-739 1

MULTIPLE REGRESSION Re-examined

The Classical Regression Model in Matrix Notation

[put on bd & use]

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \epsilon_t \qquad t = 1, \ldots, T$$

where
- y_t = t^{th} obs. on dependent variable
- x_{ti} = t^{th} obs. on i^{th} explanatory variable
- ϵ_t = t^{th} value of disturbance
- β_i = i^{th} regression coefficient

- T = # of observations
- K = # of regressors

In order to include a constant term, let the first regressor = 1 for all T obs.
$(x_{t1} = 1 \;\; \forall\, t)$

We have:

$$\left.\begin{aligned}
y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_K x_{1K} + \epsilon_1 \\
y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_K x_{2K} + \epsilon_2 \\
&\;\;\vdots \\
y_T &= \beta_1 x_{T1} + \beta_2 x_{T2} + \cdots + \beta_K x_{TK} + \epsilon_T
\end{aligned}\right\} T \text{ obs.}$$

$$\underset{T\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}} = \underset{T\times K}{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & & & \\ x_{T1} & x_{T2} & \cdots & x_{TK} \end{bmatrix}} \underset{K\times 1}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}} + \underset{T\times 1}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{bmatrix}}$$

OR:

$$Y = X\beta + \epsilon$$

ALGEBRAIC RESULTS

no assumptions;
no statistics

Def$_n$: The Ordinary Least Squares (OLS) estimator, $\hat{\beta}$ of β, is the value that minimizes

$$SSE = (Y - X\beta)'(Y - X\beta)$$
$$= \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{K} \beta_i x_{ti} \right)^2$$

Take a minute. Change mode of thinking from cumbersome Σ - notation to Matrix notation.
– mathematically equivalent !!!
Review Matrix Algebra!

$$(Y - X\beta) = \underset{T \times 1}{\begin{bmatrix} y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \cdots - \beta_K x_{1K} \\ y_2 - \beta_1 x_{21} - \beta_2 x_{22} - \cdots - \beta_K x_{2K} \\ \vdots \\ y_T - \beta_1 x_{T1} - \beta_2 x_{T2} - \cdots - \beta_K x_{TK} \end{bmatrix}}$$

Hence, $SSE = (Y - X\beta)'(Y - X\beta)$

$$= \underset{1 \times T}{\begin{bmatrix} y_1 - \sum_{i=1}^{K} \beta_i x_{1i} & y_2 - \sum_{i=1}^{K} \beta_i x_{2i} & \cdots & y_T - \sum_{i=1}^{K} \beta_i x_{Ti} \end{bmatrix}} \underset{T \times 1}{\begin{bmatrix} y_1 - \sum_i \beta_i x_{1i} \\ y_2 - \sum_i \beta_i x_{2i} \\ \vdots \\ y_T - \sum_i \beta_i x_{Ti} \end{bmatrix}}$$

$$= \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{K} \beta_i x_{ti} \right)^2 \qquad \text{OLS } \hat{\beta} \text{ minimizes this.}$$

Thm: The OLS estimator of $\hat{\beta}$ is
$$\hat{\beta} = (X'X)^{-1} X'Y$$

Proof:
$$SSE := (Y - X\beta)'(Y - X\beta)$$
$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

$$\frac{\partial SSE}{\partial \beta} = -X'Y - X'Y + 2X'X\beta \qquad \text{from Exam 1}$$

*

$$-2X'Y + 2X'X\hat{\beta} = 0$$
$$\underset{(K\times 1)}{X'X\hat{\beta}} = \underset{(K\times 1)}{X'Y} \qquad \text{(the Normal Equations)}$$

K equations, K unknowns = $\hat{\beta}$

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

Note: $(X'X)$ is square, $K\times K$; must be nonsingular. (invertible)

Defn: The vector of fitted values is $\hat{Y} = X\hat{\beta}$.

Defn: The vector of residuals is
$$\underset{T\times 1}{e} = \underset{T\times 1}{Y - X\hat{\beta}} = Y - \hat{Y}$$

Note: True disturbances $= \epsilon = Y - X\beta \neq e$.

Note: $e_t = y_t - \sum_{i=1}^{K} \hat{\beta}_i x_{ti}$

$$SSE = \sum_{t=1}^{T} e_t^2 = \mathbf{e}'\mathbf{e}$$

<u>Proposition:</u> $X'e = 0$; the residuals are orthogonal to the X's.

Proof:

$$
\begin{aligned}
X'e &= X'(Y - X\hat{\beta}) \\
&= X'Y - X'X\hat{\beta} \\
&= X'Y - X'X[(X'X)^{-1}X'Y] \\
&= X'Y - X'Y \\
&= 0
\end{aligned}
$$

Note: $\underset{K\times T}{X'}\ \underset{T\times 1}{e} = \underset{K\times 1}{0}$

→ The K Normal Equations again!

In Σ-notation:

$$\sum_{t=1}^{T} X_{ti}\, e_t = 0 \qquad i = 1, \ldots, K$$

Recall, for constant term let $X_{t1} = 1 \ \forall t$.

Thus, the first Normal Equation, $\sum_{t=1}^{T} e_t = 0$.

Notation:

$$\hat{Y} = X\hat{\beta} \qquad \text{fitted values}$$

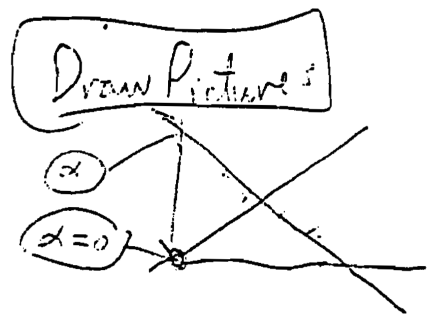
$$\left.
\begin{aligned}
SST &= \sum_t (y_t - \bar{y})^2 \\
SSR &= \sum_t (\hat{y}_t - \bar{y})^2 \\
SSE &= \sum_t (y_t - \hat{y}_t)^2 = \sum_t e_t^2 = e'e
\end{aligned}
\right\} \text{ as before}$$

$$R^2 = 1 - \frac{SSE}{SST} \quad ; \quad \bar{R}^2 = 1 - \frac{SSE / T-K}{SST / T-1}$$

<u>Thm:</u> If one of the regressors is a constant term,
$SST = SSR + SSE$.

Proved already in Simple Regression case.

Note: If $\nexists$ constant term,
SST may <u>not</u> $= SSR + SSE$.



<u>Thm:</u> If one of the regressors is a constant term,
$0 \le R^2 \le 1$.

--- In that case, $R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$;

With no constant term, the last equality need not hold!

Note: Computer programs may use the formula,
$R^2 = \frac{SSR}{SST}$.
w/o a constant term R^2 may be > 1.
($SSR + SSE$ may be $> SST$)

Note: $R^2 = \rho^2(Y, \hat{Y})$ → always works (between 0 & 1).

Keep this in mind when using no constant term.
SAS uses the formula, $R^2 = \frac{SSR}{SST}$! Warns you if $\exists$ no intercept.

Statistical Results

Defn: The model $Y = X\beta + \epsilon$ satisfies the <u>full ideal conditions</u> if:

1) $\epsilon \sim N(0, \sigma^2 I_T)$
2) X is a nonstochastic matrix of Rank K $\ni \lim_{T \to \infty} \left(\frac{X'X}{T}\right)$ is finite and nonsingular.

explanation:

1) $\rightarrow$ $$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \ddots \\ & & & \sigma^2 \end{bmatrix} \right)$$

or ϵ_t iid $N(0, \sigma^2)$.

(i) Normal $\forall t$
(ii) $E(\epsilon_t) = 0 \quad \forall t$
(iii) $Var(\epsilon_t) = \sigma^2 \quad \forall t$ (constant varianc
(iv) $Cov(\epsilon_t, \epsilon_s) = 0 \quad \forall t \neq s$ (indep.)

2) (i) nonstochastic $\rightarrow$ fixed or nonrandom $\forall$ observation.
* Generally untrue with economic data.

(ii) X is of Rank K; K columns linearly independant; allows inverse $(X'X)^{-1}$ to exist. & $\therefore$ unique estimate.

(iii) Consider $\lim_{T\to\infty} \frac{(X'X)}{T}$.

$$\frac{(X'X)_{ij}}{T} = \frac{1}{T}\sum_{t=1}^{T} x_{ti}\, x_{tj}$$

= avg. of the cross-products of the observations on two of the X's.

