ESTIMATING AND TESTING THE MEANS OF ASYMMETRIC DISTRIBUTIONS USING M-ESTIMATORS

Alfio Marazzi*

No. 20, gennaio 1997

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

Stime e test statistici della media di distribuzioni asimmetriche con l'uso delle stime M

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 alcune stime M per la media di distribuzioni asimmetriche basate sui modelli di distribuzione Lognormale, Weibull e Gamma. Queste stime vengono poi usate per definire alcuni test statistici (basati sul metodo del bootstrap) per confrontare le medie di distribuzioni asimmetriche.

INDICE Pagina

Estimating and testing the means of asymmetric distributions using M-estimators

Alfio Marazzi*

Abstract. Statistical analysis of simple data sets (e.g. the one sample and the two sample problems), is often based on classical parametric procedures (e.g., the t-test) or on nonparametric procedures based on ranks (e.g., the Wilcoxon test). Both approaches are usually inappropriate for the purpose of estimating and comparing the means of asymmetric distributions in the presence of outliers (such as gross errors or abnormal behaviours). This paper discusses the application of bootstrap and M-estimators of three popular asymmetric models – Lognormal, Gamma and Weibull – to the estimation as well as to the testing problem. S functions for computing the described procedures are made available.

Keywords. Asymmetric distributions, robust estimation, M-estimators, robust tests, bootstrap tests, Lognormal distribution, Gamma distribution, Weibull distribution.

1 Introduction

Asymmetric distributions of positive random variables occur in many statistical applications concerning, for example, survival data (medicine), yield data (agriculture, industry), failure times (industry), income and resource consumption measures (econometric studies).

We focus on two particular aspects of this kind of data. First, the population mean or the total of a finite population is the characteristic of interest; because the total is a multiple of the mean, we concentrate on estimation and testing of the population mean. Second, the data may contain values that are markedly different from most others. When some of these values are observed, the sample mean can be much larger than when none is observed and, therefore, it varies strongly from sample to sample.

Various rules for removing or downweighing discrepant values are used in practice in order to improve the stability, i.e., the robustness, of the estimation and the associated inferences. For example, a practitioner will often try to use a simple transformation, (e.g., the log) and model the transformed data as a usual symmetric location problem. Once cast in this way, many robust procedures are available. Unfortunately, their application is often incorrect: for example, the hypothesis that the shift between two transformed distributions is null, is often tested in place of the hypothesis that the untransformed population means are equal.

* Institut universitaire de médecine sociale et préventive, University of Lausanne, Bugnon 17, CH 1005 Lausanne, Switzerland

The purpose of this paper is to discuss estimation and testing procedures based on robust M-estimators and three widely used asymmetric parametric models: the Lognormal, the Weibull, and the Gamma distributions. The procedures (that can be extended to a wider class of distributions) are based on the possibility of fitting the model to the majority of the data, the fit being not unduly influenced by a minority of extreme observations. The mean estimate is then defined as the mean of the adjusted model and tests are derived from this estimate. We illustrate the procedures in the following example.

Example. Figure 1 (a) shows the histogram of 315 lengths of stay (LOS) in days of patients hospitalized in Belgium during 1988 for certain "disorders of the nervous system", and Figure 1 (b) the histogram of 32 LOS of patients hospitalized during the same year in Switzerland for the same kind of illness. The original data are summarized in Table 1. Both displays (but not the data) are truncated at LOS=50. LOS is an important indicator of hospital costs (Lave and Leinhardt, 1976). The LOS means of "medically homogeneous" groups of patients (e.g., so-called "diagnosis related groups") within a hospital are used for budgeting (the cost of n patients is proportional to n times their mean LOS), and the means of different hospitals are used to compare costs and explore possible reductions.

LOS distributions are typically asymmetric and contain outliers, whose value and frequency fluctuate from sample to sample, e.g., from year to year. This makes the usual statistical summaries largely unreliable. For example, the arithmetic means of the Belgian and Swiss data shown in Figure 1 are 7.87 and 25.47 days, respectively. Both distributions contain outliers (not shown on the histograms); for example, the Swiss data contain two stays of 374 and 198 days; removing them reduces the Swiss mean to 8.10. Removing all LOS larger than 50 reduces the means to 6.58 (Belgium) and 4.41 (Switzerland).

Usually, around 500 groups of patients are routinely analyzed in the management of a single hospital. This clearly requires an automatic procedure for conveniently treating the outliers and producing reliable statistics.

The three models studied in this paper describe the data of Table 1 equally well. (We are not bothered by the use of continuous models for discrete data, because LOS is essentially a rounded continuous variable with a large number of values.) For example, the density functions drawn with solid lines in Figure 1 have been obtained by robustly fitting a Gamma distribution to each of the data sets according to the procedure described in Section 3. They conveniently summarize the pattern of the majority of the data. The means of the fitted densities are 5.89 (Belgium) and 4.81 (Switzerland) when the complete data set is used, while they are 5.66 and 4.22 when outliers are removed. The dotted density functions have been fitted to the entire data set by maximum likelihood.

The usual pooled t-test for comparing means – inappropriate, but often used for asymmetric distributions – has a one-sided attained significance level (ASL = $P(t \le t_{obs})$, where t_{obs} is the observed value of t) of 0.9999 on the entire data set, and of 0.08 when outliers are removed. The corresponding ASL's of a bootstrap test based on the robust Gamma fits (Section 4) are 0.09 and 0.04. To study the sensitivity of these results to model selection, the Weibull and the Lognormal models have been fitted to the same data; results are summarized in Table 2. We note that the robust procedures based on the different models give similar results that are not substantially influenced by large LOS values, whereas those given by the t-test are completely distorted.

For the reader convenience we summarize in the remaining part of this section some basic concepts and formulae of robustness that will be used in later parts of the paper. Section 2 describes the models, Section 3 the estimation procedures, and Section 4 the testing procedure. Section 5 completes the examples.

Background. Let $y_1, \ldots, y_n$ be independent observations of a continuous one-dimensional random variable Y with unknown distribution F. Let Δ_y denote the distribution function of a point mass at y and $F_n(y) = (1/n) \sum_{i=1}^{n} \Delta_{y_i}(y)$ the empirical distribution function (e.d.f.) of $y_1, \ldots, y_n$. Classical statistics assumes that Y is distributed according to some distribution F_θ, with density f_θ, for an unknown value of a parameter $\theta \in \mathbb{R}^p$. The maximum likelihood (ML) criterion is often used to estimate this value of θ.

In robustness theory, one assumes that F belongs to a neighborhood $\mathcal{P}$ of one of the F_θ, say, F_{θ^*}. For example, one can use the *ϵ-contamination model* (Huber, 1981)

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

where ϵ is a given percentage of *gross errors*, e.g., $\epsilon = 5\%$. The purpose of robust methods is to infer the value θ^* that characterizes the main part of the population in such a way that the *contamination* ϵH has a limited influence on the inference.

As estimators we consider statistics $T_n(y_1, \ldots, y_n)$ that can be represented as functionals of F_n, i.e., $T_n = T(F_n)$. One way of assessing the robustness of an estimator is by means of the *influence function* (IF), which is defined, at the model F_θ, by

$$\mathrm{IF}(y; T, F_\theta) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F_\theta + \epsilon\Delta_y) - T(F_\theta)}{\epsilon}. \tag{2}$$

It describes the effect of a small contamination ($\epsilon\Delta_y$) at the point y on the estimate, standardized by the mass of the contamination (Hampel et al., 1986). Thus, the bias of the estimator caused by the contamination can be approximated by $\epsilon IF(y; T, F_\theta)$. A desirable property of an estimator is a bounded IF. Such an estimator is called *B-robust* (or bias-robust). The IF of a ML estimator is proportional to its score function $s(y, \theta) = (\partial/\partial\theta) \ln f_\theta(y)$. Unfortunately most models, including Lognormal, Gamma, and Weibull, have unbounded score functions.

We use robust *M-estimators* defined as solutions $\hat{\theta}$ of

$$\sum_{i=1}^{n} \psi(y_i, \theta) = 0, \tag{3}$$

where ψ denotes a given vector of (bounded) functions of (y, θ) with values in $\mathbb{R}^p$. Equation (3) is a generalization of the ML equation. Results concerning the asymptotic properties of M-estimators are available in the literature (see, e.g., Rieder, 1994). In particular, let $\hat{\theta}(F)$ denote the solution of

$$\int \psi(y, \theta) dF(y) = 0 \tag{4}$$

and note that $\hat{\theta}$ is the solution of (4) for $F = F_n$, i.e., $\hat{\theta} = \hat{\theta}(F_n)$.

Under certain regularity conditions (including observation independence) $\hat{\theta}$ is asymptotically normally distributed with mean $\hat{\theta}(F)$ and *asymptotic covariance matrix*

$$V(\hat{\theta}, F) = M(\psi, F)^{-1} Q(\psi, F) M(\psi, F)^{-T}, \tag{5}$$

where

$$Q(\psi, F) = \int \psi(y, \hat{\theta}(F)) \psi(y, \hat{\theta}(F))^{\mathrm{T}} dF(y), \tag{6}$$

$$M(\psi, F) = -\int \left[\frac{\partial}{\partial \theta} \psi(y, \theta)\right]_{\theta=\hat{\theta}(F)} dF(y). \tag{7}$$

In applications, $V(\hat{\theta}, F_n)/n$ or $V(\hat{\theta}, F_{\hat{\theta}})/n$ are used as approximations for the covariance matrix of $\hat{\theta}$. Moreover, the IF of an M-estimator is

$$IF(y; \hat{\theta}, F_\theta) = M(\psi, F_\theta)^{-1} \psi(y, \theta), \tag{8}$$

which is bounded when ψ is a bounded function of y.

