A sign and rank based semiparametrically efficient estimator for regression analysis

V. Verardi[1,2] and C. Vermandele[2]

[1]FNRS, UNamur

[2]ULB

UK Stata Conference, 2018

Regression estimators

Gauss-Markov assumptions

- **Ordinary Least Squares** (OLS) is undoubtedly the simplest and **most commonly used estimator** for linear regression analysis.

- Under a set of hypotheses, called **Gauss-Markov** (GM) assumptions, this estimator is the **most efficient** linear unbiased estimator.

- One of these assumptions is that errors are **normally distributed**. In case of heavy-tailed and/or asymmetrical distribution of the error term, OLS is not the most efficient estimator anymore.

- Errors with a **heavier tailed distribution** can result in extreme observations and can significantly affect the OLS estimates of regression coefficients; the **loss in efficiency** can be very large.

Gauss-Markov assumptions

- If the **innovation** distribution is **known** and **not gaussian**, the OLS estimator is outperformed by the **maximum likelihood estimator**.
- If the **true** error **distribution** is **unknown**, a nice solution is to **approximate it**, relying on the information available in the sample, and to estimate the regression coefficients by **pseudo-maximum likelihood**.
- Unfortunately, this often leads to a **rather complex maximum likelihood optimization** problem as the density function is either very complicated or non-explicit.
- The optimization problem needed to fit the model is generally **difficult to handle**.

Gauss-Markov assumptions

- *Xu and Genton (2015)* consider that the distribution of the error term belongs to the family of the **Tukey** g**-and-**h **distributions** ($T_{g,h}$) and propose a computationally efficient numerical procedure to estimate jointly, by maximum likelihood, the **parameters of the error distribution and the regression coefficients**.
- The $T_{g,h}$ distribution, proposed by *Tukey(1977)*, is **extremely flexible** and **approximates well** a large number of **commonly used densities**.
- In essence, this distribution is defined as a **transformation of the standard normal** allowing to introduce skewness and to obtain larger tail heaviness.
- The difficulty encountered by *Xu and Genton (2015)* is that **there is no explicit expression for the density function** of a $T_{g,h}$

Tukey g-and-h

Definition

Let Z be a random variable from the standard normal distribution $\mathcal{N}(0, 1)$. Define the random variable Y through the transformation

$$Y = \xi + \omega\tau_{g,h}(Z)$$

where $\xi \in \mathbb{R}$, $\omega > 0$, and

$$\tau_{g,h}(z) = \frac{1}{g}\left(e^{gz} - 1\right)e^{hz^2/2}$$

with $g \in \mathbb{R}$ and $h \geq 0$.

Variable Y is said to have a Tukey's g-and-h distribution with location parameter ξ and scale parameter ω: $Y \sim T_{g,h}(\xi, \omega)$. Parameter g controls the direction and the degree of skewness and h controls the tail thickness.

Tukey g-and-h

It is easy to show that the **density** function of the $T_{g,h}(\xi,\omega)$-**distributed** random variable Y takes the form:

$$f_{Y|\boldsymbol{\theta}}(y) = \frac{\phi\left(\tau_{g,h}^{-1}\left(\frac{y-\xi}{\omega}\right)\right)}{\omega\tau'_{g,h}\left(\tau_{g,h}^{-1}\left(\frac{y-\xi}{\omega}\right)\right)}, \qquad y \in \mathbb{R},$$

where $\phi(\cdot)$ is the standard normal density function, and $\tau_{g,h}^{-1}(\cdot)$ and $\tau'_{g,h}(\cdot)$ are the inverse and the first derivative of function $\tau_{g,h}(\cdot)$, respectively.

By defining $Q_{g,h}(u) = \tau_{g,h}\left(\Phi^{-1}(u)\right)$, the density function can be rewritten as

$$f_{Y|\boldsymbol{\theta}}(y) = \frac{1}{\omega Q'_{g,h}\left(Q_{g,h}^{-1}\left(\frac{y-\xi}{\omega}\right)\right)}, \qquad y \in \mathbb{R},$$

where $Q_{g,h}^{-1}(\cdot)$ and $Q'_{g,h}(\cdot)$ are the inverse and the first derivative of the quantile function $Q_{g,h}(\cdot)$ of the standardized $T_{g,h}(0,1)$-distribution.

Maximum Likelihood

Let $y_1, \ldots, y_n$ be the realizations of n i.i.d. random variables of **unknown density** f, possibly **skewed and/or heavy tailed**.