Note: If X's measured as deviations from means,

$$\frac{(X'X)_{ij}}{T} = \mathrm{Cov}(X_i, X_j) = \frac{1}{T}\sum_{t=1}^{T}(x_{ti}-\bar{x}_i)(x_{tj}-\bar{x}_j)$$

$$= \frac{1}{T}\sum_{t=1}^{T} x_{ti}\, x_{tj} - T\bar{x}_i\,\bar{x}_j$$

✓ → $\frac{1}{T}\sum_{t=1}^{T} x_{ti}\, x_{tj}$ is like a "sufficient statistic" for $\mathrm{Cov}(X_i, X_j)$;
it contains the relevant information.

Requiring $\lim_{T\to\infty} \frac{(X'X)}{T}$ to be <u>finite & nonsingular</u>

means that we don't want these "Covariances" to blow up as T ↑es;
i.e. we don't want any two X's to be highly correlated with <u>each</u> <u>other</u>.

⟶ if problem exists, Multicollinearity; problems!

nonsingularity → $(X'X)^{-1}$ exists.

Classical Least Squares

Consider the properties of OLS $\hat{\beta}$.

Lemma: $\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$

Proof:
$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon \\ &= \beta + (X'X)^{-1}X'\epsilon\end{aligned}$$

Look familiar?

Write in Σ-notation, with $K=1$, x & y as deviations from mean

$$\hat{\beta} = \beta + \left([x_1 \cdots x_T] \begin{bmatrix} x_1 \\ \vdots \\ x_T \end{bmatrix} \right)^{-1} [x_1 \cdots x_T] \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix}$$

$$= \beta + \frac{\sum_{t=1}^{T} x_t \epsilon_t}{\sum_{t=1}^{T} x_t^2} \equiv \beta + \frac{\sum_{t=1}^{T} (x_t - \bar{x}) \epsilon_t}{\sum_{t=1}^{T} (x_t - \bar{x})^2}$$

Proved in Simple Regression case!

Note: Proof here very simple in matrix notation.

<u>Thm:</u> $\hat{\beta}$ is unbiased.

Proof:
$$E(\hat{\beta}) = E[\beta + (X'X)^{-1}X'\epsilon]$$
$$= \beta + (X'X)^{-1}X'E(\epsilon) \qquad (X\text{'s nonstochastic})$$
$$= \beta$$

(41)

<u>Thm:</u> The Covariance Matrix of $\hat{\beta}$ is $\sigma^2(X'X)^{-1}$.

Proof:
$$\mathrm{Cov}(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\}$$
$$= E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$$
$$= E[(X'X)^{-1}X'\epsilon][(X'X)^{-1}X'\epsilon]'$$
$$= E[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}]$$
$$= (X'X)^{-1}X'\, E(\epsilon\epsilon')\, X(X'X)^{-1} \qquad (X\text{'s nonstochastic})$$
$$= (X'X)^{-1}X' \; \sigma^2 I_T \; X(X'X)^{-1} \qquad (\mathrm{Cov}(\epsilon) = \sigma^2 I_T)$$
$$= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} \qquad (\sigma^2 \text{ is a scalar})$$
$$= \sigma^2 (X'X)^{-1}$$

Note: Assumption 1) says $\epsilon \sim N(0, \sigma^2 I_T)$;
i.e. $\mathrm{Cov}(\epsilon) = E(\epsilon\epsilon') = \sigma^2 I_T$.

Again, look familiar?
Write in Σ-notation, with $K=1$, x & y as deviations from mean.

$$K=1; \quad \underset{1\times 1}{\mathrm{Cov}(\hat{\beta})} = \mathrm{Var}(\hat{\beta}) = \sigma^2\left([x_1 \cdots x_T]\begin{bmatrix} x_1 \\ \vdots \\ x_T \end{bmatrix}\right)^{-1} = \frac{\sigma^2}{\sum_{t=1}^{T} x_t^2} = \frac{\sigma^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

In multiple regression case, we considered diagonal: this has more!! $\Rightarrow$

In general case, $Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ is $K \times K$.
Along diagonal, $Var(\hat{\beta}_i)$, $i = 1, \dots, K$.
off diagonal, $Cov(\hat{\beta}_i, \hat{\beta}_j)$, $i, j = 1, \dots, K$.

Most computer programs provide this information.
(SAS does if you ask for it.)

Thm.: (Gauss - Markov)
$\hat{\beta}$ is the BLUE of β.

Proof: Consider any other linear estimate of β;
$\tilde{\beta} = CY$ where C is a $K \times T$ matrix
($\tilde{\beta}$ is a linear combination of the y's).

Notation: $D = C - (X'X)^{-1}X'$
or $C = (X'X)^{-1}X' + D$

Note: D = difference between C and its value if $\tilde{\beta} = \hat{\beta}$.
If $D = 0$, $C = (X'X)^{-1}X'$ and $\tilde{\beta} = \hat{\beta}$.

$$
\begin{aligned}
E(\tilde{\beta}) = E(CY) &= C\,E(Y) \\
&= C\,E(X\beta + \epsilon) \\
&= CX\beta \\
&= [(X'X)^{-1}X' + D]\,X\beta \\
&= (X'X)^{-1}X'X\beta + DX\beta \\
&= \beta + DX\beta
\end{aligned}
$$

Thus for unbiasedness, require $DX = 0$ (or $CX = I$).

Hence we are requiring $\tilde{\beta}$ to be <u>linear</u> and <u>unbiased</u>; if $\hat{\beta}$ is "better" than any such $\tilde{\beta}$, $\hat{\beta}$ is BLUE. Compare their covariance matrices!

$$\mathrm{Cov}(\tilde{\beta}) = E(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'$$

Note:
$$\begin{aligned}
\tilde{\beta} = CY &= [(X'X)^{-1}X' + D][X\beta + \epsilon] \\
&= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon + DX\beta + D\epsilon \\
&= \beta + (X'X)^{-1}X'\epsilon + D\epsilon \qquad (\text{since } DX\beta = 0)
\end{aligned}$$

$$\rightarrow \quad \begin{aligned}
\tilde{\beta} - \beta &= (X'X)^{-1}X'\epsilon + D\epsilon \\
&= [(X'X)^{-1}X' + D]\epsilon
\end{aligned}$$

Thus,
$$\begin{aligned}
\mathrm{Cov}(\tilde{\beta}) &= E\{[(X'X)^{-1}X' + D]\epsilon\}\{[(X'X)^{-1}X' + D]\epsilon\}' \\
&= E[(X'X)^{-1}X' + D]\epsilon\epsilon'[X(X'X)^{-1} + D'] \\
&= [(X'X)^{-1}X' + D]\, E(\epsilon\epsilon')\, [X(X'X)^{-1} + D'] \\
&= [(X'X)^{-1}X' + D]\, \sigma^2 I_T\, [X(X'X)^{-1} + D'] \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} + \sigma^2(X'X)^{-1}X'D' \\
&\quad + \sigma^2 DX(X'X)^{-1} + \sigma^2 DD' \\
&= \sigma^2(X'X)^{-1} + \sigma^2 DD' \qquad (\text{since } DX = X'D' = 0)
\end{aligned}$$

Comparing $\mathrm{Cov}(\hat{\beta})$ and $\mathrm{Cov}(\tilde{\beta})$;

$$\mathrm{Cov}(\tilde{\beta}) - \mathrm{Cov}(\hat{\beta}) = \sigma^2 DD'$$ and this is <u>psd</u>.

Thus $\hat{\beta}$ is efficient relative to any other linear unbiased estimator $(\tilde{\beta})$; $\hat{\beta}$ is BLUE.

Thm: $\hat{\beta}$ is consistent.

Proof: (easy way) – $\hat{\beta}$ is unbiased. Thus $\hat{\beta}$ is consistent if its covariance matrix goes to 0.

Note: $(X'X)^{-1} = \frac{T}{T}(X'X)^{-1} = \frac{1}{T}(T^{-1}X'X)^{-1} = \frac{1}{T}\left(\frac{X'X}{T}\right)^{-1}$

Take
$$\lim_{T\to\infty} \sigma^2 (X'X)^{-1} = \lim_{T\to\infty} \sigma^2 \frac{1}{T}\left(\frac{X'X}{T}\right)^{-1}$$
$$= \underbrace{\lim_{T\to\infty} \frac{\sigma^2}{T}}_{0} * \underbrace{\lim_{T\to\infty}\left(\frac{X'X}{T}\right)^{-1}}_{\substack{\text{(finite limit)} \\ \text{by ideal conditions}}}$$
$$= 0$$

Thus, $Cov(\hat{\beta}) \to 0$ as $T \to \infty$, so $\hat{\beta}$ is consistent.

