

A Smoothed Three-Part Redescending M-Estimator

Alistair J. Martin  and Brenton R. Clarke * 

School of Mathematics, Statistics, Chemistry and Physics, Murdoch University, Murdoch, WA 6150, Australia

* Correspondence: b.clarke@murdoch.edu.au

Abstract: A smoothed M-estimator is derived from Hampel's three-part redescending estimator for location and scale. The estimator is shown to be weakly continuous and Fréchet differentiable in the neighbourhood of the normal distribution. Asymptotic assessment is conducted at asymmetric contaminating distributions, where smoothing is shown to improve variance and change-of-variance sensitivity. Other robust metrics compared are largely unchanged, and therefore, the smoothed functions represent an improvement for asymmetric contamination near the rejection point with little downside.

Keywords: robust estimation; asymptotic variance; weak continuity; gross error sensitivity; change-of-variance sensitivity

1. Introduction

In this paper, we describe a refinement of the Hampel three-part redescenders [1] that smooths sharp corners in the spirit of [2,3]. In doing so, all the asymptotic theories of weak continuity and Fréchet differentiability apply to the introduced M-estimators. The central parametric family studied here is assumed to be

$$\mathcal{F} = \left\{ F_{(\mu, \sigma)^\top}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) : -\infty < \mu < \infty, \sigma > 0 \right\},$$

where $\Phi(x)$ is the standard cumulative normal distribution on the real line. Without loss of generality estimators including M-estimators described in [4–6], it can be studied at the standard normal distribution where $(\mu, \sigma)^\top = (0, 1)^\top$. This paper differs from many in that the asymptotics and the performance of estimators are discussed in a neighbourhood of underlying distributions. The main distribution is standard normal, with a contamination parameterised by μ_2 and α , as follows:

$$G_0(x) = (1 - \epsilon)\Phi(x) + \epsilon\Phi\left(\frac{x - \mu_2}{\alpha}\right), \quad (1)$$

where $0 < \epsilon < 1/2$ is considered small, $\mu_2 \in \mathbb{R}$ (the real line), and $\alpha \in \mathbb{R}^+$. The distribution is symmetric when $\mu_2 = 0$ and asymmetric otherwise. Contamination processes themselves may be subject to the central limit theorem, whence the mixture model is apt. Moreover, regarding alternative heavy-tailed contamination, this may be represented by choosing μ_2 to be arbitrarily large. Since there is no restriction placed on μ_2 and α , very large outliers can be represented using the mixture.

In a sense [1], following the publication of [7], contaminated populations with the description of a rejection point are predicated, beyond which observations are to have zero



Academic Editor: Wei Zhu

Received: 22 March 2025

Revised: 25 April 2025

Accepted: 26 April 2025

Published: 30 April 2025

Citation: Martin, A.J.; Clarke, B.R. A Smoothed Three-Part Redescending M-Estimator. *Stats* **2025**, *8*, 33. <https://doi.org/10.3390/stats8020033>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

weight on the estimation procedure. This came accompanying the introduction of the first ψ -function in robust estimation to continuously redescend to zero, as follows:

$$\psi_{1;a,b,c}(x) = \begin{cases} x & \text{for } 0 \leq |x| \leq a \\ a \operatorname{sign}(x) & \text{for } a \leq |x| \leq b \\ a \frac{c-|x|}{c-b} \operatorname{sign}(x) & \text{for } b \leq |x| \leq c \\ 0 & \text{for } c \leq |x| \end{cases}, \quad (2)$$

where $0 < a \leq b \leq c$. The tuning constant a defines a region about which the estimator is equivalent to the maximum likelihood estimate at $\mathcal{F} = \Phi$, $\psi(x) = x$. The value of b can be used to “Winsorise” the observations between a and b standard deviations from the true location. The rejection point is set by the tuning constant c . Figure 1 plots $\psi_{1;a,b,c}$. Setting $b = c = \infty$ yields Huber’s ψ -function. See [8].

The study of the Hampel three-part redescender considered estimating location only and made use of an auxiliary estimator of scale. The work in [1,8–10] addresses the question of the scale of the central distribution by suggesting using either a robust estimator of scale such as the median of the absolute deviations (MAD) estimator or its consistent version, that is,

$$\text{MADN} = \text{median}\{|X_i - \text{median}(X_1, X_2, \dots, X_n)|\} / 0.6745,$$

where the value 0.6745 is to make the scale estimator consistent for the scale when the underlying distribution is indeed normal. See [11] for a recent synthesis of MADN. The estimator of location is then solved via a scaled location equation. Given the sample $X_1, X_2, \dots, X_n$ from a hypothesised distribution in $\mathcal{F}$ and an estimate of the standard deviation σ , say, $\hat{\sigma}$, the estimating equation for location is then of the form

$$\sum_{i=1}^n \psi_{1;a,b,c} \left(\frac{X_i - \mu}{\hat{\sigma}} \right) = 0.$$

Alternatively, the estimator of both location and scale can be solved through the use of the solution of the simultaneous equations

$$\mathbf{K}_n(\tau_1, \tau_2) = \frac{1}{n} \sum_{i=1}^n \mathbf{\Psi} \left(\frac{X_i - \tau_1}{\tau_2} \right) = \int \mathbf{\Psi} \left(\frac{x - \tau_1}{\tau_2} \right) dF_n(x) = \mathbf{0}. \quad (3)$$

Here, $F_n(x)$ is the empirical distribution of the sample, and $\mathbf{\Psi} = (\psi_1, \psi_2)^\top$ is a vector function used in the estimation for location and scale. In this work, we propose estimators for both location and scale and, as such, focus on joint estimation. A three-part redescending M-estimator of scale may be constructed in the following form:

$$\psi_{2;a,b,c}(x) = \begin{cases} x^2 - 1 - P & \text{for } 0 \leq |x| \leq a \\ a^2 - 1 - P & \text{for } a \leq |x| \leq b \\ (a^2 - 1 - P) \frac{c-|x|}{c-b} & \text{for } b \leq |x| \leq c \\ 0 & \text{for } c \leq |x| \end{cases}, \quad (4)$$

where P is defined implicitly from the equation $\int \psi_{2;a,b,c}(x) d\Phi(x) = 0$. See, for example, [12]. Setting $b = c = \infty$ yields Huber’s Proposal 2 estimator for scale, defined in [8]. Equation (4) is illustrated in Figure 1. Then to obtain the estimator, one solves (3) with $\mathbf{\Psi} = \mathbf{\Psi}_{a,b,c} = (\psi_{1;a,b,c}, \psi_{2;a,b,c})^\top$. Note that these functions are continuous but not smooth at a, b , and c .

Important robustness properties are a result of Fréchet differentiability, including uniform asymptotic consistency of the estimate's variability assessment based on the influence function. See [2]. Proofs of Fréchet differentiability in the neighbourhood of a central distribution are challenging for non-smooth ψ functions; hence, smoothing to achieve a proof of Fréchet differentiability is desirable.

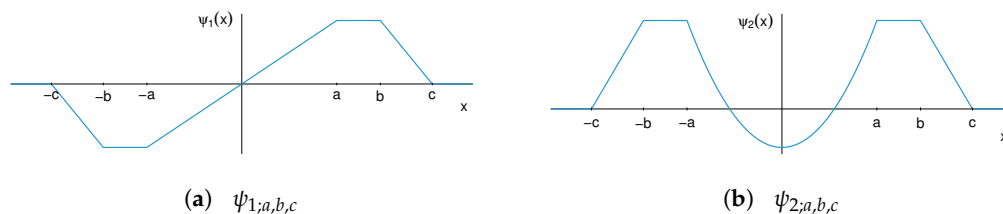


Figure 1. Plots of the three-part redescenders.

The remainder of this paper is structured as follows: In Section 2.1, we describe the smoothed M-estimator. Section 2.2 describes Fréchet differentiability and introduces theorems necessary to prove weak continuity and Fréchet differentiability of the introduced estimator. Additional robustness measures are introduced in Section 2.3, before comparing the smoothed and non-smoothed variants against these measures and considering their asymptotic behaviour at G_0 in Section 3.