If density f is **assumed to be a** $T_{g,h}(\xi, \omega)$, we may try to estimate $\boldsymbol{\theta} = (\xi, \omega, g, h)^{\mathrm{T}}$ by maximizing log-likelihood:

$$\begin{aligned}
\ell^{(n)}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} ln(f_{Y|\boldsymbol{\theta}}(y_i)) \\
&= \sum_{i=1}^{n} \left[\ln \phi \left(\tau_{g,h}^{-1} \left(\frac{y_i - \xi}{\omega} \right) \right) - \ln \omega - \ln \tau_{g,h}' \left(\tau_{g,h}^{-1} \left(\frac{y_i - \xi}{\omega} \right) \right) \right] \\
&= \sum_{i=1}^{n} \left[- \ln \omega - \ln Q_{g,h}' \left(Q_{g,h}^{-1} \left(\frac{y_i - \xi}{\omega} \right) \right) \right].
\end{aligned}$$

However, since $\tau_{g,h}^{-1}(\cdot)$ and $Q_{g,h}^{-1}(\cdot)$ do not have a closed form, numerical evaluation of $\ell^{(n)}(\boldsymbol{\theta})$ is needed.

Order statistics

Here, we suggest to **minimize the squared difference** between **theoretical quantiles** of the Tukey's distribution and the **empirical order statistics**:

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left[y_{(i)} - \left\{ \xi + \omega Q_{g,h}\left(\frac{i}{n+1} \right) \right\} \right]^2 \\
&= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left[y_{(i)} - \left\{ \xi + \omega \tau_{g,h}\left(z_{i/(n+1)} \right) \right\} \right]^2 ,
\end{aligned}
$$

where $y_{(i)}$ is the ith order statistics among $y_1, \ldots, y_n$ and $z_{i/(n+1)} = \Phi^{-1}\left(i/(n+1)\right)$ is the quantile of order $i/(n+1)$ of the standard normal distribution.

The **score function to minimize** is of the **non linear least squares** type and the minimization problem can easily be solved by a classic algorithm (of Gauss Newton type, for instance).

Semiparametric model

- We propose here a **quite different approach**, that finds its foundation in *Vermandele (2000)*, *Hallin et al. (2006)* and *Hallin et al. (2008)*.
- We **consider** median-restricted regression model, that is a **regression model** where the error term has zero median, but otherwise **unspecified density** f.
- This model is a *semiparametric* model, with the unknown **innovation density** playing the role of an **infinite dimensional nuisance parameter**.
- In this context, semiparametric theory leads us to **define a sign and rank based estimator** of the regression coefficients as a one-step update of an initial root n consistent estimator.
- The **score function**, initially defined on the basis of the exact underlying innovation density f, is **estimated** using the fact that f **can be well adjusted by a Tukey** g**-and-**h distribution.

Semiparametric median-restricted regression model

Let us consider the following linear **regression model**: for $i = 1, \ldots, n$,

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i \qquad (1)$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathrm{T}}$, $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ and where the i.i.d error terms ε_i have zero median, but otherwise unspecified density f and distribution function F.

Let $\mathcal{F}_0 = \left\{ f : \mathbb{R} \to [0, \infty) \text{ such that } \int_{-\infty}^{0} f(z)dz = \int_{0}^{\infty} f(z)dz = 1/2 \right\}$ denote the set of all **densities** on the real line **that have median 0**. Since the innovation density is unknown, it plays the role of a nonparametric nuisance.

Model $y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i$ defines a *semiparametric* model. The residuals $e_1(\boldsymbol{\beta}) = y_1 - \mathbf{x}_1^{\mathrm{T}} \boldsymbol{\beta}, \ldots, e_n(\boldsymbol{\beta}) = y_n - \mathbf{x}_n^{\mathrm{T}} \boldsymbol{\beta}$ are i.i.d. with (marginal) density $f \in \mathcal{F}_0$.

Local Asymptotic Normality

Definition

In statistics, **local asymptotic normality**, introduced by *Le Cam (1960)*, is a property of a sequence of statistical models.

A sequence of statistical models is "locally asymptotically normal" if, asymptotically, their **likelihood ratio processes** are similar to those for a **normal location parameter**.

Technically, if the **log likelihood is approximately quadratic** with constant Hessian, then the **MLE is approximately normal**.

Intuitively speaking, statistical **inference in a LAN model** is asymptotically locally **equivalent to inference in a Gaussian shift experiment**.

