Estimation of Error Variance Under the Gauss-Markov Model
We begin by stating a fact that will be used twice in this set of notes.
If $A$ is a symmetric and idempotent matrix, then $\text{rank}(A) = \text{tr}(A)$. Result 1.11 from Ken Koehler's notes can be used to prove this fact.
Under the Gauss-Markov model,

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

where \( E(\varepsilon) = 0 \) and

\[ \text{Var}(\varepsilon) = \sigma^2 I. \]
The parameter $\sigma^2$ is an unknown positive real value.

$\sigma^2$ is called the "error variance" because

$$\text{Var}(\varepsilon_i) = \sigma^2 \quad \forall \ i = 1, \ldots, n.$$
We have already discussed the problem of estimating an estimable $C\beta$.

We showed $\text{Var}(C\hat{\beta})$ is $\sigma^2 C(x'x)^{-1}c'$. 
Thus, the variance of $\beta$ is unknown but can be estimated if we have an estimate of $\sigma^2$. 
Note that

\[ E\left( \frac{1}{n} \sum \epsilon_i \right) = \frac{1}{n} E\left( \sum \epsilon_i \right) \]

\[ = \frac{1}{n} E\left( \sum_{i=1}^{n} \epsilon_i^2 \right) = \frac{1}{n} \sum_{i=1}^{n} E\left( \epsilon_i^2 \right) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \text{Var}(\epsilon_i) = \frac{1}{n} \sum_{i=1}^{n} \sigma^2 \]

\[ = \sigma^2. \]
Thus, if we could observe 
\[ \varepsilon = y - x_\beta, \]
we could use 
\[ \frac{1}{n} \sum \varepsilon \varepsilon \]
as an unbiased estimator of \( \sigma^2 \).
However, \( \hat{\varepsilon} = y - X\hat{\beta} \) is not observable because we do not know \( X\beta \).

Instead of using \( \varepsilon \) to form an estimator of \( \sigma^2 \), consider using \( \hat{\varepsilon} = y - X\hat{\beta} = y - \hat{y} \).
\( \hat{\epsilon} = y - \hat{X}' \hat{\beta} = y - \bar{y} \) is known as the "residual vector."

To obtain an unbiased estimator of \( \sigma^2 \), we need to determine \( E(\hat{\epsilon}' \hat{\epsilon}) \).
If $\mathbf{w}$ is an $n \times 1$ random vector with $E(\mathbf{w}) = \mathbf{\mu}$ and $\text{Var}(\mathbf{w}) = \mathbf{\Sigma}$ and $\mathbf{A}$ is an $n \times n$ non-random matrix, then

$$E(\mathbf{w}' \mathbf{A} \mathbf{w}) = \mathbf{\mu}' \mathbf{A} \mathbf{\mu} + \text{tr}(\mathbf{A} \mathbf{\Sigma}).$$
Proof:

\[ E(ww'Aw) = E[\text{tr}(ww'Aw)] \]
\[ = E[\text{tr}(Aww')] = \text{tr}[E(Aww')] \]
\[ = \text{tr}[AE(ww')] \]
\[ = \text{tr}[A(\text{Var}(w) + \mu\mu')] \]
\[ = \text{tr}[A\Sigma + A\mu\mu'] \]
\[ = \text{tr}(A\Sigma) + \text{tr}(A\mu\mu') \]
\[ = \text{tr}(\mathbf{A} \Sigma) + \text{tr}(\mu' \mathbf{A} \mu) \]
\[ = \text{tr}(\mathbf{A} \Sigma) + \mu' \mathbf{A} \mu. \]
Now note that
\[ \hat{\varepsilon} = y - X\hat{\beta} = y - X\beta = y - P_x y \]

\[ = (I - P_x) y. \]

Furthermore,
\[ I - P_x \text{ is Symmetric:} \]
\[ (I - P_x)' = I' - P_x' = I - P_x \]
$I - P_x$ is idempotent:

$$(I - P_x)(I - P_x) = I - P_x - P_x + P_x P_x$$

$$= I - P_x - P_x + P_x$$

$$= I - P_x.$$
Thus,

\[ \hat{\Sigma} = \left[ (I-P_x)y \right]' \left[ (I-P_x)y \right] \]

\[ = y' (I-P_x)' (I-P_x) y \]

\[ = y' (I-P_x) (I-P_x) y \]

\[ = y' (I-P_x) y \]
It follows that

\[ E(\hat{\beta}^2) = E(y'(I-P_x)y) \]

\[ = [E(y)]'(I-P_x)E(y) \]

\[ + \text{tr}[(I-P_x)\text{Var}(y)] \]

\[ = \beta'X'(I-P_x)X\beta + \text{tr}[(I-P_x)\sigma^2I] \]
\[ = \beta' (X-P_x X) \beta + \sigma^2 \text{tr}(I-P_x) \]

