A stochastic expansion of the Huber-skip estimator for multiple regression

Joint meeting between DSTS and DSBS
Novo Alle, 2880 Bagsværd.

Søren Johansen [1]

Joint work with Bent Nielsen[2]

November 10-11 2015

[1]Department of Economics, University of Copenhagen and CREATES, University of Aarhus

[2]Nuffield College & Department of Economics, University of Oxford

List of contents

Some examples of outliers

Outliers and robust statistics, example 1

What is an outlier, how do you model them and how do you avoid them?

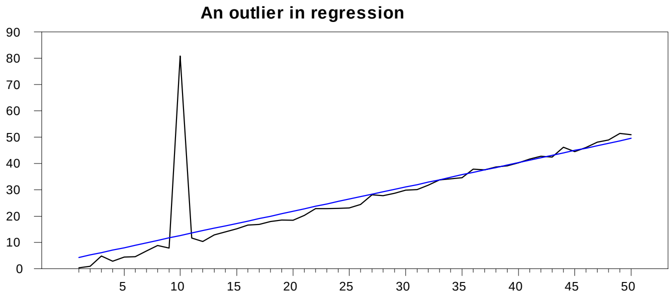


Figure: A regression $y_i = \beta x_i + \varepsilon_i, \varepsilon_i$ i.i.d. $(0, \sigma^2)$. Is the outlier just an influential observation?

Outliers and robust statistics, example 2

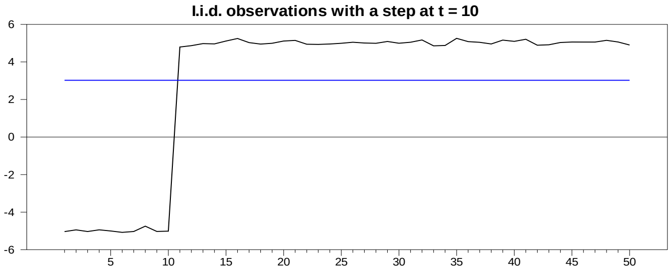


Figure: A step dummy at $t = 10$. The variance $\sigma_0^2 = 0.01$ is estimated by $\hat{\sigma}^2 = 16.37$. Some $t-$ratios are $t = (X(1) - \bar{X})/\hat{\sigma} = -1.97$ and $t = (X(50) - \bar{X})/\hat{\sigma} = 0.49$.

Outliers and robust statistics, example 3

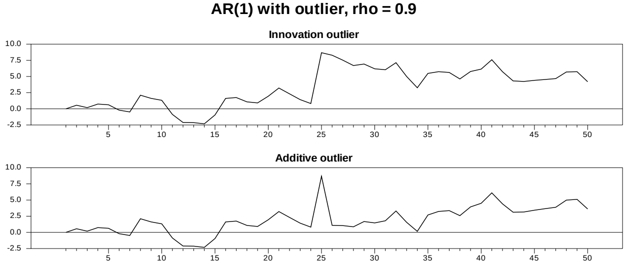


Figure: An innovative (added to the equation) and an additive (added to the process) outlier in the model $y_i = \beta y_{i-1} + \varepsilon_i$ i.i.d. $(0, \sigma^2)$

Modelling of outliers and some robust statistics

Modelling of outliers

One can model them as either innovative or additive outliers and interpret them as either influential or completely wrong observations.

One can model them deterministically or randomly:

$$f(\varepsilon) = (1-\alpha) f_0(\varepsilon) + \alpha f_1(\varepsilon)$$

Two well known robust estimators are (Huber, 1964)

$$\text{Huber-skip}: \min_\beta \sum_{i=1}^n \min\{(y_i - \beta' x_i)^2, c^2\}$$

and the Least Trimmed Squares (Rousseeuw, 1984, Víšek 2006)

$$LTS: \min_\beta \sum_{i=1}^h (y-\beta' x)^2_{(i)}.$$

Such robust estimators are analysed for

1. Asymptotic distribution when there are no outliers to evaluate the loss of efficiency
2. Influence function and breakdown point when there are outliers

