = E[(Nu)(Nu)'|X]$$

$$\underset{N \times N}{E} \rightarrow \text{mean } E(\hat{u}) = \mu_{\hat{u}}$$

$$V(\hat{u}) = E[(\hat{u} - \mu_{\hat{u}})(\hat{u} - \mu_{\hat{u}})']$$

$$\therefore V(\hat{u}|X) = N E(uu'|X) N' = N V(u|X) N = N \sigma^2 I_n N = \sigma^2 N^2 =$$

$$\therefore V(\hat{u}_i|x) = \sigma^2 N_{ii} \neq \sigma^2 = V(\hat{u}_i|x)$$

$$\text{cov}(\hat{u}_i, \hat{u}_j|x) = \sigma^2 N_{ij} \neq 0 = \text{cov}(\hat{u}_i, \hat{u}_j|x)$$

$$\text{cov}(\hat{\beta}, \hat{u}|X) = 0 \quad \hat{\beta}_i \text{ is uncorr with each } \hat{u}_s, s=1, \dots, n$$

$$\text{cov}(\hat{\beta}, \hat{u}|X) = E[(\hat{\beta} - \beta)(\hat{u} - 0)|X] = 0$$

$$E[(X'X)^{-1}X'u|X] = (X'X)^{-1}X'E(u|X) = 0$$

$$\text{cov}(\hat{\beta}, \hat{u}) = 0$$

GM Thm

Version 1: $\hat{\beta}$ is BLUE for β (A)

Corollary to 1: $\hat{\beta}_i$ is BLUE for β_i , $i=1, \dots, k$ (B)

Note: A implies B, but B generally does not imply A

Version 2: Let $\theta = c_1\beta_1 + c_2\beta_2 + \dots + c_k\beta_k = C'\beta$ (where $C' = (c_1, c_2, \dots, c_k)$) for any arbitrary constant vector C ($\neq 0$), Then $\hat{\theta} = C'\hat{\beta} + C'\hat{u}$. $C'\hat{\beta}$ is BLUE for θ .

Prove: version 1.

Let $\tilde{\beta}$ be any other estimator of β which, like $\hat{\beta}$, (1) linear in $y_1, \dots, y_n$, (2) unbiased for β .

Since $\hat{\beta}$ is already linear in y , condition (1) on $\tilde{\beta}$ is readily met by taking $\tilde{\beta}$ as $\tilde{\beta} = \hat{\beta} + AY$, where A is a constant matrix. Then A needs to meet certain restrictions

if condition (2) is to be met. Since $E(\tilde{\beta}|X) = E(\hat{\beta}|X) + E(AY|X)$ must be β , i.e. $\beta + E(AY|X) = \beta$.

it follows that $E(AY|X) = 0$, i.e. $E[A(X\beta + u)|X] = AX\beta = 0$ for every β , which gives the required restriction on A as $AX = 0$ all β , i.e. $AX = 0$ (The unbiasedness restriction)

Oct 23th

GM Thm

Version 1

$$\tilde{\beta} = \hat{\beta} + Ay = \beta + Au \quad (AX=0) = \beta + (X'X)^{-1}X'u + Au$$

$$\begin{aligned} V(\tilde{\beta}|X) &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'|X] = E[(X'X)^{-1}X'u + Au][(X'X)^{-1}X'u + Au]'|X] \\ &= E[(X'X)^{-1}X'u u' X (X'X)^{-1} + (X'X)^{-1}X'u u' A' + A u u' X (X'X)^{-1} + A u u' A'|X] \end{aligned}$$

$$= (X'X)^{-1}X'E(uu'|X)X(X'X)^{-1} + (X'X)^{-1}X'E(uu'|X)A' + A E(uu'|X)X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1}X'A' + \sigma^2AX(X'X)^{-1} + \sigma^2AA'$$

$$= \sigma^2(X'X)^{-1} + \sigma^2AA' \quad (=0 \text{ if } AX=0)$$

$$\therefore V(\tilde{\beta}|X) = V(\hat{\beta}|X) + \sigma^2AA'$$

Def of efficiency

Let $\hat{\alpha}_1, \hat{\alpha}_2$ be two linear unbiased estimator of α

Then $\hat{\alpha}_1 \geq \hat{\alpha}_2$ (in efficiency sense) iff $V(\hat{\alpha}_2) - V(\hat{\alpha}_1)$ is at least

$V(\tilde{\beta}|X) - V(\hat{\beta}|X) = \sigma^2AA'$, which is at least PSD for any $A \neq 0$.

$$\therefore \tilde{\beta} \geq \hat{\beta}$$

Comment #1

$$V(\tilde{\beta}|X)_{ii} - V(\hat{\beta}|X)_{ii} \geq 0$$

$$\therefore \tilde{\beta}_i \geq \hat{\beta}_i, \text{ each } i=1, \dots, K$$

Q&A $\Rightarrow \tilde{\beta}_i$ is BLUE for $\beta_i, i=1, \dots, K$. (H) can we say $\tilde{\beta}$ is BLUE for

No, because (H) does not take into account the covariance of β_i 's

$$\sigma^2 (X'X)^{-1} X' A' = 0$$

$$\text{cov}(\hat{\beta}, \underline{Au}) = 0$$

$$\tilde{\beta} = \hat{\beta} + \underline{Au}$$

$$\text{cov} = 0.$$

$\tilde{\beta} - \hat{\beta}$ is uncorrelated with $\hat{\beta}$, the efficient est

version II. $\hat{\theta} = c'\hat{\beta}$ $\theta = c'\beta$, $c' = (c_1, c_2, \dots, c_n) \neq 0$, c 's are constants, c arbitrary

$\hat{\theta}$ is BLUE for θ

Corollary: Take $c' = (0, \dots, 1, 0, \dots, 0)$. Then $\hat{\beta}_i$ is BLUE for β_i , $i=1, \dots, n$

Proof. Let $a'y = a_1 y_1 + a_2 y_2 + \dots + a_n y_n$ be an unbiased estimator of θ ; note that $a'y$ is, like $\hat{\theta}$, linear in y . The unbiasedness of $a'y$ for θ requires:

$$E(a'y|x) = \theta = c'\beta$$

$$\text{ie. } a'E(X\beta|x) = a'X\beta = c'\beta$$

$$\text{ie. } a'X = c' \quad (\text{unbiased condition}) \quad (1)$$

Also, under (1) $a'y$ becomes $a'X\beta + a'u = c'\beta + a'u = \theta + a'u$

$$V(a'y|x) = E[(a'y - \theta)^2|x] = E[(a'u)^2|x] = E[a'u u' a|x]$$

$$= a'E(u u'|x)a = \sigma^2 a'a$$

compare $V(a'y|x)$ with $V(\hat{\theta}|x)$

$$\downarrow$$

$$\sigma^2 a'a \quad E[(\hat{\theta} - \theta)^2|x] = E\{[c'(\hat{\beta} - \beta)]^2|x\}$$

$$= E\{c'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'c|x\}$$

$$= c'E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']c = \sigma^2 c'(X'X)^{-1}c$$

becomes $\sigma^2 a'a$ vs $\sigma^2 c'(X'X)^{-1}c = a'X'(X'X)^{-1}a$

$a'a$ vs $a'X(X'X)^{-1}X'a$

difference: $\hat{\sigma}^2 = a'[I_n - X(X'X)^{-1}X']a$

Idemp because $N^2 = N$, N is p.s.D

$\hat{\sigma}^2$ is BLUE for σ^2 .

i.e. $c'\hat{\beta}$ is BLUE for $c'\beta$ for any $c \neq 0$.

$\therefore$ "dropping c "

$\hat{\beta}$ is BLUE for β — (Version 1)

Explicitly estimating subvectors of β

Given: $y = X\beta + u$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \begin{matrix} k_1 \\ -k_2 \end{matrix} > K \Rightarrow y = X_1\beta_1 + X_2\beta_2 + u$$

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{matrix} n \times k_1 & n \times k_2 \end{matrix}$$

The normal eq^{ns} are $X'X\hat{\beta} = X'y$

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

$$X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 = X_1'y \quad (1)$$

$$X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 = X_2'y \quad (2)$$

solve (2) for $\hat{\beta}_2$, substitute it in (1), then we can

solve for $\hat{\beta}_1$, $\hat{\beta}_1 = (X_1'N_2X_1)^{-1}X_1'N_2y$

$$\hat{\beta}_2 = (X_2'N_2X_2)^{-1}X_2'N_2y$$

where $N_i = I_n - X_i(X_i'X_i)^{-1}X_i'$, $i=1,2$

$$N = I_n - X(X'X)^{-1}X'$$

N_1, N_2 idemp p.s.D, symmetric

$$M_i = X_i (X_i' X_i)^{-1} X_i'$$

M_1 & $M_2 \rightarrow$ idemp, p.s.d, symmetric

$$M_i N_i = N_i M_i = 0$$

$$M_i X_i = X_i, N_i X_i = 0$$

$$\text{check: } M M_1 = M_1 M, N N_1 = N_1 N$$

Oct 28th

$$V(\hat{\beta}_1) = \sigma^2 (X'X)^{-1} = \sigma^2 \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} = \sigma^2 \begin{bmatrix} (X_1'N_2X_1)^{-1} & \text{covariance} \\ \text{covariance} & (X_2'N_1X_2)^{-1} \end{bmatrix}$$

$$\therefore V(\hat{\beta}_1|X) = \sigma^2 (X_1'N_2X_1)^{-1}$$

$$V(\hat{\beta}_2|X) = \sigma^2 (X_2'N_1X_2)^{-1}$$

"sweepout" method

"parse out" method

focus on β_1 , $y = X_1\beta_1 + X_2\beta_2 + u \quad (1)$

choose $P = N_2$

$$N_2 y = N_2 X_1 \beta_1 + N_2 X_2 \beta_2 + N_2 u \quad (2)$$

Apply LS to (2)

$$\hat{\beta}_1 = (X_1'N_2N_2X_1)^{-1} X_1'N_2N_2y = (X_1'N_2X_1)^{-1} X_1'N_2y = \hat{\beta}_1 \text{ from before}$$

$$V(\hat{\beta}_1) = \hat{\beta}_1 = \beta_1 + (X_1'N_2X_1)^{-1} X_1'N_2u$$

$$V(\hat{\beta}_1|X) = E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1)' | X] = E[(X_1'N_2X_1)^{-1} X_1'N_2u u' N_2X_1 (X_1'N_2X_1)^{-1} | X]$$

$$= (X_1'N_2X_1)^{-1} X_1'N_2 E[uu'|X] N_2X_1 (X_1'N_2X_1)^{-1}$$

$$= \sigma^2 (X_1'N_2X_1)^{-1} X_1'N_2X_1 (X_1'N_2X_1)^{-1} = \sigma^2 (X_1'N_2X_1)^{-1}$$

$$y = a + bx_1 + cx_2 + u$$

Regress x_2 on x_1 , residual $\rightarrow s$'s purified x_2

— y on x_1 , residual $\rightarrow y$'s purified y

Stepwise Least Squares:

$$y = x_1 \beta_1 + x_2 \beta_2 + u \rightarrow \hat{\beta}_2 = (x_2' x_2)^{-1} x_2' (y - x_1 \hat{\beta}_1)$$

known

step 1, regress y on x_1 alone $\rightarrow \hat{\beta}_1 = (x_1' x_1)^{-1} x_1' y = \hat{\beta}_1$?