2. Materials and Methods

2.1. A Smoothed Three-Part Redescender

The Hampel three-part redescending ψ -functions are continuous but not continuously differentiable at $x = i \in \{\pm a, \pm b, \pm c\}$. To rectify this, one may insert smoothing functions in the 2δ neighbourhood of each i , where $\delta \leq \min \frac{1}{2}\{a, b-a, c-b\}$. As the three-part redescender is defined in terms of piecewise polynomials with heuristic tuning constants, it is desirable to retain the same properties while attaining twice differentiable functions such that higher-order asymptotics could be explored in later works. Interpolation methods, such as that outlined by ([13] p. 319), may be extended to fifth-degree polynomials, $g_{ji}(x)$, with the following constraints to ensure the resultant ψ functions are twice differentiable:

$$\begin{aligned} g_{ji}(i-\delta) &= \psi_{j;a,b,c}(i-\delta) \\ g_{ji}(i+\delta) &= \psi_{j;a,b,c}(i+\delta) \\ g'_{ji}(i-\delta) &= \psi'_{j;a,b,c}(i-\delta) \\ g'_{ji}(i+\delta) &= \psi'_{j;a,b,c}(i+\delta) \\ g''_{ji}(i-\delta) &= \psi''_{j;a,b,c}(i-\delta) \\ g''_{ji}(i+\delta) &= \psi''_{j;a,b,c}(i+\delta) \end{aligned}$$

where $\psi_{j;a,b,c}$ for $j \in \{1, 2\}$ denotes Hampel's location and Clarke and Milne's scale functions for the three-part redescender with a, b , and c as given in [12]. From these boundary conditions, it is simple to formulate and solve a system of linear equations, as follows:

$$\mathbf{A}_{ji}\mathbf{x}_{ji} = \mathbf{b}_{ji} \quad (5)$$

to identify the polynomial coefficients $\mathbf{x}_{ji}$ for each $g_{ji}(x)$. Here, $\mathbf{A}_{ji}\mathbf{x}_{ji}$ represents the left side of the aforementioned constraints, and $\mathbf{b}_{ji}$ the right side. For example, the coefficients for

$$\begin{aligned}
g_{1a}(x) &= \alpha_{1,5}x^5 + \alpha_{1,4}x^4 + \alpha_{1,3}x^3 + \alpha_{1,2}x^2 + \alpha_{1,1}x + \alpha_{1,0} \\
g'_{1a}(x) &= 5\alpha_{1,5}x^4 + 4\alpha_{1,4}x^3 + 3\alpha_{1,3}x^2 + 2\alpha_{1,2}x + \alpha_{1,1} \\
g''_{1a}(x) &= 20\alpha_{1,5}x^3 + 12\alpha_{1,4}x^2 + 6\alpha_{1,3}x + 2\alpha_{1,2}
\end{aligned}$$

are the solution of (5), where

$$\mathbf{A}_{1a} = \begin{bmatrix} (a-\delta)^5 & (a-\delta)^4 & (a-\delta)^3 & (a-\delta)^2 & (a-\delta) & 1 \\ (a+\delta)^5 & (a+\delta)^4 & (a+\delta)^3 & (a+\delta)^2 & (a+\delta) & 1 \\ 5(a-\delta)^4 & 4(a-\delta)^3 & 3(a-\delta)^2 & 2(a-\delta) & 1 & 0 \\ 5(a+\delta)^4 & 4(a+\delta)^3 & 3(a+\delta)^2 & 2(a+\delta) & 1 & 0 \\ 20(a-\delta)^3 & 12(a-\delta)^2 & 6(a-\delta) & 2 & 0 & 0 \\ 20(a+\delta)^3 & 12(a+\delta)^2 & 6(a+\delta) & 2 & 0 & 0 \end{bmatrix},$$

$$\mathbf{x}_{1a} = [\alpha_{1,5} \quad \alpha_{1,4} \quad \alpha_{1,3} \quad \alpha_{1,2} \quad \alpha_{1,1} \quad \alpha_{1,0}]^T,$$

and

$$\mathbf{b}_{1a} = \begin{bmatrix} \psi_{1;a,b,c}(a-\delta) \\ \psi_{1;a,b,c}(a+\delta) \\ \psi'_{1;a,b,c}(a-\delta) \\ \psi'_{1;a,b,c}(a+\delta) \\ \psi''_{1;a,b,c}(a-\delta) \\ \psi''_{1;a,b,c}(a+\delta) \end{bmatrix} = \begin{bmatrix} (a-\delta) \\ a \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

2.1.1. Smoothed Three-Part Redescender for Location

Inserting polynomial smoothing functions in the 2δ regions around a , b , and c yields the following ψ function for location:

$$\psi_{1;a,b,c,\delta} = \begin{cases} x & |x| \leq (a-\delta) \\ g_{1a}(x) & (a-\delta) < |x| \leq (a+\delta) \\ a \operatorname{sign}(x) & (a+\delta) \leq |x| \leq (b-\delta) \\ g_{1b}(x) & (b-\delta) < |x| \leq (b+\delta) \\ a \frac{c-|x|}{c-b} \operatorname{sign}(x) & (b+\delta) \leq |x| \leq (c-\delta) \\ g_{1c}(x) & (c-\delta) < |x| \leq (c+\delta) \\ 0 & \text{otherwise} \end{cases}$$

where

$$\begin{aligned}
g_{1a}(x) &= (\alpha_{1,4}x^4 + \alpha_{1,2}x^2 + \alpha_{1,0}) \operatorname{sign}(x) + \alpha_{1,3}x^3 + \alpha_{1,1}x \\
g_{1b}(x) &= (\beta_{1,4}x^4 + \beta_{1,2}x^2 + \beta_{1,0}) \operatorname{sign}(x) + \beta_{1,3}x^3 + \beta_{1,1}x \\
g_{1c}(x) &= (\gamma_{1,4}x^4 + \gamma_{1,2}x^2 + \gamma_{1,0}) \operatorname{sign}(x) + \gamma_{1,3}x^3 + \gamma_{1,1}x
\end{aligned}$$

and

$$\alpha_{1,4} = \zeta$$

$$\alpha_{1,3} = -4a\zeta$$

$$\alpha_{1,2} = 6(a^2 - \delta^2)\zeta$$

$$\alpha_{1,1} = -4(a^3 - 3a\delta^2 - 2\delta^3)\zeta$$

$$\alpha_{1,0} = (a^4 - 6a^2\delta^2 + 8a\delta^3 - 3\delta^4)\zeta$$

$$\beta_{1,4} = -a\zeta$$

$$\beta_{1,3} = 4ab\zeta$$

$$\beta_{1,2} = -6a(b^2 - \delta^2)\zeta$$

$$\beta_{1,1} = 4a(b^3 - 3b\delta^2 + 2\delta^3)\zeta$$

$$\beta_{1,0} = -a(b^4 - 6b^2\delta^2 - 8b\delta^3 - 3\delta^4 + 16c\delta^3)\zeta$$

$$\gamma_{1,4} = a\zeta$$

$$\gamma_{1,3} = -4ac\zeta$$

$$\gamma_{1,2} = 6a(c^2 - \delta^2)\zeta$$

$$\gamma_{1,1} = -4a(c^3 - 3c\delta^2 - 2\delta^3)\zeta$$

$$\gamma_{1,0} = a(c^4 - 6c^2\delta^2 - 8c\delta^3 - 3\delta^4)\zeta$$

for $\zeta = (16\delta^3)^{-1}$, and $\zeta = (16\delta^3(b-c))^{-1}$. Figure 2 plots the smoothed three-part redescender for location. In this general form, the functional may appear to suffer from too many parameters; however, the coefficients $\alpha_{1,\cdot}, \beta_{1,\cdot}, \gamma_{1,\cdot}$ are wholly dependent upon a, b, c , and δ . The resultant ψ -function is piecewise polynomial.

