Advanced Econometrics

Finite Sample Properties of Least Squares (Hansen Chapter 5)

Instructor: Ma, Jun

Renmin University of China

October 25, 2021

Normal Regression Model

- The normal regression model is the linear regression model with an independent normal error

$$\begin{aligned} Y &= \boldsymbol{X}'\boldsymbol{\beta} + e \\ e &\sim \mathrm{N}\left(0, \sigma^2\right). \end{aligned}$$

- The likelihood is the name for the joint probability density of the data, evaluated at the observed sample, and viewed as a function of the parameters.
- The maximum likelihood estimator is the value which maximizes this likelihood function.

- The conditional density of Y given $\boldsymbol{X}$:

$$f\left(Y \mid \boldsymbol{X}\right) = \frac{1}{\left(2\pi\sigma^2\right)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}\left(Y - \boldsymbol{X}'\boldsymbol{\beta}\right)^2\right).$$

- The conditional density of $\boldsymbol{Y}$ given $\boldsymbol{X}$:

$$\begin{aligned} f_{\boldsymbol{Y}\mid\boldsymbol{X}}\left(\boldsymbol{Y} \mid \boldsymbol{X}\right) &= \prod_{i=1}^{n} f_{Y_i\mid\boldsymbol{X}_i}\left(Y_i \mid \boldsymbol{X}_i\right) \\ &= \prod_{i=1}^{n} \frac{1}{\left(2\pi\sigma^2\right)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta}\right)^2\right) \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta}\right)^2\right) \\ &= L\left(\boldsymbol{\beta}, \sigma^2\right). \end{aligned}$$

$L\left(\boldsymbol{\beta}, \sigma^2\right)$ is called the likelihood function.

- Work with the natural logarithm:

$$\begin{aligned}\log f(\boldsymbol{Y} \mid \boldsymbol{X}) &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta}\right)^2 \\ &= \log L\left(\boldsymbol{\beta}, \sigma^2\right).\end{aligned}$$

- The MLE:

$$\left(\hat{\boldsymbol{\beta}}_{\text{mle}}, \hat{\sigma}^2_{\text{mle}}\right) = \underset{\boldsymbol{\beta}\in\mathbb{R}^k, \sigma^2>0}{\operatorname{argmax}} \log L\left(\boldsymbol{\beta}, \sigma^2\right).$$

- In most applications of maximum likelihood, the MLE must be found by numerical methods. However, in the case of the normal regression model we can find an explicit expression.
- FOC:

$$
\begin{aligned}
0 =& \left.\frac{\partial \log L\left(\boldsymbol{\beta}, \sigma^2\right)}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{\mathrm{mle}}, \sigma^2=\hat{\sigma}^2_{\mathrm{mle}}} = \frac{1}{\hat{\sigma}^2_{\mathrm{mle}}} \sum_{i=1}^{n} \boldsymbol{X}_i \left(Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}_{\mathrm{mle}}\right) \\
0 =& \left.\frac{\partial \log L\left(\boldsymbol{\beta}, \sigma^2\right)}{\partial \sigma^2}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{\mathrm{mle}}, \sigma^2=\hat{\sigma}^2_{\mathrm{mle}}} = -\frac{n}{2\hat{\sigma}^2_{\mathrm{mle}}} + \frac{1}{\hat{\sigma}^4_{\mathrm{mle}}} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}_{\mathrm{mle}}\right).
\end{aligned}
$$

- The MLE:

$$\hat{\boldsymbol{\beta}}_{\mathrm{mle}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i'\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right) = \hat{\boldsymbol{\beta}}_{\mathrm{ols}}.$$

- The MLE for σ^2:

$$\hat{\sigma}^2_{\mathrm{mle}} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}_{\mathrm{mle}}\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}_{\mathrm{ols}}\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 .$$

- Maximized log-likelihood is a measure of goodness of fit:

$$\log L \left(\hat{\boldsymbol{\beta}}_{\mathrm{mle}}, \hat{\sigma}^2_{\mathrm{mle}}\right) = -\frac{n}{2} \log \left(2\pi \hat{\sigma}^2_{\mathrm{mle}}\right) - \frac{n}{2}.$$

