Robust estimation of a location parameter with the integrated Hogg function



DOI: http://doi.org/10.1016/j.spl.2020.108812

# Robust Estimation of a Location Parameter with the Integrated Hogg Function

Leopoldo Catania[a], Alessandra Luati[b]

[a] *Aarhus BSS and CREATES*
[b] *University of Bologna*

**Abstract**

We study the properties of an M-estimator arising from the minimisation of an integrated version of the quantile loss function. The estimator depends on a tuning parameter which controls the degree of robustness. We show that the sample median and the sample mean are obtained as limit cases. Consistency and asymptotic normality are established and a link with the Hodges-Lehmann estimator and the Wilcoxon test is discussed. Asymptotic results indicate that high levels of efficiency can be reached by specific choices of the tuning parameter. A Monte Carlo analysis investigates the finite sample properties of the estimator. Results indicate that efficiency can be preserved in finite samples by setting the tuning parameter to a low fraction of a (robust) estimate of the scale.

*Keywords:* M-estimators; scoring rules; quantile function; robustness.

## 1. Introduction

Consider the following location scale model:

$$Y_i = \mu + \sigma\varepsilon_i \tag{1}$$

where $\varepsilon_i$ has cumulative density function (cdf) $F_\varepsilon \in \mathcal{F}$ where $\mathcal{F}$ is the set of all continuous symmetric cdfs with location equal to 0 and scale equal to 1. Assume that the $\varepsilon_i$ form an independent sequence. Our aim is to estimate the location parameter $\mu$ of the cdf of $Y_i$, from a sample of observations $\mathbf{y} = (y_i, i = 1, \ldots, N)$. Throughout the paper, we shall assume that the scale, $\sigma$, is known. Since $\varepsilon_i$ is symmetric with location 0, $\mu$ can be estimated by the sample median of $(y_i, i = 1, \ldots, N)$. Alternatively, the empirical mean is also a consistent estimator of $\mu$, provided that $\varepsilon_i$ has finite first moment.

Let $Q_Y(\tau)$ and $Q_\varepsilon(\tau) = F_\varepsilon^{-1}(\tau)$ be the quantile functions of $Y_i$ and $\varepsilon_i$, respectively, evaluated at the probability level $\tau \in (0, 1)$. Clearly, the following relation holds:

$$Q_Y(\tau) = \mu + \sigma Q_\varepsilon(\tau),$$

with $Q_\varepsilon(0.5) = 0$ and $Q_Y(0.5) = \mu$. Given this parameterization, a loss function which minimizes the distance between the theoretical quantile function and its empirical counterpart, evaluated at $y_i$, can be constructed as follows:

$$\begin{aligned} IHF(y_i, \mu, \sigma, F_\varepsilon) &= \int_0^1 (y_i - Q_Y(\tau))(\tau - \mathbb{1}(y_i \leq Q_Y(\tau)))d\tau, \\ &= \int_0^1 (y_i - \mu - \sigma F_\varepsilon^{-1}(\tau)) \left(\tau - \mathbb{1}\left(\frac{y_i - \mu}{\sigma} \leq F_\varepsilon^{-1}(\tau)\right)\right) d\tau \end{aligned} \quad (2)$$

where $\mathbb{1}(\cdot)$ is the indicator function. Equation (2) is obtained as the limit of the objective function which defines the Hogg estimator introduced by Koenker (1984), see also Koenker (2005, Section 5.5). For this reason, we label it "integrated Hogg function" (IHF). Evidently, IHF integrates all quantile losses $\rho_\tau(u) = u(\tau - \mathbb{1}(u \leq 0)$, by exploiting the location scale parametrization of $Y_i$.

With a different label, the IHF has been employed as a proper scoring rule for quantiles (Gneiting and Raftery, 2007); early works include the articles of Matheson and Winkler (1976) and Cervera and Munoz (1996). Gneiting and Ranjan (2011) considered a weighted version of IHF by including a positive weighting function, in order to emphasise different quantiles (for instance those in the tails). In the seminal work of Koenker (1984), these weights were also considered, however with the goal of minimizing the asymptotic variance of the L-Estimator proposed therein. Laio and Tamea (2007) proved that equation (2) is actually proportional to the continuous rank probability score (CRPS) introduced by Brown (1974) and Matheson and Winkler (1976), see also Hersbach (2000), which is another widely used proper scoring rule, defined as:

$$CRPS(y_i, \mu, \sigma, F_\varepsilon) = \int_{-\infty}^{\infty} \left[F_\varepsilon\left(\frac{z - \mu}{\sigma}\right) - \mathbb{1}\left(\frac{y_i - \mu}{\sigma} \leq \frac{z - \mu}{\sigma}\right)\right]^2 dz. \quad (3)$$

