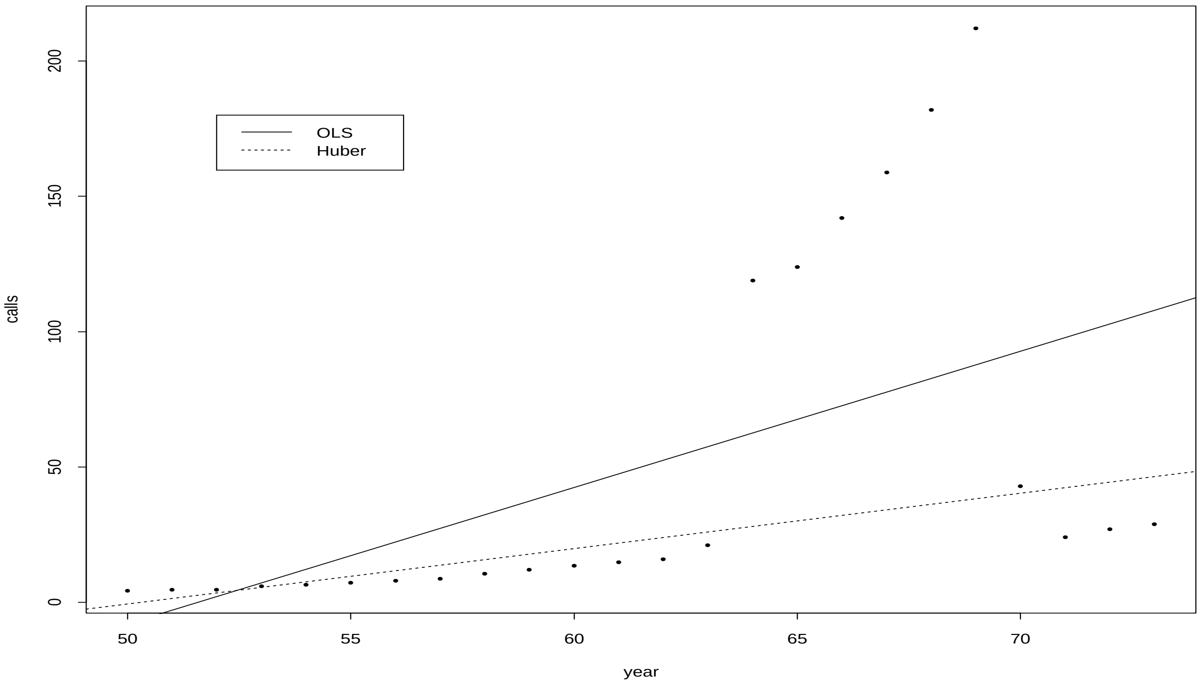
36-401 Class Notes and Exercise for 11/14/00
Robust Regression (RwG ch. 6)
Howard Seltman

1. Main ideas

   - Eliminating outliers (or using a single-case dummy variable) is an all-or-nothing proposition. Robust regression *down-weights* outliers according to one of several possible approaches.
   - The method is iterative, re-weighted least squares (IRLS).
   - When all regression assumptions are met, ordinary least squares performs better than robust regression. Robust regression often performs better for heavy-tailed distributions, e.g. in the presence of outliers.



2. Comparison to weighted least square (WLS): Weighted regression, e.g. lm(y∼x, weights=1/n), is based on user supplied weights. Robust regression uses *weight functions* to estimate and re-compute weights *iteratively* until convergence.