(hard way) – Show $\operatorname{plim} \hat{\beta} = \beta$.

$$\begin{aligned}
\operatorname{plim} \hat{\beta} &= \beta + \operatorname{plim} (X'X)^{-1} X'\epsilon \\
&= \beta + \operatorname{plim} \left(\frac{X'X}{T}\right)^{-1} \frac{X'\epsilon}{T} \\
&= \beta + \underbrace{\left[\operatorname{plim} \frac{X'X}{T}\right]^{-1}}_{\text{(finite limit)}} \underbrace{\operatorname{plim} \frac{X'\epsilon}{T}}_{\substack{\text{must show} \\ \text{this } = 0}}
\end{aligned}$$

→ intuition: X's & ϵ's uncorrelated.

44 $\llcorner$* $\operatorname{plim} \frac{X'\epsilon}{T}$ is like the "asymptotic correlation matrix" between X & ϵ.

Note: $\operatorname{plim} \frac{X'\epsilon}{T} = 0$ iff $E\left(\frac{X'\epsilon}{T}\right) = 0$ and $\operatorname{Cov}\left(\frac{X'\epsilon}{T}\right) \to 0$ as $T \to \infty$.

$$E\left(\frac{X'\epsilon}{T}\right) = \frac{X'}{T} E(\epsilon) = 0$$

$$\begin{aligned}
\operatorname{Cov}\left(\frac{X'\epsilon}{T}\right) &= E\left(\tfrac{1}{T}X'\epsilon\right)\left(\tfrac{1}{T}X'\epsilon\right)' \\
&= \frac{1}{T^2} E(X'\epsilon\epsilon' X) \\
&= \frac{1}{T^2} X' E(\epsilon\epsilon') X \\
&= \frac{1}{T^2} X' \sigma^2 I_T X \\
&= \frac{\sigma^2}{T^2} X'X
\end{aligned}$$

and since we know $\frac{X'X}{T} \to$ finite limit, by ideal conditions,

$$\sigma^2 \frac{X'X}{T^2} \to 0 \quad \text{as } T \to \infty.$$

Hence, $\operatorname{plim} \frac{X'\epsilon}{T} = 0$ and $\operatorname{plim} \hat{\beta} = \beta$. QED

Estimation of σ^2

Lemma: The matrix, $M = I_T - X(X'X)^{-1}X'$, is $T \times T$ and:

i) symmetric; $M = M'$
ii) idempotent; $MM = M$
iii) orthogonal to X; $MX = 0$

—— prove these yourself

Proposition: $e = M\epsilon$

Proof:
$$\begin{aligned} e &= Y - X\hat{\beta} \\ &= Y - X(X'X)^{-1}X'Y \\ &= MY \\ &= M(X\beta + \epsilon) \\ &= M\epsilon \qquad \text{since } MX = 0 \end{aligned}$$

Proposition: $SSE = \epsilon' M \epsilon$

Proof:
$$\begin{aligned} SSE &= e'e \\ &= (M\epsilon)'(M\epsilon) \\ &= \epsilon' M' M \epsilon \\ &= \epsilon' M M \epsilon \qquad M \text{ is symmetric} \\ &= \epsilon' M \epsilon \qquad M \text{ is idempotent} \end{aligned}$$

Digression on Trace:

If C is a $p \times p$ matrix, $\text{Trace}(C) = Tr(C) = \sum_{i=1}^{p} c_{ii}$

Useful Facts:

① $Tr(A+B) = Tr(A) + Tr(B)$

② If A is $m \times n$ and B is $n \times m$,
$Tr(AB) = Tr(BA)$ -----(should be able to prove)

③ $\text{Trace}(M) = T - K \Rightarrow$
$$\begin{aligned} Tr(M) &= Tr(I_T - X(X'X)^{-1}X') \\ &= Tr(I_T) - Tr[\underbrace{X(X'X)^{-1}X'}_{T\times T}] \\ &= T - Tr[\underbrace{(X'X)^{-1}X'X}_{K\times K}] \\ &= T - Tr(I_K) \\ &= T - K \end{aligned}$$

End of Digression

6 *

Thm: $E(SSE) = \sigma^2(T-K)$

Proof:

$$\begin{aligned}
E(SSE) &= E(\epsilon' M \epsilon) && \epsilon' M \epsilon \text{ is } 1\times 1; \\
&= E[Tr(\epsilon' M \epsilon)] && \therefore \epsilon' M \epsilon = Tr(\epsilon' M \epsilon) \\
&= E[Tr(\underbrace{M\epsilon\epsilon'}_{T\times T})] && \text{by Fact ②, Digression} \\
&= Tr[E(M\epsilon\epsilon')] && E(Tr) = Tr(\text{Exp. Value}) \\
&= Tr[M\, E(\epsilon\epsilon')] && M \text{ is fixed} \\
&= Tr(M\sigma^2 I_T) && \\
&= \sigma^2\, Tr(M) && \\
&= \sigma^2(T-K) && \text{by Fact ③, Digression}
\end{aligned}$$

Thm: $s^2 = \frac{1}{T-K}(SSE)$ is an unbiased estimator of σ^2.

Proof:

$$\begin{aligned}
E(s^2) &= E\left(\frac{1}{T-K} SSE\right) \\
&= \frac{1}{T-K} E(SSE) \\
&= \frac{1}{T-K}(T-K)\sigma^2 \\
&= \sigma^2
\end{aligned}$$

*7

Thm: s^2 is consistent.

Proof: show $\text{plim } s^2 = \sigma^2$.

$$
\begin{aligned}
s^2 &= \frac{1}{T-K} SSE \\
&= \frac{1}{T-K} \epsilon' M \epsilon \\
&= \frac{1}{T-K} \epsilon' [I_T - X(X'X)^{-1}X'] \epsilon \\
&= \frac{1}{T-K} \epsilon'\epsilon - \frac{1}{T-K} \epsilon' X (X'X)^{-1} X' \epsilon
\end{aligned}
$$

first term: $\text{plim } \frac{1}{T-K} \epsilon'\epsilon = \text{plim } \frac{1}{T} \epsilon'\epsilon$

$$
\begin{aligned}
&= \text{plim } \frac{1}{T} \sum_{t=1}^{T} \epsilon_t^2 \\
&= \sigma^2 \quad \text{by Law of Large Number.}
\end{aligned}
$$

$$
\left[\text{Recalling that } \begin{aligned} \sigma^2 &= Var(\epsilon_t) \quad \forall\, t \\ &= E(\epsilon_t^2) \end{aligned} \right]
$$

second term: $\text{plim } \frac{1}{T-K} \epsilon' X (X'X)^{-1} X' \epsilon$

$$
\begin{aligned}
&= \text{plim } \frac{1}{T} \epsilon' X (X'X)^{-1} X' \epsilon \\
&= \text{plim } \frac{\epsilon' X}{T} \left(\frac{X'X}{T} \right)^{-1} \frac{X'\epsilon}{T} \\
&= \underbrace{\text{plim } \frac{\epsilon' X}{T}}_{\text{plim} = 0} \underbrace{\left[\text{plim } \frac{X'X}{T} \right]^{-1}}_{\text{finite plim}} \underbrace{\text{plim } \frac{X'\epsilon}{T}}_{\text{plim} = 0}
\end{aligned}
$$

Thus, $\text{plim } s^2 = \sigma^2$; s^2 is consistent.

SUMMARY of results so far:

$\hat{\beta}$ is unbiased, BLUE, consistent.

s^2 is unbiased, consistent.

8*

Up to now, proved facts under the ideal conditions w/o using the assumption of Normality for ϵ. We'll begin to use this assumption now in the following attempts to show that $\hat{\beta}$ & s^2 are efficient.

First, with the Normality assumption we'll find the MLE's of β & σ^2 in that distribution.

