Prediction

<u>Setup</u>: Model relating Y to $(1, x_1, \ldots, x_p)^T = \underline{x}$:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon = \underline{\beta}^T \underline{x} + \varepsilon$$

and OLS estimates $\hat{\underline{\beta}}$ (plus something for $\hat{\sigma}^2$),

there are two new quantities we want to predict at a new set of features $\underline{x}_* = (1, x_{1*}, \ldots, x_{*p})^T$:

- Average response / mean value $\quad \underline{\beta}^T \underline{x}_* \quad$ [fit]

- New obs $\quad \underline{\beta}^T \underline{x}_* + \varepsilon_* \quad$ [prediction]

Both cases use same point predictor:

$$\hat{\underline{\beta}}^T \underline{x}_* = \underline{x}_*^T \hat{\underline{\beta}} = \underline{x}_* (X^T X)^{-1} X^T \underline{Y}$$

but, uncertainty depends on the case!

Recall standard error is an estimate of standard deviation

$$SE(\cdot) = \sqrt{\hat{Var}(\cdot)}$$

and

$$Var(\underline{x}^T \hat{\beta}) = Cov(\underline{x}^T \hat{\beta}, \underline{x}^T \hat{\beta}) = \underline{x}^T (Var\,\hat{\beta}) \underline{x}.$$

So,

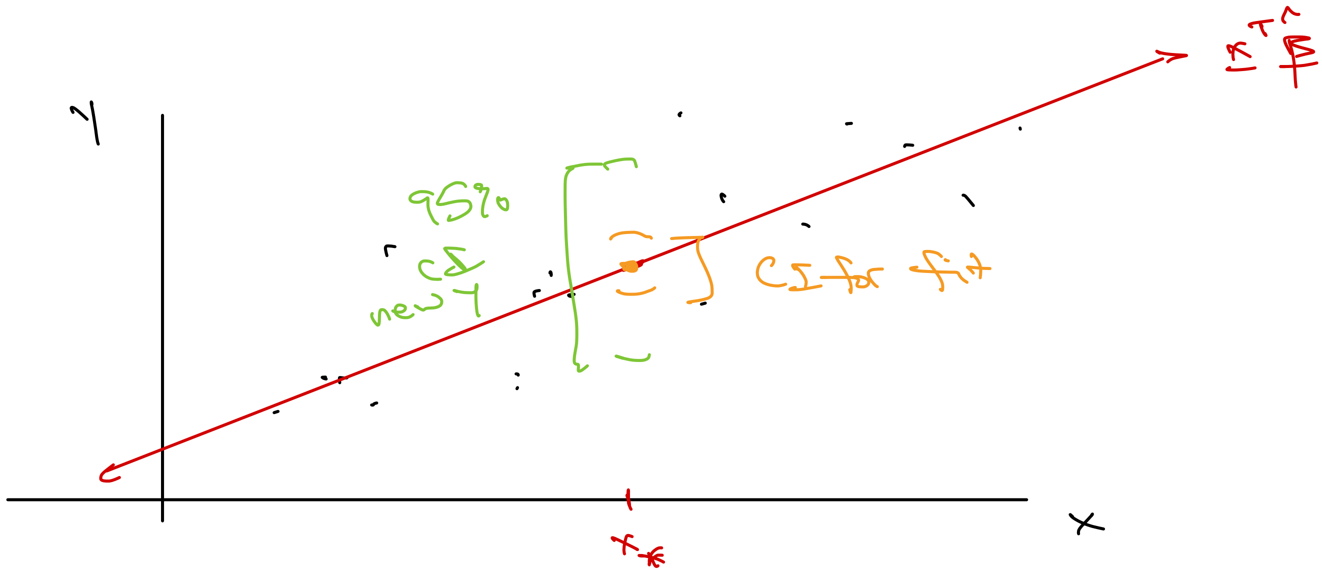
$$\hat{Var}(\underline{x}_*^T \hat{\beta}) = \hat{\sigma}^2 \underline{x}_*^T (X^T X)^{-1} \underline{x}_* \quad \text{[fit]}$$

$$\hat{Var}(\underline{x}_*^T \hat{\beta} + \varepsilon_*) = \hat{\sigma}^2 \underline{x}_*^T (X^T X)^{-1} \underline{x}_* + \hat{\sigma}^2 \quad \text{[prediction]}$$

Thus 95% approx conf. intervals are:

fit: $\quad \underline{x}_*^T \hat{\beta} \pm 1.96\, SE(\underline{x}_*^T \hat{\beta})$

prediction: $\quad \underline{x}_*^T \hat{\beta} \pm 1.96\, SE(\underline{x}_*^T \hat{\beta} + \varepsilon_*)$



