ROBUST ESTIMATION OF THE MEAN OF AN ASYMMETRIC DISTRIBUTION: The Gamma distribution case

Alfio Marazzi*

No. 17, marzo 1995

* Docente di "Teoria degli investimenti e gestione del portafoglio" al Centro di Studi Bancari. Dottore in matematica SPF Zurigo, Professore associato all'Università di Losanna, Capo della divisione di statistica e informatica, Istituto universitario di medicina sociale e preventiva, Losanna.

Stime robuste della media di una distribuzione asimmetrica: il caso della distribuzione Gamma

Le distribuzioni asimmetriche sono molto frequenti nelle applicazioni economiche. Esempi comuni sono le distribuzioni del reddito e le distribuzioni della durata di degenza ospedaliera (un importante indicatore dei costi).

Spesso i dati contengono osservazioni anomale, ad esempio redditi decisamente eccezzionali o durate di degenza assurde per una data diagnosi. Questi dati hanno molto influsso sulle stime tradizionali della media (media aritmetica) e della sua imprecisione (varianza). Siccome la loro frequenza varia da campione a campione (ad esempio, di anno in anno) le stime tradizionali di media non sono affidabili.

Un ramo della teoria statistica, la "statistica robusta", studia metodi statistici, detti "robusti", più affidabili dei metodi tradizionali. Fra i metodi di stima più noti in questo campo vi sono le stime M.

In questo quaderno vengono studiate e confrontate alcune stime M per la media di distribuzioni asimmetriche basate sul modello Gamma di distribuzione. Lo scopo è quello di selezionarne una che possa essere usata in applicazioni pratiche. Essa è implementata nella libreria di programmi ROBETH e nel sistema statistico S-Plus.

Robust estimation of the mean of an asymmetric distribution: The Gamma distribution case

A. Marazzi, C.Ruffieux

March 13, 1995

Abstract

Several types of M-estimators for multiparameter models have been proposed in Hampel et al. (1986). This paper considers 6 options and specializes them to the Gamma distribution, with the purpose of estimating its mean. It discusses their implementation, i.e., the numerical computation of the estimates, the determination of their tuning parameters, and the evaluation of measures that characterize their distribution (e.g., bias, variance, breakdown point, etc.). One of the options is selected for practical use.

1 Introduction

We consider the gamma distribution defined by its density

$$f_{\sigma,\alpha}(y) = \frac{1}{\sigma\Gamma(\alpha)}\left(\frac{y}{\sigma}\right)^{\alpha-1}\exp\left(-\left(\frac{y}{\sigma}\right)\right), \quad y > 0, \sigma > 0, \alpha > 0, \qquad (1)$$

where Γ is the Gamma function. When $\sigma = 1$, we write f_α in place of $f_{\sigma,\alpha}$. The cumulative distribution function is denoted by $F_{\sigma,\alpha}$ or by F_α.

The pair (σ, α) is the usual parameter vector; α is the shape parameter, and σ is the scale parameter. However, we often prefer to work with the pair (τ, α), where $\tau = \ln(\sigma)$, or with (ν, α), where $\nu = \ln(\alpha\sigma)$. Often it is be convenient to use the abbreviation $\theta = (\sigma, \alpha)^{\mathrm{T}}$, or $\theta = (\tau, \alpha)^{\mathrm{T}}$, or $\theta = (\nu, \alpha)^{\mathrm{T}}$, where θ is a column vector. Our main objective is to estimate the function $\alpha\sigma$ of the parameters, i.e. the mean $E(Y)$ of a Gamma distributed random variable Y. Thus, in the (ν, α) parametrization, α may be considered as a nuisance parameter.

In classical statistics, one assumes that the observations y_i are distributed exactly according to the density f_θ, for a certain unknown value of θ, and estimates this value using the available data. One of the most widely used procedures is maximum likelihood, with the estimate defined as the value $\hat{\theta}$ which maximizes $\Pi_{i=1}^{n} f_\theta(y_i)$, or equivalently as the value $\hat{\theta}$ which minimizes $-\sum \ln f_\theta(y_i)$. Equating to zero the derivatives of this function, one obtains

$$\sum_{i=1}^{n} s(y_i, \theta) = 0, \tag{2}$$

where $s(y, \theta) = (s_1(y, \theta), s_2(y, \theta))^{\mathrm{T}}$ denotes the vector of likelihood scores,

$$s(y, \theta) = \frac{\partial}{\partial \theta} \ln f_\theta(y) = \left(\frac{\partial}{\partial \theta_1} \ln f_\theta(y), \frac{\partial}{\partial \theta_2} \ln f_\theta(y) \right)^{\mathrm{T}}. \tag{3}$$

For the Gamma distribution in the (τ, α) parametrization, the score functions are:

$$s_1(y, \theta) = \left[\frac{y}{\sigma} - \alpha \right], \quad \text{and} \quad s_2(y, \theta) = \ln \left(\frac{y}{\sigma} \right) - \dot{\Gamma}(\alpha), \tag{4}$$

where $\dot{\Gamma}$ denotes the Digamma function, $\dot{\Gamma}(\alpha) = d \ln \Gamma(\alpha)/d\alpha$, and $\sigma = \exp(\tau)$. The score functions in the other parametrizations are easily obtained observing that, if $\vartheta = g(\theta)$ is a 1-1 parameter transformation (e.g. $\vartheta = (\sigma, \alpha) = (\exp(\tau), \alpha)$ or $\vartheta = (\nu, \alpha) = (\ln(\alpha) + \tau, \alpha)$) with Jacobian matrix $B(\theta) = \partial g/\partial \theta$, the new score functions are $B(\theta)^{-\mathrm{T}} s(y, \theta)$. Note that, in the (τ, α) and the (ν, α) parametrization, the score functions depend on σ via $z = y/\sigma$, and, therefore, the estimation problem is scale invariant. Unfortunately, the problem is not shape invariant.

In robustness theory one realizes that the model f_θ is a mathematical abstraction which is only an idealized approximation of reality. It is therefore necessary to use procedures that behave fairly well under deviations from the assumed model. Such procedures are called robust. One of the best known classes of robust estimators are the M-estimators. In the next sections, we consider some of the M-estimators described in Hampel et al. (1986, Chapter 4) and specialize them to the parameters of the Gamma distribution. These estimators are defined as solutions $\hat{\theta}$ of

$$\sum_{i=1}^{n} \psi(y_i, \theta) = 0, \tag{5}$$

where ψ denotes a given vector function of (y, θ) with values in $\mathbb{R}^2$. Equation (5) is a generalization of the maximum likelihood equation (2). The function ψ is a composition of the score functions s and Huber's function h_a with tuning parameter a, defined as

$$\begin{aligned} h_a(z) &= \max(-a, \min(z, a)), \qquad &\text{when } z \in \mathbb{R}^1, \\ h_a(z) &= z \min(1, a/||z||), \qquad &\text{when } z \in \mathbb{R}^2, \end{aligned}$$

where $||z||$ denotes the Euclidean norm of z. Many theoretical results concerning the asymptotic and the robustness properties of M-estimators are available in the literature.

We discuss the numerical computation of M-estimators $\hat{\theta}$ defined by (5) for various choices of the function ψ and a few parametrizations of interest. Then, we evaluate and compare numerically some of the measures that characterize the distribution of the corresponding estimates of $E(Y)$. This includes bias and breakdown point, and estimates of the maximum asymptotic bias and variance over ϵ-contamination neighborhoods of the model. We aim to select an estimation procedure for $E(Y)$ with known robustness and asymptotic properties to be used in practice; thus, ease of computation is a major criterion of choice. The selected procedures have been included into the library ROBETH (Marazzi, 1993) and interfaced to S. Programs are available from the authors.

Remark 1.1. Many of the techniques described in this paper can be applied (with obvious changes) to the estimation of the parameters of the Weibull distribution with density

$$f_{\sigma,\alpha}(y) = \frac{\alpha}{\sigma} \left(\frac{y}{\sigma}\right)^{\alpha-1} \exp(-(y/\sigma)^{\alpha}), \quad y > 0, \sigma > 0, \alpha > 0. \tag{6}$$

Note however that, if Y is distributed according to $f_{\sigma,\alpha}$, the density of $T = \ln Y$ is

$$f_{\tau,v}(t) = \frac{1}{v} \exp\left[\left(\frac{t-\tau}{v}\right) - e^{(t-\tau)/v}\right], \qquad -\infty < t < \infty, \tag{7}$$

where $\tau = \ln \sigma$ and $v = 1/\alpha$. In this parametrization, τ is a location, v is a scale parameter, and the corresponding estimation problem is location and scale invariant. This is an important simplification compared to the Gamma distribution.

