Regression Diagnostics: Ch 9

15.1

To detect problems with the model, like

- outliers
- Curvilinear trends ← not the same as non-linear
- Non-constant variance

Scatterplots are great if there is only one independent variable, but not practical for 3D+

Can look at many plots including

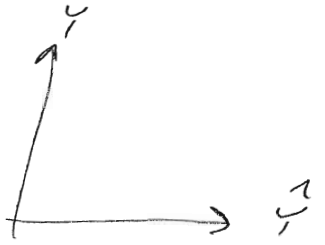
- $\hat{Y}$ vs Y
- $\hat{Y}$ vs $\hat{\varepsilon}$
- X vs $\hat{\varepsilon}$
- Time sequence vs $\hat{\varepsilon}$

Will need

$$r_{xy} = \frac{\sum (X_i - \bar{x})(Y_i - \bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

Round up

$\hat{Y}_i$ vs Y_i



15.2

What should we see?

Squared sample correlation between $\hat{Y}$ & Y equals $R^2 = \frac{SSR}{SST}$

Proof: Recall in decomposition of SS, that

$$SST = SSE + 0 + SSR$$

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{Y}_i)^2 + 2\sum (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) + \sum (\hat{Y}_i - \bar{Y})^2$$

And, since $\sum \hat{\varepsilon}_i = \sum (Y_i - \hat{Y}_i) = 0$,

$\bar{\hat{Y}} = \bar{Y}$, so sample correlation between $\hat{Y}$ & Y is

15.3

$$R^2 = \left(\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\Sigma (Y_i - \bar{Y})^2} \sqrt{\Sigma (\hat{Y}_i - \bar{Y})^2}} \right)^2$$

$$= \left(\frac{\Sigma (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{SST \times SSR}} \right)^2$$

$$= \frac{\left(\Sigma (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) + \Sigma (\hat{Y}_i - \bar{Y})^2 \right)^2}{SST \times SSR}$$

($\Sigma (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) \to 0$)

$$= \frac{SSR \times SSR}{SST \times SSR} = R^2$$

So the plot of $\hat{Y}$ vs Y is the closest thing we have to a single scatterplot

Could reveal

Outliers
curvilinear trends,
Non-constant variance

15.4

What if it looks like there are outliers?

When is a residual "too big", i.e. absolute value >

→ One problem is that while ε_i all have same variance, $\hat{\varepsilon}_i$ do NOT

Recall $\hat{\varepsilon} = (y - \hat{y}) = (I - H)\, y$

Where $H = X(X'X)^{-1}X'$ is the "Hat" matrix, so called because it puts a hat on y, by

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy$$

So $\operatorname{cov}(\hat{\varepsilon}) = (I - H)\sigma^2 I_n (I - H)'$

$$= \sigma^2 (I - H)$$

Writing $H = [h_{ij}]$, $\operatorname{Var}(\hat{\varepsilon}_i) = \sigma^2(1 - h_{ii})$

Could standardize it

$$\pi_i \quad \frac{\hat{\varepsilon}_i - 0}{\sqrt{\sigma^2(1 - h_{ii})}} \approx \frac{\hat{\varepsilon}_i}{\sqrt{s^2(1 - h_{ii})}} = \frac{\hat{\varepsilon}_i}{s\sqrt{1 - h_{ii}}}$$

The "Studentized" residual

15.5

Studentized residual has (for large samples) an approximate standard normal dist, but not t, bec num & MSE_h are not $\beta_{1 \times p}$ independent.

Better is the STUDENTIZED DELETED RESIDUAL.

Do a regression omitting observation i, obtaining $\hat{\beta}_{(i)}$. Calculate (see 9.3, p. 234)

Let $\hat{\varepsilon}_{(i)} = Y_i - x_i' \hat{\beta}_{(i)}$

Notice parallel to prediction interval

$$\hat{\varepsilon}_{(i)} \sim N\left(0, \sigma^2 + \sigma^2 x_i'(X_{(i)}' X_{(i)})^{-1} x_i\right)$$

$$Z = \frac{\hat{\varepsilon}_{(i)}}{\sqrt{\sigma^2\left(1 + x_i'(X_{(i)}' X_{(i)})^{-1}\right)}}$$

And $t_i = \frac{Z_i}{\sqrt{W/(n-1-k-1)}} = \frac{Z_i}{S_{(i)}}$, $S^2_{(i)} = \frac{SSE_{(i)}}{n-k-2}$

$$W = \frac{SSE_{(i)}}{\sigma^2} \sim \chi^2(n-k-2)$$

ind of Y_i & $\hat{\beta}_{(i)}$

15.6

$$t_i = \frac{\hat{\varepsilon}_{(i)}}{\sqrt{\sigma^2(1 + x_i'(X_{(i)}' X_{(i)})^{-1} x_i)} \sqrt{\frac{1}{\sigma^2} s^2_{(i)}}}$$

$$= \frac{\hat{\varepsilon}_{(i)}}{s_{(i)} \sqrt{1 + x_i'(X_{(i)}' X_{(i)})^{-1} x_i}} \sim A(n-k-2)$$

If the model is correct.