Distribution of OLS Coefficient Vector

- The normality assumption $e_i \mid \boldsymbol{X}_i \sim \mathrm{N}\left(0, \sigma^2\right)$ and iid assumption imply

$$\boldsymbol{e} \mid \boldsymbol{X} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{I}_n \sigma^2\right).$$

- The OLS estimator satisfies

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{e},$$

  which is a linear function of $\boldsymbol{e}$.
- Conditional on $\boldsymbol{X}$,

$$\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid_{\boldsymbol{X}} &\sim \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\mathrm{N}\left(0, \boldsymbol{I}_n \sigma^2\right) \\
&\sim \mathrm{N}\left(0, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \\
&= \mathrm{N}\left(0, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)
\end{aligned}$$

  or

$$\hat{\boldsymbol{\beta}} \mid_{\boldsymbol{X}} \sim \mathrm{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right).$$

- This shows that under the assumption of normal errors, the OLS estimate has an exact normal distribution.

Theorem

In the linear regression model,

$$\hat{\boldsymbol{\beta}} \mid_{\boldsymbol{X}} \sim \mathrm{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)$$

- Any linear function of the OLS estimate is also normally distributed, including individual estimates:

$$\hat{\beta}_j \mid_{\boldsymbol{X}} \sim \mathrm{N}\left(\beta_j, \sigma^2 \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{jj}\right).$$

Distribution of OLS Residual Vector

- The OLS residual vector: $\hat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e}$. $\hat{\boldsymbol{e}}$ is linear in $\boldsymbol{e}$.
- Conditional on $\boldsymbol{X}$,

$$\hat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e} \mid \boldsymbol{X} \sim \mathrm{N}\left(0, \sigma^2 \boldsymbol{M}\boldsymbol{M}\right) = \mathrm{N}\left(0, \sigma^2 \boldsymbol{M}\right).$$

- The joint distribution of $\widehat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{e}}$:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{e}} \end{pmatrix} = \begin{pmatrix} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{e} \\ \boldsymbol{M}\boldsymbol{e} \end{pmatrix} = \begin{pmatrix} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}' \\ \boldsymbol{M} \end{pmatrix} \boldsymbol{e}.$$

- So

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{e}} \end{pmatrix} \mid \boldsymbol{X} \sim \mathrm{N}\begin{pmatrix} \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} & 0 \\ 0 & \sigma^2 \boldsymbol{M} \end{pmatrix}.$$

Theorem

In the linear regression model, $\hat{\boldsymbol{e}} \mid \boldsymbol{X} \sim \mathrm{N}\left(0, \sigma^2 \boldsymbol{M}\right)$ *and is independent of* $\hat{\boldsymbol{\beta}}$.

Distribution of Variance Estimate

- $s^2 = \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}/(n-k) = \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}/(n-k)$.
- The spectral decomposition of $\boldsymbol{M}$: $\boldsymbol{M} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ with $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_n$ and $\boldsymbol{\Lambda}$ is diagonal with the eigenvalues of $\boldsymbol{M}$ on the diagonal.
- Since $\boldsymbol{M}$ is idempotent with rank $n-k$, it has $n-k$ eigenvalues equalling 1 and k eigenvalues equalling 0:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{bmatrix}.$$

- $\boldsymbol{U} = \boldsymbol{H}'\boldsymbol{e} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{I}_n\sigma^2\right)$.

$$\begin{aligned} (n-k)\, s^2 &= \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} \\ &= \boldsymbol{e}'\boldsymbol{H}\begin{bmatrix} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{bmatrix}\boldsymbol{H}'\boldsymbol{e} \\ &= \boldsymbol{U}'\begin{bmatrix} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{bmatrix}\boldsymbol{U} \\ &\sim \sigma^2\chi^2_{n-k}. \end{aligned}$$

