IAOS 2014 Conference – Meeting the Demands of a Changing World
Da Nang, Vietnam, 8-10 October 2014

ROBUST REGRESSION IMPUTATION:
CONSIDERATION ON THE INFLUENCE OF WEIGHT FUNCTIONS AND THE SCALE OF CONVERGENCE

Tatsuo Noro and Kazumi Wada
National Statistics Center of Japan

Notes: The views and opinions expressed in this presentation are the authors' own, not necessarily those of the institution.

OUTLINE

1. Objective
2. Methodology
3. Monte Carlo experiments
4. Summary of Results
5. Conclusions

References

OBJECTIVE

This presentation will be on the regression imputation focusing on the existence of outliers.

Resistant regression described by Bienias et al.(1997) is classical M-estimation by the Iteratively Reweighted Least Squares (IRLS) algorithm.

Bienias et al. (1997) recommended

Weight function : Tukey's biweight

Scale parameter : Average Absolute Deviation (AAD)

However, Tukey's biweight function does not promise to give the global solution.

OBJECTIVE

The aims of presentation is to shed light upon the influence of these settings of IRLS to the outcome,

so that we are able to make a suitable choice according to the purpose of the estimation and/or the data set treated.

METHODOLOGY

Linear Regression Model

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \Lambda + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_\mathbf{i}'\boldsymbol{\beta} + \varepsilon_i \,,$$

$$i = 1, \Lambda \;, n$$

where

y_i : response variable,

$\mathbf{x}_i$: explanatory variables,

ε_i : error term

METHODOLOGY

Fitted Model

$$\hat{y}_i = a + b_1 x_{i1} + b_2 x_{i2} + \Lambda + b_p x_{ip} = \mathbf{b}'\mathbf{x_i}$$

$$\mathbf{b} = \begin{pmatrix} a \\ b_1 \\ \mathrm{M} \\ b_p \end{pmatrix} : \text{estimated regression parameter}$$

METHODOLOGY

The residuals are given by

$$e_i = y_i - \hat{y}_i = y_i - \mathbf{b}'\mathbf{x_i}$$

M-estimator for **b** minimizes

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{b}'\mathbf{x_i}}{\sigma} \right),$$

where σ : scale parameter,

ρ : loss function

METHODOLOGY

Let $\psi = \rho'$ (influence function)

Then system of equations

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{b'x_i}}{\sigma} \right) \mathbf{x_i}' = \mathbf{0}$$

This equations is not able to solve. We define

$$w(e) = \frac{\psi(e)}{e} \text{ : weight function}$$

and let

$$w_i = w(e_i)$$

METHODOLOGY

Then estimator **b** is obtained by a solution of

$$\sum_{i=1}^{n} w_i \left(\frac{y_i - \mathbf{b}'\mathbf{x_i}}{\sigma} \right) \mathbf{x_i}' = \mathbf{0} \quad .$$

