Recall the regression setting. Testing with

\[ 1 = \begin{pmatrix} 1 \end{pmatrix}, \quad X_i = \begin{pmatrix} 1 \mid x_1 \mid x_2 \mid \ldots \mid x_i \end{pmatrix} \]

Let \( P \) be the projection matrix onto \( C(X_i) \).

\[ H_0: \beta_{p+1} = \beta_{p+2} = \ldots = \beta_r = 0 \quad \Leftrightarrow \quad H_0: \mathbb{E}[Y] = C(X_i) \]

Also, I can write this hypothesis in (testable) form:

\[ H_0: \beta = 0 \]

For

\[ C = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ (r-p) \times (r+1) & (r-p) \times (r-p) \end{pmatrix} \]

and we've agreed to call

\[ \text{SS}_H = \left( \widehat{\beta}_{OLS} - \mathbf{0} \right)' \left( C(X'X)^{-1} \right)^{-1} \left( \widehat{\beta}_{OLS} - \mathbf{0} \right) \]

for testing \( H_0: \beta = 0 \). This is equivalent to the "full model / reduced model" paradigm e.g. of Neter, Wasserman, and friends. (See posted "handout" for a proof.)
Here elaborate on this paradigm:

\[
Y'Y = Y'(P - \frac{1}{n} \sum_{i=1}^{p} P_i) + (P - P_x) Y + Y'(P - P_x) Y + Y'(I - P_x) Y
\]

\[
= Y'P_y Y + Y'(P - P_x) Y + Y'(P - P_x) Y + Y'(I - P_x) Y
\]

\[
\frac{Y'Y - Y'P_y Y}{2} = \frac{Y'(P - \frac{1}{n} \sum_{i=1}^{p} P_i) Y}{2} + \frac{Y'(P - P_x) Y}{2} + \frac{Y'(I - P_x) Y}{2}
\]

\[
\begin{align*}
SST_{tot} & = Y'(I - P_x) Y \\
SSR_{full} & = \frac{Y'(P - \frac{1}{n} \sum_{i=1}^{p} P_i) Y}{2} \\
SSR_{reduced} & = \frac{Y'(P - P_x) Y}{2} \\
SSR_{error} & = \frac{Y'(I - P_x) Y}{2}
\end{align*}
\]

And it's common standard to organize this in an ANOVA table (for testing \( H_0: \beta_{p+1} = \ldots = \beta_r = 0 \) in \( MCR \))

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression ( (x_1, \ldots, x_r) )</td>
<td>( SSR_{full} )</td>
<td>( \text{rank}(P - \frac{1}{n} \sum_{i=1}^{p} P_i) = (r+1) - 1 = r )</td>
</tr>
<tr>
<td>( (x_{p+1}, \ldots, x_r) )</td>
<td>( SSR_{reduced} )</td>
<td>( \text{rank}(P - P_x) = (p+1) - 1 = p )</td>
</tr>
<tr>
<td>Error ( (x_1, \ldots, x_r) )</td>
<td>( SSR_{error} )</td>
<td>( \text{rank}(P - P_x) = r - p )</td>
</tr>
<tr>
<td>Total ( Y )</td>
<td>( \text{SSS}_{full} )</td>
<td>( n - \text{rank}(X) = n - (r+1) )</td>
</tr>
<tr>
<td>Total ( Y )</td>
<td>( \text{SSS}_{tot} )</td>
<td>( n - 1 )</td>
</tr>
</tbody>
</table>
and sometimes people "reduction in SS notation" to describe

\[ \mathbf{Y}'Y = \mathbf{Y}'\mathbf{P}_0 \mathbf{Y} + \mathbf{Y}'(\mathbf{P}_1 - \mathbf{P}_0) \mathbf{Y} + \mathbf{Y}'(\mathbf{P}_2 - \mathbf{P}_1) \mathbf{Y} + \mathbf{Y}'(\mathbf{I} - \mathbf{P}_2) \mathbf{Y} \]

\[ \mathbb{R}(\beta_0) \quad \mathbb{R}(\beta_1 | \beta_0) \quad \mathbb{R}(\beta_2 | \beta_1, \beta_0) \]

and we can even take this business of breaking down \( \mathbf{Y}' \mathbf{Y} \) in pieces further — i.e., people talk about "Type I" or "sequential" sums of squares.

\[ \mathbf{Y}' \mathbf{Y} = \mathbf{Y}' \mathbf{P}_0 \mathbf{Y} \quad \mathbb{R}(\beta_0) \]

\[ + \mathbf{Y}'(\mathbf{P}_1 - \mathbf{P}_0) \mathbf{Y} \quad \mathbb{R}(\beta_1 | \beta_0) \]

\[ + \mathbf{Y}'(\mathbf{P}_2 - \mathbf{P}_1) \mathbf{Y} \quad \mathbb{R}(\beta_2 | \beta_1, \beta_0) \]

\[ + \ldots \]

\[ + \mathbf{Y}'(\mathbf{I} - \mathbf{P}_r) \mathbf{Y} \quad \mathbb{R}(\beta_r | \beta_{r-1}, \ldots, \beta_0) \]

\[ + \mathbf{Y}'(\mathbf{I} - \mathbf{P}_X) \mathbf{Y} \quad \text{SSE} \quad \text{"hierarchical"} \]

And e.g., \( \mathbb{R}(\beta_2 | \beta_1, \beta_0) \) is an appropriate SS for testing.
\( H_0: \beta_2 = 0 \) in a model that includes only (a constant and) predictor \( x_1, x_2 \) — i.e., in a LM where the model matrix \( X_2 \) — it is not the correct numerator SS for testing \( H_0: \beta_2 = 0 \) in the full model —

A different possibility is the "SAS type III SS's" — if

\[
X_{\cdot i} = X \text{ with } x_i \text{ deleted}
\]

\[
Y' (P_X - P_{X_{\cdot i}}) Y = R (\beta_i; \beta_0, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_r)
\]

is the appropriate numerator SS for testing \( H_0: \beta_i = 0 \) in full model.

It's easy to apply our theorem about independence of quadratic forms (SS's) in the normal Gauss-Markov model to conclude that pairs of sequential sums of squares are independent — it's not much harder to argue that the whole set of sequential sums of squares are mutually independent — (see Kshohler 4.5 on page 333) —

This is the famous Cochran Theorem.