This is the element in the (j, j) position of

$$\frac{\mathbf{Q}_r\,\mathbf{Q}_r^\top\mathbf{y}}{\alpha} = \frac{\mathbf{I} - \mathbf{Q}_f\,\mathbf{Q}_f^\top\mathbf{y}}{\alpha}$$

i.e., it equals e_j/α, where e_j is the residual from the model that includes the j^{th} row. In place of $\mathbf{R}\boldsymbol{\beta} = \mathbf{f}$, the regression parameters are obtained by solving

$$\begin{bmatrix} \mathbf{R} & \mathbf{q}_{j(f)} \\ 0\ldots0 & \alpha \end{bmatrix}\begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \zeta \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \frac{e_j}{\alpha} \end{bmatrix}$$

Hence

$$e_{(j)} = \frac{e_j}{\alpha^2} \tag{19}$$

$$= \frac{e_j}{||\mathbf{q}_{j(r)}||^2} \tag{20}$$

$$= \frac{e_j}{1 - a_{jj}} \tag{21}$$

where a_{jj} is the j^{th} diagonal element of the matrix

$$\mathbf{A} = \mathbf{Q}_f\mathbf{Q}_f^\top = \mathbf{I} - \mathbf{Q}_r\mathbf{Q}_r^\top \tag{22}$$

Notice that we have not made any assumption of non-singularity.

2.6.1. The OCV mean square error

It follows from Equation 21 that the *ordinary cross-validation* (OCV) mean square error statistic:

$$OCV = \frac{1}{n}\sum_{j=1}^{n} e_{(j)}^2 = \frac{1}{n}\sum_{j=1}^{n}\frac{e_j^2}{(1 - a_{jj})^2} \tag{23}$$

This is a weighted sum of the squares of the ordinary residuals. High leverage points (large $1 - a_{ii}$ get the greatest weight.

In Subsection 2.6.2 that now follows, it will be argued that the *generalized cross-validation* (GCV) mean square prediction error estimate is preferable.

2.6.2. The GCV statistic

This continues the discussion above. The GCV statistic is based on an orthonormal rotation of the row space of $\mathbf{X}$, where the rotation is chosen to yield pseudo-observations that all have equal leverage.

If $\mathbf{P}$ is any orthonormal matrix

$$\begin{aligned} ||\mathbf{P}\mathbf{y} - \mathbf{P}\mathbf{X}\boldsymbol{\beta}||^2 &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top P^\top P(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 \end{aligned}$$

Thus minimization of $||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2$ is equivalent to minimization of $||\mathbf{P}\mathbf{y} - \mathbf{P}\mathbf{X}\boldsymbol{\beta}||^2$. The effect is to replace columns of $\mathbf{X}$ by columns of $\mathbf{PX}$, and elemement of $\mathbf{y}$ by elements of $\mathbf{Py}$. Where the ordinary least squares residuals were $\boldsymbol{\epsilon}$ they are now $\mathbf{P}\boldsymbol{\epsilon}$. We have created a set of pseudo-observations and pseudo-residuals, iid normal if the elements of $\boldsymbol{\epsilon}$ are iid normal. If the observations are replaced by these pseudo-observations, the OCV statistic changes.

The GCV statistic is the OCV statistic that results from choosing $\mathbf{P}$ so that all pseudo-observations have equal leverage. It is not necessary to find such a $\mathbf{P}$ in order to calculate the GCV statistic; all that is necessary is to show that such a choice is possible.[2]

Now suppose that, as before, premultiplication by $\mathbf{Q}_f^\top$ reduces $\mathbf{X}$ to upper triangular form. Then

$$\mathbf{Q}_f^\top\mathbf{P}^\top\mathbf{P}\mathbf{X} = \mathbf{Q}_f^\top\mathbf{P}^\top(\mathbf{P}\mathbf{X})$$

i.e., premultiplication by $\mathbf{Q}_f^\top\mathbf{P}^\top$ reduces $\mathbf{PX}$ to upper triangular form. The leverages for these pseudo-observations are diagonal elements of

$$\mathbf{P}\mathbf{Q}_r\mathbf{Q}_r^\top\mathbf{P}^\top = \mathbf{I}_n - \mathbf{P}\mathbf{Q}_f\mathbf{Q}_f^\top\mathbf{P}^\top$$

Moreover the sum of the leverages is

$$\begin{aligned}\text{trace}(\mathbf{P}\mathbf{Q}_r\mathbf{Q}_r^\top\mathbf{P}^\top) &= \text{trace}(\mathbf{Q}_r\mathbf{Q}_r^\top\mathbf{P}^\top\mathbf{P}) \\ &= \text{trace}(\mathbf{Q}_r\mathbf{Q}_r^\top) \\ &= \text{trace}(\mathbf{I} - \mathbf{Q}_f\mathbf{Q}_f^\top) \\ &= n - \sum_{i=1}^{n} a_{ii}\end{aligned}$$

The sum of the leverages is $n - \sum_{i=1}^{n} a_{ii}$, just as for the unrotated configuration. In the equal leverages configuration, the leverages are thus all equal to

$$\frac{1}{n}\,(n - \sum_{i=1}^{n} a_{ii})$$

The sum of squares of the estimated pseudo-residuals is unchanged. Ordinary cross-validation gives, for these pseudo-observations, the *generalized cross-validation* (GCV) estimate

$$\frac{n\sum_{i=1}^{n} e_i^2}{(n - \sum_{i=1}^{n} a_{ii})^2}$$

Not only does this have better theoretical properties than OCV; it is simpler computationally. It is, as for the OLS estimate, derived from the residual sum of squares.

The divisor $n - \sum_{i=1}^{n} a_{ii}$ has the same role as the degrees of freedom in the classical theory.

[2]One way to show this is to note that if two rows have different L_2 norms, then we can find a Givens rotation that equalizes the lengths. This process can be repeated until all rows are, to within numerical error, of equal length.