Thm: The MLE's of β & σ^2 are

$$\hat{\beta} = (X'X)^{-1} X'Y \quad \& \quad \hat{\sigma}^2 = \frac{SSE}{T}. \leftarrow (\text{almost } s^2)$$

Proof: The log likelihood function is -

$$\ln L = -\frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \epsilon_t^2$$

$$= -\frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \epsilon'\epsilon$$

$$= -\frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)$$

*(Parenthetical note - The Jacobian of the transformation from ϵ to Y is 1; $\therefore$ the $\ln(J)$ is zero.)

$$= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}\left[Y'Y - 2Y'X\beta + \beta'X'X\beta\right]$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{1}{2\sigma^2}\left[-2X'Y + 2X'X\beta\right] = -\frac{1}{\sigma^2}\left[X'X\beta - X'Y\right]$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\left[Y'Y - 2Y'X\beta + \beta'X'X\beta\right]$$

Setting the first derivatives equal to zero:

$$-\frac{1}{\hat\sigma^2}\left[X'X\hat\beta - X'Y\right] = 0$$

$$X'X\hat\beta = X'Y$$

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

$$-\frac{T}{2\hat\sigma^2} + \frac{1}{2\hat\sigma^4}(Y-X\hat\beta)'(Y-X\hat\beta) = 0$$

$$-\frac{T}{2\hat\sigma^2} + \frac{1}{2\hat\sigma^4}SSE = 0$$

$$\frac{T}{\hat\sigma^2} = \frac{SSE}{\hat\sigma^4}$$

$$\hat\sigma^2 = \frac{SSE}{T}$$

QED

Note: Since MLE's are
i) consistent
ii) asymptotically efficient
iii) have asymptotic covariance matrix, $\mathcal{I}^{-1}$

then these are true of $\hat\beta$ and $\hat\sigma^2 = \frac{SSE}{T}$.

Note also that $\hat\sigma^2 = \frac{SSE}{T}$ is biased,
recalling that $s^2 = \frac{SSE}{T-K}$ is unbiased.

Furthermore, $s^2 = \frac{1}{T-K}SSE = \frac{T}{T-K}\hat\sigma^2 \to \hat\sigma^2$ as $T \to \infty$.
Hence these are all true of s^2 as well.

To show small-sample efficiency of MLE's $\hat\beta$ & $\hat\sigma^2$ [or s^2],
employ the Cramer-Rao Thm
and calculate the information matrix, $\mathcal{I}$.

Recall, $\mathcal{I} = -E\left[\frac{\partial \ln L}{\partial\theta\,\partial\theta'}\right]$

— from the partals, take the second order partials.

Computing the Cramer-Rao Lowerbound, $\mathcal{I}^{-1}$.

$$\frac{\partial^2 \ln L}{\partial \beta \, \partial \beta'} = -\frac{1}{\sigma^2} X'X$$

$$\left.\begin{array}{l} \dfrac{\partial^2 \ln L}{\partial \beta \, \partial \sigma^2} = \dfrac{1}{\sigma^4}\left[X'X\beta - X'Y\right] \\[2ex] \dfrac{\partial^2 \ln L}{\partial (\sigma^2)^2} = \dfrac{T}{2\sigma^4} - \dfrac{2}{2\sigma^6}\epsilon'\epsilon \end{array}\right\} \quad \mathcal{I} = -E\left[\text{these}\right]$$

$$-E\left[\frac{\partial^2 \ln L}{\partial \beta \, \partial \beta'}\right] = \frac{1}{\sigma^2} X'X$$

$$-E\left[\frac{\partial^2 \ln L}{\partial \beta \, \partial \sigma^2}\right] = \frac{1}{\sigma^4} X' E[Y - X\beta] = 0$$

$$-E\left[\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\right] = \frac{-T}{2\sigma^4} + \frac{T\sigma^2}{\sigma^6} = \frac{T}{2\sigma^4}$$

$$\Longrightarrow \quad \mathcal{I} = \begin{bmatrix} \frac{1}{\sigma^2}(X'X) & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

$$\mathcal{I}^{-1} = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{bmatrix} = \text{Cramer-Rao Lowerbound}$$

* Since $\mathrm{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$, $\hat{\beta}$ is <u>efficient</u>.

$\hat{\sigma}^2$ is biased;
$\mathrm{Var}(s^2) = \dfrac{2\sigma^4}{T-K} > \dfrac{2\sigma^4}{T}$; $\therefore$ need more to show efficiency.

50
10 ×

Use Blackwell - Rao Thm:
Show that s^2 is a function of a sufficient statistic.

Lemma: $\epsilon'\epsilon = e'e + (\hat\beta-\beta)'X'X(\hat\beta-\beta)$

Proof: $\epsilon'\epsilon = (Y-X\beta)'(Y-X\beta)$

$= Y'Y - 2\beta'X'Y + \beta'X'X\beta$

→ add & subtract $(-2\hat\beta'X'Y + \hat\beta'X'X\hat\beta)$

$= Y'Y - 2\hat\beta'X'Y + \hat\beta'X'X\hat\beta + 2\hat\beta'X'Y - 2\beta'X'Y +$
$+ \beta'X'X\beta - \hat\beta'X'X\hat\beta$

→ but the first three terms $= (Y-X\hat\beta)'(Y-X\hat\beta) = e'e$

$= e'e + 2\hat\beta'X'Y - 2\beta'X'Y + \beta'X'X\beta - \hat\beta'X'X\hat\beta$

→ substitute $X'Y = X'X\hat\beta$

$= e'e + 2\hat\beta'X'X\hat\beta - 2\beta'X'X\hat\beta + \beta'X'X\beta - \hat\beta'X'X\hat\beta$
$= e'e + \hat\beta'X'X\hat\beta - 2\beta'X'X\hat\beta + \beta'X'X\beta$

→ note that $\beta'X'X\hat\beta = \hat\beta'X'X\beta$ is a scalar

$= e'e + \hat\beta'X'X\hat\beta - \hat\beta'X'X\beta - \beta'X'X\hat\beta + \beta'X'X\beta$

→ last 4 terms can be factored

$= e'e + (\hat\beta-\beta)'X'X(\hat\beta-\beta)$

Implications:

1. This holds for any β.
2. $(\hat\beta-\beta)'X'X(\hat\beta-\beta)$ is psd.
3. Since 1 & 2 are true, $\beta = \hat\beta$ is a global minimum of SSE.

④ $$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\,\epsilon'\epsilon$$

$$= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left[e'e + (\hat\beta-\beta)'X'X(\hat\beta-\beta)\right]$$

$$= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left[SSE + (\hat\beta-\beta)'X'X(\hat\beta-\beta)\right]$$

$$= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left[(T-K)\,s^2 + (\hat\beta-\beta)'X'X(\hat\beta-\beta)\right]$$

$$= g(t;\theta)\cdot h(y_1, y_2, \ldots, y_T)$$

where $t = \begin{bmatrix}\hat\beta \\ s^2\end{bmatrix}$, $\theta = \begin{bmatrix}\beta \\ \sigma^2\end{bmatrix}$,

and where $g(t;\theta)$ does not depend on the y_t's except through t, and $h(y_1,\ldots,y_T)$ does not depend on θ.

$\Longrightarrow$ t is a <u>sufficient statistic</u> for θ;

& by the Blackwell-Rao Thm, since $\hat\beta$ & s^2 are <u>unbiased</u> & <u>sufficient statistics</u>, they are <u>efficient</u> for β and σ^2.

* (i) Already knew $\hat\beta$ efficient; now know s^2 is.

* (ii) $\hat\sigma^2 = \frac{SSE}{T}$ is also a sufficient statistic for σ^2, but is biased, and therefore not efficient.

* (iii) with small samples. $\hat\sigma^2$ understates σ^2 → false security!

Summary:

$\hat{\beta}$ is unbiased, BLUE, consistent, asymptotically efficient, and effieient

s^2 is unbiased, consistent, asymptotically efficient, and efficient.

Note: In general, efficiency $\not\Leftrightarrow$ asymptotic efficiency. There may be some estimators which are not efficient with small samples (perhaps due to bias), but with second moments that become small "very quickly" as $T \to \infty$.

Thm: The asymptotic distributions of $\hat{\beta}$ & s^2 are:

$$\sqrt{T}(\hat{\beta} - \beta) \to N\left(0, \sigma^2 \lim_{T\to\infty}\left(\frac{X'X}{T}\right)^{-1}\right)$$

$$\sqrt{T}(s^2 - \sigma^2) \to N(0, 2\sigma^4)$$

Proof: If $\hat{\theta}$ is MLE, of θ,

$$\sqrt{T}(\hat{\theta} - \theta) \to N\left(0, \lim_{T\to\infty}\left(\frac{\mathcal{I}}{T}\right)^{-1}\right);$$

$$\theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}; \quad \hat{\theta}_{MLE} = \begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix}; \quad \mathcal{I} = \begin{bmatrix} \frac{1}{\sigma^2}(X'X) & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

s^2 and $\hat{\sigma}^2$

Their limits are equal. so

(s^2)

$$\left(\frac{\mathcal{I}}{T}\right)^{-1} = \begin{bmatrix} \sigma^2\left(\frac{X'X}{T}\right)^{-1} & 0 \\ 0 & 2\sigma^4 \end{bmatrix}$$

[illegible]

QED

Very Important Thm:

The (small sample) distributions of $\hat{\beta}$ & s^2 are –

(i) $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$

(ii) $\dfrac{(T-K)s^2}{\sigma^2} \sim \chi^2_{T-K} \qquad \Rightarrow s^2 \sim \left(\dfrac{\sigma^2}{T-K}\right)\chi^2_{T-K}$

(iii) $\hat{\beta}$ & s^2 are independent.