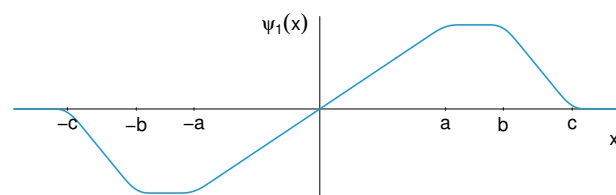


Figure 2. Plot of the smoothed $\psi_{1;a,b,c,\delta}$ redescender for location.

2.1.2. Smoothed Three-Part Redescender for Scale

The smoothing procedure for the scale ψ function yields the following definition:

$$\psi_{2;a,b,c,\delta} = \begin{cases} x^2 - 1 - P & |x| \leq (a - \delta) \\ g_{2a}(x) & (a - \delta) \leq |x| \leq (a + \delta) \\ \eta & (a + \delta) \leq |x| \leq (b - \delta) \\ g_{2b}(x) & (b - \delta) < |x| \leq (b + \delta) \\ \eta \frac{c - |x|}{c - b} & (b + \delta) \leq |x| \leq (c - \delta) \\ g_{2c}(x) & (c - \delta) \leq |x| \leq (c + \delta) \\ 0 & \text{otherwise} \end{cases}$$

where

$$\begin{aligned} g_{2a}(x) &= (\alpha_{2,5}x^5 + \alpha_{2,3}x^3 + \alpha_{2,1}x) \operatorname{sign}(x) \\ &\quad + \alpha_{2,4}x^4 + \alpha_{2,2}x^2 + \alpha_{2,0} \\ g_{2b}(x) &= \beta_{2,4}x^4 + \beta_{2,2}x^2 + (\beta_{2,3}x^3 + \beta_{2,1}x) \operatorname{sign}(x) + \beta_{2,0} \\ g_{2c}(x) &= \gamma_{2,4}x^4 + \gamma_{2,2}x^2 + (\gamma_{2,3}x^3 + \gamma_{2,1}x) \operatorname{sign}(x) + \gamma_{2,0} \end{aligned}$$

and

$$\begin{aligned} \alpha_{2,5} &= \zeta \\ \alpha_{2,4} &= -3a\zeta \\ \alpha_{2,3} &= 2(a^2 - 3\delta^2)\zeta \\ \alpha_{2,2} &= 2(a^3 + 3a\delta^2 + 4\delta^3)\zeta \\ \alpha_{2,1} &= -3(a^4 - 2a^2\delta^2 + \delta^4)\zeta \\ \alpha_{2,0} &= -(16P\delta^3 + 3a\delta^4 - a^5 + 16\delta^3 - 8a^2\delta^3 + 6a^3\delta^2)\zeta \\ \beta_{2,4} &= -\eta\zeta \\ \beta_{2,3} &= 4b\eta\zeta \\ \beta_{2,2} &= -6(b^2 - \delta^2)\eta\zeta \\ \beta_{2,1} &= 4(b^3 + 2\delta^3 - 3b\delta^2)\eta\zeta \\ \beta_{2,0} &= (6b^2\delta^2 + 8b\delta^3 - 16c\delta^3 + 3\delta^4 - b^4)\eta\zeta \\ \gamma_{2,4} &= \eta\zeta \\ \gamma_{2,3} &= -4c\eta\zeta \\ \gamma_{2,2} &= 6(c^2 - \delta^2)\eta\zeta \\ \gamma_{2,1} &= 4(2\delta^3 - c^3 + 3c\delta^2)\eta\zeta \\ \gamma_{2,0} &= -(3\delta^4 - c^4 + 6c^2\delta^2 + 8c\delta^3)\eta\zeta \end{aligned}$$

for $\zeta = (16\delta^3)^{-1}$ and $\xi = (16\delta^3(b-c))^{-1}$, as before, and $\eta = a^2 - 1 - P$. Once again, the coefficients $\alpha_{2,\cdot}, \beta_{2,\cdot}, \gamma_{2,\cdot}$ are not free variables; they are dependent upon a, b, c , and δ .

As for the non-smoothed three-part redescender, P may be found as the root of the implicit equation $\int \psi_2(x)\phi(x)dx = 0$, or can be expanded to an explicit form in terms of the standard normal distribution pdf and cdf (as shown in the Supplementary Materials). Figure 3 plots the above smoothed function for scale.

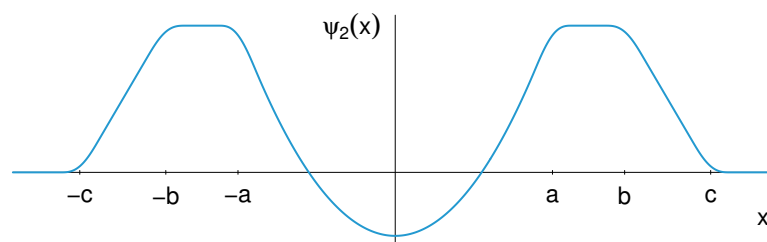


Figure 3. Plot of the smoothed $\psi_{2;a,b,c,\delta}$ redescender for scale.

2.2. The Influence Function and Fréchet Derivative

To compare robustness and the behaviour of estimators, one requires appropriate measures and tools. The estimator, which is a solution of Equation (3), is related here in its implicit functional form. Quoting from [6] (p. 29), where now $\mathcal{F}$ is any suitable parametric family, “A single functional root of Equation (3) may be written $\mathbf{T}[\Psi, F_n]$ where F_n is the empirical distribution function. More generally $\mathbf{T}[\Psi, G]$ can be defined as a functional root of equations

$$\mathbf{K}_G(\tau) = \int_{\mathbb{R}} \Psi(x, \tau) dG(x) = \mathbf{0}. \quad (6)$$

$\mathbf{T}[\Psi, G] = +\infty$ if no root exists. In the event of several roots of (6), a distance criterion ρ_o is employed to select the estimator from them. Classically, the roots correspond to extremes of the distance, but this is not assumed here. Specifically, the general functional $\mathbf{T}[\Psi, \rho_o, \cdot]$ is defined as the solution to

$$\inf_{\tau \in S(\Psi, G)} \rho_o(G, \tau) = \rho_o(G, \mathbf{T}[\Psi, \rho_o, G]), \quad (7)$$

where

$$S(\Psi, G) = \left\{ \tau \in \mathcal{T} \mid \int_{\mathbb{R}} \Psi(x, \tau) dG(x) = \mathbf{0} \right\}, \quad (8)$$

if a solution exists. Otherwise $\mathbf{T}[\Psi, \rho_o, G] = +\infty$. Conditions on both Ψ and ρ_o determine consistency, the very minimum requirement being Fisher consistency, which means that $\mathbf{T}[\Psi, \rho_o, F_{\tau}] = \tau$ for all $\tau \in \mathcal{T}$. That is if you have the whole population you must get the corresponding true parameter. This implicitly also assumes the parametric model for the population is identifiable. A parametric family $\mathcal{F}$ is identifiable if and only if $F_{\tau_1} = F_{\tau_2}$ implies $\tau_1 = \tau_2$ for all $\tau_1, \tau_2 \in \mathcal{T}$. That is, two different parameters cannot give the same distribution”.

An easily afforded selection functional ρ_o at the location and scale parametric model is

$$\rho_o(G, \tau) = \left\| \tau - \left[G^{-1} \left(\frac{1}{2} \right), \text{MADN}(G) \right]^{\top} \right\|^2 \quad (9)$$

Here, $\|\cdot\|$ is the Euclidean norm. The functional form of MADN and the influence function of MAD are discussed in [10] Section 5.2 and also [11,14].