constant $y - x_1 \hat{\beta}_1 = u$ $(x_2' x_2)^{-1} x_2' (y - x_1 \hat{\beta}_1) = (x_2' x_2)^{-1} x_2' (y - x_1 (x_1' x_1)^{-1} x_1' y)$

step 2, regress u on $x_2 \rightarrow \hat{\beta}_2 = (x_2' x_2)^{-1} x_2' N_1 y = \hat{\beta}_2$? $= (x_2' x_2)^{-1} x_2' N_1 y$

both $\hat{\beta}_1 \neq \beta_1$ $\hat{\beta}_2 \neq \beta_2$ it does not work when doing stepwise

if $\hat{\beta}_1 = \beta_1$, we need $(x_1' x_1)^{-1} x_1' y = (x_1' N_2 x_2)^{-1} x_1' N_2 y \Rightarrow x_1' = x_1' N_2$

$\Rightarrow x_1' = x_1' [I - x_2 (x_2' x_2)^{-1} x_2'] \Rightarrow x_1' = x_1' - x_1' x_2 (x_2' x_2)^{-1} x_2' \Rightarrow$

$= 0$ required $= 0$ required

$\Rightarrow x_1' x_2 (x_2' x_2)^{-1} x_2' x_2 = 0$ (post-multiply by x_2) $\Rightarrow x_1' x_2 = 0$

i.e. x_1 is orthogonal to x_2

$\therefore$ under $x_1' x_2 = 0$, it is true that $\hat{\beta}_1 = \beta_1$

But when x_1, x_2 are orthogonal, stepwise LS works $\hat{\beta}_1 = \beta_1, \hat{\beta}_2 = \beta_2$
(if first-step is correct, then second step will automatically be correct)

$y = x_1 \beta_1 + x_2 \beta_2 + \text{error}$

regress: $(x_2' x_2)^{-1} x_2' (y - x_1 \hat{\beta}_1) = \hat{\beta}_2$?

Yes

The "second" block of normal eq^{ns} is $X_2'X_1\hat{\beta}_1 + (X_2'X_2)\hat{\beta}_2 = X_2'y$
 (answer for the above question) $\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'(y - X_1\hat{\beta}_1)$

$$\hat{u} = y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2$$

specification error analysis

error of omission

True: $y = X_1\beta_1 + \beta_k X_k + u$

where $X = (X_1' ; X_k')$; $\beta = \begin{pmatrix} \beta_1 \\ \beta_k \end{pmatrix}$

False: $y = X_1\beta_1 + u$

$$LS: b_1 = (X_1'X_1)^{-1}X_1'y$$

Issue: knowing b_1 , what can be said about β_1 ?

(is b_1 unbiased about β_1 ? unless X_k is orthogonal to X_1)

$$E(b_1|X) = (X_1'X_1)^{-1}X_1'E(X_1\beta_1 + \beta_k X_k + u|X)$$

$$= (X_1'X_1)^{-1}X_1'X_1\beta_1 + \beta_k(X_1'X_1)^{-1}X_1'X_k + 0$$

$$= \beta_1 + \beta_k(X_1'X_1)^{-1}X_1'X_k = \beta_1 + \beta_k \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{pmatrix}$$

where $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{k-1} \end{pmatrix}$ are the LS estimates of coefficients in the regression of

X_k on X_1 . $X_k = X_1(\) + \text{Error}$

⌘ auxiliary regression of omitted variables on all included variables

so b_1 is biased for β_1 , $E(b_1) = \beta_1 + \underbrace{\beta_k}_{\neq 0} a_i$

Oct 30th

Error of Commission

True $y = X\beta + u$

False $y = X\theta + Z\gamma + u$
 $\begin{matrix} N \times K & N \times 1 \end{matrix}$

We work with false. Apply LS to the false by the sweepout method

$$\hat{\theta} = (X'N_2X)^{-1}X'N_2Y$$

$$N_2 = I_n - Z(Z'Z)^{-1}Z'$$

Knowing $\hat{\theta}$, what can be said about β ?

$$E(\hat{\theta}|X, Z) = (X'N_2X)^{-1}X'N_2(X\beta + u) = \beta + (X'N_2X)^{-1}X'N_2u$$

$$\therefore E(\hat{\theta}|X, Z) = \beta + 0 = \beta$$

$\hat{\theta}$ is unbiased for β , but $\hat{\theta}$ is not as efficient as $\hat{\beta}$ (from True)

$$\hat{\beta} \succ \hat{\theta} \text{ iff}$$

$V(\hat{\theta}|X) - V(\hat{\beta}|X)$ is at least P.S.D

$$\sigma^2(X'N_2X)^{-1} - \sigma^2(X'X)^{-1}$$

ie. $[(X'X)^{-1}]^{-1} - [(X'N_2X)^{-1}]^{-1}$ is at least P.S.D

Suppose A & B are p.d.

Such that $A - B$ is at least P.S.D

Then $B^{-1} - A^{-1}$ is at least P.S.D

ie. $X'X - X'N_2X$ is at least P.S.D

$$X'(I_n - N_2)X \text{ is } -$$

which is true since $I_n - N_2$ is idempotent hence P.S.D

If H is p.d / P.S.D

Then $G'HG$ is at least P.S.D ($G \neq 0$)

Note: $\hat{\theta} = \beta$ if $X'Z = 0$

Comments:

1) Forecast based on $\hat{\theta}$ will not be as efficient as those based on β

2) The false estimate γ by $\hat{\gamma}$, given by $\hat{\gamma} = (Z'NZ)^{-1} Z'NY$, with $V(\hat{\gamma}|X, Z) = \sigma^2 (Z'NZ)^{-1} \neq 0$

That is the false estimates γ known to be zero by a random variable $\hat{\gamma}$ with non-zero variance. Clearly, $\hat{\gamma}$ is an inefficient est of $\gamma = 0$, the efficient est being 0. The false ignore the prior info that γ is 0.

3) When $X'Z \neq 0$, $\text{cov}(\hat{\theta}, \hat{\gamma}) \neq 0$, the inefficiency implicit in the use of $\hat{\gamma}$ carries over to $\hat{\theta}$, causing it also to be inefficient.

Result 1: When a regression is added to the true model, the resulting RSS is either reduced or unchanged.

$$RSS_T = Y'NY$$

$$RSS_F = Y'SY$$

$$\text{where } S = I_n - W(W'W)^{-1}W'$$

$$W = [X \ Z]$$

$$RSS_T - RSS_F = Y'(N-S)Y \geq 0? \text{ Yes, since } N-S \text{ is idemp}$$

$$(N-S)(N-S) = N + S - NS - SN = N + S - S - S = N - S \text{ idemp}$$

$$NS = (I_n - X(X'X)^{-1}X')(I_n - W(W'W)^{-1}W') = I - W(W'W)^{-1}W' - X(X'X)^{-1}X' + X(X'X)^{-1}X'W(W'W)^{-1}W' = S$$

Result 2: Let R_T^2 , R_F^2 be the R^2 's for the true & the false model. Then $R_F^2 > R_T^2$

$$RSS_T = y'N\hat{y} = TSS(1-R^2)$$

$$RSS_F = y'Sy = TSS(1-R_F^2)$$

$$\therefore R_F^2 > R^2, \text{ since } RSS_F \leq RSS_T$$

From result 1 & 2, it is evident that one can increase R^2 by adding irrelevant regressors.

Inference:

Testing of Hypothesis

Confidence Regions (or CI's)

for more than coeffs

$$H_0: \beta_i = \beta_i^0 \text{ (specified)}$$

$$H_1: \beta_i \neq \beta_i^0$$

H_0 is not true.

Normality Assumptions: u 's are iid $(0, \sigma^2)$, $u \sim N(0, \sigma^2 I_n)$

$$(i) y \sim N(X\beta, \sigma^2 I_n)$$

Any linear combination distribution y is also normal

Since $\hat{\beta}$ is linear in y , it follows that $\hat{\beta}$ is also normal, with

mean β , and $V(\hat{\beta}) = \sigma^2(X'X)^{-1}(X) = W$

$$\therefore \hat{\beta} \sim N[\beta, \sigma^2(X'N_2X)^{-1}]$$

$$\text{Under } H_0, \hat{\beta}_i \sim (\beta_i^0, \sigma^2(X'N_2X)^{-1})$$

THM 1: If $\sum_{p=1}^p \sim N(\mu_{p1}, \sigma^2 S_{ss})$, then -

$$(S - \mu_s)' (\sigma^2 S_{ss})^{-1} (S - \mu_s) \sim \chi_p^2 \text{ p.d.f} = \chi(S_{ss})$$

$$\frac{(S - \mu_s)' (S_{ss})^{-1} (S - \mu_s)}{\sigma^2} \sim \chi_p^2$$

∴ by THM 1,

$$\frac{(\beta_1 - \beta_0)' (X_1' N_1 X_1)^{-1} (\beta_1 - \beta_0)}{\sigma^2} \sim \chi_{k_1}^2 \quad \text{d.f.} = k_1 = r(X_1' N_1 X_1)$$

Nov 4th

Substitute σ^2 by LS est S_f^2

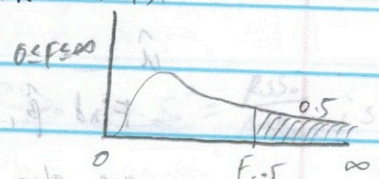
We have
$$\frac{(\beta_1 - \beta_0)' (X_1' N_1 X_1)^{-1} (\beta_1 - \beta_0)}{S_f^2} = \frac{\hat{u}' \hat{u}}{n-k}$$

especially we consider

$$F = \frac{(\beta_1 - \beta_0)' (X_1' N_1 X_1)^{-1} (\beta_1 - \beta_0) / k_1}{\hat{u}' \hat{u} / n-k}$$
 which has the F-distribution

with $\left. \begin{array}{l} \text{d.f. in the } N_1 = k_1 \\ \text{d.f. in the } D_1 = n-k \end{array} \right\} F \sim F_{k_1, n-k} \quad P(F)$

reject H_0 if $F \geq F_{\alpha}$, accept



$$F.p.g. = \frac{X_1^2 / p}{X_2^2 / q}$$

when $X_1^2 \sim \chi^2$ with d.f. = r

and X_1^2, X_2^2 are indep

write $F = \frac{(\hat{\beta}_1 - \beta_0)' (X_1' N_1 X_1)^{-1} (\hat{\beta}_1 - \beta_0) / k_1}{\hat{u}' \hat{u} / n-k} = \frac{Q_1 / k_1}{Q_2 / n-k} = \frac{Q_1 / k_1}{Q_2 / n-k}$

$\hat{u}' \hat{u} = \hat{u}' N_1 \hat{u} \Rightarrow \hat{u}' \hat{u} / n-k = \frac{Q_2}{n-k}$

(i) we always know that $\frac{Q_1}{k_1} \sim \chi_{k_1}^2$

(ii) To prove that $\frac{Q_2}{n-k} \sim \chi^2$, we use

THM 2: Let $\sum_{p \times 1} \epsilon_i \sim N(0, \sigma^2 I_p)$ ϵ_i 's are standardized normal dist, which means indep.