Specifically, the following relation holds:

$$IHF(y_i, \mu, \sigma, F_\varepsilon) = \frac{1}{2} CRPS(y_i, \mu, \sigma, F_\varepsilon). \quad (4)$$

Pfanzagl (1969) and Birgé and Massart (1993) derive consistency results for parameter estimation under minimisation of proper scoring rule, such as IHF and CRPS, under the heading of minimum contrast estimators, see Gneiting and Raftery (2007) for a discussion.

In this paper, we consider Maximum likelihood-like estimators (M-Estimators), introduced by Huber (1964) as minimisers of some criterion function, for the location parameter of model (1), under the IHF loss function. Robust inference based on scoring rules as M-estimators has been recently considered also in Dawid et al. (2016), Kanamori and Fujisawa (2014, 2015), and Ovcharov (2015). In a related paper, Ferrari and La Vecchia (2012) introduce a robust estimator by maximising a surrogate likelihood function. As in our case, their estimator depends on a tuning parameter that balances robustness and efficiency and can be selected through a function of the contamination error.

Recently, Zou and Yuan (2008) introduced the composite quantile regression (CQR) for model selection, later extended to local polynomial regression by Kai et al. (2010), by minimizing a discretized version of the IHF loss function. The latter authors explicitly refer to the Hogg estimator in a remark where they quote Koenker (1984). In both cases, the CQR is not integrated, though Zou and Yuan (2008) point out a possible extension of their method in the direction of the general CQR (GCQR) which resembles the IHF considered in this paper. The GCQR is developed by Wu et al. (2019) who focus on the computational advantages with respect to CQR, and do not address the robustness issue.

Similarly to the aforementioned papers, we do not require any specific knowledge about the shape of $F_\varepsilon$, except that it belongs to $\mathcal{F}$. We show that the M-Estimator arising from minimisation of (2) or of (3) is robust, consistent, and asymptotically Normal. We also show that the sample median and the sample mean arise as limit cases for a tuning parameter $c$ which controls the rate of robustness of the estimator. Results show that, by picking any $\tilde{F}_\varepsilon \in \mathcal{F}$, with $\tilde{F}_\varepsilon$ possibly different from $F_\varepsilon$, the resulting estimator for $\mu$ is robust, consistent and asymptotically Normal. A link with the Hodges-Lehmann estimator and the Wilcoxon test is detailed in the case $F_\varepsilon = \tilde{F}_\varepsilon$ and $\sigma = c$. We do not require the existence of any moment and consider the Cauchy and Gaussian models for examples and discussion.

The paper is organized as follows. Section 2 details the estimator and establishes its consistency and asymptotic normality. Section 3 is concerned with efficiency and discusses three examples, for different choices of $F_\varepsilon$ and $\tilde{F}_\varepsilon$. Section 4 reports a Monte Carlo simulation study in order to assess the finite sample properties of the estimator. Conclusions are drawn in Section 5. The Appendix contains the proofs of the main results.

## 2. Robust location estimator

Let $\tilde{F}_\varepsilon$ be any cdf that belongs to $\mathcal{F}$, and let $c > 0$ be a tuning parameter. We consider the following estimator for the location of (1) according to a sample of $N$ observations $\mathbf{y} = (y_i, i = 1, \ldots, N)$:

$$T_N = \arg\min_{\mu} \quad N^{-1}\sum_{i=1}^{N} IHF(y_i, \mu, c, \tilde{F}_\varepsilon). \tag{5}$$

The estimator is equivalently defined by the implicit equation

$$N^{-1}\sum_{i=1}^{N} \psi(y_i, \mu, c, \tilde{F}_\varepsilon) = 0. \tag{6}$$

where

$$\psi(y_i, \mu, c, \tilde{F}_\varepsilon) = \tilde{F}_\varepsilon\left(\frac{y_i - \mu}{c}\right) - \frac{1}{2}$$

is equal to minus the derivative of $IHF(y_i, \mu, c, \tilde{F}_\varepsilon)$ with respect to $\mu$.[1] Note that, in the limit case $c \to 0$, equation (6) becomes

$$N^{-1}\sum_{i=1}^{N} \mathbb{1}(y_i \leq \mu) = \frac{1}{2},$$

for which $T_N$ is the empirical median. Hence, we expect that increasing $c$ will imply a deterioration in the robustness of $T_N$. In the following we provide rigorous arguments about this.

