3.5.3 Partitioning the Total Sum of Squares

The total sum of squares, which we shall call SST, is

$$SST = \boldsymbol{Y}^\top \boldsymbol{Y} = \sum_{i=1}^{n} y_i^2,$$

where $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)^\top \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$. It can be easily shown that

$$\frac{\boldsymbol{Y}^\top \boldsymbol{Y}}{\sigma^2} \sim \chi^2_{(n, \frac{(\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta})}{2\sigma^2})}.$$

Recall that the sum of squares of deviations of observed $y_i's$ from their predicted values is $SSE = \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}^\top \boldsymbol{Y} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y}$. The difference

$$\begin{aligned} \text{SSR} &= \text{SST} - \text{SSE} \\ &= \boldsymbol{Y}^\top \boldsymbol{Y} - (\boldsymbol{Y}^\top \boldsymbol{Y} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y}) \\ &= \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y} = \boldsymbol{Y}^\top \boldsymbol{H} \boldsymbol{Y} \end{aligned}$$

represents that portion of SST that is attributed to having fitted the regression, and hence it is called the *sum of squares due to regression*, SSR. It is also often called the *reduction in sum of squares*. Similarly, we can show

$$\frac{SSR}{\sigma^2} \sim \chi^2_{(r(\boldsymbol{X}), \frac{(\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta})}{2\sigma^2})}.$$

The partitioning of SST can be summarized in a manner that serves as a foundation for developing the traditional analysis of variance (ANOVA) table.

Table 1: Traditional ANOVA table.

Source	df	SS	MS	F-statistics
Regression	$r(\boldsymbol{X})$	$\hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y}$	$\frac{SSR}{r(\boldsymbol{X})}$	$F(R) = \frac{MSR}{MSE}$
Error	$n - r(\boldsymbol{X})$	$\boldsymbol{Y}^\top \boldsymbol{Y} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y}$	$\frac{SSE}{n - r(\boldsymbol{X})}$	
Total	n	$\boldsymbol{Y}^\top \boldsymbol{Y}$		

Note: Since $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$ and

$$\begin{aligned} \text{SSR} &= \boldsymbol{Y}^\top \boldsymbol{H} \boldsymbol{Y} \quad (\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top), \\ \text{SSE} &= \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{Y}, \end{aligned}$$

we have $\boldsymbol{H}(\sigma^2 \boldsymbol{I})(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$,

$$\Rightarrow \quad \text{SSR is independent of SSE}$$

$$\Rightarrow \quad F(R) \sim F_{[r(\boldsymbol{X}), n - r(\boldsymbol{X}), \frac{\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}}{2\sigma^2}]}.$$

Thus under $H_0 : \boldsymbol{\beta}_0 = \boldsymbol{0}$,

$$F(R) \sim F_{[r(\boldsymbol{X}), n - r(\boldsymbol{X}), 0]}.$$

Next, suppose the model had no predictors in it but had simply been $y_i = c_0 + \varepsilon_i \quad i = 1, \cdots, n$. Then, the least square estimate of c_0 is

$$\hat{c}_0 = (\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \boldsymbol{Y} = \frac{1}{n} \mathbf{1}^\top \boldsymbol{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i \equiv \bar{y}.$$

The SSM is defined to be the SSR for the linear model without x. Thus,

$$\text{SSM} = \hat{c}_0 \mathbf{1}^\top \boldsymbol{Y} = \frac{1}{n} \boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y} = n\bar{y}^2.$$

$$\frac{\text{SSM}}{\sigma^2} = \boldsymbol{Y}^T \left(\frac{1}{n\sigma^2} \mathbf{1}\mathbf{1}^T \right) \boldsymbol{Y}$$

Define $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ as

$$\boldsymbol{A} \triangleq \frac{1}{n\sigma^2} \mathbf{1}\mathbf{1}^T$$

and

$$\boldsymbol{\Sigma} \triangleq \sigma^2 \boldsymbol{I}$$

Note that

$$\boldsymbol{A}\boldsymbol{\Sigma} = \left(\frac{1}{n\sigma^2} \mathbf{1}\mathbf{1}^T \right) (\sigma^2 \boldsymbol{I}) = \frac{1}{n} \mathbf{1}\mathbf{1}^T.$$

