This is a note from Yongmiao Hong's book, I've changed the subscript t into i, this note covers the entire content of chap1.

Assumptions

Ass 1: Linearity

$$Y_i = X_i'\beta + \varepsilon_i$$

Ass 2: Strict Exogeneity

$$E(\varepsilon_i | X_i) = 0$$

By the law of iterated expectations(IIE):
$E(X_i\varepsilon_i) = 0$
$E(\varepsilon_i) = 0$
then $\text{Cov}(X_i, \varepsilon_i) = 0$

Ass 3: Nonsingularity

$$Rank(\mathbf{X}'\mathbf{X}) = K \leq n$$

Ass 4: Spherical Error Variance

$$E(\varepsilon_i^2 | \mathbf{X}) = \sigma^2$$

We can write Ass 2 and Ass 4 compactly as follows:

$$E(\varepsilon | \mathbf{X}) = 0 \text{ and } E(\varepsilon\varepsilon' | \mathbf{X}) = \sigma^2 \mathbf{I}$$

Ass 5: Conditional Normality

$$\varepsilon | \mathbf{X} \sim N(0, \sigma^2 \mathbf{I})$$

Ass 5 implies both Ass 2 and Ass 4.

OLS Estimator

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i Y_i \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i Y_i
\end{aligned}$$

The FOC implies that the estimated residual e is **orthogonal** to regressors X:

$$\mathbf{X}'e = \sum_{i=1}^{n} X_i e_i = 0$$

Sample error:

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

Define an $n \times n$ projection matrix:

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

and

$$\mathbf{M} = \mathbf{I} - \mathbf{P}$$

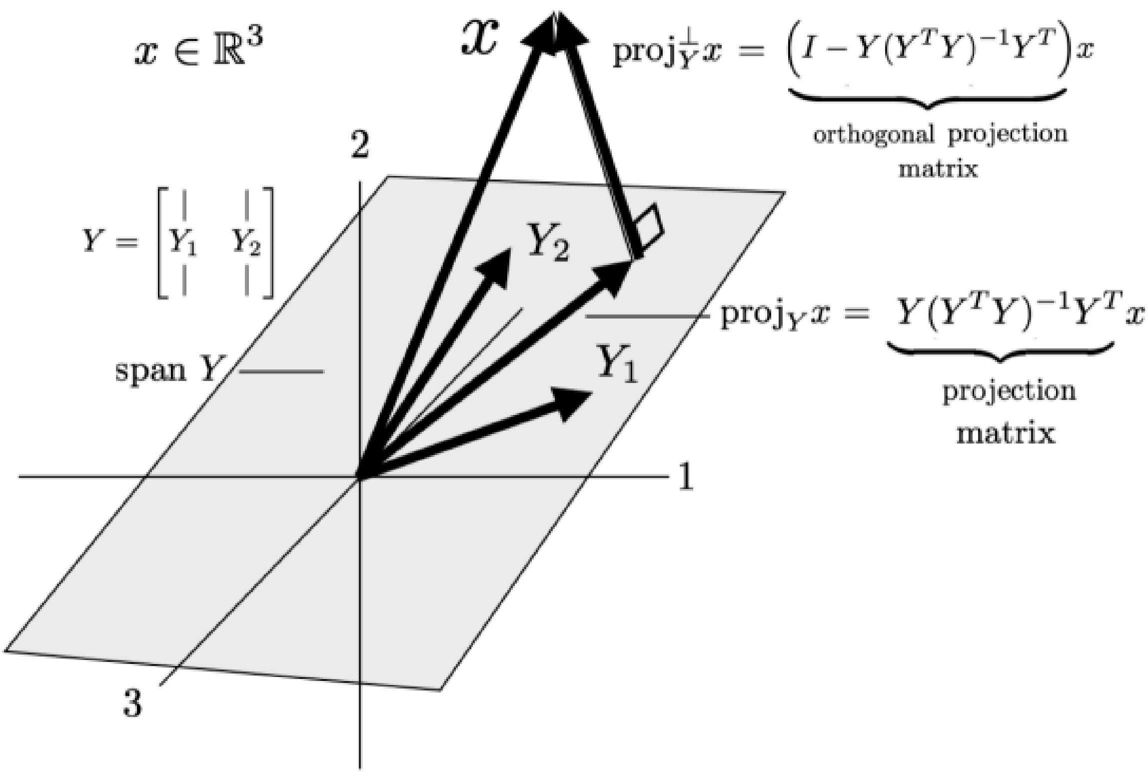
Then both matrices $\mathbf{P}$ and $\mathbf{M}$ are symmetric (i.e., $\mathbf{P} = \mathbf{P}'$ and $\mathbf{M} = \mathbf{M}'$) and idempotent (i.e., $\mathbf{P}^2 = \mathbf{P}, \mathbf{M}^2 = \mathbf{M}$), with