A way of looking at the robustness of $T_N$ is to evaluate its breakdown point. As discussed in Huber and Ronchetti (2009), the breakdown point is "*the smallest fraction of bad observations that may cause an estimator to take on arbitrarily large aberrant values*". In our case, $\psi$ is skew-symmetric, i.e. $\psi(-\infty, \mu, c, \tilde{F}_\varepsilon) = -\psi(\infty, \mu, c, \tilde{F}_\varepsilon) < \infty$, and the breakdown point of $T_N$ reaches its best possible values of $\frac{1}{2}$, irrespectively on the choice of $\tilde{F}_\varepsilon$ and $c$. This means that $T_N$ is robust against Lévy and Prohorov neighborhoods-type of contamination for $F_\varepsilon \in \mathcal{F}$, when contamination is up to 50%, see Huber and Ronchetti (2009).

Another way to look at robustness, is to consider the influence curve (IC) introduced by Hampel (1968, 1974) and defined as

$$IC(x; F_\varepsilon, T_N) = \frac{\psi(x, T_N, c, \tilde{F}_\varepsilon)}{\mathbb{E}_{F_\varepsilon}\left[\psi'\left(x, T_N, c, \tilde{F}_\varepsilon\right)\right]},$$

[1] Here and throughout the paper we will assume that technical conditions to exchange integral and derivatives hold.

where $\psi' = \partial\psi/\partial\mu$. The IC allows one to assess the relative influence of individual observations toward the value of $T_N$ and, in our case, it has the following form

$$IC(x; F_\varepsilon, T_N) = \frac{c\left(\tilde{F}_\varepsilon\left(\frac{x-T_N}{c}\right) - \frac{1}{2}\right)}{\mathbb{E}_{F_\varepsilon}\left[\tilde{f}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)\right]}. \tag{7}$$

We have already showed that for $c \to 0$, $T_N$ coincides with the sample median. On the other hand, to investigate the properties of the estimator when $c \to \infty$, we analyse the behaviour of $IC(x; F_\varepsilon, T_N)$ for large values of $c$. By doing this, we first note that when $c$ is large, $\tilde{F}_\varepsilon\left(\frac{x-T_N}{c}\right) - \frac{1}{2} \approx \frac{1}{2} + \text{sgn}(x - T_N)\epsilon$, and $\mathbb{E}_{F_\varepsilon}\left[\tilde{f}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)\right] \approx \tilde{f}_\varepsilon(0)$, such that:

$$IC(x; F_\varepsilon, T_N) \approx d(c), \tag{8}$$

where $|d(c)| = c(|\epsilon|/\tilde{f}_\varepsilon(0))$ and $|\epsilon|$ small, for an arbitrarily large set of values $x$ around $T_N$. From (8) we recognise that the behaviour of $IC$ is that of the influence function of the (non robust) empirical mean estimator which assigns equal weight to all observations.

### *2.1. Asymptotic results*

We now establish consistency and asymptotic Normality of the estimator (5). Proofs are reported in an online supplementary material file, see Appendix.

**Theorem 1.** *(Consistency) Let $\tilde{F}_\varepsilon \in \mathcal{F}$, and let $c > 0$ be a tuning parameter. Then $T_N \to \mu$ in probability and almost surely for $N \to \infty$.*

**Theorem 2.** *(Asymptotic Normality) Under the assumptions of Theorem 1,*

$$\sqrt{N}(T_N - \mu) \to_d \mathcal{N}(0, v^2),$$

*where*

$$v^2 = \frac{c^2\left[\mathbb{E}_{F_\varepsilon}\left[\tilde{F}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)^2\right] - \frac{1}{4}\right]}{\mathbb{E}_{F_\varepsilon}\left[\tilde{f}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)\right]^2} \tag{9}$$

Theorems 1 and 2 indicate that estimation and inference on $\mu$ can be carried out based on any choice of $\tilde{F}_\varepsilon$. Consistency does not depend on $\sigma$ or $c$, while the asymptotic variance of $T_N$ does.

