Chapter 11: Robust & Quantile regression

Adapted from Timothy Hanson

Department of Statistics, University of South Carolina

Stat 704: Data Analysis I

11.3: Robust regression

- Leverages h_{ii} and deleted residuals t_i are useful for finding outlying $\mathbf{x}_i$ and Y_i (w.r.t. the model) cases.
- Cook's D_i and DFFIT_i indicate which cases influence the fit of the model, i.e. the OLS $\mathbf{b}$.
- What to do with influential and/or outlying cases? Are they transcription errors or somehow unrepresentative of the target population?
- Outliers are often interesting in their own right and can help guide the building of a better model.
- Robust regression dampens the effect of outlying cases on estimation to provide a better fit to the majority of cases.
- Useful in situations when there's no time for "influence diagnostics" or a more careful analysis.

- Robust regression is effective when the error distribution is not normal, but heavy-tailed.
- M-estimation is a general class of estimation methods. Choose β to minimize

$$Q(\beta) = \sum_{i=1}^{n} \rho(Y_i - \mathbf{x}_i'\beta),$$

  where $\rho(\cdot)$ is some function.
- $\rho(u) = u^2$ gives OLS $\mathbf{b}$.
- $\rho(u) = |u|$ gives L_1 regression, or *least absolute residual* (LAR) regression.
- Huber's method – described next – builds a $\rho(\cdot)$ that is a compromise between OLS and LAR. It looks like u^2 for u close to zero and $|u|$ further away.

Iteratively Reweighted Least Squares

Outlying values of $r_i^j = y_i - \mathbf{x}_i'\mathbf{b}^j$ are (iteratively) given less weight in the estimation process.

0 (Starting $\mathbf{b}^0$) OLS: $w_i^0 = 1/e_i^2$ from $\mathbf{e} = \mathbf{Y} - \mathbf{Xb}$. Set $j = 1$.

1 (WLS for $\mathbf{b}^j$ using $\mathbf{W}^{j-1}$): $\mathbf{b}^j = (\mathbf{X}'\mathbf{W}^{j-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{j-1}\mathbf{Y}$.

2 $\hat{\sigma}^j = \text{median}\{|Y_i - \mathbf{x}_i'\mathbf{b}^j|/\Phi^{-1}(0.75) : i = 1, \ldots, n\}$.

3 $w_i^j = W\left(\frac{Y_i - \mathbf{x}_i'\mathbf{b}^j}{\hat{\sigma}^j}\right)$, where

$$W(u) = \left\{ \begin{array}{cc} 1 & \text{if } |u| \leq 1.345 \\ \frac{1.345}{|u|} & \text{if } |u| > 1 \end{array} \right\}.$$

Set j to $j + 1$.

1. Repeat steps 1 through 3 until $\hat{\sigma}^j$ and $\mathbf{b}^j$ stabilize.

There are other weight functions (SAS default is bisquare, pp. 439-440) and other methods for updating $\hat{\sigma}^j$ – see book and SAS documentation if interested.

SAS code: M-estimation w/ Huber weight

```
data prof;
 input state$ mathprof parents homelib reading tvwatch absences;
datalines;
 Alabama                252  75  78  34  18  18
 Arizona                259  75  73  41  12  26
                   ...et cetera...
 Wisconsin              274  81  86  38   8  21
 Wyoming                272  85  86  43   7  23
;

proc robustreg data=prof method=m (wf=huber);
 model mathprof=parents homelib reading tvwatch absences; run;

proc robustreg data=prof method=m (wf=huber);
 model mathprof=parents homelib reading tvwatch absences;
 test absences reading; run;

proc robustreg data=prof method=m (wf=huber);
 model mathprof=parents homelib tvwatch ;
 id state; output out=out p=p sr=sr; run;
```

SAS output...

The ROBUSTREG Procedure

Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	178.1630	28.2897	122.7162	233.6097	39.66	<.0001
parents	1	0.3574	0.2007	-0.0360	0.7507	3.17	0.0750
homelib	1	0.7702	0.1403	0.4952	1.0452	30.14	<.0001
reading	1	0.1495	0.2100	-0.2620	0.5611	0.51	0.4763
tvwatch	1	-0.8218	0.2752	-1.3612	-0.2824	8.92	0.0028
absences	1	-0.0121	0.2058	-0.4155	0.3912	0.00	0.9530

Robust Linear Tests

Test	Test Statistic	Lambda	DF	Chi-Square	Pr > ChiSq
Rho	0.2201	0.8646	2	0.25	0.8805
Rn2	0.5073		2	0.51	0.7760

Quantile Regression

Quantile regression

- Another option is general *quantile regression*.
- L_1 regression (p. 438), described above, is *median regression* and can be carried out in SAS PROC QUANTREG. Other quantiles can similarly be regressed upon.
- Robust quantile regression is particularly helpful for modeling responses with non-constant error variance.
- QUANTREG solves minimization problem using simplex algorithm, details in documentation.