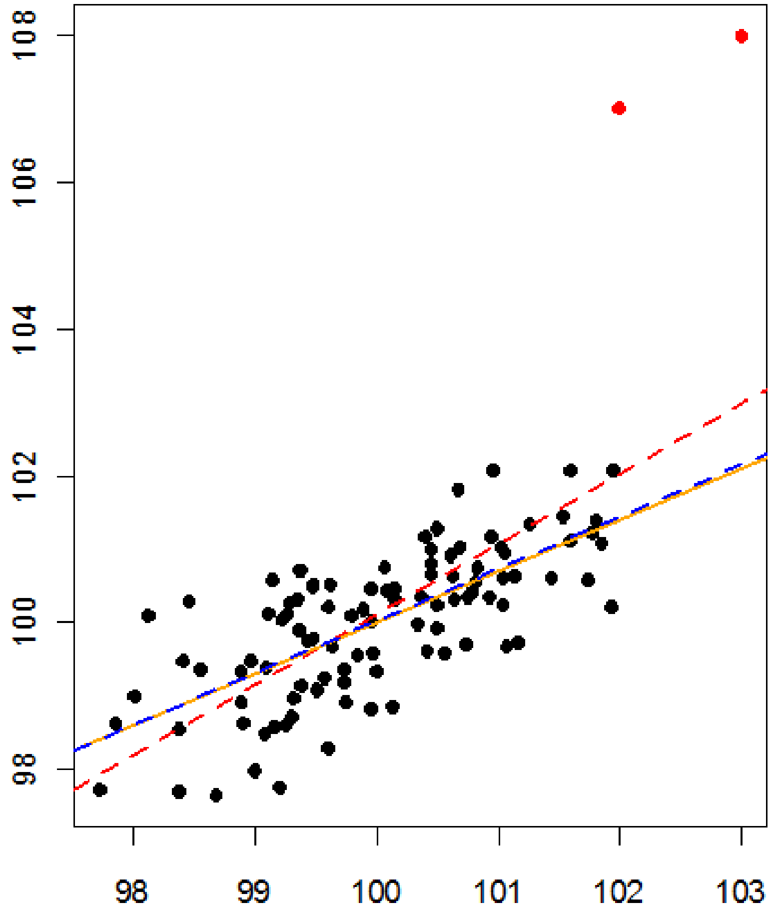
IRLS Algorithm

Iteratively Reweighted Least Squares (IRLS) algorithm by Holland & Welsch(1977).

To control an influence of outliers by downweight (w_i<1)

It is easy to calculate, but not so robust for an explanatory variables.

IRLS Algorithm



Data set used

Number of data n=102

100 data

〜bivariate normal (cor=0.7, var=1)

2 data

artificial outliers

Results of regression Analysis

OLS: **α** 4.74 **β** 0.95

OLS without outliers:

α 30.0 **β** 0.7

IRLS: **α** 29.10 **β** 0.71

IRLS Algorithm

1st step (Initial Estimation)

We estimate $\mathbf{b}^{(0)}$ by ordinary least squares(OLS).

$$\mathbf{b}^{(0)} = \left[\mathbf{X'}\mathbf{X}\right]^{-1}\mathbf{X'}\mathbf{y} : \text{initial estimate}$$

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} : \text{data matrix}$$

IRLS Algorithm

2nd step

Calculate residuals $e_i^{(j-1)}$

and average absolute deviation (AAD) $s^{(j-1)}$

$$s^{(j-1)} = \frac{1}{n}\sum_{i=1}^{n}\left|e_i^{(j-1)} - \overline{e}^{(j-1)}\right|, \text{where } \overline{e}^{(j-1)} = \frac{1}{n}\sum_{i=1}^{n} e_i^{(j-1)}$$

according to weight function

$$w_i^{(j-1)} = w\left(e_i^{(j-1)}\right)$$

IRLS Algorithm

3rd step

New weighted least squares estimates

$$\mathbf{b}^{(j)} = \left[\mathbf{X}'\mathbf{W}^{(j-1)}\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{W}^{(j-1)}\mathbf{y}$$

where diagonal matrix $\mathbf{W}^{(j-1)} = \mathrm{diag}\left\{w_i^{(j-1)}\right\}$

Steps 2nd and 3rd are repeated

IRLS Algorithm

Convergence criteria

We stopped iterating when the proportionate change in *s* was less than 0.01.