### 2.2. Robustness measures

The gross error sensitivity of $T_N$ at $F_\varepsilon$ measures the (normalized) infinitesimal effect of an outlier in linear approximation and is given by

$$\gamma^* = \sup_{x\in\Re} |IC(x; F_\varepsilon, T_N)| = \frac{c}{2\mathbb{E}_{F_\varepsilon}\left[\tilde{f}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)\right]},$$

which is bounded for finite values of $c$, and implies bias robustness (B–robustness) of $T_N$. This measure alone does not give a complete picture of the robustness of $T_N$ and it is usually paired with the change-of-variance sensitivity $\kappa^*$, which measures the (normalized) infinitesimal effect of an outlier in linear approximation to the variance of $T_N$. We first introduce the change of variance function (CVF) as in Hampel et al. (2005, Ch. 2.5)

$$CVF(x, \psi, F_\varepsilon) = v^2\left[1 + \frac{\left(\tilde{F}_\varepsilon\left(\frac{x}{c}\right) - \frac{1}{2}\right)^2}{\mathbb{E}_{F_\varepsilon}\left[\tilde{F}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)^2\right] - \frac{1}{4}} - \frac{2\tilde{f}_\varepsilon\left(\frac{x}{c}\right)}{\mathbb{E}_{\tilde{F}_\varepsilon}\left[\tilde{f}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)\right]}\right],$$

and then the sensitivity as

$$\kappa^* = \sup_{x\in\Re}\frac{CVF(x, \psi, F_\varepsilon)}{v^2} = \frac{4\mathbb{E}_{F_\varepsilon}\left[\tilde{F}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)^2\right]}{4\mathbb{E}_{F_\varepsilon}\left[\tilde{F}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right)^2\right] - 1}.$$

We note that $\kappa^*$ is increasing in $c$ and reaches its minimum at 2 for $c \to 0$, because $\lim_{c\to 0}\tilde{F}_\varepsilon\left(\varepsilon\frac{\sigma}{c}\right) = \mathbb{1}(\varepsilon > 0)$, which is the change-of-variance sensitivity of the median estimator. For finite values of $c$, $\kappa^*$ is finite which implies that $T_N$ is variance robust (V-robust).

### 2.3. Link with the Hodges-Lehmann estimator, the Wilcoxon test, and the ML logistic estimator

As noted in Huber and Ronchetti (2009), the rank estimate (R-Estimate) of location based on the Wilcoxon signed-rank test (Wilcoxon, 1945) leads to the Hodges-Lehmann estimator (Hodges Jr and Lehmann, 1963; Sen, 1963) for the location as the solution to:

$$\int_{-\infty}^{\infty} F_\varepsilon(2T_N - x) f_\varepsilon(x) dx = \frac{1}{2},$$

where we have assumed without loss of generality that $\mu = 0$ and $\sigma = 1$. Notably, in our case of symmetric $F_\varepsilon$, the influence function of this estimator is:

$$IC(x, F_\varepsilon, T_N) = \frac{F_\varepsilon(x) - \frac{1}{2}}{\mathbb{E}_{F_\varepsilon}[f_\varepsilon(x)]},$$

which is equivalent to our equation (7) in the case of $\tilde{F}_\varepsilon = F_\varepsilon$, and $c = \sigma$. Furthermore, we note that, if we set $\tilde{F}_\varepsilon(x) = \frac{1}{2} + \frac{1}{2}\tanh(x)$, i.e. the cdf of a logistic distribution, the M estimator (6) coincides with the maximization of logistic likelihood. If also $F_\varepsilon = \tilde{F}_\varepsilon$ and $\sigma = c$ we have correct specification and equivalence between the M estimator and logistic Maximum Likelihood estimator.

## 3. Efficiency

Efficiency depends on the true distribution function $F_\varepsilon$, the scale $\sigma$, the choices of $\tilde{F}_\varepsilon$ and $c$. Thus, explicit results cannot be obtained. Clearly, if $F_\varepsilon = \tilde{F}_\varepsilon$, the value of $c$ which maximizes efficiency will be the one that, if exists, ensures the equivalence between the Maximum Likelihood and the M-Estimator. For example, if $F_\varepsilon = \tilde{F}_\varepsilon = \Phi$, where $\Phi$ is the distribution function of a Gaussian random variable, then the limit case $c \to \infty$ (empirical mean) implies maximum efficiency. A second case is given by the logistic case $F_\varepsilon(x) = \tilde{F}_\varepsilon(x) = \frac{1}{2} + \frac{1}{2}\tanh(x)$, where $c = \sigma$ leads to maximum efficiency. However, these results seem to be somehow related only to the Gaussian and logistic cases. In other cases, especially when $F_\varepsilon \neq \tilde{F}_\varepsilon$, our results indicate that the asymptotic variance (9) is minimized for a positive and finite value of $c$. In the following, we discuss in detail some cases of particular interest.