2 Estimators

Let $s(\theta, y)$ denote the vector of score functions (4) for $\theta = (\tau, \alpha)$. For a fixed value of the parameter b (see below), define $\hat{\theta} = (\hat{\tau}, \hat{\alpha})^{\mathrm{T}}$ as a solution of

$$\sum_{i=1}^{n} \psi_b \left[A_b(\theta)(s(y_i, \theta) - c_b(\theta))\right] = 0. \tag{8}$$

We consider two options for the function $\psi_b(z)$, $z = (z_1, z_2)$:

(i) The *shrinking norm estimator* with

$$\psi_b(z) = h_b(z), \qquad b \in \mathbb{R}^1,$$

(ii) The *shrinking component estimator* with

$$\psi_b(z) = (h_{b_1}(z_1), h_{b_2}(z_2))^{\mathrm{T}}, \quad b = (b_1, b_2) \in \mathbb{R}^2.$$

The parameter b, called *tuning parameter*, is fixed by the user (see Section 6). $A_b(\theta)$ is a 2×2 nonsingular lower triangular matrix; its inverse may be considered as the transposed Jacobian of a parameter transformation $\vartheta = g(\theta)$. We consider the following options:

(a) The (τ, α) *parametrization* with

$$A_b(\theta) = I;$$

(b) The (ν, α) *parametrization* with

$$A_b(\theta) = \begin{pmatrix} 1 & 0 \\ -1/\alpha & 1 \end{pmatrix};$$

(c) The *standard parametrization* implicitly defined by

$$\int \psi_b \left[A_b(\theta)(s(y, \theta) - c_b(\theta))\right] \psi_b \left[A_b(\theta)(s(y, \theta) - c_b(\theta))\right]^{\mathrm{T}} f_\theta(y) dy = I. \tag{9}$$

Finally, the function $c_b(\theta)$ is defined by:

$$\int \psi_b \left[A_b(\theta)(s(y,\theta) - c_b(\theta))\right] f_\theta(y) dy = 0. \tag{10}$$

Remark 2.1. The special case $A = I$ of the shrinking component estimator is a simple generalization of the maximum likelihood equations, where the likelihood scores are truncated and "re-centered" by c_b. The "correction" c_b makes the M-estimate asymptotically unbiased if the observations are distributed according to the model (Fisher consistency). Note that all estimators reduce to the maximum likelihood estimator for $b = b_1 = b_2 = \infty$.

Remark 2.2. Option (c) defines the *standardized* estimators of Hampel et al. (1986, Chapter 4), where further classes of M-estimators are proposed. In that book, the shrinking norm estimator is called the *optimal B_s-robust estimator*, and the shrinking component estimator is called the *optimal B_s^p-robust estimator for partitioned parameters.* Note however that, despite the name, no optimality result is known for the second proposed estimator, whereas a weak optimality property is known for the optimal B_s-robust case (Hampel et al., 1986, p.244). Both estimators truncate the re-centered score functions of the standard parameter vector $\vartheta = g(\theta)$. Due to the form of the asymptotic covariance matrix of $g(\hat{\theta})$ (Sections 3 and 5), the affine transformation A_b defined by equation (9) makes the standard parameters approximately orthogonal. An application of the shrinking norm estimator of the parameters of the Gamma distribution is described in Victoria-Feser (1993).

Remark 2.3. Previous experience (Willimann, 1980) suggests that the unstandardized estimators of Hampel et al. (1986) are computationally less reliable than the standardized ones. We do not consider them in this paper.

Remark 2.4. Thanks to the scale invariance, the transformation $A_b(\theta)$ and the correction $c_b(\theta)$ defined by (9) and (10) do not depend on τ. Therefore, we will write $A_b(\alpha)$ and $c_b(\alpha)$ in place of $A_b(\theta)$, $c_b(\theta)$. In the case of the Weibull distribution (7), the dependence of on both the parameters τ and v is trivial.

3 Evaluation criteria

Let $\mathcal{Y}$ denote the sample space, $\Theta = \{\theta\}$ the parameter space (of dimension p), and let $y_1, y_2, \ldots, y_n$ denote n observations of the random variable Y. We consider the unknown distribution F of Y, the model distributions F_θ ($\theta \in \Theta$), the empirical distribution F_n, and the distribution Δ_y which puts probability 1 at the point y. The densities of F and F_θ are denoted by f and f_θ. As F_θ is only an idealized approximation of F it is useful to assume that F belongs to a neighborhood $\mathcal{P}$ of one of the F_θ, say, F_{θ^*}. We use the *ϵ-contamination model*:

$$F \in \mathcal{P}_\epsilon = \{G \mid G = (1-\epsilon)F_{\theta^*} + \epsilon H,\ H \text{ arbitrary}\}, \tag{11}$$

where ϵ is a given percentage of *gross errors*, e.g., $\epsilon = 5\%$.

In order to evaluate and compare estimates of $E(Y)$, for the Gamma model, we compute asymptotic bias and variances, estimates of the maximum asymptotic bias over $\mathcal{P}_\epsilon$, estimates of the maximum asymptotic variances over $\mathcal{P}_\epsilon$, and asymptotic relative efficiencies with respect to the maximum likelihood estimate. We seek for estimates with small bias and variance over some set of challenge distributions. In particular, we seek for estimates with a small maximum asymptotic bias, a small maximum asymptotic variance and a large asymptotic efficiency with respect to the maximum likelihood estimate at the model F_θ. This section summarizes the definitions of these quantities (see Hampel et al., 1986, for more details). The computational aspects are discussed in Section 5.

Let $\hat{\theta}(F_n)$ be a p-dimensional M-estimate, where $\hat{\theta}(F)$ is a solution of

$$\int \psi(y, \hat{\theta}(F))dF(y) = 0, \tag{12}$$

and ψ is a user defined vector valued function $\psi : \mathcal{Y} \times \Theta \to \mathbb{R}^p$. In particular, we consider the cases introduced in Section 2. We use the abbreviation $\hat{\theta}$ in place of $\hat{\theta}(F_n)$. Under certain regularity conditions, $\hat{\theta}$ is asymptotically normally distributed, with asymptotic covariance matrix

$$V(\hat{\theta}, F) \quad = \quad M(\psi, F)^{-1}Q(\psi, F)M(\psi, F)^{-T}, \tag{13}$$

where

$$Q(\psi, F) = \int \psi(y, \hat{\theta}(F))\psi(y, \hat{\theta}(F))^{\mathrm{T}} dF(y), \tag{14}$$

$$M(\psi, F) = -\int \left[\frac{\partial}{\partial \theta}\psi(y, \theta)\right]_{\theta=\hat{\theta}(F)} dF(y). \tag{15}$$

In applications, $V(\hat{\theta}, F_n)/n$ or $V(\hat{\theta}, F_{\hat{\theta}})/n$ are often used as approximations for the covariance matrix of $\hat{\theta}$. If $\hat{\theta}$ is Fisher consistent at F_θ, i.e., if it satisfies $\int \psi(x, \theta) dF_\theta(x) = 0$ for all θ (see equation (10)), one obtains:

$$M(\psi, F_\theta) = \int \psi(y, \theta) s(y, \theta) dF_\theta(y), \tag{16}$$

where $s(y, \theta)$ is the vector of likelihood scores. This simplifies the computation of $V(T, F_\theta)$.

Let ξ be a one-dimensional parameter defined as a function $\xi = h(\theta)$, e.g., $\xi = \alpha \exp(\tau)$. Let $\hat{\xi}$ be an estimator of ξ, and $V(\hat{\xi}, F)$ its asymptotic variance. We assume that $\hat{\xi}$ satisfies $\hat{\xi}(F_\theta) = \xi$ for all θ, and let $\xi^* = h(\theta^*)$. The *asymptotic bias* of $\hat{\xi}$ under a point contamination $\epsilon\Delta_y$ of F_{θ^*} is

$$B(\hat{\xi}, \xi^*, \epsilon, y) = |\hat{\xi}((1-\epsilon)F_{\theta^*} + \epsilon\Delta_y) - \xi^*|. \tag{17}$$

The *maximum asymptotic bias* of $\hat{\xi}$ over $\mathcal{P}_{\epsilon^*}$ is

$$\mathcal{B}(\hat{\xi}, \xi^*, \epsilon) = \sup_{G \in \mathcal{P}_\epsilon} |\hat{\xi}(G) - \xi^*|. \tag{18}$$

The *maximum asymptotic variance* of $\hat{\xi}$ over $\mathcal{P}_{\epsilon^*}$ is

$$\mathcal{V}(\hat{\xi}, \xi^*, \epsilon) = \sup_{G \in \mathcal{P}_\epsilon} V(\hat{\xi}, G). \tag{19}$$

Let $\check{\theta}$ denote the maximum likelihood estimate of θ and let $\check{\xi} = h(\check{\theta})$. The *asymptotic relative efficiency* of $\hat{\xi}$ with respect to $\check{\xi}$ at the distribution F is

$$\mathrm{ARE}(\hat{\xi}, \check{\xi}, F) = V(\check{\xi}, F)/V(\hat{\xi}, F). \tag{20}$$

4 Computation of the estimates

General suggestions for simultaneously solving (8)–(10) by means of iterative algorithms can be found in Hampel et al. (1986, Section 4.6b). Given initial values for (τ, α), one computes c_b from (10), A_b from (9) and improved values for (τ, α) from (8). This scheme is usually memory-efficient but typically quite slow, especially when used within a resampling scheme. Therefore, we use the following two step procedure:

(a) For fixed b, solve (9) and (10) with respect to A_b and c_b for a discrete set of α-values and store the solutions. More precisely, choose an interval $[\alpha_1, \alpha_2]$ and an integer k (e.g., $k = 100$), and tabulate the solutions $A_b(\alpha)$ and $c_b(\alpha)$ of (9) and (10) for $\alpha = \alpha_1 + (i-1)(\alpha_2 - \alpha_1)/(k-1)$, $i = 1, \ldots, k$.

(b) Solve (8) with respect to θ, using a linear interpolation of the tables obtained in step (a) in order to compute $A_b(\alpha)$ and $c_b(\alpha)$ for the required values of α.

