Advanced Econometrics

Lecture 5: Normal Regression and Maximum Likelihood (Hansen Chapter 5)

Instructor: Ma, Jun

Renmin University of China

October 30, 2018

正态回归模型

Normal Regression Model

极大似然估计首先假设

误差项是正态分布的.

目的是构造出一个关于β

的置信区间.

且还假设了同方差.

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

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

抽样回归.

假设误差项正态分布

$Y_i = X_i'\beta + e_i$

$e_i | X_i \sim N(0, \sigma^2)$

$\Downarrow$

e与X相互独立.

相当于假设 ① 同方差

② e关于X独立

那么 $Y_i | X_i \sim N(X_i'\beta, \sigma^2)$

(Y_i, X_i) i.i.d. $i = 1, 2, \cdots, n$

- 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.

所有方差-协方差矩阵一定是半正定的.

Normal Regression Model

$\Rightarrow \underset{n\times 1}{Y} \mid \underset{n\times k}{X} \sim N(X\beta, \sigma^2 I_n)$

- 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 $\boldsymbol{X}$:

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

似然函数.

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

Normal Regression Model

- Work with the natural logarithm:

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

log 似然函数，

其最大点就是极大似然估计.

- The MLE:

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

Normal Regression Model

- 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:

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

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

极大似然估计是一阶条件的解

Normal Regression Model

- The MLE:

极大似然中最重要的估计

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

可以看到最小二乘估计和极大似然估计相等.

- The MLE for σ^2:

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

有偏的

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

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

$$\log(\hat{\sigma}^2_{mle}) = \log\left(\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2\right)$$
$$= \log\sum_{i=1}^{n}\hat{e}_i^2 - \log n$$
$$= \log SSR - \log n$$

Distribution of OLS Coefficient Vector

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

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

※记住结论就行.

- The OLS estimator satisfies

正态变量的线性变换 ←

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

which is a linear function of $\boldsymbol{e}$.

- Conditional on $\boldsymbol{X}$,

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

or

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

$e|X \sim$ normal

$\Rightarrow Y|X \sim$ normal

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

$\Rightarrow \hat{\beta}|X \sim$ normal

$\Rightarrow \hat{\beta}|X \sim N(E(\hat{\beta}|X), Var(\hat{\beta}|X))$

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

Distribution of OLS Coefficient Vector

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

Theorem

In the linear regression model,

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

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

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

$$\xi_j = (0, 0, \overset{\overset{j\text{-th}}{\uparrow}}{1}, 0, \cdots, 0)$$

$$\hat{\beta}_j = \xi_j' \beta$$

$$Var(\hat{\beta}_j | X) = Var(\xi_j' \hat{\beta} | X)$$

$$= \xi_j' \, Var(\hat{\beta} | X) \, \xi_j$$

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

Distribution of OLS Residual Vector

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

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

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

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

- So

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

e|X ~ normal

$\begin{pmatrix} (X'X)^{-1}X' \\ M \end{pmatrix} e \Rightarrow \begin{pmatrix} (X'X)^{-1}X' \\ M \end{pmatrix} e|_X$

~ normal

Var

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

Distribution of Variance Estimate

$s^2 = \frac{1}{n-k}\hat{e}'\hat{e} = \frac{1}{n-k}e'Me$

- $s^2 = \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}/\left(n-k\right) = \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}/\left(n-k\right)$.
- 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:

$$\boldsymbol{\Lambda} = \left[\begin{array}{cc} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{array}\right].$$

$U = H'e \sim \text{normal}$

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

$\mathrm{Var}(H'e) = H'\,\mathrm{Var}(e)\,H$

$= \sigma^2 H'H = \sigma^2 I_n$

$$\begin{aligned}
\left(n-k\right)s^2 &= \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} \\
&= \boldsymbol{e}'\boldsymbol{H}\left[\begin{array}{cc} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{array}\right]\boldsymbol{H}'\boldsymbol{e} \\
&= \boldsymbol{U}'\left[\begin{array}{cc} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{array}\right]\boldsymbol{U} \\
&\sim \sigma^2\chi^2_{n-k}.
\end{aligned}$$

Distribution of Variance Estimate

Theorem

In the linear regression model,

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

自由度为n-k的卡方分布

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

与X相互独立，可以理解成条件分布

也可以理解成非条件分布

T

t-statistic

- The "z-statistic":

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

真实的抽样方差→

Z统计量与X相互独立.

$Z \sim N(0,1)$ 可以是条件分布也可以是非条件分布.

- 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:

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

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{[\hat{V}_0^W]_{jj}}} \quad vs. \quad \frac{\hat{\beta}_j - \beta_j}{\sqrt{[s^2(X'X)^{-1}]_{jj}}} \sim t_{n-k}$$

方差估计是异方差稳健的,

不稳健的

t-statistic

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.

置信区间

Confidence Intervals for Regression Coefficients

- An OLS estimate $\widehat{\boldsymbol{\beta}}$ 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 $\widehat{\beta}$ a fixed multiple of the standard error:

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

置信区间的上下界是随机的.

