Likelihood Ratio Test of a General Linear Hypothesis
Consider the Likelihood Ratio Test of

\[ H_0 : C\beta = d \quad \text{vs} \quad H_A : C\beta \neq d. \]

Suppose

\[ y \sim N(X\beta, \sigma^2 I). \]

The likelihood function is

\[ L(\beta, \sigma^2 | y) = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{1}{2\sigma^2} (y-X\beta)'(y-X\beta)} \quad \text{for } \beta \in \mathbb{R}^p \text{ and } \sigma^2 > 0. \]
Note that

\[ L(\beta, \sigma^2 | y) = (2\pi \sigma^2)^{-n/2} e^{-\frac{1}{2}\sigma^2 Q(\beta)}, \]

where

\[ Q(\beta) = (y - X\beta)'(y - X\beta) \]
\[ = \| y - X\beta \|^2. \]
The parameter space under the null hypothesis $H_0 : C\beta = d$ is

$$\Omega_0 = \{ (\beta, \sigma^2) : C\beta = d, \sigma^2 > 0 \}.$$ 

The parameter space corresponding to the union of the null and alternative parameter spaces is

$$\Omega = \{ (\beta, \sigma^2) : \beta \in \mathbb{R}^p, \sigma^2 > 0 \}.$$
The likelihood ratio test rejects $H_0$ iff

$$\Lambda(y) = \frac{\sup_{\Omega_0} L(\beta, \sigma^2 | y)}{\sup_{\Omega} L(\beta, \sigma^2 | y)}$$

is sufficiently small.
To conduct a significance level $\alpha$ likelihood ratio test, we reject $H_0$ iff

$$\Lambda(y) \leq c_\alpha,$$

where $c_\alpha$ satisfies

$$\sup\{P(\Lambda(y) \leq c_\alpha | \beta, \sigma^2) : (\beta, \sigma^2) \in \Omega_0\} \leq \alpha.$$
To find $\Lambda(y)$, we must maximize the likelihood over $\Omega_0$ and $\Omega$.

For any fixed $\beta \in \mathbb{R}^p$, we can find the value of $\sigma^2 > 0$ that maximizes $L(\beta, \sigma^2 | y)$ as follows.
Because log is a strictly increasing function, the value of $\sigma^2$ that maximizes $L(\beta, \sigma^2 | y)$ is the same as the value of $\sigma^2$ that maximizes

$$l(\beta, \sigma^2 | y) \equiv \log L(\beta, \sigma^2 | y).$$
\begin{align*}
l(\beta, \sigma^2 | y) &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} Q(\beta) \\
\frac{\partial l(\beta, \sigma^2 | y)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{Q(\beta)}{2\sigma^4}.
\end{align*}

Equating to 0 and solving for $\sigma^2$ yields

\begin{align*}
\hat{\sigma}^2(\beta) &= \frac{Q(\beta)}{n}.
\end{align*}
Furthermore, examination \( \frac{\partial l(\beta, \sigma^2 | y)}{\partial \sigma^2} \) as a function of \( \sigma^2 \) shows that \( l(\beta, \sigma^2 | y) \) as a function of \( \sigma^2 \) is increasing to the left of \( Q(\beta)/n \) and decreasing to the right of \( Q(\beta)/n \).

Thus, for any fixed \( \beta \in \mathbb{R}^p \), the likelihood is maximized over \( \sigma^2 > 0 \) at \( Q(\beta)/n \).
Now note that

\[
L(\beta, \hat{\sigma}^2(\beta)|y) = \left(2\pi Q(\beta)/n\right)^{-n/2} e^{-\frac{n}{2Q(\beta)}Q(\beta)}
= \left(2\pi eQ(\beta)/n\right)^{-n/2},
\]

which is clearly maximized over \( \beta \) by minimizing \( Q(\beta) \) over \( \beta \).
Let \( \hat{\beta} \) be any minimizer of \( Q(\beta) \) over \( \beta \in \mathbb{R}^p \).

We know from previous results that \( \hat{\beta} \) minimizes \( Q(\beta) \) over \( \beta \in \mathbb{R}^p \) iff \( \hat{\beta} \) is a solution to NE. \( (X'X)^{-1}X'y \) is one solution.
Thus, the maximum likelihood estimator (MLE) of $\sigma^2$ is

$$\hat{\sigma}^2_{\text{MLE}} = \frac{Q(\hat{\beta})}{n} = \frac{||y - X\hat{\beta}||^2}{n},$$

where $\hat{\beta}$ is any solution to the NE.