In practice, it is necessary to fix α_1 and α_2 so that $[\alpha_1, \alpha_2]$ contains the solution $\hat{\alpha}$ of (8). However, the preliminary computation of step (a) substantially accelerates step (b). In the following, we discuss both steps in more detail.

4.1 Computation of the vector c_b

The computation of c_b for given b, α and $A_b(\alpha)$, can be performed by means of the usual algorithms for M-estimates of location.

4.2 Computation of the matrix A_b

For the shrinking norm case, the problem of computing the matrix A_b, for given b, α, and a given value c of $c_b(\alpha)$, can be formulated as the problem of solving for A the equation

$$\text{ave}\{u(|z|)zz^{\mathrm{T}}\} = I, \tag{21}$$

where $z = (z_1, z_2)^{\mathrm{T}} = Ax$, $x = s(\theta, y) - c$, $A = A_b(\alpha)$, $\theta = (\tau = 0, \alpha)$, $\text{ave}\{.\} = \int \{.\} f_\alpha(y) dy$ and $u(|z|) = \min(1, b/|z|)^2$. A fixed point algorithm for solving (21) is described by Huber (1981, Chapter 8), who also gives a convergence proof.

For the shrinking component case, one has to solve for A:

$$\text{ave}\left\{ \begin{pmatrix} u_1(|z_1|) & 0 \\ 0 & u_2(|z_2|) \end{pmatrix} zz^{\mathrm{T}} \begin{pmatrix} u_1(|z_1|) & 0 \\ 0 & u_2(|z_2|) \end{pmatrix} \right\} = I, \tag{22}$$

where $u_1(|z_1|) = \min(1, b_1/|z_1|)$, and $u_2(|z_2|) = \min(1, b_2/|z_2|)$. A simple iterative algorithm is obtained as follows. Let $A^{(0)}$ be an initial and $A^{(1)}$ an improved value of A. Replace the arguments of u_1 and u_2 by $|z_1^{(0)}|$ and $|z_2^{(0)}|$, where $z^{(0)} = A^{(0)}x$, and the middle product zz^{T} by $A^{(1)}x[A^{(1)}x]^{\mathrm{T}}$. Then solve for $A^{(1)}$. One obtains:

$$\begin{aligned} a_{11}^{(1)} &= \text{ave}\{u_1(|z_1^{(0)}|)^2 x_1^2\}^{-2}, \\ a_{22}^{(1)} &= \text{ave}\{u_2(|z_2^{(0)}|)^2(\kappa x_1 + x_2)^2\}^{-2}, \\ a_{21}^{(1)} &= \kappa a_{22}^{(1)}, \end{aligned} \tag{23}$$

where $\kappa = -\text{ave}\{u_1(|z_1^{(0)}|)u_2(|z_2^{(0)}|)x_1x_2\}/\text{ave}\{u_1(|z_1^{(0)}|)u_2(|z_2^{(0)}|)x_1^2\}$. A convergence proof of this algorithm is still lacking, but the procedure works well in practice.

Figure 1 shows the elements of the matrix A_b for varying α. They have been obtained by simultaneously solving (9) and (10) for the shrinking norm and the shrinking component cases. The precision of the two algorithms is similar (but the vertical scales on the diagrams of a_{21} are different).

4.3 Computation of the parameter estimates

For $\theta = (\tau, \alpha)$, $\sigma = \exp(\tau)$, and $j = 1, 2$ define

$$S_{pj}(\sigma, \alpha) = \frac{1}{n} \sum \psi_{b_j}[\tilde{s}_j(y_i, \theta)], \tag{24}$$

$$S_{nj}(\sigma, \alpha) = \frac{1}{n} \sum \tilde{s}_j(y_i, \theta) \min[1, b/|\tilde{s}(y_i, \theta)|], \tag{25}$$

where

$$\begin{aligned} \tilde{s}_1(y, \theta) &= a_{11}(\alpha)[y/\sigma - \alpha - c_1(\alpha)], \\ \tilde{s}_2(y, \theta) &= a_{21}(\alpha)[y/\sigma - \alpha - c_1(\alpha)] + a_{22}(\alpha)[\ln(y/\sigma) - \dot{\Gamma}(\alpha) - c_2(\alpha)]. \end{aligned}$$

The components of $\tilde{s}(\theta, y)$ are the recentered score functions of the standard parameters. The shrinking component estimate is the solution of $S_{pj}(\sigma, \alpha) = 0$, $j = 1, 2$, and the shrinking norm estimate is the solution of $S_{nj}(\sigma, \alpha) = 0$, $j = 1, 2$.

The diagrams in Figure 2 show the functions S_{pj} and S_{nj} for varying σ (some fixed values of α and artificial data sets have been used). For the shrinking component case, the function S_{p1} is monotone in σ and changes sign. This guarantees the existence of a zero. Thus, the following simple idea works:

(a) Use a regula–falsi method to invert S_{p1} with respect to σ over the interval $[\alpha_1, \alpha_2]$, for which tables of $c_b(\alpha)$ and $A_b(\alpha)$ are available. In other words, for any desired $\alpha \in [\alpha_1, \alpha_2]$ obtain an highly accurate solution $\bar{\sigma}(\alpha)$ of $S_{p1}(\sigma, \alpha) = 0$.

(b) For $\alpha = \alpha_1 + (i-1)(\alpha_2 - \alpha_1)/(k-1)$, $i = 1, \dots, k$, compute $S_{p2}(\bar{\sigma}(\alpha), \alpha)$. Any time this function changes sign, use a bisection method to solve $S_{p2}(\bar{\sigma}(\alpha), \alpha) = 0$.

This procedure usually finds the solutions of the system $S_{pj}(\sigma, \alpha) = 0$. It can be complemented with a plot of $S_{p2}(\bar{\sigma}(\alpha), \alpha)$, in order to circumscribe the solutions of the second equation, if necessary. On the contrary, the behaviour of S_{nj} is quite complex for $j = 1$ and 2. It suggests that both these equations may have multiple solutions, and solving their system is not as straightforward as in the previous case. In our implementation, we replace the regula-falsi method with a (blind) bisection method in order to (hopefully) find the smallest zero of S_{n1}.

4.4 A note on parameter transformations

The matrix A_b and the correction c_b depend on the parametrization; in Section 2, they have been defined for $\theta = (\tau, \alpha)$. We want to determine them, when equations (8), (9) and (10) are solved in a different parametrization. We use the following elementary lemma (chain rule).

Lemma. Let $\vartheta = g(\theta)$ and $\theta = h(v)$ (i.e., $v = h^{-1}(\theta)$) define two one-to-one parameter transformations. Let $B(\theta) = \partial g(\theta)/\partial\theta$ be the Jacobian of g, and $C(v) = \partial h(v)/\partial v$ be the Jacobian of h. Then, the Jacobian of $\vartheta = g(h(v))$ is $B(h(v))C(v)$.

Suppose now, for example, that $\vartheta = g(\theta)$ is the standard parameter vector defined by (9) and that $v = (\nu, \alpha)$ is the new parametrization. Then $A_b(\theta) = B(\theta)^{-T}$, and $\theta = (\tau, \alpha) = (\nu - \ln(\alpha), \alpha) = h(v)$. Thus,

$$C(v) = \begin{pmatrix} 1 & -1/\alpha \\ 0 & 1 \end{pmatrix},$$

and A_b must be replaced with

$$B(h(v))^{-T}C(v)^{-T} = A_b(\theta)\begin{pmatrix} 1 & 0 \\ 1/\alpha & 1 \end{pmatrix}.$$

It is easy to see that c_b must be replaced with $C(v)^{-T}c_b(\theta)$. As a corollary, the estimates of the standard parameters ϑ obtained from (8)–(10), using the (τ, α) or the (ν, α) parametrization, coincide.

5 Computation of asymptotic quantities

Let $\hat{\theta}(F_n)$ be an estimate of θ, as defined in Sections 2 and 3. Moreover, let $\xi = h(\theta)$ be a one-dimensional function of θ, e.g., $\xi = \alpha \exp(\tau) = E(Y)$. We estimate ξ with $h(\hat{\theta})$. In order to evaluate the quantities $\mathcal{B}$ and $\mathcal{V}$, defined in (18) and (19), we use the influence function and the change of variance function of $\hat{\theta}$ and $h(\hat{\theta})$ (Hampel et al., 1986).