First introduced by [1,15], the Influence Function (IF) is a heuristic tool that indicates the infinitesimal influence of contamination at a given point, $x = z_0$, on the asymptotic value $\mathbf{T}[\psi, G_0]$. In our discussion, $\tau_0 = \mathbf{T}[\Psi, \rho_o, G_0]$. For the estimating functional $\mathbf{T}$ at a distribution G_0 , the IF is defined as

$$\mathbf{IF}(z_0; G_0, \mathbf{T}) = \lim_{\varepsilon \rightarrow 0^+} \frac{\mathbf{T}[(1 - \varepsilon)G_0 + \varepsilon\Delta_{z_0}] - \mathbf{T}[G_0]}{\varepsilon} \quad (10)$$

where Δ_{z_0} puts mass 1 at the point z_0 . In [5,16], the authors write an equivalent definition, as follows:

$$\mathbf{IF}(z_0; G_0, \mathbf{T}) = -\mathbf{M}(\tau_0, G_0)^{-1} \Psi(z_0, \tau_0). \quad (11)$$

The matrix

$$\mathbf{M}(\tau_0, G_0) = \int_{\mathbb{R}} \left\{ \frac{\partial}{\partial \tau} \Psi(x, \tau) \right\} \Big|_{\tau=\tau_0} dG_0(x) \quad (12)$$

is assumed to be nonsingular. In this form, we see that the Influence Function is proportional to Ψ , and indeed, the Fréchet derivative in Theorem 2 to be expressed below is exactly

$$\mathbf{T}'_{G_0}(G - G_0) = \int_{\mathbb{R}} \mathbf{IF}(x; G_0, \mathbf{T}) d(G - G_0)(x).$$

See [17] for an earlier chronological account of the development of the Fréchet derivative in statistics.

So long as we have a compact set D that excludes the solution for scale $\tau_2 = 0$ in which we search for the roots, the conditions set out below are easy to check and are sufficient to provide weak continuity and Fréchet differentiability.

Conditions $\mathcal{W}$

W_0 : $\mathbf{T}[\Psi, \rho_0, G_0] = \tau_0$.

W_1 : Ψ is a 2×1 vector function on $\mathbb{R} \times \Theta$ and has continuous partial derivatives on $\mathbb{R} \times D$, where $D \subset \Theta$ is some non-degenerate compact interval containing $\tau_0 = (\tau_{01}, \tau_{02})^\top$ in its interior and for which the scale τ_{02} is bounded away from zero.

W_2 : $\{\Psi(x, \tau) | \tau \in D\}$, $\left\{\frac{\partial}{\partial \tau} \Psi(x, \tau) | \tau \in D\right\}$ are bounded above in Euclidean norm $\|\mathbf{A}\| = \{\text{trace}(\mathbf{A}^\top \mathbf{A})\}^{1/2}$ by a constant.

W_3 : The matrix $\mathbf{M}(\tau_0, G_0)$, given by (12), is nonsingular.

The innovation in the conditions made here is to describe the Conditions at G_0 , rather than an assumed F_τ . Note that we discuss the behaviour of the estimating functional at a solution $\mathbf{T}[\Psi, \rho_0, G_0] = \tau_0$ different from $\mathbf{T}[\Psi, \rho_0, F_\tau] = \tau$. The restriction to a compact set D is for convenience rather than necessity, since it is easier to check the conditions. Since G_0 is chosen to be from the mixture distribution (1) which has for any given G_0 a bounded density and is an absolutely continuous distribution, the Conditions $\mathcal{W}$ imply Conditions $\mathcal{A}$ analogous to those defined in [6] (pp. 31,32) for the neighbourhoods generated by any one of the metric distances, these being Kolmogorov (also known as the Kolmogorov–Smirnov), Lévy, or Prohorov metrics. For example, the Kolmogorov–Smirnov metric or supremum norm between the two distributions F and G defined on the real line is $d_{KS}(F, G) = \sup_x |F(x) - G(x)|$. At an absolutely continuous distribution G_0 with a density that is bounded by a constant $C > 0$, it is noted, compare [6] problem 2.3, as follows:

$$d_{KS}(G, G_0) \leq (C + 1)d_L(G, G_0) \leq (C + 1)d_P(G, G_0),$$

where d_L is the Lévy metric and d_P is the Prohorov metric. The Prohorov metric governs the weak topology, as discussed in [18], and it is known that the empirical distribution generated by G_0 converges weakly to G_0 on a set of probability one. Hence, for a weakly continuous estimator, that is, an estimator that is continuous with respect to the Prohorov metric, we have automatic consistency, as explained further below. Moreover, the smoothed M-estimators discussed in this paper will be robust at G_0 in the sense of Theorem 2 of [18], as demonstrated in Theorem 2.5 and Corollary 2.3 in [6].

However, for Ψ choices with continuous partial derivatives, as is the case of $\Psi = (\psi_{1;a,b,c,\delta}, \psi_{2;a,b,c,\delta})^\top$, and noting that the functions are of total bounded variation, we can claim the existence of a weakly continuous and Fréchet differentiable root as a corollary to [6] [Theorem 2.3, Theorem 2.6] and [6] [Theorem 2.5] using the auxiliary selection functional $\rho(G, \tau) = |\tau - \tau_0|$.

Theorem 1 (Adapting Theorem 2.5 of [6]). *Assume Conditions $\mathcal{W}$. Then for ϵ small and non-negative, there exists a Prohorov neighbourhood $U(G_0)$ of G_0 such that the functional defined via $\mathbf{T}[\Psi, \rho, \cdot]$ is weakly continuous at each $G \in U(G_0)$.*

See Appendix A for the proof of Theorem 1. In addition, we can write

Theorem 2 (Adapting Theorem 2.6 of [6]). Assume Conditions $\mathcal{W}$ and assume for all $G \in \mathcal{G}$, the space of distributions on the real line, that

$$\int_{\mathbb{R}} \Psi(x, \tau_0) d(G - G_0)(x) = O(d_{KS}(G, G_0)) \quad (13)$$

as $d_{KS}(G, G_0) \rightarrow 0$. Then

$$\left| \mathbf{T}[\Psi, \rho, G] - \mathbf{T}[\Psi, \rho, G_0] - \mathbf{T}'_{G_0}(G - G_0) \right| = o(d_{KS}(G, G_0)),$$

where

$$\mathbf{T}'_{G_0}(G - G_0) = -M(\tau_0, G_0)^{-1} \int_{\mathbb{R}} \Psi(x, \tau_0) d(G - G_0)(x)$$

That is, $\mathbf{T}[\Psi, \rho, \cdot]$ is Fréchet differentiable at G_0 with respect to the Kolmogorov–Smirnov metric.

See the proof of Theorem 2 in Appendix B.

As argued in [10] (pp. 38–39) and [6], Corollary 2.5, with F replaced by G_0 , it is possible to state that there exists a unique consistent root of equations $\mathbf{T}[\Psi, \rho, F_n]$ to $\mathbf{T}[\Psi, \rho, G_0]$ for which the asymptotic distribution

$$\sqrt{n}(\mathbf{T}[\Psi, \rho, F_n] - \mathbf{T}[\Psi, \rho, G_0])$$

is multivariate normal with mean zero and asymptotic variance $\Sigma(\mathbf{T}, G_0)$, which is given below in (14).

For example, the estimator is Fréchet differentiable at G_0 with respect to the supremum norm, essentially using the Kolmogorov–Smirnov metric distance to measure the distance between distributions. This can therefore give the asymptotic distribution of $\sqrt{n}(\mathbf{T}[\Psi, F_n] - \mathbf{T}[\Psi, G_0])$ as multivariate normal with a mean zero and variance–covariance matrix, given by $\Sigma(\mathbf{T}, G_0)$, which is

$$\mathbf{M}(\tau_0, G_0)^{-1} \int_{\mathbb{R}} \Psi(x, \tau_0) \Psi(x, \tau_0)^\top dG_0(x) \{ \mathbf{M}(\tau_0, G_0)^{-1} \}^\top, \quad (14)$$

where we interpret $\Psi(x, \tau)$ to be $\Psi\left(\frac{x - \tau_1}{\tau_2}\right)$ when $(\tau_1, \tau_2)^\top = \tau$. See [6] (p. 46), Corollary 2.4 and Corollary 2.5.