Check that

$$(\boldsymbol{A}\boldsymbol{\Sigma})^2 = \left(\frac{1}{n} \mathbf{1}\mathbf{1}^T \right) \left(\frac{1}{n} \mathbf{1}\mathbf{1}^T \right) = \frac{1}{n^2} \mathbf{1}(\mathbf{1}^T \mathbf{1}) \mathbf{1}^T = \frac{1}{n^2} \mathbf{1}(n) \mathbf{1}^T = \frac{1}{n} \mathbf{1}\mathbf{1}^T = \boldsymbol{A}\boldsymbol{\Sigma},$$

which is idempotent. It follows that

$$\frac{SSM}{\sigma^2} \sim \chi^2_{(1, \frac{(\mathbf{1}^\top \boldsymbol{x}\boldsymbol{\beta})^2}{2n\sigma^2})}.$$

The *total sum of squares corrected for the mean* is

$$\text{SST}_m = SST - SSM = \boldsymbol{Y}^\top \boldsymbol{Y} - \boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{I} - \frac{\mathbf{1}}{\boldsymbol{n}} \mathbf{1}^\top) \boldsymbol{Y}.$$

Then, we can show that

$$\frac{SST_m}{\sigma^2} \sim \chi^2_{(n-1, \frac{\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} - \frac{1}{n}(\mathbf{1}^\top \boldsymbol{X}\boldsymbol{\beta})^2}{2\sigma^2})}.$$

Similarly, the *sum of squares of regression corrected for the mean* is

Centered and excluded the 1 columns

$$\text{SSR}_m = SSR - SSM = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y} - \boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y} = \hat{\boldsymbol{b}}^\top \boldsymbol{Z}^\top \boldsymbol{Y}.$$

SS explained by the covariables only.

This is because, we can write

$$\begin{aligned} SSR = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y} &= (\hat{\beta}_0 \quad \hat{\boldsymbol{b}}) \begin{pmatrix} \mathbf{1}^\top \\ \boldsymbol{X}_1^\top \end{pmatrix} \boldsymbol{Y} \\ &= \hat{\beta}_0 \mathbf{1}^\top \boldsymbol{Y} + \hat{\boldsymbol{b}}^\top X_1^\top Y \\ &= \frac{1}{n}(\boldsymbol{Y}^\top \mathbf{1} - \bar{\boldsymbol{X}}^\top \hat{\boldsymbol{b}}) \mathbf{1}^\top \boldsymbol{Y} + \hat{\boldsymbol{b}}^\top \boldsymbol{X_1 Y} \\ &= \frac{1}{n} \boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y} + \hat{\boldsymbol{b}}^\top (\boldsymbol{X}_1^\top - \bar{\boldsymbol{X}} \mathbf{1}^\top) \boldsymbol{Y} \\ &= \frac{1}{n} \boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y} + \hat{\boldsymbol{b}}^\top \boldsymbol{Z}^\top \boldsymbol{Y} \\ &\equiv \text{SSM} + \text{SSR}_m. \end{aligned}$$

Note that $SSR_m = \hat{\boldsymbol{b}}^\top \boldsymbol{Z}^\top \boldsymbol{Y} = \hat{\boldsymbol{b}}^\top (\boldsymbol{Z}^\top \boldsymbol{Z}) \hat{\boldsymbol{b}}$ and

$$\hat{\boldsymbol{b}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{Y} \sim N[\boldsymbol{b}, (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \sigma^2],$$

thus

$$\frac{\text{SSR}_\text{m}}{\sigma^2} \sim \chi^2_{(k, \frac{\boldsymbol{b}^\top (\boldsymbol{Z}^\top \boldsymbol{Z}) \boldsymbol{b}}{2\sigma^2})}.$$

Therefore, the analysis of variance corrected for the mean for fitting linear regression can be summarized in the following table:

Table 2 ANOVA table corrected for the mean.