5.1 Influence and change of variance functions of a multidimensional M-estimate

The *influence function* of $\hat{\theta}$ at the distribution F is defined as the p-vector

$$\mathrm{IF}(y;\hat{\theta},F) = \left[\frac{\partial}{\partial t}\hat{\theta}((1-t)F + t\Delta_y)\right]_{t=0} = \lim_{t\to 0}\frac{\hat{\theta}((1-t)F+t\Delta_y) - \hat{\theta}(F)}{t} \tag{26}$$

in those $y \in \mathcal{Y}$ where this limit exists. One obtains

$$\mathrm{IF}(y;\hat{\theta},F) = M(\psi,F)^{-1}\psi(y,\hat{\theta}(F)). \tag{27}$$

The *change of variance function* of $\hat{\theta}$ at the distribution F is defined as the $p \times p$-matrix

$$\mathrm{CVF}(y;\hat{\theta},F) = \left[\frac{\partial}{\partial t}V(\hat{\theta},(1-t)F + t\Delta_y)\right]_{t=0}, \tag{28}$$

for all y where this expression exists, where $V(\hat{\theta},F)$ denotes the asymptotic covariance matrix of $\hat{\theta}$ given by (13). The elements (h,k) of M and Q are

$$m_{hk}(\psi,F) \quad = \quad -\int \left[\frac{\partial\psi_h(y,\theta)}{\partial\theta_k}\right]_{\theta=\hat{\theta}(F)} dF(y), \tag{29}$$

$$q_{hk}(\psi, F) = \int \psi_h(y, \hat{\theta}(F))\psi_k(y, \hat{\theta}(F))dF(y). \tag{30}$$

Thus,

$$\left[\frac{\partial}{\partial t} m_{hk}(\psi, (1-t)F + t\Delta_y))\right]_{t=0} = -m_{hk}(\psi, F) - \left[\frac{\partial \psi_h(y,\theta)}{\partial \theta_k}\right]_{\theta=\hat{\theta}(F)}$$

$$- \mathrm{IF}(y; \hat{\theta}, F)^{\mathrm{T}} \int \left[\nabla\left(\frac{\partial \psi_h(y,\theta)}{\partial \theta_k}\right)\right]_{\theta=\hat{\theta}(F)} dF(y), \tag{31}$$

where $\nabla(\partial\psi_h/\partial\theta_k)$ denotes the gradient of $\partial\psi_h/\partial\theta_k$, and

$$\left[\frac{\partial}{\partial t} q_{hk}(\psi, (1-t)F + t\Delta_y))\right]_{t=0} = -q_{hk}(\psi, F) + \psi_h(y, \hat{\theta}(F))\psi_k(y, \hat{\theta}(F))$$

$$+ \mathrm{IF}(y; \hat{\theta}, F)^{\mathrm{T}} \int \left[\psi_h(y,\theta)\nabla\psi_k(y,\theta) + \psi_k(y,\theta)\nabla\psi_h(y,\theta)\right]_{\theta=\hat{\theta}(F)} dF(y), \tag{32}$$

where $\nabla\psi_h$ and $\nabla\psi_k$ denote the gradients of ψ_h and ψ_k. Using these expressions we evaluate

$$\frac{\partial}{\partial t}V = \frac{\partial}{\partial t}\left(M^{-1}QM^{-T}\right) =$$

$$\left(\frac{\partial}{\partial t}M^{-1}\right)QM^{-T} + M^{-1}\left(\frac{\partial}{\partial t}Q\right)M^{-T} + M^{-1}Q\left(\frac{\partial}{\partial t}M^{-T}\right), \tag{33}$$

where

$$\frac{\partial}{\partial t}M^{-1} = -M^{-1}\left(\frac{\partial}{\partial t}M\right)M^{-1}, \quad \text{and} \quad \frac{\partial}{\partial t}M^{-T} = \left(\frac{\partial}{\partial t}M^{-1}\right).^{\mathrm{T}} \tag{34}$$

5.2 Influence and change of variance functions of functions of estimates

Let $\vartheta = g(\theta)$ be a differentiable one-to-one parameter transformation with nonsingular Jacobian $B(\theta) = \partial g(\theta)/\partial\theta$, and let $\hat{\vartheta} = g(\hat{\theta})$ be an estimate of ϑ. In this section, $\hat{\theta} = \hat{\theta}(F)$. One obtains:

$$\mathrm{IF}(y; \hat{\vartheta}, F) = B(\hat{\theta})^{\mathrm{T}}\mathrm{IF}(y; \hat{\theta}, F), \tag{35}$$

$$V(\hat{\vartheta}, F) = B(\hat{\theta})^{\mathrm{T}}V(\hat{\theta}, F)B(\hat{\theta}), \tag{36}$$

$$\begin{aligned}\mathrm{CVF}(y; \hat{\vartheta}, F) = {} & \mathrm{IF}(y; B(\hat{\theta}), F)^{\mathrm{T}} \cdot V(\hat{\theta}, F) \cdot B(\hat{\theta}) \\ & + B(\hat{\theta})^{\mathrm{T}} \cdot \mathrm{CVF}(y; \hat{\theta}, F) \cdot B(\hat{\theta}) \\ & + B(\hat{\theta})^{\mathrm{T}} \cdot V(\hat{\theta}, F) \cdot \mathrm{IF}(y; B(\hat{\theta}), F).\end{aligned} \tag{37}$$

We obtain similar expressions for $\hat{\xi} = h(\hat{\theta})$, where h is a one-dimensional function of θ. It suffices to replace $\hat{\vartheta}$ by $\hat{\xi}$, and $B(\theta)$ with $\nabla h(\theta)$ (the gradient of $h(\theta)$) in (34)–(36).

5.3 Influence and change of variance function of the mean estimate

Assume now that $\hat{\theta} = (\hat{\tau}, \hat{\alpha})$, $\mathrm{IF}(y; \hat{\theta}, F)$, $V(\hat{\theta}, F)$ and $\mathrm{CVF}(y; \hat{\theta}, F)$ are available. Let $\vartheta = g(\theta) = (\exp(\tau), \alpha)^{\mathrm{T}}$, and $\xi = h(\theta) = \sigma\alpha$, with $\sigma = \exp(\tau)$. Let $\hat{\theta} = (\hat{\sigma}, \hat{\alpha})^{\mathrm{T}} = (\exp(\hat{\tau}), \hat{\alpha})^{\mathrm{T}}$, and $\hat{\xi} = \hat{\sigma}\hat{\alpha}$. In this section, $\hat{\alpha} = \hat{\alpha}(F)$, $\hat{\sigma} = \hat{\sigma}(F)$, and $\hat{\tau} = \hat{\tau}(F)$. Then,

$$B(\theta) = \begin{pmatrix} \exp(\tau) & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, $\mathrm{IF}(y; \hat{\alpha}, F)$ remains unchanged, and

$$\begin{aligned} \mathrm{IF}(y; \hat{\sigma}, F) &= \exp(\hat{\tau}) \cdot \mathrm{IF}(y; \hat{\tau}, F), \\ V(\hat{\vartheta}, F) &= \begin{pmatrix} \exp(2\hat{\tau}) v_{\tau\tau} & \exp(\hat{\tau}) v_{\tau\alpha} \\ \exp(\hat{\tau}) v_{\alpha\tau} & v_{\alpha\alpha} \end{pmatrix}, \end{aligned}$$

where $v_{..}$ denote the elements of $V(\hat{\theta}, F)$. The function $\mathrm{CVF}(y; \hat{\vartheta}, F)$ follows from (36), with

$$\mathrm{IF}(y; B(\hat{\theta}), F) = \begin{pmatrix} \mathrm{IF}(y; \hat{\sigma}, F) & 0 \\ 0 & 0 \end{pmatrix}.$$

Finally,

$$\begin{aligned} \mathrm{IF}(y; \hat{\xi}, F) &= \hat{\alpha}\mathrm{IF}(y; \hat{\sigma}, F) + \hat{\sigma}\mathrm{IF}(y; \hat{\alpha}, F), \\ V(\hat{\xi}, F) &= (\hat{\alpha}, \hat{\sigma}) \cdot V(\hat{\vartheta}, F) \cdot (\hat{\alpha}, \hat{\sigma})^{\mathrm{T}}, \\ \mathrm{CVF}(y; \hat{\xi}, F) &= (\mathrm{IF}(y; \hat{\alpha}, F), \mathrm{IF}(y; \hat{\sigma}, F)) \cdot V(\hat{\vartheta}, F) \cdot (\hat{\alpha}, \hat{\sigma})^{\mathrm{T}} \\ &\quad + (\hat{\alpha}, \hat{\sigma}) \cdot \mathrm{CVF}(y; \hat{\vartheta}, F) \cdot (\hat{\alpha}, \hat{\sigma})^{\mathrm{T}} \\ &\quad + (\hat{\alpha}, \hat{\sigma}) \cdot V(\hat{\vartheta}, F) \cdot (\mathrm{IF}(y; \hat{\alpha}, F), \mathrm{IF}(y; \hat{\sigma}, F))^{\mathrm{T}}. \end{aligned}$$

Figure 3 shows the influence function of the standardized shrinking norm and shrinking component estimates of $E(Y)$.

5.4 Maximum bias and variance

The *maximum asymptotic bias* $\mathcal{B}(\hat{\xi}, \xi^*, \epsilon)$ is approximated by

$$\hat{\mathcal{B}}(\hat{\xi}, \xi^*, \epsilon) = \epsilon \cdot \sup_y \mathrm{IF}(y; \hat{\xi}, F_{\theta^*}), \tag{38}$$

where $\gamma^*(\hat{\xi}, F) = \sup_y \mathrm{IF}(y; \hat{\xi}, F)$ is the *gross-error sensitivity* of $\hat{\xi}$ at the distribution F. The *maximum asymptotic variance* $\mathcal{V}(\hat{\xi}, \xi^*, \epsilon)$ is approximated by

$$\hat{\mathcal{V}}(\hat{\xi}, \xi^*, \epsilon) = V(\hat{\xi}, F_{\theta^*}) \exp\left(\epsilon \cdot \sup_y \left[\mathrm{CVF}(y; \hat{\xi}, F_{\theta^*})/V(\hat{\xi}, F_{\theta^*})\right]\right), \tag{39}$$

where $\kappa^*(\hat{\xi}, F) = \sup_y \left[\mathrm{CVF}(y; \hat{\xi}, F)/V(\hat{\xi}, F)\right]$ is the *change-of-variance sensitivity* of $\hat{\xi}$ at the distribution F (Hampel et al., 1986).

