

M -Estimators for Regression with Changing Scale

Christopher S. Withers

Industrial Research Limited, Lower Hutt, New Zealand

Saralees Nadarajah

University of Manchester, Manchester, UK

Abstract

M -estimation provides a class of estimators for the ‘signal plus noise’ problem, where the signal has a parametric form and the distribution of the noise is unspecified. Here, we extend this to modeling observations subject to trends in both location and scale, that is, to the model

observation = (location signal) + (scale signal) \times (noise),

where the *location signal* and *scale signal* are smooth functions of an unknown q -vector θ say, and the components of the noise have some unknown cumulative distribution function (cdf) F say. We define the *scaled M -estimator* of θ with respect to a given smooth function $\rho : \mathbb{R} \rightarrow \mathbb{R}$. When the scale is not changing this reduces to the usual *unscaled M -estimator* requiring that F be suitably centered with respect to ρ .

AMS (2000) subject classification. Primary 62G20, Secondary 62F12.

Keywords and phrases. M -estimator, Regression, Robust, Trend in scale

1 Introduction

The asymptotic theory of M -estimators for regression models has been the subject of many papers. The papers are too many to cite. For comprehensive reviews, we refer the readers to the following most excellent books and book chapters: Andrews et al. (1972), Rey (1978), Randles and Wolfe (1979, Chapter 5), Serfling (1980, Chapter 7), Bierens (1981), Huber (1981), Hampel et al. (1986), Carroll and Ruppert (1988), Staudte and Sheather (1990), Koul (1992, Chapter 4), Bickel et al. (1993), Marazzi (1993), Pfanzagl (1994), Rieder (1994), Verboon (1994), Rao and Toutenburg (1995, Chapter 9), Huber (1996, Chapter 3), van der Vaart and Wellner (1996, part III), Prakasa Rao (1999, Chapter 2), van de Geer (2000), Peracchi (2001, Chapter 1), Koul (2002), Maronna et al. (2006), Liese and Miescke (2008, Chapter 7), Huber and Ronchetti (2009), and Lee (2010).

Table 1: p -values of the six models fitted to annual minimum temperature-data

Model	p -value
const linear	0.042
linear const	0.044
linear linear	0.045
const exponential	0.092
exponential const	0.046
exponential exponential	0.111

M -estimators have received applications in almost every area of science, engineering and medicine. Some recent application areas include: nonlinear regression models with long-memory errors (Chen et al., 2011), high breakdown robust estimation in computer vision (Hoseinnezhad and Bab-Hadiashar, 2011), and detection / compensation of GPS measurement biases (Faurie and Giremus, 2010).

This paper develops M -estimation theory for a richer class of models than the usual ‘signal + noise’ class of regression models. This richer class is generally able to fit data better than the ‘signal + noise’ class of models, see Tables 1, 2, 3, 4, 5 and 6 in Section 7 and Figs. 5 to 7 in Section 7 for illustrations. Instead of having only a ‘location signal’, we allow also a ‘scale signal’. That is the class of models we shall provide M -estimator theory for is

$$Y_\ell = \mu_\ell + \sigma_\ell e_\ell \tag{1.1}$$

in \mathbb{R} , μ_ℓ in \mathbb{R} , $\sigma_\ell > 0$ and e_ℓ in \mathbb{R} for $\ell = 1, \dots, n$, where $\{\mu_\ell, \sigma_\ell : \ell = 1, \dots, n\}$ are “smooth functions” of an unknown q -vector θ and the ‘residuals’ $\{e_\ell, \ell = 1, \dots, n\}$ are independent and identically distributed (i.i.d.) with some unknown cdf $F(x)$. One can also take $\{\mu_\ell, \sigma_\ell : \ell = 1, \dots, n\}$ to be functions

Table 2: p -values of the six models fitted to annual maximum temperature data

Model	p -value
const linear	0.032
linear const	0.039
linear linear	0.040
const exponential	0.086
exponential const	0.035
exponential exponential	0.109

Table 3: p -values of the six models fitted to annual average temperature data

Model	p -value
const linear	0.021
linear const	0.045
linear linear	0.046
const exponential	0.025
exponential const	0.046
exponential exponential	0.121

of some parameter in a lower dimensional space, say \mathbb{R}^r , $r < q$. Examples of this kind are considered later, see, for example, Section 2.4. By “smooth function”, we mean that the function is infinitely differentiable with respect to its arguments.

The i.i.d. assumption of the residuals may sound restrictive. But models of the kind (1.1) with the i.i.d. assumption have been fitted to many real data sets: annual maximum daily rainfall in west central Florida (Nadarajah, 2005), annual flood peaks of the Pachang River, Taiwan (Nadarajah and Shiau, 2005); annual maximum daily windrun in New Zealand (Withers and Nadarajah, 2006); annual maximum precipitation in China (Feng et al. 2007); annual maximum temperatures in New Zealand (Withers et al. 2009); to mention just a few. The theorems presented will not hold if the i.i.d. assumption is not satisfied. Exactly how the results will change is a possible future work.

M -estimation theory for a particular case of (1.1) has been considered by Carroll and Ruppert (1982). The theory developed in this paper is more general. For other work extended by (1.1), we refer the readers to Carroll (1982), Welsh et al. (1994), Bianco et al. (2000) and references therein.

Table 4: Measures of robustness for models fitted to annual minimum temperature data

Parameter	Exp-Const model	Exp-Exp model
α_1	0.024	0.010
α_2	0.040	0.014
β_1	0.027	0.012
β_2		0.014

Table 5: Measures of robustness for models fitted to annual maximum temperature data

Parameter	Linear-Const model	Exp-Exp model
α_1	0.043	0.001
α_2	0.061	0.002
β_1	0.010	0.009
β_2		0.015

We shall see that in general for the M -estimator of θ to be consistent the noise cdf $F(x)$ must be centered and scaled so that

$$\int_{-\infty}^{\infty} \rho^{(1)}(e)dF(e) = 0, \int_{-\infty}^{\infty} e\rho^{(1)}(e)dF(e) = 1, \tag{1.2}$$

where $\rho^{(r)}(e)$ is the r th derivative of $\rho(e)$ and $\rho : \mathbb{R} \rightarrow \mathbb{R}$ is some smooth function with respect to which M -estimators are defined. Unless otherwise stated, we assume throughout that $\rho(\cdot), \mu_\ell(\cdot), \sigma_\ell(\cdot)$ are smooth functions and that $F(\cdot)$ is differentiable.

Before exploring the implications of the constraints, (1.2), let us review the constraints on the familiar “observation = signal + noise” class of models

$$Y_\ell = \mu_\ell + e_\ell \tag{1.3}$$

in \mathbb{R} for $\ell = 1, 2, \dots, n$, where $\mu_\ell = \mu_\ell(\theta)$ is a function of an unknown vector θ , and $\{e_\ell\}$ are i.i.d. with some unknown cdf $F(x)$.

Parametric methods assume that $F(x)$ has a parametric form, say $F(x) = F_\theta(x)$ and then apply a method such as maximum likelihood (ML) or the method of moments to estimate θ . However, if $F(x)$ does not in fact have the parametric form assumed, then the confidence regions (CRs) for θ will generally not have the correct level even for large sample size n .

Semi-parametric methods do not assume any particular form for the noise cdf $F(x)$ but it must satisfy a constraint *depending on the estimator* in order that the estimator is consistent. The same is true for the usual M -estimator

Table 6: Measures of robustness for models fitted to annual average temperature data

Parameter	Exp-Const model	Exp-Exp model
α_1	0.020	0.001
α_2	0.030	0.002
β_1	0.019	0.007
β_2		0.011

Table 7: p -values for the fit of Edgeworth type approximation

m	p -value
1	0.000
2	0.000
3	0.001
4	0.001
5	0.001
6	0.001
7	0.001
8	0.001
9	0.002
10	0.002

with respect to a given smooth function $\rho : \mathbb{R} \rightarrow \mathbb{R}$, typically convex, that is $\hat{\theta}$ minimizing

$$\Lambda_{0,n}(\theta) = n^{-1} \sum_{\ell=1}^n \rho(Y_{\ell} - \mu_{\ell}(\theta)),$$

where $\mu_{\ell}(\theta) = m_{\ell}(\gamma) + \xi$. The cdf F must be centered so that

$$\int_{-\infty}^{\infty} \rho^{(1)}(e) dF(e) = 0.$$

We now rewrite (1.3) in the ‘basic’ form

$$Y_{\ell} = m_{\ell}(\gamma) + \epsilon_{\ell} \tag{1.4}$$

in \mathbb{R} for $\ell = 1, 2, \dots, n$ with $\{\epsilon_{\ell}\}$ i.i.d. $G(x)$ say. Unlike F , ξ depends on the choice of ρ . The parameter $\xi = \xi(\rho, G)$ is given implicitly by

$$\int_{-\infty}^{\infty} \rho^{(1)}(x - \xi) dG(x) = 0,$$

where $e_{\ell} = \epsilon_{\ell} - \xi$ and $F(x) = G(x + \xi)$. Much statistical theory has been developed assuming that residuals are normally distributed with mean zero. For example, Seber and Wild (1989, page 111) consider the model $\mu_{\ell} = \theta_1 \{1 - \exp(-\theta_2 x_{\ell})\}$, $e_{\ell} \sim \mathcal{N}_1(0, \theta_3)$. However, unless there is a compelling reason for believing residuals are symmetrically distributed - and indeed normal, a safer approach would be to add in a location parameter and use a semi-parametric estimator.

A simple choice for usual M -estimators is to take $\rho(e)$ proportional to $|e|^r$ for some $r > 0$ since a choice like $\rho(e) = |e| + e^2$ is not dimensionally sound. The alternative of $\rho(e) = \rho_0(e/\sigma_0)$ for some given σ_0 of the same dimension as the observations $\{Y_\ell\}$ raises the question of how to choose σ_0 : Section 5 gives one answer to this question.

The classical choice $\rho(e) = e^2/2$ gives the *least squares estimator* (LSE), while the choice $\rho(e) = |e|^r/r$, r even gives the more general L_r estimator. We shall see in Tables 4 to 6 in Section 7 that our more general model (allowing scale to change, for the location parameters - those of $\{\mu_\ell\}$, but *not* for the remaining parameters - typically those of $\{\sigma_\ell\}$) is more robust to outliers than the LSE for the “signal + noise” model.

If F were *known*, say $F(x) = F_0(x)$ with probability density function (pdf) f_0 , then the M -estimator with $\rho(x) = -\log f_0(x)$, is the ML estimator (MLE).

The main results of this paper are given in Section 2: the case, where there is a trend in the scale of the observations, as well as a trend in the location. So, instead of (1.4) the basic model is

$$Y_\ell = m_\ell + s_\ell \epsilon_\ell \tag{1.5}$$

in \mathbb{R} for $\ell = 1, 2, \dots, n$, where $m_\ell = m_\ell(\gamma)$ and $s_\ell = s_\ell(\gamma) > 0$ are given functions of an unknown vector γ , and $\{\epsilon_\ell\}$ are i.i.d. $G(x)$, unknown. For estimators to be consistent we now need in general to introduce not only a location parameter ξ as before but also a scale parameter τ say. In the case of M -estimators these two parameters are certain functions of ρ and $G(x)$. Rewriting (1.5) as the ‘working’ model (1.1) with $e_\ell = (\epsilon_\ell - \xi) / \tau$ i.i.d. $F(x) = G(\xi + \tau x)$, we have

$$\mu_\ell = m_\ell + \xi s_\ell, \quad \sigma_\ell = \tau s_\ell, \quad \theta' = (\gamma', \xi, \tau). \tag{1.6}$$

(More generally, θ is some one-to-one function of (γ', ξ, τ) . We denote the pdf of $F(x)$ by $f(x)$ when it exists.) If $F(x)$ had a known form $F_0(x)$ with pdf $f_0(x)$, one could use the MLE for θ . That is, one could choose $\hat{\theta}$ to minimize

$$\Lambda_n(\theta) = n^{-1} \sum_{\ell=1}^n \lambda_\ell(\theta),$$

where

$$\lambda_\ell(\theta) = \rho(e_\ell(\theta)) + \log \sigma_\ell(\theta), \tag{1.7}$$

$$e_\ell(\theta) = (Y_\ell - \mu_\ell(\theta)) / \sigma_\ell(\theta), \tag{1.8}$$

and $\rho(e) = -\log f_0(e) = \rho(e, F_0)$ say. The *scaled M-estimator* of θ with respect to ρ is defined in the same way with the form of ρ now allowed to be arbitrary. For the case when the observations form a random sample, that is $\mu_\ell = \theta_1, \sigma_\ell = \theta_2$, this idea appears on pages 39-40 of Andrews et al. (1972): see Section 2.2 to follow.

By (1.2) the parameters $(\xi, \tau) = (\xi, \tau)(\rho, G)$ are defined implicitly by

$$\int_{-\infty}^{\infty} \rho^{(1)}((x - \xi)/\tau) dG(x) = 0, \quad \int_{-\infty}^{\infty} ((x - \xi)/\tau) \rho^{(1)}((x - \xi)/\tau) dG(x) = 1. \tag{1.9}$$

For example, for the L_r -estimator with r even, ξ and τ are given implicitly and explicitly, respectively, by $\mathbb{E}_G \left[|e_1 - \xi|^{r-1} \text{sign}(e_1 - \xi) \right] = 0$ and $\tau = (\mathbb{E}_G [|e_1 - \xi|^r])^{1/r}$, so $\xi = \mathbb{E}_G [e_1]$ for $r = 2$. Example 2.1 gives some simplifications for the L_r -estimator when r is even and the parameters of $\{\mu_\ell\}$ and those of $\{\sigma_\ell\}$ are distinct. We call γ the *fundamental parameter* and ξ, τ the *working parameters* as only they depend on the choice of ρ .

Section 3 continues with the case, where the parameters of $\{\mu_\ell\}$ are different from those of $\{\sigma_\ell\}$. Section 4 relates the M -estimator to the quasi-likelihood estimator and gives conditions for likelihood ratio type tests to exist. If we want to compare different estimators of the fundamental parameter γ we must ensure that μ_ℓ and σ_ℓ must have the form (1.6). Section 5 explores these constraints. For example, we identify what constraints this puts on the “linear-linear model” $\mu_\ell = \alpha'x_\ell, \sigma_\ell = \beta'y_\ell$. Section 6 shows how to apply our model to obtain a simultaneous analysis for independent series sharing some common parameters. The motivation is modeling trend in New Zealand temperature series. By combining series one is more likely to pick up a weak scale signal if one exists. By using series that do not overlap in time, one effectively increases the length of the series without having to deal with the problem of correlation between series. Section 7 illustrates a data application. Finally, some future work are noted in Section 8.

2 Main Results

Here, we assume the location-scale model (1.1) with $\{e_\ell\}$ i.i.d. $F(x)$, and we obtain asymptotic normality of $\hat{\theta}$ minimizing

$$\Lambda_n(\theta) = n^{-1} \sum_{\ell=1}^n \lambda_\ell(\theta) \tag{2.1}$$

for $\lambda_\ell(\theta)$ of (1.7), (1.8) for given smooth functions $\{\mu_\ell(\theta), \sigma_\ell(\theta)\}$, and $\rho(e)$. These are typically taken as one, but alternatively may be chosen to reflect the credibility of the observations.

Section 2.1 derives the asymptotic distribution of $\hat{\theta}$ for the general case, see Theorem 2.1. The particular case for $q = 2$ is given in Section 2.2. Robustness of $\hat{\theta}$ is discussed in Section 2.3. CRs for smooth functions of θ are constructed in Section 2.4. Estimation of θ and related issues for $\rho(e) = |e|^r/d$ are given in Section 2.5. Estimation of θ and related issues for $\rho(e) = |e|^r/r$ are given in Section 2.6. Section 2.7 derives the asymptotic behavior of the unscaled M -estimator when the scale is changing.

2.1. *Asymptotic distribution of $\hat{\theta}$.* We first introduce some notation that we use throughout. Set

$$\rho_{i,j} = \int_{-\infty}^{\infty} \rho^{(i)}(e)e^j dF(e), \quad \rho_{i,j,k} = \int_{-\infty}^{\infty} \rho^{(i)}(e)\rho^{(j)}(e)e^k dF(e). \tag{2.2}$$

For example, (1.2) can be written $\rho_{1,0} = 0, \rho_{1,1} = 1$. For any real or compatible vector sequences $\mathcal{F} = \{\mathcal{F}_\ell, \ell = 1, \dots, n\}, \mathcal{G} = \{\mathcal{G}_\ell, \ell = 1, \dots, n\}, \dots$, set

$$[\mathcal{F}\mathcal{G}\dots] = n^{-1} \sum_{\ell=1}^n \mathcal{F}_\ell \mathcal{G}_\ell \dots$$

For example, $[\sigma^{-2}\mu.\sigma'] = n^{-1} \sum_{\ell=1}^n \sigma_\ell^{-2}\mu_\ell.\sigma'_\ell$. Set

$$a_{1,1} = \rho_{2,0}, \quad a_{1,2} = a_{2,1} = \rho_{2,1}, \quad a_{2,2} = \rho_{2,2} + 2\rho_{1,1} - 1, \quad a = (a_{i,j}), \tag{2.3}$$

$$b_{1,1} = \rho_{1,1,0} - \rho_{1,0}^2, \quad b_{1,2} = \rho_{1,1,1} - \rho_{1,0}\rho_{1,1}, \quad b_{2,2} = \rho_{1,1,2} - \rho_{1,1}^2, \quad b = (b_{i,j}). \tag{2.4}$$

Theorem 2.1. (1.1) with F having finite moments and with all relevant matrices invertible. Let $\hat{\theta}$ minimize (2.1). We have the following:

(I) If F is centered and scaled so that $\rho_{1,0} = 0, \rho_{1,1} = 1$, then

$$n^{1/2} (\hat{\theta} - \theta) \dot{\sim} \mathcal{N}_q(0, V), \tag{2.5}$$

where $V = A^{-1}BA^{-1}$ for

$$A = a_{1,1} \lim_{n \rightarrow \infty} [\sigma^{-2}\mu.\mu'] + a_{1,2} \lim_{n \rightarrow \infty} \sum^2 [\sigma^{-2}\mu.\sigma'] + a_{2,2} \lim_{n \rightarrow \infty} [\sigma^{-2}\sigma.\sigma'], \tag{2.6}$$

$$B = b_{1,1} \lim_{n \rightarrow \infty} [\sigma^{-2}\mu.\mu'] + b_{1,2} \sum^2 \lim_{n \rightarrow \infty} [\sigma^{-2}\mu.\sigma'] + b_{2,2} \lim_{n \rightarrow \infty} [\sigma^{-2}\sigma.\sigma'],$$

and $\sum^2 C = C + C'$ for C a square matrix.

- (II) If $\{\sigma_\ell\}$ do not depend on θ and $\rho_{1,0} = 0$, then (I) holds, only the first term in the above expressions for A , B is non-zero, also $V = c_{1,1} \lim_{n \rightarrow \infty} [\sigma^{-2} \mu \cdot \mu']^{-1}$ for $c_{1,1} = a_{1,1}^{-2} b_{1,1} = \rho_{2,0}^{-2} \rho_{1,1,0}$.
- (III) If $\{\mu_\ell\}$ do not depend on θ and $\rho_{1,1} = 1$, then (2.5) holds, only the last term in the above expressions for A , B is non-zero; also $V = c_{2,2} \lim_{n \rightarrow \infty} [\sigma^{-2} \sigma \cdot \sigma']^{-1}$ for $c_{2,2} = a_{2,2}^{-2} b_{2,2} = (\rho_{2,2} + 1)^{-2} (\rho_{1,1,2} - 1)$.
- (IV) If for $i = 1, 2$, γ_i is an q_i -vector, $\{\sigma_{\ell,0}\}$ are given scalars, $\{y_\ell\}$ are given q_2 -vectors such that $[y \cdot] = 0$, $\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$, $\mu_\ell = m_\ell(\gamma) + \xi \sigma_\ell(\gamma)$, $\sigma_\ell = \sigma_{\ell,0} \exp(y'_\ell \gamma_2)$, and $\rho_{1,0} = 0$, then (2.5) holds with $a_{2,2}$ in (2.6) replaced by $\rho_{2,2} + \rho_{1,1}$.