$$\frac{|s^{(j)} - s^{(j-1)}|}{s^{(j-1)}} < 0.01$$

Weight functions

We compare two weight functions

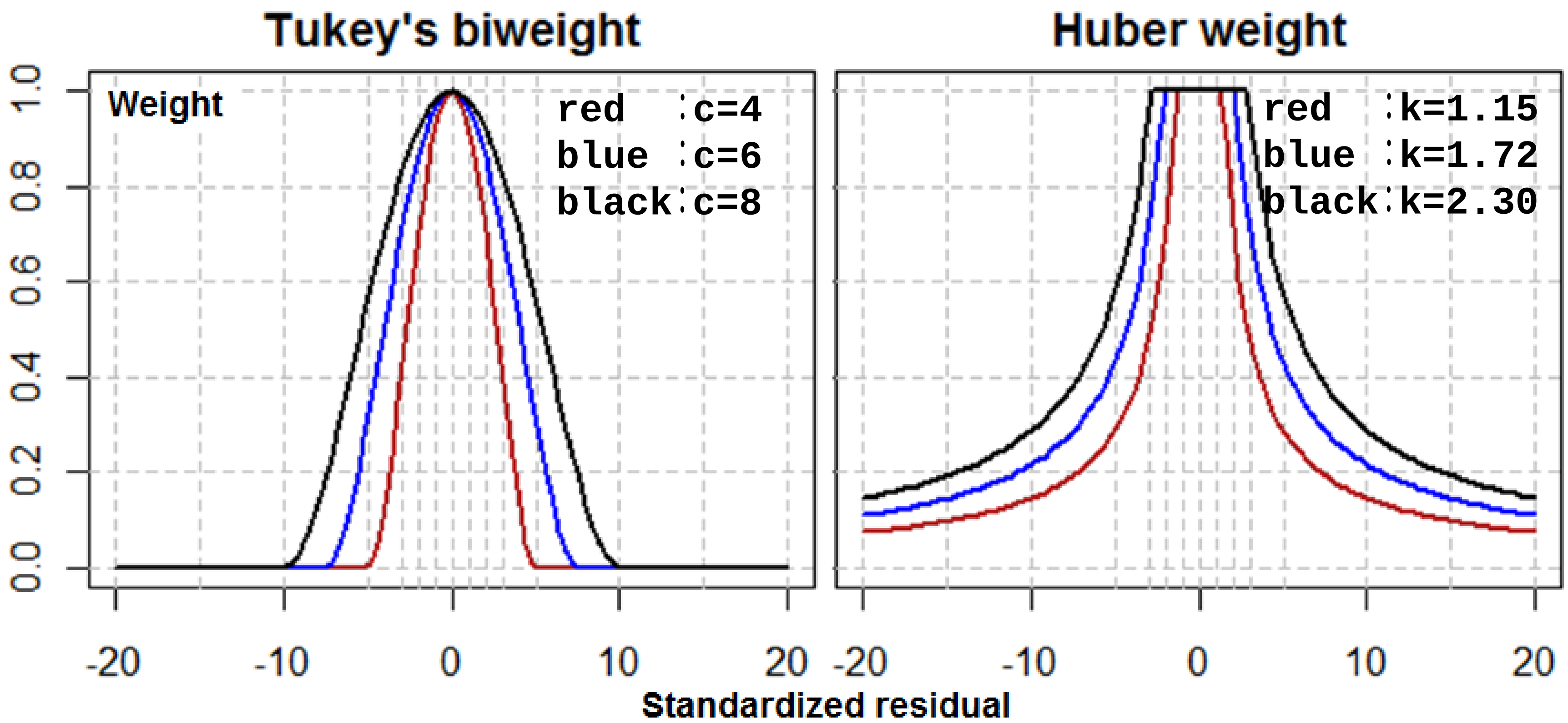
Tukey's biweight(bisquare) function

$$w_i = \begin{cases} \left(1 - \left(\dfrac{e_i}{cs}\right)^2\right)^2 & |e_i| \le cs \\ 0 & \text{else} \end{cases}$$

and Huber weight function

$$w_i = \begin{cases} 1 & |e_i| \le ks \\ \dfrac{ks}{|e_i|} & \text{else} \end{cases}$$

Weight Functions



$$\begin{cases} \text{if } |r_i| \le (c \times s) & w_i = (1-(\frac{r_i}{c \times s})^2)^2 \\ \text{if } |r_i| > (c \times s) & w_i = 0 \end{cases}$$

$$\begin{cases} \text{if } |r_i| \le (k \times s) & w_i = 1 \\ \text{if } |r_i| > (k \times s) & w_i = \frac{k \times s}{|r_i|} \end{cases}$$

Weight Functions

Tuning Constant (Tukey's *c* and Huber's *k*)

This is to control the robustness of the estimator depending on the user's preference.

Bienias et al.(1997) recommended Tukey's *c* from 4 to 8 regarding the AAD scale.

