Multi-variate Statistics

Extension of Bi-variate Statistics

$(Y, \boldsymbol{X})$~ random variables

where

$\boldsymbol{X}$~ vectors of K random variables

$$\boldsymbol{X} = [X_1, X_2, \ldots, X_K]$$

Y ~ a single random variable

Multi-variate Analyses

- Pair-wise Covariance or Correlation
- Multi-way ANOVA
- Multiple Regression

Multiple Regression Analysis

Focus on the dependency of Y on the **X** vector, e.g.,

$$\mu_{Y|X} = m(X_1, X_2, \ldots, X_K) = m(X)$$

$$\sigma^2_{Y|X} = v(X_1, X_2, \ldots, X_K) = v(X)$$

X_k - explanatory or independent variable, $k = 1,\ldots,K$

Y - dependent variable

Multiple Linear Regression

Assumptions

1) linearity $\mu_{Y|X} = X\boldsymbol{\beta}$

where $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \ldots \quad \beta_K]^{\mathrm{T}}$ are unknown parameters

2) variance-independent or $\sigma^2_{Y|X} = \sigma^2$

3) normality, i.e. $Y \mid X \sim N(X\boldsymbol{\beta}, \sigma^2)$

CLNRM (1)

Classical **L**inear **N**ormal **R**egression **M**odel is based upon the assumptions

$$Y_i = \boldsymbol{X}_i\beta + \varepsilon_i$$

where i = index of the observation

ε_i = identical and independent normal error term

$\varepsilon_i \sim \text{N}(0, \sigma^2)$ for all $i = 1, \ldots, n$

CLNRM (2)

$\boldsymbol{X}_i$ is pre-selected or non-random but Y_i or ε_i is randomly sampled.

$\boldsymbol{X}_i\beta$ is the non-random component of Y_i

ε_i is the random component of Y_i.

Note that X_1 can be intentionally set to one for all observations so that its coefficient β_1 becomes the y-intercept.

CLNRM
Matrix Representation (1)

Define

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{K1} \\ X_{12} & X_{22} & \dots & X_{K1} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{Kn} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

CLNRM
Matrix Representation (2)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim MVN\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right)$$

where

$\mathbf{0}$ is a nx1 column vector of zeroes

$\mathbf{I}_n$ is an nxn identity matrix.

CLNRM

Matrix Representation (3)

X is non-random. It is required that the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is invertible. Why?

Remember why we need $\sum_{i=1}^{n}(X_i - \overline{X})^2 > 0$ in Simple Linear Regression?

OLS Estimation for CLNRM (1)

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n}[Y_i - (X_{1i}\beta_1 + X_{2i}\beta_2 + \ldots + X_{Ki}\beta_K)]^2$$

or

$$\min_{\boldsymbol{\beta}} \left[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]^{\mathrm{T}}\left[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]$$

OLS Estimation for CLNRM (2)

First-Order Conditions

$$2\left[-\mathbf{X}\right]^{\mathrm{T}}\left[\mathbf{Y}-\mathbf{X\beta}\right]=\mathbf{0}$$

$$-\mathbf{X}^{\mathrm{T}}\mathbf{Y}+\mathbf{X}^{\mathrm{T}}\mathbf{X\beta}=\mathbf{0}$$

$$\hat{\boldsymbol{\beta}}=\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

OLS Estimation for CLNRM (3)

Estimator for σ^2

$$\hat{\sigma}^2=\frac{1}{n-K}\left[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right]^{\mathrm{T}}\left[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right]$$

$$=\frac{1}{n-K}\left[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}-\mathbf{Y}^{\mathrm{T}}\hat{\mathbf{Y}}\right]$$

where $\hat{\mathbf{Y}}=\mathbf{X}\hat{\boldsymbol{\beta}}$ is called the fitted value of $\mathbf{Y}$

Why *n-K*?

Properties of OLS estimators (1)

Theorem $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}$$

Does not require normality assumption.