Note that (II) of Theorem 2.1 corresponds to constrained M -estimators of Mendes and Tyler (1996). The proof of Theorem 2.1 is given in Appendix F. We use the notation $\partial_i = \partial/\partial\theta_i$, $\lambda_{\ell,i,j,\dots} = \partial_i \partial_j \cdots \lambda_\ell(\theta)$, and similarly for $\Lambda_{n,i}$, $e_{\ell,i,j,\dots}$, $\mu_{\ell,i,j,\dots}$, $\sigma_{\ell,i,j,\dots}$. We set $\Delta = \widehat{\theta} - \theta$,

$$A_{i,j} = \lim_{n \rightarrow \infty} \mathbb{E} [\Lambda_{n,i,j}] = \lim_{n \rightarrow \infty} \mathbb{E} \left[n^{-1} \sum_{\ell=1}^n \lambda_{\ell,i,j} \right],$$

and

$$B_{i,j} = \lim_{n \rightarrow \infty} n^{-1} \text{covar} \left[\sum_{\ell=1}^n \lambda_{\ell,i}, \sum_{\ell=1}^n \lambda_{\ell,j} \right] = \lim_{n \rightarrow \infty} n^{-1} \sum_{\ell=1}^n \text{covar} [\lambda_{\ell,i}, \lambda_{\ell,j}].$$

Furthermore, we let $(A^{i,j})$ be the inverse of $(A_{i,j})$, $q \times q$ and let $\sum_{i,j}^2 \mathcal{F}_{i,j} = \mathcal{F}_{i,j} + \mathcal{F}_{j,i}$.

If $\sigma_\ell(\theta) \equiv 1$, then (II) of Theorem 2.1 reduces to the result of Huber (1964): $n^{1/2} (\widehat{\theta} - \theta) \sim \mathcal{N}_q(0, c_{1,1} V^{-1})$, where $c_{1,1} = \rho_{2,0}^{-2} \rho_{1,1,0}$ and $V = \lim_{n \rightarrow \infty} [\mu \cdot \mu']$.

Corollary 2.1. *Note that A and B are proportional for (II) and (III) of Theorem 2.1 but not for the general case (I) unless either $(\rho_{2,0}, \rho_{2,1}, \rho_{2,2} + 1)$ is proportional to $(\rho_{1,1,0}, \rho_{1,1,1}, \rho_{1,1,2} - 1)$, or $\sigma_\ell(\theta) = \nu_1 + \nu_2 \mu_\ell(\theta)$ for some known constants ν_1, ν_2 . In this case, $A = \kappa(a, \nu_2) \lim_{n \rightarrow \infty} [\sigma^{-2} \mu \cdot \mu']$, $B = \kappa(b, \nu_2) \lim_{n \rightarrow \infty} [\sigma^{-2} \mu \cdot \mu']$, $V = \kappa(a, \nu_2)^{-2} \kappa(b, \nu_2) \lim_{n \rightarrow \infty} [\sigma^{-2} \mu \cdot \mu']^{-1}$, where $\kappa(a, \nu_2) = a_{1,1} + 2a_{1,2} \nu_2 + a_{2,2} \nu_2^2$. Firth (1987, Sections 2 and 3) considers the cases $\nu_2 = 0$ and $\nu_1 = 0$ under slightly different assumptions.*

The proof of Corollary 2.1 is given in Appendix F.

2.2. *Asymptotic distribution for $q = 2$.* Suppose $\mu_\ell \equiv \theta_1, \sigma_\ell \equiv \theta_2$, so $q = 2$ and the observations form a random sample. Then, for a, b of (2.3), (2.4), $A = a/\theta_2^2$ and $B = b/\theta_2^2$. So, $V = \theta_2^2 c$, where

$$c = a^{-1}ba^{-1} = (c_{i,j}) \tag{2.7}$$

say. In terms of $d_a = \det a = a_{1,1}a_{2,2} - a_{1,2}^2$, and $d_b = \det b = b_{1,1}b_{2,2} - b_{1,2}^2$, c and $c^{-1} = (c^{i,j})$ are given by

$$\begin{aligned} d_a^2 c_{1,1} &= d_b c^{2,2} = b_{1,1}a_{2,2}^2 - 2b_{1,2}a_{1,2}a_{2,2} + b_{2,2}a_{1,2}^2, \\ d_a^2 c_{1,2} &= -d_b c^{1,2} = -b_{1,1}a_{1,2}a_{2,2} + b_{1,2}(a_{1,1}a_{2,2} + a_{1,2}^2) - b_{2,2}a_{1,1}a_{1,2}, \\ d_a^2 c_{2,2} &= d_b c^{1,1} = b_{1,1}a_{1,2}^2 - 2b_{1,2}a_{1,1}a_{1,2} + b_{2,2}a_{1,1}^2. \end{aligned}$$

The estimator $\hat{\theta}$ was proposed for this example on pages 39-40 of Andrews et al. (1972); c.f. page 136 of Huber and Ronchetti (2009). Let G be the cdf of $\{Y_\ell\}$. Their formulas (29), (30) for the influence function (first von Mises derivative) of $\theta(G)$ defined by $\hat{\theta} = \theta(G_n)$, where G_n is the sample empirical cdf, can be written $a\theta_G(x) = \theta_2 t(y)$, where $t_1(y) = \rho^{(1)}(y)$, $t_2(y) = y\rho^{(1)}(y) - 1$ and $y = (x - \theta_1)/\theta_2$. So, $n^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}_2(0, V)$, where $V = \int_{-\infty}^{\infty} \theta_G(x)\theta_G(x)'$ and $dG(x) = \theta_2^2 a^{-1}b^*a^{-1}$ for $b^* = \int_{-\infty}^{\infty} t(y)t(y)'$, $dF(y) = b$, in agreement with our result.

2.3. *Robustness of $\hat{\theta}$.* When is $\hat{\theta}$ robust to outliers? Assuming ρ is bounded near zero, B is bounded in F if $\rho_{1,1,2}$ is bounded, that is if $e\rho^{(1)}(e)$ is bounded. This means $\rho(e)$ must be a ‘redescending’ estimator (that is $\rho^{(1)}(e)$ may increase but for large $|e|$ must decrease). This condition rules out $\rho(e) = |e|^r/r$.

Let $\rho_{1,1,i}^* = \int_{-\infty}^{\infty} f(e)^{-1}f^{(1)}(e)^2e^i de$, where f is the pdf of F . If $\mu_\ell = \mu_\ell(\alpha), \sigma_\ell = \sigma_\ell(\beta), \theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ then the MLEs for α and β are asymptotically independent if $\rho_{1,1,1}^* = 0$, in particular if F is symmetric about zero, as for the normal and the double-exponential distributions.

In the *parametric* case, where $F(e) = G((\xi + \tau e)/\tau)$ for some given cdf G , (ξ, τ) are determined by (1.2), that is by (1.9). So, $\xi = 0$ if ρ and G are symmetric about zero.

Let \hat{F} be the empirical cdf of the estimated residuals

$$\hat{e}_\ell = \{Y_\ell - \mu_\ell(\hat{\theta})\} / \sigma_\ell(\hat{\theta})$$

for $\ell = 1, \dots, n$. Set $V(F, \theta) = V$ of (2.5) and $\hat{V} = V(\hat{F}, \hat{\theta})$. Under mild conditions $\hat{V} \xrightarrow{p} V$ as $n \rightarrow \infty$. For example, writing $\rho_{i,j}$ or $\rho_{i,j,k}$ of (2.2) as

$\int_{-\infty}^{\infty} g(e)dF(e)$, a consistent estimator is $\int_{-\infty}^{\infty} g(e)d\widehat{F}(e) = n^{-1} \sum_{\ell=1}^n g(\widehat{e}_{\ell})$. The α -quantile of the cdf of Y_{ℓ} is $Q_{\ell}(\alpha) = \mu_{\ell} + \sigma_{\ell}F^{-1}(\alpha)$, $0 < \alpha < 1$. If $\zeta_n/n \rightarrow \alpha$ in $(0, 1)$ as $n \rightarrow \infty$ for a sequence of integers ζ_n , then a consistent estimator of this quantile is

$$\widehat{Q}(\alpha) = \mu_{\ell}(\widehat{\theta}) + \sigma_{\ell}(\widehat{\theta})\widehat{e}_{(\zeta_n)},$$

where $\widehat{e}_{(1)} \leq \dots \leq \widehat{e}_{(n)}$ are the ordered values of $\widehat{e}_1, \dots, \widehat{e}_n$.

2.4. *Confidence regions for θ* . Suppose we want an asymptotic CR for some function $t(\theta)$ in \mathbb{R}^r with $r \leq q$. This is obtained by Studentizing

$$n^{1/2}(\widehat{t} - t(\theta)) \xrightarrow{d} \mathcal{N}_r(0, V_t),$$

where $\widehat{t} = t(\widehat{\theta})$,

$$V_t = V_t(\rho, F, \theta) = tVt' = \sum_{i,j=1}^q t_i V_{i,j} t_j$$

is $r \times r$, $\dot{t} = \partial t(\theta)/\partial \theta'$ is $r \times q$, and $t_{.i} = \partial t(\theta)/\partial \theta_i$. So, a CR for $t = t(\theta)$ with asymptotic coverage 0.95 is

$$n(t - \widehat{t})' \widehat{V}_t^{-1} (t - \widehat{t}) \leq \chi_{r,0.95}^2, \tag{2.8}$$

where $\widehat{V}_t = V_t(\rho, \widehat{F}, \widehat{\theta})$. A p -value for $(H_0 : t = 0)$ is

$$p_n = p(t) \approx Ch_r(q_n) \tag{2.9}$$

for $q_n = n\dot{t}'\widehat{V}_t^{-1}\widehat{t}$, where $Ch_r(x) = \Pr(\chi_r^2 < x)$. So, if $t(\theta)$ is univariate, a confidence interval (CI) for t with asymptotic coverage 0.95 is

$$|t - \widehat{t}| \leq 1.96 \left(\widehat{V}_t/n\right)^{1/2}, \tag{2.10}$$

and a one-sided p -value for $(H_0 : t = 0)$ is

$$p_n = p(t) \approx \Phi(q_n) \tag{2.11}$$

for $q_n = n^{1/2}\dot{t}'\widehat{V}_t^{-1}\widehat{t}$, where $\Phi(\cdot)$ is the standard normal cdf. Typically μ_{ℓ} and σ_{ℓ} have different parameters.

Example 2.1. Suppose that $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, μ_ℓ depends only on α in \mathbb{R}^p and σ_ℓ depends only on β in $\mathbb{R}^{p'}$. Set

$$x_{\ell,1} = \partial\mu_\ell/\partial\alpha, \quad x_{\ell,2} = \partial\sigma_\ell/\partial\beta, \quad D_{i,j} = \lim_{n \rightarrow \infty} [\sigma^{-2} x_i x'_j]. \tag{2.12}$$

So, A, B can be written as the 2×2 partitioned matrices $A = (a_{i,j} D_{i,j})$, $B = (b_{i,j} D_{i,j})$ for $\{a_{i,j}, b_{i,j}\}$ of (2.3), (2.4). If $\rho_{2,1} = 0$ (for example, if $\rho(e) \propto |e|^r$, r even) then $a_{1,2} = 0$ so that V, V^{-1} have the 2×2 partitions $V = (V_{i,j})$ with

$$V_{i,j} = c_{i,j} D^{i,j}, \quad D^{i,j} = D_{i,i}^{-1} D_{i,j} D_{j,j}^{-1}, \tag{2.13}$$

where

$$c_{i,j} = a_{i,i}^{-1} b_{i,j} a_{j,j}^{-1}. \tag{2.14}$$

Also $V^{-1} = (V^{i,j})$ with $V^{i,j} = A_{i,i} B^{i,j} A_{j,j}$, $(B^{i,j}) = B^{-1}$, so that CRs and p -values for smooth $t = t(\alpha) : \mathbb{R}^p \rightarrow \mathbb{R}^r$ are given by (2.8)-(2.11) with $V_t = c_{1,1} \dot{t}(\alpha) D_{1,1}^{-1} \dot{t}(\alpha)'$, and CRs and p -values for $t = t(\beta) : \mathbb{R}^{p'} \rightarrow \mathbb{R}^r$ are given by (2.8)-(2.11) with $V_t = c_{2,2} \dot{t}(\beta) D_{2,2}^{-1} \dot{t}(\beta)'$. For example, if $t(\beta) = \beta_2/\beta_1 = \delta$ say then $V_t = c_{2,2} \beta_1^{-2} (\delta^2 D_{2,2}^{1,1} - 2\delta D_{2,2}^{1,2} + D_{2,2}^{2,2})$, where $D_{i,i}^{-1} = (D_{i,i}^{a,b})$, and similarly for $t(\alpha) = \alpha_2/\alpha_1$. By Theorem 2.1 (II), if $\{\sigma_\ell\}$ do not depend on θ then (2.13), (2.14) remain true for $i = j = 1$ without the conditions $\rho_{1,1} = 1, \rho_{2,1} = 0$. By Theorem 2.1 (III), if $\{\mu_\ell\}$ do not depend on θ then (2.13), (2.14) remain true for $i = j = 2$ without the conditions $\rho_{1,0} = 0, \rho_{2,1} = 0$. For general ρ , (2.13) remains true if $\mu_\ell = x'_\ell \alpha, \sigma_\ell = x'_\ell \beta, c = a^{-1} b a^{-1}$ that is for c as in (2.7). In any of these cases it follows that

- (I) for a CR for any smooth function of α , an asymptotically optimal choice of ρ among a fixed selection is that which minimizes $\widehat{c}_{1,1}$ with respect to ρ , where $\widehat{c}_{1,1}$ is a consistent estimator of $c_{1,1}$;
- (II) for a CR for any smooth function of β , an asymptotically optimal choice of ρ among a fixed selection is that which minimizes $\widehat{c}_{2,2}$ with respect to ρ , where $\widehat{c}_{2,2}$ is a consistent estimator of $c_{2,2}$;
- (III) the asymptotic efficiencies of $\widehat{\alpha}$ and $\widehat{\beta}$ relative to the weighted MLEs are

$$\eta_1 = 1/\{\rho_{1,1,0}^* c_{1,1}\}, \quad \eta_2 = 1/\{(\rho_{1,1,2}^* - 1) c_{2,2}\}, \tag{2.15}$$

respectively, where the weighted MLEs are defined to be $\widehat{\alpha}$ and $\widehat{\beta}$ with $\rho(e) = -\log f(e)$ if f were known.

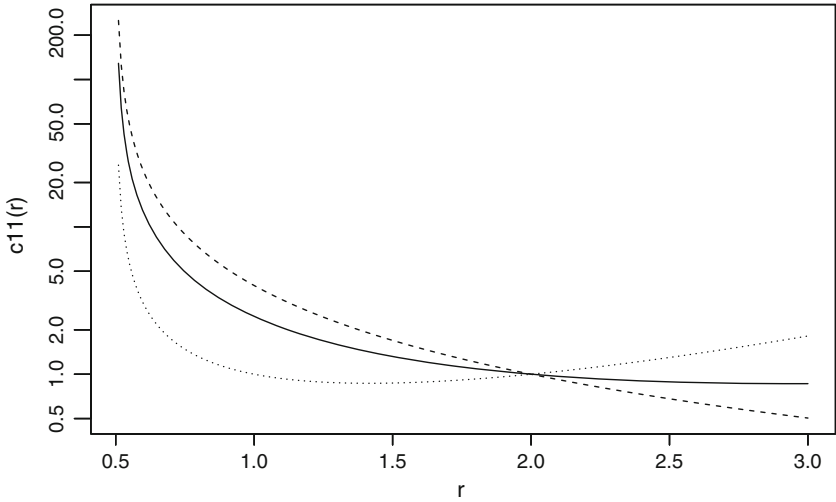


Figure 1: $c_{1,1}(r)$ for the normal (*solid curve*), uniform (*curve of dashes*), and double-exponential (*curve of dots*) distributions. The y axis is in log scale

For $\rho(e) = |e|^r/r$, r even, $c_{i,i}(r) = c_{i,i}$ are given by (2.18)-(2.19) to follow and are plotted against r in Figs. 1 to 4 for some examples along with η_1, η_2 .

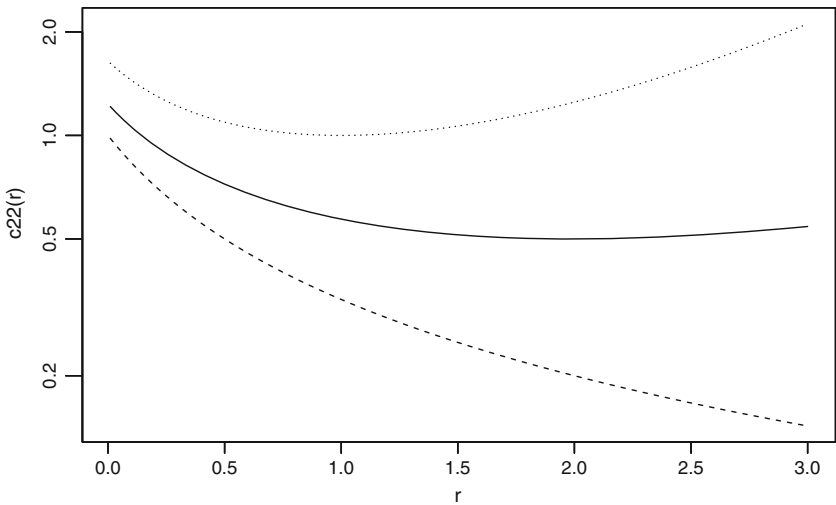


Figure 2: $c_{2,2}(r)$ for the normal (*solid curve*), uniform (*curve of dashes*), and double-exponential (*curve of dots*) distributions. The y axis is in log scale

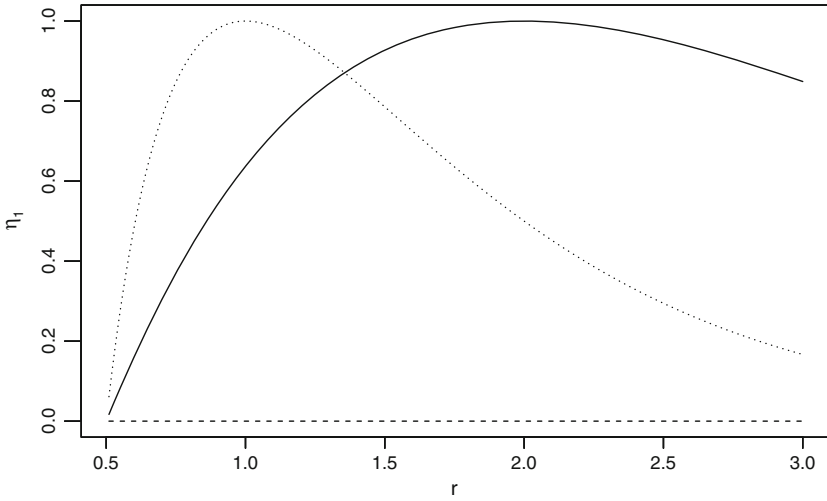


Figure 3: Asymptotic efficiency η_1 for the normal (solid curve), uniform (curve of dashes), and double-exponential (curve of dots) distributions

So, in any of these cases, a CR for smooth $t = t(\alpha) : \mathbb{R}^p \rightarrow \mathbb{R}^r$, or for smooth $t = t(\beta) : \mathbb{R}^{p'} \rightarrow \mathbb{R}^r$, is given by (2.8) with corresponding p-value (2.9) or if $r = 1$ by (2.10), (2.11), where for $t(\alpha)$, $V_t = \dot{t}V_{1,1}\dot{t}'$ with $\dot{t} = \partial t(\alpha)/\partial \alpha'$, $r \times p$, and for $t(\beta)$, $V_t = \dot{t}V_{2,2}\dot{t}'$ with $\dot{t} = \partial t(\beta)/\partial \beta'$, $r \times p'$.