Source	df	SS	MS	F-statistics
Mean	1	$\frac{1}{n}\boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y}$	$\frac{SSM}{1}$	$F(M) = \frac{MSM}{MSE}$
Regression	$r(\boldsymbol{X}) - 1$	$\hat{\boldsymbol{b}}^\top \boldsymbol{Z}^\top \boldsymbol{Y}$	$\frac{SSR_m}{r(\boldsymbol{X})-1}$	$F(R) = \frac{MSR_m}{MSE}$
Error	$n - r(\boldsymbol{X})$	$\boldsymbol{Y}^\top \boldsymbol{Y} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}^\top \boldsymbol{Y}$	$\frac{SSE}{n-r(\boldsymbol{X})}$	
Total	n	$\boldsymbol{Y}^\top \boldsymbol{Y}$		

Note:

(1)

B Σ A

Since $\mathbf{1}\mathbf{1}^\top (\boldsymbol{I}\sigma^2)(\boldsymbol{I} - \boldsymbol{H}) = 0$

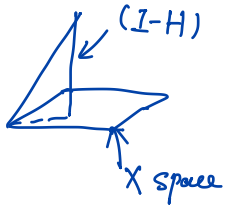
$\Rightarrow$ SSM is independent of SSE

$\Rightarrow \quad F(M) \sim F'_{[1, n-r(\boldsymbol{X}), \frac{(\mathbf{1}^\top \boldsymbol{X}\boldsymbol{\beta})^2}{2n\sigma^2}]}$

Under $H_0 : \mathbf{1}^\top \boldsymbol{X}\boldsymbol{\beta} = 0$ $\quad (H_0 : E(\bar{y}) = 0)$

$F(M) \sim F'_{[1, n-r(\boldsymbol{X}), 0]}$.

1 is in the space of X then 1 ⊥ I-H.



(2)

$$\begin{aligned}
\text{Since } \mathrm{SSR_m} &= \hat{\boldsymbol{b}}^\top \boldsymbol{Z}^\top \boldsymbol{Y} \\
&= \boldsymbol{Y}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-\mathbf{1}} \boldsymbol{Z}^\top \boldsymbol{Y} \\
\text{and } \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-\mathbf{1}} \boldsymbol{Z}^\top (\boldsymbol{I}\sigma^2)(\boldsymbol{I}-\boldsymbol{H}) &= 0, \\
\Rightarrow \quad & \mathrm{SSR_m} \text{ is independent of SSE} \\
\Rightarrow \quad & F(R_m) \sim F'_{[r(\boldsymbol{X})-1, n-r(\boldsymbol{X}), \frac{\boldsymbol{b}^\top (\boldsymbol{Z}^\top \boldsymbol{Z})\boldsymbol{b}}{2\sigma^2}]}
\end{aligned}$$

Construct a statistic that contain parameter to be tested.

$$\begin{aligned}
&\text{Under } H_0 : \boldsymbol{b} = 0, \\
&F(R_m) \sim F_{[r(\boldsymbol{X})-1, n-r(\boldsymbol{X}), 0]}.
\end{aligned}$$

(3)

$$\begin{aligned}
\mathrm{SSM} &= \frac{1}{n} \boldsymbol{Y}^\top \mathbf{1}\mathbf{1}^\top \boldsymbol{Y} \\
\mathrm{SSR_m} &= \boldsymbol{Y}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-\mathbf{1}} \boldsymbol{Z}^\top \boldsymbol{Y} \\
\text{Since } \mathbf{1}\mathbf{1}^\top (\boldsymbol{I}\sigma^2) \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-\mathbf{1}} \boldsymbol{Z}^\top &= 0,
\end{aligned}$$

(this is because $\boldsymbol{Z} = \mathbf{X_1} - \mathbf{1}\bar{\boldsymbol{X}}^\top \Rightarrow \mathbf{1}^\top \boldsymbol{Z} = n(\bar{\boldsymbol{X}}^\top - \bar{\boldsymbol{X}}^\top) = 0$)

Thus, SSM and $\mathrm{SSR_m}$ are independent.

3.5.4 General linear hypothesis

The literature of linear models abounds with discussions of different kinds of hypotheses that can be of interest in widely different fields of application. Four hypothesis of particular interest are:

1. $H_0 : \boldsymbol{\beta} = \mathbf{0}$, the hypothesis that all of the elements of $\boldsymbol{\beta}$ are zero;
2. $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0}$, the hypothesis that $\beta_i = \beta_{i0}$ for $i = 1, 2, \ldots, k$, that is, that each β_i is equal to some specified value β_{i0};
3. $H_0 : \lambda^\top \boldsymbol{\beta} = m$, that some linear combination of the elements of $\boldsymbol{\beta}$ equals a specified constant;
4. $H_0 : \boldsymbol{\beta_q} = \mathbf{0}$, that some of $\beta_i's$, q of them where $q < k$ are zero.

