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

Aarhus Conference of Probability, Statistics and their Applications – Celebrating the Scientific Achievements of Ole E. Barndorff-Nielsen,

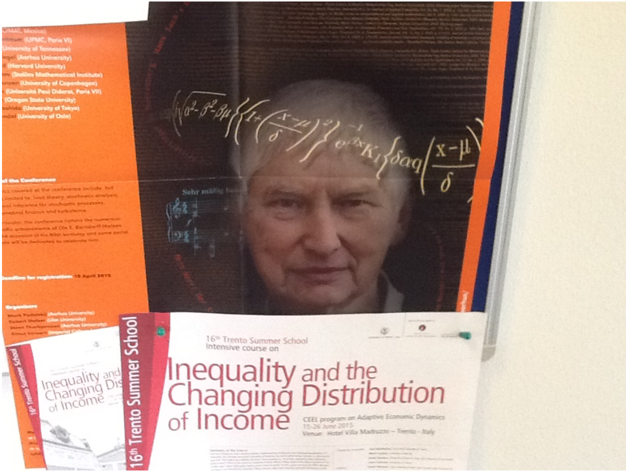
15-19 June at Aarhus University.

Søren Johansen [1]

Joint work with Bent Nielsen[2]

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

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


16th Trento Summer School
Intensive course on
Inequality and the Changing Distribution of Income
CEEL program on Adaptive Economic Dynamics
15-26 June 2015
Venue: Hotel Villa Madruzzo – Trento - Italy
16th Trento Summer School

List of contents

- The multiple regression model
- Huber-skip and M-estimators
- Why Huber-skip ?
- Main results for M-estimators:
  - Tightness,
  - Consistency
  - Stochastic expansion
- Main martingale result
- Conclusion

The setup

The multiple regression model:

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

where ε_i is 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 $R_n(\beta) = n^{-1} \sum_{i=1}^n \rho(y_i - \beta' x_i)$, ρ continuous, "increasing", $\rho(u) \geq 0$ with right and left derivatives. The minimizer is an $M-estimator$. Leading case Huber-skip

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

The project: To find a first order asymptotic expansion of a general class of M-estimators
The technique: We apply martingale techniques to study weighted and marked empirical processes

The Huber-skip estimator

The Huber-skip suggested by Huber (1964) is a special case of an M-estimator chosen because of its robust properties.

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

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

The asymptotic properties are given by Jurečková, Sen, and Picek (2012) for the location model, but only few results has been given for regression

Why Huber-skip ?
1. Difficult to computation 2. Requires known scale 3. More robust estimators exist 4. The mathematics too difficult

M-estimators

Least squares: $\rho(v) = \frac{1}{2}v^2$
Quantile regression: $\rho(v) = -(1-p)v1_{(v<0)} + pv1_{(v\geq 0)}$
Maximum likelihood: $\rho(v) = -\log f(v)$
Huber-skip: $\rho(v) = \frac{1}{2}\min(v^2, c^2)$

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.

Main results

A few definitions
$R_n(\beta) = n^{-1}\sum_{i=1}^n \rho(y_i - \beta' x_i),\ \hat{\Sigma}_n = N'\sum_{i=1}^n x_i x_i' N = O_P(1)$,
$h(\mu) = E(\rho(\varepsilon - \mu)) \geq h(\mu_\rho)$ for all μ, $\dot{h}(\mu_\rho) = 0$,
$\ddot{h}(\mu_\rho) = -\int \dot{\rho}(u - \mu_\rho)\dot{f}(u)du > 0$, $\beta_\rho = (\mu_0 + \mu_\rho, \alpha_0')'$

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

Theorem 2 *Under more Assumptions, $N^{-1}(\hat{\beta} - \beta_\rho) = O_P(n^{1/2-\eta})$ for* $0 < \eta < 1/4$.