### *3.1. Gaussian errors and Gaussian $\tilde{F}_\varepsilon$*

In the case when $F_\varepsilon = \tilde{F}_\varepsilon = \Phi$, the IC becomes

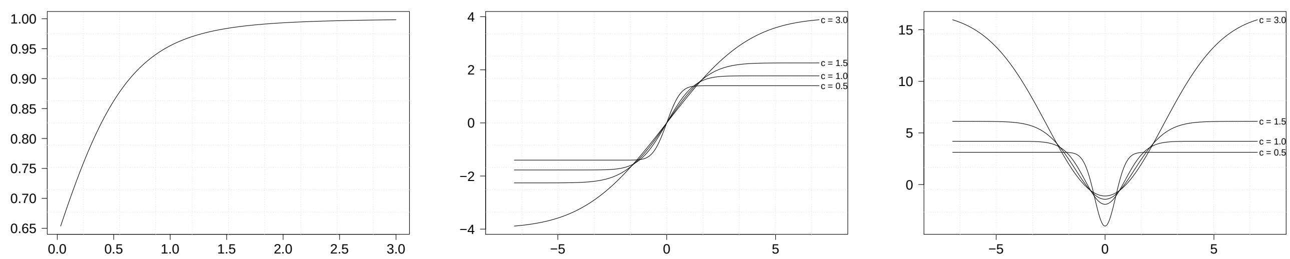
$$IC(x, \Phi, \mu) = \left[\Phi\left(\frac{x-\mu}{c}\right) - \frac{1}{2}\right]\sqrt{2\pi(\sigma^2 + c^2)}.$$

A closed form expression for the asymptotic variance of $\sqrt{N}T_N$ is not available if not when $c = \sigma$. In the latter case we have

$$v^2 = \sigma^2 \frac{\pi}{3},$$

such that the asymptotic relative efficiency (ARE) of the estimator is approximately $3/\pi \approx 95\%$. Note that, the median for a Gaussian distribution has an ARE of approximately $2/\pi \approx 64\%$. When $c \neq \sigma$ we have that $v^2$ is decreasing in $c$. To see this, we can compute the derivative of $v^2$ with respect to $c$ which is

$$\frac{\partial v^2}{\partial c} \propto -\int_{-\infty}^{\infty} z\phi\left(z\frac{c}{\sigma}\right) 2\phi(z)\Phi(z)dz,$$




(a) ARE (b) IC (c) CVF

Figure 1: Asymptotic relative efficiency with respect to $c$ (a), influence function (b), and change of variance function (c), in the case $F_\varepsilon = \tilde{F}_\varepsilon = \Phi$.

where $\phi$ indicates the density of a standard Normal random variable. We recognise that the integral is positive since it is equivalent to the expectation of a function which is symmetric about 0, namely $z\phi\left(z\frac{c}{\sigma}\right)$, with respect to a skew normal distribution with positive skewness coefficient equals to 1 and location equal to 0, see Azzalini (1985). Numerical analysis indicates that $v^2$ decreases to $\sigma^2$ as long as $c$ increases. This suggests that in the limit case when $c \to \infty$ we recover the variance of the non-robust ML estimator for the location of the Gaussian, i.e., the empirical mean. Figure 1 reports the ARE, IC, and CVF for this case.

### *3.2. Cauchy errors and Cauchy $\tilde{F}_\varepsilon$*

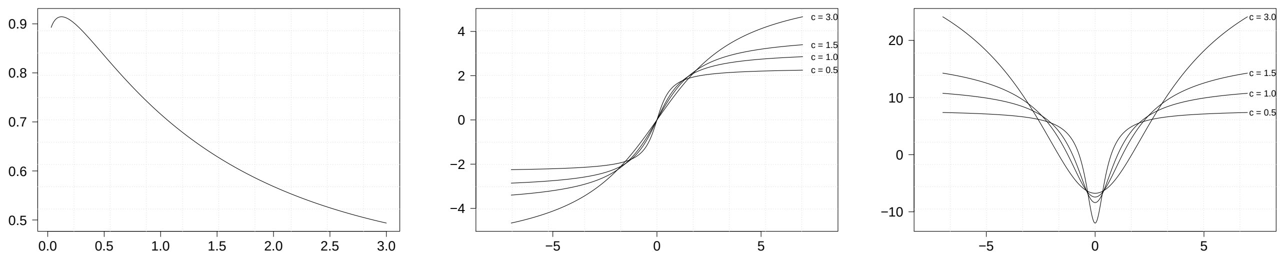
In the case of a Cauchy distribution, $F_\varepsilon(\varepsilon) = \tilde{F}_\varepsilon(\varepsilon) = \frac{1}{\pi}\arctan(\varepsilon) + \frac{1}{2}$ and the IC becomes