3. Terminology
    - *Robust:* Performs well under small violations of assumptions.
    - *Resistant:* Not changed much when there are small changes in the sample data (small changes in many cases or large changes in a few cases). Usually implies robustness.
    - OLS is inefficient for violated assumptions, e.g. heavy tailed errors (not robust). Slopes can change large amounts if a single data point is moved (not resistant).
4. Goals of Robust Regression
    - Consistency and reasonable efficiency when the assumed model is true
    - Only slightly impaired by small departures from the assumptions
    - Not drastically affected by larger departures from the assumptions
5. Example: Measures of spread
    - Mean Absolute Deviation:

$$d_n = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

    $E(d_n) = \sqrt{2/\pi}\sigma \approx 0.80\sigma$ for $x \sim N(\mu, \sigma^2)$
    - Mean Squared Deviation:

$$s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

    $E(s_n) = \sigma$
    - Asymptotic relative efficiency (ARE) of d_n relative to s_n

$$\text{ARE}(\epsilon) = \lim_{n\to\infty} \frac{\text{var}(s_n)/[E(s_n)]^2}{\text{var}(d_n)/[E(d_n)]^2}$$

    Data: $(1-\epsilon)$ from N(0,σ^2) mixed with ϵ from N(0,9σ^2)

ϵ	**ARE(ϵ)**
0	0.876
0.001	0.948
0.002	1.016
0.005	1.198
0.02	1.752
0.10	1.903
0.50	1.017
1.00	0.876

6. M-Estimation and Iteratively Re-weighted Least Squares

- M-estimators minimize the sum of a function, ρ, of the scaled residuals, u_i where $u_i = \frac{e_i}{s}$. OLS is an M-estimator where $s = 1$ and $\rho(u_i) = u_i^2$.
- We seek

$$\underset{b}{\operatorname{argmin}} \sum_{i=1}^{n} \rho(u_i)$$

  where u_i depends on the b_k's. This is equivalent to finding the b_k's that satisfy

$$\sum \psi(u_i) = \sum w_i u_i = 0$$

  where $w_i = \psi(u_i)/u_i$ and $\psi()$ is the derivative of $\rho()$.
- Resistant estimate of s: Median absolute deviation is

$$\text{MAD} = \text{median}|e_i - \text{median}(e_i)|$$

  which is a resistant measure of spread. A resistant estimate of the residual standard error is $s = \text{MAD}/0.6745$.
- Iterative procedure
  (a) Obtain starting values for the coefficients, e.g. from OLS.
  (b) Calculate scaled residuals. Use the residuals to calculate weights.
  (c) Apply WLS using the most recent set of weights to get new coefficient estimates.
  (d) Go to step b if things are still changing.

Breakout Session for 11/14/00

1. For OLS, $\rho(u_i) = u_i^2$. Calculate $\psi(u_i)$ and w_i. Intuitively, how will the regression coefficients change if we re-run the regression with these weights?

2. For Huber estimation,

$$\begin{array}{ll} \rho(u_i) = u_i^2 & \text{if} \;\; |u_i| \leq c \\ \rho(u_i) = c(2|u_i| - c) & \text{if} \;\; |u_i| > c \end{array}$$

(Assume c=1.345.) Calculate $\psi(u_i)$ and w_i. Intuitively, how will the regression coefficients change if we re-run the regression with these weights?

3. Sketch the weight function vs. the residual for the above two weight functions. Interpret.

4. Consider the data x=1, 2, 3, 4, 5, and y=7.384, 14.265, 11.490, 12.580, 15.442. Plot the data including your best guess for the regression line. Interpret the following where wn, rn and bn are the weights, the scaled residuals, and the coefficients for the n^{th} iteration of a robust regression based on the Huber weight function.

w1	r1	w2	r2	w3	r3	w4	r4
1	-1.84	0.73	-1.67	0.80	-1.16	1.00	-0.97
1	3.26	0.41	5.66	0.24	7.42	0.18	9.90
1	-0.70	1.00	-0.46	1.00	-0.16	1.00	0.03
1	-1.03	1.00	-1.19	1.00	-1.27	1.00	-1.54
1	0.30	1.00	0.54	1.00	0.59	1.00	0.70

	(Intercept)	x
b0	7.90	1.44
b1	7.90	1.44
b2	6.97	1.62
b3	6.32	1.75
b4	6.02	1.82