Theorem 3 *Under still more Assumptions, $N^{-1}(\hat{\beta} - \beta_\rho) = O_P(1)$ and has a first order expansion*

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

The results for the Huber-skip

For the Huber-skip for symmetric density and stationary regressors:

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

$$n^{1/2}(\hat{\beta} - \beta_0) = \ddot{h}(0)^{-1}\hat{\Sigma}_n^{-1} n^{-1/2} \sum_{i=1}^{n} x_i \varepsilon_i 1_{(|\varepsilon_i| \leq c)} + o_{\mathsf{P}}(1)$$

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

random walk regressors:

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

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

1-step estimators and their iteration

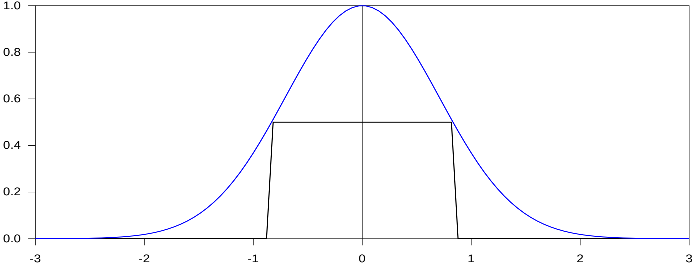
Score function: $n^{-1}\sum_{i=1}^{n}(y_i-\beta' x_i)x_i' 1_{(|y_i-\beta' x_i|\leq c)}=0$ is difficult to solve.
Take some estimator $\breve{\beta}$ and solve $n^{-1}\sum_{i=1}^{n}(y_i-\beta' x_i)x_i' 1_{(|y_i-\breve{\beta}' x_i|\leq c)}=0$

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

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

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

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

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

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 Let $\lambda = |\beta - \beta_\rho|$ and $\delta = (\beta - \beta_\rho)/|\beta - \beta_\rho|$ and
$\rho(x) = \frac{1}{2}\min(x^2, c^2)$,
If $|\delta' x_i| \geq a > 0$, $|\varepsilon_i| \leq A$ and $\lambda \geq (c + A + |\mu_\rho|)/a$, then
$\rho(y_i - \beta' x_i) = \frac{1}{2}c^2$:

$$\begin{aligned} y_i - \beta' x_i &= \varepsilon_i - (\beta - \beta_\rho)' x_i - \mu_\rho = \varepsilon_i - \lambda \delta' x_i - \mu_\rho \\ |y_i - \beta' x_i| &\geq \lambda |\delta' x_i| - |\varepsilon_i| - |\mu_\rho| \geq \lambda a - A - |\mu_\rho| \geq c \end{aligned}$$

A lower bound for $R_n(\beta)$ is

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \rho(y_i - \beta' x_i) 1_{(|\varepsilon_i| \leq A)} 1_{(|\delta' x_i| \geq a)} \\ = & \frac{1}{2}c^2 n^{-1} \sum_{i=1}^n 1_{(|\varepsilon_i| \leq A)} 1_{(|\delta' x_i| \geq a)} \geq \frac{1}{2} c^2 n^{-1} \sum_{i=1}^n \{1 - 1_{(|\varepsilon_i| \geq A)} - 1_{(|\delta' x_i| \leq a)}\} \end{aligned}$$

Tightness continued

Thus the objective function is bounded below for $|\beta - \beta_\rho| \geq (A+c)/a$

$$R_n(\beta) = n^{-1}\sum_{i=1}^n \rho(y_i - \beta' x_i) \geq \frac{1}{2}c^2 n^{-1}\sum_{i=1}^n (1 - 1_{(|\varepsilon_i| \geq A)} - 1_{(|\delta' x_i| \leq a)})$$

Now define $F_n(a) = \sup_{|\delta|=1} n^{-1}\sum_{i=1}^n 1_{(|\delta' x_i| \leq a)}$ and assume that for some $\xi > 0$ and for all $\epsilon > 0$, there exist (a_0, n_0) such that

$$P(F_n(a) \leq \xi) \geq 1 - \epsilon, \;\; n \geq n_0, \;\; a \leq a_0.$$

"The fraction of small regressors" is bounded by ξ with large probability.
$R_n(\beta) - R_n(\beta_\rho)$ is zero for $\beta = \beta_\rho$, and bounded below for
$\lambda = |\beta - \beta_\rho| \geq (A + c + |\mu_\rho|)/a$,

$$n^{-1}\sum_{i=1}^n \{\rho(y_i - \beta' x_i) - \rho(y_i - \beta_\rho' x_i - \mu_\rho)\} \geq \frac{1}{2}c^2(1 - \delta - \xi) - h(\mu_\rho) - \delta > 0.$$

if we choose $0 < \xi < 1 - h(\mu_\rho)/(\frac{1}{2}c^2) < 1$. Hence a minimizer exists with large probability and $|\hat{\beta} - \beta_\rho| \leq (A + c + |\mu_\rho|)/a$.

