## Identities for the Leave-one-out LS estimates

To calculate these jackknife (or leave-one-out) LS estimates, such as $\hat{\boldsymbol{\beta}}_{(i)}$ and $\hat{\sigma}_{(i)}$, we don't need to run the model $n$ times. Using the Sherman-Morrison formula (a.k.a. Woodbury matrix identity), we have

$$\left[\mathbf{X}_{(i)}^t \mathbf{X}_{(i)}\right]^{-1} = (\mathbf{X}^t\mathbf{X})^{-1} + \frac{(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}_i^t(\mathbf{X}^t\mathbf{X})^{-1}}{1-h_i}.$$

Consequently,

$$\mathbf{x}_i^t(\mathbf{X}_{(i)}^t\mathbf{X}_{(i)})^{-1}\mathbf{x}_i = \frac{h_i}{1-h_i} \qquad \hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\beta}} - \frac{(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}_i r_i}{1-h_i}$$

$$y_i - \hat{y}_{(i)} = \frac{1}{1-h_i}(y_i - \hat{y}_i), \qquad \hat{\sigma}_{(i)}^2 = \hat{\sigma}^2 \frac{(n-p)-(r_i^*)^2}{n-p-1}$$

$$\begin{aligned}
\text{Jackknife Residual } t_i &= \frac{y_i - \hat{y}_{(i)}}{\hat{\sigma}_{(i)}\left[1+\mathbf{x}_i^t(\mathbf{X}_{(i)}^t\mathbf{X}_{(i)})^{-1}\mathbf{x}_i\right]^{1/2}} \\
&= \frac{r_i}{\hat{\sigma}_{(i)}\sqrt{1-h_i}} = r_i^*\left[\frac{n-p-1}{n-p-(r_i^*)^2}\right]^{1/2} \\
\text{Cook's distance } D_i &= \frac{\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}\|^2}{p\hat{\sigma}^2} = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^t(\mathbf{X}^t\mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})}{p\hat{\sigma}^2} \\
&= \frac{r_i^2}{p\hat{\sigma}^2}\frac{h_i}{(1-h_i)^2} = \frac{(r_i^*)^2}{p}\left(\frac{h_i}{1-h_i}\right)
\end{aligned}$$

where $r_i^* = r_i/(\hat{\sigma}\sqrt{1-h_i})$ is the standardized residual for the $i$th sample.

## The Leave-one-out Prediction Error

We provide an alternative proof for the leave-one-out prediction error

$$y_i - \hat{y}_{(i)} = \frac{1}{1-h_i}(y_i - \hat{y}_i). \tag{1}$$

Consider two regression models:

- *Model I*: the regression model on the $n$ data points $\mathbf{y}_{n\times 1}$, where

$$\hat{y}_i = h_i y_i + \sum_{j\neq i} H_{ij} y_j. \tag{2}$$

- *Model II*: the regression model on a set of new data points $\mathbf{y}^*_{n\times 1}$, which are almost the same as the original data $\mathbf{y}$ except the $i$th element,

$$y_i^* = \hat{y}_{(i)} = \mathbf{x}_i^t\hat{\boldsymbol{\beta}}_{(i)},$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ denotes the leave-one-out (leave the $i$th sample out) LS coefficient. Note that the two regression models have the same projection matrix since their design matrices are the same. We have

$$\hat{y}_i^* = h_i y_i^* + \sum_{j \neq i} H_{ij} y_j^* = h_i y_i^* + \sum_{j \neq i} H_{ij} y_j, \tag{3}$$

where $H_{ij}$'s and $h_i$ are the same as the ones in (2).

We can show that the LS line for *Model II* passes the $i$-th sample, that is

$$\hat{y}_i^* = y_i^* = \hat{y}_{(i)}. \tag{4}$$

Then (3) becomes

$$\hat{y}_{(i)} = h_i \hat{y}_{(i)} + \sum_{j \neq i} H_{ij} y_j = h_i \hat{y}_{(i)} - h_i y_i + \sum_{j} H_{ij} y_j = h_i \hat{y}_{(i)} - h_i y_i + \hat{y}_i.$$

Then you can get the conclusion (1) by re-arranging the terms above.

**Proof for eq (4)**: Let $\hat{\boldsymbol{\alpha}}$ denote the corresponding LS coefficient for *Model II*, that is,

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \arg\min_{\boldsymbol{\alpha}} \sum_{j=1}^{n} (y_j^* - \mathbf{x}_j^t \boldsymbol{\alpha})^2 \\ &= \arg\min_{\boldsymbol{\alpha}} \Big[ (\hat{y}_{(i)} - \mathbf{x}_i^t \boldsymbol{\alpha})^2 + \sum_{j \neq i} (y_j - \mathbf{x}_j^t \boldsymbol{\alpha})^2 \Big]. \end{aligned}$$

It is not difficult to show that $\hat{\boldsymbol{\beta}}_{(i)}$ the minimizer of the optimization above since it'll make the first term equal to zero and meanwhile minimize the 2nd term. So we have

$$\hat{\boldsymbol{\beta}}_{(i)} = \hat{\boldsymbol{\alpha}}.$$

So the LS fit for the $i$th sample (in *Model II*) is

$$\hat{y}_i^* = \mathbf{x}_i^t \hat{\boldsymbol{\alpha}} = \mathbf{x}_i^t \hat{\boldsymbol{\beta}}_{(i)} = \hat{y}_{(i)}.$$