Theorem: Given \( y = \mathbf{w}\beta + \epsilon \)

with \( E(\epsilon \epsilon^T) = \mathbf{Q} \)

\[ E(\epsilon) = 0, \quad E(\mathbf{w}\beta^T) = \mathbf{R} \]

Then the minimum variance linear estimator of \( \beta \) depending on \( y \) is given by the following

\[ \hat{\beta} = \mathbf{Rw}^T (\mathbf{wRw}^T + \mathbf{Q})^{-1} y \]

with the error covariance

\[ E(\beta - \hat{\beta})(\beta - \hat{\beta})^T = \mathbf{R} - \mathbf{Rw}^T (wRw^T + \mathbf{Q})^{-1} \mathbf{wR} \]

Proof: The minimum variance linear estimate (mule) of a random vector \( x \) depending on \( y \) is given by

\[ \hat{x} = E(\mathbf{xy}^T) E(\mathbf{yy}^T)^{-1} y \]

\[ \hat{\beta} = E(\beta \mathbf{y}^T) E(\mathbf{yy}^T)^{-1} y \]

\[ \mathbf{By}^T = \beta [ \mathbf{w}\beta + \epsilon ]^T = \beta (\beta^T \mathbf{w}^T + \mathbf{e}^T) \]

\[ = \beta \beta^T \mathbf{w}^T + \beta \beta^T \]

\[ E[\mathbf{By}^T] = E(\beta \beta^T) \mathbf{w}^T + E(\beta \beta^T) \]
\[ \mathbb{E} [yy^T] = ( \omega \beta + \varepsilon)( \beta \omega^T + \varepsilon^T) \]

\[ = \omega \beta \beta^T \omega^T + \omega \beta \varepsilon^T + \varepsilon \beta \omega^T + \varepsilon \varepsilon^T \]

\[ \therefore \mathbb{E} [yy^T] = \mathbb{W} \mathbb{R} \omega^T + \Omega \]

\[ \therefore \widehat{\beta} = \mathbb{R} \omega^T(\mathbb{W} \mathbb{R} \omega^T + \Theta)^{-1} \gamma. \]

\[ \mathbb{E}[(\beta - \widehat{\beta})(\beta - \widehat{\beta})^T] = \mathbb{E}[(\beta - \widehat{\beta})\beta^T] \]

\[ = \mathbb{E}[(\beta - \widehat{\beta})\beta^T] \quad \text{(since \((\beta - \widehat{\beta})(\beta - \widehat{\beta})^T\) is symmetric)} \]

\[ = \mathbb{E}[(\beta \beta^T) - \mathbb{E}[(\widehat{\beta} \beta^T)] \]

\[ = \mathbb{E}[(\beta \beta^T) - \mathbb{E}[(\widehat{\beta} (\beta - \widehat{\beta} + \widehat{\beta})^T)] \]

\[ = \mathbb{E}[(\beta \beta^T) - \mathbb{E}[(\widehat{\beta} \beta^T)] \]

\[ = \mathbb{R} \]

\[ \mathbb{E}[(\widehat{\beta} \widehat{\beta}^T)] = \mathbb{E} \left\{ \mathbb{R} \omega^T(\mathbb{W} \mathbb{R} \omega^T + \Theta)^{-1} \gamma \gamma^T(\mathbb{W} \mathbb{R} \omega^T + \Theta)^{-1} \right\} \]

\[ = \mathbb{E} \left\{ \mathbb{R} \omega^T(\mathbb{W} \mathbb{R} \omega^T + \Theta)(\mathbb{W} \mathbb{R} \omega^T + \Theta)^{-1} \right\} \]

\[ = \mathbb{R} \omega^T(\mathbb{W} \mathbb{R} \omega^T + \Theta)^{-1} \mathbb{W} \omega \]

\text{Error covariance is given by}

\[ \cdots \]
\[ E[(\beta - \hat{\beta})(\beta - \hat{\beta})^T] = R - R^T (WRU^T + \theta)^{-1}WR. \]

**Corollary:** Consider the problem in the theorem above, then \( \hat{\beta} \) can be written as

\[ \hat{\beta} = (W^T \Omega^{-1}W + R^{-1})^{-1}W^T \Omega^{-1}y \]

with \( E[(\beta - \hat{\beta})(\beta - \hat{\beta})^T] = (W^T \Omega^{-1}W + R^{-1})^{-1} \)

**Remark:** Note that with the above relation if \( P \) denotes the error covariance that

\[ \hat{\beta} = P \Omega^{-1}y \]

or in other words

\[ P \hat{\beta} = W \Omega^{-1}y \]

Note the RKS \((W \Omega^{-1}y)\) does not depend on the statistics \(R\) on \( \beta \).