Theorem

In the linear regression model, conditional on $\boldsymbol{X}$,

$$\frac{(n-k)\, s^2}{\sigma^2} \sim \chi^2_{n-k}$$

and is independent of $\hat{\boldsymbol{\beta}}$.

t-statistic

- The "z-statistic":

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 \left[(\boldsymbol{X}'\boldsymbol{X})^{-1} \right]_{jj}}} \sim \mathrm{N}(0, 1) .$$

- Replace the unknown variance σ^2 with its estimate s^2:

$$T = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 \left[(\boldsymbol{X}'\boldsymbol{X})^{-1} \right]_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{s\left(\hat{\beta}_j\right)}.$$

- Write the t-statistic as the ratio of the standardized statistic and the square root of the scaled variance estimate:

$$\begin{aligned} T &= \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 \left[(\boldsymbol{X}'\boldsymbol{X})^{-1} \right]_{jj}}} \Big/ \sqrt{\frac{(n-k)s^2}{\sigma^2} \Big/ (n-k)} \\ &\sim \frac{\mathrm{N}(0, 1)}{\sqrt{\chi^2_{n-k} / (n-k)}} \\ &\sim t_{n-k}. \end{aligned}$$

Theorem

In the normal regression model, $T \sim t_{n-k}$.

- This derivation shows that the t-statistic has a sampling distribution which depends only on the quantity $n - k$. The distribution does not depend on any other features of the data.
- In this context, we say that the distribution of the t-statistic is pivotal, meaning that it does not depend on unknowns.
- The theorem only applies to the t-statistic constructed with the homoskedastic standard error estimate. It does not apply to a t-statistic constructed with the robust standard error estimates.

Confidence Intervals for Regression Coefficients

- An OLS estimate $\hat{\boldsymbol{\beta}}$ is a point estimate for the coefficients $\boldsymbol{\beta}$.
- An interval estimate takes the form $\widehat{C} = \left[\widehat{L}, \widehat{U}\right]$. The goal of an interval estimate $\widehat{C}$ is to contain the true value with high probability.
- The interval estimate $\widehat{C}$ is a function of the data and hence is random.
- An interval estimate $\widehat{C}$ is called a $1-\alpha$ confidence interval when $\Pr\left(\boldsymbol{\beta} \in \widehat{C}\right) = 1-\alpha$.
- A good choice for a confidence interval is by adding and subtracting from the estimate $\hat{\beta}$ a fixed multiple of the standard error:

$$\widehat{C} = \left[\hat{\beta} - c \cdot s\left(\hat{\beta}\right), \hat{\beta} + c \cdot s\left(\hat{\beta}\right)\right].$$

- $\widehat{C}$ is the set of parameter values for β such that the t-statistic $T(\beta)$ is smaller than some constant c:

$$\widehat{C} = [\beta : |T(\beta)| \leq c] = \left\{\beta : -c \leq \frac{\hat{\beta} - \beta}{s(\hat{\beta})} \leq c\right\}.$$

- The coverage probability is

$$\begin{aligned}\Pr\left(\beta \in \widehat{C}\right) &= \Pr(|T(\beta)| \leq c) \\ &= \Pr(-c \leq T(\beta) \leq c) \\ &= 2 \cdot F(c) - 1\end{aligned}$$

where F is the t distribution with $n - k$ degrees of freedom ($F(-c) = 1 - F(c)$).

Theorem

In the normal regression model, $\widehat{C}$ with $c = F^{-1}(1 - \alpha/2)$ has coverage probability $\Pr\left(\beta \in \widehat{C}\right) = 1 - \alpha$.