The model and the problem

The multiple regression model:

$$y_i = \beta' x_i + \varepsilon_i = \mu + \alpha' z_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where ε_i is an "innovation" independent of $(x_1, \ldots, x_i, \varepsilon_1, \ldots, \varepsilon_{i-1})$ with finite variance, distribution function F, density f, and derivative $\dot{f}$
$E(\varepsilon_i)$ need not be zero
The regressors: deterministic, or stochastic; stationary or random walk
M-estimators: The objective function

$$R_n(\beta) = n^{-1} \sum_{i=1}^{n} \rho(y_i - \beta' x_i),$$

where $\rho(u) \geq 0$ is continuous, increasing for $u \geq 0$ and decreasing for $u \leq 0$ with right and left derivatives. The minimizer is an $M-$**estimator**.
Leading case is Huber-skip

$$\rho(u) = \frac{1}{2} \min(u^2, c^2)$$

The model and the problem

The problem: To find conditions for the M-estimator to exist, be consistent and have a first order asymptotic expansion, which allows us to find asymptotic distributions

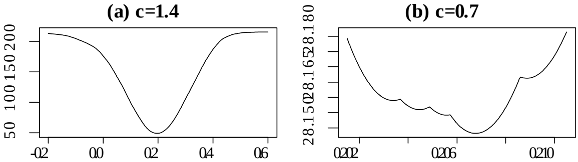
The technique: We apply martingale techniques to study weighted and marked empirical processes, containing estimation uncertainty

The Huber-skip estimator

The Huber-skip Huber (1964) is a robust M-estimator, (Jurečková, Sen, and Picek, 2012 for the location model)

$$\text{Objective function}: n^{-1}\sum_{i=1}^{n}\frac{1}{2}\min\{(y_i-\beta' x_i)^2, c^2\}$$

$$\text{Score function}: n^{-1}\sum_{i=1}^{n}(y_i-\beta' x_i)x_i' 1_{(|y_i-\beta' x_i|\leq c)}$$



Why Huber-skip ?

1. Difficult to compute 2. Requires known scale 3. More robust estimators exist 4. The mathematics too difficult

A few definitions

$$
\begin{aligned}
R_n(\beta) &= n^{-1}\sum_{i=1}^{n}\rho(y_i-\beta' x_i), \\
\hat{\Sigma}_n &= N'\sum_{i=1}^{n}x_i x_i' N \text{ and } (\hat{\Sigma}_n, \hat{\Sigma}_n^{-1}) = O_P(1)
\end{aligned}
$$

$$
\begin{aligned}
h(\mu) &= E(\rho(\varepsilon-\mu)) \geq h(\mu_\rho) \\
\dot{h}(\mu_\rho) &= 0 \\
\ddot{h}(\mu_\rho) &= -\int \dot{\rho}(u-\mu_\rho)\dot{f}(u)du > 0,
\end{aligned}
$$

$$
\beta_0 = (\mu_0, \alpha_0')' \text{ and } \beta_\rho = (\mu_0+\mu_\rho, \alpha_0')'
$$

M-estimators

Least squares: $\rho(u) = \frac{1}{2}u^2$
Quantile regression: $\rho(u) = -(1-p)u1_{(u<0)} + pu1_{(u\geq 0)}$
Maximum likelihood: $\rho(u) = -\log f(u)$
Huber-skip: $\rho(u) = \frac{1}{2}\min(u^2, c^2)$

Some literature:

- Huber, P.J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73–101.
- Maronna, R.A., Martin, D.R., and Yohai, V.J. (2006) *Robust Statistics: Theory and Methods.* New York: Wiley.
- Huber, P.J. and Ronchetti, E.M. (2009) *Robust Statistics*. New York: Wiley.
- Jurečková, J., Sen, P.K. and Picek, J. (2012) *Methodological Tools in Robust and Nonparametric Statistics.* London: Chapman & Hall/CRC Press.

Results

Main results

Theorem 1 *Under Assumptions, a minimizer* $\hat{\beta}$ *of* $R_n(\beta)$ *exists with large probability and* $N^{-1}(\hat{\beta}-\beta_\rho) = O_P(n^{1/2-\eta})$ for $0 < \eta < 1/4$