**Proof:** \( \hat{\beta} = RW^T (WRU^T + \theta)^{-1}y \).

To show that \( \hat{\beta} = (W^T \Omega^{-1}W + R^{-1})^{-1}W^T \Omega^{-1}y \)

This can be shown by proving

\[ RW^T (WRU^T + \theta)^{-1} = (W^T \Omega^{-1}W + R^{-1})^{-1}W^T \Omega^{-1}y \]

\[ \equiv (W^T \Omega^{-1}W + R^{-1}) RW^T (WRU^T + \theta)^{-1} \]
\[ (\omega^T \Theta^{-1} \omega + R^{-1}) R \omega^T (\omega R \omega^T + \Theta) \\
= (\omega^T \Theta^{-1} \omega + R^{-1})(\omega^T \Theta^{-1} \omega + R^{-1}) \omega^T \\
(\omega R \omega^T + \Theta) \\
= (\omega^T \Theta^{-1} \omega + R^{-1}) R \omega^T = \omega^T \Theta^{-1} (\omega R \omega^T + \Theta) \\
= \omega^T \Theta^{-1} \omega R \omega^T + \omega^T = \omega^T \Theta^{-1} \omega R \omega^T + \omega \\
which \ is \ true \\
\therefore \hat{\beta} = (\omega^T \Theta^{-1} \omega + R^{-1})^{-1} \omega^T \Theta^{-1} \gamma \\
EC(\beta - \hat{\beta})(\beta - \hat{\beta}) = R - R \omega^T (\omega R \omega^T + \Theta) \omega \\
[ (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} = A_{11}^{-1} + A_{12}^{-1} A_{22} (A_{22}^{-1} A_{21} A_{11})^{-1} A_{12}^{-1} A_{22}^{-1} A_{21} ] \\
Identify \ A_{11} = R^{-1}, \ A_{12} = -\omega R^{-1} \omega^T, \ A_{21} = \omega \\
A_{22} \ \text{with} \ \Theta \\
we \ have \\
EC(\beta - \hat{\beta})(\beta - \hat{\beta}) = (R^{-1} + \omega^T \Theta^{-1} \omega)^{-1} \\
Remark: \ \frac{1}{\mu} = \frac{1}{R} + 2 \omega \Theta^{-1} \omega^T \\
\text{Theorem: Suppose} \\
Ja = Ha \beta + Va \\
Jb = Hb \beta + Vb \\
are \ two \ relations \ of \ \beta \ \text{on} \ Ja, \ \text{and} \ \ Jb.
Jo are that are available. \( E(V_a^T V_a^T) = 0 \)
\[
E(V_a V_{a^T}) = Q_a; \quad E(U_b V_b^T) = Q_b
\]

Let \( \hat{\beta}_a \) be the mule of \( \beta \) based on \( y_a \).

Let \( \hat{\beta}_b \) be the mule of \( \beta \) based on \( y_b \).

What will be the mule of \( \beta \) based on \( (y_a, y_b) \)?

\[
y = \begin{pmatrix} y_a \\ y_b \end{pmatrix} = \begin{bmatrix} H_a \\ H_b \end{bmatrix} \beta + \begin{bmatrix} V_{a^T} \\ V_{b^T} \end{bmatrix}
\]

Let \( \hat{\beta} \) be the mule of \( \beta \) based on \( y \)

\[
\hat{\beta} = H^T \begin{bmatrix} Q_a^{-1} & 0 \\ 0 & Q_b^{-1} \end{bmatrix} \begin{bmatrix} y_a \\ y_b \end{bmatrix}
\]

\[
= P^{-1} \hat{\beta} = H^T Q_a^{-1} y_a + H_b^T \alpha_b^T y_b
\]

\[
P^{-1} = (H^T \alpha_a^{-1} H + R^{-1})
\]

\[
= (H_a^T H_b^T) \begin{bmatrix} Q_a^{-1} & 0 \\ 0 & Q_b^{-1} \end{bmatrix} [H_a H_b] + R^{-1}
\]

\[
= H_a^T Q_a^{-1} H_a + H_b^T \alpha_b^{-1} H_b + R^{-1}
\]
Gauss Markov Estimate: Consider the

\[ y = \mathbf{X}\beta + \mathbf{e} \]

where \( \beta \) is a constant vector of parameters with \( \beta \in \mathbb{R}^n \).
\( \mathbf{e} \) is a random vector with
\[ E(\mathbf{e}\mathbf{e}^T) = \mathbf{Q} \in \mathbb{R}^{n \times n} \]
\[ E(\mathbf{e}) = \mathbf{0} \]
\( y \) is a random vector in \( \mathbb{R}^n \)