Hypothesis Testing

- Let $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ be a parameter of interest. Some examples of $\boldsymbol{\theta}$ include:
  - The coefficient of one of the regressors: $\theta = \beta_1$, $d = 1$, $\Theta = \mathbb{R}$.
  - A vector of coefficients: $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_l)'$, $d = l$, $\Theta = \mathbb{R}^l$.
  - The variance of errors: $\theta = \sigma^2$, $d = 1$, $\Theta = (0, \infty)$.
- A statistical hypothesis is an assertion about $\boldsymbol{\theta}$. Usually, we have two competing hypotheses, and we want to draw a conclusion, based on the data, as to which of the hypotheses is true. Let $\Theta_0 \subset \Theta$ and $\Theta_1 \subset \Theta$ such that $\Theta_0 \cap \Theta_1 = \varnothing$ and $\Theta_0 \cup \Theta_1 = \Theta$. The two competing hypotheses are:
  - Null hypothesis $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$. This is a hypothesis that is held as true, unless data provides sufficient evidence against it.
  - Alternative hypothesis $\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_1$. This is a hypothesis against which the null is tested. It is held to be true if the null is found false.

- The subsets Θ_0 and Θ_1 are chosen by the econometrician and therefore are known. Their union defines the maintained hypothesis, i.e. the space of values that $\boldsymbol{\theta}$ can take. For example, when $\Theta = \mathbb{R}$, one may consider $\Theta_0 = \{0\}$, and $\Theta_1 = \mathbb{R} \setminus \{0\}$. Another example is $\Theta_0 = (-\infty, 0]$ and $\Theta_1 = (0, \infty)$.
- When Θ_0 has exactly one element (Θ_0 is a singleton), we say that $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$ is a simple hypothesis. Otherwise, we say that $\mathbb{H}_0$ is a composite hypothesis. Similarly, $\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_1$ can be simple or composite depending on whether Θ_1 is a singleton or not.
- Let $S \in \mathcal{S}$ denote a statistic and the range of its values. A decision rule is defined by a partition of $\mathcal{S}$ into acceptance region $\mathcal{A}$ and rejection (critical) region $\mathcal{R}$ ($\mathcal{A} \cap \mathcal{R} = \emptyset$ and$\mathcal{A} \cup \mathcal{R} = \mathcal{S}$).
- $\mathbb{H}_0$ is rejected when the test statistic falls in to the rejection region $\mathcal{R}$.

Type I and Type II Errors

- There are two types of errors that the econometrician can make:

		Truth	
		$\mathbb{H}_0$	$\mathbb{H}_1$
Decision	$\mathbb{H}_0$	✓	Type II error
	$\mathbb{H}_1$	Type I error	✓

- Type I error is the error of rejecting $\mathbb{H}_0$ when $\mathbb{H}_0$ is true.
- Type II error is the error of accepting $\mathbb{H}_0$ when $\mathbb{H}_1$ is true.

Power Function

- The probabilities of Type I and II errors can be described using the power function.
- Consider a test based on S that rejects $\mathbb{H}_0$ when $S \in \mathcal{R}$. The power function of this test is defined as:

$$\pi(\boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}\left(S \in \mathcal{R}\right),$$

  where $\Pr_{\boldsymbol{\theta}}(\cdot)$ denotes that the probability must be calculated under the assumption that the true value of the parameter is $\boldsymbol{\theta}$.
- The largest probability of Type I error (rejecting $\mathbb{H}_0$ when it is true) is

$$\sup_{\boldsymbol{\theta}\in\Theta_0} \pi(\boldsymbol{\theta}) = \sup_{\boldsymbol{\theta}\in\Theta_0} \Pr_{\boldsymbol{\theta}}\left(S \in \mathcal{R}\right).$$

  The expression above is also called the *size* of a test.
- When $\mathbb{H}_0$ is simple, i.e. $\Theta = \{\boldsymbol{\theta}_0\}$, the size can be computed simply as $\pi(\boldsymbol{\theta}_0) = \Pr_{\boldsymbol{\theta}_0}(S \in \mathcal{R})$.

- The probability of Type II error is:

$$1 - \pi(\boldsymbol{\theta}) = 1 - \Pr_{\boldsymbol{\theta}}(S \in \mathcal{R}) \quad \text{for } \boldsymbol{\theta} \in \Theta_1.$$

- Typically, Θ_1 has many elements, and therefore the probability of Type II error depends on the true value $\boldsymbol{\theta}$.
- One would like to have the probabilities of Type I and II errors to be as small as possible, but unfortunately, they are inversely related.
- To reduce the probability of Type I error, one should make $\mathcal{R}$ smaller. This, however, will increase the probability of Type II error.