Hampel et al. (1986) define *optimal B-robust M-estimators* for general parametric models, that minimize the trace of the asymptotic covariance matrix under the constraint that the IF is bounded. There are several versions of optimal B-robust estimators, depending on the way one chooses to bound IF, but they have a common feature: the function ψ is a composition of the score function s and *Huber's function* h_b *with tuning constant* b,

$$h_b(z) = z \min(1, b/||z||), \qquad z \in \mathbb{R}^p,$$

where $||z||$ denotes the Euclidean norm of z. Among the proposals of Hampel et al. (1986), are the optimal standardized B-robust estimator (B_s-estimator) and the optimal standardized B-robust estimator for partitioned parameters (B_s^p-estimator). These are studied and compared, in the case of the Gamma distribution, by Marazzi and Ruffieux (1996), who conclude that the statistical performances (variance and contamination bias) of the B_s- and the B_s^p-estimators are similar; yet, the B_s^p-estimator is computationally simpler and more reliable, and should be preferred for practical use. The definition of the B_s^p-estimator is given in Section 3. An application of the B_s-estimator is described in Victoria-Feser and Ronchetti (1994).

2 The models

A random variable X is distributed according to a *Lognormal distribution* if $Y = \ln X$ has a Gaussian density

$$f_{\lambda,\sigma}(y) = (\sqrt{2\pi}\sigma)^{-1} \exp\left(-(y-\lambda)^2/(2\sigma^2)\right), \quad -\infty < y < \infty, \ -\infty < \lambda < \infty, \ \sigma > 0, \tag{9}$$

with location λ, scale σ, and score functions

$$s_1(y, \theta) = z/\sigma, \quad \text{and} \quad s_2(y, \theta) = (z^2 - 1)/\sigma, \tag{10}$$

where $\theta = (\lambda, \sigma)^{\mathrm{T}}$ (a column vector) and $z = (y - \lambda)/\sigma$. The functions σs_1 and σs_2 are invariant under location and scale trasformations. We define $\mu = E(X) = \exp(\lambda)\exp(\sigma^2/2)$.

A random variable X has a *Weibull distribution* if the density of $Y = \ln X$ is

$$f_{\tau,v}(y) = \exp[(y-\tau)/v - \exp((y-\tau)/v)]/v, \quad -\infty < y < \infty, \ -\infty < \tau < \infty, \ \sigma > 0, \tag{11}$$

where τ is a location and v is a scale parameter. The score functions are

$$s_1(y,\theta) = (\exp(z) - 1)/v \quad \text{and} \quad s_2(y,\theta) = [z(\exp(z) - 1) - 1]/v, \tag{12}$$

where $\theta = (\tau, v)^{\mathrm{T}}$ and $z = (y - \tau)/v$. The functions vs_1 and vs_2 are invariant under location and scale transformations. The distribution of X, which is usually expressed by means of the parameters $\alpha = 1/v$ (shape) and $\sigma = \exp(\tau)$ (scale), does not share this property. We define, $\mu = E(X) = \exp(\tau)\Gamma(1+v)$, where $\Gamma(\cdot)$ is the Gamma function.

A random variable Y has a *Gamma distribution* if its density is

$$f_{\tau,\alpha}(y) = (\sigma\Gamma(\alpha))^{-1} (y/\sigma)^{\alpha-1} \exp(-y/\sigma), \quad y > 0, \ \sigma > 0, \ \alpha > 0, \tag{13}$$

where α is the shape and $\sigma = \exp(\tau)$ the usual scale parameter. With $\theta = (\tau, \alpha)^{\mathrm{T}}$, the score functions are

$$s_1(y,\theta) = z - \alpha \quad \text{and} \quad s_2(y,\theta) = \ln z - \dot{\Gamma}(\alpha), \tag{14}$$

where $\dot{\Gamma}(\alpha) = (d/d\alpha)\ln\Gamma(\alpha)$ denotes the Digamma function and $z = y/\sigma$. These score functions are scale invariant but not shape invariant. We define, $\mu = E(Y) = \alpha\exp(\tau)$.

A common feature of the models described above is that

$$\mu = \kappa_1(\theta_1)\kappa_2(\theta_2), \tag{15}$$

where κ_1 and κ_2 are two real functions of the first resp., the second component of θ. This property, which is shared by other models – for example, the Pareto distribution of first kind (Johnson et al., 1994, p.574) – characterizes the testing procedures described in Section 4. We define $\zeta(\theta_2, \mu) = \kappa_1^{-1}(\mu/\kappa_2(\theta_2))$, thus, $\theta_1 = \zeta(\theta_2, \mu)$.

3 The estimators

Definition. Let $s(\theta, y)$ denote the (column) vector of score functions. For a fixed value of a parameter $\underline{b}$ (see below), the *B_s^p-estimate* $\hat{\theta}$ of θ is defined as a solution of

$$\sum_{i=1}^{n} \psi^{\underline{b}} \left[A^{\underline{b}}(\theta)(s(y_i,\theta) - c^{\underline{b}}(\theta)) \right] = 0, \tag{16}$$

where:

– The function $\psi^{\underline{b}}(z)$, $z = (z_1, z_2) \in \mathbb{R}^2$, is defined by

$$\psi^{\underline{b}}(z) = (h_{b_1}(z_1), h_{b_2}(z_2))^{\mathrm{T}}, \quad \underline{b} = (b_1, b_2) \in \mathbb{R}^2; \tag{17}$$

– $A^{\underline{b}}(\theta)$ is a 2×2 nonsingular lower triangular matrix defined by the equation

$$\int \psi^{\underline{b}} \left[A^{\underline{b}}(\theta)(s(y,\theta) - c^{\underline{b}}(\theta))\right] \psi^{\underline{b}} \left[A^{\underline{b}}(\theta)(s(y,\theta) - c^{\underline{b}}(\theta))\right]^{\mathrm{T}} f_\theta(y)dy = I; \qquad (18)$$

– The function $c^{\underline{b}}(\theta)$ is defined by

$$\int \psi^{\underline{b}} \left[A^{\underline{b}}(\theta)(s(y,\theta) - c^{\underline{b}}(\theta))\right] f_\theta(y)dy = 0; \qquad (19)$$

– The parameter $\underline{b}$ is fixed by the user and called the *tuning constant*. Note that $A_{\underline{b}}(\theta)$ does not exist if b_1 or b_2 or both are smaller than certain values (close to 1).
  The estimator $\hat{\theta}$ defined by (16)-(19) is an M-estimator according to (3) with

$$\psi(y,\theta) = \psi^{\underline{b}}[A^{\underline{b}}(\theta)(s(y,\theta) - c^{\underline{b}}(\theta))].$$

Therefore, (4)-(8) completely specify $\hat{\theta}(F)$, $V(\hat{\theta}, F)$, and $\mathrm{IF}(y; \hat{\theta}, F_\theta)$.

Interpretation. The matrix $A^{\underline{b}}(\theta)$ and the vector $c^{\underline{b}}(\theta)$ can be viewed as Lagrange multipliers for the constraints resulting from the optimality problem (Hampel et al., 1986). The inverse of $A^{\underline{b}}(\theta)$ may also be interpreted as the transposed Jacobian of a parameter transformation $\vartheta = g(\theta)$; ϑ is called the *standard parameter*. Note that $A^{\underline{b}}$ makes $Q(\psi, F) = I$ and that Q coincides with the Fisher information matrix when $b_1 = b_2 = \infty$; thus, the components of $\hat{\vartheta}$ are less correlated than those of $\hat{\theta}$. The introduction of $A_{\underline{b}}(\theta)$ greatly improves the stability of the numerical procedures for solving (16), with respect to the case $A^{\underline{b}}(\theta) \equiv 1$. The B^p_s-estimator is a modification of the ML estimator, where the corrected likelihood scores $A^{\underline{b}}(\theta)[s(y_i,\theta) - c^{\underline{b}}(\theta)]$ of ϑ are truncated at $\pm b_1$ and $\pm b_2$, respectively. The "correction" $c^{\underline{b}}$ is needed to ensure that the truncated transformed likelihood scores have expectation zero at the model. Figure 2 shows, for the Gamma model, the domains of Y-values that correspond to unmodified and truncated values of $A^{\underline{b}}(\theta)[s(y_i,\theta) - c^{\underline{b}}(\theta)]$. Observations which do not belong to the central interval in the density diagram are downweighted by the estimation procedure.