Wikipedia

Parametric modelling

- Parametric theory has received a lot of attention in the literature. Indeed, most well-known elementary statistical methods are parametric.
- One of the most important results concerns the asymptotic normality and efficiency of the **maximum likelihood estimator** (MLE), rooted in the work by **Fisher in the 1920**s.
- Another key result is the **lower bound theory** rooted in the work by **Cramer and Rao in the 1940s**
- The **semiparametric approach** to misspecification is to allow the functional form of **some components** of the model to be **unrestricted**.
- Therefore, **solutions**, if they exist and are reasonable, will have **greater applicability and robustness**.

Semiparametric efficiency

Semiparametric modelling

- **Efficiency bounds** quantify the efficiency loss that can result from a **semiparametric**, **rather than a parametric**, approach.
- These bounds provide a **guide to estimation** methods of the **parametric components** of the model
- Any $\sqrt{n}$-**consistent** and asymptotically normal **under** the **semiparametric assumptions**, is actually in the **same class as the maximum-likelihood** estimator of the parameter in the parametric submodel, and therefore has an asymptotic variance no smaller than the bound for the parametric submodel.
- Since this comparison holds for each parametric submodel that one could consider, it follows that the **asymtpotic variance** of any semiparametric estimator is **no smaller than the supremum of the Cramer-Rao bounds** for all parametric submodels.

Efficient estimator

- Classical **likelihood inference** for β can be **based on the parametric Rao score** (log-likelihood derivatives), **or**, in **Le Cam's "uniform local asymptotic normality"** terminology, on the *central sequence:* $\mathbf{\Delta}_f^{(n)}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f(e_i(\beta))\, \mathbf{x}_i$ with $\phi_f(e) = -\frac{f'(e)}{f(e)}$.
- As $n \to \infty$, under $\mathrm{P}_{f;\beta}^{(n)}$, $\mathbf{\Delta}_f^{(n)}(\beta) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{I}_f)$ where $\mathbf{I}_f$ is the (parametric) **Fisher information** matrix for β
- In particular, if $\widetilde{\beta}^{(n)}$ is a $\sqrt{n}$-consistent—but possibly inefficient—estimator of β, then

$$\widehat{\beta}_f^{(n)} = \widetilde{\beta}^{(n)} + \frac{1}{\sqrt{n}} (\mathbf{I}_f)^{-1} \mathbf{\Delta}_f^{(n)}(\widetilde{\beta}^{(n)})$$

is an *efficient* estimator of β: under $\mathrm{P}_{f;\beta}^{(n)}$, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\beta}_f^{(n)} - \beta\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, (\mathbf{I}_f)^{-1}\right)$$

Efficient estimator

As $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\beta})$ in general is not properly centred under density $g \neq f$, inference based on this central sequence is not valid when density f used for the score function $\phi_f(\cdot)$ does not coincide with the true error density; the estimator $\widehat{\boldsymbol{\beta}}_f^{(n)}$ is no longer $\sqrt{n}$-consistent.

However, in the presence of a suitable group invariance structure, semiparametrically efficient central sequence can be obtained by conditioning $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\beta})$ on the *maximal* invariant.

Proposition 1 Under $\mathrm{P}^{(n)}_{f;\boldsymbol{\beta}}$, as $n \to \infty$,

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\beta})\,\middle|\,\mathbf{N}^{(n)}(\boldsymbol{\beta}), \mathbf{R}^{(n)}(\boldsymbol{\beta})\right] &= \boldsymbol{\Delta}_f^{(n)*}(\boldsymbol{\beta}) + o_{\mathrm{P}}(1) \\ &= \underline{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\beta}) + o_{\mathrm{P}}(1) \end{aligned}$$

Efficient estimator

Define

$$\underline{\widehat{\boldsymbol{\Delta}}}^{(n)*}(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{\varphi}^{(n)}\left(\underline{R}_i^{(n)}(\boldsymbol{\beta})\right)\left[\mathbf{x}_i - \bar{\mathbf{x}}^{(n)}\right] + 2\widehat{O}^{(n)}\frac{1}{\sqrt{n}}\left(N_+^{(n)}(\boldsymbol{\beta}) - N_-^{(n)}(\boldsymbol{\beta})\right)\bar{\mathbf{x}}^{(n)}$$