Definition

A test with power function $\pi(\boldsymbol{\theta})$ is said to be a level α test if $\sup_{\boldsymbol{\theta}\in\Theta_0} \pi(\boldsymbol{\theta}) \leq \alpha$. We say it is a size α test if $\sup_{\boldsymbol{\theta}\in\Theta_0} \pi(\boldsymbol{\theta}) = \alpha$.

- Significance level of a test us the largest Type I error probability one tolerates. Typically, the significance level is chosen to be a small number close to zero: for example, $0.01, 0.05, 0.10$.
- By convention, a valid test must control the probability of Type I error (level α test, where α is equal to the significance level).
- We want the probability of a Type II error probability to be as small as possible for given Type I error probability.

Steps of Hypothesis Testing

1. Specify $\mathbb{H}_0$ and $\mathbb{H}_1$.
2. Choose the significance level α.
3. Define a decision rule (a test statistic S and a rejection region $\mathcal{R}_\alpha$) so that the resulting test is a level α test. Note that $\mathcal{R}_\alpha$ typically depends on α.
4. Perform the test.

p-Value

- The lowest significance level consistent with rejecting $\mathbb{H}_0$ is called the p-value:

$$p\text{-value}=\min\left\{0 < \alpha < 1 : S \in \mathcal{R}_\alpha\right\}.$$

- Note that p-value is a statistic and a measure of the evidence against $\mathbb{H}_0$.
- If the p-value is smaller than our tolerance(significance level), then we reject $\mathbb{H}_0$.

Power of a Test

- The power of a test with the power function $\pi(\boldsymbol{\theta})$ is defined as

$$\pi(\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in \Theta_1.$$

- Given two level α tests, we should prefer a more powerful test.
- We say that a level α test with power function $\pi_1(\boldsymbol{\theta})$ is uniformly more powerful than a level α test with power function $\pi_2(\boldsymbol{\theta})$ if $\pi_1(\boldsymbol{\theta}) \geq \pi_2(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta_1$.

A Simple Example

- Consider a sample $(X_1, X_2, ..., X_n)$ from a normal population with mean μ and variance σ^2. Suppose we know σ^2 for now. Consider $\mathbb{H}_0 : \mu = 0$ against $\mathbb{H}_1 : \mu > 0$.
- Consider $T = \frac{\overline{X}}{\sigma/\sqrt{n}}$, where $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$, which is a $\mathrm{N}(0, 1)$ random variable under $\mathbb{H}_0$.
- Consider $\mathcal{R}_\alpha = [z_{1-\alpha}, \infty)$, where $\Pr(\mathrm{N}(0, 1) > z_{1-\alpha}) = \alpha$. Under $\mathbb{H}_0$, $\Pr[T \in \mathcal{R}_\alpha] = \alpha$.
- For any given value of μ, define

$$T_\mu = T - \frac{\mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

  which is always $\mathrm{N}(0, 1)$ if the true mean is μ.
- We reject the test when $T \geq z_{1-\alpha}$, which holds if and only if

$$T_\mu = T - \frac{\mu}{\sigma/\sqrt{n}} \geq z_{1-\alpha} - \frac{\mu}{\sigma/\sqrt{n}}.$$

  It follows for this simple example, that the power function is

$$\pi(\mu) = 1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{n}\mu}{\sigma}\right).$$

- We notice that $\pi(\mu)$ is smaller for all μ if the significance level α is smaller (and hence $z_{1-\alpha}$ is larger). This reflects the trade-off between Type I error and Type II error probabilities: we cannot reduce both simultaneously.
- $\pi(\mu)$ is increasing in μ. For μ's that are farther away from 0, the test can detect such deviation at a higher probability.
- As $\mu \to \infty$, the power converges to 1. The test is very likely to reject $\mathbb{H}_0$ if the true mean is very large.
- $\pi(\mu)$ increases with the sample size n. The test can detect falseness of $\mathbb{H}_0$ at a higher probability if our sample contains more information.