Special cases. In the Lognormal case, thanks to the invariance of the scores (10), one obtains

$$A^{\underline{b}}(\theta) = \sigma \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \qquad c^{\underline{b}}(\theta) = \frac{1}{\sigma} \begin{pmatrix} 0 \\ c_2 \end{pmatrix}, \qquad (20)$$

where the constants a_{11}, a_{22}, and c_2 are defined by (18)-(19) and (9) with $\theta^{\mathrm{T}} = (\lambda, \sigma) = (0, 1)$ and are numerically computed before solving (16). For symmetry reasons, the covariance matrix $V(\hat{\theta}, F)$ is diagonal; in practice one evaluates $V(\hat{\theta}, F) \approx \hat{\sigma}^2 V(\hat{\theta}, F_{0,1})$ numerically. We set $\hat{\mu} = \exp(\hat{\lambda} + \hat{\sigma}^2)$ and obtain

$$V(\hat{\mu}, F) = \delta(\hat{\theta})^{\mathrm{T}} V(\hat{\theta}, F)\delta(\hat{\theta}) = \hat{\mu}(F)^2 V(\hat{\lambda}, F) + (\hat{\mu}(F)\hat{\sigma}(F))^2 V(\hat{\sigma}, F), \qquad (20)$$

where $\delta(\theta) = (\partial\mu/\partial\lambda, \partial\mu/\partial\sigma)^{\mathrm{T}} = (\mu, \mu\sigma)^{\mathrm{T}}$.

In the Weibull case, thanks to the invariance of the scores (12), one obtains $A^{\underline{b}}(\theta) = v\dot{A}^{\underline{b}}$ and $c^{\underline{b}}(\theta) = \dot{c}^{\underline{b}}/v$, where $\dot{A}^{\underline{b}}$ and $\dot{c}^{\underline{b}}$ are the solutions of (18)-(19) and (11) with $\theta^{\mathrm{T}} = (\tau, v) = (0,1)$ and are numerically computed before solving (16). We set $\hat{\alpha} = 1/\hat{v}$, $\hat{\sigma} = \exp(\hat{\tau})$, $\hat{\mu} = \hat{\sigma}\Gamma(1+1/\hat{\alpha})$ and obtain

$$V(\hat{\mu}, F) = \delta(\hat{\sigma},\hat{\alpha})^{\mathrm{T}} V((\hat{\sigma},\hat{\alpha}), F)\delta(\hat{\sigma},\hat{\alpha}) = \delta(\hat{\sigma},\hat{\alpha})^{\mathrm{T}} J(\hat{\theta})^{\mathrm{T}} V(\hat{\theta}, F) J(\hat{\theta})\delta(\hat{\sigma},\hat{\alpha}), \tag{21}$$

where

$$\delta(\sigma,\alpha) = (\partial\mu/\partial\sigma, \partial\mu/\partial\alpha)^{\mathrm{T}} = (\Gamma(1+1/\alpha), -\sigma\dot{\Gamma}(1+1/\alpha)/\alpha^2)^{\mathrm{T}},$$
$$J(\theta) = (\partial/\partial\theta)(\exp(\tau), 1/v))^{\mathrm{T}} = \begin{pmatrix} \exp(\tau) & 0 \\ 0 & -1/v^2 \end{pmatrix},$$

and $V(\hat{\theta}, F)$ is defined by (5)-(7). In practice one sets $V(\hat{\theta}, F) \approx \hat{v}^2 V(\hat{\theta}, F_{0,1})$.

In the Gamma case with $\theta = (\tau, \alpha)^{\mathrm{T}}$, thanks to scale invariance, $A^{\underline{b}}(\theta)$ and $c^{\underline{b}}(\theta)$ do not depend on τ. They depend, however, on α and must be computed and stored before solving (16). We set $\hat{\sigma} = \exp(\hat{\tau})$, $\hat{\mu} = \hat{\alpha}\hat{\sigma}$ and obtain

$$V(\hat{\mu}, F) = (\hat{\alpha}, \hat{\sigma}) J^{\mathrm{T}}(\hat{\theta}) V(\hat{\theta}, F) J(\hat{\theta})(\hat{\alpha}, \hat{\sigma})^{\mathrm{T}}, \tag{22}$$

where

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

is the Jacobian of $g(\theta) = (\exp(\tau), \alpha)^{\mathrm{T}}$ and $V(\hat{\theta}, F) \approx \hat{\sigma}^2 V(\hat{\theta}, F_{0,\hat{\alpha}})$ is defined by (5)-(7).

Tuning constants. If we let $\check{\mu}$ denote the ML estimate of μ, then the *asymptotic relative efficiency* (ARE) of $\hat{\mu}$ with respect to $\check{\mu}$ at the distribution F is

$$\mathrm{ARE}(\hat{\mu}, \check{\mu}, F) = V(\check{\mu}, F)/V(\hat{\mu}, F). \tag{23}$$

ARE is usually evaluated at the model F_θ. In the Lognormal case, $\mathrm{ARE}(\hat{\mu}, \check{\mu}, F_\theta)$ depends on σ but not on λ; in the Weibull and Gamma cases it depends on α but not on τ. Moreover, we define the *maximum asymptotic variance* (MAV) of $\hat{\mu}$ over the ϵ-contamination neighborhood (1), for a given mean μ^*, as

$$\mathcal{V}(\hat{\mu}, \mu^*, \epsilon) = \sup_{G \in \mathcal{P}_\epsilon} V(\hat{\mu}, G). \tag{24}$$

The most common rule for determining the tuning constant of an M-estimator is to require that the ARE with respect to the ML estimator, at the model, equals a given value, e.g., 95%; the higher the value of ARE, the more the estimate is sensitive to outliers. There are, however, several pairs of values of (b_1, b_2) with the same ARE, and the rule does not determine $\underline{b}$ uniquely. Marazzi and Ruffieux (1996) choose to minimize an approximation of the MAV as a function of $\underline{b}$ and give a few optimal values $\underline{b}(\epsilon)$, for varying ϵ and α, in the Gamma case. The sensitivity of the estimate to contamination depends then on ϵ and on the optimal value $\underline{b}(\epsilon)$ of $\underline{b}$. The choice of ϵ is made on the grounds of collateral information about the frequency of outliers. It turns out, however, that minimization of

MAV under the constraint $b_1 = b_2$ gives almost the same minimum as the unconstrained minimization. This suggests restricting attention to simplified estimators whith a single tuning constant $b = b_1 = b_2$. Table 3 (or available programs) allows one to choose b such that ARE has a desired value, knowing a rough preliminary estimate of α (Gamma and Weibull cases) or σ (Lognormal case).

REMARK 1. The most popular M-estimators of λ for the Gaussian model (9) are Huber's proposal 2 (Hp2) with a single tuning constant b (Hampel et al. 1986, p.234) and Huber's location estimator combined with the rescaled median absolute deviation (MAD) (Hampel et al. 1986, p.236). The latter makes sense when σ is a nuisance parameter (because MAD is very robust but rather inefficient) but is not appropriate for estimating the Lognormal mean μ, which depends on both λ and σ. When used to estimate μ, Hp2 and B_s^p have very similar behaviours (see, for example, Figure 3). The former, however, is computationally simpler and has a slightly higher breakdown point – that can be computed using the results of Huber 1980, p. 142-143 – for the same ARE (Table 3).

REMARK 2. In the Lognormal case, a small bias in σ^2 or λ can easily become a large one in μ. If this is a source of concern, a simple modification of the MM-estimator described in Yohai (1987, p. 648) can be helpful. The modification consists in the use of Q_n (Rousseeuw and Croux, 1993) as a final estimate of σ (Q_n being defined as 2.219 times the first quartile of the absolute differences $|\ln y_i - \ln y_j|$, $i < j$). The MM-estimator $\tilde{\lambda}$ has breakdown point (Bdp) 50% and, by conveniently setting a tuning constant k_1, it can be made highly efficient at the Gaussian model (e.g., ARE=0.95 for $k_1 = 4.687$). The Bdp of Q_n is 50% and its efficiency 0.82. The resulting estimate $\tilde{\mu} = \exp(\tilde{\lambda} + Q_n^2/2)$ has Bdp 50%; its ARE is high over a wide range of σ (e.g., 0.95 for $\sigma = 0.1$, 0.90 for $\sigma = 1$, 0.86 for $\sigma = 2$); finally, $\tilde{\mu}$ has a lower bias than both Hp2 and B_s^p (Figure 3). It requires, however, more computations.

REMARK 3. Although the problem of analytically computing the Bdp of the B_s^p-estimators for the Gamma and the Weibull models is still unsolved, some exploratory numerical computations suggest that the contamination bias of these estimators is reasonably low for practical problems.

REMARK 4. By parametrizing the Gamma distribution with α and $\nu = \ln(\alpha\sigma)$, the parameter vector splits up in a natural way into a main part $\nu = \ln(\mu)$ and a nuisance part α. Therefore, it would seem natural to use the (ν, α) parametrization in (16)-(19) (Hampel et al., 1986, Section 4.4). However, since $A^{\underline{b}}(\theta)$ is a multiplicative factor of $s(y, \theta)$, the estimates of ϑ would not change.

4 Confidence intervals and tests

Confidence intervals and tests for the one- and the two-sample problems can be computed using well known asymptotic results concerning M-estimates or according to the bootstrap rules described in Efron and Tibshirani (1993). The following descriptions are limited to the points that characterize the application of these rules to a class of models with two parameters θ_1, θ_2 and mean μ such that $\mu = \kappa_1(\theta_1)\kappa_2(\theta_2)$ (condition (15)).