4.5 Diagnostics

Have model

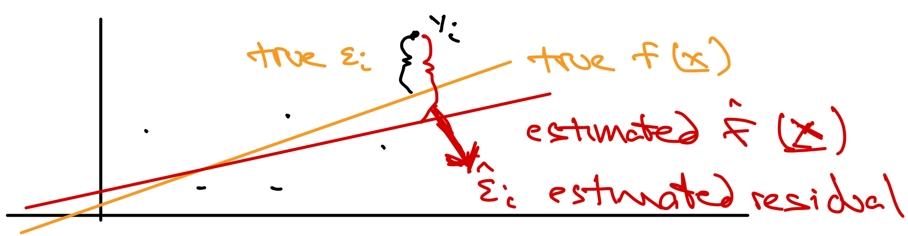
$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

+ estimators $\hat{\underline{\beta}}, \hat{\sigma}^2$. Is the model any good, and are $\{\varepsilon_i\}$ approximately normal?

DEF

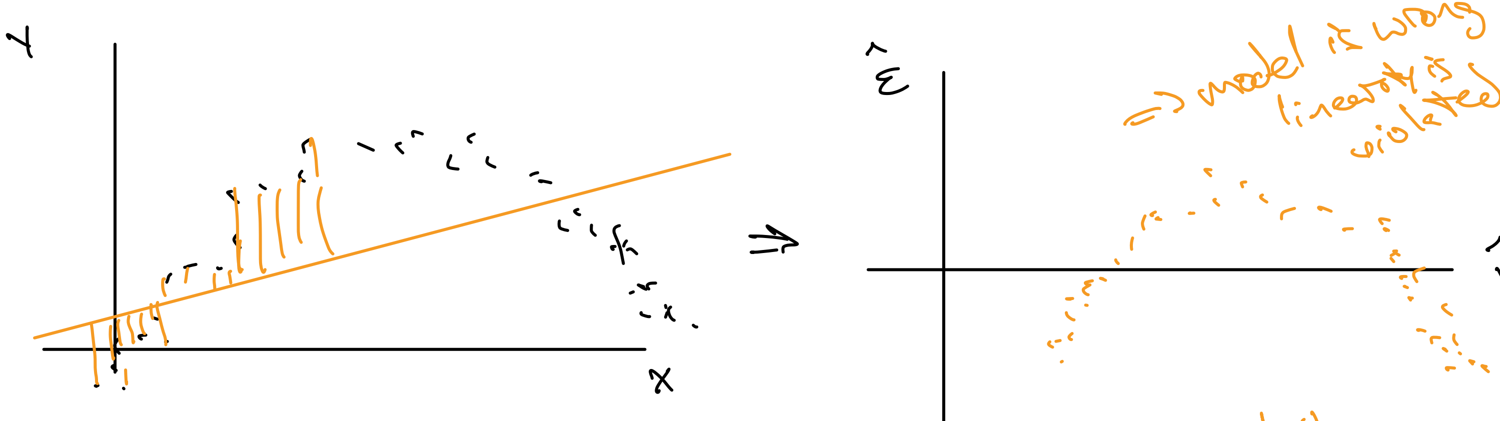
<u>Fitted values</u> $\hat{Y}_i = \underline{x}_i^T \hat{\underline{\beta}}$, $\quad \hat{\underline{Y}} = X\hat{\underline{\beta}}$ $\quad [\neq \underline{Y} = X\underline{\beta} + \underline{\varepsilon}]$

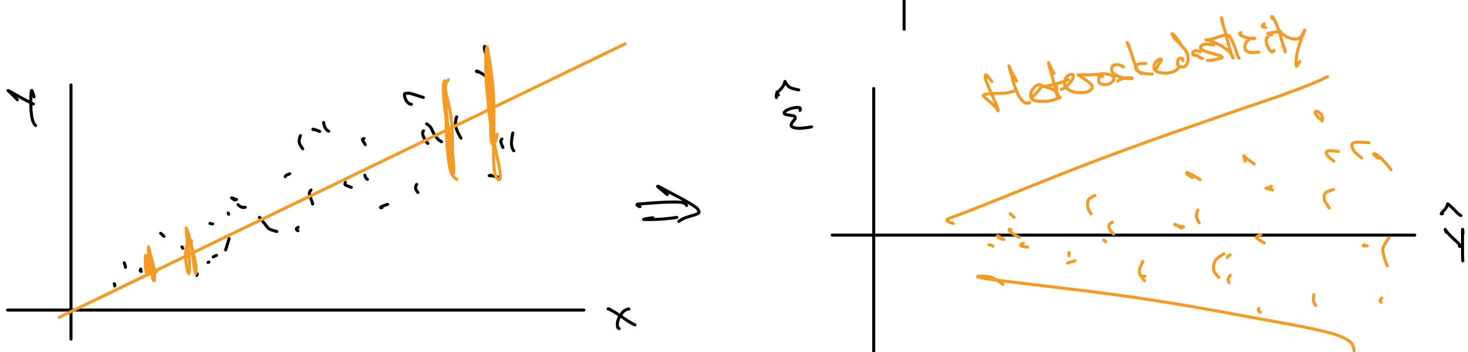
<u>Estimated residuals</u> $\hat{\underline{\varepsilon}} = \underline{Y} - X\hat{\underline{\beta}}$ $\quad [\neq \underline{\varepsilon} = \underline{Y} - X\underline{\beta}]$