t Test

- The null hypothesis:

$$\mathbb{H}_0 : \beta_j = \beta_{j,0}.$$

- The alternative hypothesis:

$$\mathbb{H}_1 : \beta_j \neq \beta_{j,0}.$$

- The standard testing statistic is

$$|T| = \left| \frac{\hat{\beta}_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)} \right|.$$

- If $\mathbb{H}_0$ is true, we expect $|T|$ to be small, but if $\mathbb{H}_1$ is true, then we would expect $|T|$ to be large. Hence the standard rule is to reject $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ for large values of the t-statistic $|T|$:

$$\text{Reject } \mathbb{H}_0 \text{ if } |T| > c.$$

- c is called the critical value. Its value is selected to control the probability of false rejections.
- When the null hypothesis is true, T has an exact student distribution. The probability of false rejection is

$$\begin{aligned}\Pr(\text{Reject } \mathbb{H}_0 \mid \mathbb{H}_0) &= \Pr(|T| > c \mid \mathbb{H}_0) \\ &= \Pr(T > c \mid \mathbb{H}_0) + \Pr(T < -c \mid \mathbb{H}_0) \\ &= 1 - F(c) + F(-c) \\ &= 2(1 - F(c)).\end{aligned}$$

- We select the value c so that this probability equals the significance level: $F(c) = 1 - \alpha/2$.
- The p-value of a t-statistic is $p = 2(1 - F(|T|))$.

Theorem

In the normal regression model, if the null hypothesis is true, then $|T| \sim t_{n-k}$. If c is set so that $\Pr(|t_{n-k}| \geq c) = \alpha$, then the test "Reject $\mathbb{H}_0$ in favor of $\mathbb{H}_1$ if $|T| > c$" has level α.

Power

- Assume that the true value is given by β_j. Assume for simplicity that σ^2 is known, so that $s\left(\hat{\beta}_j\right) = \sqrt{\sigma^2\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]_{jj}}$.
- Write

$$T = \frac{\hat{\beta}_j - \beta_j}{s\left(\hat{\beta}_j\right)} + \frac{\beta_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)}. \tag{1}$$

- We have that

$$Z = \frac{\hat{\beta}_j - \beta_j}{s\left(\hat{\beta}_j\right)} \mid \boldsymbol{X} \sim \mathrm{N}(0,1)$$

  and

$$\frac{\hat{\beta}_j - \beta_j}{s\left(\hat{\beta}_j\right)} + \frac{\beta_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)} \mid \boldsymbol{X} \sim \mathrm{N}\left(\frac{\beta_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)}, 1\right).$$

- The critical value is $c = \Phi^{-1}(1 - \alpha/2)$ $(\Pr(Z > c) = \alpha/2)$.

- If the null hypothesis is false, the distribution of the test statistic is not centered around zero, and we will see rejection rates higher than α. The probability to reject is a function of the true value β_j and depends on the magnitude of $\left|\beta_j - \beta_{j,0}\right| / s\left(\hat{\beta}_j\right)$.
- Now

$$\pi\left(\beta_1\right) = \Pr\left(\left|\frac{\hat{\beta}_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)}\right| > c\right) = \Pr\left(\left|\frac{\hat{\beta}_j - \beta_j + \beta_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)}\right| > c\right)$$

$$= \Pr\left(\left|Z + \frac{\beta_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)}\right| > c\right).$$

One-sided Test

- In the case of one-sided tests, the null and alternative hypotheses may be specified as

$$\begin{aligned}\mathbb{H}_0 &: \beta_j \leq \beta_{j,0}, \\ \mathbb{H}_1 &: \beta_j > \beta_{j,0}.\end{aligned}$$

- In this case, a valid test should satisfy the following condition:

$$\sup_{\beta_j \leq \beta_{j,0}} \Pr\left(\text{reject } \mathbb{H}_0 \mid \beta_j\right) \leq \alpha, \tag{2}$$

  i.e. the maximum probability to reject H_0 when it is true should not exceed α.