In the one-sample problem, let $y = (y_1, \ldots, y_n)$ be a sample from an unknown distribution F. In the two-sample problem, let $y = (y_1, \ldots, y_n)$ and $z = (z_1, \ldots, z_m)$ be samples from two unknown distributions F and G, respectively. We will define parametric and non-parametric estimates $\hat{F}$ and $\hat{G}$ of F and G. For the definition of the parametric estimates we will consider the following families of distributions:

$$\mathcal{O}_u = \{F_\theta \mid \theta \in \Theta\},$$
$$\mathcal{O}_c = \{F_\theta \mid \theta \in \Theta,\ \hat{\mu}(F_\theta) = \text{a given value } \mu_0\},$$
$$\mathcal{T}_u = \{(F_{\theta^y}, F_{\theta^z}) \mid \theta^y \in \Theta,\ \theta^z \in \Theta\},$$
$$\mathcal{T}_c = \{(F_{\theta^y}, F_{\theta^z}) \mid \theta^y \in \Theta,\ \theta^z \in \Theta,\ \hat{\mu}(F_{\theta^y}) = \hat{\mu}(F_{\theta^z})\},$$

where F_θ is defined by (9), (11), (13), $\theta = (\theta_1, \theta_2)^{\mathrm{T}}$, $\theta^y = (\theta_1^y, \theta_2^y)^{\mathrm{T}}$, $\theta^z = (\theta_1^z, \theta_2^z)^{\mathrm{T}}$, Θ is the parameter space, and the functional $\hat{\mu}(F)$ is defined in the preceding sections. We will denote by $K_{n,y}$ the e.d.f of a sample y of size n drawn from a distribution K, with $K = \hat{F}$ or $K = F$ or $K = \hat{G}$ or $K = G$.

The one-sample problem: Confidence intervals for the mean. For the parametric bootstrap approach we set $\hat{F} = F_{\hat{\theta}} \in \mathcal{O}_u$; for the non-parametric bootstrap approach we set $\hat{F} = F_{n,y}$. According to Huber (1981, p. 6-7), we define the parameter to be estimated in terms of the limiting value $\hat{\mu}(F)$ of the estimate $\hat{\mu}(\hat{F})$. In other words, the robust approach replaces the expected mean μ by the "majority mean" $\hat{\mu}(F)$. A bootstrap confidence interval for $\hat{\mu}(F)$ can be computed by means of simulated values $\hat{\mu}(\hat{F}_{n,y^*})$, where y^* is drawn from $\hat{F}$.

The one-sample problem: A test for the mean. In order to test the hypothesis that the majority mean equals a given value μ_0, one can check whether the confidence interval for $\hat{\mu}(F)$ includes μ_0, but a direct method is also available. This method requires an estimate $\hat{F}$ of F under the hypothesis. Within the parametric framework, we set $\hat{F} = F_{\hat{\theta}} \in \mathcal{O}_c$, where $\hat{\theta}$ is the solution of (16)-(19) under the constraint $\hat{\mu}(F_\theta) = \mu_0$. Thanks to (15), equations (16)-(19) reduce to

$$\sum_{i=1}^{n} h_b[s_2(y, (\zeta(\theta_2, \mu_0), \theta_2)) - c_b(\theta_2)] = 0,$$

where $c_b(\theta)$ is defined by

$$\int h_b[s_2(y, (\zeta(\theta_2, \mu_0), \theta_2)) - c_b(\theta_2)] f_\theta(x) dx = 0,$$

and b is a user-defined tuning constant. For the non-parametric approach we set $\hat{F} = F_{n,\tilde{y}}$, where $\tilde{y} = (\mu_0/\hat{\mu})y$. Note that $\hat{\mu}(\hat{F}) = \mu_0$, for both the parametric and the non-parametric approaches, and that $\hat{\mu}(F)$ replaces μ. Thus, we test the hypothesis $H : \hat{\mu}(F) = \mu_0$, in place of $\mu = \mu_0$, and we use the test statistic $\hat{\mu}(\hat{F}) - \mu_0$. The bootstrap null distribution is based on simulated values $\hat{\mu}(\hat{F}_{n,y^*}) - \mu_0$, where y^* is drawn from $\hat{F}$.

The two-sample problem: Confidence intervals for the mean difference. We set $(\hat{F}, \hat{G}) = (F_{\hat{\theta}^y}, G_{\hat{\theta}^z}) \in \mathcal{T}_u$ (parametric bootstrap) or $(\hat{F}, \hat{G}) = (F_{n,y}, G_{m,z})$ (non-parametric bootstrap). The estimates $\hat{\theta}^y$ and $\hat{\theta}^z$ are independently obtained using y and z and the methods of Section 3. A confidence interval for $\hat{\mu}(G) - \hat{\mu}(F)$ is based on simulated values $\hat{\mu}(\hat{G}_{m,z^*}) - \hat{\mu}(\hat{F}_{n,y^*})$, where y^* and z^* are drawn from $\hat{F}$ and $\hat{G}$, respectively.

The two-sample problem: Testing equality of means. We test $H: \hat{\mu}(F) = \hat{\mu}(G)$ using a test statistic of the form $h(\hat{\mu}(\hat{G})/\hat{\mu}(\hat{F}))$, where h is a monotone function, e.g., $h(\cdot) = \ln(\cdot)$. We do not assume identical shapes or scales. Within the parametric framework, we need an estimate $(\hat{F}, \hat{G}) \in \mathcal{T}_c$, where the constrained parameter space is three-dimensional. Unfortunately, M-estimates of $(\theta_2^y, \theta_2^z, \mu)$ are computationally very heavy and programs are not available. As a remedy, we propose to use the unconstrained estimates $\hat{\theta}_2^y$ and $\hat{\theta}_2^z$ obtained for $\mathcal{T}_u$, set $\tilde{\theta}_1^y = \zeta(\hat{\theta}_2^y, \tilde{\mu})$, $\tilde{\theta}_1^z = \zeta(\hat{\theta}_2^z, \tilde{\mu})$, $\hat{G} = F_{\tilde{\theta}_1^z, \hat{\theta}_2^z}$, $\hat{F} = F_{\tilde{\theta}_1^y, \hat{\theta}_2^y}$, where $\tilde{\mu}$ is an arbitrary value, and note that $\hat{\mu}(\hat{G})/\hat{\mu}(\hat{F})$ does not depend on $\tilde{\mu}$, With the non-parametric approach, we set $(\hat{F}, \hat{G}) = (F_{n,\tilde{y}}, G_{m,\tilde{z}})$, where $\tilde{y} = (\tilde{\mu}/\hat{\mu}_y)y$ and $\tilde{z} = (\tilde{\mu}/\hat{\mu}_z)z$, and $\tilde{\mu}$ is arbitrary. The bootstrap null distribution is based on simulated values $h(\hat{\mu}(\hat{G}_{m,z^*})/\hat{\mu}(\hat{F}_{n,y^*}))$, where y^* and z^* are drawn from $\hat{F}$ and $\hat{G}$, respectively.

REMARK 1. It is usually convenient to standardize the test statistic, so that its null distribution is approximately standard normal. For example, for the two-sample test we can use

$$t = \frac{\ln(\hat{\mu}(\hat{G})/\hat{\mu}(\hat{F}))}{\sqrt{w}},$$

where

$$w = (1/n)V(\hat{\mu}, \hat{F})\hat{\mu}(\hat{F})^{-2} + (1/m)V(\hat{\mu}, \hat{G})\hat{\mu}(\hat{G})^{-2}$$

is an estimate of the variance of $\ln(\hat{\mu}(\hat{G})/\hat{\mu}(\hat{F}))$.

REMARK 2. The two sample problem test can be easily extended to the multisample problem. For $j = 1, \ldots, k$, let $x_j = (x_{j,1}, \ldots, x_{j,n_j})$ be a sample from an unknown distribution F_j and let $\hat{\mu}_j = \hat{\mu}(F_{n_j,x_j})$. We define parametric and non-parametric estimates $\hat{F}_j$ $(j = 1, \ldots, k)$ of F_j. For the definition of the parametric estimates we consider

$$\mathcal{M}_c = \{(F_{\theta^1}, \ldots, F_{\theta^k}) \mid \theta^1 \in \Theta, \ldots, \theta^k \in \Theta,\ \hat{\mu}(F_{\theta^1}) = \ldots = \hat{\mu}(F_{\theta^k})\}.$$

We test $H:\ \hat{\mu}(F_1) = \ldots = \hat{\mu}(F_k)$ using a test statistic based on functionals of the form

$$S(F_j) = h\left(\frac{\hat{\mu}(F_j)}{\sum_{j=1}^k d_j \hat{\mu}(F_j)}\right), \qquad j = 1, \ldots, k,$$

where $d_1, \ldots, d_k$ are fixed coefficients (for example, $d_j = 1/k$ for all j, or $d_j = 1$ for one particular $j = j'$ and $d_j = 0$ for $j \neq j'$) and $h(\cdot)$ is a monotone function, e.g., $h(\cdot) = \ln(\cdot)$.