5. $\Sigma' A \Sigma \sim \chi^2$ iff A is idemp with d.f. $= r(A) < p$

corollary: $u \sim N(0, \sigma^2 I_n)$, Then $(\frac{u}{\sigma})' \sim N(0, I_n)$

Then $(\frac{u}{\sigma})' A (\frac{u}{\sigma}) \sim \chi^2$ with d.f. $= r(A)$

i.e. $\frac{u' A u}{\sigma^2} \sim \chi^2$ with d.f. $= r(A)$

Using THM / corollary

we find that

$\frac{Q_1}{\sigma^2} = \frac{u' N u}{\sigma^2} \sim \chi^2$ because N is idemp, also d.f. $= r(N) = n - k$

(iii) To prove idemp of Q_1 & Q_2 ,

a) $Q_1 = f(\hat{\beta})$

$Q_2 = f(\hat{u})$

$\hat{\beta}$ is linear in u , hence a normal distribution variable.

$\hat{u} = \dots$

Find $\hat{\beta}, \hat{u}$ are uncorrelated, and hence under normality,

are also indemp / complete the proof

b) Q_1 is $f(X' N u)$

Q_2 is $f(N u)$

THM: Let $X_1 = \Sigma' A \Sigma \sim \chi^2$, $X_2 = \Sigma' B \Sigma \sim \chi^2$. Then X_1, X_2 are IND $\iff AB = 0$

$X_1 = \Sigma' A \Sigma = \Sigma' A A \Sigma = \Sigma' A' A \Sigma = J' J$ say

$\therefore X_1 = f(J)$

$X_2 = \Sigma' B \Sigma = \Sigma' B B \Sigma = \Sigma' B' B \Sigma = e' e$ say

$\therefore X_2 = f(e)$

show J & e are IND $\text{cov}(J, e) = E(J e') = E(\Sigma \Sigma' B') =$

$= A E(\Sigma \Sigma' B') = AB' = AB$

which must be 0 if X_1, X_2 are to be IND

$$\text{cov}(X_1'N, u, Nu) = E(X_1'N, uu'N) = X_1'N \cdot E(uu'|X)N$$

$$= X_1'N \cdot \sigma^2 I_N = X_1'N + (N_2'N - NN_2) = 0$$

$$= X_1'X + (X_2'X - XX_2) = 0 \quad (\text{since } XN=0, NX \Rightarrow XX_1=0)$$

$H_0: \text{slope} = 0$
 $H_1: \text{Not } H_0$ } Goodness of Fit-test

Wald - ~~the~~ Method

Likelihood Ratio Related Test

$$H_0: \beta_1 = \beta_1^0$$

$$H_1: \text{Not } H_0$$

$$y = X_1\beta_1 + X_2\beta_2 + u \quad u \sim N(0, \sigma^2 I_n)$$

$$\text{under } H_0: y = X_1\beta_1^0 + X_2\beta_2 + u$$

i.e. $y = X_1\beta_1^0 + X_2\beta_2 + u \rightarrow \text{get } RSS_0 \rightarrow S_0^2 = \frac{RSS_0}{n-k_2}$ is the restricted estⁿ of σ^2 .

$RSS_0 \Rightarrow$ restricted RSS restricted by H_0 .

under H_1 , the unrestricted model is as is - RSS - unrestricted residual

sum of square $\Rightarrow S^2 = \frac{RSS}{n-k}$ is the unrestricted estⁿ of σ^2

$$\frac{S_0^2}{S^2} \approx 1$$

$$F = \frac{(RSS_0 - RSS)/f_1}{RSS/f_2} \sim F_{f_1, f_2}$$

$f_1: \#$ of restriction on β imposed by $H_0 = k$,

$$f_2: n-k$$

$$RSS_0 - RSS = u'N_2u - u'Nu = u' \underbrace{(N_2 - N)}_{\text{indep}} u = u'Nu = (1 - R^2)TSS$$

$$\text{tr}(N_2 - N) = \text{tr}(N_2) - \text{tr}(N) = n - k_2 - (n - k) = k,$$

$$Q = U' [N_2 (X_1' N_2 X_1)^{-1} X_1' N_2] U$$

$$M = M_1 + N_1 X_2 (X_2' N_1 X_2)^{-1} X_2' N_1$$

$$M - M_2 = N_2 X_1 (X_1' N_2 X_1)^{-1} X_1' N_2$$

$$M - M_2 = N_2 - N$$

Nov 6th

$$\beta \sim N(\beta, \sigma^2 (X'X)^{-1} = \sigma^2 A, \text{ say})$$

$$H_0: \beta_j = \beta_j^0$$

$$H_1: \beta_j \neq \beta_j^0$$

$$t = \frac{\hat{\beta}_j - \beta_j^0}{SE_j} \quad SE_j = \sqrt{\text{var}(\hat{\beta}_j)} = \sqrt{\sigma^2 a_{jj}} = \sqrt{\frac{\hat{\sigma}^2}{n-k} a_{jj}}$$

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 a_{jj})$$

$$Z_j = \frac{\hat{\beta}_j - \beta_j^0}{\sigma \sqrt{a_{jj}}} \sim N(0, 1)$$

$\therefore$ under H_0

$$Z_j = \frac{\hat{\beta}_j - \beta_j^0}{\sigma \sqrt{a_{jj}}} \sim N(0, 1)$$

$$SE_j \cdot t = \frac{\hat{\beta}_j - \beta_j^0}{SE_j} = \frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\frac{\hat{\sigma}^2}{n-k} a_{jj}}} = \frac{(\hat{\beta}_j - \beta_j^0) / \sigma \sqrt{a_{jj}}}{\sqrt{\hat{\sigma}^2 / (n-k)}} = \frac{Z_j}{\sqrt{X_{n-k}^2 / (n-k)}}$$

$$\text{Def } t = \frac{Z}{\sqrt{X_q^2/q}} \sim t_q \text{ where } Z \sim N(0, 1) \text{ and indep of } X_q^2$$

95% CR for β_1 : $\beta_{K \times 1} = \begin{pmatrix} \beta_1 \\ K_{1 \times 1} \\ \beta_2 \\ K_{2 \times 1} \end{pmatrix}$

$H_0: \beta_1 = \beta_1^0$

$H_1: \beta_1 \neq \beta_1^0$

$F = \frac{(\beta_1 - \beta_1^0)' (X_1' N_2 X_1) (\beta_1 - \beta_1^0) / K_1}{S^2}$

Accept H_0 if $F \leq F_{0.5}$

i.e. if $(\beta_1 - \beta_1^0)' (X_1' N_2 X_1) (\beta_1 - \beta_1^0) \leq K_1 S^2 F_{0.5}$

$\leq f_0$, say

the 95% CR is given by $(\hat{\beta}_1 - \beta_1^0)' X_1' N_2 X_1 (\hat{\beta}_1 - \beta_1^0) \leq f_0$ when the LHS is tested as a function of β_1 , for given $\hat{\beta}_1$

$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$
 β_1 $\rightarrow$ scalar intercept
 β_2 $\rightarrow$ slopes $K-1$

$X = \begin{pmatrix} L_n & X_2 \\ \frac{1}{N} & \end{pmatrix}$

$y = X\beta + u$

$= L_n \beta_1 + X_2 \beta_2 + u$

$H_0: \beta_2 = 0 \rightarrow$ slopes are zero

$H_1: \text{not } H_0$

Goodness of Fit test

proceed only if H_0 is rejected

Under H_0 : The restricted model is

$y = \beta_1 L_n + u \rightarrow$ restricted residuals are: $u_0 = y - \hat{\beta}_1 L_n = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}$

$RSS_0 = u_0' u_0 = \sum (y_i - \bar{y})^2 = TSS$

Under H_1 , the unrestricted RSS is

$RSS = \hat{u}' \hat{u} = u' N u = (1 - R^2) TSS$

$F = \frac{(RSS_0 - RSS) / (K-1)}{RSS / (n-k)} \sim F_{K-1, n-k}$

$$F = \frac{(TSS - (1-R^2)TSS)/k-1}{(1-R^2)TSS/n-k} = \frac{R^2/k-1}{(1-R^2)/n-k}$$

ANOVA Table

Regression

MSS

$$RSS/k-1 = R^2 TSS/k-1 \quad (A)$$

Residual

$$RSS/n-k = (1-R^2)TSS/n-k \quad (B)$$

$$F = \frac{A}{B}$$

Forecasting

$$y_1 = x_1' \beta + u_1$$

$$\vdots$$

$$y_n = x_n' \beta + u_n$$

β, σ^2, u

Forecasting period:

Given x_{n+1}

$$y_{n+1} = x_{n+1}' \beta + u_{n+1}$$

not known known

$$\text{est } y_{n+1} = \text{est } x_{n+1}' \beta + \text{est } u_{n+1}$$

$$= x_{n+1}' \hat{\beta} + 0$$

Forecast error

$$e_f = x_{n+1}' \hat{\beta} - (x_{n+1}' \beta + u_{n+1}) = x_{n+1}' (\hat{\beta} - \beta) - u_{n+1}$$

Note that $E(e_f |) = 0$

$$V(e_f |) = V[x_{n+1}' (\hat{\beta} - \beta)] + V(u_{n+1})$$

$$= E[x_{n+1}' (\hat{\beta} - \beta) (\hat{\beta} - \beta)' x_{n+1}] + \sigma^2$$

$$= E[x_{n+1}' (\hat{\beta} - \beta) (\hat{\beta} - \beta)' x_{n+1}] + \sigma^2$$

$$= x_{n+1}' V(\hat{\beta}) x_{n+1} + \sigma^2$$

$$= \sigma^2 x_{n+1}' (X'X)^{-1} x_{n+1} + \sigma^2 = \sigma_f^2, \text{ say}$$

$$\epsilon_t \sim N(0, \sigma_f^2)$$

$$Z = \frac{\epsilon_t}{\sigma_f} \sim N(0, 1)$$

$$\frac{\epsilon_t}{\sigma_f} \sim t_{n-k}$$

$$\frac{X'_{n+1} \hat{\beta} - y_{n+1}}{\sigma_f} \sim t_{n-k}$$

$$[T_1 = X'_{n+1} \hat{\beta} - t_{\alpha/2} \sigma_f \quad T_2 = X'_{n+1} \hat{\beta} + t_{\alpha/2} \sigma_f] \quad 95\% \text{ CI for } y_{n+1}$$