Proof:

First (i); already proved unbiasedness of $\hat{\beta}$;
already proved $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$;
$\hat{\beta}$ is Normal,
since $(X'X)^{-1}X'Y$ is a l.c. of the ϵ's,
which are Normal.

QED(i)

To prove (ii) & (iii) we need some Lemmas.

Lemma ①: If $\epsilon \sim N(0, \sigma^2 I)$, and B is a symmetric, idempotent matrix, then

$$\frac{\epsilon' B \epsilon}{\sigma^2} \sim \chi^2_{Trace(B)}$$

(Proved in Schmidt)
p.11

To prove (ii);

$$s^2 = \frac{SSE}{T-K} = \frac{\epsilon' M \epsilon}{T-K}$$

where $M = I_T - X(X'X)^{-1}X'$ is symmetric & idempotent, and Trace $(M) = T-K$.

$$\Longrightarrow \frac{(T-K)s^2}{\sigma^2} = \frac{\epsilon' M \epsilon}{\sigma^2} \sim \chi^2_{T-K}$$

by Lemma (1).

QED(ii)

<u>Lemma (2):</u> Let ϵ & B be as in Lemma (1), and C be any matrix $\ni$ CB is defined. Then $C\epsilon$ and $\epsilon' B \epsilon$ are independent iff $CB = 0$.

(also proved in Schmidt p.12)

To prove (iii);

consider $\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$

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

Let $C = (X'X)^{-1}X'$ and $B = M\ [= I_T - X(X'X)^{-1}X']$.

Noting that $(X'X)^{-1}X'M = 0$,

$(X'X)^{-1}X'\epsilon$ is independent of $\epsilon' M \epsilon$ by Lemma (2).

i.e. $(\hat{\beta} - \beta)$ is independent of $SSE = (T-K)s^2$;

i.e. $\hat{\beta}$ is independent of s^2. QED (iii)

Intuition behind (iii);

Even though $s^2 = \frac{1}{T-K}(Y - X\underline{\hat{\beta}})'(Y - X\underline{\hat{\beta}})$
(<u>depends</u> <u>on</u> $\hat{\beta}$)

there are ϵ's in $\hat{\beta}$ and in Y,
and they cancel in such a way that
$\hat{\beta}$ & s^2 are indeed independent. (as just proven)

Alternatively put;
estimating $\hat{\beta}$ uses K d.f. ;
estimating s^2 uses T-K d.f. ;
and the d.f. that each depends on are exclusive
of those that the other depends on.
Hence, independence.

Note; no other estimates from this
distribution will be independent
of both $\hat{\beta}$ & s^2.

This is a gory proof!
— much worse w/o matrix notation.

With the small-sample distributions of $\hat{\beta}$ & s^2,
we can test hypotheses.

Tests of Hypotheses

A. Definitions

<u>Defn:</u> Let a be an arbitrary $K \times 1$ vector. of numbers

Then $s_a^2 = s^2 a'(X'X)^{-1}a$.

$\mathrm{Var}(a'\hat{\beta})$

<u>Defn:</u> A statistic is distributed t_{T-K} if it can be expressed as the ratio of two statistics with:

(i) the numerator $\sim N(0,1)$

(ii) the denominator $\sim \sqrt{\left(\frac{1}{T-K}\right)\chi^2_{T-K}}$

(iii) the numerator & denominator independent.

<u>Defn:</u> A statistic is distributed $F_{m,\,T-K}$ if it can be expressed as the ratio of two statistics with:

(i) the numerator $\sim \left(\frac{1}{m}\right)\chi^2_m$

(ii) the denominator $\sim \left(\frac{1}{T-K}\right)\chi^2_{T-K}$

(iii) the numerator & denominator independent.

B. Using the t distribution

Thm: $\dfrac{a'(\hat{\beta}-\beta)}{s_a} \sim t_{T-K}$

Proof: must satisfy the 3 conditions.

First consider the numerator, $a'(\hat{\beta}-\beta)$.
This is Normal – ($\hat{\beta}$ is).
The mean is zero – ($E(\hat{\beta}) = \beta$).

$$
\begin{aligned}
Var[a'(\hat{\beta}-\beta)] &= E[a'(\hat{\beta}-\beta)]^2 \\
&= E[a'(\hat{\beta}-\beta)][a'(\hat{\beta}-\beta)]' \\
&= E[a'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'a] \\
&= a'E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'a \qquad \text{(a fixed)} \\
&= a'[\sigma^2(X'X)^{-1}]a \\
&= \sigma^2 a'(X'X)^{-1}a \equiv \sigma_a^2
\end{aligned}
$$

(See Defn for s_a^2)

$$\therefore \frac{a'(\hat{\beta}-\beta)}{\sigma_a} \sim N(0,1)$$

[Since we don't know σ_a^2 we'll estimate it with s_a^2]

Second, consider as the denominator,

$$\frac{s_a^2}{\sigma_a^2} = \frac{s^2 a'(X'X)^{-1}a}{\sigma^2 a'(X'X)^{-1}a} = \frac{s^2}{\sigma^2}$$

$$\therefore \frac{(T-K)\,s_a^2}{\sigma_a^2} \sim \chi^2_{T-K}$$

[Since we already proved that $\dfrac{(T-K)s^2}{\sigma^2} \sim \chi^2_{T-K}$]

55

Putting the numerator & denominator together,

$$\frac{a'(\hat{\beta}-\beta)}{s_a} = \frac{\dfrac{a'(\hat{\beta}-\beta)}{\sigma_a}}{\sqrt{\dfrac{s_a^2}{\sigma_a^2}}} \quad \frac{\}\sim N(0,1)}{\}\sim\sqrt{\dfrac{\chi^2_{T-K}}{T-K}}}$$

Third, we previously proved that
$(\hat{\beta}-\beta)$ and $(T-K)\,s^2$ are independent.
This implies that
$a'(\hat{\beta}-\beta)$ and s_a are independent.

Hence, all conditions are satisfied, and

$$\frac{a'(\hat{\beta}-\beta)}{s_a} \sim t_{T-K}.$$

Thm: Test of a single linear restriction on β.

Let a be a known $K\times 1$ vector,
b be a known scalar.
The linear restriction considered is

$$a'\beta = b\ ; \quad a_1\beta_1 + a_2\beta_2 + \cdots + a_K\beta_K = b.$$

Then under H_0: $a'\beta = b$, $\dfrac{a'\hat{\beta}-b}{s_a} \sim t_{T-K}$.

Proof: $$\frac{a'\hat{\beta}-b}{s_a} \underset{\substack{\uparrow \\ \text{(Under } H_0)}}{=} \frac{a'\hat{\beta}-a'\beta}{s_a} = \frac{a'(\hat{\beta}-\beta)}{s_a} \underset{\substack{\uparrow \\ \text{(just proved)}}}{\sim} t_{T-K}.$$

Intuition:

$$a'(\hat{\beta}-\beta) = [a_1 \ a_2 \ \dots \ a_K] \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \\ \vdots \\ \hat{\beta}_K - \beta_K \end{bmatrix}$$

$$= a_1(\hat{\beta}_1 - \beta_1) + a_2(\hat{\beta}_2 - \beta_2) + \dots + a_K(\hat{\beta}_K - \beta_K)$$

This is a scalar, a linear function of the differences between our best estimates of the parameters and their true values.

a is arbitrary; it could be any test of any single coefficient estimate, or a combination of coefficient estimates.

This test statistic makes sense!
It is the difference between our predicted value, $a'\hat{\beta}$, and the true value under H_0, $a'\beta = b$,
divided by the standard error of the linear combination of estimated parameters, s_a [= our estimate of $\sqrt{Var(a'(\hat{\beta}-\beta))}$]

This is the statistic achieved in both, the Likelihood Ratio Test and the Wald Test.

Explain with examples.