Note that $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

Properties of OLS estimators (2)

Proof
$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}E(\mathbf{Y}) \\ &= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}E(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}) \\ &= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}[\mathbf{X}\boldsymbol{\beta}+E(\boldsymbol{\varepsilon})] \\ &= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} \end{aligned}$$

Properties of OLS estimators (3)

Proof
$$
\begin{aligned}
V(\hat{\boldsymbol{\beta}}) &= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}V(\mathbf{Y})\left[\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\right]^{T} \\
&= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}V(\mathbf{Y})\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1} \\
&= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}V(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1} \\
&= \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}V(\boldsymbol{\varepsilon})\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1} \\
&= \sigma^{2}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{I}_{n}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1} \\
&= \sigma^{2}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}
\end{aligned}
$$

Properties of OLS estimators (4)

Theorem Due to the normality assumption

of $\mathcal{E}$,
$$\hat{\boldsymbol{\beta}} \sim MVN\left(\boldsymbol{\beta}, \sigma^{2}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\right)$$

and
$$(n-K)\frac{\widehat{\sigma^{2}}}{\sigma^{2}} \sim \chi^{2}(n-K)$$

Properties of OLS estimators (5)

Variance-Covariance Matrix of $\hat{\boldsymbol{\beta}}$

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}$$

$$= \begin{bmatrix} V(\hat{\beta}_1) & C(\hat{\beta}_1, \hat{\beta}_2) & \cdots & C(\hat{\beta}_1, \hat{\beta}_K) \\ C(\hat{\beta}_2, \hat{\beta}_1) & V(\hat{\beta}_2) & \cdots & C(\hat{\beta}_2, \hat{\beta}_K) \\ \vdots & \vdots & \ddots & \vdots \\ C(\hat{\beta}_K, \hat{\beta}_1) & C(\hat{\beta}_K, \hat{\beta}_2) & \cdots & V(\hat{\beta}_K) \end{bmatrix}$$

σ^2 is generally unknown.

Properties of OLS estimators (6)

Estimated Variance-Covariance Matrix of $\hat{\boldsymbol{\beta}}$

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \widehat{\sigma^2} (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

$$= \begin{bmatrix} \hat{V}(\hat{\beta}_1) & \hat{C}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \hat{C}(\hat{\beta}_1, \hat{\beta}_K) \\ \hat{C}(\hat{\beta}_2, \hat{\beta}_1) & \hat{V}(\hat{\beta}_2) & \cdots & \hat{C}(\hat{\beta}_2, \hat{\beta}_K) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}(\hat{\beta}_K, \hat{\beta}_1) & \hat{C}(\hat{\beta}_K, \hat{\beta}_2) & \cdots & \hat{V}(\hat{\beta}_K) \end{bmatrix}$$

Properties of OLS estimators (7)

Standard Deviation of $\hat{\beta}_k$

$$sd(\hat{\beta}_k) = \sqrt{V(\hat{\beta}_k)}$$

Standard Error of $\hat{\beta}_k$

$$se(\hat{\beta}_k) = \sqrt{\hat{V}(\hat{\beta}_k)}$$

Properties of OLS estimators (8)

$$t_{cal} = \frac{\dfrac{\hat{\beta}_k - \beta_k}{sd(\hat{\beta}_k)}}{\sqrt{\dfrac{(n-K)\dfrac{\widehat{\sigma^2}}{\sigma^2}}{n-K}}}$$

$$= \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \sim t(n-K)$$

<<Basis for statistical inference>>

Central Limit Theorem (1)

Similar to that for the Simple Linear Regression Model. Even though the error terms are <u>not</u> normal, the properties of OLS estimators asymptotically hold when the sample size is very large.

Central Limit Theorem (2)

In mathematical term,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \overset{A}{\sim} \mathrm{MVN}\left(\mathbf{0}, \sigma^2 \left[\frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n}\right]^{-1}\right)$$

Gauss-Markov Theorem (1)

Similar to that for the Simple Linear Regression Model. Given that $\mathbf{X}$ is non-random, OLS estimator is Best Linear Unbiased Estimator.

Gauss-Markov Theorem (2)

$\hat{\boldsymbol{\beta}}$ is OLS estimator of β

$\tilde{\boldsymbol{\beta}}$ is a non-OLS linear unbiased estimator of β

$$\mathbf{h}\mathbf{V}(\hat{\boldsymbol{\beta}})\mathbf{h}^{T} \le \mathbf{h}\mathbf{V}(\tilde{\boldsymbol{\beta}})\mathbf{h}^{T}$$

for any vector $\mathbf{h} \neq \mathbf{0}$

Coefficient of Determination (1)

R^2 is a measure for goodness-of-fit. How well does the model fit the observed data? Low R^2 implies "bad" fit.