H_0 : No structural break

H_1 : Not H_0

Accept if $T_1 \leq y_{n+1} \leq T_2$

Reject otherwise

$\hat{\beta}_n = \beta$ for n obs

$\hat{\beta}_{n+1}$ for $n+1$ obs

$\hat{\beta}_{n+1} \geq \hat{\beta}_n$

Restricted LS

$y = X\beta + u$ — GM by itself

Prior restrictions on β :

$$r = R \cdot \beta \Rightarrow \begin{matrix} r_1 = R_1 \beta \\ r_2 = R_2 \beta \\ r_3 = R_3 \beta \end{matrix}$$

where r, R are known, and $r(R, R) = r(R) = q < k$

Apply LS to (1) Subject to (2)

constitute valid prior info that could be incorporated in estimation

$$\lg Q = \beta_1 + \beta_2 \lg K + \beta_3 \lg L + u$$

$\beta = \beta_2 + \beta_3 = \text{scale parameter}$

If impose restriction $1 = \beta_2 + \beta_3$

$$1 = (0, 1, 1) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

$$r = R'\beta$$

$$\min_{\beta, \lambda} f = u'u + 2\lambda'(r - R\beta)$$

$$\text{Nov 11th} = x_i\beta + u_i$$

$$y = X\beta + u \quad (1)$$

$$r = R\beta \quad (2)$$

$$\min_{\beta, \lambda} f = (y - X\beta)'(y - X\beta) + 2\lambda'(r - R\beta)$$

Forecasting $X R^2$

$$y_{n+1} = x'_{n+1}\beta + u_{n+1}, \text{ Forecast is } y_{n+1}^f = x'_{n+1}\hat{\beta}$$

$$y = X\beta + u \text{ — True}$$

$$y = X\hat{\beta} + Zr + u \text{ — False} \Rightarrow \text{Forecast is } \hat{y}_{n+1} = x'_{n+1}\hat{\beta} \text{ because } \hat{\beta} \geq \hat{\theta}$$

Efficient forecasting is better than a higher R^2

$$\frac{\partial f}{\partial \beta} = 0 \Rightarrow -2X'y + 2X'X\beta - 2R'\lambda = 0 \quad (3)$$

$$\frac{\partial f}{\partial \lambda} = 0 \Rightarrow r = R\beta \quad (4)$$

Let $\beta_{r, \hat{\lambda}}$ be the solution, The β_r is the restricted LS est of β

from 3) : $\beta_r = (X'X)^{-1}(X'y + R'\hat{\lambda})$

since $r = R\beta_r$ (from (4)), we have

$$R\beta_r = R(X'X)^{-1}X'y + R(X'X)^{-1}R'\hat{\lambda}$$

$$\therefore r = R\hat{\beta} + R(X'X)^{-1}R'\hat{\lambda}$$

$$\therefore R(X'X)^{-1}R'\hat{\lambda} = r - R\hat{\beta}$$

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

$$\beta_r = (X'X)^{-1}(X'y + R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta}))$$

$$\beta_r = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta}) \quad (6)$$

β_r improves on $\hat{\beta}$ by taking into account the discrepancy by which $\hat{\beta}$ fails to satisfy $r - R\hat{\beta} = 0$ restriction; indeed, $\beta_r = \hat{\beta}$ if the discrepancy is zero.

$$E(\beta_r | X, R) = E(\hat{\beta} | X, R) + A E(r - R\hat{\beta} | X, R)$$

$$= \beta + A \cdot 0 = \beta \quad \text{no bias}$$

$$\beta_r = \beta + (X'X)^{-1}X'u + A[r - R\hat{\beta} - R(X'X)^{-1}X'u]$$

$$\beta_r - \beta = [I - AR](X'X)^{-1}X'u = (I - AR)(\hat{\beta} - \beta)$$

$$V(\beta_r | X, R) = E[(\beta_r - \beta)(\beta_r - \beta)' | X, R]$$

$$= (I - AR)V(\hat{\beta})(I - AR)'$$

$$= (I - AR)V(\hat{\beta}) - (I - AR)V(\hat{\beta})R'A'$$

$$= V(\hat{\beta}) - ARV(\hat{\beta}) - (I - AR)V(\hat{\beta})R'A'$$

$$= V(\hat{\beta})R'A' - ARV(\hat{\beta})R'A'$$

$$= \sigma^2 X'X^{-1}R'A' - \sigma^2 X'X^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'A'$$

$$= 0$$

$$= V(\hat{\beta}) - X'X^{-1}R'[R(X'X)^{-1}R']^{-1}R \cdot \sigma^2 (X'X)^{-1}$$

$$V(\hat{\beta}_0) - V(\hat{\beta}_r) = \sigma^2 X'X^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}, \text{ which is a PSD,}$$

$$\text{so } \beta_r \geq \hat{\beta}$$

if $\Gamma - R\beta = \delta$ testing for $\Gamma - R\beta = 0$ i.e. null

$$H_0: \Gamma - R\beta = \delta = 0$$

$$H_1: \Gamma - R\beta = \delta \neq 0$$

(i) Wald Test: Use $\Sigma = \Gamma - R\hat{\beta} = \Gamma - R\beta - R(X'X)^{-1}X'u$

$$= 0 - R(X'X)^{-1}X'u$$

under H_0 : $E(\Sigma/X) = 0$

$$V(\Sigma/X) = R(X'X)^{-1}X'E(uu')X(X'X)^{-1}R' = \sigma^2 R(X'X)^{-1}R'$$

$$\text{By THM 1: } \frac{\Sigma' [R(X'X)^{-1}R']^{-1} \Sigma / q}{\sigma^2} \sim F_{q, n-k}$$

(ii) LM Test on $\hat{\lambda} = [R(X'X)^{-1}R']^{-1}(\Gamma - R\hat{\beta})$

$\hat{\lambda}$ is a non-singular transformation of Σ . Therefore the test based on $\hat{\lambda}$ is identical to the test based on Σ .

$$\hat{\lambda} = G\Sigma$$

$$E(\hat{\lambda}) = GE(\Sigma) = 0$$

$$V(\hat{\lambda}) = GV(\Sigma)G' = G\sigma^2 R(X'X)^{-1}R'G'$$

$$\frac{\hat{\lambda}' [G R(X'X)^{-1}R' G']^{-1} \hat{\lambda} / q}{\sigma^2} \sim F_{q, n-k}$$

Compatibility Test (i) (ii)

Nov 13th

Likelihood Ratio related test

$$F = \frac{\text{Restricted RSS} - \text{Unrestricted RSS} / q}{\text{Unrestricted RSS} / n - k} \sim F_{q, n-k}$$

$$\beta_r, \hat{\beta}$$

$$u_r = y - X\beta_r \rightarrow \text{Restricted RSS} = u_r' u_r$$

$$\hat{u} = y - X\hat{\beta} \rightarrow \text{Unrestricted RSS} = \hat{u}' \hat{u}$$

$$u_r = y - X\hat{\beta} + X\hat{\beta} - X\beta_r = \hat{u} + X(\hat{\beta} - \beta_r) = \hat{u} + X d_r$$

$$\text{Since } X\hat{u} = 0,$$

$$u_r' u_r = \hat{u}' \hat{u} + d_r' X' X d_r$$

$$F = \frac{d_r' X' X d_r / q}{\hat{u}' \hat{u} / n - k}$$

Testing of equality of coeff in two regression (Chow Test)

$$y_1 = X_1 \beta_1 + u_1 \rightarrow \text{GM by itself}, u_1 \sim N(0, \sigma_1^2 I_{n_1})$$

$$y_2 = X_2 \beta_2 + u_2 \rightarrow \text{GM by itself}, u_2 \sim N(0, \sigma_2^2 I_{n_2})$$

$$H_0: \beta_1 = \beta_2 = \beta, \text{ say } (K_1 = K_2 = K, \text{ say}) \quad H_1: \beta_1 \neq \beta_2$$

$$\text{Assumption: } n_1 > K, n_2 > K$$

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ (say)}$$

Note: Asymptotic tests can be developed ($n_1 \rightarrow \infty, n_2 \rightarrow \infty$)

even $\sigma_1^2 \neq \sigma_2^2$, in this scenario, finite sample tests are not available

(GM) The restricted model, restricted under H_0 is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_2 \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ \vdots \\ u_2 \end{pmatrix}$$

$$y = X_0 \beta + u, u \sim \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{2n} & 0 \\ 0 & \sigma^2_{2n} \end{pmatrix}$$

BLUE $\rightarrow \hat{\beta}_0 = (X_0' X_0)^{-1} X_0' y \rightarrow \hat{u}_0 = y - X_0 \hat{\beta}_0 \rightarrow RSS_0 = \hat{u}_0' \hat{u}_0$

The unrestricted model is

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \Rightarrow y = X \beta + u \Rightarrow \hat{u} = y - X \hat{\beta}$$

$$(y = x_1 \beta_1 + x_2 \beta_2 + u, x_1' x_2 = 0)$$

$$\therefore F = \frac{(\hat{u}_0' \hat{u}_0 - \hat{u}' \hat{u}) / f_1}{\hat{u}' \hat{u} / f_2} \sim F_{f_1, f_2}$$

$$\begin{aligned} f_1 &= k \\ f_2 &= n_1 + n_2 - 2k \\ df \hat{u}_0' \hat{u}_0 &= n_1 + n_2 - k \\ df \hat{u}' \hat{u} &= n_1 + n_2 - 2k \\ f_1 &= df \hat{u}_0' \hat{u}_0 - df \hat{u}' \hat{u} = k \end{aligned}$$

Pooling

$$X_0' X_0 \hat{\beta} = X_0' y$$

$$(x_1' x_1 + x_2' x_2) \hat{\beta} = (x_1' y_1 + x_2' y_2)$$

Two way of pooling

1) pool sample and apply LS like $y = X_0 \beta + u$ $X_0 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

2) Apply LS separately to the two models and then pool this

$\hat{\beta}_1, \hat{\beta}_2$ pool them, take it as $\hat{b}$, then $\hat{b} = \hat{\beta}$

$$x_1' x_1 \hat{\beta}_1 = x_1' y_1 \Rightarrow (x_1' x_1 + x_2' x_2) \hat{b} = x_1' y_1 + x_2' y_2$$

$$x_2' x_2 \hat{\beta}_2 = x_2' y_2$$

$$\hat{b} = \hat{\beta}$$

Dummy variable technique.

$$y_1 = a_1 I_1 + b x_1 + u_1$$

$$y_2 = a_2 I_2 + b x_2 + u_2$$

$$\Rightarrow y_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix} + b \begin{bmatrix} x_1 \end{bmatrix} + u_1$$

$$y_2 = \begin{bmatrix} a_2 \\ 1 \\ \vdots \end{bmatrix} + b \begin{bmatrix} x_2 \end{bmatrix} + u_2$$

$$\Rightarrow y_1 = \begin{bmatrix} a_1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} a_2 \\ 0 \\ b \end{bmatrix} + b \begin{bmatrix} x_1 \end{bmatrix} + u_1$$

$$y_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix} + b \begin{bmatrix} x_2 \end{bmatrix} + u_2$$