Theorem 2 *Under more Assumptions,* $N^{-1}(\hat{\beta}-\beta_\rho) = O_P(1)$ *and has a first order expansion*

$$N^{-1}(\hat{\beta}-\beta_\rho) = \ddot{h}(\mu_\rho)^{-1}\hat{\Sigma}_n^{-1}N'\sum_{i=1}^{n} x_i\dot{\rho}(\varepsilon_i-\mu_\rho) + o_P(1)$$

$$\begin{aligned} h(\mu) &= E(\rho(\varepsilon-\mu)),\ \dot{h}(\mu_\rho) = E(\dot{\rho}(\varepsilon_i-\mu_\rho)) = 0, \\ \ddot{h}(\mu_\rho) &= -\int \dot{\rho}(u-\mu_\rho)\dot{f}(u)du > 0,\ \beta_\rho = (\mu_0+\mu_\rho, \alpha_0')' \end{aligned}$$

The main results for the Huber-skip

For the Huber-skip for symmetric density

$$\ddot{h}(0) = F(c) - F(-c) - 2cf(c)$$

$$N^{-1}(\hat{\beta}_H - \beta_0) = \ddot{h}(0)^{-1}\hat{\Sigma}_n^{-1} N \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \leq c)} + o_P(1)$$

Stationary regressors: $N = n^{-1/2}$

$$n^{1/2}(\hat{\beta}_H - \beta_0) \xrightarrow{D} \ddot{h}(0)^{-1} N_{\dim x}(0, \int_{-c}^{c} u^2 f(u) du \Sigma^{-1})$$

Random walk regressors: $N = n^{-1}$

$$n^{-1/2} x_{[nu]} \xrightarrow{D} W_x(u), \;\; n^{-1/2} \sum_{i=1}^{[nu]} \varepsilon_i 1_{(|\varepsilon_i| \leq c)} \xrightarrow{D} W_\varepsilon^c(u)$$

$$n(\hat{\beta}_H - \beta_0) \xrightarrow{D} \ddot{h}(0)^{-1}(\int_0^1 W_x W_x')^{-1} \int_0^1 W_x (dW_\varepsilon^c)'$$

Calculation of the Huber-skip by iteration

A 1-step estimator

The score equation is difficult to solve numerically:

$$\sum_{i=1}^{n}(y_i - \beta' x_i)x_i' 1_{(|y_i - \beta' x_i| \leq c)} = 0$$

A 1-step estimator: Take some initial estimator $\breve{\beta}$ and regress on observations with $|y_i - \breve{\beta}' x_i| \leq c$:

$$\begin{aligned} 0 &= \sum_{i=1}^{n}(y_i - \beta' x_i)x_i' 1_{(|y_i - \breve{\beta}' x_i| \leq c)} \\ N^{-1}(\hat{\beta} - \beta) &= (N' \sum_{i=1}^{n} x_i x_i' 1_{(|y_i - \breve{\beta}' x_i| \leq c)} N)^{-1} N' \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|y_i - \breve{\beta}' x_i| \leq c)} \end{aligned}$$

A "Taylor's" expansion in terms of the estimation error $N^{-1}(\breve{\beta} - \beta)$ gives

$$\frac{1}{F(c) - F(-c)} \hat{\Sigma}_n^{-1} N' \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \leq c)} + \frac{2cf(c)}{F(c) - F(-c)} N^{-1}(\breve{\beta} - \beta) + o_P(1)$$

Iteration of the 1-step estimator

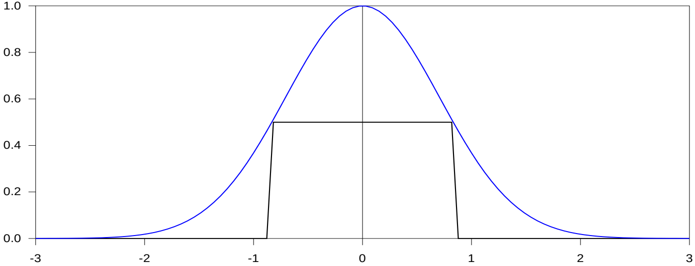
For the $(k+1)'$st step we find

$$\begin{aligned} N^{-1}(\hat{\beta}^{(k+1)} - \beta) &= \frac{1}{F(c)-F(-c)}\hat{\Sigma}_n^{-1} N' \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \leq c)} \\ &+ \frac{2cf(c)}{F(c)-F(-c)} N^{-1}(\hat{\beta}^{(k)} - \beta) + o_P(1) \end{aligned}$$