Recall that $X\hat{\beta} = P_Xy$ is the same for all solution to NE.

Thus, $\hat{\sigma}^2_{\text{MLE}}$ is the same for any $\hat{\beta}$ that solves the NE.
Note that the MLE of $\sigma^2$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{Q(\hat{\beta})}{n} = \frac{Q(\hat{\beta})}{n - \text{rank}(X)} \cdot \frac{n - \text{rank}(X)}{n}$$

$$= \frac{\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}}{n - \text{rank}(X)} \cdot \frac{n - \text{rank}(X)}{n}$$

$$= \hat{\sigma}^2 \frac{n - \text{rank}(X)}{n}.$$

Recall

$$E(\hat{\sigma}^2) = \sigma^2.$$

Thus,

$$E(\hat{\sigma}^2_{\text{MLE}}) = \frac{n - \text{rank}(X)}{n} \sigma^2 < \sigma^2.$$
If $X$ is of full-column rank, then

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

is the maximum likelihood estimator (MLE) of $\beta$. 
We have

\[
\sup_{\Omega} L(\beta, \sigma^2 | y) = L(\hat{\beta}, \hat{\sigma}^2_{\text{MLE}} | y)
\]

\[
= (2\pi e Q(\hat{\beta}) / n)^{-n/2}.
\]
If we let $\tilde{\beta}$ denote the minimizer of $Q(\beta)$ over $\beta \in \mathbb{R}^p$ satisfying $C\beta = d$, then

$$
\sup_{\Omega_0} L(\beta, \sigma^2 | y) \\
= L(\tilde{\beta}, Q(\tilde{\beta})/n | y) \\
= (2\pi e Q(\tilde{\beta})/n)^{-n/2}.
$$
Thus

\[
\Lambda(y) = \frac{(2\pi e Q(\tilde{\beta})/n)^{-n/2}}{(2\pi e Q(\hat{\beta})/n)^{-n/2}} = \left[ \frac{Q(\tilde{\beta})}{Q(\hat{\beta})} \right]^{-n/2}.
\]
Now note that

\[ \Lambda(y) \leq c_\alpha \iff \left[ \frac{Q(\tilde{\beta})}{Q(\hat{\beta})} \right]^{-n/2} \leq c_\alpha \]

\[ \iff \frac{Q(\tilde{\beta})}{Q(\hat{\beta})} \geq c_\alpha^{-2/n} \]

\[ \iff \frac{Q(\tilde{\beta})}{Q(\hat{\beta})} - 1 \geq c_\alpha^{-2/n} - 1 \]

\[ \iff \frac{Q(\tilde{\beta}) - Q(\hat{\beta})}{Q(\hat{\beta})} \geq c_\alpha^{-2/n} - 1. \]
\[ \frac{[Q(\hat{\beta}) - Q(\tilde{\beta})]}{Q(\hat{\beta})/(n - r)} \leq \frac{n - r}{q}(c_\alpha^{-2/n} - 1), \]

where

\[ q = \text{rank}(C) \quad \text{and} \quad n - r = n - \text{rank}(X). \]
Example:

Suppose

\[ y = X\beta + \varepsilon, \]

where

\[ \varepsilon \sim N(0, \sigma^2 I). \]

Furthermore, suppose \( \text{rank}(X) = p \). Partition

\[ X = [X_1, X_2] \quad \text{and} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \]

where \( X_1 \) is \( n \times p_1 \) and \( \beta_1 \) is \( p_1 \times 1 \).

Suppose we wish to test \( H_0 : \beta_2 = 0 \).
$H_0 : \beta_2 = 0$ is a GLH $H_0 : C\beta = d$ with

$$C = \begin{bmatrix} 0 & I \end{bmatrix}_{q\times q} \quad \text{and} \quad d = 0_{q\times 1},$$

where

$$q = p - p_1.$$

This GLH is testable because $C$ has rank $q$ and $C\beta$ is estimable due to full-column rank of $X$. 
\[ Q(\tilde{\beta}) = \min\{ Q(\beta) : C\beta = d \} \]
\[ = \min\{ \|y - X\beta\|^2 : \beta \in \mathbb{R}^p \ni \beta_2 = 0 \} \]
\[ = \min\{ \|y - X_1\beta_1\|^2 : \beta_1 \in \mathbb{R}^{p_1} \} \]
\[ = \|y - X_1\hat{\beta}_1\|^2 = \|y - P_{X_1}y\|^2 \]
\[ = \| (I - P_{X_1})y \|^2 = y'(I - P_{X_1})y \]
\[ = \text{SSE}_{\text{Reduced}}. \]
\[ Q(\hat{\beta}) = \| y - X\hat{\beta} \|^2 = \| y - P_X y \|^2 = \| (I - P_X)y \|^2 = y'(I - P_X)y = \text{SSE}_{\text{Full}}. \]
The DF for $\text{SSE}_{\text{Full}}$ is