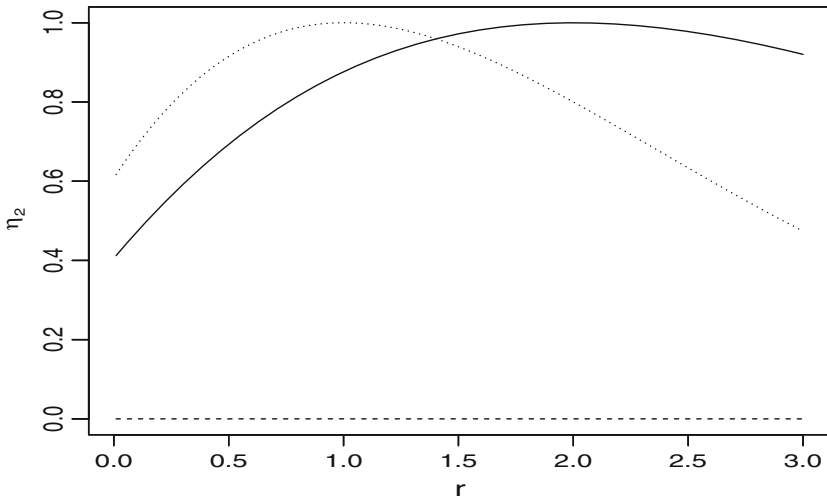


Figure 4: Asymptotic efficiency η_2 for the normal (solid curve), uniform (curve of dashes), and double-exponential (curve of dots) distributions

For an example, where $V_{i,j} = c_{i,j}D^{i,j}$ but c is not given by $a^{-1}ba^{-1}$, see (3.12), (3.13) to follow.

Suppose that $Y_\ell = m_\ell(\gamma) + \epsilon_\ell$ with $\{\epsilon_\ell\}$ i.i.d. $G(x)$. Then,

(i) $Y_\ell = \mu_\ell + e_\ell$ for $\mu_\ell = m_\ell + \xi$, $e_\ell = \epsilon_\ell - \xi$ and Theorem 2.1 (II) applies with $\theta = \alpha$, where $\alpha = \begin{pmatrix} \gamma \\ \xi \end{pmatrix}$.

(ii) $Y_\ell = \mu_\ell + \tau e_{\ell,\tau}$ for $e_{\ell,\tau} = e_\ell/\tau$ and Theorem 2.1 (I) applies with $\theta = \begin{pmatrix} \alpha \\ \tau \end{pmatrix}$. For general ρ , $\hat{\alpha}$ under (i) and (ii) may differ! But for $\rho(e) = |e|^r/r$, r even they are the same and

$$\tau = (\mathbb{E}[|e_1|^r])^{1/r}, \quad \hat{\tau} = \{\Lambda_{0,n}(\hat{\alpha})\}^{1/r}, \tag{2.16}$$

where

$$\Lambda_{0,n}(\alpha) = n^{-1} \sum_{\ell=1}^n |Y_\ell - \mu_\ell|^r.$$

By (2.13), $\hat{\alpha}, \hat{\tau}$ are asymptotically normal with covariance $V = (g_{i,j}G_{i,j})$, 2×2 , where $g_{i,j} = \tau^2 c_{i,j}$, $G_{1,1} = \lim_{n \rightarrow \infty} [\mu \cdot \mu \cdot]^{-1}$, $G_{1,2} = \lim_{n \rightarrow \infty} [\mu \cdot \mu \cdot]^{-1} [\mu \cdot]$, and $G_{2,2} = 1$.

(iii) More generally, suppose that $Y_\ell = m_\ell(\gamma) + s_\ell e_\ell$, $s_\ell = \sigma_{\ell,0} \exp(\gamma_2^j y_\ell)$ for known $\{y_\ell, \sigma_{\ell,0}\}$ with $\{\epsilon_\ell\}$ i.i.d. $G(x)$ and that $[y] = 0$. Then, for α as in (i), Theorem 2.1 (IV) applies with $\sigma_\ell = s_\ell$ so that $e_\ell = \epsilon_\ell - \xi$.

(iv) Theorem 2.1 (I) applies to $Y_\ell = \mu_\ell + \sigma_\ell e_{\ell,\tau}$ for $\sigma_\ell = s_\ell \tau$, $e_{\ell,\tau} = e_\ell$.

For $\rho(e) = |e|^r/r$, r even, $\hat{\alpha}$ under (iii) and (iv) are the same, and (2.16) holds with $|Y_\ell - \mu_\ell|^r$ in the expression for $\Lambda_{0,n}(\alpha)$ multiplied by s_ℓ^{-r} .

2.5. *Estimation of θ for $\rho(e) = |e|^r/d$.* Take $\rho(e) = |e|^r/d$, where r even, $d > 0$ are given constants. Suppose $\sigma_\ell = \tau s_\ell(\gamma)$ and $\mu_\ell = m_\ell(\gamma) + \xi s_\ell(\gamma) = \mu_\ell(\gamma, \xi)$. Set

$$g = d\tau^r,$$

$$K(\gamma, \xi) = n^{-1} \sum_{\ell=1}^n s_\ell(\gamma)^{-r} |Y_\ell - \mu_\ell(\gamma, \xi)|^r,$$

$$L(\gamma) = n^{-1} \sum_{\ell=1}^n \log s_\ell(\gamma).$$

Then, $\Lambda_n(\theta) = g^{-1}K(\gamma, \xi) + L(\gamma) + r^{-1} \log(g/d)$. So, changing d rescales τ but does not affect $\gamma, \xi, \hat{\gamma}, \hat{\xi}$. So, we may without loss of generality take $d = r$. Note that $\Lambda(\theta)$ is minimized over g (that is over τ) at $g = rK(\gamma, \xi)$, i.e., at

$$\tau \begin{pmatrix} \gamma \\ \xi \end{pmatrix} = \left\{ n^{-1} \sum_{\ell=1}^n |(Y_\ell - \mu_\ell) / s_\ell|^r \right\}^{1/r}. \tag{2.17}$$

So, finding $\hat{\theta}$ by iterating in θ -space as in the following note can be reduced to iterating in $\begin{pmatrix} \gamma \\ \xi \end{pmatrix}$ -space. If $r = 2$ we can also take out of the iteration process the linear parameters of μ_ℓ not in s_ℓ . Also,

$$\inf_{\theta_2} \Lambda_n(\theta) = r^{-1} + \Lambda_0(\gamma, \xi),$$

where $\Lambda_0(\gamma, \xi) = r^{-1} \log K(\gamma, \xi) + L(\gamma)$. Now $\rho^{(1)}(e) = |e|^{r-1} \text{sign}(e)$, and $\rho^{(2)}(e) = (r - 1)|e|^{r-2} + 2|e|^{r-1}\delta(e)$, where $\delta(e)$ is the Dirac function. So,

$$\rho_{1,1} = \mathbb{E}[|e_1|^r] = \int_{-\infty}^{\infty} |e|^r dF(e), \quad \rho_{1,0} = \int_{-\infty}^{\infty} |e|^{r-1} \text{sign}(e) dF(e),$$

$$a_{1,1} = \rho_{2,0} = (r - 1)\mathbb{E}[|e_1|^{r-2}],$$

$$a_{1,2} = \rho_{2,1} = (r - 1)\mathbb{E}[|e_1|^{r-1} \text{sign}(e_1)] = (r - 1)\rho_{1,0} = 0 \text{ if } \rho_{1,0} = 0,$$

$$a_{2,2} = \rho_{2,2} + \rho_{1,1} = r \text{ if } \rho_{1,1} = 1,$$

$$b_{1,1} = \rho_{1,1,0} - \rho_{1,0}^2 = \mathbb{E}[|e_1|^{2r-2}] - \rho_{1,0}^2 = \text{var}[|e_1|^{r-1} \text{sign}(e_1)],$$

$$b_{1,2} = \rho_{1,1,1} - \rho_{1,0}\rho_{1,1} = \mathbb{E}[|e_1|^{2r-1} \text{sign}(e_1)] - \rho_{1,0}\rho_{1,1}$$

$$= \text{covar}[|e_1|^r, |e_1|^{r-1} \text{sign}(e_1)],$$

$$b_{2,2} = \rho_{1,1,2} - \rho_{1,1}^2 = \mathbb{E}[|e_1|^{2r}] - 1 = \text{var}[|e_1|^r].$$

So, if $\mu_\ell = \mu_\ell(\alpha)$ and $\sigma_\ell = \sigma_\ell(\beta)$ then V is given by (2.13) in terms of $c_{i,j} = c_{i,j}(r)$, where

$$c_{1,1}(r) = (r - 1)^{-2} \left(\mathbb{E}[|e_1|^{r-2}] \right)^{-2} \mathbb{E}[|e_1|^{2r-2}], \tag{2.18}$$

$$c_{2,2}(r) = r^{-2} \left(\mathbb{E}[|e_1|^{2r}] - 1 \right). \tag{2.19}$$

So, $\rho_{1,1,2}$ is not robust to outliers. Nor is $\rho_{1,1,0}$. Note that $\rho_{2,0}$ is robust to outliers for $1 \leq r \leq 2$ but not for $r > 2$. These results are used in Appendix

A to provide Figs. 1 to 4. This also gives $c_{1,1}, c_{2,2}, \eta_1, \eta_2$ for the gamma distribution for $r = 2$. If $r = 2, \rho_{1,0} = 0$ and $\rho_{1,1} = 1$ then $\mathbb{E}[e_1] = 0, \mathbb{E}[e_1^2] = 1, b_{1,1} = 1, b_{1,2} = \mathbb{E}[e_1^3], b_{2,2} = \mathbb{E}[e_1^4] - 1, a_{1,1} = 1, a_{1,2} = 0, a_{2,2} = 2, c_{1,1} = 1, c_{1,2} = \mathbb{E}[e_1^3/2]$ and $c_{2,2} = (\mathbb{E}[e_1^4] - 1)/4$.

Some other choices of ρ are discussed in Appendix D.

Provided $\Lambda_{n\cdots}$ is continuous near $\hat{\theta}, \hat{\theta}$ can be found by Newton's method (for example, one can use the NAG routine E04LBF), that is by iteration from an initial value $\theta_{(0)}$ in terms of $e_\ell = e_\ell(\theta)$ of (1.8):

$$\theta_{(i+1)} = \theta - \Lambda_{n\cdots}^{-1} \Lambda_n.$$

at $\theta = \theta_{(i)}$, where

$$\Lambda_n = n^{-1} \sum_{\ell=1}^n \lambda_\ell.$$

for $\lambda_\ell = (\lambda_{\ell,i}), q \times 1$ of (F.2), (F.3), and

$$\Lambda_{n\cdots} = n^{-1} \sum_{\ell=1}^n \lambda_{\ell\cdots}$$

for $\lambda_{\ell\cdots} = (\lambda_{\ell,i,j}), q \times q$ of (F.4), (F.5). Convergence to $\hat{\theta}$ is quadratic, i.e., $|\theta_{(i+1)} - \hat{\theta}| = O(|\theta_{(i)} - \hat{\theta}|^2)$. It is often simpler is to replace $\Lambda_{n\cdots}$ by \hat{A} , a consistent estimator of its mean $A(\theta, F) = A$. For the L_r -estimator with r even one can use $\hat{A} = A(\hat{\theta}, \hat{F})$. If $r = 2, A$ does not depend on F .

2.6. *Estimation of θ for $\rho(e) = |e|^r/r$.* Suppose that $\rho(e) = |e|^r/r$, where r is even and that we can reparameterize θ so that $\mu = (\mu_\ell)$ and $\sigma = (\sigma_\ell)$ have different parameters, say $\mu_\ell = \mu_\ell(\alpha), \sigma_\ell = \sigma_\ell(\beta), \theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Then, $a_{1,2} = \rho_{2,1} = 0, A = \text{diag}(A_{1,1}, A_{2,2})$ with $A_{i,i} = a_{i,i} D_{i,i}$ for $D_{i,i}$ of (2.12), $a_{1,1} = (r - 1)\mathbb{E}[|e_1|^{r-2}], a_{2,2} = r$. We can iterate using

$$\begin{aligned} \alpha_{(i+1)} &= \alpha - \hat{A}_{1,1}^{-1} \Lambda_{n\cdot\alpha} \text{ at } \theta = \theta_{(i)}, \\ \beta_{(i+1)} &= \beta - \hat{A}_{2,2}^{-1} \Lambda_{n\cdot\beta} \text{ at } \theta = \theta_{(i)}, \end{aligned}$$

where

$$\begin{aligned} \Lambda_{n\cdot\alpha} &= -n^{-1} \sum_{\ell=1}^n |e_\ell(\theta)|^{r-1} \text{sign}(e_\ell(\theta)) \mu_{\ell\cdot\alpha} / \sigma_\ell, \\ \Lambda_{n\cdot\beta} &= n^{-1} \sum_{\ell=1}^n (1 - |e_\ell(\theta)|^r) \sigma_{\ell\cdot\beta} / \sigma_\ell. \end{aligned}$$

If $\sigma_\ell = \exp(y'_\ell \beta)$ this gives $\beta_{(i+1)} = \beta - \widehat{A}_{2,2}^{-1} [yy']^{-1} \{[y] - [y|e(\theta)|^r]\}$ at $\theta = \theta_{(i)}$. If $r = 2$ one has an exact formula for the linear parameters of μ . For example, if $r = 2$ and $\mu_\ell = x'_\ell \alpha$, then instead of iterating using $\alpha_{(i+1)}$ above, one uses the exact formula

$$\alpha(\beta) = [\sigma^{-2}xx']^{-1} [\sigma^{-2}xY]. \tag{2.20}$$

If $\sigma_\ell = y'_\ell \beta$ with $y_{\ell,1} \equiv 1$, then $D_{2,2} = \lim_{n \rightarrow \infty} [\sigma^{-2}yy']$ and iteration can begin at $\alpha = \alpha_{(0)}$, $\beta_1 = \tau(\alpha_{(0)})$ and $\beta_2 = \dots = \beta_{p'} = 0$, where $\alpha_{(0)} = [x_1x'_1]^{-1} [x_1Y]$, the unscaled LSE, and $\beta_{1,0}(\alpha)$ minimizes Λ_n at $\beta_2 = \dots = \beta_{p'} = 0$, i.e., $\tau(\alpha) = \{n^{-1} \sum_{\ell=1}^n |Y_\ell - \mu_\ell|^r\}^{1/r}$. The same is true if $\sigma_\ell = \exp(y'_\ell \beta)$ with $y_{\ell,1} \equiv 1$, except that $D_{2,2} = \lim_{n \rightarrow \infty} [yy']$ and $\tau(\alpha)$ is replaced by $\log \tau(\alpha)$.

2.7. *Asymptotic Behavior of the Unscaled M-estimator.* We end this section with the asymptotic behavior of the usual unscaled M -estimator when the scale is changing, for the case $\rho(e) = |e|^r/r$, r even. The proof is similar to that for the case when the model contains no change in scaling.

Theorem 2.2. *Suppose the observations satisfy the location-scale model (1.1) with F having finite moments and with all relevant matrices invertible. Suppose $\rho(e) = |e|^r/d$ for some r even, $d > 0$. Let $\hat{\theta}$ minimize $\sum_{\ell=1}^n \rho(Y_\ell - \mu_\ell(\theta))$ or more generally $\sum_{\ell=1}^n \rho(Y_\ell - \mu_\ell(\theta))$. Let F be centered so that $\int_{-\infty}^{\infty} \rho^{(1)}(e)dF(e) = 0$. Then, (2.5) holds with*

$$A_{i,j} = \lim_{n \rightarrow \infty} n^{-1} \sum_{\ell=1}^n \mu_{\ell,i} \mu_{\ell,j} \rho_{\ell,2,0},$$

$$B_{i,j} = \lim_{n \rightarrow \infty} n^{-1} \sum_{\ell=1}^n \mu_{\ell,i} \mu_{\ell,j} \rho_{\ell,1,1,0},$$

where

$$\rho_{\ell,2,0} = \int_{-\infty}^{\infty} \rho^{(2)}(\sigma_\ell x) dF(x) = \sigma_\ell^{r-2} \rho_{2,0},$$

$$\rho_{\ell,1,1,0} = \int_{-\infty}^{\infty} \rho^{(1)}(\sigma_\ell x)^2 dF(x) = \sigma_\ell^{2r-2} \rho_{1,1,0}.$$

So, for $\rho(e) \propto |e|^r$, r even, if $\{\sigma_\ell\}$ is not changing much compared with μ_ℓ , then one does not lose much efficiency by estimating the parameters of μ_ℓ in this way. This can be useful for selecting a model for μ_ℓ before focusing on a model for $\{\sigma_\ell\}$.

3 When Trend and Variability have Different Parameters

Here, we consider the case when θ can be reparameterized so that the parameters of the trend $\{\mu_\ell\}$ will be different from those of the variability $\{\sigma_\ell\}$. Call them α in \mathbb{R}^p and β in $\mathbb{R}^{p'}$ say as in our examples above. So,

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \mu_\ell = \mu_\ell(\alpha), \sigma_\ell = \sigma_\ell(\beta).$$

By Example 2.1, we can write A, B of Theorem 2.1 as 2×2 partitioned matrices

$$A \text{ or } B = (\mathcal{F}_{i,j} D_{i,j}) \tag{3.1}$$

with $D_{i,j} = \lim_{n \rightarrow \infty} [\sigma^{-2} x_i x_j']$, $x_{\ell,1} = \partial \mu_\ell / \partial \alpha$, $x_{\ell,2} = \partial \sigma_\ell / \partial \beta$, and

$$\mathcal{F}_{i,j} = \begin{cases} a_{i,j} \text{ of (2.3) for } A, \\ b_{i,j} \text{ of (2.4) for } B, \end{cases}$$

so one may write V as a 2×2 partitioned matrix explicitly in terms of $\{a_{i,j}, b_{i,j}, D_{i,j}\}$. The general results are messy: see Appendix B. But first we consider some special cases. Set

$$C_1 = D_{1,2} D_{2,2}^{-1} D_{2,1}, \quad C_2 = D_{2,1} D_{1,1}^{-1} D_{1,2}. \tag{3.2}$$

Example 3.1. *Suppose $\rho_{2,1} = 0$. By Example 2.1,*

$$V_{i,j} = c_{i,j} D^{i,j}, \tag{3.3}$$

where

$$c_{i,j} = a_{i,i}^{-1} b_{i,j} a_{j,j}^{-1}, \quad D^{i,j} = D_{i,i}^{-1} D_{i,j} D_{j,j}^{-1},$$

so CIs for α_i and β_j of level about 0.95 are

$$n^{1/2} |\widehat{\alpha}_i - \alpha_i| < 1.96 \widehat{\sigma}(\widehat{\alpha}_i), \tag{3.4}$$

$$n^{1/2} \left| \widehat{\beta}_j - \beta_j \right| < 1.96 \widehat{\sigma} \left(\widehat{\beta}_j \right), \tag{3.5}$$

where $\hat{\sigma}(\hat{\alpha}_i)^2$ and $\hat{\sigma}(\hat{\beta}_j)^2$ are consistent estimators of $\sigma(\hat{\alpha}_i)^2 = c_{1,1} \left(D_{1,1}^{-1} \right)_{i,i}$ and $\sigma(\hat{\beta}_j)^2 = c_{2,2} \left(D_{2,2}^{-1} \right)_{j,j}$, respectively. The p -value of $(H_0 : \alpha_i = 0)$ or $(H_0 : \beta_j = 0)$ is given by (2.11) with $(t, V_t) = \left(\alpha_i, \sigma(\hat{\alpha}_i)^2 \right)$ or $\left(\beta_j, \sigma(\hat{\beta}_j)^2 \right)$. If for some scalar λ ,

$$C_1 = \lambda D_{1,1}, \quad C_2 = \lambda D_{2,2} \tag{3.6}$$

then

$$V^{-1} = (c^{i,j} D_{i,j}), \quad 2 \times 2, \tag{3.7}$$

where

$$c^{1,1} = \left(c_{1,1} - \lambda c_{1,2}^2 c_{2,2}^{-1} \right)^{-1}, \tag{3.8}$$

$$c^{2,2} = \left(c_{2,2} - \lambda c_{1,2}^2 c_{1,1}^{-1} \right)^{-1}, \tag{3.9}$$

$$c^{1,2} = c^{2,1} = -c^{1,1} c_{1,2} c_{2,2}^{-1} = -c^{2,2} c_{1,2} c_{1,1}^{-1}. \tag{3.10}$$

So, if $\lambda = 1$ then $(c^{i,j}) = (c_{i,j})^{-1}$, i.e.,

$$c^{i,j} = a_{i,i} b^{i,j} a_{j,j} \tag{3.11}$$

for $(b^{i,j}) = (b)^{-1}$, i.e., $b^{1,1} = b_{2,2}/d_b$, $b^{2,2} = b_{1,1}/d_b$, $b^{1,2} = -b_{1,2}/d_b$ and $d_b = b_{1,1}b_{2,2} - b_{1,2}^2$.