<u>Definition</u> $$R^2 \equiv 1 - \frac{SSR}{SST}$$

SSR = Sum of Squared Residuals

SST = Sum of Squared Totals

Coefficient of Determination (2)

where $$SSR = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = [\mathbf{Y} - \hat{\mathbf{Y}}]^{\mathrm{T}} [\mathbf{Y} - \hat{\mathbf{Y}}]$$

$$SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

Note that, in general, R^2 cannot be greater than one but could be negative.

Coefficient of Determination (3)

Low R^2 or a bad fit does <u>not</u> mean a bad model. It simply implies a large uncertainty in the nature. It is mainly used as a criterion to select various "candidate" models.

Coefficient of Determination (4)

If an X_i has constant value or a linear combination of X_i 's is equivalent to a constant value, then, $0 \le R^2 \le 1$ always

and
$$R^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$

Coefficient of Determination (5)

Interpretation if $0 \le R^2 \le 1$

1-R^2 or SSR/SST can be interpreted as the fraction of total variation of Y due to the random component (ε).

R^2 is generally regarded as the fraction of total variation of *Y* explained by the explanatory variables or due to the non-random component.

Adjusted- R^2 (1)

We can cheat on R^2 by adding more irrelevant independent variables on the right-hand side, especially when sample is small.

Higher *K* ==> smaller *SSR* ==>higher R^2

Adjusted- R^2 (2)

Definition

$$\bar{R}^2 \equiv 1 - \frac{SSR/(n-K)}{SST/(n-1)} = 1 - \frac{\widehat{\sigma^2}}{s_Y^2}$$

Concept

Penalize R^2 by dividing with (n-K)when an irrelevant variable is added.

Adjusted- R^2 (3)

Purpose

For a small sample, it is a better measure for goodness-of-fit than R^2. It is also used as criterion to add or remove an explanatory variable from the model if it does not contradict theories.

Statistical Inference about β_k

Confidence Interval for β_k

(1-α)100% CI for β_k = $\hat{\beta}_k \pm t_{\frac{\alpha}{2}}(n-K)se(\hat{\beta}_k)$

Hypothesis Testing for β_k

$H_0 : \beta_k = 0.6$

$H_1 : \beta_k \neq 0.6$

$$t_{cal} = \frac{\hat{\beta}_k - 0.6}{se(\hat{\beta}_k)}$$

$\left|t_{cal}\right| < t_{\frac{\alpha}{2}}(n-K) \Rightarrow$ accept H_0. Otheriwse, reject H_0.

Testing for Effect of X_k on Y

Mean-independence of Y on X_k

$H_0 : \beta_k = 0$

$H_1 : \beta_k \neq 0$

$$t_{cal} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)}$$

Accept H_0 => X_k has no significant effect on Y

Overall F-test (1)

<u>Assumption</u>

There is a constant term in the model or X_1 is a vector of one. Why?

Test for mean-independence of Y on $[X_2, X_3, \ldots, X_K]$

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_K = 0$$

$$H_1 : \beta_2 \neq \beta_3 \neq \ldots \neq \beta_K \neq 0$$

Overall F-test (2)

We are choosing between

$Y = \beta_1 + \varepsilon$ ---- (H_0)

expect low R^2 when all X_k's are included

$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_K X_K + \varepsilon$ ---- (H_1)

expect higher R^2

Overall F-test (3)