5.5 Notes on the numerical evaluation of integrals

Several integrals must be evaluated in order to compute the matrices Q and M (according to (14) and (16)), their derivatives (according to (31) and (32)), as well as the averages (23), and those needed for the determination of $c_b(\alpha)$. We use subroutines for numerical integration taken form the library QUADPACK (Piessens et al., 1980) and pay attention to the following points, in order to maximize accuracy.

1. As the integrands depend on ψ-functions with discontinuous derivatives, it is convenient to partition the integration range $[0, \infty)$ into subintervals limited by the discontinuity points. The upper limit of the numerical integration is set such that $f_\alpha(y)$ is smaller than a specified tolerance.

2. Several integrals can be simplified, by noting that the second derivative of ψ is a combination of Dirac δ-functions. For example, one of the expressions found in (31) becomes:

$$\int \frac{\partial^2}{\partial \tau^2} \psi_{b_1}(\tilde{s}_1(y, \theta) f_\alpha(y) dy =$$
$$a_{11}(\alpha) e^{-\tau} y^2 f_\alpha(y)|_{y_2}^{y_1} + a_{11}(\alpha) e^{-\tau} \int_{y_1}^{y_2} y f_\alpha(y) dy, \tag{40}$$

where y_1 and y_2 are the solutions of $\tilde{s}_1(y,\theta) = \pm b_1$, There are similar opportunities for $\int \frac{\partial^2}{\partial\tau^2}\psi_{b_2}(\tilde{s}_2))$, $\int \frac{\partial^2}{\partial\alpha\partial\tau}\psi_{b_1}(\tilde{s}_1))$, $\int \frac{\partial^2}{\partial\alpha\partial\tau}\psi_{b_2}(\tilde{s}_2))$. ($\tilde{s}_1$, $\tilde{s}_2$ are defined in Section 4.3).

3. In several cases, integration by parts avoids computation of the second derivative of $A_b(\alpha)$. For example, another integral in (31) is:

$$\int \frac{\partial^2}{\partial\alpha^2}\psi_{b_1}(\tilde{s}_1(y,\theta))f_\alpha(y)dy =$$

$$\int [-2\frac{\partial}{\partial\alpha}\psi_{b_1}(\tilde{s}_1(y,\theta))s_2(y,\theta) + \psi_{b_1}(\tilde{s}_1(y,\theta))\ddot{\Gamma}(\alpha) + \psi_{b_1}(\tilde{s}_1(y,\theta))s_2(y,\theta)^2]f_\alpha(y)dy, \quad (41)$$

where $\ddot{\Gamma}$ is the second derivative of Γ. The first derivative of $A_b(\alpha)$ can be approximated with the slope of the linear interpolation described at the beginning of this section.

4. A few series developments and continued fractions are useful to evaluate:

$$G(t,\alpha) = \int_0^t f_\alpha(y)dy, \quad (42)$$

$$J(t,\alpha) = \int_0^t \ln(y)f_\alpha(y)dy, \quad (43)$$

$$K(t,\alpha) = \int_0^t y^k f_\alpha(y)dy. \quad (44)$$

$$(45)$$

We use (Abramowitz and Stegun, 1965; Bhattacharjee, 1970):

$$G(t,\alpha) = \sum_0^\infty t f_{\alpha+n}(t)/(\alpha+n) \quad (46)$$

$$= \frac{e^{-t}t^\alpha}{\Gamma(\alpha)}\left[1 + \sum_{n=1}^\infty \frac{t^n}{(\alpha+1)(\alpha+2)\dots(\alpha+n)}\right] \quad (47)$$

$$= 1 - \frac{e^{-t}t^\alpha}{\Gamma(\alpha+1)}\left[\frac{1}{x+}\,\frac{1-\alpha}{1+}\,\frac{1}{x+}\,\frac{2-\alpha}{1+}\,\frac{2}{x+}\dots\right], \quad (48)$$

$$J(t,\alpha) = \ln(t)G(t,\alpha) - \int_0^t \frac{1}{y}G(y,\alpha)dy \quad (49)$$

$$= \ln(t)G(t,\alpha) - \left[\sum_{j=1}^\infty f_{\alpha+j}(t)\sum_{i=0}^{j-1}\frac{1}{\alpha+i}\right], \quad (50)$$

$$K(t,\alpha) = \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)}G(t,k+\alpha). \quad (51)$$

The developments are terminated when the contribution to the development is not greater than the value of a specified accuracy.

6 Tuning parameters

The most common rule for determining the tuning parameter of an M-estimator is to require that the asymptotic relative efficiency with respect to the maximum likelihood estimator, at the model, equals a given value, e.g., 95%. It is simple to apply this rule to the shrinking norm estimators of $E(Y)$ because b is a single real number. A few values of b and the corresponding relative efficiencies for the standardized shrinking norm estimator of $E(Y)$ are given in Table 1.

For the shrinking component estimators there are several pairs of values of (b_1, b_2) whith the same relative asymptotic efficiency, and the rule does not determine $b = (b_1, b_2)$ uniquely. As a remedy, we minimize the approximation $\hat{\mathcal{V}}(\hat{\xi}, \sigma^*\alpha^*, \epsilon)$ of the maximum asymptotic variance, as a function of (b_1, b_2), where α^* and σ^* are rough preliminary estimates of α and σ. The sensitivity of the estimate to contamination depends then on ϵ and on the optimal value $b(\epsilon)$ of b: the higher the value of ϵ, the less is the estimate sensitive to contamination. The choice of ϵ and $b(\epsilon)$ is made on the ground of collateral information about the frequency of outliers, or by requiring that the asymptotic relative efficiency of the estimate defined by $\psi_{b(\epsilon)}$ equals a given value, e.g., 95%. Table 2 gives a few optimal values $(b_1(\epsilon), b_2(\epsilon))$ for varying ϵ, using the standard parametrization with $\alpha^* = 1, 5, 10$ $(\sigma^* = 1)$. It also gives the corresponding values of $V(\hat{\xi}, F_\alpha)$, $\text{ARE}(\hat{\xi}, \check{\xi})$, and $\hat{\mathcal{V}}(\hat{\xi}, \alpha^*, \epsilon)$.

7 Empirical results

Table 3 and 4 summarize a few Monte Carlo experiments for comparing the shrinking component and the shrinking norm estimates in the standard parametrization. Samples of size 200 were generated from

$$F = (1 - \epsilon)F_\alpha + \epsilon H, \tag{52}$$

using various values of ϵ and α. The tuning parameters were chosen such that both estimates have the same asymptotic relative efficiency (i.e. the same asymptotic variance) at the model. In Table 3, $H = F_{\sigma=2,\alpha}$. In Table 4, $H = \Delta_{y_0}$, where

$$y_0 \approx \text{argmax}_y B(\hat{\xi}, \alpha, \epsilon, y). \tag{53}$$

Note that y_0 is different for the shrinking component and the shrinking norm cases, and that, due to the form of the influence function, it is difficult to determine y_0 with good accuracy in the shrinking norm case. $\hat{\xi}(F)$ was computed using $(1-\epsilon)500$ percentile points of F_α and 500ϵ data points concentrated on y. $\bar{\xi}$ is the arithmetic mean of the values of $\hat{\xi} = \exp(\hat{\tau})\hat{\alpha}$ over 200 experiments; $\text{se}(\bar{\xi})$ is the usual estimate of the standar error of the mean, and $\text{sd}(\bar{\xi}) = (V(\hat{\xi}, F_\alpha)/200)^{1/2}$ is the standard error obtained from the asymptotic variance. (Therefore, it is the same for the shrinking norm and the shrinking component cases, and does not depend on ϵ.) sd and se also estimate the standard error of $\hat{\xi}$. The values of $\text{sd}(\bar{\xi})$ appears to be systematically smaller than those of $\text{se}(\bar{\xi})$.

According to the results in Table 3, the shrinking component estimate performs as well as the shrinking norm estimate under a moderate contamination. However, the results of Table 4, suggest that the shrinking component estimate has a smaller bias than the shrinking norm estimate, while maintaining a higher precision, under a very unfavorable contamination.

Table 5 gives values of $B(\hat{\xi}, \alpha, \epsilon, y_0)$, values of the estimate $\hat{\mathcal{B}}(\hat{\xi}, \alpha, \epsilon)$ of the maximum asymptotic bias, and values of the estimate $(\hat{\mathcal{V}}(\hat{\xi}, \alpha, \epsilon)/200)^{1/2}$ of the maximum asymptotic standard error of $\hat{\xi}$. These standard errors can be compared with those based on the asymptotic variance at the model, i.e., with column $\text{sd}(\bar{\xi})$ in Table 3. $(\hat{\mathcal{V}}(\hat{\xi}, \alpha, \epsilon)/200)^{1/2}$ slightly overestimates the standard error $\text{se}(\bar{\xi})$ given in Table 4, but is closer to $\text{se}(\bar{\xi})$ than $\text{sd}(\bar{\xi})$.

Figure 4 shows empirical bias curves $(\epsilon, B(\hat{\xi}, \alpha, \epsilon, y_0))$ of the shrinking component estimates with options (a), (b), and (c) for the matrix A_b, for $\alpha = 5$. We obtain similar curves for other values of α. The standard parametrization clearly produces the best bias curves. They suggest that the breakdown point of the corresponding estimate is surprisingly high.