If $p = p'$ and $\det D_{1,2} \neq 0$ then $C_1 = \lambda D_{1,1}$ if and only if $C_2 = \lambda D_{2,2}$.

Example 3.2. Suppose we have p' independent random samples of sizes $n_1, \dots, n_{p'}$. Set $n = \sum_{j=1}^{p'} n_j$. Suppose the observations in the j th sample are $Y_{\ell,j} = \mu_j(\alpha) + \beta_j e_{\ell,j}$, $1 \leq \ell \leq n_j$, for $1 \leq j \leq p'$ with the n combined residuals $\{e_{\ell,j}\}$ having the same cdf F . Suppose $\rho_{2,1} = 0$. By (3.3),

$$n^{1/2} \left(\hat{\beta} - \beta \right) \sim \mathcal{N}_{p'}(0, V_{2,2}),$$

where $V_{2,2} = c_{2,2} D_{2,2}^{-1}$, $D_{2,2} = \lim_{n \rightarrow \infty} \text{diag} \left(\lambda_j \beta_j^{-2} \right)$ and $\lambda_j = n_j/n$. That is, $\{\hat{\beta}_j\}$ are asymptotically independent normals with means $\{\beta_j\}$ and variances $\{\beta_j^2/n_j\}$. Now suppose that $\rho(e) = e^2/2$ and $\mu_j(\alpha) = z'_j \alpha$, where $z_1, \dots, z_{p'}$

are given. Then, $(\hat{\alpha}, \hat{\beta})$ are given by iterating between $\hat{\alpha} = \alpha(\hat{\beta})$ and $\hat{\beta} = \beta(\hat{\alpha})$, where

$$\alpha(\beta) = \left(\sum_{j=1}^{p'} \lambda_j z_j z'_j \beta_j^{-2} \right)^{-1} \sum_{j=1}^{p'} \left(\lambda_j z_j [Y]_j \beta_j^{-2} \right),$$

$$\beta_j(\alpha) = [Y^2]_j - 2[Y]_j z'_j \alpha + (z'_j \alpha)^2,$$

where $[Y^k]_j = n_j^{-1} \sum_{\ell=1}^{n_j} Y_{\ell,j}^k$.

If for some scalar λ , $C_1 = \lambda D_{1,1}$, (for example, $p = 1$), then by (B.1)

$$V_{1,1} = c_{1,1} D_{1,1}^{-1},$$

where

$$c_{1,1} = a_{1,1}^{-2} (1 - \lambda r)^{-2} (b_{1,1} + l_1 \lambda r), \quad r = a_{1,2}^2 a_{1,1}^{-1} a_{2,2}^{-1}. \tag{3.12}$$

If for some scalar λ , $C_2 = \lambda D_{2,2}$, (for example, $p' = 1$), then by (B.2)

$$V_{2,2} = c_{2,2} D_{2,2}^{-1},$$

where

$$c_{2,2} = a_{2,2}^{-2} (1 - \lambda r)^{-2} (b_{2,2} + l_2 \lambda r). \tag{3.13}$$

Suppose (3.6) holds. Then, by (B.1)-(B.3), $V_{i,j}$ is given by (3.3) with $c_{i,i}$ of (3.12), (3.13) and

$$c_{1,2} = c_{1,2} = a_{1,1}^{-1} a_{2,2}^{-1} (1 - \lambda r)^{-2} (l_{1,2} + \lambda r b_{1,2}).$$

Also V^{-1} is given by (3.7)-(3.10). If $\lambda = 1$ then these expressions for $c_{i,j}$, $c^{i,j}$ reduce to those in Section 2.2.

Example 3.3. Suppose $p' = p$, $\mu_\ell = x'_\ell \alpha$ and $\sigma_\ell = x'_\ell \beta$. So, $x_{\ell,1} = x_{\ell,2} = x_\ell$. Then, (3.6) holds with $\lambda = 1$, $D_{i,j} \equiv D_{1,1} = \lim_{n \rightarrow \infty} n^{-1} \sum_{\ell=1}^n (x'_\ell \beta)^{-2} x_\ell x'_\ell$, $V = c \otimes D_{1,1}^{-1}$, and $V^{-1} = c^{-1} \otimes D_{1,1}$ with c and c^{-1} given in Section 2.2 or if $\rho_{2,1} = 0$ by (2.14), (3.11). The distributions of $(\hat{\alpha} - \alpha) / \beta_1$ and $\hat{\beta} / \beta_1$, \hat{e}_ℓ depend on θ only through $(\beta_2, \dots, \beta_p) / \beta_1$. If $\rho_{2,1} = \rho_{1,1,1} = 0$, then $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically independent. This holds if $\rho(e)$ and $F(e)$ are symmetric about zero: such models may be called symmetric semi-parametric models.

An important application of this example is for μ_ℓ and σ_ℓ linear in ℓ :

Example 3.4. Set $z_\ell = \ell/n$. Suppose $\mu_\ell = \alpha_1 + \alpha_2 u_\ell$, $\sigma_\ell = \beta_1 + \beta_2 u_\ell$ with $u_\ell = z_\ell$ or $z_\ell - [z]$. So, Example 3.3 holds with $x_\ell = (1, u_\ell)'$. Also $\beta_1 > 0$ and $\beta_1 + \beta_2 > 0$ since $\sigma_\ell > 0$. Put $\delta = \beta_2/\beta_1$. As $n \rightarrow \infty$, $D_{1,1} \rightarrow \beta_1^{-2}G$ or $\beta_1^{-2}H$, respectively, for G, H of Appendix C. So, if $\rho_{2,1} = 0$ then in the CIs (3.4), (3.5) for α_2, β_2 , we may take $\sigma(\widehat{\alpha}_2)^2 = c_{1,1}\beta_1^2 U_{2,1}^{-1}$, $\sigma(\widehat{\beta}_2)^2 = c_{2,2}\beta_1^2 U_{2,1}^{-1}$, for $U = G$ or H , respectively, where for $L = (L_{i,j})$ a 2×2 partitioned matrix,

$$(L^{i,j}) = L^{-1}, L_{2,1} = L_{2,2} - L_{2,1}L_{1,1}^{-1}L_{1,2}, L_{1,2} = L_{1,1} - L_{1,2}L_{2,2}^{-1}L_{2,1}. \tag{3.14}$$

In the p -value of (2.11) for $(H_0 : \beta_2 = 0)$ we can take $q_n = n^{1/2}\widehat{\delta}/(12\widehat{c}_{2,2})^{1/2}$. The distribution of q_n is determined by n, F, δ . For $t(\beta) = \beta_2/\beta_1 = \delta$, and $x_\ell = (1, z_\ell)'$, $V_t = c_{2,2}\beta_1^2 v(\delta)$, where $v(\delta) = (2 - (1 + \delta)^{-1}) / \det G$.

Example 3.5. Suppose that $\rho_{2,1} = 0$, $\mu_\ell = \alpha_1 + d_\ell(\gamma)$, $\sigma_\ell = \tau s_\ell(\delta)$, $\alpha = \begin{pmatrix} \alpha_1 \\ \gamma \end{pmatrix}$, $\beta = \begin{pmatrix} \tau \\ \delta \end{pmatrix}$ and $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$. Set $d_\ell = \partial d_\ell / \partial \gamma$ and $s_\ell = \partial s_\ell / \partial \delta$. By Example 2.1, $V = (V_{i,j})$, 2×2 is given by (2.13) in terms of $D_{i,j}$ which we partition as $(D_{i,j,k,l}, 1 \leq k, l \leq 2)$ with $D_{1,1,1,1} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-2}]$, $D_{1,1,1,2} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-2}d.]'$, $D_{1,1,2,2} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-2}d.d.]'$, $D_{2,2,1,1} = \tau^{-2}$, $D_{2,2,1,2} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-2}s.]'$, $D_{2,2,2,2} = \lim_{n \rightarrow \infty} [s^{-2}s.s.]'$, $D_{1,2,1,1} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-1}]$, $D_{1,2,1,2} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1}s.]'$, $D_{1,2,2,1} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1}d.]$, $D_{1,2,2,2} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-2}d.s.]'$. So, with $P = p - 1, Q = p' - 1$,

$$n^{1/2}(\widehat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}_P(0, c_{1,1}G_{2,1}^{-1}), \tag{3.15}$$

$$n^{1/2}(\widehat{\delta} - \delta) \xrightarrow{d} \mathcal{N}_Q(0, c_{2,2}H_{2,1}^{-1}), \tag{3.16}$$

in the notation of (3.14), where $G = D_{1,1}$ and $H = D_{2,2}$, so that

$$G_{2,1} = \tau^{-2} \lim_{n \rightarrow \infty} ([s^{-2}d.d.]' - [s^{-2}d.] [s^{-2}d.]' / [s^{-2}]),$$

a function of τ, γ, δ and

$$H_{2,1} = \lim_{n \rightarrow \infty} ([s^{-2}s.s.]' - [s^{-1}s.] [s^{-1}s.]'),$$

a function of δ . So, a CR for γ and a p -value for $(H_0 : \gamma = 0)$ are given by $n(\gamma - \widehat{\gamma})' \widehat{G}_{2,1}(\gamma - \widehat{\gamma}) \leq \widehat{c}_{1,1} \chi_{p-1,0.95}^2$ and $p_n = p(\gamma) \approx Ch_{p-1}(q_n)$ for

$q_n = n\widehat{\gamma}'\widehat{G}_{2,1}\widehat{\gamma}/\widehat{c}_{1,1}$. The same holds for δ with $p - 1$, $G_{2,1}$, $c_{1,1}$ replaced by $p' - 1$, $H_{2,1}$, $c_{2,2}$. For μ_ℓ linear, say $d_\ell = x'_\ell\gamma$, one replaces d . above by x . For μ_ℓ exponential, say $d_\ell = \exp(x'_\ell\gamma)$, one replaces d . above by qx . For σ_ℓ linear, say $s_\ell = y'_\ell\delta$, one replaces s . above by y . For σ_ℓ exponential, say $s_\ell = \exp(y'_\ell\delta)$, one replaces s . above by y so that $H_{2,1} = \lim_{n \rightarrow \infty} ([yy'] - [y][y'])$ is known.

We end this section with some simulation result.

3.1. *Simulation Results.* The quantile (2.11) was calculated for $t = \beta_2$ of Example 3.4 with $\rho(e) = |e|^r/r$ for $r = 2$ on simulated data with $n = 100$, $y_\ell = x_\ell = (1, \ell/n)'$ and $e_1 = U(-1, 1)/\tau_r$, where $\tau_r = (r + 1)^{-r}$ and $c_{2,2}(F) = c_{2,2} = \left(\int_{-\infty}^{\infty} |x|^{2r} dF(x) - 1\right) / r^2$ was estimated by $c_{2,2}(\widehat{F})$. Iteration was stopped when the change in each of the four parameter values was less than one percent. This took very few iterations. For $\delta = 0.5$ and 0.2 , we obtained $p_n = 0.9995$ and 0.45 . So, the hypothesis $\beta_2 = 0$, i.e., $\delta = 0$, would be rejected in the case $\delta = 0.5$, and accepted in the case $\delta = 0.2$. For $\theta' = (20, 10, 10, 5)$, we obtained $\widehat{\theta}' = (19.7, 12.7, 8.1, 7.0)$ and $\widehat{c}_{2,2} = 0.289$ compared to $c_{2,2} = 0.2$. For $\theta' = (20, 20, 10, 2)$, we obtained $\widehat{\theta}' = (19.95, 20.61, 10.74, -0.210)$ and $\widehat{c}_{2,2} = 0.247$. We are grateful to C. P. Pearson of National Institute of Water and Atmospheric Research, Christchurch for these calculations.

4 Likelihood Ratio Type Tests

Although we have provided CRs and tests for θ in Section 2, the convenience of likelihood ratio (LR) tests raises the question whether tests like this exist for M -estimators. Suppose $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ with θ_i a p_i vector and we wish to test $H_0 : \theta_1 = 0$ versus $H_1 : \theta_1 \neq 0$. Let the M -estimator of θ under H_0 and H_1 be $\widehat{\theta}_0 = \begin{pmatrix} 0 \\ \widehat{\theta}_{2,0} \end{pmatrix}$ and $\widehat{\theta}$. How does $\Lambda_n(\widehat{\theta}) - \Lambda_n(\widehat{\theta}_0)$ behave? Note that

$$\begin{aligned} \Lambda_n(\widehat{\theta}) &= \Lambda_n + \Lambda_n \cdot \Delta + \Delta' \Lambda_n \dots \Delta / 2 + O_p(n^{-3/2}), \\ 0 &= \Lambda_n \cdot (\widehat{\theta}) = \Lambda_n \cdot + \Lambda_n \dots \Delta + O_p(n^{-1}), \end{aligned}$$

so

$$\begin{aligned} \Lambda_n(\widehat{\theta}) &= \Lambda_n - \Lambda_n' \cdot \Lambda_n^{-1} \Lambda_n \dots \Delta / 2 + O_p(n^{-3/2}) \\ &= \Lambda_n - \Lambda_n' \cdot A^{-1} \Lambda_n \dots \Delta / 2 + O_p(n^{-3/2}). \end{aligned}$$

But $\Lambda_n = n^{-1/2}BN + O_p(n^{-1})$, where $N \sim \mathcal{N}_p(0, I)$. So, $\Lambda_n(\hat{\theta}) = \Lambda_n - n^{-1}N'CN/2 + O_p(n^{-3/2})$ for $C = B^{1/2}A^{-1}B^{1/2}$. Similarly, under H_0 ,

$$\begin{aligned} \Lambda_n(\hat{\theta}) &= \Lambda_n - \Lambda'_{n,\theta_2}A_{2,2}^{-1}\Lambda_{n,\theta_2}/2 + O_p(n^{-3/2}) \\ &= \Lambda_n - n^{-1}N'_2DN_2/2 + O_p(n^{-3/2}) \end{aligned}$$

for $D = B_{2,2}^{1/2}A_{2,2}^{-1}B_{2,2}^{1/2}$. So, setting $R_n = 2n \{ \Lambda_n(\hat{\theta}_0) - \Lambda_n(\hat{\theta}) \}$, under H_0 , $R_n = R + O_p(n^{-1/2})$, where $R = N'CN - N'_2DN_2 = N'\Omega N$ say. So, R has the same distribution as $\sum_{j=1}^p \lambda_j N_j^2$, where $\{\lambda_j\}$ are the eigenvalues of Ω . So, the distribution of R may be obtained by inverting its characteristic generating function $K_R(t) = -2^{-1} \log |I - 2\Omega t| = -2^{-1} \log(1 - 2\lambda_j t)$ for $\{\lambda_j Re(t) < 1/2\}$. By replacing Ω by its estimator one can now estimate the distribution of R_n and so obtain p -values for H_0 . However, we now show that in several special cases R is proportional to $\chi_{p_1}^2$, giving us an analogue of the likelihood-ratio test.

Suppose A and B are proportional as in Corollary 2.1, say

$$A = \nu B. \tag{4.1}$$

Then,

$$\Omega = \begin{pmatrix} \nu^{-1}I_{p_1} & 0 \\ 0 & 0 \end{pmatrix}, \tag{4.2}$$

so under H_0 , if $\hat{\nu}$ is a consistent estimator of ν then

$$\hat{\nu}R_n \xrightarrow{d} \chi_{p_1}^2 \tag{4.3}$$

as $n \rightarrow \infty$. From Theorem 2.1, we have

Corollary 4.1. *We have the following:*

- (i) *If $\rho_{1,0} = 0$ and $\{\sigma_\ell\}$ do not depend on θ then (4.1) holds with $\nu = a_{1,1}/b_{1,1} = \rho_{2,0}/\rho_{1,1,0}$.*
- (ii) *If $\rho_{1,1} = 1$ and $\{\mu_\ell\}$ do not depend on θ then (4.1) holds with $\nu = a_{2,2}/b_{2,2} = (\rho_{2,2} + 1) / (\rho_{1,1,2} - 1)$.*
- (iii) *Suppose $\sigma_\ell(\theta) \equiv \nu_1 + \nu_2\mu_\ell(\theta)$. Then, by Corollary 2.1, (4.1) holds with $\nu = (a_{1,1} + 2a_{1,2}\nu_2 + a_{2,2}\nu_2^2) / (b_{1,1} + 2b_{1,2}\nu_2 + b_{2,2}\nu_2^2)$.*

We now give a case, where A, B are not proportional but R is still proportional to $\chi_{p_1}^2$.

Suppose also that $\rho_{2,1} = \rho_{1,1,1} = 0$, as is the case for F and ρ symmetric about zero, and the parameters of $\{\mu_\ell\}$ and $\{\sigma_\ell\}$ are different, say $\mu_\ell = \mu_\ell(\alpha)$, $\sigma_\ell = \sigma_\ell(\beta)$ and $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, then $A = \text{diag}(a_{1,1}M, a_{2,2}S)$, $B = \text{diag}(b_{1,1}M, b_{2,2}S)$ and $V = \text{diag}(c_{1,1}M^{-1}, c_{2,2}S^{-1})$, where $M = \lim_{n \rightarrow \infty} [\sigma^{-2} \mu_{\cdot\alpha} \mu'_{\cdot\alpha}]$, $S = \lim_{n \rightarrow \infty} [\sigma^{-2} \mu_{\cdot\beta} \mu'_{\cdot\beta}]$, $c_{1,1} = \rho_{2,0}^{-2} \rho_{1,1,0}$ and $c_{2,2} = (\rho_{2,2} + 1)^{-2} (\rho_{1,1,2} - 1)$. Now suppose $\theta_1 = \alpha_1$ and $\theta_2 = \begin{pmatrix} \alpha_2 \\ \beta \end{pmatrix}$ for $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$. Then, setting $p_2 = \dim(\alpha) - p_1$, $C = \text{diag}(b_{1,1}a_{1,1}^{-1}I_{p_2}, b_{2,2}a_{2,2}^{-1}I_{p'})$ and $D = \text{diag}(b_{1,1}a_{1,1}^{-1}I_{p_2}, b_{2,2}a_{2,2}^{-1}I_{p'})$, (4.2) and (4.3) hold with $\nu = a_{1,1}/b_{1,1} = \rho_{2,0}/\rho_{1,1,0}$. Now suppose instead that $\theta_1 = \beta_1$ and $\theta_2 = \begin{pmatrix} \alpha \\ \beta_2 \end{pmatrix}$ for $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$. Then, (4.2) and (4.3) hold with $\nu = a_{2,2}/b_{2,2} = (\rho_{2,2} + 1) / (\rho_{1,1,2} - 1)$. For the LSE, $a_{1,1}/b_{1,1} = 1$ and $a_{2,2}/b_{2,2} = 2 / (\mathbb{E}[e_1^4] - 1)$. Finally, if $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ and we wish to test $H_0 : \alpha_1 = 0, \beta_1 = 0$, where α_1 and β_1 have dimensions p_1 and q_1 , then $R = b_{1,1}a_{1,1}^{-1}\chi_{p_1}^2 + b_{2,2}a_{2,2}^{-1}\chi_{q_1}^2$.