We will show later that all of the linear hypothesis above and others are special cases of a general procedure even though the calculation of the F-statistics may appear to differ from one hypothesis to another. The general hypothesis we consider is

$$H_0 : \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}, \quad \text{linear}$$

where $\boldsymbol{\beta}$ is the $(k+1)$-dimensional vector of parameters of the model, $\boldsymbol{K}^\top$ is any matrix of size $s \times (k+1)$ and $\boldsymbol{m}$ is a $s \times 1$ vector of specified constants. There is ONLY one restriction on $\boldsymbol{K}^\top$; $\boldsymbol{K}^\top$ *is assumed to be of full row rank.* This means that the linear function of $\boldsymbol{\beta}$ must be linearly independent. The hypothesis being tested must be made up of linearly independent functions of $\boldsymbol{\beta}$ and must contain no functions that are linear functions of others therein.

If counting the number of non-zero coef in β, H_0 will not be linear H_0.

(a) Testing linear hypothesis

Recall that in parametric linear model with normal error assumption, $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$. Then

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}, \qquad \hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \sigma^2).$$

$\Rightarrow \boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m} \sim N[\boldsymbol{K}^\top \boldsymbol{\beta} - \boldsymbol{m}, \boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K} \sigma^2]$. Next, since

$$\frac{(\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K})^{-1}}{\sigma^2} (\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}) \sigma^2 = \boldsymbol{I} \quad \text{(symmetric and idempotent)},$$

therefore, $(\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K})^{-1}$ is symmetric (why?). Thus, we define

$$Q = (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m})^\top [\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}]^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m}).$$

Then,

$$\frac{Q}{\sigma^2} \sim \chi^2_{s, \frac{1}{2\sigma^2}(\boldsymbol{K}^\top \boldsymbol{\beta} - \boldsymbol{m})^\top [\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}]^{-1} (\boldsymbol{K}^\top \boldsymbol{\beta} - \boldsymbol{m})}.$$

$\frac{Q}{\sigma^2}$: unknown. under $H_0 : \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$ $\Rightarrow 0$.

Next, we intend to show that Q and SSE are independent.

$$\begin{aligned} Q &= (\boldsymbol{K}^\top\hat{\boldsymbol{\beta}} - \boldsymbol{m})^\top[\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}^\top\hat{\boldsymbol{\beta}} - \boldsymbol{m}) \\ &= (\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} - \boldsymbol{m})^\top[\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} - \boldsymbol{m}). \end{aligned}$$

But

$$\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} - \boldsymbol{m} = \boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m}).$$

a shift of Y. $=: b$.

Therefore,

$$Q = (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m})^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}(\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m}).$$

$d^\top$ B $d^\top$

Note that if we let

$$\boldsymbol{d} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m},$$

$$\begin{aligned} Q &= \boldsymbol{d}^\top[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}(\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top]\boldsymbol{d} \\ &\equiv \boldsymbol{d}^\top\boldsymbol{B}\boldsymbol{d}, \end{aligned}$$

and

$$\boldsymbol{d} \sim N(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m}, \sigma^2\boldsymbol{I}).$$

On the other hand,

$$\begin{aligned} SSE &= \boldsymbol{Y}^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y} \\ &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m})^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{K}(\boldsymbol{K}^\top\boldsymbol{K})^{-1}\boldsymbol{m}) \\ &= \boldsymbol{d}^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{d} \\ &\equiv \boldsymbol{d}^\top\boldsymbol{A}\boldsymbol{d} \end{aligned}$$

this is because

$$\begin{aligned} \boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) &= \boldsymbol{0}, \\ (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{X} &= \boldsymbol{0}. \end{aligned}$$

Since

$$\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}[\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{0},$$

We have shown Q and SSE are independent. Finally, the test statistics

$$\begin{aligned} F(H) &= \frac{Q/s}{SSE/[N - r(\boldsymbol{X})]} \\ &= \frac{Q}{s\hat{\sigma}^2} \\ &\sim F'_{s,N-r(\boldsymbol{X}),\frac{1}{2\sigma^2}(\boldsymbol{K}^\top\boldsymbol{\beta}-\boldsymbol{m})^\top[\boldsymbol{K}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{K}]^{-1}(\boldsymbol{K}^\top\boldsymbol{\beta}-\boldsymbol{m})}, \end{aligned}$$

where $\hat{\sigma}^2 = \frac{SSE}{N-r(\boldsymbol{X})}$ is the unbiased estimator of σ^2.
Under $H_0 : \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$,

$$F(\boldsymbol{H}) \sim F'_{(s,N-r(\boldsymbol{X}))}.$$

This F-statistic is thus to be used to test the general linear hypothesis. This F-test can be shown to stem from the likelihood ratio test.

