Exercise D

Suppose that $y = X\beta + z\gamma + \varepsilon$ but we "omit" the variable z, using the OLS estimator $\hat{\beta} = (X^T X)^{-1} X^T y$, obtaining $r = (I - H)y$.

a) The residuals are

Because of orthogonality of $I - H$ with X.

$$r = (I - H)y = \cancel{(I-H)X\beta} + (I - H)z\gamma + (I - H)\varepsilon$$
$$= (I - H)z\gamma + (I - H)\varepsilon.$$

Therefore $\mathbb{E}[r] = (I - H)z\gamma$. In particular, this is not 0! If z lies in the column space of X, say $z \in \mathcal{C}(X)$, then the matrix $I_n - H$ is orthogonal to z and

$$r = \cancel{(I_n - H)}z\gamma + (I - H)\varepsilon = (I_n - H)\varepsilon,$$

The usual structure!

Consequently $\mathbb{E}[r] = 0$. Intuitively, if $z \in \mathcal{C}(X)$, then there is no "missing information", because z is obtained as a linear combination of the other variables. Even though γ cannot be estimated due to identifiability, the linear predictor is correctly specified.

In the opposite case, i.e. $z \perp\!\!\!\perp X$ meaning that z is orthogonal to X, then:

$$r = (I - H)z\gamma + (I - H)\varepsilon = z\gamma + (I - H)\varepsilon.$$

Note that $Hz = 0$, so $(I - H)z = z\gamma$. Moreover

$$\mathbb{E}[r] = z\gamma.$$

b. The added-variable plot is useful because it compares the (noisy) residuals r with their mean $(I_n - H)z$ (up to a proportionality constant). Note that

$$(I_n - H)z = \text{"residuals of } z \text{ using } X \text{ as covariates"}.$$

The "naïve" plot z vs residuals works only if the missing covariate is orthogonal to X, otherwise the signal is less evident.

c. The new variable z should definitely be included in the model. The added-variable plot indicates that much more clearly that the naive plot.