5 Model Constraints

Here, we investigate when μ and σ can be reparameterized with different parameters. If we want θ not to depend on our choice of ρ , then only certain types of models are available: those equivalent by reparameterization to the *fundamental form*

$$\mu_\ell(\theta) = m_\ell(\gamma) + \xi s_\ell(\gamma), \quad \sigma_\ell(\theta) = \tau s_\ell(\gamma), \tag{5.1}$$

where $m_\ell(\gamma)$ contains no location parameters, and $s_\ell(\gamma)$ contains no scale parameters. For example, for (x_ℓ, y_ℓ) given vector covariates, we shall see that the model

$$\mu_\ell = \theta_1 + x'_\ell \theta_2, \quad \sigma_\ell = \theta_3 + y'_\ell \theta_4$$

is only available if we can take $y_\ell = x_\ell$ or if $\dim(y_\ell) = 1$.

Let us call a model a *fundamental model* if its parameters should not depend on the choice of estimator. However, in practice this requirement is violated by the location parameter of the model unless we unrealistically

restrict ourselves to residuals symmetric about zero. Consider the usual model (no scale change)

$$Y_\ell = \mu_\ell + e_\ell \tag{5.2}$$

for $1 \leq \ell \leq n$ with $\mu_\ell = \xi + m_\ell(\gamma)$ and e_1, \dots, e_n i.i.d. F say. Here, $m_\ell(\gamma)$ contains no location parameters, otherwise, ξ would not be identifiable, and θ is any one-to-one function of (ξ, γ') .

Example 5.1. *The model (5.2) with $\mu_\ell = x'_\ell \theta$, that is $\mu = X\theta$ say with θ in \mathbb{R}^p is allowed if there exists H , $p \times p$ with $\det(H) \neq 0$ such that XH has ones in its first column. That is, this model is allowed if X has rank $p \leq n$.*

We call (5.2) a *working model*. We can write (5.2) as a fundamental model in the form

$$Y_\ell = m_\ell(\gamma) + e_\ell^*$$

with e_1^*, \dots, e_n^* i.i.d. G say. So, $e_\ell = e_\ell^* - \xi$, and $F(e) = G(e + \xi)$ depend on ρ through ξ , and $\xi = \xi(\rho, G)$ is a functional of ρ, G defined by $0 = \rho_{1,0} = \int_{-\infty}^\infty \rho^{(1)}(e) dG(e + \xi)$. For example, if $\rho(e) = e^2/2$, then $\xi(\rho, G) = \text{mean}(G)$.

So, the working model has one more parameter than the fundamental model. The working model has one constraint on its error cdf F , while the fundamental model has no constraints on its error cdf G . If we were to delete the location parameter from our working model, we would be restricting ourselves to a special class of (ρ, F) , a class that includes (ρ, F) both symmetric about zero, since then $\rho_{1,0} = 0$. In this paper, we do not want to restrict ourselves to F symmetric about zero, since this is an artificial constraint seldom satisfied in reality. The simplest example of such a restricted model is regression through the origin with one independent variable:

Example 5.2. *$Y_\ell = x_\ell \theta + e_\ell$ with $\theta \in \mathbb{R}$, $x_\ell = \ell/n$ say, and $\{e_\ell\}$ i.i.d. F symmetric about zero. It has no location parameters. Assume ρ is symmetric about zero so $\rho_{1,0} = 0$. Then, $n^{1/2} (\hat{\theta} - \theta) \sim \mathcal{N}_1(0, V(F))$ for $V(F) = c_{1,1} / \lim_{n \rightarrow \infty} [x^2]$, where $c_{1,1} = \rho_{1,1,0} \rho_{2,0}^{-2}$. So, for smooth V , an approximate 95 percent CI for $\hat{\theta}$ is $\hat{\theta} \pm 1.96 \left(V(\hat{F}) / n \right)^{1/2}$, where \hat{F} is the empirical cdf of $\{\hat{e}_\ell = Y_\ell - x_\ell \hat{\theta}\}$.*

Now let us return to our working model $Y_\ell = \mu_\ell + \sigma_\ell e_\ell$ with $\{e_\ell\}$ i.i.d. F . The constraints $\rho_{1,0} = 0, \rho_{1,1} = 1$ imply that *either* we restrict ρ and F to be symmetric about zero *or* we can write it as the *fundamental model*

$$Y_\ell = m_\ell(\gamma) + s_\ell(\gamma) e_\ell^* \tag{5.3}$$

with $\{e_\ell^*\}$ i.i.d. G , where $m_\ell(\gamma)$ has no location parameters, $s_\ell(\gamma)$ has no scale parameters, G has no constraints, and $\dim(\theta) = \dim(\gamma) + 2$. Here, $e_\ell = (e_\ell^* - \xi)/\tau$, so (5.3) implies (5.1). So, $F(e) = G(\xi + \tau e)$, and the two extra parameters (ξ, τ) are the *location and scale parameters of G with respect to ρ* defined by

$$0 = \rho_{1,0} = \int_{-\infty}^{\infty} \rho^{(1)}((x - \xi)/\tau) dG(x),$$

$$1 = \rho_{1,1} = \int_{-\infty}^{\infty} ((x - \xi)/\tau) \rho^{(1)}((x - \xi)/\tau) dG(x).$$

We call γ the *fundamental parameter* and ξ, τ the *working parameters* as only they depend on the choice of ρ . For example, if $\rho(e) = e^2/2$ then $\xi = \mathbb{E}[e_1^*] = \text{mean } G$, $\tau^2 = \text{var } G$. To summarize, if the constraints $\rho_{1,0} = 0, \rho_{1,1} = 1$ are not to restrict us to (ρ, F) symmetric about zero, then $\mu_\ell(\theta), \sigma_\ell(\theta)$ *must* have the form (5.1), where $s_\ell(\gamma)$ has no scale parameters. Otherwise, τ would not be identifiable. Of course, any given model can be parameterized in many ways. So, (ξ, τ) can be taken as components of θ , or θ can be taken as any one-to-one function of (ξ, τ, γ) .

The presence of the parameters (ξ, τ) in the model (5.3) implies that $\rho(e) = r(e)$ and $\rho(e) = r((e - \xi_0)/\tau_0)$ for given (ξ_0, τ_0) will give exactly the same estimator $\hat{\gamma}$ for γ .

We call two models *equivalent* if one model is a reparameterization of the other.

Note that $m_\ell \equiv x'_\ell \gamma$ has no location parameters if and only if

$$\text{rank} \begin{pmatrix} x_1 & \cdots & x_n \\ 1 & & 1 \end{pmatrix} = \dim(x_\ell) + 1.$$

Note also that $s_\ell \equiv y'_\ell \gamma + 1$ has no scale parameters if and only if

$$\text{rank} \begin{pmatrix} y_1 & \cdots & y_n \\ 1 & & 1 \end{pmatrix} = \dim(y_\ell) + 1.$$

What models does (5.1) allow and disallow? Firstly, μ linear is allowed, that is $\mu = (\mu_\ell) = X\alpha$ with X of full rank less than or equal to n . For if $m_\ell = \alpha'_1 y_\ell, s_\ell = \beta'_1 z_\ell$ then $\xi s_\ell = \alpha'_2 z_\ell$, where $\alpha_2 = \xi$. Taking $m_\ell = 0$ gives the model

Example 5.3. $\mu_\ell = \xi s_\ell(\gamma), \sigma_\ell = \tau s_\ell(\gamma)$, so μ_ℓ/σ_ℓ does not depend on ℓ but does depend on ρ .

Taking s_ℓ known gives the model

Example 5.4. $\mu_\ell = m_\ell(\gamma) + \xi s_\ell$, $\sigma_\ell = \tau s_\ell$. This includes the ‘ μ linear, σ linear’ model

$$(\mu_1, \dots, \mu_n)' = X\alpha \text{ in } \mathbb{R}^n, \quad \sigma_\ell = \tau s_\ell \tag{5.4}$$

if only only if X has rank p .

Theorem 5.1. Note that (5.4) and $\mu_\ell = x'_\ell \alpha$, $\sigma_\ell = x'_\ell \beta$ or their equivalents are the only linear models allowed by (5.3).

The proof of Theorem 5.1 is given in Appendix F.

Taking $s_\ell = \exp(y'_\ell \theta_2)$ gives the model

Example 5.5. $\mu_\ell = m_\ell(\gamma) + \exp(\theta_1 + y'_\ell \theta_2)$, $\sigma_\ell = \exp(\theta_3 + y'_\ell \theta_2)$ if $y'_\ell \theta_2$ has no location parameters.

Example 5.6. Here, we give A, B needed for V of (2.5) as 4×4 partitioned matrices $(A_{i,j})$, $(B_{i,j})$ for the case when the parameters of $\{m_\ell\}$ and $\{s_\ell\}$ are different, say γ in \mathbb{R}^p and δ in \mathbb{R}^p . Suppose then that

$$\theta' = (\gamma', \delta', \xi, \tau), \quad \mu_\ell = m_\ell(\gamma) + \xi s_\ell(\delta), \quad \sigma_\ell = \tau s_\ell(\delta).$$

Then, A of (2.5) has the form $(A_{i,j})$, 4×4 with $A_{1,1} = a_{1,1} \tau^{-2} \lim_{n \rightarrow \infty} [s^{-2} m.m.]$, $A_{1,2} = (a_{1,1} \xi \tau^{-1} + a_{1,2}) \tau^{-1} \lim_{n \rightarrow \infty} [s^{-2} m.s.]$, $A_{1,3} = a_{1,1} \tau^{-2} \lim_{n \rightarrow \infty} [s^{-1} m.]$, $A_{1,4} = a_{1,2} \tau^{-2} \lim_{n \rightarrow \infty} [s^{-1} m.]$, $A_{2,2} = (a_{1,1} \xi^2 \tau^{-2} + a_{2,2}) \lim_{n \rightarrow \infty} [s^{-2} s.s.] + a_{1,2} \tau^{-1} \sum^2 \lim_{n \rightarrow \infty} [s^{-2} m.s.]$, $A_{2,3} = a_{1,2} \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1} s.]$, $A_{2,4} = a_{2,2} \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1} s.]$, $A_{3,3} = a_{1,1} \tau^{-2}$, $A_{3,4} = a_{1,2} \tau^{-2}$ and $A_{4,4} = a_{2,2} \tau^{-2}$ with $\{a_{i,j}\}$ of (2.3), (2.4), where $m_\ell = \partial m_\ell / \partial \gamma$ and $s_\ell = \partial s_\ell / \partial \gamma$. Note that $B = (B_{i,j})$, 4×4 is the same with $\{a_{i,j}\}$ replaced by $\{b_{i,j}\}$.

Finally, we give the components of A, B for the general case $\theta' = (\gamma', \xi, \tau)$, $\mu_\ell = m_\ell(\gamma) + \xi s_\ell(\gamma)$ and $\sigma_\ell = \tau s_\ell(\gamma)$. Writing them as

$$A, B = \mathcal{F}_{1,1} MM + \mathcal{F}_{1,2} \sum^2 MS + \mathcal{F}_{2,2} SS \tag{5.5}$$

with $\mathcal{F}_{i,j} = a_{i,j}$, $b_{i,j}$, respectively, of (2.3), (2.4), and partitioning them as 3×3 , their components are given by $MM_{i,3} = 0$, $MM_{2,2} = \tau^{-2}$, $MM_{3,1} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-1} (\dot{m} + \xi \dot{s})]$, $MM_{3,3} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-1} (\dot{m} + \xi \dot{s}) (\dot{m} + \xi \dot{s})']$, $SS_{i,2} = 0$, $SS_{3,3} = \tau^{-2}$, $SS_{3,1} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1} \dot{s}]$, $SS_{1,1} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1} \dot{s} \dot{s}']$, $MS_{i,2} = MS_{3,j} = 0$, $MS_{2,3} = \tau^{-2}$, $MS_{2,1} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-1} \dot{s}]$, $MS_{1,3} = \tau^{-2} \lim_{n \rightarrow \infty} [s^{-1} (\dot{m} + \xi \dot{s})]$ and $MS_{1,1} = \tau^{-1} \lim_{n \rightarrow \infty} [s^{-2} (\dot{m} + \xi \dot{s}) \dot{s}']$, where $m_\ell = \partial m_\ell / \partial \gamma$ and $s_\ell = \partial s_\ell / \partial \gamma$.

6 Combining Independent Series

Combination of independent data to improve statistical estimation is common place. Some published examples include: combination of comparable information collected independently from multiple surveys of the same population (Merkouris, 2004); combination of molecular, morphological and ecological data to infer species boundaries in organisms that are difficult to sample in large numbers Sanders et al. (2006); combination of independent microarray datasets pertaining to chronic allograft nephropathy (Kong et al., 2008); combination of independent samples from different laboratories to correct for measurement error Huang et al. (2012). Other examples are annual maximum sea levels observed at costal points sufficiently apart, annual maximum flood peak levels observed at rivers sufficiently apart and Example 3.2.

Here, we show how to apply the method when series (assumed to be independent of each other) need to be combined, because the models for the various series have some parameters in common. In practice, this may mean if two series overlap in time, discarding one of the series during the overlap period. Suppose then we have K series of lengths n_1, \dots, n_K , modeled as

$$Y_{\ell,k} = \mu_{\ell,k} + \sigma_{\ell,k}e_{\ell,k}$$

for $\ell \in T_k, k = 1, \dots, K$, where T_k is a set of n_k integers, with $\{T_k\}$ disjoint, and $\{e_{\ell,k}, \ell \in T_k, k = 1, \dots, K\}$ independently distributed as F unknown. We set $n = n_1 + \dots + n_K$ and $\lambda_k = n_k/n$. Note that $T_k = \{\ell_k + 1, \dots, \ell_{k+1}\}$ with $\ell_k = n_1 + \dots + n_k$. Define $\delta_{\ell,k} = 1$ for ℓ in T_k and zero otherwise. Then, we can write our model as (1.1) with $Y_\ell = Y_{\ell,k}$ for ℓ in T_k and

$$\mu_\ell = \sum_{k=1}^K \delta_{\ell,k} \mu_{\ell,k}, \quad \sigma_\ell = \sum_{k=1}^K \delta_{\ell,k} \sigma_{\ell,k}.$$

We give V for two types of models:

Type I: $\mu_{\ell,k} = \xi_k + m_\ell, \sigma_{\ell,k} = \tau_k s_\ell$ for ℓ in T_k and $1 \leq K$.

Type II: $\mu_{\ell,k} = \xi_k s_\ell + m_\ell, \sigma_{\ell,k} = \tau_k s_\ell$ for ℓ in T_k and $1 \leq K$.

In each case, we have $m_\ell = m_\ell(\gamma), s_\ell = s_\ell(\delta)$ with γ in \mathbb{R}^P , and δ in \mathbb{R}^Q say,

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \alpha = \begin{pmatrix} \xi \\ \gamma \end{pmatrix}, \quad \beta = \begin{pmatrix} \tau \\ \delta \end{pmatrix},$$

so α has dimension $p = P + K, \beta$ has dimension $p' = Q + K$, and θ has dimension $q = p + p' = P + Q + 2K$. So, each series has its own location

and scale parameters (ξ_k, τ_k) but the other parameters γ in \mathbb{R}^P say and δ in \mathbb{R}^Q say, are shared. We denote the k th unit vector in \mathbb{R}^K as $v_{k,K}$ and summation over ℓ in T_k by $\sum_{\ell=1}^{n_k}$. We set

$$\begin{aligned} m_\ell &= \partial m_\ell / \partial \gamma, \\ s_\ell &= \partial s_\ell / \partial \delta, \\ [\mathcal{F}, \mathcal{G}, \dots]_k &= n_k^{-1} \sum_{\ell=1}^{n_k} \mathcal{F}_\ell \mathcal{G}_\ell \dots \end{aligned}$$

for compatible vector sequences $\{\mathcal{F}_\ell, \mathcal{G}_\ell, \dots\}$.

6.1. *The Type I Model.* By Example 2.1 if $\rho_{2,1} = 0$, then $V = (V_{i,j})$, 2×2 with $V_{i,j} = c_{i,j} D^{i,j}$, $D^{i,j} = D_{i,i}^{-1} D_{i,j} D_{j,j}^{-1}$, $D_{i,j} = \lim_{n \rightarrow \infty} [\sigma^{-2} x_i x'_j]$ for $x_{\ell,1} = \partial \mu_\ell / \partial \alpha$, $x_{\ell,2} = \partial \sigma_\ell / \partial \beta$, and $c_{i,j}$ in Section 2.2. Partition $D_{i,j}$ as $(D_{i,j \cdot k,l}, 1 \leq k, l \leq 2)$ corresponding to the above 2×1 partitions of α, β . The expressions for $D_{i,j \cdot k,l}$ are given in Appendix E.

So, the k th rows of $D_{1,1,1,2}$ and $D_{2,2,1,2}$ are $\lim_{n \rightarrow \infty} \lambda_k \tau_k^{-2} [s^{-2} m.']_k$ and $\lim_{n \rightarrow \infty} \lambda_k \tau_k^{-1} [s^{-2} s.']_k$, and the asymptotic distributions of $\hat{\gamma}, \hat{\delta}$ given by (3.15), (3.16) hold for $G = D_{1,1}, H = D_{2,2}$, so that

$$G_{2,1} = \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-2} ([s^{-2} m.m.']_k - [s^{-2} m.]_k [s^{-2} m.']_k / [s^{-2}]_k)$$

is a function of τ, γ, δ and

$$H_{2,1} = \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k ([s^{-2} s.s.']_k - [s^{-1} s.]_k [s^{-1} s.']_k)$$

is a function of δ . Typically $s_\ell = 1$ when $\delta = 0$, simplifying the p -value of $(H_0 : \delta = 0)$.

Example 6.1. Set $m_\ell = x'_\ell \gamma$, $s_\ell = 1 + y'_\ell \delta$ for given x_ℓ, y_ℓ . So, $m_\ell = x_\ell$, $s_\ell = y_\ell$, and

$$G_{2,1} = \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-2} ([s^{-2} x x']_k - [s^{-2} x]_k [s^{-2} x']_k / [s^{-2}]_k),$$

a function of τ, δ .

For $s_\ell = \exp(y'_\ell \delta)$ we can always change τ_k to $\tau_k^* = \tau_k \exp(u'_\ell \delta)$ for $u_\ell = y_\ell - [y]$. With this parameterization $D_{2,2}$ is diagonal so that $\{\hat{\tau}_i, \hat{\delta}_j\}$ are asymptotically independent.