Examples of the F_n condition for small regressors

For deterministic regressors: the condition $F_n(a) \leq \xi$ is satisfied in the examples

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

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

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

For continuous regressors, $F_n(a) \xrightarrow{P} 0$, for $(a, n) \rightarrow (0, \infty)$ in case

3. 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)

4. 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)

Theorem 2 *Under more Assumptions (x_i stationary),*
$n^{1/2}(\hat{\beta}-\beta_\rho) = O_P(n^{-\eta})$ *for* $0<\eta<1/4$.
Proof: Define $h(\mu) = E_{i-1}\rho(\varepsilon_i - \mu) \geq h(\mu_\rho)$,
$\dot{h}(\mu_\rho) = 0,\ \beta_\rho = (\mu_0 + \mu_\rho, \alpha_0')'$

$$\begin{aligned} & R_n(\beta) - R_n(\beta_\rho) \\ = \ & n^{-1}M_n(\beta) + n^{-1}\sum_{i=1}^n [h\{(\beta-\beta_\rho)'x_i + \mu_\rho\} - h(\mu_\rho)] \end{aligned}$$

$$M_n(\beta) = \sum_{i=1}^n \{\rho(\varepsilon_i - (\beta-\beta_\rho)'x_i - \mu_\rho) - \rho(\varepsilon_i - \mu_\rho)\} - E_{i-1}(\cdots)$$

We show that $\sup_{|\beta-\beta_\rho|\leq B|} n^{-1}|M_n(\beta)| = o(1)$, and assume
$h(\mu) \geq h(\mu_\rho) + \min(\epsilon, (\mu-\mu_\rho)^2)$, and show that
$n^{-1}\sum_{i=1}^n [h\{\lambda\delta' x_i + \mu_\rho\} - h(\mu_\rho)]1_{(|\delta' x_i|\geq a)} \geq c(\delta' x_i)^2 n^{-2\eta}(1 - F_n(a))$ for
$\lambda \geq n^{-\eta}$.

Theorem 3 *Under still more Assumptions (x_i stationary), $\hat{\beta}$ has a first order expansion*

$$n^{1/2}(\hat{\beta}-\beta_\rho) = \ddot{h}(\mu_\rho)^{-1}\hat{\Sigma}_n^{-1} n^{-1/2}\sum_{i=1}^n x_i \dot{\rho}(\varepsilon_i-\mu_\rho) + o_P(1)$$

Proof: For $\psi(u) = u1_{(|u|\leq c)}$ we have $\rho(x) = \int_0^x \psi(u)du$. The same proof works with $\dot{R}_n(\beta)$ instead of $R_n(\beta)$

$$\begin{aligned} & n^{1/2}\{\dot{R}_n(\beta)-\dot{R}_n(\beta_\rho)\} \\ = \; & n^{-1/2}\{M_n^*(\beta)-M_n^*(\beta_\rho)\} + n^{-1/2}\sum_{i=1}^n [\dot{h}\{(\beta-\beta_\rho)'x_i+\mu_\rho\}-\dot{h}(\mu_\rho)]x_i' \end{aligned}$$

$$M_n^*(\beta) = \sum_{i=1}^n \{\dot{\rho}(\varepsilon_i-(\beta-\beta_\rho)'x_i-\mu_\rho)-\dot{\rho}(\varepsilon_i-\mu_\rho)\}x_i' - E_{i-1}\{\dots\}$$

We show that for $0<\eta<1/4$, $\sup_{|\beta-\beta_\rho|\leq Bn^{-\eta}|} n^{-1/2}|M_n^*(\beta)| = o_P(1)$.