When the functions Ψ have sharp corners, as for the three-part redescender, the weak continuity and asymptotic robustness arguments are only at G_0 and not at all distributions in a neighbourhood of G_0 . Essentially, only Theorem 1 and not Theorem 2 of [18] can be established. Asymptotic normality at G_0 does follow if one uses Conditions $\mathcal{A}'$ in [19], again for ϵ small and again replacing the parametric distribution by G_0 and noting that the aforesaid Ψ functions are Lipschitz.

On the other hand, with the suggested smoothed three-part redescending ψ -functions described in Section 2.1, all the asymptotic theory, including both Theorem 1 and Theorem 2 of [18], is covered as Conditions $\mathcal{W}$ are satisfied.

It remains to show that, for small ϵ , the auxiliary selection functional ρ can be replaced by the practical selection functional ρ_0 in the above results. It can be noted that, in every open neighbourhood N of $(\mu, \sigma)^\top$ for $F_{(\mu, \sigma)^\top} = \Phi\left(\frac{x - \mu}{\sigma}\right)$,

$$\inf_{\tau \notin N} \rho_0\left(F_{(\mu, \sigma)^\top}, \tau\right) - \rho_0\left(F_{(\mu, \sigma)^\top}, (\mu, \sigma)^\top\right) > 0 \quad (15)$$

and since both the median and MADN are weakly continuous at absolutely continuous distributions, then it follows that, for every $\eta > 0$, there exists an $\varepsilon > 0$ such that $G \in n(\varepsilon, F_{(\mu, \sigma)^\top})$ implies that $\rho_0(G, \tau)$ is continuous in $\tau \in \mathcal{T}$ and satisfies

$$\sup_{\tau \in \mathcal{T}} |\rho_0(G, \tau) - \rho_0(F_{(\mu, \sigma)^\top}, \tau)| < \eta. \quad (16)$$

Here, the neighbourhoods $n(\varepsilon, F_{(\mu, \sigma)^\top})$ are defined in [6] (p. 31) and can be defined by the metric distance between distributions or, for example, an epsilon contaminated neighbourhood, as we deal with here. Since (15) and (16) are the equivalent of [6] Equations (2.12) and (2.13), where, here, ρ_0 is defined in (9), it follows that this is a valid selection functional. Moreover, according to [10] both the Median and MADN estimators are weakly continuous at the normal and mixed normal distributions considered in this paper, since they are both absolutely continuous distributions. The consequent continuity of the function ρ_0 in both its arguments assures a weakly continuous functional $T[\Psi, \rho_0, \cdot]$. By [20], the empirical distribution function F_n generated by G_0 converges weakly to the distribution G_0 on a set of probability one, and a result of [21] gives that $d_p(F_n, G_0) \rightarrow 0$ almost surely. Thus, we have almost sure consistency in the sense that $T[\Psi, \rho_0, F_n] \rightarrow_{a.s.} T[\Psi, \rho_0, G_0]$.

2.3. Quantifiable Measures of Robustness

Gross-error sensitivity quantifies the worst possible influence an infinitesimal gross error may have on the estimator, and is defined for a one-dimensional estimate by

$$\gamma^*(T, G_0) = \sup_{x \in \mathbb{R}} |\text{IF}(x; G_0, T)|. \quad (17)$$

A bounded IF has a finite gross-error sensitivity, and the estimator is then said to be *bias robust*, or B-Robust [22]. There exists a trade-off between B-Robustness and efficiency. For example, the sample mean is the most efficient estimator at Φ but is not robust, as $\gamma^* = \infty$.

Several methods exist for multi-dimensional γ^* ; however, for simplicity, we consider $\gamma^*(T, \Phi)$ for location and scale estimation separately, as does [5] in the discussion of the tanh estimator.

Local-shift sensitivity measures local fluctuations, such as that which might arise from rounding, grouping, or observational inaccuracies, and compliments the global worst-case measure provided by gross-error sensitivity. It is defined by the standardised, worst-case slope, as follows:

$$\lambda^* = \sup_{x \neq y} \frac{|\text{IF}(y; G_0, T) - \text{IF}(x; G_0, T)|}{|y - x|}. \quad (18)$$

The **change of variance function** (CVF) introduced by [22,23] provides insight into the robustness of the estimator's asymptotic variance, providing complimentary assessment of asymptotic value robustness provided by the IF. For continuous ψ -functions at G_0 , under conditions outlined in [5] (pp. 125–126), the CVF is defined as

$$\text{CVF}(z_0; \psi, G_0) = \frac{\partial}{\partial \epsilon} \left[V\left(\psi, (1 - \epsilon)G_0 + \frac{\epsilon}{2}(\Delta_{z_0} + \Delta_{-z_0})\right) \right]_{\epsilon=0}, \quad (19)$$

where $V(\psi, G_0)$ is the asymptotic variance of the ψ estimator at G_0 . All estimators under consideration in this study adhere to the aforementioned conditions. A more complete introduction to the CVF may be found in [5], Chapter 2.5.

Similar to the IF, the CVF indicates the influence of contamination at z_0 upon the asymptotic variance of the estimator. Unlike the IF, where all large deviations from zero are

indicative of potential bias, positive and negative, CVF values are interpreted differently; negative CVF values indicate a decrease in asymptotic variance, so one is more concerned with positive CVF values.

The standardised **change of variance sensitivity** for continuous ψ -functions is defined by

$$\kappa^*(\psi, G_0) = \sup_{x \in \mathbb{R}} \frac{\text{CVF}(x; \psi, G_0)}{V(\psi, G_0)}. \quad (20)$$

In [5], the authors provide a more complete definition, including for the case of piecewise ψ -functions with jump discontinuities. A finite κ^* is said to be variance robust, or V-Robust.

3. Results

We compute asymptotic efficiency, γ^* , λ^* , and κ^* at Φ , and asymptotic variance (14) at asymmetric distributions G_0 for each estimator. $\mathbf{T}[\Psi, G_0]$ and $\Sigma(\mathbf{T}, G_0)$ for polynomial-defined ψ -functions can be computed more efficiently via simple-but-elaborate calculus, which we provide in the Supplementary Materials.

3.1. Smoothed and Non-Smoothed Estimator Comparison

In Table 1, the asymptotic variance of the smoothed and non-smoothed three-part estimator at Φ is tabulated with a , b , and c values as used in [12]. δ was chosen to be the highest value in $\frac{1}{2} \min\{a, b - a, c - b\}$. ν_{11} and ν_{22} represent the asymptotic variance, which are the diagonals of (14). While ν_{11} is by and large equivalent for the smoothed estimator, ν_{22} is consistently improved.

Table 2 lists measures of robustness for a smaller set of the three-part functional parameters. Estimator efficiency and local-shift sensitivity are seen to be equivalent for the smoothed and non-smoothed estimators. Smoothing greatly improves the change-of-variance sensitivity, at a much smaller expense of increased gross-error sensitivity. This improvement may be visualised by plots of the CVF, as shown in Figure 4 for location and scale, where smoothing blunts the sharp corners of the maxima. The smoothed CVFs for all parameter values in Table 2, for location and scale, are compared in Figure 5. Note that more efficient parameter choices yield higher change-of-variance peaks (which are lower than what would appear for the non-smoothed estimator).