$$\begin{aligned} \mathbf{PX} &= \mathbf{X}, \\ \mathbf{MX} &= 0 \end{aligned}$$

then:

$$SSR(\hat{\beta}) = e'e = Y'MY = \varepsilon' M \varepsilon$$

understanding from the projecting perspective:



proof: $P = \underset{n\times k}{X}\underset{k\times k}{(X'X)^{-1}}\underset{k\times n}{X'}$ $M = I_n - P$

① $P' = P$: $P' = (X(X'X)^{-1}X')'$

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

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

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

$= P$

$M' = M$: $M' = (I_n - P)'$

$= I_n' - P'$

$= I_n - P$

$= M$

② $P^2 = P$: $P^2 = X(X'X)^{-1}\cancel{X'X(X'X)^{-1}}X'$

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

$= P$

$M^2 = M$: $M^2 = (I_n - P)^2$

$= I_n^2 - I_nP - PI_n + P^2$

$= I_n - 2P + P$

$= I_n - P$

$= M$

③ $PX = X$: $PX = X\cancel{(X'X)^{-1}X'X}$

$\downarrow$
projection

$= X$

$MX = 0$: $MX = (I_n - P)X$

$= X - PX$

$= 0$

④ $e = M\varepsilon$: $e = Y - X\hat{\beta}$

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

$= (I - X(X'X)^{-1}X)Y$

$= MY$

$= M(X\beta^{\circ} + \varepsilon)$

$= M\varepsilon$

⑤ $SSR(\hat{\beta}) = e'e$

$= (M\varepsilon)'M\varepsilon$

$= \varepsilon'M'M\varepsilon$

$= \varepsilon'MM\varepsilon$

$= \varepsilon'M\varepsilon$

Goodness of Fit

Uncentered R^2:

$$R_{uc}^2 = \frac{\hat{Y}'\hat{Y}}{Y'Y} = 1 - \frac{e'e}{Y'Y}$$

Centered R^2/Coefficient of Determination:

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

$$R^2 = \rho_{Y\hat{Y}}^2, 0 \leq R^2 \leq 1$$

If the regression function doesn't contain the intercept, then the R^2 may be negative.

Consistency and Efficiency of the OLS Estimator

- Unbiasedness:
  $E(\hat{\beta}|\mathbf{X}) = \beta$ and $E(\hat{\beta}) = \beta$
- Vanishing Variance:
  var($\hat{\beta}|\mathbf{X}$) = $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
- Orthogonality:
  cov($\hat{\beta}, e|\mathbf{X}$) = 0
- Gauss-Markov Theorem:
  var($\hat{b}|\mathbf{X}$) - var($\hat{\beta}|\mathbf{X}$) ~ PSD(positive semi-definite)
- Sample Residual Variance Estimator:

$$s^2 = e'e/(n-K) = \frac{1}{n-K}\sum_{i=1}^n e_i^2$$

Sampling Distribution of the OLS Estimator

Ass 5: Conditional Normality

$$\varepsilon|\mathbf{X} \sim N(0, \sigma^2\mathbf{I})$$

Normality of the OLS Estimator:

$$(\hat{\beta} - \beta)|\mathbf{X} \sim N[0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$$
$$R(\hat{\beta} - \beta)|\mathbf{X} \sim N[0, \sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R']$$

The $J \times K$ matrix R is a selection matrix.

Variance Estimation for the OLS Estimator

Residual Variance Estimator:

$$\left.\frac{(n-K)s^2}{\sigma^2}\right|_{\mathbf{X}} = \left.\frac{e'e}{\sigma^2}\right|_{\mathbf{X}} \sim \chi^2_{n-K}$$

Hypothesis Testing

$$\mathbf{H}_0 : R\beta = r$$

we can consider the statistic $R\hat{\beta} - r$ and check if this difference is significantly different from zero. Under H_0, we have:

$$R\hat{\beta} - r = R\hat{\beta} - R\beta = R(\hat{\beta} - \beta) \to 0 \text{ as } n \to \infty$$

Under the alternative to $\mathbf{H}_0$: $R\beta \neq r$, but we still have $\hat{\beta} - \beta \to 0$. It follows that:

$$R\hat{\beta} - r = R(\hat{\beta} - \beta) + R\beta - r \to R\beta - r \neq 0 \text{ as } n \to \infty$$

In other words, $R\hat{\beta} - r$ will converge to a nonzero limit $R\beta - r$.