It is convenient to use a standardized test statistic, so that its null distribution is approximately a χ^2 distribution with $k-1$ degrees of freedom, for example,

$$Z(\hat{F}_1, \ldots, \hat{F}_k) = \sum_{j=1}^{k} \left(\frac{\ln(\hat{\mu}(\hat{F}_j)/\hat{\mu}(\hat{F}_{j'}))}{\sqrt{w_j}}\right)^2,$$

where j' is a particular value of j and

$$w_j = (1/n_j)V(\hat{\mu}, \hat{F}_j)\hat{\mu}(\hat{F}_j)^{-2} + (1/n_j')V(\hat{\mu}, \hat{F}_{j'})\hat{\mu}(\hat{F}_{j'})^{-2}.$$

Within the parametric framework, $(\hat{F}_1, \ldots, \hat{F}_k) \in \mathcal{M}_c$. Unfortunately, the constrained M-estimates are computationally very heavy. As a remedy, we use the unconstrained estimates $\hat{\theta}_2^1, \ldots, \hat{\theta}_2^k$, set a common mean estimate $\tilde{\mu}$ to an arbitrary value, compute $\tilde{\theta}_1^1, \ldots, \tilde{\theta}_1^k$ from these estimates, and set $\hat{F}_j = F_{\tilde{\theta}_1^j, \hat{\theta}_2^j}$. With the non-parametric approach, we set $(\hat{F}_1, \ldots, \hat{F}_k) = (F_{n_1, \tilde{x}_1}, \ldots, F_{n_k, \tilde{x}_k})$, where $\tilde{x}_j = (\tilde{\mu}/\hat{\mu}_j)x_j$, $(j = 1, \ldots, k)$ and $\tilde{\mu}$ is arbitrary. The bootstrap null distribution is based on simulated values $Z(\hat{F}_{n_1, x_1^*}, \ldots, \hat{F}_{n_k, x_k^*})$, where $x_1^*, \ldots, x_k^*$ are drawn from $\hat{F}_1, \ldots, \hat{F}_k$. respectively.

REMARK 3. The two-sample procedure described above contains two arbitrary steps. First, the unconstrained estimates $\hat{\theta}_2^z$ and $\hat{\theta}_2^y$ have been used to avoid programming and tuning the constrained estimates. Second, $\hat{\theta}_1^z$ and $\hat{\theta}_1^y$ could have been used in place of $\hat{\theta}_2^z$ and $\hat{\theta}_2^y$. These choices deserve further investigation.

5 Examples

After specifying the parameter values used for the introductory example of Section 1, we consider three examples with simulated data of the two sample test described in Section 4.

Introductory example. Suspecting heavy contamination, all models were initially fitted using a relatively low value of a single tuning constant: $b = 1.3$. Indicating the Belgian and Swiss samples by y and z, respectively, we obtained: $(\hat{\alpha}^y, \hat{\mu}^y) = (1.09, 5.63)$, $(\hat{\alpha}^z, \hat{\mu}^z) = (2.39, 4.71)$ (Gamma); $(\hat{\alpha}^y, \hat{\mu}^y) = (1.07, 5.54)$, $(\hat{\alpha}^z, \hat{\mu}^z) = (1.71, 4.57)$ (Weibull); $(\hat{\sigma}^y, \hat{\mu}^y) = (1.08, 7.01)$, $(\hat{\sigma}^z, \hat{\mu}^z) = (0.71, 5.26)$ (Lognormal, Hp2). Tuning constants were then adjusted, so that ARE $\approx$ 0.85. We obtained: $(b^y, b^z) = (1.72, 1.44)$ (Gamma); $(b^y, b^z) = (1.79, 1.48)$ (Weibull); $(b^y, b^z) = (1.46, 1.26)$ (Lognormal, Hp2). The tuning constant of the MM-estimate described in Section 3, Remark 2, was set to 3.58 for y and 3.50 for z. Bootstrap tests were computed using 1000 simulated samples y^* and z^*.

Examples with simulated data. In the first example, we consider two samples of size 50 (data set A_0, Appendix). The first 47 y-values are drawn from a Gamma distribution with $\sigma = 1$ and $\alpha = 5$, and the last 3 values from a contaminating uniform distribution on $[20, 70]$. The 50 z-values are Gamma distributed with $\sigma = 5$ and $\alpha = 1$. Thus, the means of the uncontaminated populations are both equal to 5. Table 4 gives the one-sided ASL's (ASL $= P(t \leq t_{obs})$) of three tests that are often used in similar situations: the usual pooled t-test, the t-test on $\log(y)$ and $\log(z)$, and the Wilcoxon rank sum test. None is appropriate for comparing the means and the corresponding ASL's are very small, meaning rejection of the hypothesis of identical means at the 1% level. The fifth column refers to the robust test described in Section 4 based on the Gamma model ($b = 1.3$), which accepts the hypothesis. The same conclusion is achieved if the robust test is (erroneously) based on the Weibull model ($b = 1.3$) as shown by the ASL in the last column. By removing the outliers (data set A_1) the usual t-test accepts the hypothesis of identical means, whereas the ASL's of the other tests remain substantially unchanged.

In the second example, we consider two more samples of size 50 (data set B_0, Appendix). The first 47 y-values are Gamma distributed with $\sigma = 1$ and $\alpha = 5$, and the last 3 are outliers from uniform distribution on $[20, 70]$. The 50 z-values are Gamma distributed with $\sigma = 7.5$ and $\alpha = 1$. Thus, the means of the uncontaminated distributions are 5 and

7.5. The ASL's reported in Table 4 show that only the robust tests (based on the Gamma or the Weibull model) detect the mean difference.

In these examples, the bootstrap null distributions of the two sample tests – based on 1000 simulated values t^* of t – are approximately normal. The observed value of the test statistic based on the Gamma model is $t_{obs} = -0.79$ for data set A_0. We have ASL $= \#\{t^* \leq t_{obs}\}/1000 = 0.254$ and ASL $= P(t \leq t_{obs}) = 0.215$ with the normal approximation. With the data set B_0, we obtain $t_{obs} = 1.98$, ASL $= 0.974$ (simulated) and ASL $= 0.976$ (normal approximation).

Power functions of the tests based on the M-estimate with $b = 1.3$ and on the ML estimate for the Gamma model were computed as follows:

(a) An approximation of the null distribution of t was obtained by means of 1000 samples (y, z) of size $n = 50$ and $m = 50$ such that $y \sim F_{\ln(1),5}$, and $z \sim F_{\ln(5),1}$. Note that $E(Y) = E(Z) = 5$. Critical values t_{1-p} (the $(1-p)$th percentile of the null distribution) were computed for $p = 0.05$ and $p = 0.10$. We obtained $t_{0.90} = 1.23$, $t_{0.95} = 1.57$ for the M-estimate, and $t_{0.90} = 1.14$, $t_{0.95} = 1.48$ for the ML estimate.

(b) 1000 values of t based on samples (y, z) of size $n = 50$ and $m = 50$ such that $y \sim (1-\epsilon)F_{\ln(1),5} + \epsilon U[0, 50]$ and $z \sim F_{\ln(\sigma),1}$ ($U[0, 50]$ is the uniform distribution on $[0, 50]$) were obtained and the power approximation $q(\sigma) = \#\{t^* \geq t_{1-p}\}/1000$ was computed for $\sigma = 5, 6, 7, 8, 9$, $\epsilon = 0.00$ and 0.06. The results are summarized in Table 5 and Figure 4 and point out the power stability of the robust test.

6 Summary and conclusions

The population mean (or total) is often the target parameter an asymmetric distribution. Unfortunately, the arithmetic mean (its natural estimate) is very nonrobust. Moreover, unlike the symmetric case, there is no simple nonparametric robust estimator, like the median or the trimmed mean. Symmetrizing transformations (e.g., log) not always work and inferences for simmetrized data cannot usually replace those for the original mean.

We can think at the population as a mixture of two parts: the majority, that can be described by a parametric model, and a minority of outliers called "contamination". M-estimates can then be used in order to adjust an asymmetric model – e.g, the Lognormal, the Gamma or the Weibull distribution – to the majority of the data; the mean of the adjusted model is a robust estimate of the majority mean. We have shown how the bootstrap approach can be used for deriving robust tests for the majority mean in the one- and two-sample problems. Simulations have shown that, under contaminated data, the robust tests allow one to reach the correct decisions, whereas the usual tests can fail.

The practical impact of an unreliable mean comparison can be serious; an example with real data suggested that decisions in hospital budgeting are unfounded if exceptional patient behaviours in routine data are not treated by robust methods.

The estimation procedures described in this paper have been included in the library ROBETH (Marazzi, 1993) and interfaced to S-PLUS (StatSci, 1993). They are made available under `http://www//hospvd.ch/public/instituts/iumsp/home.htm`.