and

$$\widehat{\mathbf{I}}^{(n)*} = \widehat{\mathcal{I}}^{(n)}\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i - \bar{\mathbf{x}}^{(n)}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}^{(n)}\right)^{\mathrm{T}} + \left(2\widehat{O}^{(n)}\right)^2\bar{\mathbf{x}}^{(n)}\left(\bar{\mathbf{x}}^{(n)}\right)^{\mathrm{T}}$$

with $\varphi_f(u) = \phi_f\left(F^{-1}(u)\right)$, for $u \in (0,1)$ and denoting by $\mathbf{R}^{(n)}(\boldsymbol{\beta})$ and $\mathbf{s}^{(n)}(\boldsymbol{\beta})$ the vector of ranks and the vector of signs associated with the residuals $e_1(\boldsymbol{\beta}), \ldots, e_n(\boldsymbol{\beta})$. Define $N_+^{(n)}(\boldsymbol{\beta})$ and $N_-^{(n)}(\boldsymbol{\beta})$ as the numbers of positive and negative residuals, respectively.

Efficient estimator
Define

$$\underline{R}_i^{(n)}(\boldsymbol{\beta}) = I\left[s_i(\boldsymbol{\beta}) = -1\right]\left(\frac{1}{2}\cdot\frac{R_i^{(n)}(\boldsymbol{\beta})}{N_-^{(n)}(\boldsymbol{\beta})+1}\right) + I\left[s_i(\boldsymbol{\beta}) = +1\right]\left(\frac{1}{2}+\frac{1}{2}\cdot\frac{R_i^{(n)}(\boldsymbol{\beta}) - \left(n - N_+^{(n)}(\boldsymbol{\beta})\right)}{N_+^{(n)}(\boldsymbol{\beta})+1}\right)$$

For the estimation of $\varphi_f(\cdot)$, $\mathcal{I}_f$ and $f(0)$, we approximate error density f using a Tukey distribution with location parameter (median) equal to zero and skewness parameter g, tail heaviness parameter h and scale parameter w estimated from the residuals $e_i(\widetilde{\boldsymbol{\beta}}^{(n)})$, $i = 1, \ldots, n$ by solving
$\widehat{\boldsymbol{\theta}} \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^n \left[e_{(i)} - \left\{\xi + \omega Q_{g,h}\left(\frac{i}{n+1}\right)\right\}\right]^2$

Efficient estimator

Then, denoting by $\widehat{f}$ the density function of the $T_{\widehat{g},\widehat{h}}(0,\widehat{\omega})$-distribution and we have:

$$\widehat{O}^{(n)} = \widehat{f}(0) = \frac{1}{\widehat{\omega} Q'_{\widehat{g},\widehat{h}}\left(Q^{-1}_{\widehat{g},\widehat{h}}(0)\right)} = \frac{1}{\widehat{\omega} Q'_{\widehat{g},\widehat{h}}(1/2)} = \frac{1}{\widehat{\omega}\sqrt{2\pi}},$$

$$\widehat{\varphi}^{(n)}(u) = \varphi_{\widehat{f}}(u) = \frac{Q''_{\widehat{g},\widehat{h}}(u)}{\widehat{\omega}\left[Q'_{\widehat{g},\widehat{h}}(u)\right]^2}$$

and

$$\widehat{\mathcal{I}}^{(n)} = \int_{-\infty}^{\infty} \phi^2_{\widehat{f}}(y)\widehat{f}(y)dy = \int_0^1 \varphi^2_{\widehat{f}}(u)du$$

where the integral is determined numerically.

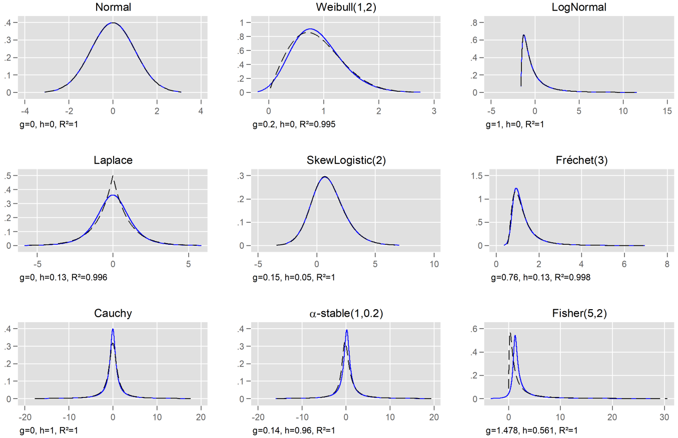
Setup (n=100, n=1000)