(b) Estimation of $\boldsymbol{\beta}$ under the null hypothesis

A natural question to ask when considering the null hypothesis $H_0 : \boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$ is "What is the estimator of $\boldsymbol{\beta}$ under the null hypothesis?" This might be especially pertinent following non-rejection of the hypothesis by the preceding F test. The desired estimator, denoted by $\tilde{\boldsymbol{\beta}}$, is readily obtainable using constrained least squares. Thus, when H_0 is true, $\tilde{\beta}$ is derived so as to minimize the least squares objective function subject to the constraint $\boldsymbol{K}^\top \boldsymbol{\beta} = \boldsymbol{m}$. With 2θ as a vector of Lagrange multipliers, we minimize

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + 2\theta^\top (\boldsymbol{K}^\top \boldsymbol{\beta} - \boldsymbol{m})$$

with repect to β and θ. Note that 2θ is a vector of Lagrange multipliers. After the minimization, it can be easily obtained that

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}[\boldsymbol{X}^\top \boldsymbol{Y} - \boldsymbol{K}(\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K})^{-1} (\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} - \boldsymbol{m})] \\ &= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}(\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K})^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m}),\end{aligned}$$

where $\hat{\boldsymbol{\beta}}$ is the ordinary LS estimate. Note that

$$\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}(\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K})^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m}). \qquad (*)$$

We have estimated $\boldsymbol{\beta}$ under H_0. We next show that the corresponding residual sum of squares is $SSE + Q$ where Q is the numerator sum of squares of the F-statistic used in testing the general linear hypothesis.

Without the null hypothesis (the full model),

$$SSE = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

Under the null hypothesis (the reduced model), the sum of squares of residual is

$$\begin{aligned}SSE_{H_0} &= (\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) \\ &= [\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})]^\top [\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})] \\ &= [\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})]^\top [\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})] \\ &= (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})^\top \boldsymbol{X}^\top \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}),\end{aligned}$$

cross term.

where the last equation is due to $(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta_0})^\top \boldsymbol{X}^\top (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0$. From $(*)$,

$$\begin{aligned} SSE_{H_0} &= SSE + (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m})^\top [\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}]^{-1} \\ &\qquad \times \boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K} [\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}]^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m}) \\ &= SSE + (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m})^\top [\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}]^{-1} (\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m}) \\ &\qquad \text{the non-centric } F \text{ is location parameter.} \\ &= SSE + Q \\ &\geq SSE. \end{aligned}$$

(c) Four Common Hypothesis.

In this section, we illustrate the expressions for $F(H)$ and $\tilde{\boldsymbol{\beta}}$ for four commonly occurring hypothesis.

Remember K^T

1. $H_0 : \quad \boldsymbol{\beta} = \boldsymbol{0}$, (therefore, $\boldsymbol{K}^\top = \boldsymbol{I}$, $s = k+1$, $\boldsymbol{m} = 0$)

   (1a)

   $$\begin{aligned} Q &= \boldsymbol{Y}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} ((\boldsymbol{X}^\top \boldsymbol{X})^{-1})^{-1} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \\ &= \boldsymbol{Y}^\top \boldsymbol{X} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \\ &= SSR. \end{aligned}$$

   (1b)

   $$F(\boldsymbol{H}) = \frac{SSR/r(\boldsymbol{X})}{SSE/[n - r(\boldsymbol{X})]}.$$

   (1c) Under the null hypothesis,

   $$F(\boldsymbol{H}) \sim F_{(r(\boldsymbol{X}), n - r(\boldsymbol{X}))}.$$

   (1d)

   $$\begin{aligned} \tilde{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1} ((\boldsymbol{X}^\top \boldsymbol{X})^{-1})^{-1} (\hat{\boldsymbol{\beta}}) \\ &= \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} \\ &= \boldsymbol{0}. \end{aligned}$$

2. $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_c}$ (therefore $\boldsymbol{K}^\top = \boldsymbol{I}, s = k+1, \boldsymbol{m} = \boldsymbol{\beta}_c$)

   (2a)

   $$\begin{aligned} Q &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta_c})^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta_c}) \\ &= \boldsymbol{d}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{d}. \end{aligned}$$

   (2b)

   $$F(\boldsymbol{H}) = \frac{Q}{(k+1)\hat{\sigma}^2}.$$

(2c) Under the null hypothesis,

$$F(\boldsymbol{H}) \sim F_{(k+1,n-(k+1))}.$$

(2d) $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_c) = \boldsymbol{\beta}_c$.