Iterating to ∞ when $2cf(c)/(F(c)-F(-c)) < 1$ gives the expansion of Huber-skip

$$N^{-1}(\hat{\beta}_H - \beta) = \frac{1}{F(c)-F(-c)-2cf(c)}\hat{\Sigma}_n^{-1} N' \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \leq c)} + o_P(1)$$

Condition for fixed point



A condition for the central part of the distribution to be non-trivial, and a fixed point in the iterated 1-step estimator

$$F(c) - F(-c) - 2cf(c) > 0 \text{ or } \frac{2cf(c)}{F(c) - F(-c)} < 1$$

A condition on small regressors and the proof of tightness

The condition for few small regressors

For δ a vector of unit length, $|\delta| = 1$, we define

$$F_n(a) = \sup_{|\delta|=1} n^{-1} \sum_{i=1}^{n} 1_{(|\delta' x_i| \leq a)}.$$

For small a this gives the fraction of regressors with small projecting on δ.
If x_i are linearly dependent, $F_n(a) = 1$. Thus $F_n(a)$ measures linear dependence.
An assumption for existence and consistency of the $M-$estimator is
Assumption *For some* $0 \leq \xi < 1$

$$\lim_{(a,n)\to(0,\infty)} P(F_n(a) \geq 1 - \xi) = 0$$

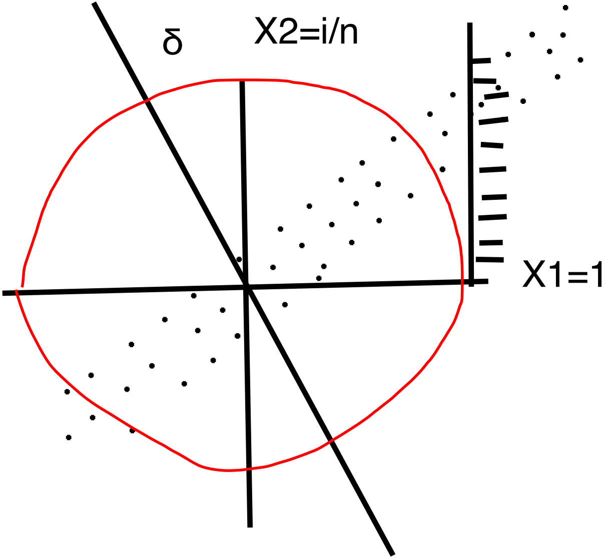
Examples of the condition for small regressors

1. If $x_i = 1_{(i \geq [n\xi_0])}$ then the frequency of zero values is $n^{-1}[n\xi_0], \xi_0 < 1$

$$n^{-1}\sum_{i=1}^{n} 1_{(|\delta' x_i| \leq a)} = n^{-1}\sum_{i=1}^{n} 1_{(|x_i| \leq a)} = n^{-1}[n\xi_0] \to \begin{cases} \xi_0 & 0 \leq a < 1 \\ 1 & 1 \leq a \end{cases}$$

2. $x_i = (1, in^{-1})'$ then $F_n(a) \leq 8a \to 0$, for $(a, n) \to (0, \infty)$
3. $x_i = (1, i^{-1})'$ then $F_n(a) \to 1$, for $(a, n) \to (0, \infty)$

Theorem $F_n(a) \overset{P}{\to} 0$, for $(a, n) \to (0, \infty)$ in case
4. If x_i is a stationary $AR(k)$, with density of $\delta' x_i | x_1, \ldots, x_{i-1}$ uniformly bounded (an example is Gaussian errors)
5. If $x_i n^{-1/2}$ is a random walk with density of $\delta' x_i / (n\delta'\Phi\delta)^{1/2}$ uniformly bounded (an example is Gaussian errors)


δ
X2=i/n
X1=1

Theorem 1 *Under Assumptions (x_i stationary), a minimizer $\hat{\beta}$ of $R_n(\beta)$ exists with large probability and $\hat{\beta} - \beta_\rho = O_P(1)$*