Before applying (3.15), (3.16), one should test the model. To do this one might first assume the more general model

$$\mu_{\ell,k} \equiv \xi_k + m_{\ell}(\gamma_k), \quad \sigma_{\ell,k} \equiv \tau_k s_{\ell}(\delta_k),$$

and test $H_1 : \delta_1 = \dots = \delta_K$ and $H_2 : \gamma_1 = \dots = \gamma_K$. Other hypotheses of interest are $H_3 : \delta_1 = \dots = \delta_K = 0$ and $H_4 : \gamma_1 = \dots = \gamma_K = 0$.

These may be tested as follows. Let $\hat{\xi}_k, \hat{\gamma}_k, \hat{\tau}_k, \hat{\delta}_k$ be the estimator of θ obtained from the k th sample, and let $\hat{\sigma}(\hat{\xi}_k)^2, \hat{c}_{1,1,k} \hat{G}_{2,1,k}^{-1}, \hat{\sigma}(\hat{\tau}_k)^2, \hat{c}_{2,2,k} \hat{H}_{2,1,k}^{-1}$ be the estimators of their covariances as given in Section 3.2. A test of H_3 is to accept if

$$\sum_{k=1}^K n_k \hat{\delta}'_k \hat{H}_{2,1,k} \hat{\delta}_k \leq \hat{c}_{2,2} \chi_{r,0.95}^2$$

for $r = (Q - 1)K$. A test of H_4 is to accept if

$$\sum_{k=1}^K n_k \hat{\gamma}'_k \hat{G}_{2,1,k} \hat{\gamma}_k \leq \hat{c}_{1,1} \chi_{r,0.95}^2$$

for $r = (P - 1)K$. To test H_1 , write it as $L(\delta'_1, \dots, \delta'_K)' = 0$, where $L = I_p \otimes (I_K - 1_K 1'_K / K)$ is idempotent of rank $P(K - 1)$. Also $N'LN = S(NN') - S(N)'S(N)/K$ for $N' = (N'_1, \dots, N'_K)$, $S(NN') = \sum_{k=1}^K N'_k N_k$, $S(N) = \sum_{k=1}^K N_k$. So, a test of H_1 is to accept if $S(NN') - S(N)'S(N)/K \leq \chi_{r,0.95}^2$, where $N_k = V_k^{-1/2} \hat{\delta}_k$ and $V_k = \hat{c}_{2,2}^{-1} \hat{H}_{2,1} / n_k$ for the k th series. The test for H_2 is exactly analogous. Suppose that one accepts H_1 . Then, instead of finding the M -estimator $\hat{\delta}$ for the combined series, one can use

$$\tilde{\delta} = \left(\sum_{k=1}^K \hat{V}_k^{-1} \right)^{-1} \left(\sum_{k=1}^K \hat{V}_k^{-1} \hat{\delta}_k \right),$$

where V_k is as above. This estimator has the same asymptotic covariance as the M -estimator for the combined series $\hat{\delta}$. This has great practical value as it means that an analysis of the combined series can be carried out in terms of the estimators for the component series. A similar result holds for testing H_2 . Replace $\delta, H, c_{2,2}$ by $\gamma, G, c_{1,1}$.

6.2. *The Type II Model.* Since the parameters of primary interest are $w = \begin{pmatrix} \gamma \\ \delta \end{pmatrix}$, it is better to redefine θ as $\theta' = (w', \xi', \tau')$. For then, using

a 2×2 partition of V corresponding to $\begin{pmatrix} \xi \\ \tau \end{pmatrix}$ and $w, n^{1/2}(\widehat{w} - w) \xrightarrow{d} \mathcal{N}_{P+Q}(0, V_{1.2}^{-1})$ with $V = A^{-1}BA^{-1}$. In the notation of (5.5), we partition A, B as 4×4 corresponding to $\theta' = (\gamma', \delta', \xi, \tau)$ with the elements of MM, MS, SS as given by Appendix E.

By Theorem 2.1 II, the results of this section hold for $D_{1,1}, MM$ without assuming that $\rho_{1,1} = 1, \rho_{2,1} = 0$ if $\{\sigma_\ell\}$ are given, i.e., if $\tau_k \equiv 1$ and $\{s_\ell\}$ are given; if also $s_\ell \equiv 1$ then for the Type I models, $\theta = \begin{pmatrix} \xi \\ \gamma \end{pmatrix}$ and (3.15) holds with

$$\begin{aligned} D_{1,1 \cdot 1,1} &= \lim_{n \rightarrow \infty} \text{diag}(\lambda_k, k = 1, \dots, K), \\ D_{1,1 \cdot 2,1} &= \lim_{n \rightarrow \infty} (\lambda_1 [m \cdot]_1, \dots, \lambda_K [m \cdot]_K), \\ D_{1,1 \cdot 2,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k [m \cdot m \cdot]'_k, \\ G_{2,1} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k ([m \cdot m \cdot]'_k - [m \cdot]_k [m \cdot]'_k), \end{aligned} \tag{6.1}$$

and for the Type II models, $\theta = \begin{pmatrix} \gamma \\ \xi \end{pmatrix}$ and (3.15) holds with

$$\begin{aligned} MM_{1,1} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k [m \cdot m \cdot]'_k, \\ MM_{1,3} &= \lim_{n \rightarrow \infty} (\lambda_1 [m \cdot]_1, \dots, \lambda_K [m \cdot]_K), \\ MM_{3,3} &= \lim_{n \rightarrow \infty} \text{diag}(\lambda_k), \\ n^{1/2}(\widehat{\theta} - \theta) &\xrightarrow{d} \mathcal{N}_p(0, G_{2,1}^{-1}) \end{aligned}$$

for $G_{2,1}$ of (6.1). By Theorem 2.1 III, (3.16) and the results of this section hold for $D_{2,2}, SS$ without assuming that $\rho_{1,0} = 0, \rho_{2,1} = 0$ if $\{\mu_\ell\}$ are given, that is if $\xi_k \equiv 0$ and $\{m_\ell\}$ are given.

7 Data Application

Here, we apply some of the models in Section 3 to three real data sets. They data sets are: annual minimum temperatures for Albert Park, Auckland, New Zealand from 1910 to 1983; annual maximum temperatures for

Albert Park, Auckland, New Zealand from 1855 to 1988; annual average temperatures for Albert Park, Auckland, New Zealand from 1910 to 1983. The data were obtained from the National Institute of Water and Atmospheric Research, Christchurch, New Zealand.

The scatter plots of the three data sets are shown in Figs. 5, 6 and 7. We can see a trend in location in all three data sets. We can also see a trend in scatter in at least the first two data sets. To account for these trends, we fitted the following models, where Year denotes the year, Year₀ denotes the start year and n denotes the number of observations:

1. the const-linear model specified by $\mu_{\text{Year}} = \alpha_1$ and $\sigma_{\text{Year}} = \beta_1 + \beta_2 \cdot (\text{Year} - \text{Year}_0) / n$, see Example 3.4;
2. the linear-const model specified by $\mu_{\text{Year}} = \alpha_1 + \alpha_2 \cdot (\text{Year} - \text{Year}_0) / n$ and $\sigma_{\text{Year}} = \beta_1$, see Example 3.4;
3. the linear-linear model specified by $\mu_{\text{Year}} = \alpha_1 + \alpha_2 \cdot (\text{Year} - \text{Year}_0) / n$ and $\sigma_{\text{Year}} = \beta_1 + \beta_2 \cdot (\text{Year} - \text{Year}_0) / n$, see Example 3.4;
4. the const-exponential model specified by $\mu_{\text{Year}} = \alpha_1$ and $\sigma_{\text{Year}} = \exp[\beta_1 + \beta_2 \cdot (\text{Year} - \text{Year}_0) / n]$, see Example 3.5;
5. the exponential-const model specified by $\mu_{\text{Year}} = \exp[\alpha_1 + \alpha_2 \cdot (\text{Year} - \text{Year}_0) / n]$ and $\sigma_{\text{Year}} = \beta_1$, see Example 3.5;

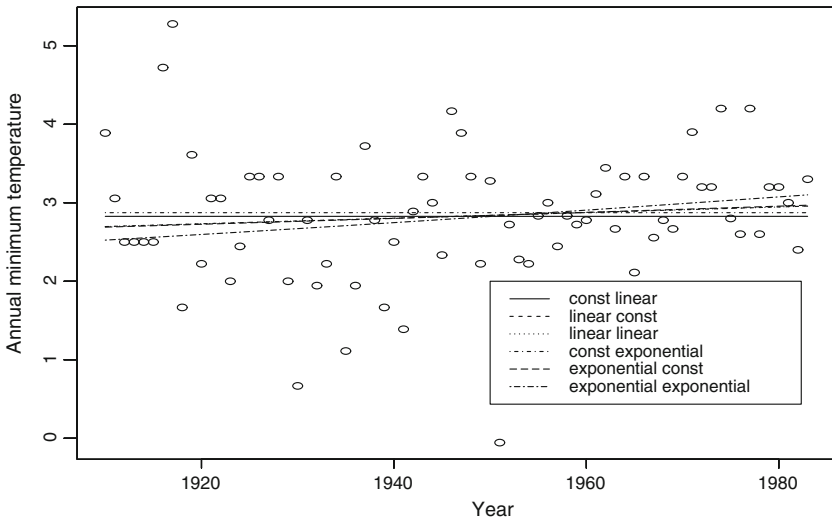


Figure 5: The six fitted models and their median for Auckland annual minimum temperatures

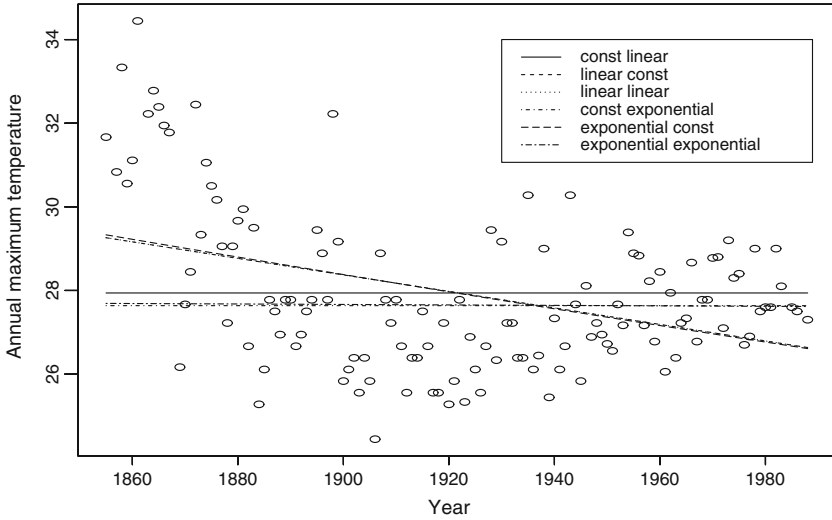


Figure 6: The six fitted models and their median for Auckland annual maximum temperatures

6. The exponential-exponential model specified by $\mu_{\text{Year}} = \exp[\alpha_1 + \alpha_2 \cdot (\text{Year} - \text{Year}_0)/n]$ and $\sigma_{\text{Year}} = \exp[\beta_1 + \beta_2 \cdot (\text{Year} - \text{Year}_0)/n]$, see Example 3.5.

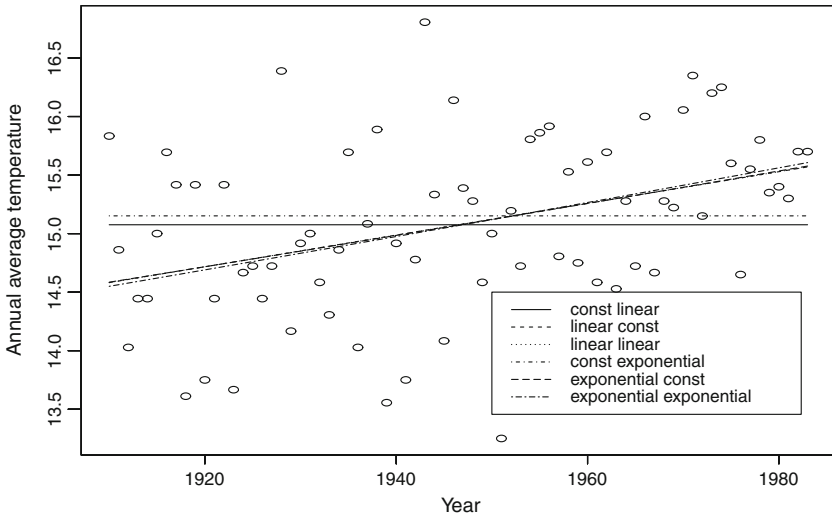


Figure 7: The six fitted models and their median for Auckland annual average temperatures

Note that $\text{Year}_0 = 1910$ for annual minimum temperatures, $\text{Year}_0 = 1855$ for annual maximum temperatures and $\text{Year}_0 = 1910$ for annual average temperatures. Note also that $n = 74$ for annual minimum temperatures, $n = 134$ for annual maximum temperatures and $n = 74$ for annual average temperatures. There are of course many other models that can be fitted to account for the trends in location and scale. But these are perhaps the simplest models suggested by Section 3.

The six models were fitted to each data set by the LSE. The discrimination among the fitted models was performed using the one-sample Kolmogorov-Smirnov test. The p -values from this test are reported in Tables 1 to 3. Since the Kolmogorov-Smirnov test assumes that the fitted model gives the “true” parameter values, the p -values were computed by simulation as follows:

- (i) fit the model to the given data and compute the corresponding Kolmogorov-Smirnov statistic;
- (ii) generate 10000 samples each of the same size as the data from the fitted model in step (i);
- (iii) refit the model to each of the 10000 samples;
- (iv) compute the Kolmogorov-Smirnov statistic for the 10000 fits in step (iii);
- (v) construct an empirical cdf of the 10000 values of the Kolmogorov-Smirnov statistic obtained in step (iv);
- (vi) compare the Kolmogorov-Smirnov statistic obtained in step (i) with the empirical cdf in step (v) to get the p -value.

The adequate fits for the annual minimum temperature data are given by the const-exponential and exponential-exponential models. The exponential-exponential model is the better of the two with the parameter estimates $\widehat{\alpha}_1 = 0.9230463$, $\widehat{\alpha}_2 = 0.2086026$, $\widehat{\beta}_1 = 0.2115011$, $\widehat{\beta}_2 = -0.8881930$, suggesting an upward trend in location and a downward trend in scatter for annual minimum temperature. None of the “signal + noise” models provide adequate fits. The fitted values of the median of Y_t for all six models are plotted in Fig. 5, which confirms the best fit of the exponential-exponential model.

The adequate fits for the annual maximum temperature data are given by the const-exponential and exponential-exponential models. The exponential-exponential model is the better of the two with the parameter estimates $\widehat{\alpha}_1 =$

3.321125398, $\widehat{\alpha}_2 = -0.002600307$, $\widehat{\beta}_1 = 1.267043083$, $\widehat{\beta}_2 = -1.602177560$, suggesting a downward trend in location and a downward trend in scatter for annual maximum temperature. Again none of the “signal + noise” models provide adequate fits. The fitted values of the median of Y_t for all six models are plotted in Fig. 6, which confirms the best fit of the exponential-exponential model.

Only the exponential-exponential model provides an adequate fit for the annual average temperature data. Its parameter estimates are $\widehat{\alpha}_1 = 2.67660708$, $\widehat{\alpha}_2 = 0.07107887$, $\widehat{\beta}_1 = -0.09020316$, $\widehat{\beta}_2 = -0.59943031$, suggesting an upward trend in location and a downward trend in scatter for annual average temperature. None of the other models (including all of the “signal + noise” models) provide adequate fits. The fitted values of the median of Y_t for all six models are plotted in Fig. 7, which confirms the best fit of the exponential-exponential model.

For all three data sets, the exponential-exponential model gives the best fit. None of the “signal + noise” models provide adequate fits for any of the data sets. However, the “signal + noise” model giving the best p -value for annual minimum temperature data is the exponential-const model. The “signal + noise” model giving the best p -value for annual maximum temperature data is the linear-const model. The “signal + noise” model giving the best p -value for annual average temperature data is the exponential-const model.

We now check to see how the robustness of the exponential-exponential model compares to the robustness of the “signal + noise” model giving the best p -value. The robustness was computed as follows:

- (i) fit the model to the data by removing its i th observation;
- (ii) repeat step (i) for $i = 1, 2, \dots, n$ (where n is the number of observations), recording each time the parameter estimates;
- (iii) as a measure of robustness, compute the standard deviation of the n estimates obtained in step (ii) for each parameter.

The measures of robustness for the exponential-exponential model and the “signal + noise” model giving the best p -value are compared in Tables 4 to 6. We can see that the measures of robustness for the exponential-exponential model are consistently smaller, suggesting that it is more robust than “signal + noise” models.

Finally, we examine the validity of Edgeworth type approximations used, for example, in Section 2.4. Edgeworth type approximations have many

others uses in statistics, for example, approximation of the Pearson chi-squared statistics (Lewis et al., 1984; Stafford, 1995). For example, the density of the Studentized mean of a random sample of size n can be approximated by

$$p_n(x) = \phi(x) + \phi(x) \sum_{r=1}^{\infty} n^{-r/2} \left[h'_r(x) - x h_r(x) \right], \quad (7.1)$$

where $\phi(x) = d\Phi(x)/dx$, $h'_r(x) = dh_r(x)/dx$ and $h_r(x)$ are certain polynomials tabulated in Withers (1984). Edgeworth type approximations are based on normality. However, not all data sets are normal or even approximately normal. For example, most financial data are heavy tailed. Normal type distributions are known to give very poor fits to financial data.

We now illustrate an example where Edgeworth type approximations fail. Figure 8 shows the histogram of the log-returns of crude oil prices for the period from 12th of March 1993 to 13th of March 2013. (The data were obtained from the database Datastream.) The sample size is 3913. Nadarajah et al. (2014) presented statistical evidence to suggest that the data can be considered as a random sample. We fitted (7.1) to the data by truncating the infinite sum at $r = m$. The p -values of the Kolmogorov-Smirnov test for the fit are shown in Table 7. We see that the fits are not close to being adequate even for m as large as 10.

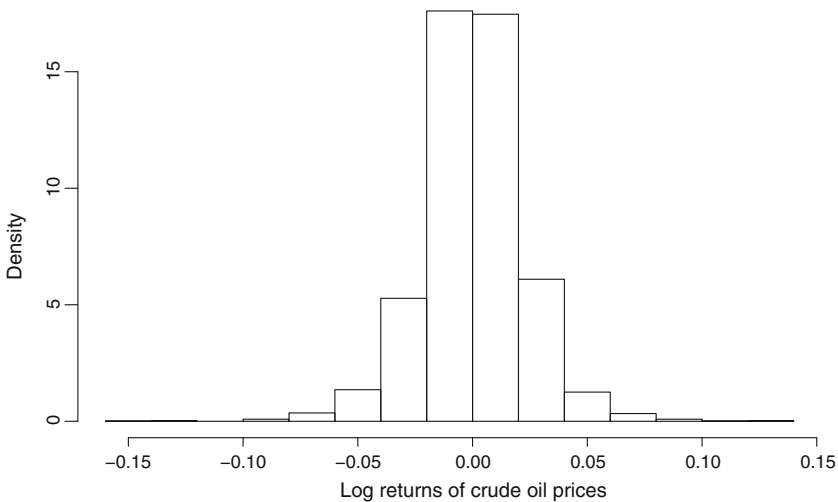


Figure 8: Histogram of the log-returns of crude oil prices

8 Future Work

The work of this paper can be extended in several ways. Throughout this paper, we have used one single $\rho(\cdot)$ function to obtain robust estimators for the location and scale parameters. An extension is to consider different $\rho(\cdot)$ functions for the location and scale parameters. Other extensions could include: use of different $\rho(\cdot)$ for every observation, as done in Bai and Wu (1997); study of robustness properties including breakdown points and influence functions of the proposed estimators, as done in Carroll and Ruppert (1982).