Generate variables x_1, x_2 and x_3 from three independent standard random normals. Then generate $y = x_1 + x_2 + x_3 + \varepsilon$ where ε is distributed according to some specific distribution:

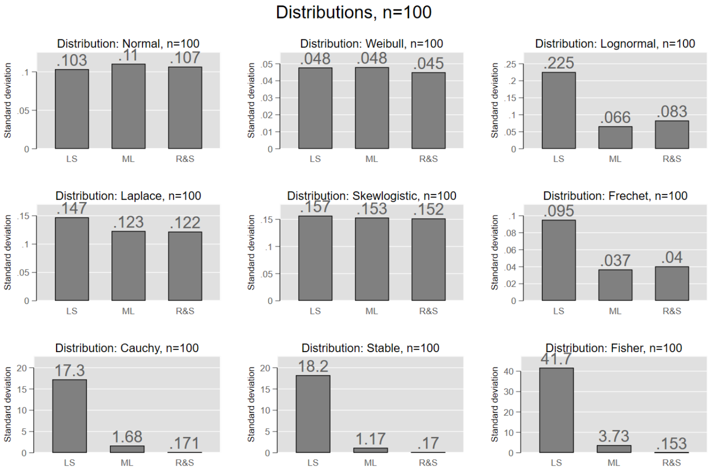
1. Heavy-tailed distributions. The distributions considered are i) Normal, ii) Weibull(1,2), iii) LogNormal, iv) standard Laplace, v) SkewLogistic(2), vi) Fréchet(3), vii) standard Cauchy, viii) Stable(1,0.2) and ix) Fisher(5,2).

2. Tgh distributions with varying g and h parameters with $g \in \{0, 0.25, 0.5, 0.75\}$ and $h \in \{0, 0.25, 0.5, 0.75\}$.