Table 1. Asymptotic variance, rounded to four decimal places, at the standard normal distribution for location and scale, comparing the smoothed and non-smoothed three-part redescender, under the same parameters as [12]. Lower variance for equal a, b, c parameters is highlighted in bold.

a	b	c	Smoothed				Non-Smoothed		
			P	δ	ν_{11}	ν_{22}	P	ν_{11}	ν_{22}
1.285	1.96	2.575	−0.35516	0.3075	1.2164	1.1292	−0.34505	1.2182	1.1493
1.31	2.039	2.575	−0.33817	0.268	1.1998	1.0862	−0.33063	1.2013	1.1000
1.31	2.039	4	−0.32757	0.3645	1.0958	0.8542	−0.31500	1.0966	0.8747
1.31	2.575	3.5	−0.33002	0.4625	1.0795	0.8104	−0.31035	1.0802	0.8381
1.5	2.5	3.5	−0.24814	0.5	1.0645	0.7367	−0.22728	1.0637	0.7513
1.645	∞	∞	−0.1748	0.3	1.0259	0.6352	−0.16868	1.0262	0.6402
1.645	2	3.3	−0.19578	0.1775	1.0942	0.7822	−0.19312	1.0943	0.7841
1.645	2.24	3.3	−0.19083	0.2975	1.0754	0.7426	−0.18377	1.0751	0.7461
1.645	2.4	4	−0.18557	0.3775	1.0470	0.6818	−0.17506	1.0466	0.6874
1.96	∞	∞	−0.09117	0.3	1.0115	0.5692	−0.08702	1.0116	0.5710
1.96	2.4	3.3	−0.1069	0.22	1.0501	0.6537	−0.10399	1.0500	0.6542
1.96	2.575	4	−0.09846	0.3075	1.0263	0.6042	−0.09352	1.0259	0.6050
∞	∞	∞	−	−	−	−	0	1.000	0.500

Table 2. Measures of robustness for the smoothed ($\delta > 0$) and non-smoothed ($\delta = 0$) three-part redescending M-estimators at Φ , where e is efficiency, γ^* is gross-error sensitivity, λ^* is local-shift sensitivity, and κ^* is change-of-variance sensitivity.

a	b	c	P	δ	Location				Scale			
					e	γ^*	λ^*	κ^*	e	γ^*	λ^*	κ^*
1.645	2	3.3	−0.1931	0	0.914	1.950	1.265	7.474	0.638	1.960	3.290	11.931
1.645	2	3.3	−0.1957	0.175	0.914	1.954	1.265	6.699	0.639	1.973	2.940	11.350
1.96	2.4	3.3	−0.1040	0	0.952	2.139	2.178	10.110	0.764	2.255	3.920	20.794
1.96	2.4	3.3	−0.1064	0.2	0.952	2.143	2.178	8.588	0.765	2.270	3.520	18.960
2	2.6	3.2	−0.0926	0	0.959	2.155	3.333	12.636	0.786	2.271	4.000	28.793
2	2.6	3.2	−0.0976	0.3	0.958	2.164	3.333	9.651	0.786	2.304	3.400	25.417

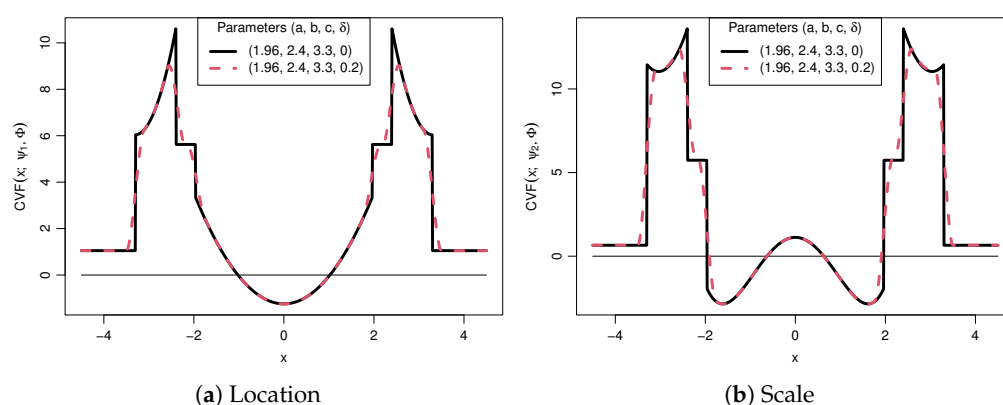


Figure 4. The $CVF(x; \psi, \Phi)$ of the three-part ψ -function, with $a = 1.96$, $b = 2.4$, $c = 3.3$, shown with and without $\delta = 0.2$ smoothing.

It is worth noting that the rejection point for the three-part redescender is $\pm c$, and at $\pm(c + \delta)$ for the smoothed variant. In practise, the value for the latter of $\Psi_{a,b,c}(c) \approx 0$ and δ is small, so this change has little impact on the estimator's robustness, as can be seen in Tables 1 and 2.

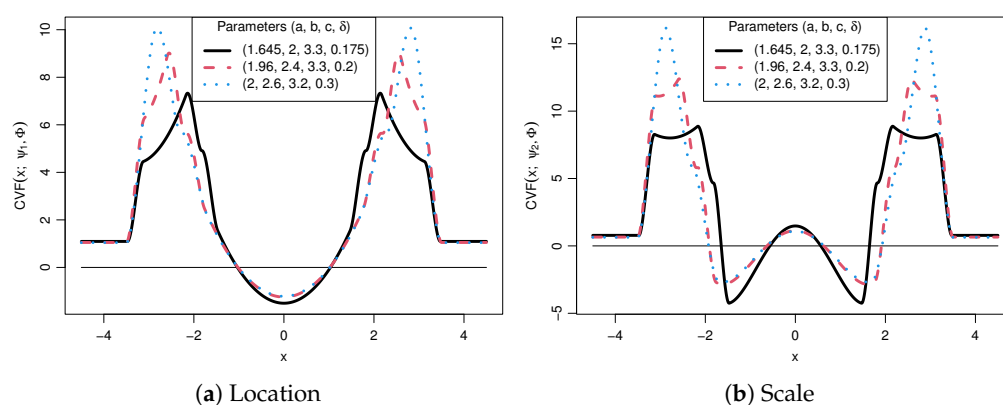


Figure 5. The smoothed three-part $CVF(x; \psi, \Phi)$ with smoothed functional parameters, as described in Table 2.

3.2. Contaminated Distribution Asymptotics

We now consider the asymptotic value and variance of $T[\psi, G_0]$, the multivariate location and scale solution of (6) under the contaminated model (1). Setting $\epsilon = 0.05$, we simulate the case where around 5% of the data pollutes our otherwise standard normal distribution. We hold $\alpha \in \{1, 2, 3\}$ and allow μ_2 to vary. Figures 6 and 7 show the asymptotic values for location and scale, respectively, for the smoothed three-part with parameters as tabulated in Table 2.

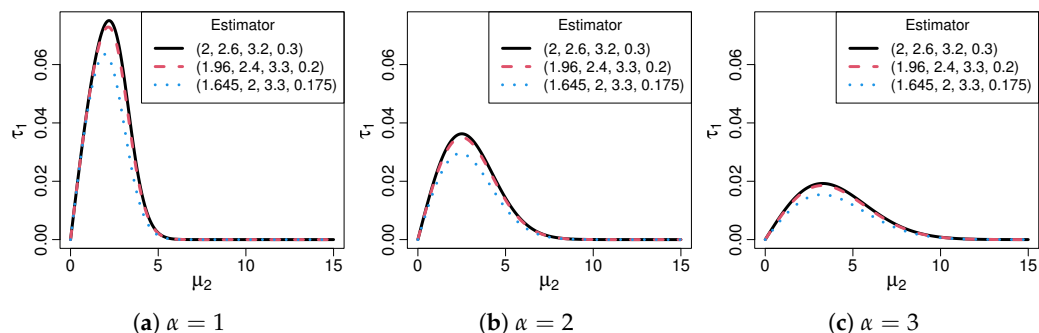


Figure 6. Plots of the asymptotic location solution of (6), τ_1 , for the estimators at (1) on $\mu_2 \in [0, 15]$. $\epsilon = 0.05$, representing an increasing scatter for the contaminating process. Smoothed ψ -function parameters are as described for Table 2.