The first order condition implies when we insert $\hat{\beta}$ and use that $\dot{h}(\mu_\rho) = -E_{i-1}\dot{\rho}(\varepsilon_i - \mu_\rho) = 0$ and $n^{-1/2}M_n^*(\hat{\beta}) = o_P(1)$ that

$$
\begin{aligned}
-n^{-1/2}\sum_{i=1}^{n}\dot{\rho}(\varepsilon_i - \mu_\rho)x_i &= n^{-1/2}\sum_{i=1}^{n}\dot{h}\{(\hat{\beta} - \beta_\rho)'x_i + \mu_\rho\}x_i' + o_P(1) \\
&= -n^{1/2}(\hat{\beta} - \beta_\rho)'\{n^{-1}\sum_{i=1}^{n}x_i x_i'\}\ddot{h}(\mu_\rho) + o_P(1)
\end{aligned}
$$

The inequality of Bercu and Touati (2008, Theorem 2.1)

For a square integrable martingale M_n :

$$P(|M_n| \geq x, \sum_{i=1}^{n} \{(\Delta M_i)^2 + E_{i-1}(\Delta M_i)^2\} \leq y) \leq 2 \exp\left(-\frac{x^2}{2y}\right)$$

Main martingale result

Let $u_{ni}(\kappa), \kappa \in \mathcal{K} \subset R^m,\ \mathcal{K}$ compact, $u_{ni}(\kappa_0) = 0,\ 1 \leq i \leq n$ and define the martingales with respect to a filtration $\mathcal{F}_i$

$$M_n(\kappa) = \sum_{i=1}^{n} u_{ni}(\kappa) - E_{i-1}(u_{ni}(\kappa))$$

Theorem Assume there exists r such that $2^r > 2 + m$, and $1 \leq p \leq 2^r$ and such that

$$n^{-1} \sum_{i=1}^{n} E[\sup_{\kappa \in B(\kappa_0, B)} E_{i-1}\{\sup_{\tilde{\kappa} \in \mathcal{B}(\kappa, Qn^{-\phi})} |u_{ni}(\kappa) - u_{ni}(\tilde{\kappa})|^p\}] \leq n^{-\phi} C$$

Then

$$\begin{aligned} \sup_{\kappa \in B(\kappa_0, B)} |n^{-1} M_n(\kappa)| &= o_P(1), \\ \sup_{\kappa \in B(\kappa_0, Bn^{-\eta})} |n^{-1/2} M_n(\kappa)| &= o_P(1) \text{ for } 0 < \eta < 1/2. \end{aligned}$$

Verification of conditions for small martingale

For the Huber-skip we take $\rho(u) = \frac{1}{2}\min(u^2, c^2), \dot{\rho}(u) = u1_{(|u|\leq c)}$

$$\begin{aligned} |\rho(u) - \rho(v)| &\leq c|u - v| \\ |u1_{(|u|\leq c)} - v1_{(|v|\leq c)}| &\leq |u - v| + c(1_{(|v-c|\leq|u-v|)} + 1_{(|v+c|\leq|u-v|)}) \end{aligned}$$

Define $u_{ni}(\beta) = \rho(y_i - \beta' x_i) - \rho(y_i - \beta_\rho' x_i)$ implies
$E_{i-1}\sup_{|\beta-\tilde{\beta}|\leq Qn^{-\phi}} |u_{ni}(\beta) - u_{ni}(\tilde{\beta})| \leq Cn^{-\phi}|x_i|$
Define $u_{ni}(\beta) = \{\dot{\rho}(y_i - \beta' x_i) - \dot{\rho}(y_i - \beta_\rho' x_i)\}x_i'$ implies

$$E_{i-1}\sup_{|\beta-\tilde{\beta}|\leq Qn^{-\phi}} |u_{ni}(\beta) - u_{ni}(\tilde{\beta})| \leq Cn^{-\phi}|x_i|^2 + C\sup_v f(v)|n^{-\phi}|x_i|^2$$

Conditions for uniformly small martingales satisfied under moment conditions on x_i and a bounded density of ε_i.

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 result 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.*