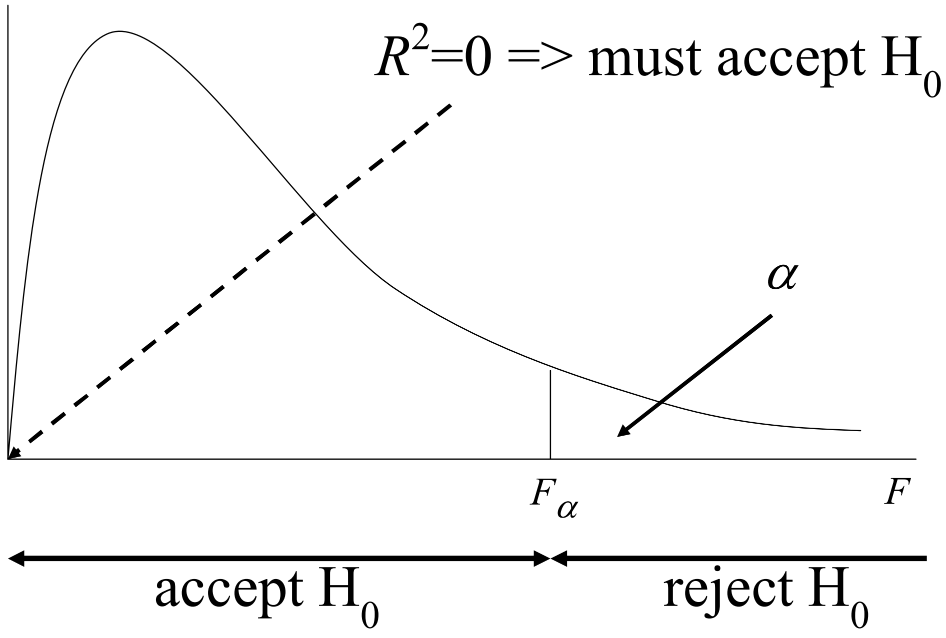
$$F_{cal} = \frac{R^2}{1-R^2}\frac{n-K}{K-1} \sim F(K-1, n-K)$$

Accept H_0 if $F_{cal} < F_{\alpha}(K\text{-}1, n\text{-}K)$. Otherwise, reject H_0. Note that

1) an F-test is always right-tailed.

2) we need a positive R^2.

Overall F-test (4)



Generalized F-test (1)

$$\mathrm{H}_0 : \mathbf{H}(\boldsymbol{\beta}) = \mathbf{0}$$

$$\mathrm{H}_1 : \mathbf{H}(\boldsymbol{\beta}) \neq \mathbf{0}$$

where

$\mathbf{H}(\boldsymbol{\beta})$ is a Mx1 vector function of $\boldsymbol{\beta}$

Note that M must be less than K.

Generalized F-test (2)

$$\mathrm{H}_0 : \begin{bmatrix} H_1(\boldsymbol{\beta}) \\ H_2(\boldsymbol{\beta}) \\ \vdots \\ H_M(\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \mathrm{H}_1 : \begin{bmatrix} H_1(\boldsymbol{\beta}) \\ H_2(\boldsymbol{\beta}) \\ \vdots \\ H_M(\boldsymbol{\beta}) \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or

$$\mathrm{H}_0 : H_1(\boldsymbol{\beta}) = 0, H_2(\boldsymbol{\beta}) = 0, \ldots, H_M(\boldsymbol{\beta}) = 0$$

$$\mathrm{H}_1 : H_1(\boldsymbol{\beta}) \neq 0, H_2(\boldsymbol{\beta}) \neq 0, \ldots, H_M(\boldsymbol{\beta}) \neq 0$$

Generalized F-test (3)

Linear Restriction

$\mathbf{H}(\beta)$ is a Mx1 vector linear function of β

$$\mathbf{H}(\beta) = \mathbf{R}\beta - \mathbf{r}$$

where $\mathbf{R}$ is an MxK coefficient matrix with Rank=M

$\mathbf{r}$ is a Mx1 constant vector

$H_0 : \mathbf{R\beta} - \mathbf{r} = \mathbf{0}$ or $\mathbf{R\beta} = \mathbf{r}$

$H_1 : \mathbf{R\beta} - \mathbf{r} \neq \mathbf{0}$ or $\mathbf{R\beta} \neq \mathbf{r}$

Generalized F-test (4)

Two approaches

- Restricted Least Square (RLS)
- Wald Test

Restricted Least Square (1)

Require two LS runs

Unrestricted run is the OLS run on the original model

==> SSR_U

where

SSR_U is the sum of squared residuals from the unrestricted run

Restricted Least Square (2)

Restricted LS run is as follows

$$\min_{\beta} [\mathbf{Y}-\mathbf{X}\beta]^T[\mathbf{Y}-\mathbf{X}\beta]$$

subject to $\mathbf{R}\beta=\mathbf{r}$

==> SSR_R

where

SSR_R is the sum of squared residuals from the restricted run

Restricted Least Square (3)