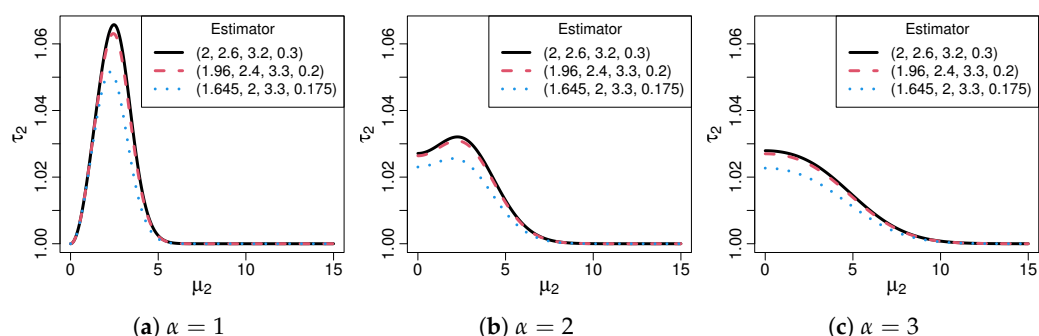


Figure 7. Plots of the asymptotic scale solution of (6), τ_2 , for the estimators at (1) on $\mu_2 \in [0, 15]$. $\epsilon = 0.05$, representing an increasing scatter for the contaminating process. Smoothed ψ -function parameters are as described for Table 2.

Note that if $\alpha \approx 0$, Figures 6 and 7 would resemble the IF and Figures 8 and 9 would resemble the CVF for the respective estimators. Instead, we are interested in the effect of scattered contamination in the vicinity of μ_2 .

Clearly, if μ_2 is non-zero and $\epsilon > 0$, the estimator, which is a solution of Equation (3), will be consistent to something other than $(\mu, \sigma)^T = (0, 1)^T$. This leads to the asymptotic bias, as given by $\mathbf{T}[\psi, G_0] - \mathbf{T}[\psi, \Phi]$. One can see that as scattered contamination moves beyond the $(c + \delta)\sigma$ rejection point, the estimates tend towards the true μ and σ .

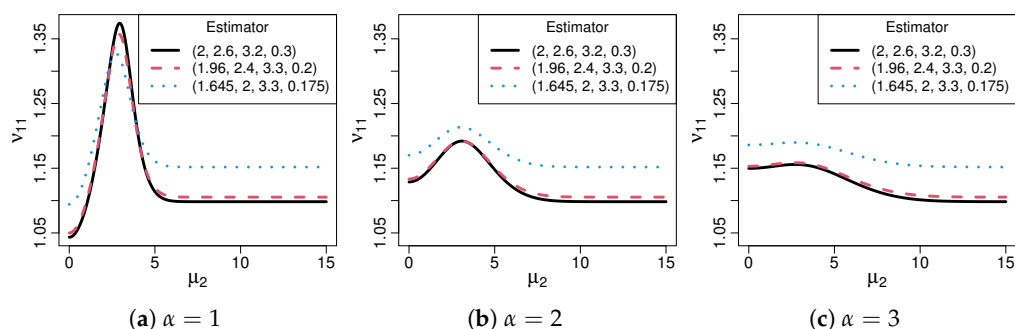


Figure 8. Plots of asymptotic variance for the location solution of (6), labelled v_{11} , for the estimators at (1) on $\mu_2 \in [0, 15]$. $\epsilon = 0.05$, representing an increasing scatter for the contaminating process. Smoothed ψ -function parameters are as described for Table 2.

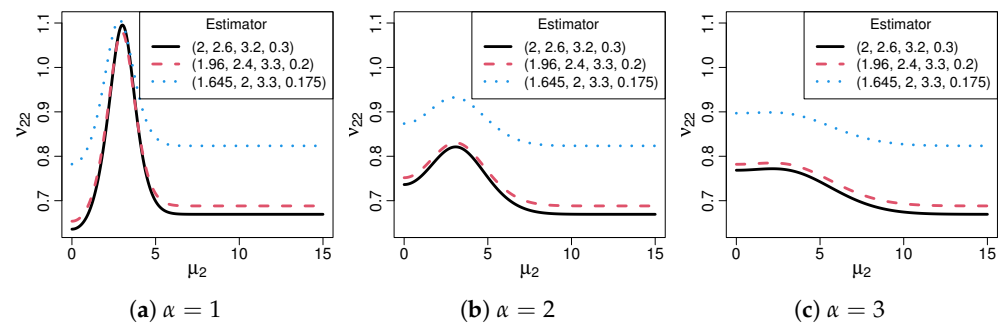


Figure 9. Plots' asymptotic variance for the scale solution of (6), labelled v_{22} , for the estimators at (1) on $\mu_2 \in [0, 15]$. $\epsilon = 0.05$, representing an increasing scatter for the contaminating process. Smoothed ψ -function parameters are as described for Table 2.

Figure 10 illustrates the benefits of smoothing for higher levels of contamination. The smoothed three-part estimator with $a = 2$, $b = 2.6$, $c = 3.3$ is compared with $\delta \in \{0, 0.3\}$. Clearly, asymptotic variance rises sharply around $\mu_2 \approx 3.6$; however, with smoothing, the peak is $\approx 75\%$ lower, indicating that smoothing is particularly useful when a higher level of contamination may be present in the vicinity of the rejection point.

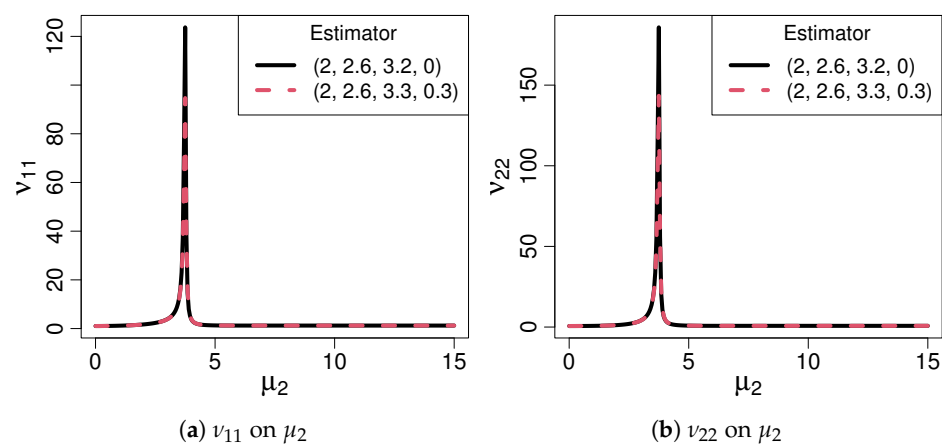


Figure 10. Asymptotic variance on $\mu_2 \in [0, 15]$ with $\epsilon = 0.15$ and $\alpha = 1$. The functional parameters are $a = 2$, $b = 2.6$, $c = 3.2$, and $\delta = 0$ versus $\delta = 0.3$.

4. Conclusions

A smoothed three-part redescender is introduced, as a refinement of Hampel's three-part estimators, with definitions based entirely upon piecewise polynomials. Smoothing is shown to yield a significant improvement in both change of variance sensitivity and asymptotic variance near the rejection point when the distribution has higher levels of asymmetric contamination. Additionally, the resultant functional is shown to be both weakly continuous and Fréchet differentiable in the neighbourhood of the normal distribution. Other robustness measures, including efficiency and asymptotic bias, are largely unchanged from the non-smoothed functions, indicating little-to-no downside to using the smoothed ψ -functions in place of the traditional non-smoothed variants. Anywhere an M-estimator for location and/or scale has been employed for statistical applications may be enhanced by using our approach.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/stats8020033/s1>. Calculus for a Smoothed Three Part Redescending M-Estimator.

Author Contributions: Conceptualisation, A.J.M. and B.R.C.; methodology, B.R.C. and A.J.M.; software, A.J.M.; validation, formal analysis, investigation, resources, and data curation, A.J.M. and B.R.C.; writing—original draft preparation, A.J.M. and B.R.C.; writing—review and editing, A.J.M. and B.R.C.; visualisation, A.J.M. and B.R.C.; supervision, B.R.C.; project administration, B.R.C.; funding acquisition, A.J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors report that there are no competing interests to declare.

