Why High Leverage Observations Tend to have Small Residuals

The hat matrix is

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}', \tag{1}$$

where $\boldsymbol{X}$ is the design matrix. The vector of fitted values can be written as

$$\hat{\boldsymbol{Y}} = \boldsymbol{H}\boldsymbol{Y}. \tag{2}$$

Residuals can be written in vector form as

$$\begin{aligned} \boldsymbol{e} &= \boldsymbol{Y} - \hat{\boldsymbol{Y}} \\ &= \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y} \qquad (3) \\ &= (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} \qquad (4) \end{aligned}$$

Since the variance-covariance matrix of $\boldsymbol{Y}$ is $\sigma^2\boldsymbol{I}$, the variance-covariance matrix of the residuals is

$$\begin{aligned} \boldsymbol{Var(e)} &= \sigma^2(\boldsymbol{I} - \boldsymbol{H})'(\boldsymbol{I} - \boldsymbol{H}) \\ &= \sigma^2(\boldsymbol{I} - \boldsymbol{H}) \qquad (5) \end{aligned}$$

One can easily show that $(\boldsymbol{I} - \boldsymbol{H})$ is idempotent, i.e. $(\boldsymbol{I} - \boldsymbol{H})'(\boldsymbol{I} - \boldsymbol{H}) = (\boldsymbol{I} - \boldsymbol{H})$. Therefore, the variance of e_i is

$$Var(e_i) = \sigma^2(1 - h_{ii}) \tag{6}$$

where h_{ii} is the ith element on the main diagonal of the hat matrix, and the covariance is

$$Cov(e_i, e_j) = -h_{ij}\sigma^2 \text{ for } i \neq j \tag{7}$$

where h_{ij} is the (i, j)th element the hat matrix. We make the following notes from (6) and (7):

- From (6), observations that high leverage (h_{ii} close to 1) will have a small variance. We know the mean of residuals is 0, so this means that residuals from high leverage observations tend to be close to 0, and so tend to be small.
- (6) also implies that while the variance of errors is constant, the variance of residuals is not exactly constant.
- From (7), residuals are not uncorrelated.

If $n >> p$, i.e. if the sample size is large relative to the number of parameters in the model, the elements in the hat matrix tend towards 0. This means that the variance of the residuals is approximately constant, and the covariance of residuals is approximately 0.