Regression Robustness (and rough rules of thumb)

Statistic	All assumptions OK $Y_i = \mu_{x_i} + \varepsilon_i$ $\mu_{x_i} = \boldsymbol{x_i} \cdot \vec{\boldsymbol{\beta}}$ $\varepsilon_i \sim N(0, \sigma^2_{x_i})$ $\sigma^2_{x_i} = \sigma^2 \forall i$ Cor$(\varepsilon_i, \varepsilon_j) = 0$ Cor$(\varepsilon_i, x_{ji}) = 0$	Hetero-skedasticity $\sigma^2_{x_i}$ depends on $\boldsymbol{x_i}$	Non-normality $\varepsilon_i \sim F$	Non-linearity $\mu_{x_i} = f(\boldsymbol{x_i} \cdot \vec{\beta})$	Non-independent errors Cor$(\varepsilon_i, \varepsilon_j) \neq 0$	X not fixed Var$(x_j) \neq 0$ for some j
$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ Unbiased: E$(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ $E(Y\|x_j+1) = E(Y\|x_j)+\beta_j$	yes	yes	yes	**N/A** except over a limited range when very mild	yes	robust up to, say, V(x)<0.2V(y)
$\widehat{\sum} = \boldsymbol{\sigma^2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ Variance estimate unbiased: E$\left(\widehat{\sum}\right) = V(\widehat{\boldsymbol{\beta}})$	yes	robust up to, say, $\sigma_{\max} < 2\,\sigma_{\min}$	robust to moderately severe non-normality	**N/A** except over a limited range when very mild	robust to mild correlation of errors, say, $cor(\varepsilon_i,\varepsilon_{i'})<0.2$	robust up to, say, V(x)<0.2V(y)
95% CI = $[\hat{\beta}_j \pm t_{0.975}\Sigma_{j+1,j+1}]$ holds 95% of the sampling distribution of $\hat{\beta}$	yes	robust up to, say, $\sigma_{\max} < 2\,\sigma_{\min}$	robust to moderately severe non-normality	**N/A** except over a limited range when very mild	robust to mild correlation of errors, say, $cor(\varepsilon_i,\varepsilon_{i'})<0.2$	robust up to, say, V(x)<0.2V(y)
95% PI holds 95% of the sampling distribution of $\hat{\beta}$	yes	less robust	**not robust**	**N/A** except over a limited range when very mild	less robust	robust up to, say, V(x)<0.2V(y)
p-value <0.05 5% of the time when H_0 is true	yes	robust up to, say, $\sigma_{\max} < 2\,\sigma_{\min}$	robust to moderately severe non-normality	**N/A** except over a limited range when very mild	robust to mild correlation of errors, say, $cor(\varepsilon_i,\varepsilon_{i'})<0.2$	robust up to, say, V(x)<0.2V(y)

Cor$(\varepsilon_i, x_{ji}) \neq 0$ (endogenicity): $\hat{\beta}$ and its variance are biased; CI, PI, and p-value are wrong; robust to small corr. of x_j and a missing x’