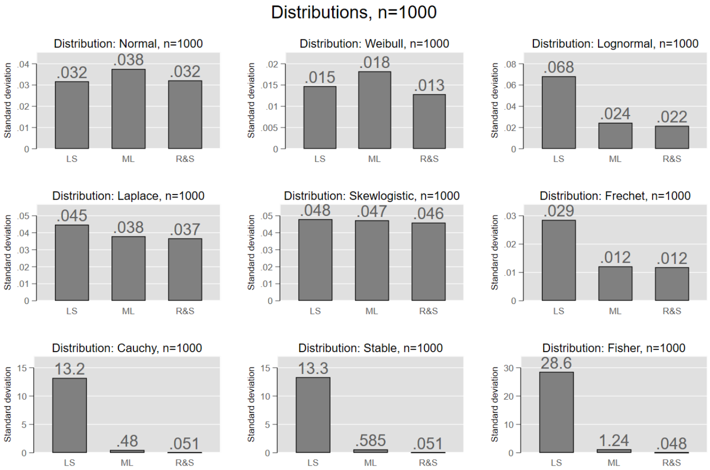
Selected distributions



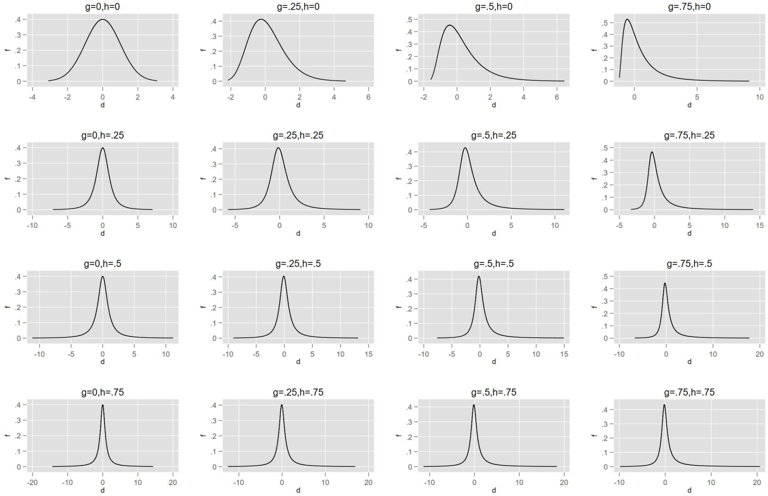
Average standard error



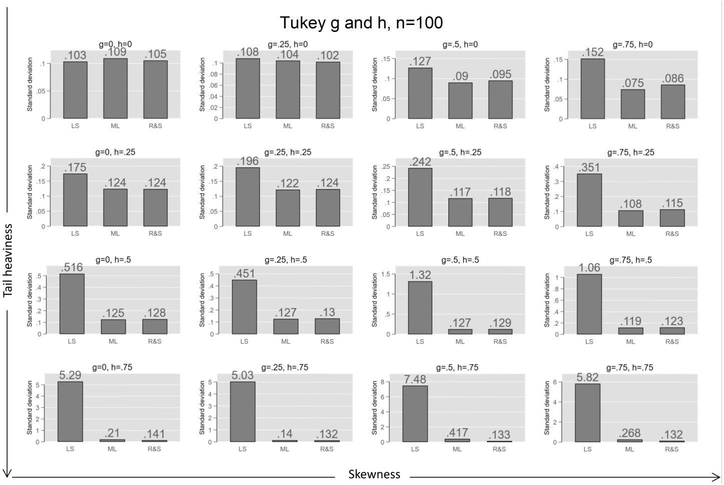
Average standard error



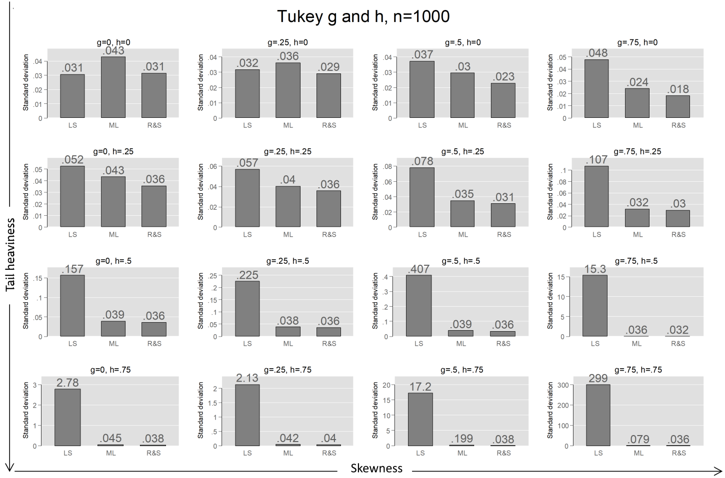
Tukey g-and-h


g=0,h=0
g=.25,h=0
g=.5,h=0
g=.75,h=0
g=0,h=.25
g=.25,h=.25
g=.5,h=.25
g=.75,h=.25
g=0,h=.5
g=.25,h=.5
g=.5,h=.5
g=.75,h=.5
g=0,h=.75
g=.25,h=.75
g=.5,h=.75
g=.75,h=.75
f
d

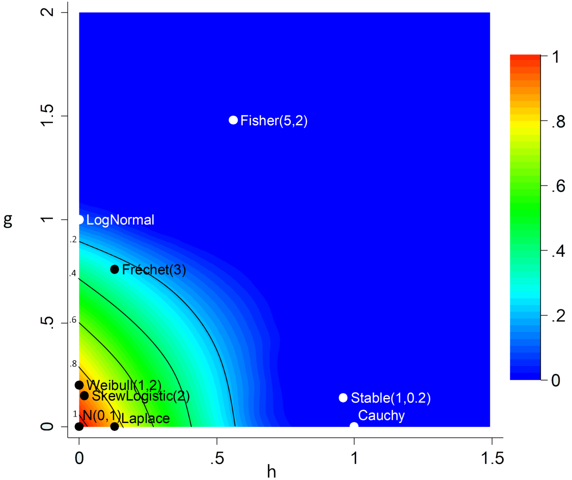
Average standard error



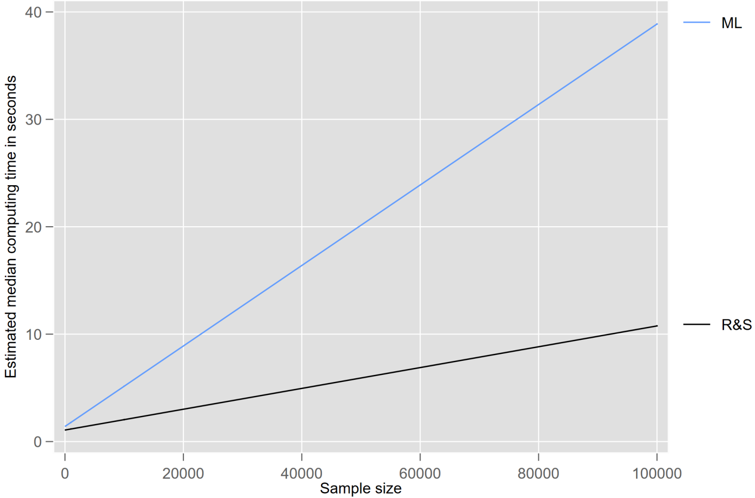
Average standard error



Relative variance w.r.t. OLS



R&S vs ML


40
30
20
10
0
Estimated median computing time in seconds
0
20000
40000
60000
80000
100000
Sample size
ML
R&S