<u>Example 1;</u> $H_0: \beta_i = 0$

or, the hard way, $a'\beta = 0$

or $$\begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} = 0$$

[i^{th} position]

In this case, $$\begin{aligned} s_a^2 &= s^2 a'(X'X)^{-1} a \\ &= s^2 (X'X)^{-1}_{ii} \\ &= s^2 * \; i^{th} \text{ element on diagonal of } (X'X)^{-1} \\ &= \text{variance estimate of } \hat{\beta}_i \end{aligned}$$

Then, the test statistic is

$$\frac{a'\hat{\beta} - b}{\sqrt{s^2 a'(X'X)^{-1} a}} = \frac{\hat{\beta}_i}{\sqrt{s^2 (X'X)^{-1}_{ii}}} = \frac{\text{estimate of } \beta_i}{\text{its std error}}$$

This is usually given for each parameter estimate.

Example 2: $H_0: \quad \beta_2 + \beta_4 = -6$

or, the hard way, $a = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$ and $b = -6$;

Then

$$a'\beta = [0\,1\,0\,1\,0\,0 \cdots] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} = \beta_2 + \beta_4$$

In this case,

$$s_a^2 = s^2 a'(X'X)^{-1} a = s^2 [0\,1\,0\,1\,0\,0 \cdots] \begin{bmatrix} & & \\ & (X'X)^{-1} & \\ & & \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

$$= s^2 (X'X)^{-1}_{22} + s^2 (X'X)^{-1}_{44} + 2 s^2 (X'X)^{-1}_{24}$$

$$= Var(\hat\beta_2) + Var(\hat\beta_4) + 2\, Cov(\hat\beta_2, \hat\beta_4)$$

$$= Var(\hat\beta_2 + \hat\beta_4)$$

Then the test statistic is

$$\frac{a'\hat\beta - b}{s_a} = \frac{\hat\beta_2 + \hat\beta_4 + 6}{\sqrt{\underbrace{s^2(X'X)^{-1}_{22}}_{Var(\beta_2)} + \underbrace{s^2(X'X)^{-1}_{44}}_{Var(\beta_4)} + \underbrace{2 s^2 (X'X)^{-1}_{24}}_{2Cov(\beta_2, \beta_4)}}} \sim t_{T-K}$$

[e.g. may be estimating Cobb-Douglas p.f., and want to test the Returns-to-Scale $(\alpha + \beta \gtreqless 1)$

C. Using the F distribution

The t-distribution is used to test any *single* linear restriction on β.

The F-distribution is used to test *several* linear restrictions on β.

Thm: Test of *several* linear restrictions on β.

Let R be a known $m \times K$ matrix,
r be a known $m \times 1$ vector,

Then under H_0: $R\beta = r$,

$$\frac{(R\hat{\beta}-r)'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta}-r)\,/\,m}{SSE\,/\,T-K} \sim F_{m,\,T-K}$$

Note: $\underset{m\times K}{R}\,\underset{K\times 1}{\beta} = \underset{m\times 1}{r}$ is m linear restrictions

of the form; $R_{i1}\beta_1 + R_{i2}\beta_2 + \cdots + R_{iK}\beta_K = r_i \qquad i = 1, \ldots, m$

Proof: must satisfy 3 conditions in defn of $F_{m,T-K}$.

First consider the denominator, $\frac{SSE}{\sigma^2} \sim \chi^2_{T-K}$ (already proved)

$$\Rightarrow \frac{SSE}{T-K} \sim \left(\frac{\sigma^2}{T-K}\right)\chi^2_{T-K}$$

QED (ii)

First consider the numerator.

$$R\hat{\beta} - r \underset{\text{(under } H_0)}{=} R\hat{\beta} - R\beta = R(\hat{\beta}-\beta) = R(X'X)^{-1}X'\epsilon$$

$$\Longrightarrow \text{numerator} = [R(X'X)^{-1}X'\epsilon]'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'\epsilon \;/m$$
$$= \epsilon'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'\epsilon \;/m$$
$$= \frac{\epsilon' Q \epsilon}{m}$$

where $Q = X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'$ is symmetric and idempotent,

and
$$\text{Trace}(Q) = Tr\{[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X(X'X)^{-1}R'\}$$
$$= Tr\{[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'\}$$
$$= Tr(I_m)$$
$$= m$$

Then by Lemma 1 stated previously,

$$\frac{\epsilon'Q\epsilon}{\sigma^2} \sim \chi^2_m$$

$$\Longrightarrow \text{numerator} = \frac{\epsilon'Q\epsilon}{m} \sim \sigma^2 * \left(\frac{\chi^2_m}{m}\right)$$

QED (i)

Second consider the denominator.

$$\frac{SSE}{\sigma^2} \sim \chi^2_{T-K} \qquad \text{(already proven)}$$

$$\Longrightarrow \text{denominator} = \frac{SSE}{T-K} \sim \sigma^2 * \left(\frac{\chi^2_{T-K}}{T-K}\right)$$

QED (ii)

To prove the third condition for an F we need another Lemma.

<u>Lemma ③ :</u> If $\epsilon \sim N(0, \sigma^2 I_T)$, and Q & M are symmetric, idempotent matrices, then $\epsilon' Q \epsilon$ and $\epsilon' M \epsilon$ are <u>independent</u> iff $MQ = 0$.

(also proved in Schmidt)

Note : (a) $\frac{\epsilon' Q \epsilon}{m}$ = numerator ;
(b) $\frac{\epsilon' M \epsilon}{T-K} = \frac{SSE}{T-K}$ = denominator ;
(c) $MQ = 0$.

QED(iii)

Thus we have our F-statistic,

$$\frac{\text{numerator}}{\text{denominator}} \Rightarrow \frac{(i)\ \sigma^2 \frac{\chi^2_m}{m}}{(ii)\ \sigma^2 \frac{\chi^2_{T-K}}{T-K}} \sim F_{m, T-K}$$

(iii) numerator & denominator independent

QED

Discussion : This is an upper tail test.
Since the numerator of the test statistic is of the form,
$(R\hat{\beta}-r)' [\text{psd matrix}] (R\hat{\beta}-r)$;
this is a quadratic form in $(R\hat{\beta}-r)$ $\rightarrow$ the measure of how far our estimate is from that hypothesized.
If $H_0: R\beta = r$ is not true. $R\hat{\beta}-r$ should be big. & F big !

* Example 1: H_0: $\left.\begin{array}{l}\beta_1 + \beta_2 = 4 \\ \beta_3 = 0\end{array}\right\} R\beta = r$

$m = 2$ restrictions

$$\underset{2\times K}{R = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \end{bmatrix}} \qquad \underset{2\times 1}{r = \begin{bmatrix} 4 \\ 0 \end{bmatrix}}$$

We have sample data; X, Y.
We have $\hat{\beta} = (X'X)^{-1}X'Y$.

Plug into the test statistic! Can do!

Another way to determine the test statistic:

General Statement:

The test statistic,

$$F_{m,T-K} \Rightarrow \frac{(R\hat{\beta}-r)'\left[R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta}-r)/m}{SSE/T-K} \quad \text{can be written as} \quad \frac{(SSE_R - SSE_u)/m}{SSE_u/T-K}.$$

write on board & ...

where 2 regressions are run — one unrestricted, and one restricted (with the restrictions, $R\beta = r$, imposed). SSE_R & SSE_u are the rest. & unrest. S.S.

In this example;

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \cdots \\ &= \beta_1 x_1 + (4-\beta_1)x_2 + \beta_4 x_4 + \cdots \\ &= 4x_2 + \beta_1(x_1 - x_2) + \beta_4 x_4 + \cdots \end{aligned}$$

Run this regression;

$(y - 4x_2) = \beta_1(x_1 - x_2) + \beta_4 x_4 + \cdots$ to get SSE_R.

Example 2: Test of equality restriction on β.

Let β^* be a known vector.

$H_0: \beta = \beta^* \quad \Longrightarrow \quad R\beta = r$ where $R = I_K$ and $r = \beta^*$.