In conclusion, the empirical results of this section and the remarks on the computational aspects point out the standardized shrinking component estimator as the best procedure among those considered. Figure 5, shows the histogram of 2385 lengths of stay of patients that where hospitalized in Switzerland during 1988 for "vaginal delivery"; the two densities of the Gamma distribution drawn on the same figure have been determined by means of the maximum likelihood and the recommended procedure. The maximum likelihood estimate was $\hat{\alpha} = 5.31$, $\hat{\sigma} = 1.46$ (mean = 7.81), and the shrinking component estimate was $\hat{\alpha} = 19.9$, $\hat{\sigma} = 0.38$ (mean = 7.59).

Acknowledgment

The authors are grateful to Alex Randriamiharisoa for his generous help with programming.

References

Abramowitz M., Stegun I.A. (1965). *Handbook of Mathematical Functions.* Dover, New York.

Bhattacharjee G.P. (1970). Algorithm AS 32: The Incomplete Gamma Integral. *Applied Statistics,* **19**, 285–287.

Hampel F.R., Ronchetti E., Rousseeuw P.J., Stahel W.A. (1986). *Robust statistics: The approach based on influence functions*, Wiley, New York.

Huber P.J. (1981). *Robust statistics.* Wiley, New York.

Marazzi A. (1993). *Algorithms, routines, and S functions for robust statistics.* Wadsworth & Brooks/Cole, Pacific Grove.

Piessens R., de Doncker-Kapenga E., Überhuber C.W., Kahaner D.K. (1980). *QUADPACK: A subroutine library for automatic integration*, Springer Verlag, Berlin Heidelberg.

Victoria-Feser M.P. (1993). Robust estimation of personal income distribution models. DARP-4, Octobre 1993. Suntory-Toyota International Centre for Economics and Related Disciplines, London School of Economics, Houghton Street, London WC2A 2AE.

Willimann B. (1980). Optimale robuste Schätzungen für die Parameter der Gamma-Verteilung. Diplomarbeit. Fachgruppe für Statistik, ETH Zürich, CH-8057 Zürich.

Table 1. ARE($\hat{\xi}, \check{\xi}, F_{\alpha^*}$) denotes the asymptotic relative efficiency of the standardized shrinking norm estimate $\hat{\xi}$ of $E(Y)$ with respect to the maximum likelihood estimate $\check{\xi}$ at F_{α^*}. b is the tuning parameter.

α^*	b	ARE($\hat{\xi}, \check{\xi}, F_{\alpha^*}$)
1	4.4	0.974
	3.7	0.952
	3.2	0.923
	3.0	0.906
	2.7	0.872
	2.6	0.857
5	3.8	0.975
	3.2	0.952
	2.8	0.923
	2.6	0.901
	2.4	0.873
	2.3	0.855
10	3.7	0.975
	3.1	0.950
	2.8	0.927
	2.6	0.906
	2.4	0.878
	2.2	0.841

Table 2. Values of b_1 and b_2 that minimize the approximation $\hat{\mathcal{V}}(\hat{\xi}, \alpha^*, \epsilon)$ of the maximum asymptotic variance of the standardized shrinking component estimate $\hat{\xi}$ of $E(Y)$ over an ϵ contamination neighborhood $\mathcal{P}_\epsilon$ of F_{α^*}. $V(\hat{\xi}, F_{\alpha^*})$ denotes the asymptotic variance of $\hat{\xi}$ at F_{α^*}, ARE$(\hat{\xi}, \check{\xi}, F_{\alpha^*})$ the asymptotic relative efficiency with respect to the maximum likelihood estimate $\check{\xi}$ at F_{α^*}.

α^*	ϵ	b_1	b_2	$V(\hat{\xi}, \alpha^*)$	ARE$(\hat{\xi}, \check{\xi}, F_{\alpha^*})$	$\hat{\mathcal{V}}(\hat{\xi}, \alpha^*, \epsilon)$
1	0.01	2.7	2.5	1.058	0.945	1.23
	0.02	2.3	2.1	1.093	0.915	1.41
	0.05	1.7	1.7	1.192	0.839	1.95
	0.07	1.3	1.5	1.313	0.762	2.32
	0.10	1.3	1.3	1.363	0.734	2.96
	0.15	1.3	1.3	1.363	0.734	4.35
	0.20	1.1	1.1	1.611	0.621	6.16
5	0.01	2.1	2.7	5.155	0.970	5.62
	0.02	1.7	2.3	5.303	9.943	6.10
	0.05	1.5	1.7	5.523	0.905	7.38
	0.07	1.5	1.7	5.523	0.905	8.29
	0.10	1.3	1.3	5.978	0.836	9.66
	0.15	1.3	1.3	5.978	0.836	12.28
	0.20	1.3	1.3	5.978	0.836	15.61
10	0.01	2.1	2.7	10.207	0.980	11.03
	0.02	1.7	2.3	10.486	0.954	11.81
	0.05	1.5	1.7	10.889	0.918	13.94
	0.07	1.5	1.7	10.889	0.918	15.39
	0.10	1.3	1.3	11.734	0.852	17.63
	0.15	1.3	1.3	11.734	0.852	21.62
	0.20	1.3	1.3	11.734	0.852	26.50

Table 3. For each value of α, 200 estimates $\hat{\xi} = \exp(\hat{\tau})\hat{\alpha}$ are computed using samples of size 200 from $(1-\epsilon)F_\alpha + \epsilon F_{\sigma=2,\alpha}$. $\bar{\xi}$ is the mean estimate over the 200 samples, se($\bar{\xi}$) the standard error of the mean, and sd($\bar{\xi}$) = $(V(\hat{\xi}, F_\alpha)/200)^{1/2}$.

			Shrinking component			Shrinking norm		
ϵ	α	sd($\bar{\xi}$)	b_1, b_2	$\bar{\xi}$	se($\bar{\xi}$)	b	$\bar{\xi}$	se($\bar{\xi}$)
0.00	1	0.077	1.7, 1.7	1.002	0.078	2.4	0.998	0.077
	5	0.166	1.5, 1.7	4.997	0.171	2.6	5.004	0.177
	10	0.234	1.5, 1.7	9.890	0.242	2.6	9.984	0.241
0.05	1	0.077	1.7, 1.7	1.039	0.078	2.4	1.038	0.077
	5	0.166	1.5, 1.7	5.149	0.183	2.6	5.139	0.186
	10	0.234	1.5, 1.7	10.289	0.252	2.6	10.256	0.256
0.10	1	0.082	1.3, 1.3	1.069	0.098	2.0	1.067	0.095
	5	0.172	1.3, 1.3	5.341	0.204	2.2	5.327	0.210
	10	0.242	1.3, 1.3	10.506	0.284	2.2	10.451	0.285

Table 4. For each value of α, 200 estimates $\hat{\xi} = \exp(\hat{\tau})\hat{\alpha}$ are computed using samples of size 200 from $(1-\epsilon)F_\alpha + \epsilon\Delta_{y_0}$, where y_0 is very unfavorable for bias. $\bar{\xi}$ is the mean estimate over the 200 samples, and se($\bar{\xi}$) the standard error of the mean.

			Shrinking component			Shrinking norm		
ϵ	α	ARE	b_1, b_2	$\bar{\xi}$	se($\bar{\xi}$)	b	$\bar{\xi}$	se($\bar{\xi}$)
0.05	1	0.84	1.7, 1.7	1.143	0.091	2.4	1.146	0.095
	5	0.90	1.5, 1.7	5.233	0.198	2.6	5.298	0.195
	10	0.91	1.5, 1.7	10.316	0.258	2.6	10.372	0.299
0.10	1	0.73	1.3, 1.3	1.260	0.131	2.0	1.298	0.124
	5	0.81	1.3, 1.3	5.508	0.215	2.2	5.583	0.231
	10	0.85	1.3, 1.3	10.630	0.288	2.2	10.752	0.306

Table 5. $B = B(\hat{\xi}, \alpha, \epsilon, y_0)$ is the maximum bias of $\hat{\xi}$ over a point mass contamination $\epsilon\Delta_y$, for varying y. $\hat{\mathcal{B}} = \hat{\mathcal{B}}(\hat{\xi}, \alpha, \epsilon)$ is the estimate of the maximum asymptotic bias over an ϵ-contamination neighborhood of F_α based on the influence function. $\hat{\mathcal{V}} = \hat{\mathcal{V}}(\hat{\xi}, \alpha, \epsilon)$ is the estimate of the maximum asymptotic variance over an ϵ-contamination neighborhood of F_α based on the change of variance function.