Transform RLS to OLS (Elimination Approach)

Define $\mathbf{R}$=[$\mathbf{A}$ $\mathbf{B}$] where

$\mathbf{A}$ is an *M*x*M* invertible sub-matrix of $\mathbf{R}$

$\mathbf{B}$ is the *M*x(*K-M*) sub-matrix containing columns of $\mathbf{R}$ not in $\mathbf{A}$

Restricted Least Square (4)

Define $\mathbf{X}$=[$\mathbf{V}$ $\mathbf{W}$] where

$\mathbf{V}$ is an *N*x*M* sub-matrix of $\mathbf{X}$

$\mathbf{W}$ is the *N*x(*K-M*) sub-matrix containing columns of $\mathbf{X}$ not in $\mathbf{V}$

Restricted Least Square (5)

Re-write the restriction as

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma} \\ \boldsymbol{\delta}\end{bmatrix} = \mathbf{r} \quad \text{or } \mathbf{A}\boldsymbol{\gamma} + \mathbf{B}\boldsymbol{\delta} = \mathbf{r}$$

where

γ is a Mx1 subset of β
δ is a $(K\text{-}M)$x1 subset of β

Restricted Least Square (6)

Re-write the model as

$$\mathbf{Y} = \begin{bmatrix}\mathbf{V} & \mathbf{W}\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma} \\ \boldsymbol{\delta}\end{bmatrix} + \boldsymbol{\varepsilon}$$

$$= \mathbf{V}\boldsymbol{\gamma} + \mathbf{W}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

Restricted Least Square (7)

Since **A** is invertible,

$$\boldsymbol{\gamma} = \mathbf{A}^{-1}[\mathbf{r} - \mathbf{B}\boldsymbol{\delta}]$$

Substitute into the model.

$$\mathbf{Y} = \mathbf{V}\mathbf{A}^{-1}[\mathbf{r} - \mathbf{B}\boldsymbol{\delta}] + \mathbf{W}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

$$\mathbf{Y} - \mathbf{V}\mathbf{A}^{-1}\mathbf{r} = [\mathbf{W} - \mathbf{V}\mathbf{A}^{-1}\mathbf{B}]\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

Restricted Least Square (8)

$$\mathbf{P} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

where $\mathbf{P} = \mathbf{Y} - \mathbf{V}\mathbf{A}^{-1}\mathbf{r}$, $\mathbf{Z} = \mathbf{W} - \mathbf{V}\mathbf{A}^{-1}\mathbf{B}$

Apply OLS

$$\hat{\boldsymbol{\delta}} = [\mathbf{Z}^{\mathrm{T}}\mathbf{Z}]^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{P}$$

$$\hat{\boldsymbol{\gamma}} = \mathbf{A}^{-1}[\mathbf{r} - \mathbf{B}\hat{\boldsymbol{\delta}}]$$

$$\hat{\sigma}_R^2 = \frac{SSR_R}{n-(K-M)} \quad SSR_R = [\mathbf{P} - \mathbf{Z}\hat{\boldsymbol{\delta}}]^{\mathrm{T}}[\mathbf{P} - \mathbf{Z}\hat{\boldsymbol{\delta}}]$$

Restricted Least Square (9)

$$V(\hat{\boldsymbol{\delta}}) = \sigma^2[\mathbf{Z}^T\mathbf{Z}]^{-1}$$

$$V(\hat{\boldsymbol{\gamma}}) = \mathbf{A}^{-1}\mathbf{B}V(\hat{\boldsymbol{\delta}})\mathbf{B}^T[\mathbf{A}^T]^{-1}$$

$$= \sigma^2\mathbf{A}^{-1}\mathbf{B}[\mathbf{Z}^T\mathbf{Z}]^{-1}\mathbf{B}^T[\mathbf{A}^T]^{-1}$$

$$COV(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\delta}}) = \sigma^2\mathbf{A}^{-1}\mathbf{B}[\mathbf{Z}^T\mathbf{Z}]^{-1}$$

$$V(\hat{\boldsymbol{\beta}}_R) = \sigma^2\begin{bmatrix} \mathbf{A}^{-1}\mathbf{B}[\mathbf{Z}^T\mathbf{Z}]^{-1}\mathbf{B}^T[\mathbf{A}^T]^{-1} & \mathbf{A}^{-1}\mathbf{B}[\mathbf{Z}^T\mathbf{Z}]^{-1} \\ [\mathbf{Z}^T\mathbf{Z}]^{-1}\mathbf{B}^T[\mathbf{A}^T]^{-1} & [\mathbf{Z}^T\mathbf{Z}]^{-1} \end{bmatrix}$$