$\Rightarrow$ write $a_1 = a_2 = a$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix} + b \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

from two intercept to one $\uparrow$ from one intercept to two

weird situation $n_1 > k, n_2 < k$

$$y_1 = x_1 \beta_1 + u_1$$

$$y_2 = x_2 \beta_2 + u_2$$

$n_1 > k, n_2 < k$

(structural test)

cannot apply LS to y_2 model

we are basically testing whether n_2 observations for the second model came from the same popl from which the first model originates

$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

estimate

$$H_0: \beta_1 = \beta_2$$

$$H_1: \beta_1 \neq \beta_2$$

(insufficient observations)

$$\text{Restricted: } y = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \Rightarrow y = x_0 \beta + u \rightarrow \hat{\beta}_0 = (x_0' x_0)^{-1} x_0' y$$

$$\Rightarrow \hat{u}_0 = y - x_0 \hat{\beta}_0 \Rightarrow RSS_0 = \hat{u}_0' \hat{u}_0$$

$$\text{unrestricted: } y_1 = x_1 \beta + u \Rightarrow \hat{\beta} = (x_1' x_1)^{-1} x_1' y_1 \Rightarrow \hat{u}_1 = y_1 - x_1 \hat{\beta}$$

$$\text{use } \hat{\beta}_1 \Rightarrow RSS = \hat{u}_1' \hat{u}_1$$

$$F = \frac{(RSS_0 - RSS) / k}{RSS / (n_1 - k)} \sim F_{k, n_1 - k}$$

$$df \text{ for } RSS_0 = n_1 + n_2 - k$$

$$RSS = n_1 - k$$

$$df_1 = df_{RSS_0} - df_{RSS} = n_2$$

Nov 18th

Testing of equality of subsets of coeffs in two regressions.

$$\begin{aligned} y_1 &= X_1 \beta_1 + Z_1 \gamma_1 + u_1 = W_1 \delta_1 + u_1 & W_1 &= (X_1, Z_1), \delta_1 = \begin{pmatrix} \beta_1 \\ \gamma_1 \end{pmatrix} \\ y_2 &= X_2 \beta_2 + Z_2 \gamma_2 + u_2 = W_2 \delta_2 + u_2 & W_2 &= (X_2, Z_2), \delta_2 = \begin{pmatrix} \beta_2 \\ \gamma_2 \end{pmatrix} \end{aligned}$$

$$H_0 = \gamma_1 = \gamma_2 = \gamma \text{ (say)} \quad (l_1 = l_2)$$

$$H_1 = \gamma_1 \neq \gamma_2$$

restricted version model:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 & Z_1 \\ 0 & X_2 & Z_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \gamma \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

unrestricted model

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 & Z_1 & 0 \\ 0 & X_2 & 0 & Z_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} Z_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ Z_2 \end{pmatrix} \quad \text{Dummy variable}$$

$$RSS_0 = u_0' u_0; \quad u_0 = y - W$$

$$RSS = \hat{u}' \hat{u}; \quad \hat{u} = y - W \hat{\beta}$$

$$F = \frac{(RSS_0 - RSS) / f_1}{RSS / f_2} \sim F_{f_1, f_2}$$

$$\begin{aligned} f_1 &= \text{# of lin restriction} \\ f_2 &= n + n_2 - (K + L) \end{aligned}$$

$$y = a + bX + cZ + d + u$$

$$H_0: b = c$$

$$E(\hat{b} - \hat{c}) = b - c = 0$$

$$H_1: b \neq c$$

$$V(\hat{b} - \hat{c}) = V(\hat{b}) + V(\hat{c}) - 2\text{Cov}(\hat{b}, \hat{c})$$

$$\text{Let } \theta = b - c$$

$$\begin{aligned} y &= a + (b + \theta)X + cZ + d + u \\ &= a + \theta X + c(X + Z) + d + u \end{aligned}$$

$$y \text{ on } \begin{matrix} \hat{\beta} & \hat{\gamma} \\ x & z \end{matrix}, S \quad \hat{\theta} = \hat{\beta} - \hat{\gamma}$$

$$\begin{pmatrix} 0 & \delta^2 \mathbf{I}_n \\ 0 & \delta^2 \mathbf{I}_n \end{pmatrix}$$

GLS

$$y = x\beta + u \quad \text{--- GM except (1)}$$

$$V(u|x) = \sigma^2 \mathbf{\Omega} \neq \sigma^2 \mathbf{I}_n \quad \mathbf{\Omega} \neq \mathbf{I}_n$$

parismony
unequal v

heteroscedastic
disturbance

i) all u 's indep but some or all have unequal variances

$$y_{ij} = a + b x_{ij} + u_{ij} \quad i=1, \dots, n \quad \sum (u_{ij}) = 0$$

$$j=1, \dots, n_i \quad V(u_{ij}) = \sigma^2$$

$$\sum_{j=1}^{n_i} y_{ij} = \sum_{j=1}^{n_i} a + b \sum_{j=1}^{n_i} x_{ij} + \sum_{j=1}^{n_i} u_{ij}$$

$$= n_i a + b(\quad) + \sum_{j=1}^{n_i} u_{ij} \rightarrow e_i \rightarrow V(e_i) = V(n_i \text{ terms})$$

$$= \sigma^2 + \sigma^2 + \dots + \sigma^2 = n_i \sigma^2$$

aggregate model is

$$(\quad) = n_i a + b(\quad) + e_i \quad \leftarrow \begin{matrix} V(e_i) = n_i \sigma^2 \\ \text{state total eq?} \end{matrix}$$

ii) equal variance of u 's, but some or all u 's are correlated

Let $P_{n \times n}$ be a non-singular matrix and consider the transformation:

$$Py = (Px)\beta + Pu$$

$$\text{i.e. } w = W(\beta + v) \text{ say (2)}$$

The non-singularity of P ensures that there is no loss information entailed in transformation.

from (1) to (2), we choose P such that

$$E(VV') = V(V|x) = E[PUU'P'] = P(\sigma^2 \mathbf{\Omega})P' = \sigma^2 \mathbf{I}_n$$

$$\text{i.e. such that } P\mathbf{\Omega}P' = \mathbf{I}_n, \text{ i.e. } \mathbf{\Omega} = P^{-1}\mathbf{I}_n P'^{-1} = (P'P)^{-1}$$

(1) Is such a P available?

(2) If available, is it unique?

Yes
No

It's not unique but it doesn't matter because we need only $P'P$, which is unique since $(P'P)^{-1} = \Omega$ (unique given)

It is of course obvious that $E(V|X) = P E(u|X) = 0$

$$W = W\beta + V \quad (2)$$

$$E(V|W) = E(V|X) = 0$$

$$V(V|W) = V(V|X) = \sigma^2 I_n$$

So, (2) is a GM, and LS on it is appropriate

Suppose Ω is known, then the LS est of β in (2) is

$$\begin{aligned} \beta_x &= (W'W)^{-1}W'u \\ &= (X'P'PX)^{-1}(X'P'u) \\ &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y) \end{aligned}$$

which is in fact the GLS est of β in (1)

The LS residual vector in (2) is

$$\hat{V} = Pu = Py - Px\beta_x = P(y - X\beta_x)$$

which defines the GLS residual vector in (1) as

$$u_x = y - X\beta_x$$

Finally, the LS est of σ^2 in (2) is

$$\frac{\hat{V}'\hat{V}}{n-k} = \frac{u'P'u}{n-k} = \frac{u_x'\Omega^{-1}u_x}{n-k}, \text{ which is the GLS est of } \sigma^2 \text{ in (1)}$$

Nov 25th

$$y = x\beta + u \rightarrow (1)$$

$$V(u|x) = \sigma^2 \Omega$$

$$py = (px)\beta + pU$$

$$N = 0 \quad W = W\beta + U \quad (2)$$

Consequences of LS on (1) treating (1) as an LS setup

$$\hat{\beta}, \hat{u}, \hat{S}^2$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

$$E(\hat{\beta}|x) = \beta$$

$$E(\hat{u}|x) = E(u|x) = 0$$

$$E(S^2|x) = E(u'u|x) = E\left[\frac{\text{tr} N u u'}{n-k}\right] = \frac{\text{tr} N E(uu'|x)}{n-k}$$

$$= \frac{\sigma^2 \text{tr} N \Omega}{n-k} \neq \frac{\sigma^2 \text{tr} N I_n}{n-k} = \sigma^2$$

$$\therefore E(S^2|x) \neq \sigma^2$$

All inferences developed in the LS framework will be invalid just because $E(S^2|x) \neq \sigma^2$

$$\begin{aligned} V_1(\hat{\beta}|x) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|x] = E[(X'X)^{-1}X'u u'X(X'X)^{-1}|x] \\ &= \sigma^2 (X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

$$V_F(\hat{\beta}|x) = \sigma^2 (X'X)^{-1}$$

$$\therefore V_1(\hat{\beta}|x) \neq V_F(\hat{\beta}|x) = \sigma^2 (X'X)^{-1}$$

$\hat{\beta}$ is unbiased for β

It's also true that $\hat{\beta}$ is consistent for β under mild regularity conditions.

we write $\text{plim } \beta = \beta$.

i.e. $\text{plim } \beta_j = \beta_j, j=1, \dots, k$

i.e. $\hat{\beta}_j \rightarrow \beta_j$ as $n \rightarrow \infty$

($\hat{\beta}_j$ is weakly convergent to β_j)

$$\therefore \text{plim } \hat{u} = \text{plim}(y - x\hat{\beta}) = \text{plim } y - x \text{plim } \hat{\beta} = y - x\beta = u$$

so, the residual vector is consistent for u . This result enables us to estimate Ω consistently.

$$\beta_o = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad \text{feasible GLS est of } \beta$$

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \dots & \omega_{nn} \end{pmatrix}$$

$$\hat{\omega}_{11} = \text{plim } \hat{u}_1^2$$

$$\omega_{11} = E(u_1^2) = V(u_1)$$

$$\therefore \Omega = \begin{pmatrix} \hat{u}_1^2 & & & \\ & \hat{u}_2^2 & & \\ & & \ddots & \\ & & & \hat{u}_n^2 \end{pmatrix}$$

forecasting: $y_{n+1} = x_{n+1}' \beta + u_{n+1}$
 $\downarrow$ not known $\downarrow$ given

$$y_{n+1}^f = \text{forecast of } y_{n+1}, \text{ given } x_{n+1}, x, u$$

$$= \text{est } x_{n+1}' \beta + \text{est } u_{n+1}$$

$$= x_{n+1}' \beta_x + \text{est}[E(u_{n+1} | u_1, \dots, u_n)]$$

$$E(U_{n+1} | \dots, u_1, u_2, \dots, u_n) = 0? \quad N_0 \quad 0 = (V) \quad \gamma$$