Under H_0:

$$(R\hat{\boldsymbol{\beta}} - r)|\mathbf{X} \sim N(0, \sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R')$$

Questions:

(1) $\mathrm{var}[(R\hat{\beta} - r)|\mathbf{X}] = \sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R'$ is a scalar or a matrix?

(2) Since σ^2 is unknown, $\frac{R\hat{\beta}-r}{\sqrt{\sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R'}}$ is no longer normally distributed.

Case 1: t-Test: $J = 1$, replace σ^2 by s^2

$$\begin{aligned} \mathrm{T} &= \frac{R\hat{\beta} - r}{\sqrt{s^2 R(\mathbf{X}'\mathbf{X})^{-1}R'}} \\ &= \frac{\frac{R\hat{\beta}-r}{\sqrt{\sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R'}}}{\sqrt{\frac{(n-K)s^2}{\sigma^2}/(n-K)}} \\ &\sim \frac{N(0,1)}{\sqrt{\chi^2_{n-K}/(n-K)}} \\ &\sim t_{n-K} \end{aligned}$$

- Decision Rule of the t-Test Based on Critical Values:

Reject H_0: $|T| > C_{t_{n-K}, \frac{\alpha}{2}}$

$$P\left(t_{n-K} > C_{t_{n-K},\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$
$$P\left(|t_{n-K}| > C_{t_{n-K},\frac{\alpha}{2}}\right) = \alpha$$