Abbreviations

The following abbreviations are used in this manuscript:

MAD median of absolute deviations
MADN normalised median of absolute deviations

Appendix A. Proof of Theorem 1

The only difference in Conditions $\mathcal{W}$ in [6] (p. 43) and Conditions $\mathcal{W}$ in this paper is condition $\mathcal{W}_0$ and Condition $\mathcal{W}_3$, but the argument is exactly the same, resulting in Theorem 1 in this paper. Note that, in $\mathcal{W}_0$, F_θ is replaced by G_0 and θ is replaced by τ_0 . In $\mathcal{W}_3$, the matrix $\mathbf{M}(\theta)$ is replaced with $\mathbf{M}(\tau_0, G_0)$ (see (12)), which is assumed to be nonsingular.

Appendix B. Description of Proof of Theorem 2

We illustrate the Proof of Theorem 2 in this paper for the assumed choice of

$$\Psi_{a,b,c,\delta}(x, \tau) = \left[\psi_{1;a,b,c,\delta}\left(\frac{x - \tau_1}{\tau_2}\right), \psi_{2;a,b,c,\delta}\left(\frac{x - \tau_1}{\tau_2}\right) \right]^\top,$$

being the smoothed Hampel three-part redescender for location and scale, when $\delta > 0$.

Note that formula (13) of this paper is satisfied since the component functions in $\Psi_{a,b,c,\delta}$ are polynomials of degree at most five in their arguments with coefficients that are bounded. The next step is to integrate by parts each component power of x^k , for $1 \leq k \leq 5$. Note that, for any partition on which x^k is evaluated, call it P_i for the i^{th} partition, evaluating at the limits of integration, which are, say, $D_{\min;P_i}$ and $D_{\max;P_i}$, which, as the assumed parameter set is compact with a scale σ that is bounded away from zero, are both less than infinity in absolute value. It follows, for example, integrating by parts, that for $A = D_{\min;P_i}$ and $B = D_{\max;P_i}$,

$$\begin{aligned} \int_A^B x^k d(G - G_0) &= B^k(G(B) - G_0(B)) - A^k(G(A) - G_0(A)) \\ &\quad - \int_A^B (G(x) - G_0(x)) kx^{k-1} dx \end{aligned}$$

Noting also

$$|(G(B) - G_0(B))| \leq d_{KS}(G, G_0)$$

and similarly

$$|(G(A) - G_0(A))| \leq d_{KS}(G, G_0)$$

and seeing also

$$\left| \int_A^B (G(x) - G_0(x)) kx^{k-1} dx \right| \leq \int_A^B d_{KS}(G, G_0) |kx^{k-1}| dx < \text{Const} \times d_{KS}(G, G_0),$$

it follows from the above integration that

$$\left| \int_A^B x^k d(G - G_0) \right| \leq \text{const} \times d_{KS}(G, G_0)$$

whence it is seen that combining the integrals over all the finite number of partitions gives that

$$\int_{\mathbb{R}} \Psi_{a,b,c,\delta}(x, \tau_0) d(G - G_0)(x) = O(d_{KS}(G, G_0))$$

as $d_{KS}(G, G_0) \rightarrow 0$.

Now, consider $G_0 \in n(\epsilon, \Phi) \subset n_p(\epsilon, \Phi)$. According to [6] Theorem 2.5 and Corollary 2.3, there exists a Prohorov neighbourhood, $U(F_\theta)$, such that $T[\Psi, \rho, \cdot]$ exists and is weakly continuous at each $G \in U(F_\theta)$. Note that the arguments needed for a vector $\theta = [0, 1]^\top$, for example, are the same as in [6]. Assuming Conditions $\mathcal{W}$, there exists a qualitatively robust M-functional estimator T_n , which is robust and consistent with $T[G]$ for all $G \in U(F_\theta)$. This includes G_0 for small ϵ .

Finally, it can be remarked that Huber's monograph [10] establishes the weak continuity of the MAD estimating functional at an absolutely continuous distribution and, combined with the weak continuity of the median, suggests that the selection functional $\rho_0(F_{(\mu, \sigma)^\top}, \tau)$ (9) is also weakly continuous at an absolutely continuous distribution such as G_0 in the main paper. Using arguments equivalent to [6] Theorem 2.4, it can be noted that $T[\Psi_{a,b,c,\delta}, \rho, \cdot] = T[\Psi_{a,b,c,\delta}, \rho_0, \cdot]$ for all G in a small-enough Prohorov neighbourhood of $\Phi(x)$, which includes the distribution G_0 for small-enough ϵ .

References

1. Hampel, F.R. The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.* **1974**, *69*, 383–393. [\[CrossRef\]](#)
2. Bednarski, T.; Zontek, S. Robust estimation of parameters in a mixed unbalanced model. *Ann. Stat.* **1996**, *24*, 1493–1510. [\[CrossRef\]](#)
3. Bachmaier, M. Consistency of completely outlier-adjusted simultaneous redescending M-estimators of location and scale. *Adv. Stat. Anal.* **2007**, *91*, 197–219. [\[CrossRef\]](#)
4. Huber, P.J.; Ronchetti, E.M. *Robust Statistics*, 2nd, ed.; John Wiley & Sons: New York, NY, USA, 2009.
5. Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A. *Robust Statistics, the Approach Based on Influence Functions*; John Wiley & Sons: New York, NY, USA, 1986.
6. Clarke, B.R. *Robustness Theory and Application*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
7. Andrews, D.F.; Bickel, P.J.; Hampel, F.R.; Huber, P.J.; Rogers, W.H.; Tukey, J.W. *Robust Estimates of Location: Survey and Advances*; Princeton University Press: Princeton, NJ, USA, 1972.
8. Huber, P.J. Robust estimation of a location parameter. *Ann. Math. Statist.* **1964**, *35*, 73–101. [\[CrossRef\]](#)
9. Huber, P.J. Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1973**, *1*, 799–821. [\[CrossRef\]](#)
10. Huber, P.J. *Robust Statistics*, 1st ed.; John Wiley & Sons: New York, NY, USA, 1981.
11. Arachchige, C.N.P.G.; Prendergast, L.A. Confidence intervals for median absolute deviations. *Commun. Stat. Simul. Comput.* **2024**, *52*, 1–10. [\[CrossRef\]](#)
12. Clarke, B.R.; Milne, C.J. A small sample bias correction and implications for inference. In Proceedings of the 59th ISI World Statistics Congress, Hong Kong, China, 25–30 August 2013.
13. Hearn, D.; Baker, M.P. *Computer Graphics*, 2nd ed.; Prentice Hall, Inc.: New York, NY, USA, 1997.
14. Rousseeuw, P.J.; Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **1993**, *88*, 1273–1283. [\[CrossRef\]](#)

15. Hampel, F.R. Contributions to the Theory Of Robust Estimation. Ph.D. Thesis, University of California, Berkeley, CA, USA, 1968.
16. Huber, P.J. *Robust Statistical Procedures*, 2nd ed.; CBMS-NSF Regional Conference Series in Applied Mathematics; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1996.
17. Bednarski, T. Fréchet differentiability and robust estimation. In *Asymptotic Statistics, Proceedings of the Fifth Prague Symposium 1994*; Mandl, P., Hušková, M., Eds.; Springer: Berlin/Heidelberg, Germany, 1993; pp. 49–58.
18. Hampel, F.R. A general qualitative definition of robustness. *Ann. Math. Statist.* **1971**, *42*, 1887–1896. [[CrossRef](#)]
19. Clarke, B.R. Nonsmooth analysis and Fréchet differentiability of M-functionals. *Probab. Theory Relat. Fields* **1986**, *73*, 197–209. [[CrossRef](#)]
20. Varadarajan, V.S. On the convergence of probability distributions. *Sankhyā A* **1958**, *19*, 23–26.
21. Prohorov, Y.V. Convergence of random processes and limit theorems in probability. *Theory Probab. Appl.* **1956**, *1*, 157–214. [[CrossRef](#)]
22. Rousseeuw, P.J. A new infinitesimal approach to robust estimation. *Z. Wahrsch. Verw. Geb.* **1981**, *56*, 127–132. [[CrossRef](#)]
23. Rousseeuw, P. New Infinitesimal Methods in Robust Statistics. Ph.D. Thesis, Vrije Universiteit Brussel, Brussels, Belgium, 1981.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.