Appendix

Data set $\mathbf{A}_0$

$$y = (6.40, 4.08, 8.12, 2.30, 2.09, 5.32, 3.43, 4.60, 5.43, 3.20, 4.99, 5.69, 4.46, 9.96, 4.67, 4.39, 2.88, 3.49, 2.23, 3.14, 6.15, 7.82, 4.95, 6.12, 8.04, 3.34, 4.20, 5.67, 8.52, 3.16, 3.72, 2.65, 4.75, 4.47, 4.74, 4.71, 8.80, 4.18, 5.79, 2.95, 11.76, 3.05, 4.59, 2.46, 6.69, 4.53, 7.20, 39.10, 65.88, 43.78),$$

$$z = (16.14, 5.76, 7.16, 6.52, 0.52, 3.92, 5.01, 3.92, 0.68, 12.60, 4.65, 0.91, 1.43, 1.52, 1.44, 1.55, 2.25, 0.22, 1.08, 0.08, 5.08, 8.92, 3.67, 4.65, 1.38, 5.05, 4.28, 3.66, 7.47, 14.65, 1.03, 1.65, 4.75, 15.06, 21.15, 4.05, 11.15, 2.57, 2.66, 2.62, 4.10, 9.61, 0.30, 6.96, 3.34, 0.63, 2.73, 3.26, 3.35, 6.51).$$

Data set $\mathbf{B}_0$

$$y = (2.97, 7.08, 2.92, 5.49, 4.41, 1.45, 3.47, 6.28, 6.69, 7.31, 4.07, 6.04, 4.35, 3.67, 3.43, 5.98, 10.67, 4.44, 6.65, 10.41, 6.41, 5.66, 4.36, 2.93, 6.38, 9.20, 6.97, 8.37, 3.81, 3.55, 5.42, 3.29, 5.18, 4.20, 6.02, 3.39, 4.37, 5.76, 3.53, 2.97, 5.52, 4.40, 6.21, 3.88, 10.14, 8.35, 1.38, 63.95, 32.74, 36.86),$$

$$z = (7.44, 13.79, 1.80, 0.58, 2.52, 3.76, 2.64, 14.82, 6.21, 5.13, 5.01, 1.23, 18.91, 0.90, 18.77, 9.44, 1.95, 3.53, 31.36, 3.16, 0.96, 5.17, 6.30, 3.55, 11.40, 8.84, 1.23, 2.18, 3.64, 7.87, 14.63, 10.80, 4.88, 19.66, 3.67, 18.72, 18.35, 1.76, 1.48, 1.41, 10.90, 9.23, 29.54, 4.80, 4.13, 17.04, 6.52, 5.07, 7.35, 1.78).$$

References

Efron B., Tibshirani R.J. (1993). *An introduction to the bootstrap.* Chapman & Hall, New York.

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

Huber P.J. (1980). *Robust statistics.* Wiley, New York.

Johnson N.L, Kotz S., Balakrishnan, N. (1994). Continuous Univariate Distributions. Volume 1, Second Edition. Wiley, New York.

Lave JR., Leinhardt S. (1976), The cost and length of a hospital stay, *Inquiry,* 13, pp. 327-343.

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

Marazzi A., Ruffieux C. (1996). Implementing M-estimators of the Gamma distribution. In: H. Rieder (Ed.), *Robust statistics, data analysis, and computer intensive methods, In honor of Peter Huber's 60th birthday,* Lecture Notes in Statistics, 109, Springer Verlag, Heidelberg.

Rieder, H. (1994). Robust Asymptotic Statistics. Springer-Verlag, New York.

Rousseeuw P.J., Croux C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association,* 88, pp. 1273-1283.

StatSci (1993). S-PLUS for Windows User's Manual, Statistical Sciences, Inc., Seattle.

Victoria-Feser M.P., Ronchetti E. (1994). Robust methods for personal-income distribution models. *The Canadian Journal of Statistics,* 22, pp. 247-258.

Yohai V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics,* 15, pp. 642-656.

Table 1. Frequency distribution (Freq.) of lengths of stay (LOS) in days of patients hospitalized in Belgium and Switzerland during 1988 for certain "disorders of the nervous system".

Belgium

LOS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	21
Freq.	51	59	34	32	32	9	6	12	9	11	11	8	4	3	0	4	6	2	1
LOS	22	26	28	29	32	33	34	35	36	37	40	43	44	49	60	68	81	96	134
Freq.	1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1

Switzerland

LOS	1	2	3	4	5	6	7	8	9	16	115	198	374
Freq.	2	6	5	5	4	2	2	1	1	1	1	1	1

Table 2. LOS means of data in Table 1 (BE=Belgium, CH=Switzerland) and one-sided attained significance levels (ASL) of two sample tests based on three models. The robust estimates are described in Sections 3 and 5. The tests are described in Section 4.

	Full data set			LOS > 50 removed		
Mean	BE	CH	ASL	BE	CH	ASL
Arithmetic	7.87	25.47	1.00	6.58	4.41	0.08
Lognormal, Hp2	6.86	5.25	0.07	6.51	4.46	0.01
Lognormal, MM	6.57	5.07	0.09	6.44	4.43	0.01
Gamma B_s^p	5.89	4.81	0.09	5.66	4.22	0.04
Weibull B_s^p	5.85	4.67	0.06	5.60	4.11	0.01

Table 3. Tuning constant b such that certain M-estimates (B_s^p and Hp2 with $b = b_1 = b_2$, see Section 3) of the Lognormal, Gamma, and Weibull mean have a given asymptotic relative efficiency (ARE) with respect to the maximum likelihood estimate. In the Lognormal case, ARE depends just on σ; in the Gamma and Weibull cases, ARE depends just on α. The breakdown point (Bdp) of Huber's proposal 2 (Hp2) and B_s^p-estimates for the Lognormal model are reported.

			Lognormal				Gamma	Weibull
			Hp2	Hp2	B_s^p	B_s^p	B_s^p	B_s^p
ARE	α	σ	b	Bdp	b	Bdp	b	b
0.85	1	2.0	1.69	0.23	2.15	0.21	1.76	1.83
	2	1.0	1.43	0.27	1.72	0.25	1.48	1.38
	3	0.8	1.32	0.29	1.58	0.26	1.39	1.26
	4	0.6	1.18	0.31	1.45	0.28	1.35	1.25
	5	0.4	1.01	0.34	1.35	0.29	1.33	1.22
	10	0.2	0.83	0.37	1.28	0.30	1.29	1.21
0.90	1	2.0	1.91	0.20	2.54	0.18	2.12	2.18
	2	1.0	1.59	0.24	2.02	0.22	1.76	1.67
	3	0.8	1.54	0.25	1.84	0.24	1.62	1.39
	4	0.6	1.40	0.27	1.66	0.26	1.55	1.36
	5	0.4	1.23	0.30	1.51	0.27	1.51	1.34
	10	0.2	1.06	0.33	1.41	0.28	1.44	1.32
0.95	1	2.0	2.22	0.16	3.27	0.14	2.77	2.81
	2	1.0	1.98	0.19	2.61	0.18	2.26	2.23
	3	0.8	1.87	0.20	2.36	0.19	2.07	2.00
	4	0.6	1.74	0.22	2.07	0.21	1.95	1.92
	5	0.4	1.58	0.25	1.82	0.24	1.88	1.88
	10	0.2	1.42	0.27	1.65	0.26	1.73	1.86

Table 4. Attained significance levels of five tests on contaminated (data sets A_0 and B_0) and uncontaminated (data sets A_1 and B_1) asymmetric two sample problems.

Data set	t-test usual	t-test on logs	Wilcoxon test	Robust test Gamma	Robust test Weibull
A_0	0.04	0.001	0.01	0.25	0.24
A_1	0.31	0.006	0.03	0.34	0.33
B_0	0.56	0.339	0.50	0.97	0.99
B_1	0.98	0.404	0.81	0.99	0.99

Table 5. $p_{ML}(\sigma)$ and $p_{ME}(\sigma)$ represents the power as a function of σ of the two-sample test based on maximum likelihood (ML) and *M*-estimates (ME, with $b_1 = b_2 = 1.3$) of a Gamma mean for $y \sim (1-\epsilon)F_{\ln(1),5} + \epsilon U[0,50]$, $z \sim F_{\ln(\sigma),1}$. The sample sizes are $n = 50$ and $m = 50$. The power functions are drawn in Figure 5.

		$\varepsilon = 0$		$\varepsilon = 0.06$	
σ	p	$p_{ML}(\sigma)$	$p_{ME}(\sigma)$	$p_{ML}(\sigma)$	$p_{ME}(\sigma)$
5	0.10	0.13	0.10	0.02	0.09
	0.05	0.05	0.04	0.00	0.04
6	0.10	0.45	0.35	0.15	0.31
	0.05	0.28	0.23	0.11	0.20
7	0.10	0.85	0.74	0.39	0.67
	0.05	0.75	0.63	0.30	0.52
8	0.10	0.95	0.91	0.63	0.86
	0.05	0.92	0.83	0.55	0.81
9	0.10	1.00	0.98	0.80	0.96
	0.05	0.98	0.95	0.73	0.95