Let $U_{n+1} = a_1 u_1 + a_2 u_2 + \dots + a_n u_n + e_{n+1}$, say, be the regression of U_{n+1} on $(u_1, \dots, u_n)$

$E(e_{n+1} | u) = 0$ and e_{n+1} is uncorr with $u_1, \dots, u_n$

$$\therefore E(U_{n+1} | u) = a_1 u_1 + \dots + a_n u_n = a' u \text{ (say)}$$

$$U_{n+1} = a' u + e_{n+1}$$

multiply by u'

$$U_{n+1} u' = a' u u' + e_{n+1} u'$$

and take expectations

$$E(U_{n+1} u') = a' E(u u') + E(e_{n+1} u') = a' \sigma^2 \Omega + 0$$

$$\sigma^2 q' = a' \sigma^2 \Omega \quad \left(\frac{N}{T} \right) + \left(\frac{X}{a} \right) = \left(\frac{P}{T} \right)$$

$$\sigma^2 q' = [\text{cov}(U_{n+1}, u_1), \text{cov}(U_{n+1}, u_2), \dots, \text{cov}(U_{n+1}, u_n)]$$

$$q' \Omega^{-1} = a'$$

$$y_{n+1}^T = x_{n+1}' \beta + a' u_x = x_{n+1}' \beta + q' \Omega^{-1} u_x$$

Mixed estimation

$$y = X\beta + u \quad \text{GM}$$

Prior inf on β (stochastic)

$$r = R\beta + v$$

$$\begin{matrix} \downarrow & \downarrow & \downarrow \\ q \times 1 & q \times k & q \times 1 \end{matrix}$$

$$r(R, R) = r(R) = q < k$$

$$E(V) = 0$$

$$V(V) = \Psi \text{ (known)}$$

V is indep of u

$$y = a + vx + u$$

$\hat{b}$ SE

$$\hat{b} = b + V, \quad E(V) = 0, \quad V(V) = \Psi = (SE)^2$$

$$r = b + v$$

$$= (0, 1) \begin{pmatrix} a \\ b \end{pmatrix} + V$$

$$\begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} x \\ R \end{pmatrix} \beta + \begin{pmatrix} u \\ v \end{pmatrix}$$

$$y_x = x_x \beta + u_x \rightarrow V(u_x) = \begin{pmatrix} V(u) & 0 \\ 0 & V(v) \end{pmatrix} = \begin{pmatrix} \sigma^2_{2n} & 0 \\ 0 & \Psi \end{pmatrix}$$

$$= \sigma^2 \begin{bmatrix} I_{2n} & 0 \\ 0 & \frac{\Psi}{\sigma^2} \end{bmatrix} = \sigma^2 \Omega$$

$$V(\beta_x) = \sigma^2 (X'_x \Omega^{-1} X_x)^{-1} = \sigma^2 \left(\frac{X'_x X_x}{\sigma^2} + \sigma^2 R \Psi^{-1} R' \right)^{-1}$$

$$\beta_x \geq \hat{\beta}$$

$$V(\hat{\beta} | \gamma) = \sigma^2 (X'X)^{-1}$$

show $\sigma^2 (X'X)^{-1} - \left(\frac{X'_x X_x}{\sigma^2} + \sigma^2 R \Psi^{-1} R' \right)^{-1}$ is p.s.d

$$\Rightarrow \frac{X'_x X_x}{\sigma^2} + \sigma^2 R \Psi^{-1} R' - \frac{X'_x X_x}{\sigma^2} \text{ is p.s.d.}$$

— Theil - Goldberger's Compatibility Test

H_0 : prior info is correlated

H_1 : Not H_0

$$\begin{aligned}\varepsilon &= r - R\beta = r - R(\beta + (X'X)^{-1}X'u) \\ &= v - R(X'X)^{-1}X'u\end{aligned}$$

$$E(\varepsilon |) = 0$$

$$\begin{aligned}V(\varepsilon |) &= V(v - R(X'X)^{-1}X'u) \\ &= V(v) + R()^{-1}X E(uu')X()^{-1}R' \\ &= \psi + \sigma^2 R(X'X)^{-1}R' \\ &= \sigma^2 \Sigma_{\varepsilon\varepsilon} \text{ say}\end{aligned}$$

$\therefore$ By THM 1

$$\frac{\varepsilon' \Sigma_{\varepsilon\varepsilon}^{-1} \varepsilon}{\sigma^2} \sim \chi_q^2, \quad q = r(\Sigma_{\varepsilon\varepsilon})$$

$$\frac{\varepsilon' \Sigma_{\varepsilon\varepsilon}^{-1} \varepsilon / q}{\sigma^2} \sim F(q, n-k)$$

Dec 1th

$$\begin{aligned}y_t &= a + bx_t + u_t \\ E(u_t |) &= 0 \\ V(u_t |) &= \sigma^2 z_t^{2\lambda}\end{aligned} \quad \left| \quad \begin{aligned}y_t &= a + bx_t + u_t \\ V(u_t) &= \sigma^2 x_t^{2\lambda}\end{aligned}\right.$$

$$\sigma^2 \Omega = \sigma^2 \begin{pmatrix} z_1^{2\lambda} & 0 & \dots & 0 \\ 0 & \Sigma & & \\ \vdots & & \ddots & \\ 0 & & & z_n^{2\lambda} \end{pmatrix}$$

$$E(u_t^2 |) = \sigma^2 z_t^{2\lambda}$$

$$u_t^2 \approx \sigma^2 z_t^{2\lambda}$$

$$\lg u_t^2 \approx \lg \sigma^2 + 2\lambda \lg z_t + \varepsilon_t$$

$$LS \rightarrow \hat{a}, \hat{b} \rightarrow \hat{u}_t = y_t - \hat{a} - \hat{b}x_t = y_t - 1 - 2x_t = 2$$

$$p \lim \hat{u}_t = u_t$$

$$\therefore p \lim \hat{u}_t^2 = u_t^2$$

$$lg \hat{u}_t^2 = lg \sigma^2 + 2 \lambda g \hat{u}_t + \hat{u}_t^2$$

Time Series

AR(1) Model

$$y_t = a + bx_t + u_t$$

$$u_t = \rho u_{t-1} + v_t \quad |\rho| < 1$$

(noise) v_t 's — i.i.d. $(0, \sigma_v^2)$

$$(i) E(u_t) = 0$$

$$(ii) V(u_t) = \sigma_u^2 \quad \left[\begin{aligned} V(u_t) &= \rho^2 V(u_{t-1}) + V(v_t) \\ \sigma_u^2 &= \rho^2 \sigma_u^2 + \sigma_v^2 \end{aligned} \right]$$

$$(iii) \text{cov}(u_t, u_{t-1}) = \text{cov}(\rho u_{t-1} + v_t, u_{t-1}) = \rho \sigma_u^2 + 0$$

$$\begin{aligned} \text{cov}(u_t, u_{t-2}) &= \text{cov}[\rho u_{t-1} + v_t, u_{t-2}] = \text{cov}(\rho u_{t-1}, u_{t-2}) + \text{cov}(v_t, u_{t-2}) \\ &= \rho^2 \text{cov}(u_{t-2}, u_{t-2}) + \rho \text{cov}(v_{t-1}, u_{t-2}) \\ &= \rho^2 \sigma_u^2 \end{aligned}$$

$$\therefore V(u)_{n \times 1} = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & \rho & 1 \end{bmatrix}$$

$$P \Omega P' = I_n$$

$$\Omega = P^{-1} P'^{-1} = (P'P)^{-1}$$

$$\therefore \Omega^{-1} = P'P$$

$$P = \frac{1}{\sqrt{1-\rho^2}} \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & -\rho & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

CO

$$\begin{cases} y_2 - \rho y_1 = a(1-\rho) + b(x_2 - \rho x_1) + (u_2 - \rho u_1) = V_2 \\ y_3 - \rho y_2 = a(1-\rho) + b(x_3 - \rho x_2) + V_3 \\ \vdots \\ y_n - \rho y_{n-1} = a(1-\rho) + b(x_n - \rho x_{n-1}) + V_n \end{cases}$$

Cochrane - Osmall method

$$y_t = a + bx_t + u_t$$

$$y_{t-1} = a + bx_{t-1} + u_{t-1}$$

$$= a + bx_t + \rho u_{t-1} + V_t$$

$$= a + bx_t + \rho [y_{t-1} - a - bx_{t-1}] + V_t$$

$$y_t - \rho y_{t-1} = a(1-\rho) + b(x_t - \rho x_{t-1}) + V_t, \quad t=0, \dots, n$$

Apply LS. $\hat{a}_\rho = \frac{\text{est}[(a(1-\rho))]}{1-\rho}$ and $\hat{b}_\rho = \text{CO est}^2$

CO est of $\rho = \hat{\rho}$

Apply LS to the original method

$$\hat{a}, \hat{b} \rightarrow \hat{u}_t = y_t - \hat{a} - \hat{b}x_t$$

Next, apply LS to

$$\hat{u}_t = \rho \hat{u}_{t-1} + \text{error}$$

$$\hat{\rho} = \frac{\sum_{t=2}^n \hat{u}_{t-1} \hat{u}_t}{\sum_{t=2}^n \hat{u}_{t-1}^2}$$

$$b_0 = b_1 \times b$$

only care about ρ is consistent
(not efficient in est b)

$$\Rightarrow y_t = a(1-\rho) + \rho y_{t-1} + b x_t - b \rho x_{t-1} + v_t \quad \text{[Durbin Application]}$$