Restricted Least Square (10)

<u>Lagrange Method</u>

FOC $$-\mathbf{X}^\mathbf{T}[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}] + \mathbf{R}^\mathbf{T}\hat{\boldsymbol{\lambda}} = \mathbf{0}$$

$$\mathbf{X}^T\mathbf{Y} - \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}}_R - \mathbf{R}^\mathbf{T}\hat{\boldsymbol{\lambda}} = \mathbf{0}$$

$$\hat{\boldsymbol{\beta}}_R = [\mathbf{X}^T\mathbf{X}]^{-1}[\mathbf{X}^T\mathbf{Y} - \mathbf{R}^\mathbf{T}\hat{\boldsymbol{\lambda}}]$$

$$= [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{Y} - [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{R}^\mathbf{T}\hat{\boldsymbol{\lambda}}$$

$$= \hat{\boldsymbol{\beta}}_U - [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{R}^\mathbf{T}\hat{\boldsymbol{\lambda}}$$

Restricted Least Square (11)

Substitute into $\mathbf{R}\beta=\mathbf{r}$

$$[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}] - \mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\lambda}} = \mathbf{0}$$

$$\hat{\boldsymbol{\lambda}} = \mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$$

where $\mathbf{S} = \mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}$

$$\hat{\boldsymbol{\beta}}_R = \hat{\boldsymbol{\beta}}_U - [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$$

$$= \left[\mathbf{I} - [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\mathbf{S}^{-1}\mathbf{R}\right]\hat{\boldsymbol{\beta}}_U$$

$$+ [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\mathbf{S}^{-1}\mathbf{r}$$

Restricted Least Square (12)

$$\mathrm{V}(\hat{\boldsymbol{\beta}}_R) = \sigma^2\mathbf{D}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{D}^{\mathbf{T}}$$

where $\mathbf{D} = \mathbf{I} - [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\mathbf{S}^{-1}\mathbf{R}$

$$\hat{\sigma}_R^2 = \frac{SSR_R}{n-(K-M)}$$

where $SSR_R = [\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R]^{\mathrm{T}}[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R]$

Prove that both RLS and LM yield identical result

Restricted Least Square (13)

$$F_{cal} = \frac{(SSR_R - SSR_U)/M}{SSR_U/(n-K)} \sim F(M, n-K)$$

where

M is the number of restriction equations or constraints or the number of rows in matrix **R**

Note that $df_U = n\text{-}K$ and $df_R = n\text{-}(K\text{-}M)$

Restricted Least Square (14)

$F_{cal} < F_{\alpha}(M,n\text{-}K)$ ==>Accept H_0

or restriction holds

$F_{cal} > F_{\alpha}(M,n\text{-}K)$ ==>Reject H_0 or restriction

does not holds

Wald Test (1)

Require only the Unrestricted run

==> $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$

$$F_{cal} = [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]^{\mathrm{T}} \left[\mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}}\right]^{-1} [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}] \frac{1}{\widehat{\sigma^2} M}$$

$$\sim F(M, n-K)$$

Accept H_0 if $F_{cal} < F_{\alpha}(M,n\text{-}K)$.
Otherwise, reject H_0.

Wald Test (2)

<u>Concept</u>

Note that, given H_0 is true,

$$[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}] \sim \mathrm{MVN}(\mathbf{0}, \sigma^2\mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}})$$

Standardize a normal vector

$$\mathbf{Z} = \left[\sigma^2\mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}}\right]^{-\frac{1}{2}} [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]$$

Wald Test (3)

Note that Z is a vector of M iid standard normal RV's

$$\mathbf{Z}^{\mathrm{T}}\mathbf{Z} = [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]^{\mathrm{T}} \left[\sigma^2 \mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}}\right]^{-1} [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]$$

$$\sim \chi^2(M)$$

Wald Test (4)

$$F_{cal} = \frac{\dfrac{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}}{M}}{\dfrac{(n-K)\dfrac{\hat{\sigma}^2}{\sigma^2}}{n-K}} \sim F(M, n-K)$$

$$= [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]^{\mathrm{T}} \left[\mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}}\right]^{-1} [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}] \frac{1}{\hat{\sigma}^2 M}$$