3. $H_0 : \boldsymbol{\lambda}^\top \boldsymbol{\beta} = m$ (therefore $\boldsymbol{K}^\top = \boldsymbol{\lambda}^\top, s = 1, \boldsymbol{m} = m$)

(3a)

$$\begin{aligned} Q &= (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} - m)^\top [\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}]^{-1} (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} - m) \\ &= (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} - m)^2 / \boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}. \end{aligned}$$

(3b)

$$F(H) = \frac{Q}{\hat{\sigma}^2}.$$

(3c) Under the null hypothesis,

$$F(H) \sim F_{(1,n-r(\boldsymbol{X}))}.$$

Note that

$$\sqrt{F(H)} = \frac{\sqrt{Q}}{\hat{\sigma}} \sim t_{n-r(\boldsymbol{X})}.$$

(3d)

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda} [\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}]^{-1} (\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} - m) \\ &= \hat{\boldsymbol{\beta}} - \frac{(\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m})}{\boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda}. \end{aligned}$$

Note that

$$\boldsymbol{\lambda}^\top \hat{\boldsymbol{\beta}} - \boldsymbol{m} \sim N(\boldsymbol{\lambda}^\top \boldsymbol{\beta} - \boldsymbol{m}, \boldsymbol{\lambda}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{\lambda} \sigma^2).$$

4. $H_0 : \boldsymbol{\beta}_q = 0$ where $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_q \\ \boldsymbol{\beta}_p \end{bmatrix}$

In this case, $\boldsymbol{K}^\top = [\boldsymbol{I}_q \quad \boldsymbol{0}]$, $p = k + 1 - q$, $\boldsymbol{m} = \boldsymbol{0}$, $s = q$.

Partition $\hat{\boldsymbol{\beta}}$ and $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ such that

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_q \\ \hat{\boldsymbol{\beta}}_p \end{bmatrix} \text{ and } (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} \boldsymbol{T}_{qq} & \boldsymbol{T}_{qp} \\ \boldsymbol{T}_{pq} & \boldsymbol{T}_{pp} \end{bmatrix}.$$

Note that $\boldsymbol{K}^\top \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{q}}$ and $[\boldsymbol{K}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{K}]^{-1} = \boldsymbol{T}_{qq}^{-1}$.

(4a) $Q = \hat{\boldsymbol{\beta}}_q^\top \boldsymbol{T}_{qq}^{-1} \hat{\boldsymbol{\beta}}_q$.

(4b) $F(H) = \frac{Q}{q\hat{\sigma}^2}$.

(4c) Under H_0, $F(H) \sim F_{(q,n-r(\boldsymbol{X}))}$.

$\tilde{\beta} = \hat{\beta}_p - T_{pq} T_{qq}^{-1} \hat{\beta}_q$.

(4d) $\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{0}_q \\ \hat{\boldsymbol{\beta}}_p - \boldsymbol{T}_{pq} \boldsymbol{T}_{qq}^{-1} \hat{\boldsymbol{\beta}}_q \end{bmatrix}$.

1. LS is not robust since it is for mean regress excluding the heavy-tail error distribution

3.6* Quantile regression and the least absolute deviation (see Koenker 2008).

Much of applied statistics may be viewed as an elaboration of the linear regression model and associated estimation methods of least squares. As a competitive alternative to the least squares regression, quantile regression (Koenker and Bassett, 1978, Econometrica) is a widely used statistical tool amongst researchers and practitioners for the modeling and inference of conditional quantile function. By allowing varying nature across different points of conditional distribution, the quantile regression model is able to accommodate heterogeneous effect in the population and is robust to the outliers of the response.