Approach 2S to this: Durbin's $\rho \approx \tilde{\rho} = \text{coeff-est of } y_{t-1}$

$$\text{Durbin's est: } \hat{a}_a = \frac{\text{est } a(1-\tilde{\rho})}{1-\tilde{\rho}} \text{ and } \hat{b}_x$$

$$\lim_{n \rightarrow \infty} (\hat{\rho} - \tilde{\rho}) = 0$$

Forecasting:

$$y_{n+1} = a + b x_{n+1} + u_{n+1}$$

↓
given

$$\text{est } y_{n+1} = \text{est } a + \text{est } b x_{n+1} + \text{est } u_{n+1}$$

$$= \hat{a}_a + \hat{b}_x \cdot x_{n+1} + \hat{\rho} u_n^*$$

$$u_{n+1} = \rho u_n + v_{n+1}$$

$$\text{est } u_{n+1} = \hat{\rho} u_n^* + 0$$

$$\Downarrow$$

$$y_n - \hat{a}_n - \hat{b}_n x_n$$

(GLS est)

$$\text{est } y_{n+3} = \hat{a}_a + \hat{b}_x \cdot x_{n+3} + \hat{\rho} \text{est}(u_{n+2})$$

given

$$= \hat{a}_a + \hat{b}_x \cdot x_{n+3} + \hat{\rho} u_{n+2}^* \quad (\text{if } y_{n+2}, x_{n+2} \text{ are known})$$

$$u_{n+3} = \rho u_{n+2} + v_{n+3} = \rho [\rho u_{n+1} + v_{n+1}] + v_{n+3} = \dots = \rho^3 u_n + \dots$$

$$u_{n+3}^* = \hat{\rho}^3 u_n^*$$

$$\text{est } u_{n+3} = \hat{\rho}^3 \cdot u_n^*$$

Recitation

Sep 15th The model

Assignment 7, 1) $y_i = a + b x_i + u_i$

$$\hat{a} = \bar{y} - \hat{b} \bar{x}$$

or $\hat{a} = \bar{y}$

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\bar{x} = 0$$

$$\Rightarrow L'X = 0$$

$$y_i = a + \text{error}$$

$$\tilde{a} = L'Y(L'L)^{-1} = \bar{y}$$

$$y_i = \beta x_i + \text{error}$$

$$\tilde{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \hat{\beta}$$

$\hat{a} = \tilde{a}$ because the assumption

2) If $L'X = 0$, then $\text{cov}(a, b) = 0$

$$\hat{a} = a + \frac{\sum u_i}{n}$$

$$\hat{b} = b + \frac{\sum x_i u_i}{\sum x_i^2}$$

$$E(\hat{a}) = a, E(\hat{b}) = b$$

$$\text{cov}(\hat{a}, \hat{b}) = E \left\{ [\hat{a} - E(\hat{a})][\hat{b} - E(\hat{b})] \right\} =$$

$$= E \left\{ \frac{\sum u_i}{n} \cdot \frac{\sum x_i u_i}{\sum x_i^2} \right\}$$

$$= E \left\{ \frac{u_1 + u_2 + \dots + u_n}{n} \cdot \frac{x_1 u_1 + x_2 u_2 + \dots + x_n u_n}{\sum x_i^2} \right\}$$

$$= E \left[\frac{x_1 u_1^2}{n \sum x_i^2} + \frac{x_2 u_2^2}{n \sum x_i^2} + \dots + \frac{x_n u_n^2}{n \sum x_i^2} \right] \quad (E$$

$$= \frac{x_1}{n \sum x_i^2} E(u_1^2) + \frac{x_2}{n \sum x_i^2} E(u_2^2) + \dots + \frac{x_n}{n \sum x_i^2} E(u_n^2)$$

$$= \frac{\sum x_i \sigma^2}{n \sum x_i^2} = \frac{n \bar{x} \sigma^2}{n \sum x_i^2} = 0$$

$$2. a) t = \frac{\hat{a} - 0.75}{SE_{\hat{a}}} \sim t_{n-3}$$

$$b) F = \frac{(RSS_r - RSS_{ur}) / \text{# of restrictions}}{RSS_{ur} / (n - (k+1))} = \frac{(RSS_r - RSS_{ur}) / 1}{RSS_{ur} / (n-3)} \sim F_{1, n-3}$$

$$RSS_r > RSS_{ur}$$

↓

$$\log Q = (0.75 - a_1) + (0.75 - a_2) \log K_i + a_2 \log L + u_i$$

$$\log Y^* = -a_1 - a_2 \log \left(\frac{L}{K_i} \right) + u_i$$

Assignment 6.

$$4. 1) R = C_0 + bAC + cP_d + dP_f + u$$

$$P_d = 2P_f$$

not satisfy full-rank condition

$$r(X) = 4$$

$$r(P_d | \text{multicollinearity}) = 3$$

$$R = C_0 + bAC + (c+d)P_f + u$$

$$\text{Or } R = C_0 + bAC + (c + \frac{d}{2})P_d + u$$

$$3. \text{ True: } F = a + b(Z_n + Z_w) + cA + u \quad (3.1)$$

$$F = a + b_1 Z_n + c_1 A + \text{error} \quad (3.2)$$

$$\text{For 3.2: } X_0 = [1_n, Z_n, A]$$

$$X = [1_n, Z_n + Z_w, A]$$

$$X = X_0 + [0, Z_w, 0]$$

$$d_0 = (q_0, b_0, c_0)' \quad d = (q, b, c)'$$

$$\begin{aligned} \hat{d}_0 &= (X_0' X_0)^{-1} X_0' E = (X_0' X_0)^{-1} X_0' [X_0 + [0, I_n, 0]] d + u \\ &= (X_0' X_0)^{-1} X_0' [X_0 d + b I_n + u] \\ &= d + (X_0' X_0)^{-1} X_0' b I_n + (X_0' X_0)^{-1} X_0' u \end{aligned}$$

$$E[\hat{d}_0 | x] = d + b \begin{bmatrix} 1 \\ g \\ h \end{bmatrix} + 0$$

(f, g, h) are the coeff. $I_n = f I_n + g I_n + h A + \text{error}$

$$2.1) \quad y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u$$

$$R_2\beta_2 = r_2$$

$$\min_{\beta, \lambda} f = (y - X\beta)'(y - X\beta) + 2\lambda (r_2 - R_2\hat{\beta}_2)$$

$$\begin{aligned} \frac{\partial f}{\partial \beta} &\Rightarrow -2X'y + 2X'X\beta - 2R_2'\lambda = 0 \\ &\Rightarrow -X'y + X'X\beta - R_2'\lambda = 0 \quad (1) \end{aligned}$$

$$\frac{\partial f}{\partial \lambda} \Rightarrow r_2 = R_2\hat{\beta}_2 \quad (2)$$

$$\text{from (1), } X'X\beta = (X'y + R_2'\lambda)$$

$$\beta = (X'X)^{-1} [X'y + R_2'\lambda] = (X'X)^{-1} X'y + (X'X)^{-1} R_2'\lambda$$

$$\beta = \hat{\beta} + (X'X)^{-1} R_2'\lambda$$

$$\text{from (2)} \quad r_2 = R_2\beta = R_2[\hat{\beta} + (X'X)^{-1} R_2'\lambda]$$

solve for λ .

$$r_2 - R_2\hat{\beta} = R_2(X'X)^{-1} R_2'\lambda$$

$$\lambda = [R_2(X'X)^{-1} R_2']^{-1} (r_2 - R_2\hat{\beta})$$

$$\beta = \hat{\beta} + (X'X)^{-1} R_2' [R_2(X'X)^{-1} R_2']^{-1} (r_2 - R_2\hat{\beta})$$

$$2) R_2 \beta_2 = r_2 \Rightarrow \beta_2 = \beta_2^0 \text{ given, find } \beta_1$$

$$y - X_2 \beta_2^0 = X_1 \beta_1 + u$$

$$\tilde{\beta}_1 = (X_1' X_1)^{-1} X_1' (y - X_2 \beta_2^0)$$

$$3) \tilde{\beta} = \beta_1 + v, \text{ cov}(\tilde{\beta}, v) = 0$$

$$\begin{aligned} \text{cov}(\tilde{\beta}_1, \hat{\beta}_1 - \tilde{\beta}_1) &= \text{cov}(\tilde{\beta}_1, \hat{\beta}_1) - \text{cov}(\tilde{\beta}_1, \tilde{\beta}_1) \\ &= \text{cov}(\tilde{\beta}_1, \hat{\beta}_1) - \text{var}(\tilde{\beta}_1) \\ &= \text{cov}(\tilde{\beta}_1, \hat{\beta}_1) - \sigma^2 (X_1' X_1)^{-1} \end{aligned}$$

$$\text{cov}(\tilde{\beta}_1, \hat{\beta}_1) = E[(\tilde{\beta}_1 - E(\tilde{\beta}_1))(\hat{\beta}_1 - E(\hat{\beta}_1))']$$

$$= E[(\tilde{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1)']$$

$$\hat{\beta}_1 = (X_1' N_2 X_1)^{-1} X_1' N_2 y$$

$$N_2 = I - X_2 (X_2' X_2)^{-1} X_2'$$

$$\text{or } \hat{\beta}_1 = \beta_1 + (X_1' N_2 X_1)^{-1} X_1' N_2 u$$

$$\text{or } \hat{\beta}_1 - \beta_1 = (X_1' N_2 X_1)^{-1} X_1' N_2 u$$

$$\tilde{\beta}_1 = \beta_1 + (X_1' X_1)^{-1} X_1' u$$

$$\text{cov}(\tilde{\beta}_1, \hat{\beta}_1) = E[(X_1' X_1)^{-1} X_1' u u' N_2 X_1 (X_1' N_2 X_1)^{-1}]$$

$$= \sigma^2 (X_1' X_1)^{-1} (X_1' N_2 X_1) (X_1' N_2 X_1)^{-1}$$

$$= \sigma^2 (X_1' X_1)^{-1}$$

$$\therefore \text{cov}(\tilde{\beta}_1, v) = 0$$

$$2) y_i = a + b x_i + u_i$$

$$\hat{b} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

if scaling up

$$S_i = k x_i, k > 1$$

$$y_i = a + b_0 S_i + u_i$$

$$\hat{b}_0 = \frac{\sum (S_i - \bar{S})(y_i - \bar{y})}{\sum (S_i - \bar{S})^2}$$

$$= \frac{\sum (k x_i - k \bar{x})(y_i - \bar{y})}{\sum (k x_i - k \bar{x})^2}$$

$$= \frac{k \sum (x_i - \bar{x})(y_i - \bar{y})}{k^2 \sum (x_i - \bar{x})^2}$$

$$= \frac{1}{k} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\hat{b}}{k}$$

3) If $\hat{b}$ is insignificant $\Rightarrow \hat{b}_0$ is insignificant also?

$$H_0 = b = 0$$

$$H_1 = b \neq 0$$

$$H_0 = b_0 = 0$$

$$H_1 = b_0 \neq 0$$

$$t_b = \frac{\hat{b}}{SE}$$

$$SE^2 = \frac{s^2}{\sum (x_i - \bar{x})^2}$$

$$t_{b_0} = \frac{\hat{b}_0}{SE_0} = \frac{\hat{b}/k}{SE_0}$$

$$t_b = t_{b_0}$$

$$SE_{t_0} = \frac{1}{k} SE_{t_b}$$

$$SE_{t_0}^2 = \frac{s^2}{\sum (S_i - \bar{S})^2} = \frac{s^2}{k^2 \sum (x_i - \bar{x})^2} = \frac{s^2}{k^2}$$

Oct 2nd

Assignment 2

$$C_i = bY_i + u_i$$

$$Y_i = dC_i + v_i$$

Why both relationship cannot exist at the same time?