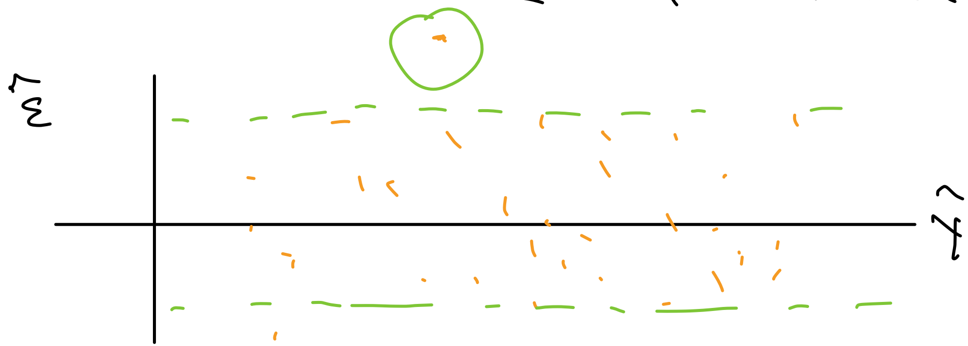
If assumptions are correct, $\hat{\varepsilon}$ should have no structure.

① Fitted values vs. estimated! residuals plot



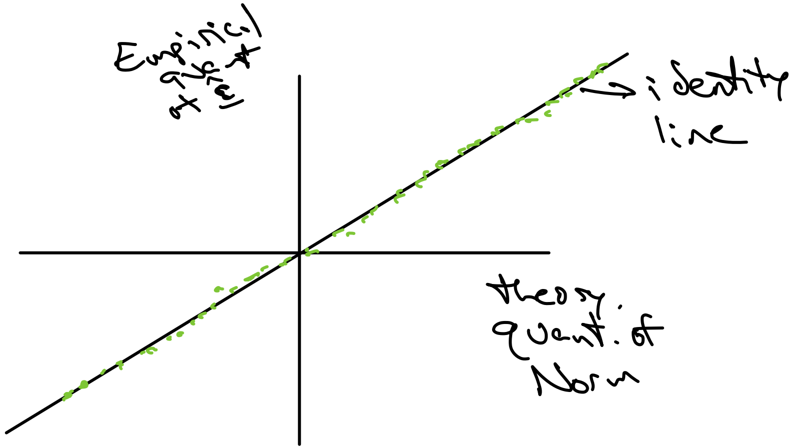


(2) Unusual values of $\hat{\varepsilon}$ may be evidence of outliers

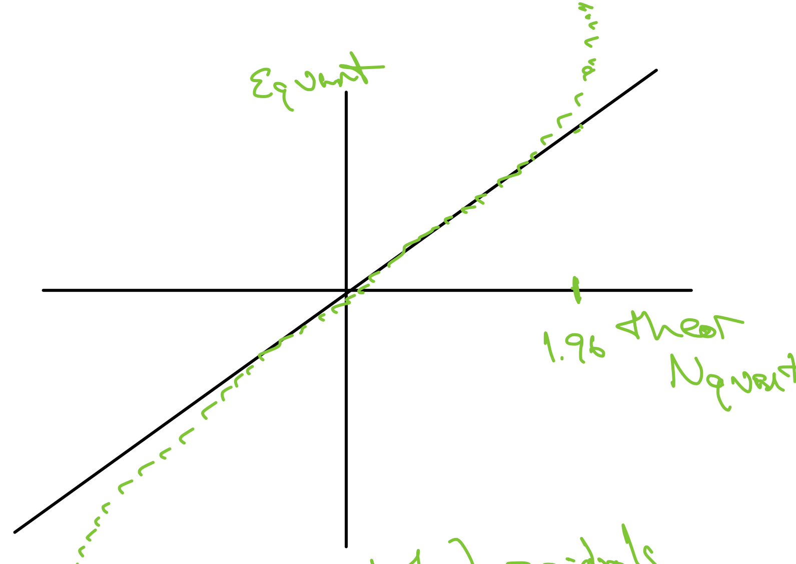


Outliers do not necessarily affect $\hat{\beta}$, but do inflate $\hat{\sigma}^2$

(3) Quantile-quantile plot can assess normality, plots theoretical quantiles of a normal against empirical quantiles of (standardized) $\hat{\varepsilon}$.



If $\{\varepsilon_i\}$ are approx normal



Heavy-tailed residuals