3.6.1 Quantile.

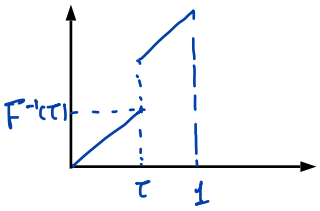
population level definition (sample level definition) ?

Any real-valued random variable Y may be characterized by its (right-continuous) distribution function

$$F(y) = P(Y \le y),$$

whereas for any $0 < \tau < 1$,

$$F^{-1}(\tau) = \inf\{y : F(y) \ge \tau\}$$



is called the τth quantile of Y. The median $F^{-1}(1/2)$ plays the central role.

say, median regress

3.6.2 Linear quantile regression.

Given a covariate vector $X \in R^{p+1}$, the τth conditional quantile function of $Y \in R$ is modeled as

$$Q_{Y|X}(\tau) = X^\top \beta_\tau, \tag{1}$$

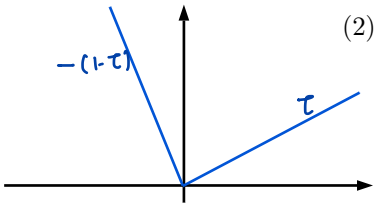
* exclude the $-\infty, +\infty$ to avoid technical problem?

for certain specific $\tau \in (0,1)$ of interest, and β_τ is $(p+1)-$vector usually including an intercept. For model (1) with complete data, a classical estimate of β_τ, denoted as $\hat{\beta}^c_\tau$, is obtained by minimizing

open interval.

$X^\top \hat{\beta}_\tau$ is the τ-th quantile of Y.

$$R(\beta_\tau) \equiv \sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \beta_\tau) \tag{2}$$



over β_τ, where (y_i, x_i) are iid copies of (Y, X), $i = 1, \ldots, n$,

$$\rho_\tau(u) = u(\tau - I(u < 0))$$

slope

is the quantile *check loss function* and $I(\cdot)$ is the indicator function. The beauty of this method rests in its simplicity and ease of inference. Recall that in least square estimator has a closed-form solution by solving the so-called normal equation. In quantile regression, we proceed similarly, but we need to exercise some caution about the differentiation step. The objective function $R(\beta_\tau)$ is piecewise linear and continuous. It is differentiable except at the points at which one or more residuals, $y_i - x_i^\top \beta_\tau$, are zero. At such points, $R(\beta_\tau)$ has directional derivatives in all directions, depending, however, on the direction of evaluation. The

$$\min_{\hat{x}} E_x \rho_\tau(X - \hat{x}) = \min_{\hat{x}} \left\{ \int_{-\infty}^{\hat{x}} (\tau-1)(x-\hat{x})\, dF(x) + \int_{\hat{x}}^{+\infty} \tau(x-\hat{x})\, dF(x) \right\}$$

only

$$\frac{\partial}{\partial \hat{x}} \downarrow = (\tau-1)\int_{-\infty}^{\hat{x}} (-1)\, dF(x) \qquad \frac{\partial}{\partial \hat{x}} \downarrow = +\tau \int_{\hat{x}}^{+\infty} (-1)\, dF(x)$$

$$-(\tau-1)\int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{+\infty} dF(x) = 0$$

$$\Rightarrow (1-\tau) F(\hat{x}) - \tau(1 - F(\hat{x})) = 0$$

$$\Rightarrow F(\hat{x}) = \tau \Rightarrow \hat{x} = \inf\{x : F(x) = \tau\}.$$

directional derivative of R in direction w is given by

$$\begin{aligned}\nabla R(\boldsymbol{\beta}_\tau, w) &\equiv \frac{d}{dt} R(\boldsymbol{\beta}_\tau + tw)|_{t=0} \\ &= \frac{d}{dt} \sum_{i=1}^{n} [y_i - x_i^\top(\boldsymbol{\beta}_\tau + tw)]\{\tau - I[y_i - x_i^\top(\boldsymbol{\beta}_\tau + tw) < 0]\}|_{t=0} \\ &= -\sum_{i=1}^{n} \psi_\tau(y_i - x_i^\top \boldsymbol{\beta}_\tau, -x_i^\top w) x_i^\top w, \end{aligned} \tag{3}$$

where

$$\psi_\tau(u, v) = \begin{cases} \tau - I(u < 0) & \text{if } u \neq 0; \\ \tau - I(v < 0) & \text{if } u = 0. \end{cases}$$