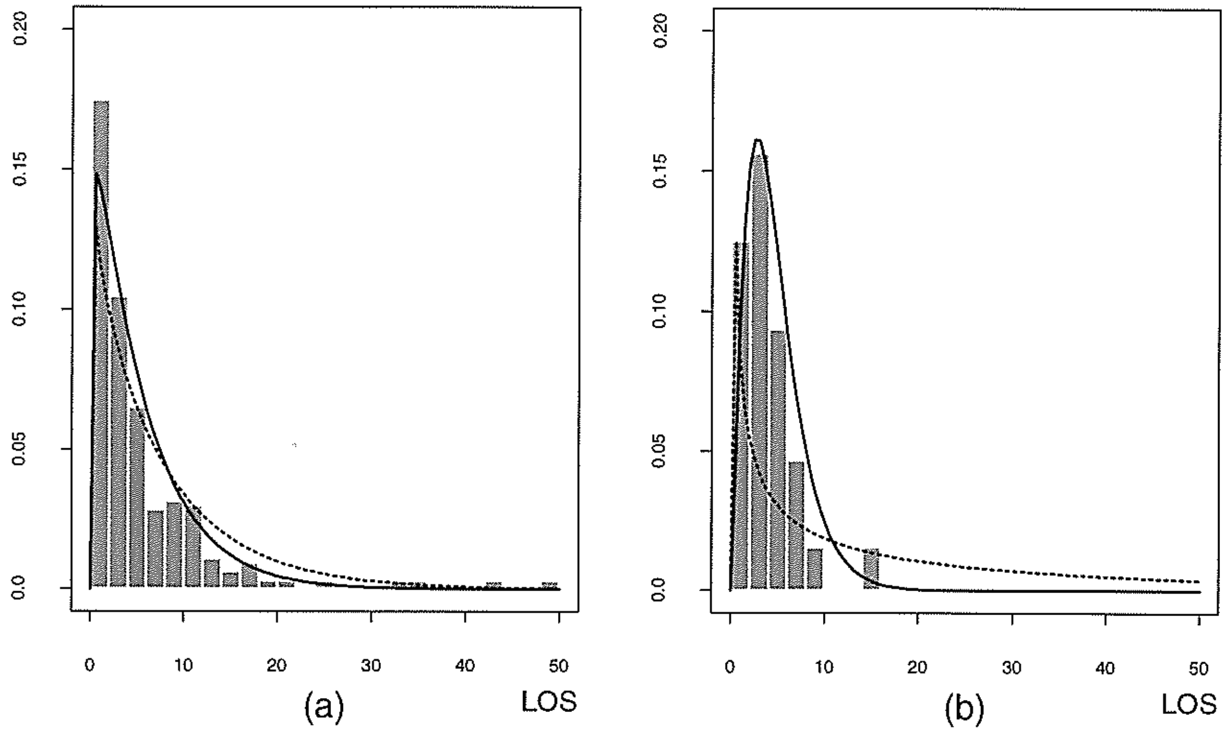

Figure 1. (a) Histogram of 315 lengths of stay (LOS) in days of patients hospitalized in Belgium during 1988 for certain "disorders of the nervous system". (b) Histogram of 32 LOS of patients hospitalized during the same year in Switzerland for the same kind of illness. Both histograms are truncated at LOS=50. The densities of the Gamma distributions have been determined by means of the maximum likelihood (dotted lines) and the robust estimates (solid lines).

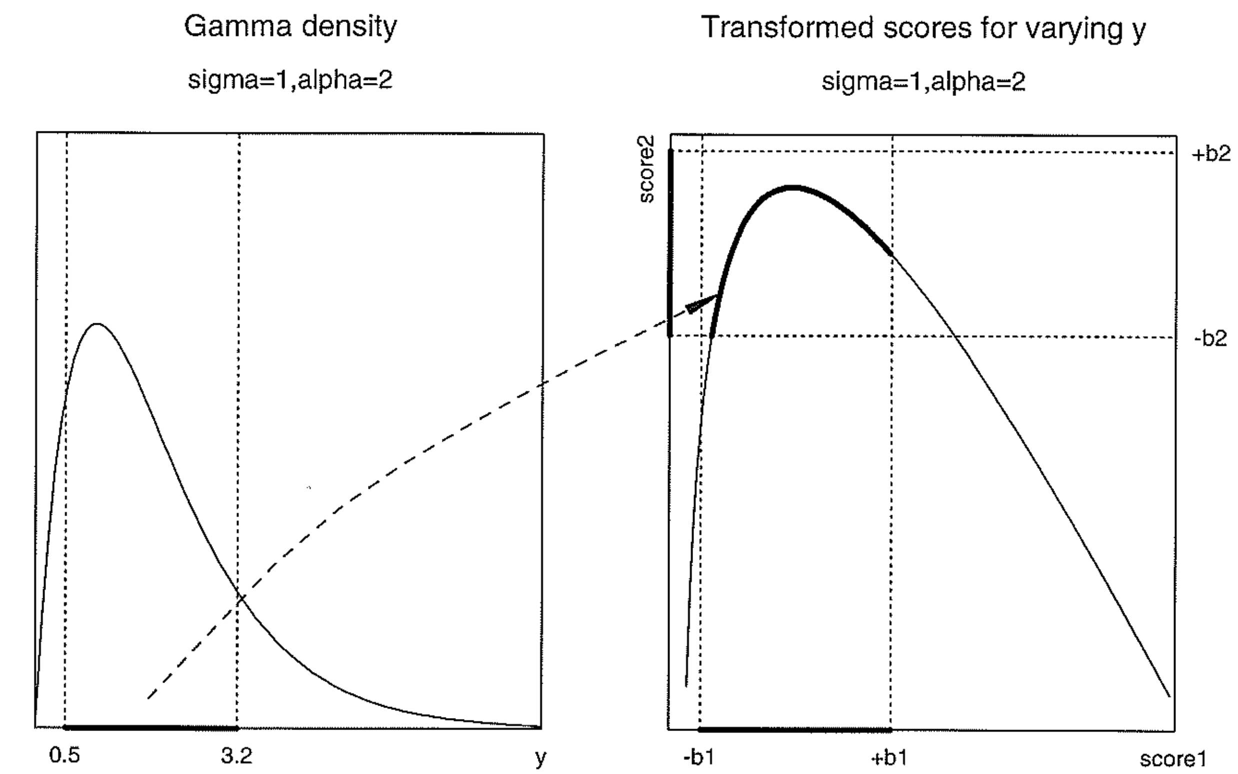

Figure 2. The left-hand diagram is the Gamma density with $\sigma = 1$ and $\alpha = 2$. The right-hand diagram represents the curve $(\tilde{s}_1(y;\theta), \tilde{s}_2(y;\theta))$, where $\tilde{s}_1$ and $\tilde{s}_2$ are the transformed likelihood scores of the Gamma distribution, for varying y and $\theta = (0,2)^{\mathrm{T}}$. The bold curve segment is not truncated in the estimation procedure and corresponds to the bold interval $(0.5, 3.2)$ of y-values in the density diagram.

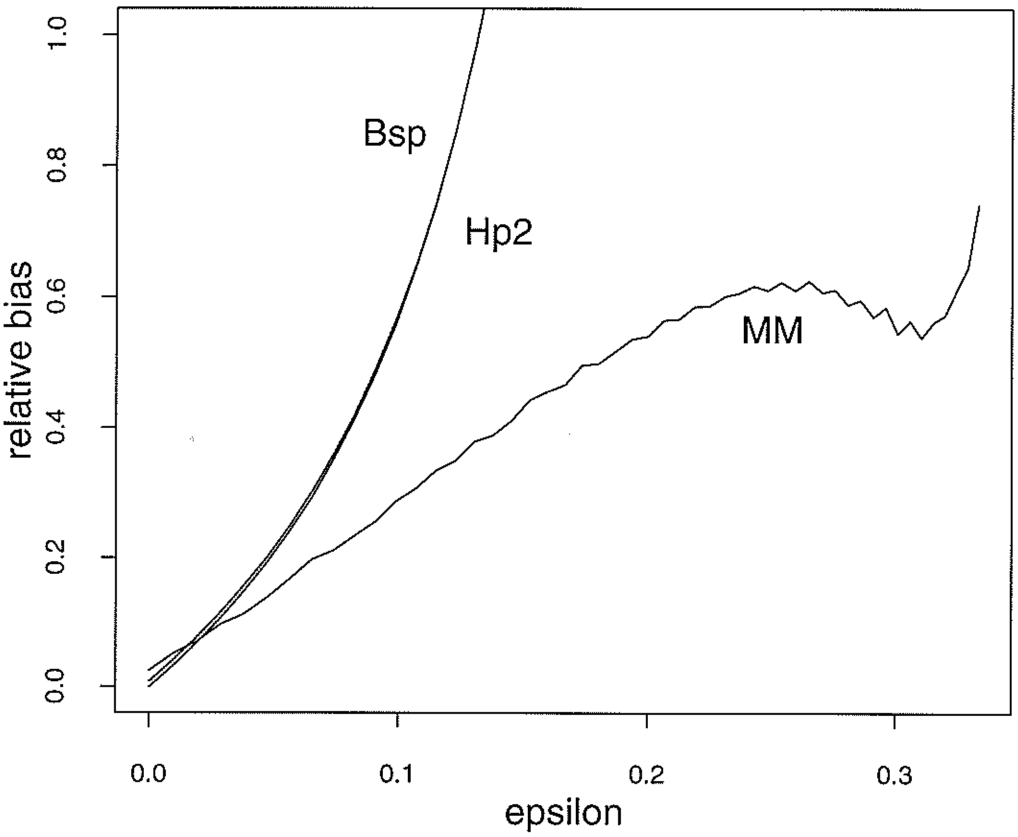

Figure 3. Relative bias curves of the Lognormal mean estimate $\mu = \exp(\lambda + \sigma^2/2)$ based on the following estimates of λ and μ: Hp2 (Huber's proposal 2, Section3, Remark 1, $b_1 = b_2 = 1.43$), B_s^p (Section 3, (17)-(20), $b_1 = b_2 = 1.72$), modified MM (Section 3, Remark 2, $k_1 = 3.56$). The relative bias for a given ϵ is computed as $[\hat{\mu}((1-\epsilon)G + \epsilon\Delta_{500}) - \exp(0.5)]/\exp(0.5)$, where G is the Lognormal c.d.f. with $\lambda = 0$, $\sigma = 1$.