Semiparametrically efficient sign and rank regression estimators

Now we only have to plug everything in an ado-file

The idea here is to estimate regression parameters and the distribution of the error term jointly.

We can assume that the distribution could be reasonable well approximated by a Tukey g-and-h distribution as expected here or rely on a kernel density estimation

- `flexrank` *depvar indepvars* `[if] [in]`
  - Tukey g-and-h based score function
- `flexnp` *depvar indepvars* `[if] [in]`
  - Kernel based score function

Both methods should be asymptotically equivalent but the former has much better small sample behaviour.

Example

```
clear
set seed 1234567
set obs 50
drawnorm x1-x3
gen e=(-ln(uniform()))^(-1/3)
gen y=x1+x2+x3+e
flexrank y x*
flexnp y x*
qreg y x*
reg y x*
```

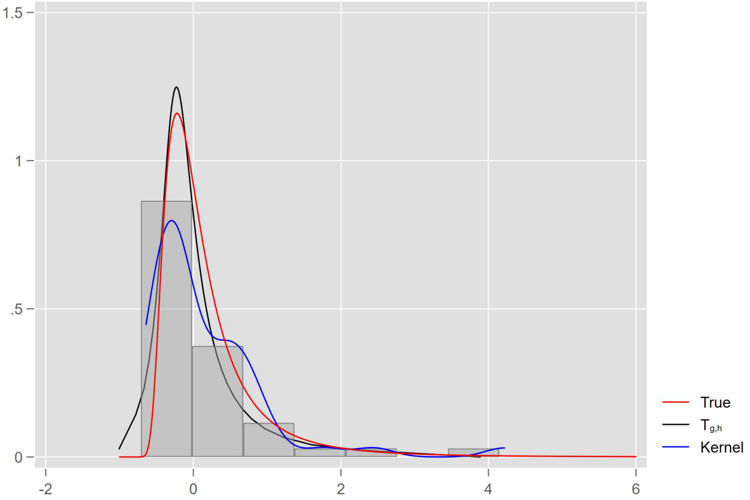
Example

VARIABLES	R&S	NP	L_1	LS
x1	1.054***	1.058***	1.030***	1.166***
	(0.050)	(0.091)	(0.111)	(0.124)
x2	0.909***	0.942***	0.960***	0.916***
	(0.048)	(0.088)	(0.107)	(0.119)
x3	1.099***	0.975***	1.210***	1.219***
	(0.050)	(0.092)	(0.112)	(0.125)
Constant	1.236***	1.245***	1.183***	1.427***
	(0.066)	(0.095)	(0.109)	(0.122)
Observations	50	50	50	50
R-squared	0.973	0.860	0.727	0.878

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

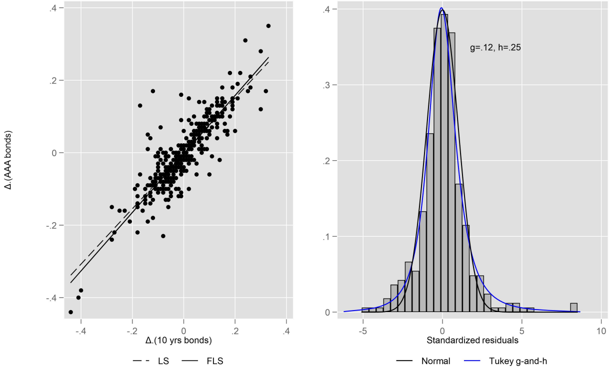
Fréchet example

Example



Example

Week-to-week differences in AAA bond rates are regressed on the difference in 10-year bond rates (period 2002-2014).