		Shrinking component			Shrinking norm		
ϵ	α	B	$\hat{\mathcal{B}}$	$(\hat{\mathcal{V}}/200)^{1/2}$	B	$\hat{\mathcal{B}}$	$(\hat{\mathcal{V}}/200)^{1/2}$
0.05	1	0.137	0.118	0.098	0.148	0.156	0.105
	5	0.243	0.215	0.192	0.288	0.280	0.211
	10	0.322	0.288	0.264	0.378	0.369	0.290
0.10	1	0.256	0.207	0.121	0.304	0.299	0.136
	5	0.477	0.388	0.219	0.605	0.530	0.263
	10	0.630	0.519	0.296	0.791	0.704	0.359

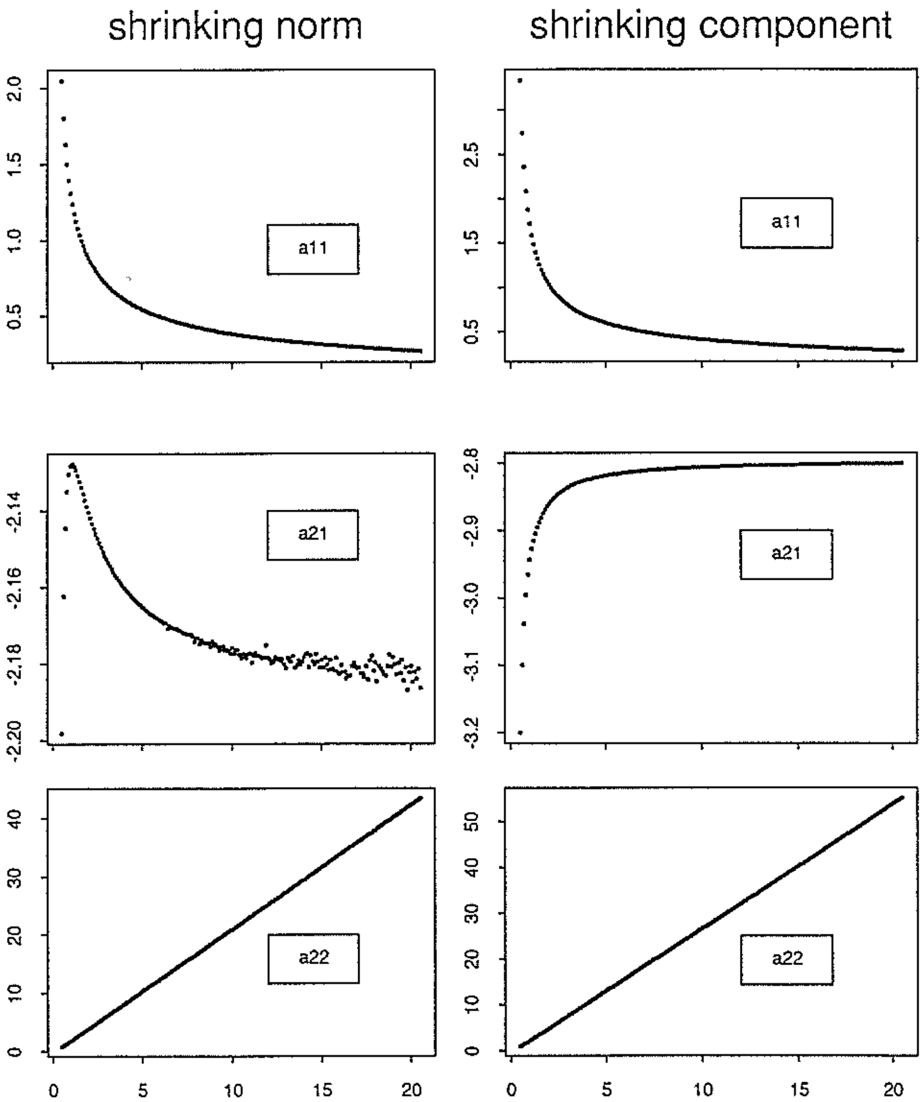

Figure 1. Elements of the matrix $A_b(\alpha)$ for varying α (shrinking norm $b = 2.6$, shrinking component $b_1 = 1.5, b_2 = 1.7$).

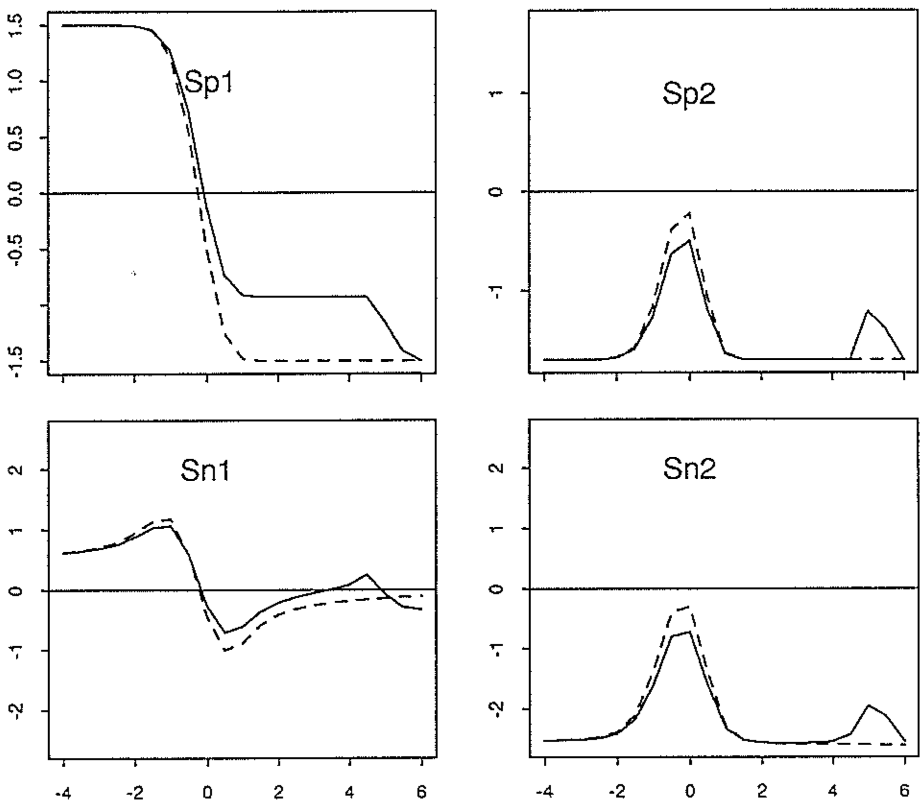

Figure 2. Functions $S_{p1}, S_{p2}, S_{n1}, S_{n2}$ for varying σ (log scale) on artificial data sets.

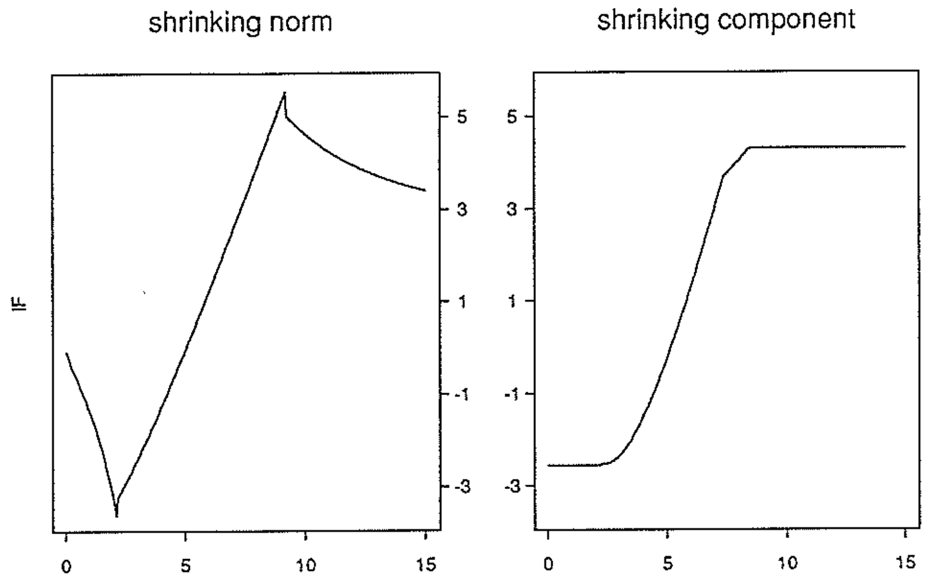

Figure 3. Influence functions of the standardized shrinking norm ($\alpha = 5$, $b = 2.6$) and shrinking component estimates ($\alpha = 5, b_1 = 1.5, b_2 = 1.7$).