$$IC(x, F_\varepsilon, \mu) = \arctan\left(\frac{x-\mu}{c}\right)(\sigma + c),$$

the asymptotic variance becomes

$$v^2 = (\sigma + c)^2 \frac{c}{\sigma\pi} \int_{-\infty}^{\infty} \frac{\arctan(\varepsilon)^2}{\sigma^2 + \varepsilon^2 c^2} d\varepsilon,$$

which has a minimum at approximately $c = 0.12\sigma$ with an ARE of approximately 91.5%. If $c = \sigma$, then $v^2 = \frac{\sigma^2\pi^2}{3}$, which implies an ARE of $6/\pi^2 \approx 60\%$. Figure 2 reports the ARE, IC, and CVF for this case.




(a) ARE (b) IC (c) CVF

Figure 2: Asymptotic relative efficiency with respect to $c$ (a), influence function (b), and change of variance function (c), in the Cauchy case $F_\varepsilon(\varepsilon) = \tilde{F}_\varepsilon(\varepsilon) = \frac{1}{\pi}\arctan(\varepsilon) + \frac{1}{2}$.

*3.3. Cauchy Errors and Gaussian $\tilde{F}_\varepsilon$*

In the case of Cauchy errors and $\tilde{F}_\varepsilon = \Phi$ we obtain:

$$IC(x, \Phi, \mu) = \frac{c^2 \phi\left(\frac{\sigma}{c}\right)\left[\Phi\left(\frac{x-\mu}{c}\right) - \frac{1}{2}\right]}{(1 - \Phi(\frac{\sigma}{c}))},$$

the asymptotic variance becomes

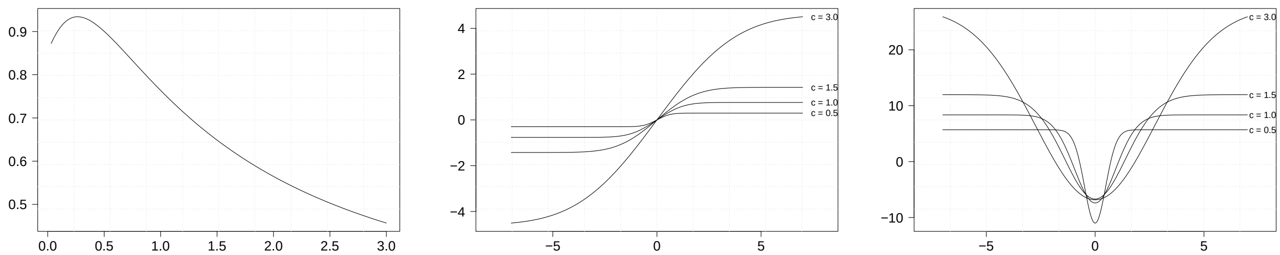
$$v^2 = \sigma c \sqrt{\frac{2}{\pi}} \int_0^1 \frac{x^2 \exp\{\frac{\Phi^{-1}(x)^2}{2}\}}{\sigma^2 + c^2 \Phi^{-1}(x)^2} dx,$$

which has a minimum for approximately $c = 0.25\sigma$ with an ARE of approximately 93%. If $c = \sigma$, the ARE decreases to approximately 76%. The ARE for the median estimator is approximately 80%. Figure 3 reports the ARE, IC, and CVF for this case.

## 4. Finite sample properties

We now investigate the finite sample properties of the M-estimator (5) with respect to different choices of $F_\varepsilon$, $\tilde{F}_\varepsilon$, and $c$. We consider the finite sample bias (BIAS) and the relative efficiency loss (RELoss) defined as one minus the ratio between the ARE and the efficiency for fixed $N$ which can be written as:

$$\text{RELoss}(N) = 1 - \frac{\text{Var}(T_N)}{v^2},$$




(a) ARE (b) IC (c) CVF

Figure 3: Asymptotic relative efficiency with respect to $c$ (a), influence function (b), and change of variance function (c), in the case $F_\varepsilon(\varepsilon) = \frac{1}{\pi}\arctan(\varepsilon) + \frac{1}{2}$ and $\tilde{F}_\varepsilon(\varepsilon) = \Phi(\varepsilon)$.

