Math 455 Class 25 March 28

6.3 "Checking Structure"

Fundamental Question: Is a linear model in the variables we have the right model?

In the simple case (1 explanatory variable) we can graph the points (scatter plot) and the line.

This isn't possible in many variables.

Added Variable and Partial Residual plots are an attempt to remedy this.

<u>Added Variable plot</u>: shows unique contribution of the predictor, given the presence of the other predictors

<u>Partial Residual plot</u>: shows nonlinearity in predictor (again in presence of other predictors)

	X-axis	Y-axis
Added Variable	Residuals of X_i ~ other pred.	Residuals of Y ~ other pred.
Partial Residual	X_i	$\hat{\beta}_i x_i + e_i$ (Y_i with effect of other predictors removed.)

Note: homework due date was extended.

New homework will be assigned today.

We also need to talk about 7.1:

errors-in-variables.

Let's say we're studying a simple linear reg.

$Y = \beta_0 + \beta_1 X + \varepsilon$

Suppose X is measured with an error.

or Y ________________

What difference does this make to the estimate $\hat{\beta}_1$?

Some possibilities:

The estimate is unchanged

The estimate becomes larger

__________________ smaller.

The estimate could increase or decrease?

Does it make a difference if the errors are in X or in Y?

Errors in X reduce the magnitude $|\hat{\beta}_1|$ of our coefficient estimate.

This is called "Attenuation Bias".

For a much more general treatment of possible errors, see formula of p. 100 (Ch 7)

"After some calculation" = Exercise math 447.

How can this be tested?

A simplified model where some errors are 0 is presented; you do the math 447 exercise and explain the result.

How does the formula of p. 100

$$E[\hat{\beta}_1] = \beta_1 \frac{\sigma_x^2 + \sigma_{x\delta}}{\sigma_x^2 + \sigma_\delta^2 + 2\sigma_{x\delta}}$$

Explain "Attenuation Bias"?

Assume we have errors only in X.

$\sigma_\delta^2 > 0$. But $\sigma_\varepsilon^2 = 0$

(no errors in Y) and $\sigma_{x\delta} = 0$.

No covariance of X with error in X.

Then:

$$E[\hat{\beta}_1] = \beta_1 \underbrace{\left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_\delta^2} \right)}_{\text{abs value } < 1}$$

"Smaller magnitude" or "Attenuation Bias".

Lots of other ideas in this formula.

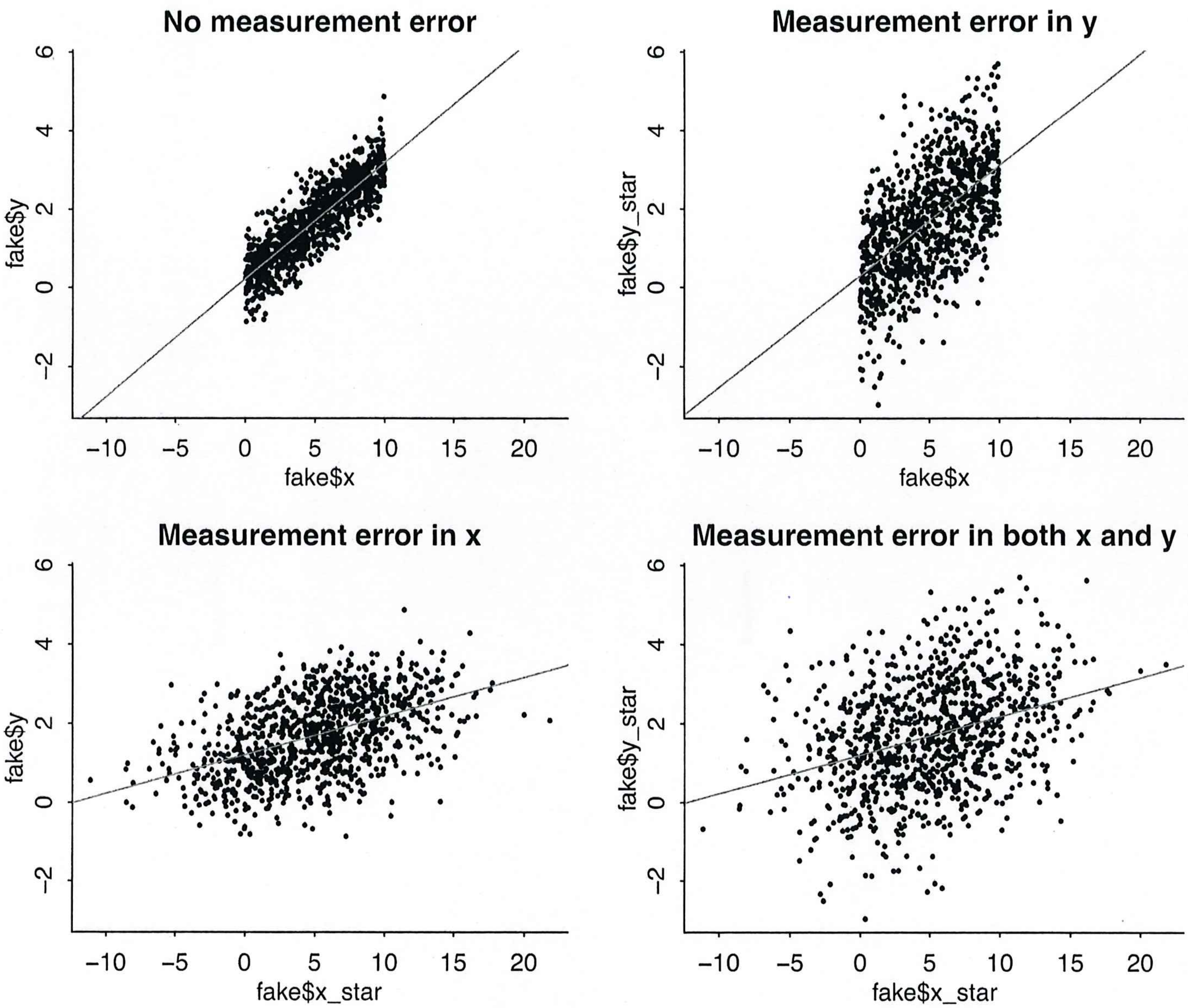
Pictures taken from A. Gelman

From "Gelman e.t. al. "Regression and other Stories".

One idea: You can learn a lot by making up data.

Special case: Before you do your experiment, make up some data and analyze that!

You might see a need for better experimental design.


No measurement error
Measurement error in y
Measurement error in x
Measurement error in both x and y
fake$y
fake$y_star
fake$x
fake$x_star
−10
−5
0
5
10
15
20
−2
2
4
6