\[ = \beta' (X-X) \beta + \sigma^2 \left[ \text{tr}(I) - \text{tr}(P_x) \right] \]

\[ = \sigma^2 \left[ n - \text{tr}(P_x) \right] \]

\[ = \sigma^2 \left[ n - \text{rank}(P_x) \right] \text{ (Follows from KK Result 1.11)} \]

\[ = \sigma^2 \left[ n - \text{rank}(X) \right] \text{ (See HW3)} \]

\[ = \sigma^2 (n-r), \text{ where } r = \text{rank}(X). \]
Thus, $E(\hat{\Sigma}^2) = \sigma^2(n-r)$

and $E[\hat{\Sigma}^2/(n-r)] = \sigma^2$.

Let $\hat{\sigma}^2 = \frac{\hat{\Sigma}^2}{n-r} = \frac{y'(I-P_x)y}{n-r}$

We have $E(\hat{\sigma}^2) = \sigma^2$. 
\[ \hat{\varepsilon}_i^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{\gamma}_i)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2} \]

is often referred to as the **Error Sum of Squares**
or **Sum of Squares for Error** and is abbreviated **SSE**.
\[ \hat{\sigma}^2 = \frac{\text{SSE}}{n-r} \] is often referred to as the Mean Square for Error and is abbreviated MSE.

\[ E(\text{MSE}) = \sigma^2 \]
We have shown that $\hat{\sigma}^2$ is an unbiased estimator of $\sigma^2$ under the Gauss-Markov model, where $E(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = \sigma^2 I$ but the distribution of $\varepsilon$ is unspecified.
If we assume $\varepsilon \sim N(0, \sigma^2 I)$, we can derive the distribution of

$$\hat{\sigma}^2 = \text{MSE}.$$
**Chi-Square Distributions**

Suppose $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1)$.

Then $Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$ has the multivariate standard normal distribution $N(0, I)$. 

Let $u$ be a random variable with the same distribution as $z'z = \sum_{i=1}^{n} z_i^2$.

Then $u \overset{d}{=} z'z = \sum_{i=1}^{n} z_i^2$ has a central chi-square distribution with $n$ degrees of freedom.
The notation $u \sim X_n^2$ indicates that $u$ has a central chi-square distribution with $n$ degrees of freedom.
Now suppose \( w \sim N(\mu, \mathbf{I}) \).

Then \( u = w'w \) has a non-central chi-square distribution with \( n \) degrees of freedom (df) and non-centrality parameter \( \mu'\mu \). We write \( u \sim X^2_n(\mu'\mu) \).
Note that a central chi-square distribution is the same as a non-central chi-square distribution with non-centrality parameter (ncp) equal to 0: $X_n^2 \overset{d}{=} X_n^2(0)$. 
Suppose $A$ is an $n \times n$ symmetric matrix of rank($A$) = $k$.

Suppose $\mathbf{w} \sim N(\mu, \Sigma)$, where $
\Sigma$ is a nonsingular $n \times n$ matrix.

If $A\Sigma$ is idempotent, then

$\mathbf{w}' A \mathbf{w} \sim \chi^2_k (\mu' A \mu)$.
A proof of this important result is provided on pages 309-314 of Ken Koehler's notes. Please read and understand the proof.
Now under the normal theory Gauss–Markov model,

\[ Y \sim N(\mathbf{X}\beta, \sigma^2 I). \]

Consider the distribution of

\[ \frac{(n-r)\hat{\sigma}^2}{\sigma^2} = Y' (I - P_X) Y. \]
\[
\left( \frac{I - P_X}{\sigma^2} \right) (6^2 I) = I - P_X,
\]

which is idempotent. Thus,

\[
l_X'(I - P_X) \xrightarrow{\text{as}} X^2 \left( \beta' X' (I - P_X) X \beta \right) \quad \text{rank} \left( \frac{I - P_X}{\sigma^2} \right)
\]
Because \((I-P_x)X = X - P_x X\)

\[= X - X\]

\[= 0,\]

the ncp

\[\beta' (I-P_x) X \beta \over \sigma^2 = 0.\]
Because

\[ \text{rank} \left( \frac{I-P_x}{\sigma^2} \right) = \text{rank} \left( I-P_x \right) \]

\[ = \text{tr} \left( I-P_x \right) = \text{tr}(I) - \text{tr}(P_x) \]

\[ = n - \text{rank}(P_x) = n - \text{rank}(X) \]

\[ = n - r, \quad \text{the df} = n - r. \]
Thus, we have shown that 

\[
\frac{(n-r) \hat{\sigma}^2}{\sigma^2} \sim X^2_{n-r}.
\]

This implies that 

\[
\hat{\sigma}^2 \sim \frac{\sigma^2}{n-r} X^2_{n-r}
\]

has a "scaled" chi-square distribution.