Then
$$\frac{(R\hat\beta - r)'[R(X'X)^{-1}R']^{-1}(R\hat\beta - r)/m}{SSE/T-K} = \frac{(\hat\beta-\beta^*)'[(X'X)^{-1}]^{-1}(\hat\beta-\beta^*)/K}{s^2}$$

$$= (\hat\beta-\beta^*)'[s^2(X'X)^{-1}]^{-1}(\hat\beta-\beta^*)/K$$

$$\sim F_{K,T-K}$$

[This is intuitively nice, since it is analagous to the case of a single linear restriction;

$$\frac{\hat\beta_i - \beta_i^*}{\sqrt{s^2(X'X)^{-1}_{ii}}} \quad .$$
]

* Example 3: Test of existence of a relationship. (developed before)

$\boxed{H_0: \beta = 0} \rightarrow \boxed{\text{almost}}$ Special case of ex. 2 with $\beta_2^* = \beta_3^* = \cdots = 0$.
(letting the first regressor be the constant term)

test $H_0: \beta_2 = \beta_3 = \cdots = \beta_K = 0$. $\quad (K-1) = m = \#$ restrictio[ns]

Applying the General Statement;

$SSE_R = SSE$ on regression with no regressors $= SST$!

Test Statistic;
$$\frac{(SST - SSE_K)/(K-1)}{SSE_K/T-K}$$

$$= \frac{SSR/K-1}{SSE/T-K} = \frac{R^2\ \cancel{SST}/K-1}{(1-R^2)\ \cancel{SST}/T-K} = \left(\frac{T-K}{K-1}\right)\frac{R^2}{1-R^2} \sim F_{K-1,T-K}$$

☆ This is easy to calculate !! This is always given

DIGRESSION on Partitioned Inverse Rule

Suppose that a square matrix, Z, can be partitioned as follows.

$$Z = \begin{bmatrix} E & F \\ G & H \end{bmatrix} \quad \text{Then} \quad Z^{-1} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} E & F \\ G & H \end{bmatrix}\begin{bmatrix} a & c \\ b & d \end{bmatrix} = I.$$

Given $E, F, G, \& H$, find $a, b, c,$ and d.

The following 4 statements hold.

① $Ea + Fb = I$

② $Ec + Fd = 0 \implies$ from ②: $Ec = -Fd$

③ $Ga + Hb = 0$ $\quad c = -E^{-1}Fd$ (if E nonsingular)

④ $Gc + Hd = I$

Substitute into ④: $G(-E^{-1}Fd) + Hd = I$

$$\underbrace{(H - GE^{-1}F)}_{D} d = I$$

$$Dd = I$$

$$\boxed{\begin{aligned} d &= D^{-1}I = D^{-1} \\ c &= -E^{-1}FD^{-1} \end{aligned}}$$

?

From ①: $Ea = I - Fb$

$$a = E^{-1}(I - Fb)$$

Substitute into ③: $GE^{-1}(I - Fb) + Hb = 0$

$$GE^{-1} - GE^{-1}Fb + Hb = 0$$

$$\underbrace{(H - GE^{-1}F)}_{D} b = -GE^{-1}$$

$$Db = -GE^{-1}$$

$$\boxed{\begin{aligned} b &= -D^{-1}GE^{-1} \\ a &= E^{-1}(I + FD^{-1}GE^{-1}) \end{aligned}}$$

$$\therefore \quad Z^{-1} = \begin{bmatrix} E^{-1} + E^{-1}FD^{-1}GE^{-1} & -E^{-1}FD^{-1} \\ -D^{-1}GE^{-1} & D^{-1} \end{bmatrix}$$

[$\ast$ put $\mathbb{D}^{-1}$ in top left corner.]

Alternatively, from ③: $\quad Hb = -Ga$

$$b = -H^{-1}Ga \qquad \text{(if } H \text{ nonsingular)}$$

Substitute into ①:

$$Ea + F(-H^{-1}Ga) = I$$
$$\underbrace{(E - FH^{-1}G)}a = I$$
$$\mathbb{D}a = I$$
$$\boxed{\begin{aligned} a &= \mathbb{D}^{-1} \\ b &= -H^{-1}G\mathbb{D}^{-1} \end{aligned}}$$

From ④:

$$Hd = I - Gc$$
$$d = H^{-1}(I - Gc)$$

Substitute into ②:

$$Ec + FH^{-1}(I - Gc) = 0$$
$$Ec + FH^{-1} - FH^{-1}Gc = 0$$
$$\underbrace{(E - FH^{-1}G)}c = -FH^{-1}$$
$$\mathbb{D}c = -FH^{-1}$$
$$c = -\mathbb{D}^{-1}FH^{-1}$$
$$d = H^{-1}(I + G\mathbb{D}^{-1}FH^{-1})$$

And Z^{-1} can also be written:

$$Z^{-1} = \begin{bmatrix} \mathbb{D}^{-1} & -\mathbb{D}^{-1}FH^{-1} \\ -H^{-1}G\mathbb{D}^{-1} & H^{-1} + H^{-1}G\mathbb{D}^{-1}FH^{-1} \end{bmatrix}$$

Thus $$Z^{-1} = \begin{bmatrix} E^{-1} + E^{-1}FD^{-1}GE^{-1} & -E^{-1}FD^{-1} \\ -D^{-1}GE^{-1} & D^{-1} \end{bmatrix} = \begin{bmatrix} \mathbb{D}^{-1} & -\mathbb{D}^{-1}FH^{-1} \\ -H^{-1}G\mathbb{D}^{-1} & H^{-1} + H^{-1}G\mathbb{D}^{-1}FH^{-1} \end{bmatrix}$$

where $D = H - GE^{-1}F$ and $\mathbb{D} = E - FH^{-1}G$

Example 4: Test of equality restriction on part of β.

Partition β — $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \begin{array}{l} (K-H)\times 1 \\ H \times 1 \end{array}$ $\qquad H_0: \beta_2 = \beta_2^*$

Restrictions — $R\beta = r$

where $R = [\underset{H\times(K-H)}{0}, I_H]$; $r = \beta_2^*$

$$R\beta \Rightarrow \Bigg[\underbrace{\begin{matrix} 0 & 0 & \cdots \\ 0 & 0 & \cdots \\ \vdots & & \\ 0 & 0 & \cdots \end{matrix}}_{H\times(K-H)} \;\; \underbrace{\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & 1 \end{matrix}}_{H\times H}\Bigg] \begin{bmatrix} \left.\begin{matrix}\beta_1 \\ \vdots \\ \beta_{K-H}\end{matrix}\right\} \beta_1 \\ \left.\begin{matrix}\beta_{K-H+1} \\ \vdots \\ \beta_K\end{matrix}\right\} \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_{K-H+1} \\ \vdots \\ \beta_K \end{bmatrix} = \beta_2$$

$m = H = \#$ of restrictions

Substitute this into the test statistic:

$$\frac{(R\hat\beta - r)'[R(X'X)^{-1}R']^{-1}(R\hat\beta - r)/m}{SSE/T-K} = \frac{(\hat\beta_2 - \beta_2^*)'[(0\ I_H)(X'X)^{-1}\binom{0}{I_H}]^{-1}(\hat\beta_2 - \beta_2^*)/}{s^2}$$

$$= \frac{(\hat\beta_2 - \beta_2^*)' D (\hat\beta_2 - \beta_2^*)/H}{s^2} \sim F_{H,T-K}$$

where $\underset{T\times K}{X} = [\underset{T\times(K-H)}{X_1}, \underset{T\times H}{X_2}]$ $\leftarrow$ (partition X)

$(X_1 \rightarrow \beta_1 \ ; \ X_2 \rightarrow \beta_2)$

$M_1 = I_T - X_1(X_1'X_1)^{-1}X_1'$

$D = X_2'M_1X_2$

by Partitioned Inverse Rule.

To see that D is appropriate, apply the Partition Inverse Rule.

In our case,

$$Z = X'X = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} [X_1 \;\; X_2] = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

and $Z^{-1} = (X'X)^{-1}$

$$= \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} E^{-1} + E^{-1}FD^{-1}GE^{-1} & -E^{-1}FD^{-1} \\ -D^{-1}GE^{-1} & D^{-1} \end{bmatrix}$$

where $D = H - GE^{-1}F$

$$= X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2$$

$$= \boxed{X_2'M_1X_2}$$

with $M_1 = I_T - X_1(X_1'X_1)^{-1}X_1'$

In this example,

$$[R(X'X)^{-1}R']^{-1} = \left[(0 \;\; I_H)(X'X)^{-1}\begin{pmatrix} 0 \\ I_H \end{pmatrix}\right]^{-1}$$

$$= [\text{bottom right partition of } (X'X)^{-1}]^{-1}$$

$$= [D^{-1}]^{-1} = D.$$

Hence, the statistic is

$$\frac{(\hat{\beta}_2 - \beta_2^*)' D (\hat{\beta}_2 - \beta_2^*) / H}{s^2} \sim F_{H,\,T-K}.$$

Example 6: Chow Test

- Test of equality of 2 regressions
- (Analysis of Covariance) ⇐ Suits mod

$$Y_1 = X_1\beta_1 + \epsilon_1 \quad ; \quad T_1 \text{ obs.} \qquad \beta_1 \to K\times 1 \qquad X_1 \to T_1\times K$$
$$Y_2 = X_2\beta_2 + \epsilon_2 \quad ; \quad T_2 \text{ obs.} \qquad \beta_2 \to K\times 1 \qquad X_2 \to T_2\times K$$

$H_0: \ \beta_1 = \beta_2 = \beta$ Assumption: $\mathrm{Cov}\begin{pmatrix}\epsilon_1\\ \epsilon_2\end{pmatrix} = \sigma^2 I_{T_1+T_2}$

i.e. all $T_1 + T_2$ disturbances have equal variances & are indep.