"**Proof**" A lower bound for $R_n(\beta)$ is, using $\beta = \beta_\rho + \lambda\delta, \quad \lambda = |\beta - \beta_\rho|$

$$R_n(\beta) = n^{-1}\sum_{i=1}^{n}\rho(y_i - \beta' x_i), \text{ here } \rho(u) = \frac{1}{2}\min(u^2, c^2)$$

$$\begin{aligned}
&\geq n^{-1}\sum_{i=1}^{n}\rho(\varepsilon_i - \mu_\rho - \lambda\delta' x_i)1_{(|\varepsilon_i|\leq A)}1_{(|\delta' x_i|\geq a)} \\
&= \frac{1}{2}c^2 n^{-1}\sum_{i=1}^{n}1_{(|\varepsilon_i|\leq A)}1_{(|\delta' x_i|\geq a)} \geq \frac{1}{2}c^2 n^{-1}\sum_{i=1}^{n}\{1 - 1_{(|\varepsilon_i|\geq A)} - 1_{(|\delta' x_i|\leq a)}\} \\
&\geq \frac{1}{2}c^2\{1 - P(|\varepsilon_1| \geq A) - F_n(a)\} \text{ for } \lambda \geq (c + A + |\mu_\rho|)/a
\end{aligned}$$

The asymptotic theory using martingales

The asymptotic theory for non-smooth objective function

Theorem 2 *Under more Assumptions (x_i stationary), $\hat{\beta}$ has a first order expansion*

$$N^{-1}(\hat{\beta}-\beta_\rho) = \ddot{h}(\mu_\rho)^{-1}\hat{\Sigma}_n^{-1}N'\sum_{i=1}^n x_i\dot{\rho}(\varepsilon_i-\mu_\rho)+\mathrm{o}_\mathsf{P}(1)$$

Proof:

1. Note that $n\dot{R}_n(\beta)=\sum_{i=1}^n \dot{\rho}(y_i-\beta' x_i)x_i'$ is not smooth
2. Define the martingale

$$M_n^*(\beta)=\sum_{i=1}^n\{\dot{\rho}(y_i-\beta' x_i)-\dot{\rho}(y_i-\beta_\rho' x_i)\}x_i' - \dot{h}\{(\beta-\beta_\rho)'x_i+\mu_\rho\}x_i'$$

3. Note that $\sum_{i=1}^n \dot{h}\{(\beta-\beta_\rho)'x_i+\mu_\rho\}x_i'$ is smooth
4. Prove that for $0<\eta<1/4$, $\sup_{|\beta-\beta_\rho|\leq Bn^{-\eta}} n^{-1/2}|M_n^*(\beta)| = o_P(1)$.
5. Thus replace the score $n\dot{R}_n(\beta)-n\dot{R}_n(\beta_\rho)$ by

$$\sum_{i=1}^n \dot{h}\{(\beta-\beta_\rho)'x_i+\mu_\rho\}x_i'$$

The asymptotic theory for non-smooth objective function

We found $n\dot{R}_n(\beta) - n\dot{R}_n(\beta_\rho) \approx \sum_{i=1}^n \dot{h}\{(\beta - \beta_\rho)' x_i + \mu_\rho\} x_i'$

6. Replace β by $\hat{\beta}$:

$$\begin{aligned}\sum_{i=1}^n \dot{\rho}(\varepsilon_i - \mu_\rho) x_i' N &\approx \sum_{i=1}^n \dot{h}\{(\hat{\beta} - \beta_\rho)' x_i + \mu_\rho\} x_i' N \\ &\approx \ddot{h}(0)(\hat{\beta} - \beta_\rho)' N'^{-1} N' \sum_{i=1}^n x_i x_i' N\end{aligned}$$

Conclude

$$N^{-1}(\hat{\beta} - \beta_\rho) = \frac{1}{\ddot{h}(0)} (N' \sum_{i=1}^n x_i x_i' N)^{-1} N \sum_{i=1}^n x_i \dot{\rho}(\varepsilon_i - \mu_\rho) x_i' + o_P(1).$$

Summary

We have defined M-estimators and in particular the Huber-skip

$$\min_{\beta} \sum_{i=1}^{n} \min\{(y_i - \beta' x_i)^2, c^2\}$$

suggested some 50 years ago.
Using recent martingale results and a "new" definition of scarcity of small regressors, we have proved tightness, consistency, and found an asymptotic expansion from which we can find asymptotic distributions depending on regressors.
The results hold for a wide class of regressors including some deterministic regressors, stationary regressors, and random walk regressors.
The assumptions for the M-estimators include conditions for the objective function ρ, the density f, and the regressors.

Johansen, S. and B. Nielsen (2013). *Asymptotic theory of M-estimators for multiple regression in time series. In progress.*