- Consider the following test (decision rule):

$$\text{Reject } \mathbb{H}_0 \text{ when } T > c.$$

  where c is set so that $\Pr\left(t_{n-k} \geq c\right) = \alpha$.

- Under $\mathbb{H}_0$, we have:

$$
\begin{aligned}
\Pr\left(\text{reject } \mathbb{H}_0 \mid \beta_j \leq \beta_{j,0}\right) &= \Pr\left(T > c \mid \beta_j \leq \beta_{j,0}\right) \\
&= \Pr\left(\frac{\hat{\beta}_j - \beta_{j,0}}{s\left(\hat{\beta}_j\right)} > c \mid \beta_j \leq \beta_{j,0}\right) \\
&\leq \Pr\left(\frac{\hat{\beta}_j - \beta_j}{s\left(\hat{\beta}_j\right)} > c \mid \beta_j \leq \beta_{j,0}\right) \\
&\quad \text{since } \beta_j \leq \beta_{j,0} \\
&= \alpha \text{ (since } \frac{\hat{\beta}_j - \beta_j}{s\left(\hat{\beta}_j\right)} \sim t_{n-k}\text{)}.
\end{aligned}
$$

- The size control condition is satisfied.

Testing a Single Linear Restriction

- Suppose we want to test

$$\begin{aligned} \mathbb{H}_0 &: \quad \boldsymbol{c}'\boldsymbol{\beta} = r, \\ \mathbb{H}_1 &: \quad \boldsymbol{c}'\boldsymbol{\beta} \neq r. \end{aligned}$$

- In this case, $\boldsymbol{c}$ is a k-vector, r is a scalar, and under the null hypothesis

$$c_1\beta_1 + \ldots + c_k\beta_k - r = 0.$$

- We have that the LS estimator of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right).$$

Then, under H_0,

$$\frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - r}{\sqrt{\sigma^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} \mid \boldsymbol{X} \sim \mathrm{N}\left(0, 1\right).$$

- The t-statistic

$$\begin{aligned} T &= \frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - r}{\sqrt{s^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} \\ &= \left(\frac{\boldsymbol{c}'\hat{\boldsymbol{\beta}} - r}{\sqrt{\sigma^2 \boldsymbol{c}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{c}}} \right) \Big/ \sqrt{\frac{\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}}{\sigma^2} / (n-k)}. \end{aligned}$$

- $\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}/\sigma^2 \mid \boldsymbol{X} \sim \chi^2_{n-k}$ and is independent of $\hat{\boldsymbol{\beta}}$. Therefore, under $\mathbb{H}_0$, $T \mid \boldsymbol{X} \sim t_{n-k}$.
- The significance level α two-sided test of $\mathbb{H}_0 : \boldsymbol{c}'\boldsymbol{\beta} = r$ is given by "reject $\mathbb{H}_0$ if $|T| > c$", where $\Pr\left(|t_{n-k}| > c\right) = \alpha$.

Testing Multiple Linear Restrictions

- Suppose we want to test

$$\begin{aligned} \mathbb{H}_0 &: \boldsymbol{R\beta} = \boldsymbol{r}, \\ \mathbb{H}_1 &: \boldsymbol{R\beta} \neq \boldsymbol{r}, \end{aligned}$$

  where $\boldsymbol{R}$ is a $q \times k$ matrix and $\boldsymbol{r}$ is a q-vector.
- $\boldsymbol{R} = \boldsymbol{I}_k$, $\boldsymbol{r} = \boldsymbol{0}$. In this case, we test that $\beta_1 = \ldots = \beta_k = 0$.
- $R = \begin{pmatrix} 1 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}$, $r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. In this case, $H_0 : \beta_1 + \beta_2 = 1, \beta_3 = 0$.
- Consider the F-statistic

$$F = \frac{(RSS_r - RSS_{ur})/q}{RSS_{ur}/(n-k)}.$$

  - RSS_r: the restricted Residual Sum of Squares.
  - RSS_{ur}: the unrestricted Residual Sum of Squares.