If, at a point $\hat{\boldsymbol{\beta}}_\tau$, the directional derivatives are all nonnegative (that is, $\nabla R(\boldsymbol{\beta}_\tau) \geq 0$ for all $w \in \mathbb{R}^p$ with $\|w\| = 1$), then $\hat{\boldsymbol{\beta}}_\tau$ minimizes $R(\boldsymbol{\beta}_\tau)$. This is a natural generalization of simply setting $\nabla R(\boldsymbol{\beta}_\tau) = 0$ when $R(\cdot)$ is smooth. It simply requires that the function is increasing as we move away from the point $\hat{\boldsymbol{\beta}}_\tau$ regardless of the direction in which we decide to move.

The computation of $\hat{\boldsymbol{\beta}}_\tau^c$ is straightforward with the help of linear programming. More details of the computation of quantile regression can be found in Koenker (2008). There is a vast literature on the estimation and inference for one or several percentile levels for model (1); see Koenker (2003, 2008), He (1997, 2003), among many others.

3.6.3 Asymptotic properties of quantile regression.

The conditional distribution function of Y_is will be written as

$$P(Y_i < y | x_i) = F_{Y_i}(y|x_i) = F_i(y)$$

and so

$$Q_{Y_i}(\tau|x_i) = F_{Y_i}^{-1}(\tau|x_i) \equiv \xi_i(\tau).$$

We will employ the following regularity conditions to explore the asymptotic behavior of the estimator $\hat{\beta}_n(\tau)$, the minimizer of

$$\sum_{i=1}^{n} \rho_\tau(y_i - x_i^\top \boldsymbol{\beta}_\tau).$$

- **Condition (A1)**. The distribution functions $\{F_i\}$ are absolutely continuous, with continuous densities $f_i(\xi)$ uniformly bounded away from 0 and ∞ at the points $\xi_i(\tau), i = 1, 2, \ldots$.

- **Condition (A2)**. There exists positive definite matrices D_0 and $D_1(\tau)$ such at

  (i) $\lim_{n\to\infty} n^{-1} \sum x_i x_i^\top = D_0$;

(ii) $\lim_{n\to\infty} n^{-1}\sum f_i(\xi_i(\tau))x_i x_i^\top = D_1(\tau)$;

(iii) $\max_{i=1,\ldots,n}\|x_i\|/\sqrt{n} \to 0$.

Conditions $A2(i)$ and $A2(iii)$ are familiar throughout the literature on M-estimators for regression models; some variant of them is necessary to ensure that a Lindeberg condition is satisfied. Condition $A2(ii)$ is really a matter of notational convenience and could be deduced from Condition $A2(i)$ and a slightly strengthened version of Condition $A1$.

Theorem 4.1 *Under Conditions A1 and A2,*

$$\sqrt{n}\{\hat{\beta}_n(\tau) - \beta_0(\tau)\} \to N(0, \tau(1-\tau)D_1^{-1}D_0D_1^{-1})$$

in distribution as $n \to \infty$. In particular, in the iid error model (that is, $Y_i = x_i^\top\beta + u_i$, $i = 1,\ldots,n$ with iid errors u_i have a common distribution function F),

$$\sqrt{n}\{\hat{\beta}_n(\tau) - \beta_0(\tau)\} \to N(0, \omega^2 D_0^{-1}),$$

in distribution as $n \to \infty$, where $\omega^2 = \tau(1-\tau)/f_i^2(\xi_i(\tau))$.

Remark: The asymptotic properties above were established under fixed design, which can be extended to handle random design without further difficulties.

3.6.4 The least absolute deviation.

In the iid error model (random design) $Y_i = x_i^\top\beta + u_i$, $i = 1,\ldots,n$, where u_i have a common distribution function F, the LAD is a special case of quantile regression when $\tau = 0.5$. Its estimate $\hat{\beta}_{LAD}$ is defined to be the minimizer of

$$\min_\beta \sum_{i=1}^n |Y_i - X_i^\top\beta|.$$

Hence, the following theorem can be established in a straightforward fashion under regularity conditions.

Theorem 4.2 *Under Conditions A1 and A2,*

$$\sqrt{n}\{\hat{\beta}_{LAD} - \beta_0\} \to N(0, \omega^2 V^{-1})$$

in distribution as $n \to \infty$, where $\omega^2 = 1/4f^2(0)$ and $V = E(XX^\top)$.