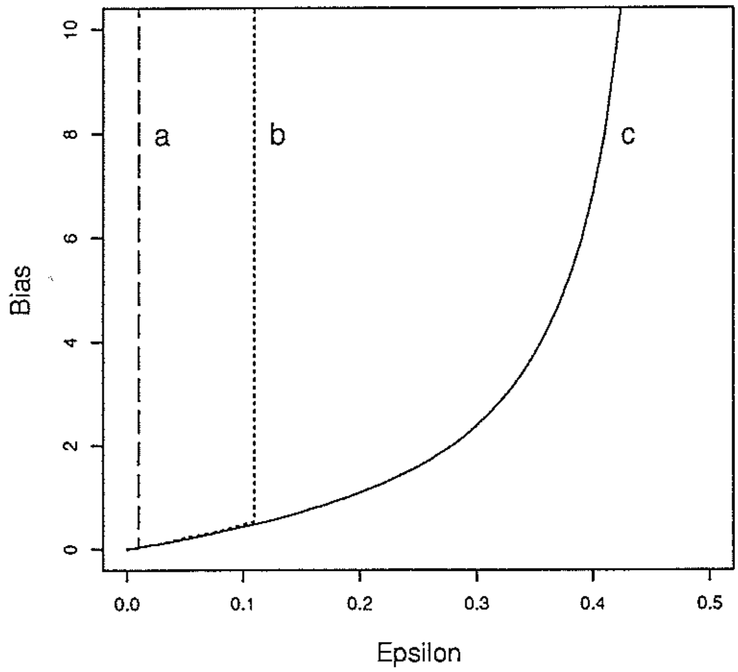

Figure 4. Bias curves of the shrinking component estimates

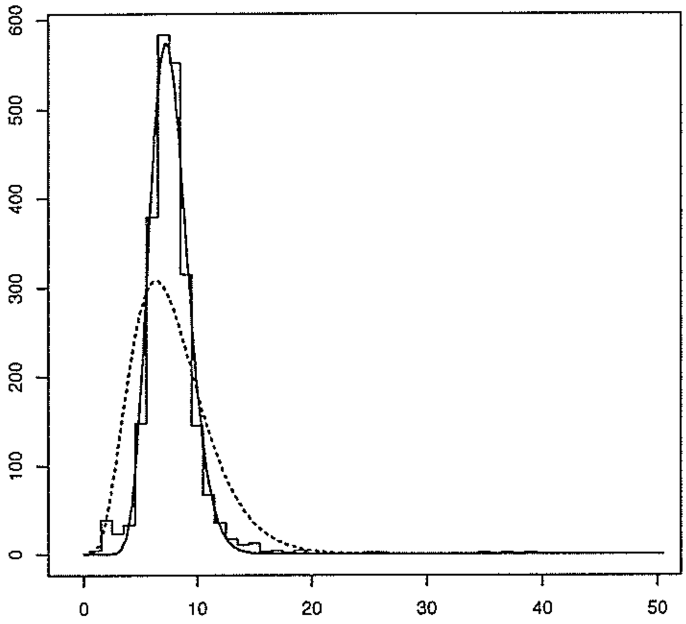

Figure 5. Histogram of 2395 lengths of stay of patients that where hospitalized in Switzerland during 1988 for "vaginal delivery"; the two densities of the Gamma distribution have been determined by means of the maximum likelihood (dotted line) and the standardized shrinking component estimate (solid line).

Pubblicazioni del Centro di Studi Bancari

Le società a tassazione speciale

Le società di partecipazioni, le società holding, le società di sede, le società ausiliarie e le società di servizio

Vi sono società che in materia di imposte beneficiano di un trattamento particolare, poiché in luogo di essere tassate sull'utile e sul capitale come le altre società anonime, lo sono in modo agevolato. Per consentire al lettore di approfondire questa problematica, gli autori esaminano i problemi connessi all'imposizione di queste società, secondo il diritto vigente, il diritto federale in materia di imposte dirette entrato in vigore il 1° gennaio 1995, le disposizioni del diritto internazionale e le norme contro l'abuso e la pianificazione fiscale internazionale.

Gli autori: Marco Bernasconi, Centro di Studi Bancari, Lugano Vezia; Nico H. Burki, Studio legale Bär & Karrer, Zurigo; Felix R. Ehrat, Studio legale Bär & Karrer, Lugano e Zurigo; Angelo Digeronimo, Amministrazione federale delle contribuzioni, Berna; Guglielmo Maisto, Studio Maisto e Miscali, Milano.

144 pp. Fr. 35.-- (banche ABT Fr. 24.--)

Moneta e banconote

Per uno studio interdisciplinare su oggetto, idee e immagini

Il volume rappresenta un tentativo originale d'integrazione della forma e del contenuto. Grazie alla riproduzione di oltre 60 schizzi è proposta la storia delle banconote svizzere in immagine.

Accanto a questo incantevole percorso estetico vengono riportati otto saggi di economisti, storici, sociologi, psicologi sul ruolo fondamentale giocato dalla moneta nella società. Il tentativo dichiarato è di far sparire dietro le quinte —anche grazie alla suggestione delle immagini— gli steccati disciplinari, lasciando sul palcoscenico l'attore principale: la moneta.

Gli autori: René Chopard, Centro di Studi Bancari, Lugano Vezia; Augusto Graziani, Università La Sapienza, Roma; Jean-Michel Servet, Università Lumière, Lione; Aldo De Maddalena, Università Bocconi, Milano; Renzo Carli, Università La Sapienza, Roma; Alessandro Cavalli, Università degli Studi, Pavia; Jérôme Baratelli, Ecole des arts décoratifs, Ginevra; Michel de Rivaz, Banca nazionale svizzera, Berna.

208 pp. Fr. 47.-- (banche ABT Fr. 33.--)

Nella stessa collana

CLAUDIO BORGHESE, PAOLO LAVEZZO, ANDREA MANZITTI, MARCO BERNASCONI, *La liberalizzazione valutaria in Italia ed il "monitoraggio" fiscale*, Centro di Studi Bancari, Meta-Edizioni, Vezia, Bellinzona, 1991, 88 pp. *Fr. 33.-- (banche ABT Fr. 26.40)*

MARKUS LUSSER, PAOLO CLAROTTI, PAOLO BERNASCONI, ALVARO CENCINI, BERNARD SCHMITT, PASCAL BRIDEL, *Europa '93! e la piazza finanziaria svizzera? Riflessioni sulle conseguenze bancarie, giuridiche e monetarie dell'integrazione comunitaria*, a cura di RENÉ CHOPARD, prefazione di CLAUDIO GENERALI, Centro di Studi Bancari, Meta-Edizioni, Vezia, Bellinzona, 1992, 144 pp. *Fr. 35.-- (banche ABT Fr. 24.--)*

RENÉ CHOPARD, *Il sistema bancario ticinese e la piazza finanziaria svizzera, Caratteristiche, evoluzione, prospettive nel contesto europeo e internazionale*, Centro di Studi Bancari, Meta-Edizioni, Vezia, Bellinzona, 1992, 233 pp., 172 tab. *Fr. 38.-- (banche ABT Fr. 25.--)*

PAOLO BERNASCONI, MARCO BORGHI, DANIEL ZUBERBÜHLER, URS PHILIPP ROTH, ALEXIS P. LAUTENBERG, OTTO STICH, *Il sistema bancario svizzero contro il riciclaggio. Le Direttive della Commissione federale delle banche; la Convenzione dell'Associazione svizzera dei banchieri; le Raccomandazioni del Gruppo d'azione finanziaria internazionale - GAFI*, a cura di RENÉ CHOPARD, prefazione di FLAVIO PEDRAZZOLI, Centro di Studi Bancari, Meta-Edizioni, Vezia, Bellinzona, 1993, 216 pp. *Fr. 47.-- (banche ABT Fr. 33.--)*

Le pubblicazioni possono essere comandate a: Centro di Studi Bancari, Biblioteca, Villa Negroni, CH - 6943 Vezia-Lugano

Quaderni di ricerca

No. 1 RENÉ CHOPARD
La banca ticinese. Alcune considerazioni sul fenomeno di apertura e il concetto di identità, febbraio 1991, 44 pp. *Fr. 5.--*

No. 2 ALVARO CENCINI
Les pays face au problème de la dette, giugno 1991, 28 pp. *Fr. 5.--*

No. 3 RENÉ CHOPARD
Sciences économiques et systèmes monétaires informels, settembre 1991, 24 pp. *Fr. 5.--*

No. 4 ALFIO MARAZZI, A. RANDRIAMIHARISOA, G. VAN MELLE
Algorithms and programs for bounded-influence estimates in discrete generalized linear models, ottobre 1991, 44 pp. *Fr. 5.--*

No. 5 a cura di RENÉ CHOPARD
Corsi speciali 1991; Bibliografie, gennaio 1992, 32 pp. *Fr. 5.--*

No. 6 MAURO BARANZINI
The theory of income distribution and the controversy between Cambridge (U.K.) and Cambridge (Mass., U.S.A.), aprile 1992, 37 pp. *Fr. 5.--*

No. 7 AURELIO MATTEI
Le previsioni congiunturali, maggio 1992, 36 pp. *Fr. 5.--*

No. 8 MAURO PICCHI
Le prescrizioni della Banca Nazionale in materia d'esportazione di capitali e il loro influsso sul mercato finanziario svizzero, settembre 1992, 38 pp. *Fr. 5.--*

No. 9 RENÉ CHOPARD
La partecipazione svizzera allo Spazio Economico Europeo: le conseguenze sul sistema bancario ticinese, novembre 1992, 54 pp. *Fr. 5.--*

No. 10 SYLVAIN MATTHEY
La liberté d'établissement et de prestation de services des banques dans l'Espace Economique Européen. Un aperçu de l'acquis communautaire, novembre 1992, 22 pp. *Fr. 5.--*

No. 11 MATTHÄUS DEN OTTER
Entraide administrative internationale et le secret bancaire suisse, novembre 1992, 16 pp. *Fr. 5.--*

No. 12 ALVARO CENCINI
L'inflation: une analyse fondamentale, maggio 1993, 19 pp. *Fr. 5.--*

No. 13 AMILCARE BERRA
Passato e presente della banca ticinese, settembre 1993, 12 pp. *Fr. 5.--*

No. 14 ROBERTO SCAZZIERI
Ciclo economico e cambiamenti strutturali dell'economia, novembre 1993, 13 pp. *Fr. 5.--*

No. 15 SAVERIO ALBERTI, ORLANDO NOSETTI
Ristrutturazione e "Turnaround", marzo 1994, 35 pp. *Fr. 5.--*

No. 16 SAVERIO ALBERTI
Banca e crisi aziendale, febbraio 1995, 48 pp. *Fr. 5.--*

No. 17 ALFIO MARAZZI
Robust estimation of the mean of an asymmetric distribution: the Gamma distribution case, marzo 1995, 27 pp. *Fr. 5.--*

No. speciale ALICE MORETTI
Villa Negroni, Cenni storici, settembre 1992, 33 pp. *Fr. 5.--*

I Quaderni possono essere comandati a: Centro di Studi Bancari, Biblioteca, Villa Negroni, CH - 6943 Vezia-Lugano