← 这个区间关于 $\hat{\beta}$ 是对称的. c 是待定的.

Confidence Intervals for Regression Coefficients

置信区间只考一个参数.

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

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

$$\Pr\left(\underbrace{\left| \frac{\hat{\beta}-\beta}{s.e.(\hat{\beta})} \right|}_{\sim t_{n-k}} < c \right) = 1-\alpha$$

- The coverage probability is

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

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

$\Rightarrow 2F(c) - 1 = 1-\alpha$

F 是自由度为 $n-k$ 的 t 分布的累积分布函数.

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

假设检验

(期中不考)

t Test

- The **null hypothesis**:

原假设 $$\mathbb{H}_0 : \beta = \beta_0.$$

- The alternative hypothesis:

备择假设 $$\mathbb{H}_1 : \beta \neq \beta_0.$$

- The standard testing statistic is

$$|T| = \left| \frac{\hat{\beta} - \beta_0}{s\left(\hat{\beta}\right)} \right|.$$

- 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 $|T|$:

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

t Test

c是临界值.

- 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

一类错误概率.

即H₀为真时拒绝了H₀.

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

- We select the value c so that this probability equals the significance level: $F(c) = 1 - \alpha/2$.

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 significance level α.

t Test

P值不是一个数.是一个统计量.
一个随机变量.

- The p-value of a t-statistic is $p = 2\left(1 - F\left(|T|\right)\right)$.
- p-value is a statistic, and is random, and is a measure of the evidence against $\mathbb{H}_0$.

Likelihood Ratio Test (F Test)

- The regression model:

$$Y_i = \boldsymbol{X}_{1i}'\boldsymbol{\beta}_1 + \boldsymbol{X}_{2i}'\boldsymbol{\beta}_2 + e_i.$$

  Let $k = \dim(\boldsymbol{X}_i)$, $k_1 = \dim(\boldsymbol{X}_{1i})$ and $q = \dim(\boldsymbol{X}_{2i})$ so $k = k_1 + q$.

- The hypothesis is

$$\mathbb{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}.$$

$\beta_1 = 0$ & $\beta_2 = 0$

- If $\mathbb{H}_0$ is true, then the regressors $\boldsymbol{X}_{2i}$ can be omitted from the regression:

$$Y_i = \boldsymbol{X}_{1i}'\boldsymbol{\beta}_1 + e_i.$$

- The alternative hypothesis is that at least one element of $\boldsymbol{\beta}_2$ is non-zero:

$$\mathbb{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}.$$

至少有一个值非零.

Likelihood Ratio Test (F Test)

- The "unconstrained" maximized likelihood:

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

- The MLE of the "constrained" model:

$$\log\widetilde{L}\left(\boldsymbol{\beta}_1, \sigma^2\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}'_{1i}\boldsymbol{\beta}_1\right)^2.$$

- The "constrained" MLE:

$$\left(\widetilde{\boldsymbol{\beta}}_1, \widetilde{\sigma}^2\right) = \underset{\boldsymbol{\beta}\in\mathbb{R}^{k_1}, \sigma^2>0}{\operatorname{argmax}} \log\widetilde{L}\left(\boldsymbol{\beta}, \sigma^2\right).$$

- The MLE is the OLS of Y_i on $\boldsymbol{X}_{1i}$:

$$\widetilde{\boldsymbol{\beta}}_1 = \left(\boldsymbol{X}'_1\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}'_1\boldsymbol{Y}$$

with

$$\widetilde{e}_i = Y_i - \boldsymbol{X}'_{1i}\widetilde{\boldsymbol{\beta}}_1 \text{ and } \widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widetilde{e}_i.$$

Likelihood Ratio Test (F Test)

- The constrained maximized log-likelihood:

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

- The likelihood ratio test rejects for large values of LR:

$$\begin{aligned} LR =& 2\left(\left(-\frac{n}{2}\log\left(2\pi\hat{\sigma}^2\right) - \frac{n}{2}\right) - \left(-\frac{n}{2}\log\left(2\pi\widetilde{\sigma}^2\right) - \frac{n}{2}\right)\right) \\ =& n\log\left(\frac{\widetilde{\sigma}^2}{\hat{\sigma}^2}\right). \end{aligned}$$

Likelihood Ratio Test (F Test)

- LR is large if and only if the F statistic is large:

$$F = \frac{\left(\widetilde{\sigma}^2 - \hat{\sigma}^2\right)/q}{\hat{\sigma}^2/(n-k)}.$$

- Under $\mathbb{H}_0$, the F statistic has an exact F distribution:

$$F = \frac{\boldsymbol{e}'\left(\boldsymbol{M}_1 - \boldsymbol{M}\right)\boldsymbol{e}/q}{\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}/(n-k)} \sim \frac{\chi_q^2/q}{\chi_{n-k}^2/(n-k)} \sim F_{q,n-k}.$$

Theorem

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