$$DF_F = \text{rank}(I - P_X) = n - p.$$ 

The DF for $\text{SSE}_{\text{Reduced}}$ is

$$DF_R = \text{rank}(I - P_{X_1}) = n - p_1.$$ 

$$DF_R - DF_F = p - p_1 = q.$$
We have shown that, in the general case, the likelihood ratio test (LRT) of

\[ H_0 : C / \beta = d \]

rejects for sufficiently large

\[
\frac{[Q(\tilde{\beta}) - Q(\hat{\beta})]/q}{Q(\hat{\beta})/(n - r)}.
\]
In this example,

\[
\frac{[Q(\hat{\beta}) - Q(\tilde{\beta})]/q}{Q(\tilde{\beta})/(n - r)} = \frac{[SSE_{Reduced} - SSE_{Full}]/(DF_R - DF_F)}{SSE_{Full}/DF_F},
\]

which should look familiar.
We now show that for a testable GLH

$$H_0 : C\beta = d,$$

the GLT statistic

$$F = \frac{(C\hat{\beta} - d)'(C(X'X)^{-1}C')^{-1}(C\hat{\beta} - d)/q}{\hat{\sigma}^2},$$

$$= \frac{[Q(\hat{\beta}) - Q(\tilde{\beta})]/q}{Q(\tilde{\beta})/(n - r)}.$$

Thus, the GLT is equivalent to the LRT.
Note that

\[ Q(\hat{\beta})/(n - r) = y' \left( \frac{I - P_X}{n - r} \right) y = \hat{\sigma}^2. \]

Thus, it remains to show

\[ Q(\tilde{\beta}) - Q(\hat{\beta}) = (C\hat{\beta} - d)'(C(X'X)^{-1}C')^{-1}(C\hat{\beta} - d). \]

From 3.10, we know that \( \tilde{\beta} \) is leading subvector of solution to RNE.
Theorem 6.1:

If $C\beta = d$ is testable and $\tilde{\beta}$ is the leading subvector of a solution to the RNE

$$
\begin{bmatrix}
X'X & C' \\
C & 0
\end{bmatrix}
\begin{bmatrix}
b \\
\lambda
\end{bmatrix} =
\begin{bmatrix}
X'y \\
d
\end{bmatrix},
$$

then

$$
Q(\tilde{\beta}) - Q(\hat{\beta}) = (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta})
= (C\hat{\beta} - d)'(C(X'X)^{-1}C')^{-1}(C\hat{\beta} - d).
$$
Proof of Result 6.1:

First show that

\[ Q(\tilde{\beta}) - Q(\hat{\beta}) = (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}). \]
Now let $\tilde{\lambda}$ denote the trailing subvector of the solution to RNE whose leading subvector is $\tilde{\beta}$; i.e., suppose

$$\begin{bmatrix} \tilde{\beta} \\ \tilde{\lambda} \end{bmatrix}$$

is solution to RNE.

Show $X'X(\hat{\beta} - \tilde{\beta}) = C'\tilde{\lambda}$. 
Now show

\[ \tilde{\lambda} = (C(X'X)^{-1}C')^{-1}(C\hat{\beta} - d). \]
We have established

(1) \[ Q(\hat{\beta}) - Q(\tilde{\beta}) = (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}) \]

(2) \[ X'X(\hat{\beta} - \tilde{\beta}) = C'\tilde{\lambda} \]

(3) \[ \tilde{\lambda} = (C(X'X)^{-1}C')^{-1}(C\hat{\beta} - d). \]

Use these results to finish the proof.
Corollary 6.4:

Suppose \( C \beta = d \) is testable, and suppose \( \hat{\beta} \) is a solution to the NE.

Then the leading subvector of a solution to the RNE with constraint

\[
C \beta = d
\]

can be found by solving for \( b \) in the equations

\[
X'Xb = X'y - C'(C(X'X)^{-1}C')^{-1}(C'\hat{\beta} - d).
\]