The joint equation: $\begin{pmatrix}Y_1\\ Y_2\end{pmatrix} = \begin{pmatrix}X_1\\ X_2\end{pmatrix}\beta + \begin{pmatrix}\epsilon_1\\ \epsilon_2\end{pmatrix}$

or $Y = X\beta + \epsilon$; $T = T_1 + T_2$ obs

Under H_0, this is what we'd do.

$SSE_{unrestricted} = SSE_{T_1} + SSE_{T_2}$ (don't impose restriction)

$SSE_{restricted} = SSE_T$ (impose restriction)

The test statistic is

$$F_{K,T-2K} \Rightarrow \frac{[SSE_T - (SSE_{T_1} + SSE_{T_2})]/K}{(SSE_{T_1} + SSE_{T_2})/T-2K}$$

degrees of freedom:

$\to (T-K) - (T_1-K) - (T_2-K) = K$

$\to (T_1-K) + (T_2-K) = T-2K$

<u>examples of uses of Chow Test:</u>

(i) In a long time series regression where $\exists$ some possible structural change in model after certain event in time. — [e.g. war, energy crisis

(ii) Cross-sectional data that is grouped in some way. — e.g. Blacks – whites, males – females

Again the "General Statement" is important.

$$\frac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/m}{SSE/T-K}$$ is sometimes gory.

It is often easier to write the statistic as

$$\frac{(SSE_R - SSE_u)/m}{SSE_u/T-K}.$$

So consider how to impose the restrictions in H_0 [$\rightarrow SSE_R$] and do it!

This works for any kind of <u>linear</u> restrictions & ∴ for any of the above examples.

<u>Example 7</u>: Test of adequacy of the model. [how well does model <u>forecast</u>]

$Y = X\beta + \epsilon$; T obs. in original sample

Get $\hat{\beta} = (X'X)^{-1}X'Y$, & use it to <u>forecast</u> when new observations on X become available.

i.e. let X_0 be E new obs. on the explanatory var.'s. Then the <u>forecast</u> is

$$\hat{Y}_0 = X_0\hat{\beta}$$

~~Define: Forecast error~~ $\cancel{= v_0 = Y_0 - \hat{Y}_0 = (X_0\beta + \epsilon_0) - X_0\hat{\beta}}$

$\cancel{= (X_0\beta_0 + \epsilon_0) - X_0(X'X)^{-1}X'(X\beta+\epsilon)}$

Define: $\quad$ forecast error $= v_0 = Y_0 - \hat{Y}_0$

$$\begin{aligned}
&= (X_0\beta + \epsilon_0) - X_0\hat{\beta} \\
&= X_0\beta + \epsilon_0 - X_0(X'X)^{-1}X'Y \\
&= \cancel{X_0\beta} + \epsilon_0 - X_0(X'X)^{-1}X'(\cancel{X\beta} + \epsilon) \\
&= \epsilon_0 - X_0(X'X)^{-1}X'\epsilon
\end{aligned}$$

This is what we're interested in;
To make statements about how well the model forecasts
we must <u>test</u> <u>hypotheses</u> about v_0. (Is error "significant"?)
$\longrightarrow$ we want confidence bands around v_0.
To do this we need the <u>distribution of v_0.</u>

$$E(v_0) = E(\epsilon_0 - X_0(X'X)^{-1}X'\epsilon) = 0$$

v_0 is distributed Normally; a l.c. of ϵ's.

$$\begin{aligned}
\mathrm{Cov}(v_0) &= E(v_0 v_0') \\
&= E[\epsilon_0 - X_0(X'X)^{-1}X'\epsilon][\epsilon_0 - X_0(X'X)^{-1}X'\epsilon]' \\
&= E[\epsilon_0\epsilon_0' + X_0(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}X_0' \\
&\qquad - \epsilon_0\epsilon'X(X'X)^{-1}X_0' - X_0(X'X)^{-1}X'\epsilon\epsilon_0'] \\
&= \sigma^2 I_E + \sigma^2 X_0(X'X)^{-1}X_0' - \underbrace{E(\epsilon_0\epsilon')}_{(=0)}X(X'X)^{-1}X_0' - X_0(X'X)^{-1}X'\underbrace{E(\epsilon\epsilon_0')}_{(=0)} \\
&= \sigma^2 I_E + \sigma^2 X_0(X'X)^{-1}X_0'
\end{aligned}$$

since ϵ & ϵ_0 independent
(big assumption possibly ∄)

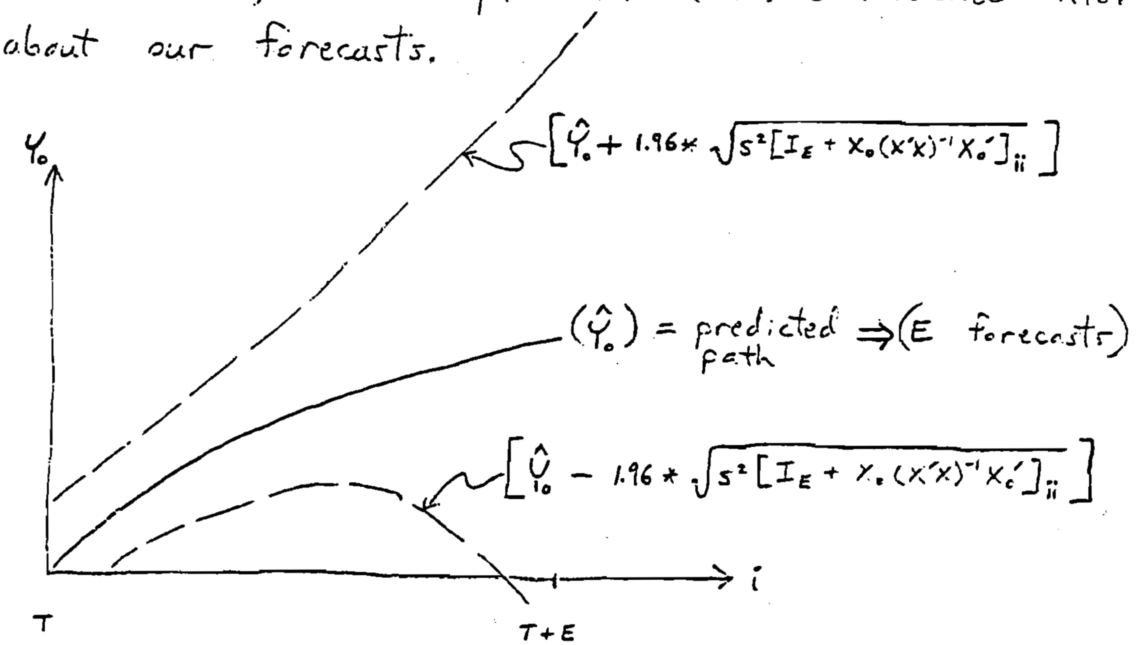
$$\Longrightarrow \quad v_0 \sim N\{0, \sigma^2[I_E + X_0(X'X)^{-1}X_0']\}$$

Under H_0: the model is adequate,

$$\frac{(v_0)_i}{\sqrt{s^2\left[I_E + X_0(X'X)^{-1}X_0'\right]_{ii}}} = \frac{(Y_0 - \hat{Y}_0)_i}{\left(\text{standard error wrt new data}\right)_i} \sim t_{T-K}$$

$$i = 1, \ldots, E .$$

Furthermore, we can plot the (95%) confidence interval about our forecasts.



Observe that the confidence interval becomes wider (quadratically) as the forecast is made farther past T. (i.e. farther past the point at which the model is estimated)

--> We have less confidence in our forecasts, the farther out the future we predict

---

* Construction of Q allows us to run just one regression, and obtain everything we want. — very useful.