Example#1 (1)

Overall F-test is a simple case of Generalized F-tests with

$$\underset{(K-1)\text{x}K}{\mathbf{R}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Example#1 (2)

<u>RLS Approach</u>

Since the restriction set is simple, the restricted model can be written as

$$Y_i = \beta_1 + \varepsilon_i$$

By OLS => $\hat{\beta}_1 = \overline{Y}$

$$SSR_R = \sum_{i=}^{n} (Y_i - \hat{\beta}_1)^2 = \sum_{i=}^{n} (Y_i - \overline{Y})^2$$

Example#1 (3)

Note that SSR_{R} = SST of the unrestricted model.

$$F_{cal} = \frac{SST_U - SSR_U}{SSR_U}\frac{n-K}{K-1}$$

$$= \frac{(SST_U - SSR_U)/SST_U}{SSR_U/SST_U}\frac{n-K}{K-1}$$

$$= \frac{R^2}{1-R^2}\frac{n-K}{K-1}$$

Example#1 (4)

<u>Wald Test</u> (single-run)

See Eviews example

Example#2 (1)

Removing X_2 and X_3

$$H_0 : \beta_2 = 0, \beta_3 = 0$$

$$H_1 : \beta_2 \neq 0, \beta_3 \neq 0$$

Use this **R** and **r** in the test

$$\underset{2\text{x}K}{\mathbf{R}} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Example#2 (2)

RLS Approach

Since the restriction set is simple, the restricted model can be written as

$$Y_i = \beta_1 + \beta_4 X_{4i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

Example#3 (1)

$$\mathrm{H}_0 : \beta_2 = 0, \beta_3 = 0, \beta_4 + \beta_5 = 1$$

$$\mathrm{H}_1 : \beta_2 \neq 0, \beta_3 \neq 0, \beta_4 + \beta_5 \neq 1$$

Use this **R** and **r** in the test

$$\underset{3\times K}{\mathbf{R}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Example#3 (2)

<u>RLS Approach</u>

Since the restriction set is simple, the restricted model can be written as

$$Y_i = \beta_1 + \beta_4 X_{4i} + (1-\beta_4) X_{5i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

$$Y_i - X_{5i} = \beta_1 + \beta_4 (X_{4i} - X_{5i}) + \beta_6 X_{6i} \ldots + \beta_K X_{Ki} + \varepsilon_i$$

See EViews example

Normality Tests

- Cumulative Normal plot
- Goodness-of-fit test (a Chi-square test)
- Jarque-Bera Test

Cumulative Normal Plot (1)

If X is normal, graph of inverse CDF of cumulative relative frequency versus X will exhibit linearity

Step 1 Sort X

Step 2 Calculate Cumulative Relative frequency F for each X. Note that

$$0 \le F \le 1$$

Cumulative Normal Plot (2)

<u>Step 3</u> Calculate (look for in the Z-table) the Z value for the area on left equal to F

<u>Step 4</u> Plot Z against standardized X

If the graph is linear with slope of +1, ==> X~Normal

Jarque-Bera Normality Test (1)

$$\mathrm{H}_0 : S = 0, \kappa = 3$$

$$\mathrm{H}_1 : S \neq 0, \kappa \neq 3$$

where S is skewedness

κ is Kurtosis

$$\chi^2_{cal} = (n-K)\left(\frac{1}{6}\hat{S}^2 + \frac{1}{24}(\hat{\kappa}-3)^2\right) \sim \chi^2_\alpha(2)$$