Holland and Welsch(1977) calculated the tuning constants for 95% asymptotic efficiency at the Gaussian distribution.

Tuning Constant

Tukey's *c* for AAD	4	6	8
Tukey's *c* for SD	5.01	7.52	10.03
Tukey's *c* for MAD	3.38	5.07	6.76
Huber's *k* for AAD	1.15	1.72	2.30
Huber's *k* for SD	1.44	2.16	2.88
Huber's *k* for MAD	0.97	1.46	1.94

MAD : Median Absolute Deviation

$$\mathrm{MAD}(e_i) = \mathrm{median}[|\, e_i - \mathrm{median}(e_j)\,|]$$

Monte Carlo experiment

$$x_i\ (i = 1, \Lambda\ ,100\ ) \sim\ U(0,\ 10),\ \text{i.i.d.}$$

The dependent variable y_i is made in accordance with the linear regression model

$$y_i = 2 + 5x_i + \varepsilon_i$$

Error term follows independently *t*-distribution with degree of freedom 1,2,3,5,10, and standard normal distribution.

Replication is 100,000 data sets for each error term.

Monte Carlo experiment

Comparison

Tukey’s biweight via Huber weight

AAD scale via MAD scale

Tuning constants according with the weight functions

Monte Carlo experiment

Conditions to be compared

A. Weight function： (1) Tukey's biweight (2) Huber weight

B. Scale parameter： (1) Average Absolute Deviation (AAD)
(2) Median Absolute Deviation (MAD)

C. Tuning constant：

Tukey[B-(1)]	(i) TK4: 4	(ii) TK6: 6	(iii) TK8: 8
Tukey[B-(2)]	(i) TK4: 5.01	(ii) TK6: 7.52	(iii) TK8: 10.03
Huber[B-(1)]	(i) HB4: 1.15	(ii) HB6: 1.72	(iii) HB8: 2.30
Huber[B-(2)]	(i) HB4: 1.44	(ii) HB6: 2.16	(iii) HB8: 2.88

D. Convergence criteria of the proportional change of scale
(a) 0.01 (b) 0.001 (c) 0.0001

Note:(1) The values for B-(2) is not for the MAD scale, but of SD since the mad function in R returns the values adjusted in accordance with SD.
(2) The limit of iteration is 150.

Summary of Results

Troubles with Tukey's biweight

(1) Infinite loop

Repeating the estimates of regression parameters in iteration

(2) Estimation impossible

Estimation fails there are two extreme outliers on the same side of the regression line.

Summary of Results

Maximum of iteration count

scale	AAD						MAD					
wt & tc	TK4	TK6	TK8	HB4	HB6	HB8	TK4	TK6	TK8	HB4	HB6	HB8
	convergence criteria 0.01						convergence criteria 0.01					
df 1	6	6	6	6	6	6	150	150	150	21	53	76
df 2	6	5	5	5	5	5	36	22	150	18	19	13
df 3	6	5	5	6	5	4	23	17	150	11	11	11
df 5	7	5	5	5	5	4	25	16	13	14	12	14
df 10	6	5	4	5	5	4	15	10	8	11	9	8
df Inf	6	5	4	6	5	4	12	9	5	10	8	6
	convergence criteria 0.0001						convergence criteria 0.0001					
df 1	11	9	10	9	9	9	150	150	150	30	63	150
df 2	13	10	9	10	8	7	150	150	150	41	54	26
df 3	13	9	8	11	8	7	46	32	150	23	20	30
df 5	13	9	7	11	8	7	150	37	26	22	25	25
df 10	15	9	7	11	8	7	33	21	14	21	17	16
df Inf	14	8	7	11	8	7	33	19	8	20	15	11