Acknowledgments. The authors would like to thank the Editor, the Associate Editor and the two referees for carefully reading and for their comments which greatly improved the paper.

Appendix A

Here, we give the asymptotic variances and efficiencies $c_{1,1}$, $c_{2,2}$, η_1 , η_2 plotted in Figs. 1 to 4 for Z normal, uniform, double-exponential. We also give these asymptotic variances and efficiencies for the gamma distribution for $r = 2$.

Suppose e_1 is a standardized form of a random variable Z with pdf $f_Z(z)$. That is $e_1 = (Z - \xi_r) / \tau_r$, where (ξ_r, τ_r) is given by

$$\mathbb{E} \left[|Z - \xi_r|^{r-1} \text{sign} (Z - \xi_r) \right] = 0, \quad \tau_r = h_r(r)^{1/r}, \quad h_r(a) = \mathbb{E} [|Z - \xi_r|^a]. \tag{A.1}$$

For a CI for $t(\alpha)$ or $t(\beta)$, the best choice of r is that which minimizes $c_{1,1}(r)$ of (2.18) or $c_{2,2}(r)$ of (2.19), respectively. So,

$$\begin{aligned} c_{1,1}(1) &= f(0)^{-2}/4 = h_1(1)^{-2} f_Z(\xi_1)^{-2} /4 \text{ since } f(0) = \tau_1 f_Z(\xi_1), \\ c_{1,1}(r) &= (r - 1)^{-2} h_r(r)^{-2/r} h_r(r - 2)^{-2} h_r(2r - 2) \text{ for } r \text{ even,} \\ c_{1,1}(2) &= 1, \\ c_{2,2}(r) &= r^{-2} \{ h_r(2r) h_r(r)^{-2} - 1 \}. \end{aligned}$$

Figures 1 and 2 plot $c_{1,1}$ and $c_{2,2}$ against r for Z normal, uniform, double-exponential. In each case, $c_{1,1}(r)$ is continuous at one. Figures 3 and 4 plot the asymptotic efficiencies of $\hat{\alpha}$, $\hat{\beta}$ relative to the MLE given by (2.15) against r for Z normal and double-exponential. For the uniform, the information is infinite.

If Z is symmetric about μ then ξ_r , $h_r(a)$ of (A.1) are given by $\xi_r = \mu$, $h_r(a) = \mathbb{E} [|Z|^a]$. Set $\rho_{1,1,i}^*(e_1) = \rho_{1,1,i}^*$ of Theorem 2.2. Then, $\rho_{1,1,i}^*(\lambda X) =$

$\lambda^{i-2} \rho_{1,1,i}^*(X)$. So, for $\lambda = \sigma/\tau_r$ if $e_1 = Z/\tau_r$ and $Z = \sigma e_0$, the asymptotic efficiency of $\hat{\alpha}$ relative to the MLE is $\eta_1 = \lambda^2 / \{\rho_{1,1,0}^*(e_0) c_{1,1}\}$, and the asymptotic efficiency of $\hat{\beta}$ relative to the MLE is $\eta_2 = 1 / \{(\rho_{1,1,2}^*(e_0) - 1) c_{2,2}\}$. For $Z \sim \mathcal{N}_1(\mu, \sigma^2)$, we have $\xi_r = \mu$, $h_r(a) = \sigma^a N_a$ for $N_a = \mathbb{E}[|\mathcal{N}_1(0, 1)|^a] = (2\pi)^{-1/2} 2^b \Gamma(b)$ at $b = (a + 1)/2$, so

$$\begin{aligned} c_{1,1}(r) &= \Gamma((r + 1)/2)^{-2-2/r} \Gamma(r-1/2) \pi^{1/2+1/r} / 4 \approx r^{-2} 2^{r-3/2} \exp(1-r) \text{ for } r \text{ large,} \\ c_{1,1}(1) &= c_{1,1}(1+) = \pi^2/4 \approx 2.467, \quad c_{1,1}(2) = 1, \quad c_{1,1}(3) = 3\pi^{4/3}/16 \approx 0.8627, \\ c_{1,1}(4) &= 6/3^{1.6} \approx 0.9623, \quad c_{1,1}(5) = 105\pi^{1.2} 2^{-8.4} \approx 1.228, \\ \rho_{1,1,0}^*(e_0) &= 1 \text{ for } e_0 \sim \mathcal{N}_1(0, 1), \\ \eta_1 &= 2\pi^{-1/2} \Gamma((r + 1)/2)^2 / \Gamma(r - 1/2) \approx r^2 2^{-1/2-r} \exp(2-r) \text{ for } r \text{ large,} \\ c_{2,2}(r) &= r^{-2} \left\{ \pi^{1/2} \Gamma(r + 1/2) / \Gamma(r/2 + 1/2)^2 - 1 \right\} \approx r^{-3} 2^{r+1.5} \text{ for } r \text{ large,} \\ c_{2,2}(1) &= \pi/2 - 1 \approx 0.571, \quad c_{2,2}(2) = 0.5, \quad c_{2,2}(3) = 5\pi/24 - 1/9 \approx 0.543, \\ c_{2,2}(4) &= 2/3, \quad c_{2,2}(6) = 113/90 \approx 1.26, \\ \rho_{1,1,2}^*(e_0) &= 3 \text{ for } e_0 \sim \mathcal{N}_1(0, 1), \\ \eta_2 &= 1 / \{2c_{2,2}\}. \end{aligned}$$

For $Z = \mu + \sigma U(-1, 1)$, we have $\xi_r = \mu$, $h_r(a) = \sigma^a / (a + 1)$, so for r even $c_{1,1}(r) = (r + 1)^{2/r} (2r - 1)^{-1} \approx \exp(2)/(2r)$ for r large. In particular, $c_{1,1}(1) = c_{1,1}(1+) = 4$, $c_{1,1}(2) = 1$, $c_{1,1}(3) = 2^{4/3}/5 \approx 0.504$ and $c_{1,1}(4) = 5^{1/2}/7 \approx 0.319$. Note that $c_{2,2}(r) = 1/(2r + 1)$. Also, $\rho_{1,1,i}^* = \infty$ for $i = 0, 2$, so $\eta_i = 0$. For $Z = \mu + \sigma D$ and D double-exponential, $\xi_r = \mu$, $\tau_r = \sigma (r!)^{1/r}$, $h_r(a) = \sigma^a a!$, and

$$\begin{aligned} c_{1,1}(r) &= (r!)^{-2/r} (r-1)!^{-2} (2r-2)! \approx \pi^{-1/2} \exp(2) 2^{2r-1.5} r^{-2} \text{ for } r \text{ large,} \\ \rho_{1,1,0}^* &= r!^{2/r}, \\ \eta_1 &= \Gamma(r)^2 / \Gamma(2r - 1) \approx \pi^{1/2} 2^{2-2r} r^{1/2} \text{ for } r \text{ large,} \\ c_{2,2}(r) &= r^{-2} \left\{ \binom{2r}{r} - 1 \right\} \approx \pi^{-1/2} 2^{2r} r^{-5/2} \text{ for } r \text{ large,} \\ \rho_{1,1,2}^* &= 2, \\ \eta_2 &= 1/c_{2,2}. \end{aligned}$$

For $f_Z(z) = \exp(-z) z^{g-1} / \Gamma(g)$ with g a positive integer, the median ξ_1 is given by $1/2 = \exp(-\xi_1) \sum_{i=0}^{g-1} \xi_1^i / i!$, so $\xi_1 = \log 2$ for $g = 1$, $\xi_1 \approx 1.68$ for $g = 2$; $c_{1,1}(1) = \log \xi_1^{-2} \approx 2.08$ for $g = 1$, $c_{1,1}(1) = \xi_1^{-6} (1 + \xi_1)^4 \approx 74.1$ for $g = 2$, $c_{1,1}(2) = 1$ for $g = 1$ or 2 ; $c_{2,2}(1) = -2/\xi_1 + 2/\xi_1^2 \approx 1.28$ for $g = 1$,

$c_{2,2}(1) = (6 - 4\xi_1 + \xi_1^2)(1 + \xi_1)^2 \xi_1^{-4} - 1 \approx 0.896$ for $g = 2$; $c_{2,2}(2) = 1$ for $g = 1$, $c_{2,2}(2) = 3.75$ for $g = 2$.

Appendix B

Here, we show how to obtain the 2×2 partition for $V = (V_{i,j})$ for the case $\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, $\mu_\ell = \mu_\ell(\alpha)$, and $\sigma_\ell = \sigma_\ell(\beta)$.

Theorem B.1. For $a_{i,j}$ and $b_{i,j}$ of (2.3), (2.4) set

$$\begin{aligned} r &= a_{1,2}^2 a_{1,1}^{-1} a_{2,2}^{-1}, \\ l_1 &= a_{1,1} (b_{2,2}/a_{2,2} - 2b_{1,2}/a_{1,2}), \\ l_2 &= a_{2,2} (b_{1,1}/a_{1,1} - 2b_{1,2}/a_{1,2}), \\ l_{1,2} &= b_{1,2} - a_{1,2} (b_{1,1}/a_{1,1} + b_{2,2}/a_{2,2}), \end{aligned}$$

and $N_i = D_{i,i} - rC_i$ for C_i of (3.2) and $D_{i,j}$ of (3.1). Then,

$$V_{1,1} = a_{1,1}^{-2} N_1^{-1} (b_{1,1} D_{1,1} + l_1 r C_1)^{-1} N_1^{-1}, \tag{B.1}$$

$$V_{2,2} = a_{2,2}^{-2} N_2^{-1} (b_{2,2} D_{2,2} + l_2 r C_2)^{-1} N_2^{-1}, \tag{B.2}$$

$$(V_{2,1})' = V_{1,2} = a_{1,1}^{-1} a_{2,2}^{-1} N_1^{-1} D_{1,2} \left(l_{1,2} I_{p'} + r b_{1,2} D_{2,2}^{-1} D_{2,1} D_{1,1}^{-1} D_{1,2} \right) N_2^{-1}. \tag{B.3}$$

PROOF. We use the formula for the inverse of a square 2×2 partitioned matrix for: $\Sigma = (\Sigma_{i,j})$, 2×2 , $\Sigma^{-1} = (\Sigma^{i,j})$, where $\Sigma^{1,1} = \Sigma_{1,2}^{-1}$, $\Sigma^{1,2} = -\Sigma^{1,1} \Sigma_{1,2} \Sigma_{2,2}^{-1}$, $\Sigma^{2,2} = \Sigma_{2,1}^{-1}$, $\Sigma^{2,1} = -\Sigma^{2,2} \Sigma_{2,1} \Sigma_{1,1}^{-1}$, $\Sigma_{1,2} = \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}$ and $\Sigma_{2,1} = \Sigma_{2,2} - \Sigma_{2,1} \Sigma_{1,1}^{-1} \Sigma_{1,2}$. So, $A^{2,2} = A_{2,1}^{-1}$, where $A_{2,1} = a_{2,2} N_2$, and

$$V_{2,2} = \sum_{i,j=1}^2 A^{2,i} B_{i,j} A^{j,2} = A^{2,2} L_{2,2} A^{2,2},$$

where

$$\begin{aligned} L_{2,2} &= A_{2,1} A_{1,1}^{-1} B_{1,1} A_{1,1}^{-1} A_{1,2} - A_{2,1} A_{1,1}^{-1} B_{1,2} - B_{2,1} A_{1,1}^{-1} A_{1,2} + B_{2,2} \\ &= r l_2 C_2 + b_{2,2} D_{2,2}, \end{aligned}$$

so $V_{2,2} = a_{2,2}^{-2} \{ (l_2 + b_{2,2}) N_2^{-1} D_{2,2} N_2^{-1} - l_2 N_2^{-1} \}$. Note that $V_{1,1}$ follows reversing 1 and 2. Similarly,

$$(V_{2,1})' = V_{1,2} = A^{1,1} L_{1,2} A^{2,2},$$

where

$$\begin{aligned} L_{1,2} &= A_{1,2}A_{2,2}^{-1}B_{2,1}A_{1,1}^{-1}A_{1,2} - A_{1,2}A_{2,2}^{-1}B_{2,2} - B_{1,1}A_{1,1}^{-1}A_{1,2} + B_{1,2} \\ &= l_{1,2}D_{1,2} + rb_{1,2}D_{1,2}D_{2,2}^{-1}C_2 \\ &= (l_{1,2} + b_{1,2})D_{1,2} - b_{1,2}\mathcal{M}_{1,2}, \end{aligned}$$

and

$$\mathcal{M}_{1,2} = D_{1,2}D_{2,2}D_{2,2}^{-1}N_2 = N_1D_{1,1}^{-1}D_{1,2}.$$

So,

$$V_{1,2} = a_{1,1}^{-1}a_{2,2}^{-1}N_1^{-1}D_{1,2} \left\{ (l_{1,2} + b_{1,2})N_2^{-1} - b_{1,2}D_{2,2}^{-1} \right\}.$$

The result follows.

Appendix C

For Example 2.1 with $\mu_\ell = x'_\ell\alpha$, we have $A_{1,1} = a_{1,1} \lim_{n \rightarrow \infty} D_n$ with $D_n = [\sigma^{-2}xx']$. For Example 2.1 with $\sigma_\ell = x'_\ell\beta$, we have $A_{2,2} = a_{2,2} \lim_{n \rightarrow \infty} D_n$. Here, we give $D = \lim_{n \rightarrow \infty} D_n$ when $\mu_\ell = x'_\ell\alpha$ and $\sigma_\ell = x'_\ell\beta$. For $x'_\ell = (1, \ell/n)$ and $D = \beta_1^{-2}H$, where $H = (h_{i+j-2})$, 2×2 , $h_i = \int_0^1 (1 + \delta t)^{-2} t^i dt$, $h_0 = (1 + \delta)^{-1}$, $h_1 = \delta^{-2}\mathcal{M}_1 - \delta^{-1}h_0$, $h_2 = \delta^{-2}(2 + \delta)h_0 - 2\delta^{-3}\mathcal{M}_1$, $\mathcal{M}_1 = \log(1 + \delta)$, $\det(H) = \delta^{-2}h_0 + \delta^{-4}\mathcal{M}_1^2$ and $\delta = \beta_2/\beta_1$. For $x'_\ell = (1, \ell/n - 1/2)$ and $D = \beta_1^{-2}G$, where $G = (g_{i+j-2})$, $g_i = \int_{-1/2}^{1/2} (1 + \delta t)^{-2} t^i dt = d_i - 2\delta d_{i+1} + 3\delta^2 d_{i+2} + \dots$ for $d_i = \int_{-1/2}^{1/2} t^i dt = 2^{-i}/(i+1)$ for i even, and zero for i odd, $g_0 = (1 - \delta^2/4)^{-1}$, $g_1 = \delta^{-2}\mathcal{M}_2 - \delta^{-1}g_0$, $\mathcal{M}_2 = \log \{ (1 + \delta/2) / (1 - \delta/2) \}$ and $g_2 = \delta^{-2}(g_0 + 1) - 2\delta^{-3}\mathcal{M}_2$.

Appendix D

This paper has mostly considered the choice $\rho(e) \propto |e|^r$, r even. Here, we give some other choices of ρ . The first two have $\rho^{(1)}$ ‘redescending’.

Example D.1. Take $\rho(e) = d_1^{-1} \log(d_2 + e^2)$, where $d_2 > 0$. Then,

$$\begin{aligned} 0 &= \rho_{1,0} = 2d_1^{-1} \int_{-\infty}^{\infty} e (d_2 + e^2)^{-1} dF(e), \\ \rho_{1,1,0} &= 4d_1^{-2} \int_{-\infty}^{\infty} e^2 (d_2 + e^2)^{-2} dF(e), \\ \rho_{2,0} &= 2d_1^{-1} \int_{-\infty}^{\infty} (d_2 - e^2) (d_2 + e^2)^{-2} dF(e). \end{aligned}$$

One can optimize the choice of d_2 over a predetermined finite set to minimize the estimator of $\rho_{2,0}^{-2}\rho_{1,1,0}$ and avoid $\rho_{2,0} = 0$. If $d_1 \geq 2$, $\hat{\tau} = 0$ for the model $\sigma_\ell = \tau$ so one needs $d_2 < 2$; also $\hat{\alpha} = \hat{\alpha}(d_2)$ for the model $\sigma_\ell = 1$. This differs from $\hat{\alpha} = \hat{\alpha}(d_2\tau^2)$ for the model $\sigma_\ell = \tau$.

Example D.2. Suppose $\rho(e) = -c^{-1} \exp(-e^2/2)$. Then, $\rho^{(1)}(e) = c^{-1}e \exp(-e^2/2)$ and $\rho^{(2)}(e) = -c^{-1}(1 - e^2) \exp(-e^2/2)$.

Example D.3. The M-estimator corresponding to $F(x) \sim$ Gumbel has $d\rho(e) = e + \exp(-e)$ for some constant $d > 0$. So, $d\rho^{(1)}(e) = 1 - \exp(-e)$ and $d\rho^{(2)}(e) = \exp(-e)$. We require that F be centered and scaled so $\mathbb{E}[\exp(-e_1)] = 1$ and $d = \mathbb{E}[e_1] - \mathbb{E}[e_1 \exp(-e_1)]$. Also $d^2\rho_{1,1,i} = \mathbb{E}[\{1 - \exp(-e_1)\}^2 e_1^i]$ and $d\rho_{2,i} = \mathbb{E}[\exp(-e_1) e_1^i]$. This choice of $\rho(\cdot)$ is very non-robust to large negative observations.

For some more choices of $\rho(e)$, see Andrews et al. (1972).