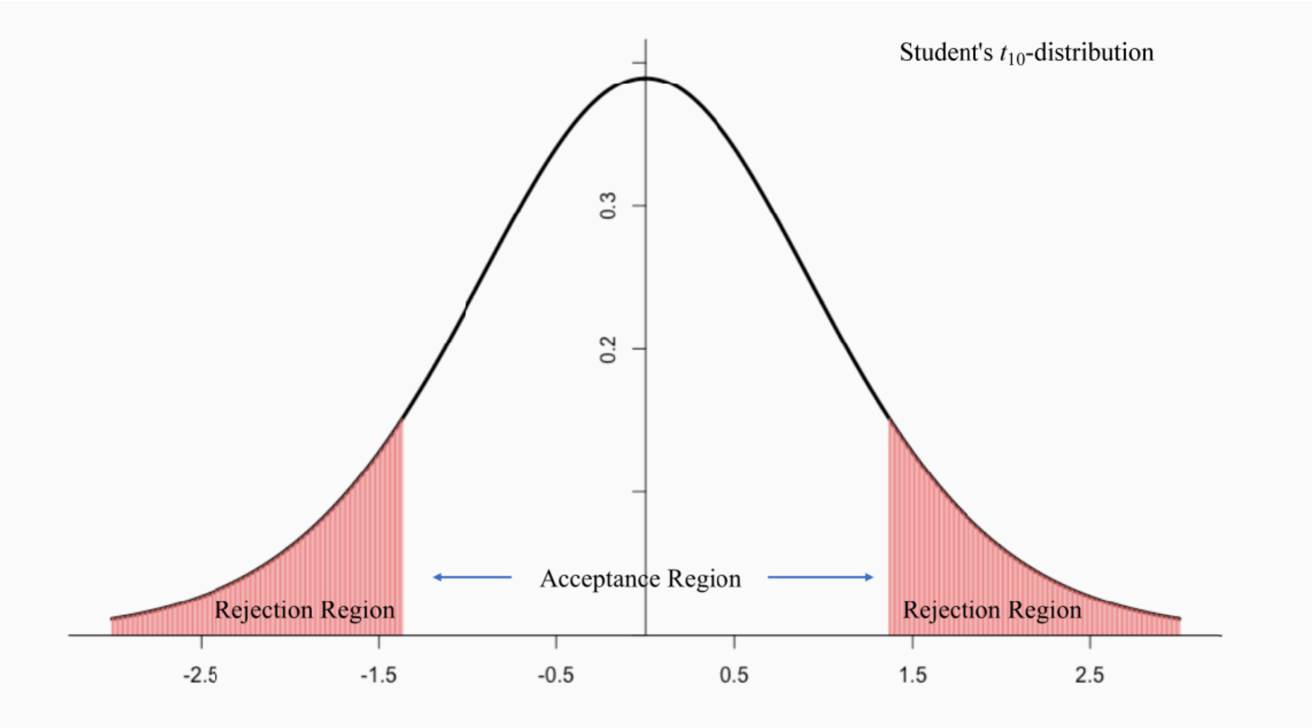


Figure 3.3 Acceptance and rejection regions of a t-test.

- Decision Rule Based on the P -value:
  - A small P-value means reject the null hypothesis.
  - A large P-value means accept the null hypothesis.

When we reject a null hypothesis, we often say there is a **statistically significant** effect. This does not mean that there is an **economic significant** effect. This is because when **large samples** are used, small and practically unimportant effects are likely to be statistically significant(since denominator of T which contains s^2 is very small when n is infinite).

Case 2: F-Test: $J > 1$, replace σ^2 by s^2

Lemma Quadratic Form of Normal Random Variables:
If a $q \times 1$ random vector $Z \sim N(0, V)$, where $V = \mathrm{var}(Z)$ is a nonsingular $q \times q$ variance-covariance matrix, then

$$Z'V^{-1}Z \sim \chi_q^2$$

Applying this lemma, and using the result that

$$(R\hat{\beta} - r)|\mathbf{X} \sim N[0, \sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R']$$

under $\mathbf{H}_0$, we have the quadratic form:

$$(R\hat{\beta} - r)'[\sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r) \sim \chi_J^2$$

$$\frac{(R\hat{\beta}-r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta}-r)}{\sigma^2} \sim \chi^2_J$$

Like in constructing a t-test statistic, we should replace σ^2 by s^2 in the left hand side:

$$\frac{(R\hat{\beta}-r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta}-r)}{s^2}$$

$$\begin{aligned}
\mathrm{F} &\equiv \frac{(R\hat{\beta}-r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta}-r)/J}{s^2} \\
&= \frac{\frac{(R\hat{\beta}-r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta}-r)}{\sigma^2}/J}{\frac{(n-K)s^2}{\sigma^2}/(n-K)} \\
&\sim F_{J,n-K}
\end{aligned}$$