Jarque-Bera Normality Test (2)

where $$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

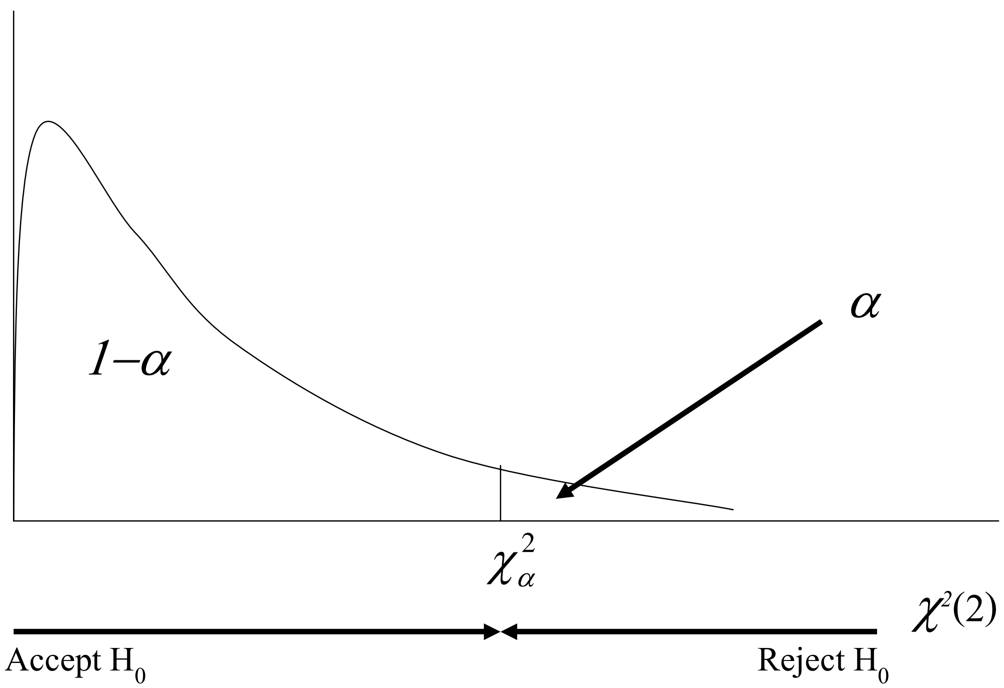
$$\hat{S} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\hat{\sigma}}\right)^3$$

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\hat{\sigma}}\right)^4$$

Perform a right-tailed χ^2-test

Note: different definition for skewedness and kurtosis

Jarque-Bera Normality Test (3)



Prediction Interval of Y (1)

$$\mathrm{E}(Y \mid \mathbf{X}_0) = \mathbf{X}_0\boldsymbol{\beta}$$

$$\widehat{\mathrm{E}(Y \mid \mathbf{X}_0)} = \mathbf{X}_0\hat{\boldsymbol{\beta}}$$

Is an <u>unbiased</u> estimator of E(Y|$\mathbf{X}_0$)

where $\mathbf{X}_0 = [X_{10}, X_{20}, \ldots, X_{K0}]$

$$\mathrm{V}\left(\widehat{\mathrm{E}(Y \mid \mathbf{X}_0)}\right) = \mathbf{X}_0 \mathbf{V}(\hat{\boldsymbol{\beta}})\left[\mathbf{X}_0\right]^{\mathrm{T}}$$

$$= \sigma^2 \mathbf{X}_0 [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\left[\mathbf{X}_0\right]^{\mathrm{T}}$$

Prediction Interval of Y (2)

(1-α)100% CI for E(Y|$\mathbf{X}_0$) =

$$= \mathbf{X}_0\hat{\boldsymbol{\beta}} + t_{\frac{\alpha}{2}}(n-K)\,\mathrm{se}\left(\widehat{\mathrm{E}(Y \mid \mathbf{X}_0)}\right)$$

where

$$\mathrm{se}\left(\widehat{\mathrm{E}(Y \mid \mathbf{X}_0)}\right) = \sqrt{\widehat{\sigma^2}\mathbf{X}_0 [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\left[\mathbf{X}_0\right]^{\mathrm{T}}}$$

Prediction Interval of Y (3)

(1-α)100% PI for Y|$\mathbf{X}_0$ =

$$= \mathbf{X}_0\hat{\boldsymbol{\beta}} + t_{\frac{\alpha}{2}}(n-K)\text{se}(Y \mid \mathbf{X}_0)$$

where

$$\text{se}(Y \mid \mathbf{X}_0) = \sqrt{\widehat{\sigma^2}\left(1 + \mathbf{X}_0[\mathbf{X}^T\mathbf{X}]^{-1}[\mathbf{X}_0]^T\right)}$$