- Consider the restricted problem

$$\min_{\boldsymbol{b}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})' (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}) \quad \text{subject to } \boldsymbol{R}\boldsymbol{b} = \boldsymbol{r}.$$

- A Lagrangian function for this problem is

$$L(\boldsymbol{b}, \boldsymbol{\lambda}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})' (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}) + 2\boldsymbol{\lambda}' (\boldsymbol{R}\boldsymbol{b} - \boldsymbol{r}),$$

  where $\boldsymbol{\lambda}$ is a q-vector.
- Let $\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\lambda}}$ be the solution, where $\tilde{\boldsymbol{\beta}}$ is the restricted LS estimator. It has to satisfy the first-order conditions

$$\frac{\partial L\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\lambda}}\right)}{\partial \boldsymbol{b}} = 2\boldsymbol{X}'\boldsymbol{X}\tilde{\boldsymbol{\beta}} - 2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{R}'\tilde{\boldsymbol{\lambda}} = \mathbf{0}, \qquad (3)$$

$$\frac{\partial L\left(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\lambda}}\right)}{\partial \boldsymbol{\lambda}} = \boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r} = \mathbf{0}. \qquad (4)$$

- The restricted LS estimator is given by

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R}' \left(\boldsymbol{R} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{R}'\right)^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - r\right),$$

  where $\hat{\boldsymbol{\beta}}$ is the LS estimator without the restriction $\boldsymbol{R}\boldsymbol{b} = \boldsymbol{r}$.

- Define the restricted residuals

$$
\begin{aligned}
\tilde{\boldsymbol{e}} &= \boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} \\
&= \left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) + \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\left(\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - r\right) \\
&= \hat{\boldsymbol{e}} + \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\left(\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - r\right),
\end{aligned}
$$

- Then,

$$
\begin{aligned}
RSS_r &= \tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}} \\
&= \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} + \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
&\quad + 2\hat{\boldsymbol{e}}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\left(\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \\
&= RSS_{ur} + \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right).
\end{aligned}
$$

- Since $s^2 = \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}/(n-k) = RSS_{ur}/(n-k)$,

$$
F = \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(s^2\boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right)/q.
$$

- We show next that under $\mathbb{H}_0$, $F \mid \boldsymbol{X} \sim F_{q,n-k}$. First,

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim \mathrm{N}\left(\boldsymbol{R}\boldsymbol{\beta}, \sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right).$$

- Then, under $\mathbb{H}_0$,

$$\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r} \mid \boldsymbol{X} \sim \mathrm{N}\left(0, \sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right).$$

  It follows that

$$\left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right)' \left(\sigma^2 \boldsymbol{R}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{R}'\right)^{-1} \left(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \mid \boldsymbol{X} \sim \chi^2_q.$$

  The result follows from $\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}/\sigma^2 \mid \boldsymbol{X} \sim \chi^2_{n-k}$ and independent of $\hat{\boldsymbol{\beta}}$ and the definition of F-distribution.
- Therefore, the test is given by "reject $\mathbb{H}_0$ if $F > c$", where $\Pr\left(F_{q,n-k} > c\right) = \alpha$.

Test of Model Significance

- Consider a model with the intercept

$$Y_i = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + U_i,$$

- Consider the null hypothesis $H_0 : \beta_2 = \ldots \beta_k = 0$. The restricted model is given by

$$Y_i = \beta_1 + U_i.$$

- In this case, the restricted LS estimator is $\tilde{\beta}_1 = n^{-1} \sum_{i=1}^n Y_i = \overline{Y}$, and $RSS_r = TSS = \sum_{i=1}^n \left(Y_i - \overline{Y}\right)^2$. In this case,

$$\begin{aligned} F &= \frac{(TSS - RSS_{ur})/(k-1)}{RSS_{ur}/(n-k)} \\ &= \frac{ESS/(k-1)}{RSS_{ur}/(n-k)} \\ &= \frac{R^2/(k-1)}{\left(1 - R^2\right)/(n-k)} \\ &\sim F_{k-1,n-k}. \end{aligned}$$