Summary of Results

Standard deviation of the estimated mean

	AAD（convergence criteria 0.01）						MAD（convergence criteria 0.01）					
	df 1	df 2	df 3	df 5	df 10	df Inf.	df 1	df 2	df 3	df 5	df 10	df Inf.
OLS	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862
TK4	0.6521	0.5944	0.5914	0.5895	0.5889	0.5879	0.7321	0.5946	0.5915	0.5893	0.5882	0.5867
TK6	0.6803	0.5963	0.5918	0.5893	0.5882	0.5867	0.6302	0.5967	0.5925	0.5897	0.5883	0.5863
TK8	0.7117	0.5986	0.5928	0.5896	0.5882	0.5864	0.6273	0.5990	0.5937	0.5902	0.5884	0.5863
HB4	2.1044	0.5954	0.5914	0.5892	0.5884	0.5872	0.6113	0.5955	0.5918	0.5893	0.5882	0.5866
HB6	3.0941	0.5981	0.5923	0.5893	0.5882	0.5866	0.6221	0.5984	0.5932	0.5899	0.5883	0.5863
HB8	4.1281	0.6010	0.5934	0.5898	0.5882	0.5864	0.6334	0.6012	0.5945	0.5905	0.5886	0.5862
	AAD（convergence criteria 0.0001）						MAD（convergence criteria 0.0001）					
	df 1	df 2	df 3	df 5	df 10	df Inf.	df 1	df 2	df 3	df 5	df 10	df Inf.
OLS	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862	167.8765	0.8680	0.6029	0.5919	0.5889	0.5862
TK4	0.6522	0.5944	0.5915	0.5898	0.5892	0.5884	0.7292	0.5945	0.5916	0.5893	0.5882	0.5867
TK6	0.6803	0.5963	0.5918	0.5893	0.5882	0.5867	0.6295	0.5967	0.5925	0.5897	0.5883	0.5863
TK8	0.7116	0.5986	0.5927	0.5896	0.5882	0.5864	0.6263	0.5990	0.5937	0.5902	0.5884	0.5863
HB4	2.1038	0.5953	0.5913	0.5891	0.5885	0.5874	0.6105	0.5954	0.5917	0.5892	0.5882	0.5866
HB6	3.0923	0.5981	0.5922	0.5893	0.5882	0.5867	0.6216	0.5984	0.5932	0.5899	0.5883	0.5863
HB8	4.1270	0.6010	0.5934	0.5898	0.5882	0.5864	0.6331	0.6011	0.5945	0.5905	0.5886	0.5863

CONCLUSIONS

(1) Huber weight is slightly faster to converge, but Tukey's biweight is capable to remove the influence of extreme outliers completely.

(2) The choice of scale parameter affects computational time. The AAD scale makes convergence faster than the MAD scale for both of the weight functions.

(3) A convergence criteria is not influenced for the accuracy.

CONCLUSIONS

We recommend

A. Huber weight function with AAD scale

If it also promises convergence of the iteration.

B. Tukey's biweight function with AAD scale

If he wish to eliminate their influence from the inference.

References

- [1] Beaton, A. E. and Tukey, J. W. (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, *Technometrics* 16, 147-185
- [2] Bienias, J. L., Lassman, D. M. Scheleur, S. A. & Hogan H. (1997) Improving Outlier Detection in Two Establishment Surveys. *Statistical Data Editing 2 - Methods and Techniques.* (UNSC and UNECE eds.), 76-83.
- [3] Fox, J. & Weisberg S. (2010) Robust Regression, Appendix to An R Companion to Applied Regression. Sage, Thousand Oaks, CA, 2nd ed. 2011
- [4] Holland, P. W. & Welsch, R. E. (1977), Robust Regression Using Iteratively Reweighted Least-Squares, *Communications in Statistics – Theory and Methods* 6(9), 813-827

References

- [5] Huber, P. J. (1964) Robust estimation of a location parameter, *Annals of Mathematical Statistics* 35, 73-101
- [6] Huber, P. J. (1973) Robust Regression: Asymptotics, Conjectures and Monte Carlo, *Annals of Statistics*.1, 799-821
- [7] Huber, P. J. & Ronchetti, Elvezio M. (2009) *Robust Statistics*, 2nd ed., John Wiley & Sons, Inc., New York
- [8] Rousseeuw, P. J. & Leroy, A. M. (1987) *Robust Regression and Outlier Detection*, John Wiley & Sons, Inc.
- [9] Tukey, J.W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

Thank you for your attention !