First requirement is that we desire an "unbiased" estimate of \( \beta \) that is linearly dependent on \( y \).
we scale \[ \hat{\beta} = K\gamma \] \text{s.t.}
\[ E(\hat{\beta}) = E(\beta) \text{ for all } \beta \in \mathbb{R}^n \]
\[ \Rightarrow E[K\gamma] = K\beta \text{ for all } \beta \in \mathbb{R}^n \]
\[ \Rightarrow E[K(\omega\beta + \epsilon)] = K\beta \text{ for all } \beta \in \mathbb{R}^n \]
\[ \Rightarrow K\omega = \beta \text{ for all } \beta \in \mathbb{R}^n \]
\[ \Rightarrow K\omega = I \]

The related optimization problem is
\[ \mu = \min \ E \left\{ \|\beta - \hat{\beta}\|_2^2 \right\} \]
\[ \hat{\beta} = K\gamma \]
\[ K\omega = I. \]

The solution to the above problem is termed the Gauss-Markov estimate, minimum-variance unbiased estimate (MVUE).
\[ \|\beta - \hat{\beta}\|_2^2 = (\beta - \hat{\beta})^T (\beta - \hat{\beta}) \]
\[ = (\beta - K\gamma)^T (\beta - K\gamma) \]
\[ = E((\beta - K\gamma)^T (\beta - K\gamma)) \]
\[ = E((\beta - K\omega\beta - K\epsilon)^T (\beta - K\omega\beta - K\epsilon)) \]
\[ = E((\beta^T - \beta^T \omega^T K T - \epsilon^T K \epsilon) (\beta - K\omega\beta - K\epsilon)) \]
\[ = E[\beta^T \beta - \beta^T K\omega\beta - \beta^T K\epsilon - \epsilon^T K \epsilon - \epsilon^T (K T \beta - \epsilon) - \epsilon^T (K T \beta - \epsilon) + \epsilon^T (K T \beta - \epsilon) + \epsilon^T (K T \beta - \epsilon)] \]
\[ = \text{lb} \beta^T \beta - \beta^T K \omega \beta - \beta^T K \varepsilon \] 
\[ - \beta^T \omega^T K^T \beta + \beta^T \omega^T K^T K \omega \beta \] 
\[ + \varepsilon^T K^T K \varepsilon \] 
\[ = \varepsilon^T \left( K^T K \varepsilon \right) = \text{trace} \left( K \Omega K^T \right), \] 
\[ \varepsilon^T \left( \varepsilon \Omega \varepsilon \right) = 0. \]

\[ \mu = \text{trace} \left( K \Omega K^T \right) \]
\[ K W = I \]
\[ K^T = \begin{bmatrix} -k_1^T \\ k_2^T \\ \vdots \\ k_n^T \end{bmatrix} \]
\[ = \mu \sum k_i^T \phi k_i \]
\[ k_i^T \omega_j = 0 \quad \forall \ j = 1 \ldots n. \]
\[ \mu_i = \mu \left\| k_i \right\|^2 \phi k_i \]
\[ k_i^T \omega_j = \delta_{ij}, \quad \forall j = 1 \ldots n. \]
\[ \forall i, j = 1 \ldots n \]
\[ \mu_i = \mu \left\| k_i \right\|^2 \phi k_i \]
\[ k_i^T \omega_j = \delta_{ij}, \quad \forall j = 1 \ldots n. \]
\[ \mu_i = \mu \left\| k_i \right\|^2 \phi k_i \]
\[ \mu = \sum_{i=1}^{n} \mu_i \]
\[ < \omega, y > = x^T \Omega x \]
\[ < k_i, k_i > = k_i^T \phi k_i \]
\[ < k_i, \Omega^T \omega_j > = \delta_{ij}, \quad \forall j = 1 \ldots n. \]
\[ \mu = \sum_{i=1}^{n} \mu_i \]
\[ < \omega, y > = x^T \Omega x \]