Median, or L_1 regression minimizes

$$Q_{0.5}(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta} \in \mathbf{R}^p} \sum_{i=1}^{n} |Y_i - \mathbf{x}_i'\boldsymbol{\beta}|.$$

Let $0 < \tau < 1$ be a probability. Koenker and Bassett (1978) define $\mathbf{b}_\tau$, the τth quantile regression effects vector, to minimize

$$Q_\tau(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta} \in \mathbf{R}^p} \left[\tau \sum_{i: y_i \geq \mathbf{x}_i'\boldsymbol{\beta}} |Y_i - \mathbf{x}_i'\boldsymbol{\beta}| + (1-\tau) \sum_{i: y_i < \mathbf{x}_i'\boldsymbol{\beta}} |Y_i - \mathbf{x}_i'\boldsymbol{\beta}| \right].$$

Let Y_h be a draw from the population with accompanying $\mathbf{x}_h$. The $\mathbf{b}_\tau$ that minimizes $Q_\tau(\boldsymbol{\beta})$ satisfies

$$P(Y_h \leq \mathbf{x}_h'\mathbf{b}_\tau) \approx \tau.$$

Note $q_\tau(\mathbf{x}) = \mathbf{x}'\mathbf{b}_\tau$. How are the elements of $\mathbf{b}_\tau$ interpreted?

SAS code: Fall 2008 GPA data

```
proc quantreg data=fall08;
 model cltotgpa=satv / quantile=0.25;
 output out=o1 p=p1;

proc quantreg data=fall08;
 model cltotgpa=satv / quantile=0.50;
 output out=o2 p=p2;

proc quantreg data=fall08;
 model cltotgpa=satv / quantile=0.75;
 output out=o3 p=p3;

data o4; set o1 o2 o3; proc sort data=o4; by quantile satv;
```

SAS output

Why is the slope for regression on $q_{.75}$ not increase?

```
               The QUANTREG Procedure

Quantile                               0.25

                                95% Confidence
Parameter DF Estimate              Limits
Intercept  1   1.3657    0.9885 1.6033
satv       1   0.0025    0.0020 0.0032

Quantile                                0.5

                                95% Confidence
Parameter DF Estimate              Limits
Intercept  1   1.5485    1.2774 1.8131
satv       1   0.0029    0.0024 0.0034

Quantile                               0.75

                                95% Confidence
Parameter DF Estimate              Limits
Intercept  1   2.0624    1.8469 2.3454
satv       1   0.0026    0.0022 0.0030
```

Normal-errors regression is also quantile regression

For our standard model:

$$Y = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim N(0, \sigma^2),$$

note that

$$P(Y \leq q_\tau) = \tau \Leftrightarrow$$

$$P(\mathbf{x}'\boldsymbol{\beta} + \epsilon \leq q_\tau) = \tau \Leftrightarrow$$

$$P\left(\frac{\epsilon}{\sigma} \leq \frac{q_\tau - \mathbf{x}'\boldsymbol{\beta}}{\sigma}\right) = \tau \Leftrightarrow$$

$$\Phi\left(\frac{q_\tau - \mathbf{x}'\boldsymbol{\beta}}{\sigma}\right) = \tau \Leftrightarrow$$

$$\frac{q_\tau - \mathbf{x}'\boldsymbol{\beta}}{\sigma} = \Phi^{-1}(\tau) \Leftrightarrow$$

$$q_\tau(\mathbf{x}) = \Phi^{-1}(\tau)\sigma + \mathbf{x}'\boldsymbol{\beta}.$$

- When x_j increases by one, *every quantile* increases by β_j.
- What does this imply about the quantile functions?
- Could we see a plot like we saw for the GPA data a few slides ago?
- Final comment: L_1 regression is obtained as the MLE from a standard regression model assuming the errors are distributed double-exponential (Laplace).