This is a good definition of "too big"

We are conducting n tests

(not independent)

If model is correct,

- Pr rejecting each H_0 is $0.05 = \alpha$
- Pr of rejecting at least one is much greater
- How much greater?

15.7

How much greater?

Suppose $\alpha = 0.05$, all k null hypotheses are true, and tests are independent (which they are not here). Then letting A_j represent event reject H_{0j},

$$P(\text{Reject at least one}) = P\left(\bigcup_{j=1}^{k} A_j\right)$$

$$= 1 - P\left\{\left(\bigcup_{j=1}^{k} A_j\right)^c\right\}$$

$$\overset{\text{De Morgan}}{=} 1 - P\left\{\bigcap_{j=1}^{k} A_j^c\right\} \overset{\text{ind}}{=} 1 - \prod_{j=1}^{k} P(A_j^c)$$

$$= 1 - \prod_{j=1}^{k} (1 - P(A_j)) = 1 - (1-\alpha)^k$$

So for $\alpha = 0.05$ & 50 tests (remember, n tests here), JOINT significance level of the 50 tests is

$$1 - (0.95)^{50} = 0.923$$

But tests are not independent

15.8

Bonferroni's Inequality

$$P\left(\bigcup_{j=1}^{k} A_j\right) \leq \sum_{j=1}^{k} P(A_j)$$

- A_j is event that test j rejects H_0
- Applies to any collection of tests
- Do the tests as usual, but with significance level α/k. Then

$$P\left(\bigcup_{j=1}^{k} A_j\right) \leq \sum_{j=1}^{k} \alpha/k = k\alpha/k = \alpha$$

A bit conservative, but flexible & SIMPLE

Use the critical value for α/k instead of α

For using Studentized deleted residuals, $k = n$. Use the α/n critical value

15.9

Plot of $\hat{Y}$ vs $\hat{\varepsilon}$.

What should you see if the model is correct?

Nothing, because they are independent.

USE
- $E(\hat{\beta}\hat{\beta}') = \sigma^2(X'X)^{-1} + \beta\beta'$
- $E(YY') = \mathrm{cov}(Y) + (EY)(EY)'$
$= \sigma^2 I_n + X\beta\beta'X'$

$$C(\hat{Y}, \hat{\varepsilon}) = E(\hat{Y} - X\beta)(\hat{\varepsilon} - 0)'$$

$$= E(\hat{Y}\hat{\varepsilon}) = E\{(X\hat{\beta})(Y - X\hat{\beta})'\}$$

$$= E\{X\hat{\beta}Y'\} - E(X\hat{\beta}\hat{\beta}'X')$$

$$= X(X'X)^{-1}X'E(YY') - XE(\hat{\beta}\hat{\beta}')X'$$

$$= X(X'X)^{-1}X'(\sigma^2 I_n + X\beta\beta'X')$$

$$- X(\sigma^2(X'X)^{-1} + \beta\beta')X'$$

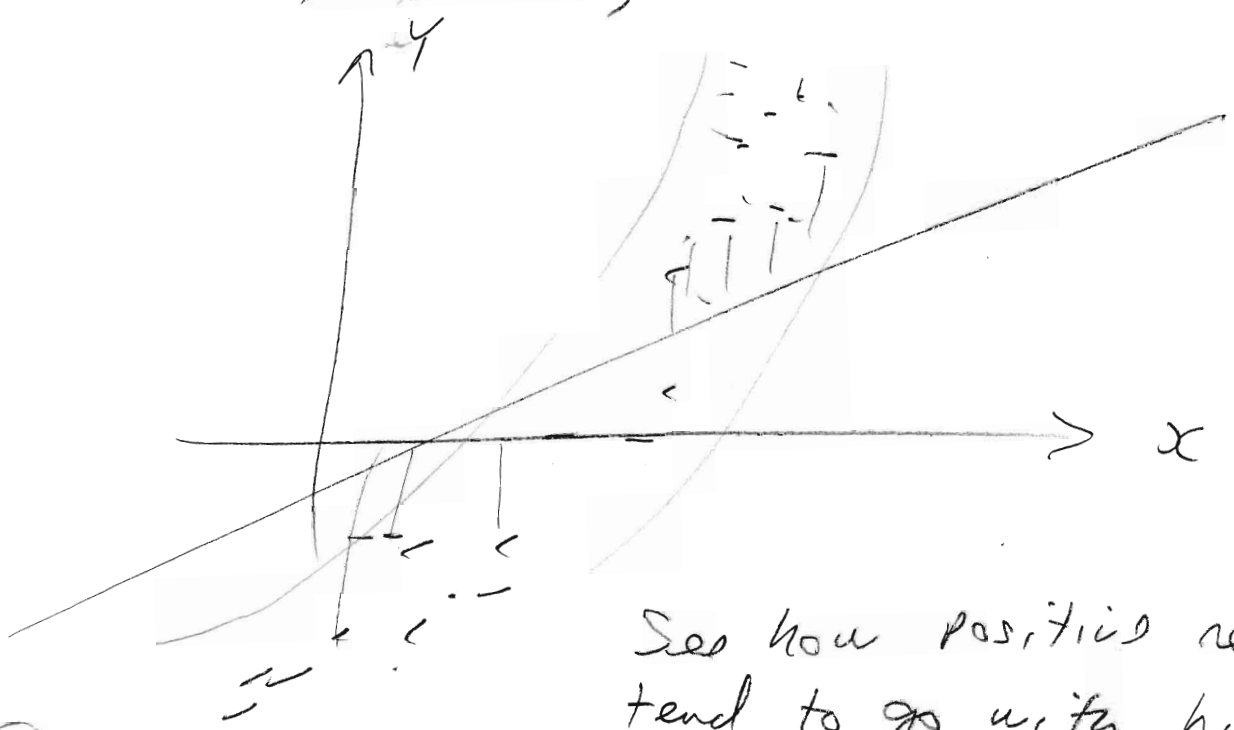
$$= \sigma^2 X(X'X)^{-1}X' + X\underbrace{(X'X)^{-1}X'X}_{I}\beta\beta'X'$$

