CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

Lecture 56
Robust Regression

Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

Robust Regression

- OLS: minimizing χ^2 (least squared errors) is not robust
  - Outliers or points from the tails of distributions are heavily weighted
  - b.p. = $1/n$: even one bad data point can make regression results meaningless
  - We have discussed remedial measures for influential outliers and non-normal distributions, but they are not always effective for large amounts of contamination and not easy to automate
- An alternative: Robust Regression

Least Absolute Deviations

- Least Absolute Deviations (LAD):
  - Minimize $S = \sum |y_i - \hat{y}_i|$
  - Weights outliers linearly, not quadratically
  - Absolute value sign is problematic since function is discontinuous at zero (linear programming required), and may not have a unique solution
  - LAD is the MLE if the residuals are independent and have the double-exponential distribution
  - For normal errors, $SE(b_k)$ is 26% bigger than OLS
  - Also called minimum L_1-norm regression

M-Estimation

- Define a general function of the residuals, $H(\varepsilon_i)$, and then minimize $S = \sum H(\varepsilon_i)$
  - For OLS, $H(\varepsilon_i) = \varepsilon_i^2$
- The properties we want for the function H
  - Always non-negative, $H(\varepsilon_i) \geq 0$
  - $H(0) = 0$
  - Symmetric, $H(-\varepsilon_i) = H(\varepsilon_i)$
  - Monotonic: if $|\varepsilon_i| > |\varepsilon_j|$ then $H(\varepsilon_i) > H(\varepsilon_j)$
  - Continuous derivative with respect to the coefficients (for numerical stability)
- Implement using iteratively reweighted LS

M-Estimation

- For least-squares regression: $S = \sum \varepsilon_i^2$, take the derivative with respect to a parameter and set = 0

$$\frac{\partial S}{\partial \beta_k} = 0 \rightarrow \sum_{i=1}^{n} \varepsilon_i x_{ki} = 0$$

- For M-estimator,

$$\frac{\partial S}{\partial \beta_k} = 0 \rightarrow \sum_{i=1}^{n} \frac{\partial H}{\partial \varepsilon_i} x_{ki} = 0$$

M-Estimation

- Define a weight as $w_i = \frac{1}{\varepsilon_i} \frac{\partial H}{\partial \varepsilon_i}$
- Giving, $\sum_{i=1}^{n} \frac{\partial H}{\partial \varepsilon_i} x_{ki} = 0 = \sum_{i=1}^{n} w_i \varepsilon_i x_{ki}$
- But this is just weighted linear regression!
- Guess the weights, fit, then calculate the residuals. Use those residuals to calculate new weights. Repeat until convergence.
  - Called iteratively reweighted Least Squares

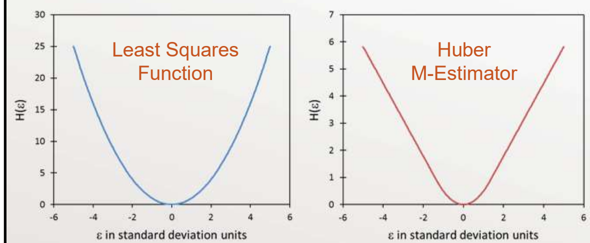
Huber M-Estimator

- The Huber M-estimator attempts to get the best of both the least-square estimator (easy to find the minimum) and the absolute deviation estimator (more robust)

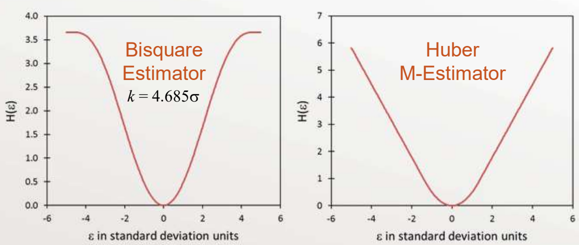
$$H(\varepsilon) = \begin{cases} \varepsilon^2/2 & for\ |\varepsilon| \le k \\ k|\varepsilon| - k^2/2 & for\ |\varepsilon| > k \end{cases}$$

- Huber picked $k = 1.345\sigma$, which gives 95% efficiency (almost the same as OLS)
- The residuals are studentized using MAD

Huber M-Estimator

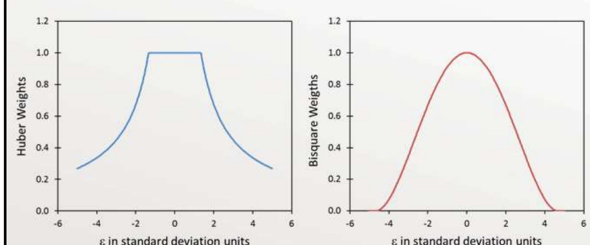


Bisquare M-Estimator



$$Bisquare\ H(\varepsilon) = \begin{cases} k^2/6(1-[1-(\varepsilon/k)^2]^3) & for\ |\varepsilon| \le k \\ k^2/6 & for\ |\varepsilon| > k \end{cases}$$

M-estimator Weights



M-Estimator Robustness

- The bisquare estimator is popular, but can suffer from local minima
  - The Huber M-estimator gives a unique solution and is often used to provide a starting point for the bisquare estimator
- Both M-estimators and LAD can tolerate large deviations in Y, so long as they are not overly influential (that is, they don't have large deviations in X)
- Robust estimators make good outlier detectors

Bounded Influence Regression

- Least Trimmed Squares deletes some percentage of the most extreme residuals (up to half), then performs OLS on the rest
  - Does not work well for small sample sizes
  - Of course, the extreme cases may be very important!
- Least Median Squares minimizes the median of the squared residuals
  - Very robust, but very low efficiency

Lecture 56: What have we learned?

- What is the breakdown point for OLS?
- Explain the basic operation of M-estimators for linear regression
- What are some of the difficulties and complications for robust regression?
- How do you choose among the robust regression options?

 Data to Decisions 13