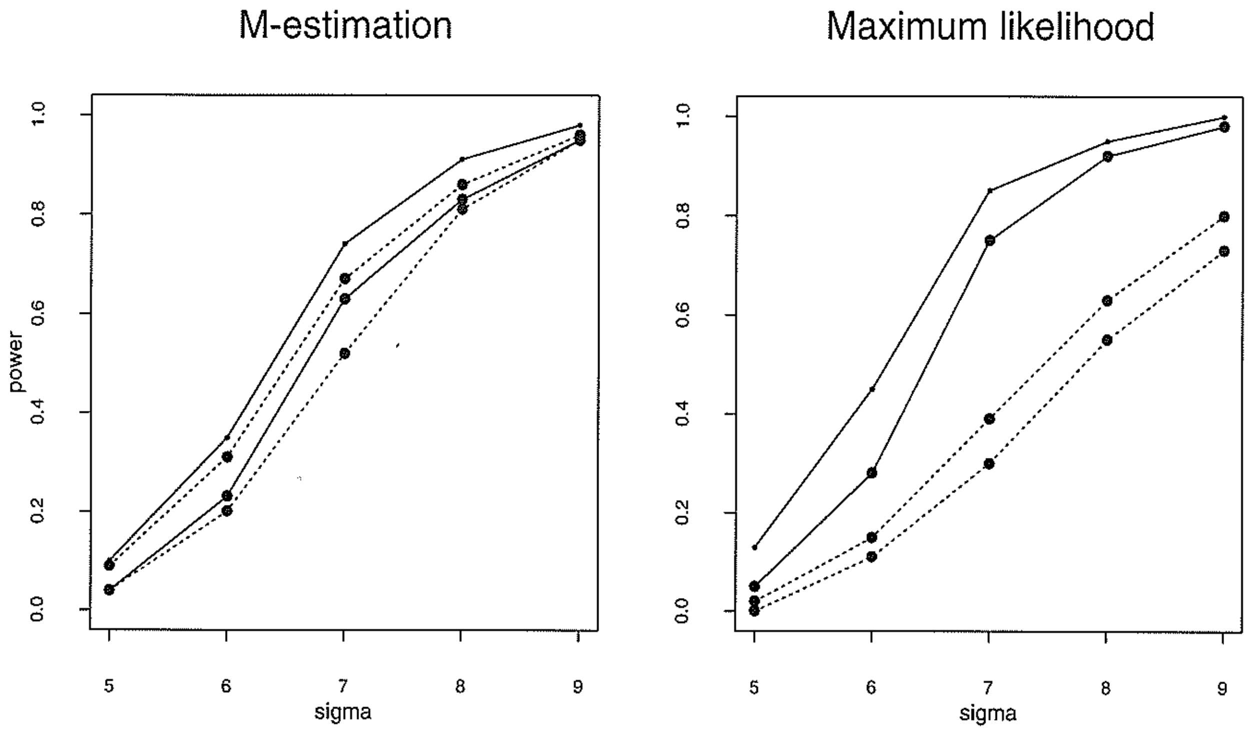

Figure =4. Approximate power functions for levels $p = 0.05$ (lower lines) and $p = 0.10$ (upper lines) of the two-sample test based on maximum Gamma-likelihood or M-estimation ($b_1 = b_2 = 1.3$) for $y \sim (1-\epsilon)F_{0,5} + \epsilon U[0,50]$, $z \sim F_{\ln(\sigma),1}$, $n = m = 50$. The solid lines have been obtained with $\epsilon = 0.06$, the dotted lines with $\epsilon = 0.00$. The values are reported in Table 4.

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.--)*

MARCO BERNASCONI, NICO H. BURKI, FELIX R. EHRAT, ANGELO DIGERONIMO, GUGLIELMO MAISTO, *Le società a tassazione speciale. Le società di partecipazioni, le società holding, le società di sede, le società ausiliarie e le società di servizio*, a cura di MARCO BERNASCONI, Centro di Studi Bancari, Meta-Edizioni, Vezia, Bellinzona, 1994, 144 pp. *Fr. 35.-- (banche ABT Fr. 24.--)*

RENÉ CHOPARD, AUGUSTO GRAZIANI, JEAN-MICHEL SERVET, ALDO DE MADDALENA, RENZO CARLI, ALESSANDRO CAVALLI, JÉRÔME BARATELLI, MICHEL DE RIVAZ, *Moneta e banconote. Per uno studio interdisciplinare su oggetto, idee e immagini*, a cura di RENÉ CHOPARD, Centro di Studi Bancari, Meta-Edizioni, Vezia, Bellinzona, 1994, 208 pp. *Fr. 47.-- (banche ABT Fr. 33.--)*

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

Nuove pubblicazioni del Centro di Studi Bancari

Les nouveaux défis au secret bancaire suisse

Enquêtes fiscales, commissions rogatoires, fusion d'entreprises

sous la direction de Paolo Bernasconi

Le fisc, le juge pénal et le juge civil, qu'ils soient suisses ou étrangers, ainsi que les héritiers, les créanciers et les actionnaires s'intéréssent aux données protégées par le secret bancaire suisse. Six avocats et professeurs d'Université ainsi que trois magistrats expliquent les lois nouvelles et la jurisprudence récente. Il en résulte un ouvrage de référence complet, utile pour résoudre les problèmes concrets qui se posent aux juristes, banquiers, magistrats et aux clients de banques suisses.

Les auteurs: Maurice Aubert; Pierre-André Béguin; Bruno Cocchi; Thomas A. Frick; Peter C. Honegger; Luca Marcellini; Xavier Oberson, Michele Rusca.

132 pp. Fr. 28.-- (banche ABT Fr. 20.--)

Sulla gestione patrimoniale

Strumenti, mercati, regole

a cura di René Chopard

Quasi 2'500 miliardi di Frs. è il patrimonio che si calcola sia amministrato dal sistema bancario svizzero ed è una banca elvetica ad essere al primo posto della graduatoria mondiale relativa all'importanza del patrimonio gestito. In questo contesto, il Ticino bancario si caratterizza per una sua ulteriore specializzazione nella gestione patrimoniale.

Non è dunque fortuito che il Centro di Studi Bancari abbia invitato ventinove specialisti (dirigenti bancari, operatori finanziari e accademici di chiara fama) a chinarsi sulle molteplici problematiche relative alla gestione patrimoniale.

Il contesto istituzionale delle attività bancarie (caratteristiche elvetiche, novità italiane ed evoluzione europea) e la struttura di produzione dei servizi finanziari (sistema bancario svizzero e ticinese) fanno da sfondo ad approfondite analisi sugli strumenti utilizzati nella gestione patrimoniale e sulle caratteristiche di questo sempre più importante mercato.

Gli autori: Paolo Clarotti; Claudio Generali; Mario Talamona; Felix Zumbach; Giorgi Ghiringhelli; Hans R. Spillmann; Marco Borghi; Alessandro Bizzozero; Marco Soldati e Massimo Pensotti; Giovanni Pirovano; José Antonio Blanco; Antonio Foglia; Aldo Visani; Giuseppe; Benelli; Angelo Porta; Gabriel Herrera; Michel Girardin; Roberto Provana; Alfredo Piacentini; Emilio Casagrande e Renato Santi; Paolo Bordi e Luca Sichel; Piergiorgio Bianchetti e Francesco Acerbi; Mauro Bini.

352 pp. Fr. 49.-- (banche ABT Fr. 34.--)

Novità legislative per le attività bancarie svizzere

Le Leggi federali sulle banche, sulle borse e sui fondi d'investimento

a cura di Paolo Bernasconi e René Chopard

Durante la seconda metà degli anni novanta la dinamica europea e l'evoluzione dei mercati finanziari hanno avuto un notevole impatto sulla legislazione bancaria svizzera che ha subito notevoli modifiche.

Oltre che riportare i testi delle nuove leggi, il volume propone delle analisi approfondite dei cambiamenti intervenuti grazie al contributo dei maggiori specialisti del campo.

Gli autori: Paolo Bernasconi; Alessandro Bizzozero; Matthäus Den Otter; Alain Hirsch; Pierre-Louis Manfrini; Henry Peter.

152 pp. Fr. 36.-- (banche ABT Fr. 24.--)

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.

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

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

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.

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

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.

No. 7 AURELIO MATTEI
Le previsioni congiunturali, maggio 1992, 36 pp.

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.

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

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.

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

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

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

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

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

No. 16 SAVERIO ALBERTI
Banca e crisi aziendale, febbraio 1995, 48 pp.

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

No. 18 ORLANDO NOSETTI
Il mercato del finanziamento degli enti pubblici locali ticinesi, aprile 1995, 50 pp.

No. 19 MATTIA L. RATTAGGI
Banking Safety and Regulation in the US: A Critical Realist Perspective, settembre 1995, 38 pp.

No. 20 ALFIO MARAZZI
Estimating and testing the means of asymmetric distributions using M-estimators, gennaio 1997, 20 pp.

No. 21 ALBERTO TOGNI
Il Rischio e la sua gestione, febbraio 1997, 15 pp.

No. 22 RENÉ CHOPARD, EDO POGLIA, RICCARDO SCARTAZZINI, VITO SVELTO, ANTONIO TAORMINA
Ricerca e formazione, pubblica e privata: costi, obiettivi e risultati, febbraio 1997, 47 pp.

No. speciale ALICE MORETTI
Villa Negroni, Cenni storici, settembre 1992, 33 pp.

I Quaderni possono essere comandati a Fr. 5.-- cadauno + spese postali a: Centro di Studi Bancari, Biblioteca, Villa Negroni, CH - 6943 Vezia-Lugano