$$1) C_i = b[dC_i + v_i] + u_i$$

$$(1 - bd)C_i = bv_i + u_i$$

$$C_i = \frac{bv_i + u_i}{1 - bd}$$

$$2) Y_i = d(bY_i + u_i) + v_i$$

$$(1 - bd)Y_i = du_i + v_i$$

$$Y_i = \frac{du_i + v_i}{1 - bd}$$

d is non-zero,

so Y_i is related with u_i , violate

so cannot be applied at the same time

$$E(u_i | Y_i) \neq 0$$

Gauss-Markov assumption

$$\text{Cov}(Y_i, u_i) = 0$$

Assignment 3

$$5. Y = \alpha + \beta X + u$$

$$X = a + bY + v$$

$$V(u) = \sigma_y^2(1 - \rho^2)$$

$$V(v) = \sigma_x^2(1 - \rho^2)$$

Whether regress Y on X if $\sigma_y^2 \leq \sigma_x^2$ and X and Y if reversible

1) Regression should always be based on causality

2) only one of the regression is valid

3. According to GM theorem, when x 's are fixed, $\hat{\beta}$ is BLUE for β ,
 $V(\hat{\beta}|x's) \leq V(\hat{\beta}^*|x's)$ where $\hat{\beta}^*$ ~~are~~ is any other linear unbiased
 estimator of β . when x 's are stochastic, does the statement hold true?

check: $V(\hat{\beta}) \leq V(\hat{\beta}^*)$?

$$V(x) = V(E(x|Y)) + E(V(x|Y))$$

$$V(\hat{\beta}) = V(\underbrace{E(\hat{\beta}|x's)}_{\text{constant}}) + E(\underbrace{V(\hat{\beta}|x's)}_{\text{constant}})$$

$$= 0 + E(V(\hat{\beta}|x's))$$

similarly: $V(\hat{\beta}^*) = E(V(\hat{\beta}^*|x's))$

so if $V(\hat{\beta}|x's) \leq V(\hat{\beta}^*|x's)$

$$V(\hat{\beta}) \leq V(\hat{\beta}^*)$$

so the statement is also true

$$V(\hat{\beta}_n) = \frac{\sigma_n^2}{\sum (x_i - \bar{x}_n)^2} \geq V(\hat{\beta}_{n+1}) = \frac{\sigma_n^2}{\sum (x_i - \bar{x}_{n+1})^2}$$

$$\sum (x_i - \bar{x}_{n+1})^2 = \sum (x_i - \bar{x}_n + \bar{x}_n - \bar{x}_{n+1})^2$$

$$= \sum (x_i - \bar{x}_n)^2 + \sum \{2(\bar{x}_n - \bar{x}_{n+1})(\bar{x}_n - \bar{x}_{n+1})\} + \sum (\bar{x}_n - \bar{x}_{n+1})^2$$

Assig 4

Oct 9th

1. $y_i = a + bx_i + cz_i + u_i$ True Model
 $y_i = p + qx_i + \text{error}$ False Model
 (midterm 2b) $z_i = v + sx_i + \text{error}$

To prove: $\tilde{q}$ is biased

$$\tilde{q} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(\hat{y}_i + u_i)}{\sum (x_i - \bar{x})^2}$$

$(y_i = \hat{y}_i + u_i)$

$$= \frac{\sum (x_i - \bar{x})(\hat{a} + \hat{b}x_i + \hat{c}z_i + u_i)}{\sum (x_i - \bar{x})^2}$$

$$\hat{S} = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (x_i - \bar{x})^2} = \hat{a} \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} + \hat{b} \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} + \hat{c} \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (x_i - \bar{x})^2} + \frac{\sum (x_i - \bar{x})u_i}{\sum (x_i - \bar{x})^2}$$

$$= 0 + \hat{b} + \hat{c}\hat{S} + 0$$

$\tilde{q}$ is biased when $\hat{S} \neq 0$

To prove: if $r_{xz} = 0$, then $\tilde{q} = \hat{b}$

to show $\hat{S} = 0$

$$\hat{S} = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{S_x^2}$$

$$r_{xz} = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{S_x \cdot S_z}$$

$$r_{xz} = \frac{\hat{S} S_x^2}{S_x \cdot S_z} = \frac{\hat{S} S_x}{S_z} \Rightarrow \hat{S} = r_{xz} \frac{S_z}{S_x}$$

so $\tilde{q} = \hat{b} + \hat{c}\hat{S} = \hat{b}$

$$3. y = x_1 \beta_1 + x_2 \beta_2 + u : \text{True Model}$$

$$y = x_1 \beta_1 + \text{error} : \text{False Model}$$

$$\hat{\beta}_1 = (x_1' x_1)^{-1} x_1' y = (x_1' x_1)^{-1} x_1' (\hat{y} + u) = (x_1' x_1)^{-1} x_1' (x_1 \hat{\beta}_1 + x_2 \hat{\beta}_2 + u)$$

$$= \hat{\beta}_1 + (x_1' x_1)^{-1} x_1' x_2 \hat{\beta}_2 + (x_1' x_1)^{-1} x_1' u$$

$$E(\hat{\beta}_1) = \hat{\beta}_1 + (x_1' x_1)^{-1} x_1' x_2 \hat{\beta}_2 \quad \text{biased}$$

$$= 0 \quad \text{unbiased}$$

$$2. y_i = a + b x_i + c z_i + u_i \quad (1.1)$$

c is known

$$y_i - c z_i = a + b x_i + u_i$$

$$\text{or } y_i' = a + b x_i + u_i \quad (1.2)$$

$$\Rightarrow \hat{b}_0, \hat{a}_0$$

$$\frac{(u_i - \bar{u})^2}{(n-1)} = \frac{(w_i - \bar{w})(v_i - \bar{v})}{(n-1)}$$

$$\text{cov}(w_i, b v_i + u_i)$$

Oct 14th.

Midterm 1

$$1. (i) y_i = a + b(2w_i + v_i) + u_i$$

$$(1.3) = a + b2w_i + bv_i + u_i \rightarrow \text{operation version of } (1.1)$$

$\tilde{b} \rightarrow \text{est of } b \text{ from 1.2}$

$$(ii) \text{GM? } bv_i + u_i = e_i$$

$$① E(e_i | W's) = 0$$

$$= E(u_i | W's) + b E(v_i | W's)$$

③ $\text{cov}(e_i, W_i)$

$$= 0 \quad \begin{matrix} \Downarrow \\ E(u_i | X's) = 0 \text{ from 1.1} \end{matrix} \quad \begin{matrix} \text{indep of } u_i \\ E(u_i | 2w_i + v_i) = 0, \text{ so } E(u_i | W's) = 0. \end{matrix}$$

② are the dist indep homoscedastic, mean 0, $\text{Var} = \sigma^2$, say

$$V(e_i | W's) = V(u_i | W's) + V(bv_i | W's) = \sigma^2 + b^2 \sigma_v^2 = \sigma^2 \text{ (say)}$$

(iii) from 1.3

$$\hat{b}_0 = \frac{\sum (w_i - \bar{w})(y_i - \bar{y})}{\sum (w_i - \bar{w})^2}$$

$$\tilde{b} = \frac{\hat{b}_0}{2}$$

$$V(\tilde{b} | X's) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

$$V(\hat{b}_0 | W's) = \frac{\sigma^2}{\sum (w_i - \bar{w})^2}$$

$$\therefore V(\tilde{b} | W's) = V\left(\frac{\hat{b}_0}{2} | W's\right) = \frac{1}{2^2} V(\hat{b}_0 | W's) = \frac{1}{2^2} \cdot \frac{\sigma^2}{\sum (w_i - \bar{w})^2}$$

$$h = \frac{V(\tilde{b} | W's)}{V(\tilde{b} | X's)} = \frac{\frac{\sigma^2}{2^2} \cdot \frac{1}{\sum (w_i - \bar{w})^2}}{\frac{\sigma^2}{\sum (x_i - \bar{x})^2}} = \frac{\sum (x_i - \bar{x})^2}{2^2 \sum (w_i - \bar{w})^2} = \frac{\sigma^2 \cdot [\sum (w_i - \bar{w})^2 + \sum (v_i - \bar{v})^2]}{\sigma^2 \sum (w_i - \bar{w})^2}$$

$$E(h|w's) = \frac{\sigma^2}{\sigma^2} \cdot \frac{\frac{2^2 \sum (w_i - \bar{w})^2 + \frac{E}{2} (v_i - \bar{v})^2}{2^2 \sum (w_i - \bar{w})^2}} \geq 1 \quad \text{if } 1 \rightarrow 0$$

3. a) $y = X\beta + u$ } $n+p$ observation on the left
 $Z\beta = Z\beta + e$

$$y_0 = X_0 \beta + u_0 \Rightarrow \hat{b}$$

(mpx1) $\downarrow$ $\downarrow$

$$\begin{pmatrix} y \\ z\beta \end{pmatrix} = \begin{pmatrix} X \\ Z \end{pmatrix} \begin{pmatrix} \beta \\ \beta \end{pmatrix}$$

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

$$Z\hat{\beta} = Z(X'X)^{-1} X'Y$$

show that $\hat{b} = \hat{\beta}$

$$(X_0' X_0) \hat{b} = X_0' Y_0$$

$$(X'X + Z'Z) \hat{b} = X'Y + Z'Z\hat{\beta} = X'X\hat{\beta} + Z'Z\hat{\beta} = (X'X + Z'Z) \hat{\beta}$$

$$\text{so } \hat{b} = \hat{\beta}$$

b) $y_i = a + bx_i + u_i$

$$V(\hat{b}|x's) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \frac{\sigma^2}{\sum x_i^2 - n\bar{x}^2}$$

under the prior $a=0$.

$$y_{i-2} = bx_i + u \quad (\text{not estimating } a)$$

$$\hat{b}_0 = \frac{\sum x_i y_{i-2}}{\sum x_i^2}$$

$$V(\hat{b}_0|x's) = \frac{\sigma^2}{\sum x_i^2}$$

$$V(\hat{b}_0|x's) < \left(\frac{\sigma^2}{\sum x_i^2 - n\bar{x}^2} \right) = V(\hat{b}|x's)$$

True, which have additional prior

Oct 30th

Assignment 5

1. $y = X'\beta + u$

(1.1) $y = x_1\beta_1 + \beta_{k-2}x + \beta_{k-1}z + \beta_k s + u$

(1.2) $y = x_1\beta_1 + 0x + \beta_{k-1}(z+2x) + \beta_k(s-x) + u$

(ii) Is $\hat{\beta}$ from (1.2) are identical to $\hat{\beta}$ (1.1)

To prove $\hat{\beta} = A^{-1}\hat{\beta}_x$, where A is non-singular matrix

(i) what is 0?

(1.2) $y = x_1\beta_1 + (0x + 2\beta_{k-1} - \beta_k)x + \beta_k s + u$

$$\beta_{k-2} = 0 + 2\beta_{k-1} - \beta_k$$

$$0 = \beta_{k-2} - 2\beta_{k-1} + \beta_k$$

(ii)
$$\begin{pmatrix} \beta_1 \\ 0 \\ \beta_{k-1} \\ \beta_k \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_{k-2} \\ \beta_{k-1} \\ \beta_k \end{pmatrix}$$

$$\hat{\beta}_x = A \hat{\beta}$$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = I \neq 0$$

$$A^{-1}\hat{\beta}_x = A^{-1}A\hat{\beta}$$

$$\hat{\beta} = A^{-1}\hat{\beta}_x$$

Yes, they are identical (not equal)