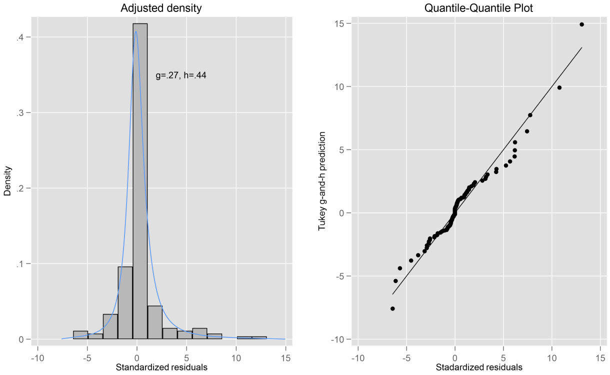
Estimated gain in efficiency w.r.t. LS: 44.4%

Example

```
. reg d*

      Source |       SS           df       MS      Number of obs   =       365
-------------+----------------------------------   F(1, 363)       =   1101.97
       Model |  2.57961138         1  2.57961138   Prob > F        =    0.0000
    Residual |  .849746427       363  .002340899   R-squared       =    0.7522
-------------+----------------------------------   Adj R-squared   =    0.7515
       Total |  3.42935781       364  .009421313   Root MSE        =    .04838

------------------------------------------------------------------------------
    daaabond |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     dbons10 |   .7639579   .0230136    33.20   0.000     .7187013    .8092146
       _cons |  -.0019117   .0025325    -0.75   0.451    -.0068919    .0030685
------------------------------------------------------------------------------


. flexrank d*
------------------------------------------------------------------------------
    daaabond |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     dbons10 |   .8069863   .0171641    47.02   0.000     .7733453    .8406273
       _cons |  -.0034377    .002088    -1.65   0.100    -.0075301    .0006547
------------------------------------------------------------------------------
```

Example

Determinants of prices of 180 economics journals at US libraries, for the year 2000.



Efficiency gain: 69.4%

Results

Regression results

VARIABLES	LS	R&S
# Pages	0.538***	0.377***
	(0.061)	(0.030)
Characters pp.	0.055	0.049***
	(0.034)	(0.017)
Total citations	-0.003	-0.029**
	(0.024)	(0.012)
First year	1.981	2.446***
	(1.248)	(0.625)
Society	-270.397**	-156.973**
	(133.634)	(66.943)
Constant	-4,157.581*	-4,994.901***
	(2,441.560)	(1,223.088)
Publisher and field F.E.		
Observations	180	180
R-squared	0.806	0.896

S.E. in parentheses, *, **, *** indicate a significance at 10%, 5% and 1% respectively

Conclusion

What to take back home

- One of the Gauss-Markov assumptions in a regression model is that errors are **normally distributed**. In case of heavy-tailed and/or asymmetrical distribution of the error term, OLS is not the most efficient estimator anymore
- **Semiparametric efficiency** can be reached using a sign and rank estimator
- The **density** of the error term can be **estimated jointly** with **regression parameters**
- Residuals **estimated density** can be easily **plotted**
- The density estimation can be done using **kernels** of more efficiently using a **Tukey g-and-h approximation**
- **Stata commands** are **available** upon request

References

Hallin, M.; Vermandele, C. and B. Werker. (2006). “Serial and nonserial sign-and-rank statistics: Asymptotic representation and asymptotic normality”. Annals of Statistics 34(1), pp. 254-289.

Hallin, M.; Vermandele, C. and B. Werker. (2008). “Semiparametrically Efficient Inference Based on Signs and Ranks for Median-Restricted Models”. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 70(2), pp. 389-412.

Le Cam, L. (1960). “Locally asymptotically normal families of distributions”. University of California publications in statistics, 3. University of California Press.

Nolan, J. P. and D. Ojeda-Revah (2013). “Linear and nonlinear regression with stable errors”. Journal of Econometrics 172, pp.186-194.

References

Stock, J. H., & M.W. Watson (2007). "Introduction to econometrics". Boston: Pearson/Addison Wesley.

Tukey, J. W. (1977). "Modern techniques in data analysis, NSF-sponsored regional research conference at Southeastern Massachusetts University, North Dartmouth, MA.

Vermandele, C. (2000). "Semiparametrically Efficient Sign-and-Rank Methods for Median Regresssion and AR Models". PhD dissertation, Université libre de Bruxelles, Brussels, Belgium.

Xu, G and Genton, M. (2015). "Efficient maximum approximated likelihood inference for Tukey's g-and-h distribution". Computational Statistics & Data Analysis 91, pp. 78-91..