| | BIAS | | | | | | RELoss | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\Phi, \Phi\}$ | | $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\mathcal{C}, \mathcal{C}\}$ | | $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\mathcal{C}, \Phi\}$ | | $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\Phi, \Phi\}$ | | $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\mathcal{C}, \mathcal{C}\}$ | | $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\mathcal{C}, \Phi\}$ | |
| $c$ | $N = 100$ | $N = 500$ | $N = 100$ | $N = 500$ | $N = 100$ | $N = 500$ | $N = 100$ | $N = 500$ | $N = 100$ | $N = 500$ | $N = 100$ | $N = 500$ |
| 0.1 | 1E-03 | -4E-04 | 2E-04 | 2E-04 | -2E-03 | 5E-04 | -0.01 | 0.00 | -0.10 | -0.07 | -0.09 | -0.06 |
| 0.5 | -8E-04 | -3E-04 | -4E-03 | -4E-04 | -5E-04 | 1E-03 | 0.00 | 0.00 | -0.15 | -0.12 | -0.11 | -0.08 |
| 1.0 | 2E-03 | -1E-03 | -5E-04 | -1E-04 | 2E-03 | -1E-03 | 0.00 | 0.00 | -0.20 | -0.18 | -0.13 | -0.11 |
| 1.5 | -2E-03 | 7E-04 | -5E-03 | 2E-03 | 5E-03 | 8E-04 | -0.01 | 0.00 | -0.27 | -0.25 | -0.17 | -0.15 |
| 2.0 | -3E-04 | -2E-04 | -4E-03 | -8E-04 | 2E-03 | 3E-03 | 0.01 | 0.00 | -0.34 | -0.33 | -0.19 | -0.19 |
| 2.5 | -6E-04 | 9E-05 | -4E-03 | 3E-04 | -3E-04 | -2E-04 | 0.00 | 0.00 | -0.42 | -0.39 | -0.24 | -0.23 |
| 3.0 | 1E-03 | -6E-04 | -3E-03 | -1E-03 | 6E-03 | -4E-04 | 0.00 | 0.00 | -0.51 | -0.48 | -0.28 | -0.27 |

Table 1: Bias and relative efficiency loss (RELoss) for of the M-estimator (5) for different choices of $\{F_\varepsilon, \tilde{F}_\varepsilon\}$, $c$, and $N$.

where $\mathrm{Var}(T_N)$ indicates the variance of the estimator for fixed $N$. We consider three cases for the pair $\{F_\varepsilon, \tilde{F}_\varepsilon\}$: i) $\{\Phi, \Phi\}$, ii) $\{\mathcal{C}, \mathcal{C}\}$, and iii) $\{\mathcal{C}, \Phi\}$, where $\mathcal{C}$ represents the distribution function of a Cauchy random variable. The tuning parameter $c$ takes values $\{0.1, 0.5, 1.0, 1.5, \ldots, 3.0\}$, while the value of the scale and the true location are set to $\sigma = 1$ and $\mu = 0$, respectively. Note that, since we fix $\sigma$, results regarding $c$ should be interpreted as in percentage of $\sigma$. For example, $c = 0.1$ means that $c$ is the 10% of $\sigma$. We consider medium and large sample size, i.e. $N = 100$ and $N = 500$, respectively. The bias and the RELoss are computed after 5000 simulations for each triplet ($\{F_\varepsilon, \tilde{F}_\varepsilon\}$, $c$, $N$).

Table 1 summarizes the results. We see that the bias is generally very low and does not depend on the shape of the true distribution, neither on the choice of $\tilde{F}_\varepsilon$, $N$, and $c$. This suggests that finite sample bias is negligible. Results for efficiency depict a different story. We find that the efficiency loss can be severe in finite samples, when the data are generated from a fat tailed distribution. On the contrary, when the data are sampled from a Gaussian distribution, there is no efficiency loss, even in small samples. In the fat-tailed case, the results strongly depend on the choice of $c$. Specifically, when $c$ is large, the efficiency loss is large as well. This can be problematic given that efficiency can be quite low even asymptotically when $c$ is large, see Section 3. On the contrary, when $c$ is low, say in the interval (0.1, 0.5), the efficiency loss is confined in the range $6\% - 15\%$. Interestingly, we find that if the data are generated from a Cauchy distribution, then the use of a Gaussian cdf $\tilde{F}_\varepsilon$ helps in reducing efficiency losses.[2]

Overall, the results suggest to be careful with the selection of $c$ if efficiency is the main concern. Being conservative and setting $c$ equal to a small fraction of a (robust) estimate of $\sigma$ seems a sensible strategy. For instance, the median absolute deviations from the median (MAD) estimator provides good results.