$$- \sigma^2 X(X'X)^{-1}X' - X\beta\beta'X' = 0$$

What might you see if model is wrong?

15.10

Well, outliers, or



See how positive residuals tend to go with higher predicted y's

Not nice linear pattern

So we might see evidence of

- OUTLIERS
- CURVILINEAR PATTERN
- Systematic non-constant variance, like more error in the prediction of large values

(Reasonable: Predict restaurant tip.)

To see systematic non constant variance, better plot standardized or (deleted) studentized Resid.

15.11

Plot of x variable vs Residuals. What should you see? Recalling

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}},$$

Look at the sample correlation between x_j & $\hat{\varepsilon}_i$ values. Assuming the model has an intercept so $\sum \hat{\varepsilon}_i = 0$, numerator is

$$\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(\hat{\varepsilon}_i - 0) = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)\hat{\varepsilon}_i = \sum_{i=1}^{n} x_{ij}\hat{\varepsilon}_i - \bar{x}_j \sum_{i=1}^{n} \hat{\varepsilon}_i$$

$$= \sum_{i=1}^{n} x_{ij}\hat{\varepsilon}_i, \text{ Element } j \text{ of } \begin{bmatrix} 1 & 1 \\ x_{11} & \\ & x_{n1} \end{bmatrix} \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = X'\hat{\varepsilon}$$

$$X'\hat{\varepsilon} = X'(Y - X\hat{\beta}) = X'Y - X'X(X'X)^{-1}X'Y = X'Y - X'Y = 0$$

So the sample correlation between IV & Residual is <u>exactly</u> zero, whether the model is correct or not.

There will be no <u>linear</u> trend, but

- Possibly curvilinear
- If curvilinear, symmetric

Regardless of the shape of the curve relating X to Y

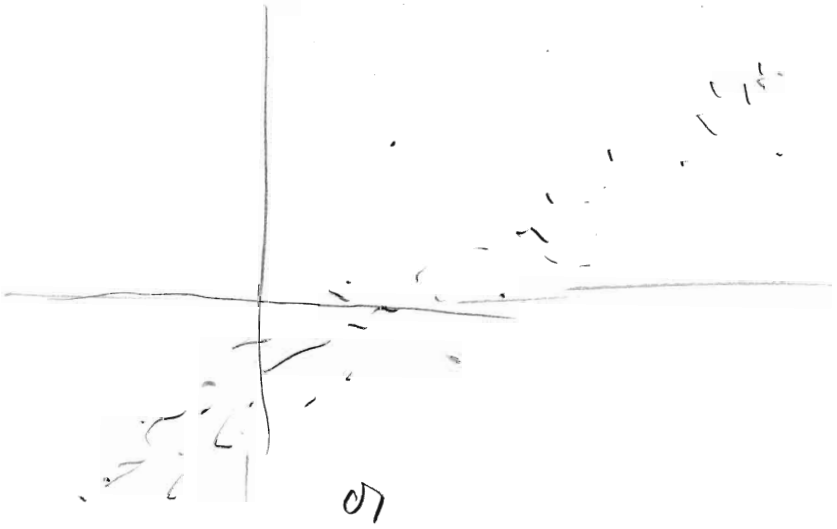
15.12

Plots of x vs $\hat{\varepsilon}$

- could also show
  - Systematically non-constant variance

If data are collected in time sequence,

- Plot time vs $\hat{\varepsilon}$ or r or t



Tendency to under-predict earlier, over-predict later.

Betrays sloppy data collection

or

Maybe prediction getting better after some practice

15.13

Standardized (or deleted Studentized) residuals are

- Easier to interpret (What's "BIG") and
- Have constant variance, so it's better to plot them rather than the raw residuals.

	X	Y	$\hat{Y}_i$	$\hat{\varepsilon}$
X	X	ORDINARY SCATTERPLOT	Pos or neg slope acc to β_1	Zero correlation, guaranteed if model is correct. Should see NO
Y		X	$R^2_{\hat{Y}\hat{Y}} = R^2$ Nice scatterplot showing overall strength of rel.	
$\hat{Y}_i$			X	Independent if model is correct. Should see NOTHING
$\hat{\varepsilon}$				X