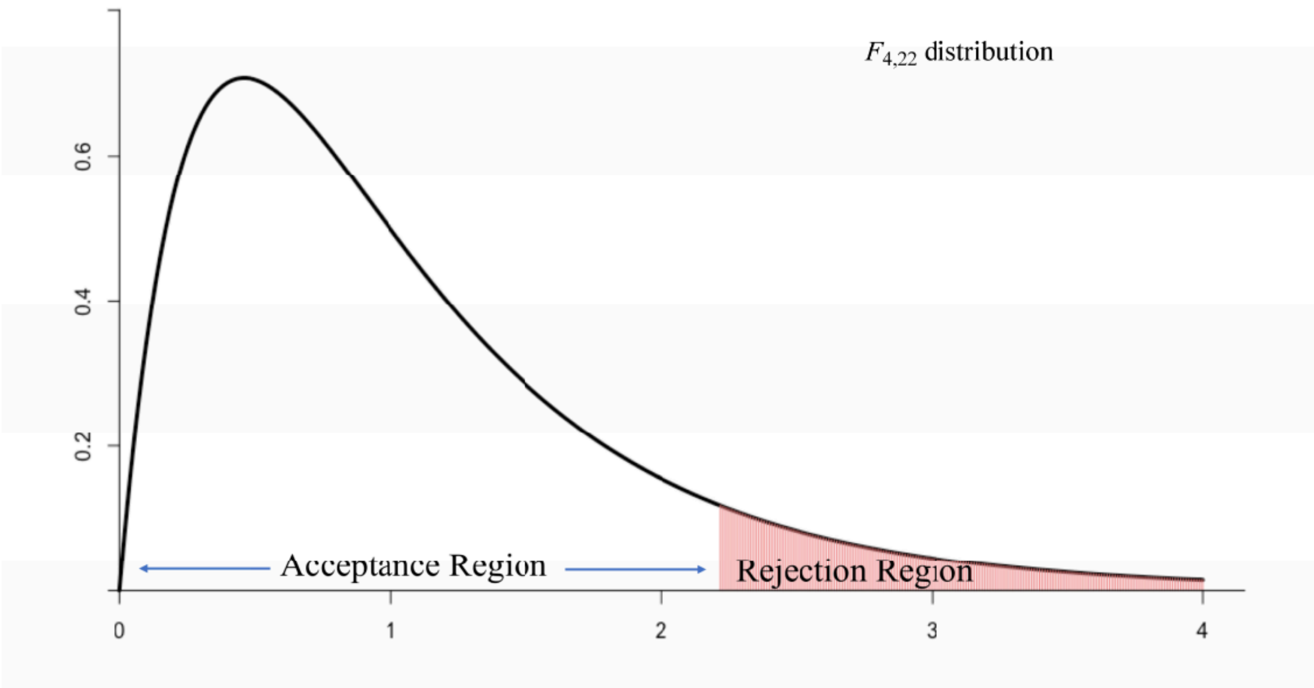


Figure 3.4 Acceptance and rejection regions of an F-test.

$$\mathrm{F} = \frac{(\tilde{e}'\tilde{e}-e'e)/J}{e'e/(n-K)} = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-K)}$$

Wald Test:

$$W = J \cdot F = \frac{(R\hat{\beta}-r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta}-r)}{s^2} \xrightarrow{d} \chi^2_J$$

Summary of test statistics:

$Z \Rightarrow t \rightarrow Z(when\, n \rightarrow \infty)$
$Z \Rightarrow \chi 2 \Rightarrow F \Rightarrow W \rightarrow \chi 2(when\, n \rightarrow \infty)$
$F = t^2$

Applications

- Case 1: Testing for Joint Significance of Explanatory Variables

- Case 2: Testing for Omitted Variables

  Examples: Granger Causality; Chow test
- Case 3: Testing for Linear Restrictions

GLS Estimation

Relax Ass 5 (no conditional heteroskedasticity and auto-correlation)

$$\varepsilon|\mathbf{X} \sim N(0, \sigma^2\mathbf{V})$$

It allows for conditional heteroskedasticity of known form $\sigma^2\mathbf{V} = \sigma^2 V(\mathbf{X})$. This is a strong assumption(V is often an unknown form), so GLS is not widely used in reality as OLS. Since Ass 5 is obeyed. T and F are useless.

- Unbiasedness:

$$E(\hat{\beta}|\mathbf{X}) = \beta \text{ and } E(\hat{\beta}) = \beta$$

- Variance:

$$\operatorname{var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \neq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

- Normal Distribution:

$$(\hat{\beta} - \beta)|\mathbf{X} \sim N(0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}).$$

- Non-Zero Correlation Between $\hat{\beta}$ and e:

$$cov(\hat{\beta}, e|\mathbf{X}) \neq 0.$$

Cholesky's Decomposition:

$$\begin{aligned} \mathbf{V}^{-1} &= \mathbf{C}'\mathbf{C}, \\ \mathbf{V} &= \mathbf{C}^{-1}(\mathbf{C}')^{-1} \end{aligned}$$

Consider the original linear regression model, if we multiply the equation by C, we obtain the transformed regression model:

$$\mathbf{CY} = (\mathbf{CX})\beta + \mathbf{C}\varepsilon \text{ or } Y^* = \mathbf{X}^*\beta + \varepsilon^*$$

GLS estimator:

$$\begin{aligned} \hat{\beta}^* &= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}Y^* \\ &= (\mathbf{X}'\mathbf{C}'\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{C}'\mathbf{C}Y) \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y \end{aligned}$$

$$var(\varepsilon^*) = var(\mathbf{C}\varepsilon) = \mathbf{C}'var(\varepsilon)\mathbf{C} = \mathbf{C}'\mathbf{C}\sigma^2\mathbf{V} = \mathbf{V}^{-1}\sigma^2\mathbf{V} = \sigma^2$$

Solutions:

(1) Approach I: Adaptive Feasible GLS Estimation

(2) Approach II: White and HAC Variance-Covariance Matrix Estimation

White's (1980): heteroskedasticity consistent variance-covariance matrix estimator

Andrews (1991) and Newey and West (1987, 1994): Heteroskedasticity and Autocorrelation Consistent (HAC) variance-covariance matrix estimation