## 5. Discussion and concluding remarks

We studied the properties of the M-estimator for location arising from minimization of the proper scoring rule $IHF/CRPS$ in the case of a location scale model with symmetric errors. We showed that the estimator introduced in the paper admits as limit cases the sample median and the sample mean, arising when $c \to 0$ and $c \to \infty$, respectively. Consistency and asymptotic normality of the estimator is established for any choice of $\tilde{F}_\varepsilon \in \mathcal{F}$, where $\mathcal{F}$ is the set of all continuous symmetric distribution function centered at zero. Notably, $\tilde{F}_\varepsilon$ can be different from the true distribution function $F_\varepsilon$. A link with the Hodges-Lehmann estimator and the logistic Maximum Likelihood estimator is discussed.

Efficiency of the estimator has been studied in detail. Results indicated that high efficiency levels can be reached asymptotically by an appropriate choice of $c$. However, in the case of Gaussian $F_\varepsilon$ and Gaussian $\tilde{F}_\varepsilon$, this strategy leads to the non robust empirical mean estimator. When the observations come from a fat tailed distribution, we found that setting $c$ to a small fraction of a (robust) estimate of the scale $\sigma$, such as the MAD, is a reasonable choice. A finite sample analysis indicates that a conservative choice of $c$ (like $c = 0.1\sigma$ or $c = 0.5\sigma$), is key to preserve acceptable efficiency levels. As a rule for selecting $c$, the minimum mean square error criterion of Ferrari and La Vecchia (2012) can be possibly adopted. Results also indicated that finite sample bias is not a cause of concern.

[2]Note that ARE is very similar in the cases $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\mathcal{C}, \mathcal{C}\}$ and $\{F_\varepsilon, \tilde{F}_\varepsilon\} = \{\mathcal{C}, \Phi\}$.

## Appendix

The proof of Theorems 1 and 2 can be found online in a supplementary material file, see Appendix A.


## Acknowledgements

We would like to thank Elvezio Ronchetti, the Editor, and Associate Editor and a Referee for their comments.


## References


Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

Birgé, L. and Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields*, 97(1-2):113–150.

Brown, T. A. (1974). Amissible scoring systems for continuous distributions. *Rand*, pages 1–28.

Cervera, J. and Munoz, J. (1996). Proper scoring rules for fractiles. *Bayesian statistics*, 5:513–519.

Dawid, A. P., Musio, M., and Ventura, L. (2016). Minimum scoring rule inference. *Scandinavian Journal of Statistics*, 43:123–138.

Ferrari, D. and La Vecchia, D. (2012). On robust estimation via pseudo-additive information. *Biometrika*, 99(1):238–244.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422.

Hampel, F. R. (1968). Contribution to the theory of robust estimation. *Ph. D. Thesis, University of California, Berkeley.*

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (2005). *Robust statistics: the approach based on influence functions*, volume 196. John Wiley & Sons.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570.

Hodges Jr, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, pages 598–611.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust statistics*. 2 edition.

Kai, B., Li, R., and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society Series B*, pages 49–69.

Kanamori, T. and Fujisawa, H. (2014). Affine invariant divergences associated with proper composite scoring rules and their applications. *Bernoulli*, 20(4):2278?2304.

Kanamori, T. and Fujisawa, H. (2015). Robust estimation under heavy contamination using unnormalized models. *Biometrika*, 102(3):559–572.

Koenker, R. (1984). A note on l-estimates for linear models. *Statistics & probability letters*, 2(6):323–325.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press, New York.

Laio, F. and Tamea, S. (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences*, 11(4):1267–1277.

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management science*, 22(10):1087–1096.

Ovcharov, E. Y. (2015). Proper scoring rules and bregman divergence. *Bernoulli*, 24(1):53?79.

Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika*, 14(1):249–272.

Sen, P. K. (1963). On the estimation of relative potency in dilution (-direct) assays by distribution-free methods. *Biometrics*, pages 532–552.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

Wu, Y., Tian, M., and Tang, M.-L. (2019). General composite quantile regression: Theory and methode. *Communications in Statistics: Theory and Methods*, pages 1–20.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, pages 1108–1126.