Appendix E

The partitions $(D_{i,j,k,l}, 1 \leq k, l \leq 2)$ of $D_{i,j}$ for the type I model in Section 6.1 are:

$$\begin{aligned}
 D_{1,1,1,1} &= \lim_{n \rightarrow \infty} \text{diag}(\lambda_k \tau_k^{-2} [s^{-2}]_k, k = 1, \dots, K), \\
 D_{1,1,1,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-2} v_{k,K} [s^{-2} m \cdot']_k, \\
 D_{1,1,2,1} &= \lim_{n \rightarrow \infty} (\lambda_1 \tau_1^{-2} [s^{-2} m \cdot]_1, \dots, \lambda_K \tau_K^{-2} [s^{-2} m \cdot]_K), \\
 D_{1,1,2,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-2} [s^{-2} m \cdot m \cdot']_k, \\
 D_{2,2,1,1} &= \lim_{n \rightarrow \infty} \text{diag}(\lambda_k \tau_k^{-2}, k = 1, \dots, K), \\
 D_{2,2,1,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} v_{k,K} [s^{-1} s \cdot']_k, \\
 D_{2,2,2,1} &= \lim_{n \rightarrow \infty} (\lambda_1 \tau_1^{-1} [s^{-1} s \cdot]_1, \dots, \lambda_K \tau_K^{-1} [s^{-1} s \cdot]_K), \\
 D_{2,2,2,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k [s^{-2} s \cdot s \cdot']_k,
 \end{aligned}$$

$$\begin{aligned}
D_{1,2,1,1} &= \lim_{n \rightarrow \infty} \text{diag} (\lambda_k \tau_k^{-2} [s^{-1}]_k, k = 1, \dots, K), \\
D_{1,2,1,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} v_{k,K} [s^{-2} s.']_k, \\
D_{1,2,2,1} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} [s^{-1} m.]_k v'_{k,K} \\
&= \lim_{n \rightarrow \infty} (\lambda_1 \tau_1^{-1} [s^{-1} m.]_1, \dots, \lambda_K \tau_K^{-1} [s^{-1} m.]_K), \\
D_{1,2,2,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} [s^{-2} m.s.']_k.
\end{aligned}$$

The elements of MM , MS , SS for the type II model in Section 6.2 are:

$$\begin{aligned}
MM_{i,4} &\equiv 0, \quad MM_{1,1} = \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-2} [s^{-2} m.m.']_k, \\
MM_{1,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \xi_k \tau_k^{-2} [s^{-2} m.s.']_k, \\
MM_{1,3} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} [s^{-1} m.]_k v'_{k,K} \\
&= \lim_{n \rightarrow \infty} (\lambda_1 \tau_1^{-1} [s^{-1} m.]_1, \dots, \lambda_K \tau_K^{-1} [s^{-1} m.]_K), \\
MM_{2,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \xi_k^2 \tau_k^{-2} [s^{-2} s.s.']_k, \\
MM_{2,3} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \xi_k \tau_k^{-2} [s^{-1} s.]_k v'_{k,K} \\
&= \lim_{n \rightarrow \infty} (\lambda_1 \xi_1 \tau_1^{-2} [s^{-1} s.]_1, \dots, \lambda_K \xi_K \tau_K^{-2} [s^{-1} s.]_K), \\
MM_{3,3} &= \lim_{n \rightarrow \infty} \text{diag} (\lambda_k \tau_k^{-2}), \\
MS_{i,1} &\equiv MS_{i,3} \equiv MS_{4,i} \equiv 0, \quad MS_{1,2} = \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} [s^{-2} m.s.']_k, \\
MM_{1,4} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-2} [s^{-1} m.]_k v'_{k,K}
\end{aligned}$$

$$\begin{aligned}
 &= \lim_{n \rightarrow \infty} \left(\lambda_1 \tau_1^{-2} [s^{-1}m.]_1, \dots, \lambda_K \tau_K^{-2} [s^{-1}m.]_K \right), \\
 MS_{2,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \xi_k \tau_k^{-1} [s^{-2}s.s.']_k, \\
 MS_{2,4} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \xi_k \tau_k^{-2} [s^{-1}s.]_k v'_{k,K} \\
 &= \lim_{n \rightarrow \infty} \left(\lambda_1 \xi_1 \tau_1^{-2} [s^{-1}s.]_1, \dots, \lambda_K \xi_K \tau_K^{-2} [s^{-1}s.]_K \right), \\
 MS_{3,2} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} v_{k,K} [s^{-1}s.']_k, \quad MS_{3,4} = \lim_{n \rightarrow \infty} \text{diag} (\lambda_k \tau_k^{-2}), \\
 SS_{i,1} \equiv SS_{i,3} \equiv 0, \quad SS_{2,2} &= \lim_{n \rightarrow \infty} [s^{-2}s.s.'] = \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k [s^{-2}s.s.']_k, \\
 SS_{2,4} &= \lim_{n \rightarrow \infty} \sum_{k=1}^K \lambda_k \tau_k^{-1} [s^{-1}s.]_k v'_{k,K} \\
 &= \lim_{n \rightarrow \infty} \left(\lambda_1 \tau_1^{-1} [s^{-1}s.]_1, \dots, \lambda_K \tau_K^{-1} [s^{-1}s.]_K \right), \\
 SS_{4,4} &= \lim_{n \rightarrow \infty} \text{diag} (\lambda_k \tau_k^{-2}).
 \end{aligned}$$

Appendix F

PROOF. of Theorem 2.1 For $1 \leq i \leq q$, θ satisfies

$$\begin{aligned}
 0 &= \Lambda_{n \cdot i}(\hat{\theta}) = n^{-1} \sum_{\ell=1}^n \lambda_{\ell \cdot i}(\hat{\theta}) \\
 &= n^{-1} \sum_{\ell=1}^n (\lambda_{\ell \cdot i} + \lambda_{\ell \cdot i, j} \Delta_j / 1! + \lambda_{\ell \cdot i, j, k} \Delta_j \Delta_k / 2! + \dots) \quad (F.1)
 \end{aligned}$$

by Taylor's expansion, where we use the tensor sum convention that repeated pairs of suffixes j, k, \dots are implicitly summed over their range $1, \dots, q$. So, $\hat{\theta}$ is consistent (that is, $\hat{\theta} \xrightarrow{p} \theta$ as $n \rightarrow \infty$) if and only if $\Lambda_{n \cdot i} \xrightarrow{p} 0$ for $i = 1, \dots, q$ since by the Law of Large Numbers each $\Lambda_{n \cdot i, j, \dots}$ is $O_p(1)$ and by the Central Limit Theorem $\Lambda_{n \cdot i} - \mathbb{E}[\Lambda_{n \cdot i}] = O_p(n^{-1/2})$. Also

$$\lambda_{\ell \cdot i} = \rho^{(1)}(e_\ell) e_{\ell \cdot i} + \sigma_\ell^{-1} \sigma_{\ell \cdot i}, \quad (F.2)$$

$$e_{\ell \cdot i} = -\mu_{\ell \cdot i} \sigma_\ell^{-1} - (Y_\ell - \mu_\ell) \sigma_\ell^{-2} \sigma_{\ell \cdot i} = -(\mu_{\ell \cdot i} + e_\ell \sigma_{\ell \cdot i}) \sigma_\ell^{-1}. \quad (F.3)$$

So,

$$\begin{aligned} \mathbb{E}[\lambda_{\ell.i}] &= (-\rho_{1,0}\mu_{\ell.i} + (1 - \rho_{1,1})\sigma_{\ell.i})\sigma_{\ell}^{-1}, \\ \mathbb{E}[\Lambda_{n.i}] &= n^{-1} \sum_{\ell=1}^n \mathbb{E}[\lambda_{\ell.i}] \\ &= -\rho_{1,0}[\sigma^{-1}\mu_{.i}] + (1 - \rho_{1,1})[\sigma^{-1}\sigma_{.i}] \\ &= 0 \text{ for (I) - (IV),} \end{aligned}$$

and $\Delta = O_p(n^{-1/2})$. II is actually just IV with $q_2 = 0$. Also

$$\Lambda_{n.i,j} = n^{-1} \sum_{\ell=1}^n \lambda_{\ell.i,j} = A_{i,j} + O_p(n^{-1/2}).$$

By (F.1), for $1 \leq i \leq q$,

$$0 = n^{-1} \sum_{\ell=1}^n \lambda_{\ell.i} + A_{i,j}\Delta_j + O_p(n^{-1}),$$

so

$$\Delta_j = -A^{j,i}n^{-1} \sum_{\ell=1}^n \lambda_{\ell.i} + O_p(n^{-1})$$

for $1 \leq j \leq q$, giving $n^{1/2}(\hat{\theta} - \theta) \sim \mathcal{N}_q(0, A^{-1}BA^{-1})$. Now

$$\begin{aligned} \text{covar}[\lambda_{\ell.i}, \lambda_{\ell.j}] &= \sigma_{\ell}^{-2} \left\{ (\rho_{1,1,0} - \rho_{1,0}^2)\mu_{\ell.i}\mu_{\ell.j} + (\rho_{1,1,1} - \rho_{1,0}\rho_{1,1}) \sum_{i,j}^2 \mu_{\ell.i}\sigma_{\ell.j} \right. \\ &\quad \left. + (\rho_{1,1,2} - \rho_{1,1}^2)\sigma_{\ell.i}\sigma_{\ell.j} \right\}. \end{aligned}$$

So,

$$\begin{aligned} B_{i,j} &= (\rho_{1,1,0} - \rho_{1,0}^2) \lim_{n \rightarrow \infty} [\sigma^{-2}\mu_{.i}\mu_{.j}] \\ &\quad + (\rho_{1,1,1} - \rho_{1,0}\rho_{1,1}) \sum_{i,j}^2 \lim_{n \rightarrow \infty} [\sigma^{-2}\mu_{.i}\sigma_{.j}] \\ &\quad + (\rho_{1,1,2} - \rho_{1,1}^2) \lim_{n \rightarrow \infty} [\sigma^{-2}\sigma_{.i}\sigma_{.j}]. \end{aligned}$$

Also,

$$\lambda_{\ell.i,j} = \rho^{(2)}(e_{\ell})e_{\ell.i}e_{\ell.j} + \rho^{(1)}(e_{\ell})e_{\ell.i,j} - \sigma_{\ell}^{-2}\sigma_{\ell.i}\sigma_{\ell.j} + \sigma_{\ell}^{-1}\sigma_{\ell.i,j}, \quad (\text{F.4})$$

and

$$\begin{aligned}
 e_{\ell,i,j} &= -\mu_{\ell,i,j}\sigma_{\ell}^{-1} + \sum_{i,j}^2 \mu_{\ell,i}\sigma_{\ell,j}\sigma_{\ell}^{-2} - (Y_{\ell} - \mu_{\ell})\sigma_{\ell,i,j}\sigma_{\ell}^{-2} & (F.5) \\
 &+ 2(Y_{\ell} - \mu_{\ell})\sigma_{\ell,i}\sigma_{\ell,j}\sigma_{\ell}^{-3} \\
 &= -\mu_{\ell,i,j}\sigma_{\ell}^{-1} + \sum_{i,j}^2 \mu_{\ell,i}\sigma_{\ell,j}\sigma_{\ell}^{-2} - e_{\ell}\sigma_{\ell,i,j}\sigma_{\ell}^{-1} + 2e_{\ell}\sigma_{\ell,i}\sigma_{\ell,j}\sigma_{\ell}^{-2}. & (F.6)
 \end{aligned}$$

So, by (F.2),

$$\begin{aligned}
 A_{i,j} &= \rho_{2,0} \lim_{n \rightarrow \infty} [\sigma^{-2}\mu_{i,j}] + \rho_{2,1} \sum_{i,j}^2 \lim_{n \rightarrow \infty} [\sigma^{-2}\mu_{i,j}] \\
 &+ (\rho_{2,2} + 2\rho_{1,1} - 1) \lim_{n \rightarrow \infty} [\sigma^{-2}\sigma_{i,j}] \\
 &- \rho_{1,0} \lim_{n \rightarrow \infty} [\sigma^{-1}\mu_{i,j}] + \rho_{1,0} \sum_{i,j}^2 \lim_{n \rightarrow \infty} [\sigma^{-2}\mu_{i,j}] \\
 &+ (1 - \rho_{1,1}) \lim_{n \rightarrow \infty} [\sigma^{-1}\sigma_{i,j}].
 \end{aligned}$$

For (IV), $[\sigma^{-1}\sigma_{i,j}] = [\sigma^{-1}\sigma_{i,j}]$.

PROOF. of Corollary 2.1 The MLE is given by Theorem 2.1 with the choice $\rho(e) = -\log f(e)$, so $\rho_{1,0} = 0$, $\rho_{1,1} = 1$, $\rho_{1,1,i} = \rho_{1,1,i}^*$, and $\rho_{2,i} = \rho_{1,1,i} - i(i-1)\mu_{i-2}$, where $\mu_j = \mathbb{E} [e_1^j]$, $(\rho_{2,0}, \rho_{2,1}, \rho_{2,2} + 1) = (\rho_{1,1,0}, \rho_{1,1,1}, \rho_{1,1,2} - 1)$ and $A = B = I_n(\theta)$.

PROOF. of Theorem 5.1 Suppose $m_{\ell} + \xi s_{\ell} = x'_{\ell}\alpha$ and $\tau s_{\ell} = y'_{\ell}\beta$. Then, m_{ℓ}, s_{ℓ} have linear covariates, say $m_{\ell} = u'_{\ell}a$, $s_{\ell} = v'_{\ell}b + 1$. So,

$$\mu_{\ell} = u'_{\ell}a + v'_{\ell}b\xi + \xi, \quad \sigma_{\ell} = v'_{\ell}b\tau + \tau.$$

This is only linearizable if $v'_{\ell}b = 0$ (Example 5.4), or $u_{\ell} = Ct_{\ell}$, $v_{\ell} = Dt_{\ell}$.

In the latter case, we may take $x_{\ell} = y_{\ell} = \begin{pmatrix} 1 \\ t_{\ell} \end{pmatrix}$, $\alpha = \begin{pmatrix} \xi \\ C'a + D'b \end{pmatrix}$,

$$\beta = \begin{pmatrix} \tau \\ D'b\tau \end{pmatrix}.$$

References

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust estimates of location: Survey and advances princeton university press*. Princeton, New Jersey.

- BAI, Z. D. and WU, Y. (1997). General M -estimation. *Journal of Multivariate Analysis* **63**, 119–135.
- BIANCO, A., BOENTE, G. and DI RIENZO, J. (2000). Some results for robust GM-based estimators in heteroscedastic regression models. *Journal of Statistical Planning and Inference* **89**, 215–242.
- BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and adaptive estimation for semiparametric models* Johns Hopkins University Press, Maryland, Baltimore.
- BIERENS, H. J. (1981). *Robust methods and asymptotic theory in nonlinear econometrics*. Springer Verlag, New York.
- CARROLL, R. J. (1982). Adapting for heteroscedasticity in linear models. *Annals of Statistics* **10**, 1224–1233.
- CARROLL, R. J. and RUPPERT, D. (1982). Robust estimation in heteroscedastic linear models. *Annals of Statistics* **10**, 429–441.
- CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and weighting in regression*. Chapman and Hall, New York.
- CHEN, J., LI, D. G. and LIN, Z. Y. (2011). Asymptotic expansion for nonparametric M -estimator in a nonlinear regression model with long-memory errors. *Journal of Statistical Planning and Inference* **141**, 3035–3046.
- FAURIE, F. and GIREMUS, A. (2010). *Combining generalized likelihood ratio and M -estimation for the detection/compensation of GPS measurement biases*, p. 4178–4181.
- FENG, S., NADARAJAH, S. and HU, Q. (2007). Modeling annual extreme precipitation in China using the generalized extreme value distribution. *Journal of the Meteorological Society of Japan* **85**, 599–613.
- FIRTH, D. (1987). On the efficiency of quasilikelihood estimation. *Biometrika* **74**, 233–245.
- HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach based on Influence Functions*. John Wiley and Sons, New York.
- HOSEINNEZHAD, R. and BAB-HADIASHAR, A. (2011). An M -estimator for high breakdown robust estimation in computer vision. *Computer Vision and Image Understanding* **115**, 1145–1156.
- HUANG, Y. D., HUANG, Y., MOODIE, Z., LI, S. and SELF, S. (2012). Comparing and combining data across multiple sources via integration of paired-sample data to correct for measurement error. *Statistics in Medicine* **31**, 3748–3759.
- HUBER, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics* **35**, 73–101.
- HUBER, P. J. (1981). *Robust statistics*. John Wiley and Sons, New York.
- HUBER, P. J. (1996). *Robust statistical procedures*, second edition, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.
- HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust statistics*, second. John Wiley and Sons, New Jersey.
- KONG, X. R., MAS, V. and ARCHER, K. J. (2008). A non-parametric meta-analysis approach for combining independent microarray datasets: application using two microarray datasets pertaining to chronic allograft nephropathy. *BMC Genomics*, 9. doi: [10.1186/1471-2164-9-98](https://doi.org/10.1186/1471-2164-9-98).
- KOUL, H. L. (1992). *Weighted empiricals and linear models*. Institute of Mathematical Statistics, Hayward, California.
- KOUL, H. L. (2002). *Weighted empirical processes in dynamic nonlinear models*, second. Institute of Mathematical Statistics, Hayward, California.

- LEE, M. -J. (2010). *Micro-Econometrics: methods of moments and limited dependent variables*, second. Springer Verlag, New York.
- LEWIS, T., SAUNDERS, I. W. and WESTCOTT, M. (1984). The moments of the Pearson chi-squared statistic and the minimum expected value in two-way tables. *Biometrika* **71**, 515–522.
- LIESE, F. and MIESCKE, K. -J. (2008). *Statistical decision theory: estimation, testing and selection*. Springer Verlag, New York.
- MARAZZI, A. (1993). *Algorithms, routines, and s functions for robust statistics wadsworth and Brooks/Cole advanced books and software*. Pacific Grove, California.
- MARONNA, R. A., MARTIN, R. D. and YOHAI, V. J. (2006). *Robust statistics: Theory and methods*. John Wiley and Sons, Chichester.
- MENDES, B. and TYLER, D. E. (1996). *Constrained M-estimation for regression*. Springer Verlag, New York, RIEDER, H. (ed.),.
- MERKOURIS, T. (2004). Combining independent regression estimators from multiple surveys. *Journal of the American Statistical Association* **99**, 1131–1139.
- NADARAJAH, S. (2005). Extremes of daily rainfall in west central Florida. *Climatic Change* **69**, 325–342.
- NADARAJAH, S., AFUECHETA, E. and CHAN, S. (2014). GARCH modeling of five popular commodities. *Empirical Economics*. doi: [10.1007/s00181-014-0845-3](https://doi.org/10.1007/s00181-014-0845-3).
- NADARAJAH, S. and SHIAU, J. T. (2005). Analysis of extreme flood events for the Pachang River, Taiwan. *Water Resources Management* **19**, 363–374.
- PERACCHI, F. (2001). *Econometrics*. John Wiley and Sons, Chichester.
- PFANZAGL, J. (1994). *Parametric statistical theory*. Walter de Gruyter and Company, Berlin.
- PRAKASA RAO, B. L. S. (1999). *Statistical inference for diffusion type processes*. Edward Arnold, London.
- RANDLES, R. H. and WOLFE, D. A. (1979). *Introduction to the theory of nonparametric statistics*. John Wiley and Sons, New York.
- RAO, C. R. and TOUTENBURG, H. (1995). *Linear models: least squares and alternatives*. Springer Verlag, New York.
- REY, W. J. J. (1978). *Robust statistical methods*. Springer Verlag, Berlin.
- RIEDER, H. (1994). *Robust asymptotic statistics*. Springer Verlag, New York.
- SANDERS, K. L., MALHOTRA, A. and THORPE, R. S. (2006). Combining molecular, morphological and ecological data to infer species boundaries in a cryptic tropical pitviper. *Biological Journal of the Linnean Society* **87**, 343–364.
- SEBER, G. A. F. and WILD, C. J. (1989). *Nonlinear regression*. John Wiley and Sons, New York.
- SERFLING, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley and Sons, New York.
- STAFFORD, J. E. (1995). Exact cumulant calculations for Pearson χ^2 and Zelterman statistics for r -way contingency tables. *Journal of Computational and Graphical Statistics* **4**, 199–212.
- STAUDTE, R. G. and SHEATHER, S. J. (1990). *Robust estimation and testing*. John Wiley and Sons, New York.
- VAN DE GEER, S. A. (2000). *Applications of empirical process theory*. Cambridge University Press, Cambridge.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer Verlag, New York.

- VERBOON, P. (1994). *A robust approach to nonlinear multivariate analysis* D.S.W.O. Press, Leiden.
- WELSH, A. H., CARROLL, R. J. and RUPPERT, D. (1994). Fitting heteroscedastic regression models. *Journal of the American Statistical Association* **89**, 100–116.
- WITHERS, C. S. (1984). In *Asymptotic expansions for distributions and quantiles with power series cumulants* vol. 46, pp. 389–396.
- WITHERS, C. S., KROUSE, D. P., PEARSON, C. P. and NADARAJAH, S. (2009). Modelling temperature trends in New Zealand. *Environmental Modeling and Assessment* **14**, 231–249.
- WITHERS, C. S. and NADARAJAH, S. (2006). Evidence of trend in return levels for daily windrun in New Zealand. *Journal of the Meteorological Society of Japan* **84**, 805–819.

CHRISTOPHER S. WITHERS
INDUSTRIAL RESEARCH LIMITED
LOWER HUTT, NEW ZEALAND
E-mail: kit.withers@gmail.com

SARALEES NADARAJAH
UNIVERSITY OF MANCHESTER
MANCHESTER M13 9PL, UK
E-mail: mbbssn2@manchester.ac.uk

Paper received: 6 January 2014.