DOI : 10.3150/08-BEJ140

Estimation of bivariate excess probabilities for elliptical models

Belkacem Abdous [a] Anne-Laure Fougères* [b] Kilani Ghoudi [c]
Philippe Soulier [b]

[a] *Département de médecine sociale et préventive, Université Laval, Québec, Canada G1K 7P4*

[b] *Université Paris X, Laboratoire MODAL'X, 92000 Nanterre, France*

[c] *Department of statistics, United Arab Emirates University, P.O Box 17555 Al-Ain, United Arab Emirates.*

Abstract

Let (X, Y) be a random vector whose conditional excess probability $\theta(x, y) := P(Y \le y \mid X > x)$ is of interest. Estimating this kind of probability is a delicate problem as soon as x tends to be large, since the conditioning event becomes an extreme set. Assume that (X, Y) is elliptically distributed, with a rapidly varying radial component. In this paper, three statistical procedures are proposed to estimate $\theta(x, y)$, for fixed x, y, with x large. They respectively make use of an approximation result of Abdous *et al.* (cf. Abdous et al. (2005, Theorem 1)), a new second-order refinement of Abdous *et al.*'s Theorem 1, and a non-approximating method. The estimation of the conditional quantile function $\theta(x, \cdot)^{\leftarrow}$ for large fixed x is also addressed, and these methods are compared via simulations. An illustration in the financial context is also given.

Key words: Asymptotic independence; Conditional excess probability; Financial contagion; Elliptic law; Rapidly varying tails.

1 Introduction

Consider two positively dependent market returns X and Y. It is of practical importance to assess the possible contagion between X and Y. Contagion formalizes the fact that for large values x, the probability $\mathrm{P}(Y > y \mid X > x)$ is greater than $\mathrm{P}(Y > y)$. See for example Abdous et al. (2005) or Bradley and Taqqu (2004, 2005), among others. Besides, this conditional probability is also related to the tail dependence coefficient which has been widely investigated in the financial risk management context, see for instance Frahm et al. (2005). Therefore, the behavior of the conditional excess probability $\theta(x, y) := P(Y \leq y \mid X > x)$ is of practical interest, especially for large values of x. Estimating this kind of probability is a delicate problem as soon as x tends to be large, since the conditioning event becomes an extreme set. Large here essentially means that x is beyond the largest value of the X observations, so that the conditional empirical distribution function then fails to be of any use, even if the probability $\theta(x, y)$ in itself is *not* a small probability, nor close to 1. Alternative methods have to be considered.

A classical approach is to call upon multivariate extreme value theory. Many refined inference procedures have been developed, making use of the structure of multivariate max-stable distributions introduced by de Haan and Resnick (1977), Pickands (1981), de Haan (1985). These procedures are successful in the rather general situation where (X, Y) are asymptotically dependent (for the maxima), which means heuristically that X and Y can be simultaneously large (see e.g. Resnick (1987) or Beirlant et al. (2004) for more details).

Efforts have recently been made to the problem in the opposite case of asymptotic independence. In some papers, attempts are made to provide models for joint tails, see e.g. Ledford and Tawn (1996, 1997), Draisma et al. (2004), Resnick (2002), Maulik and Resnick (2004). In a parallel way, Heffernan and Tawn (2004) explored modeling for multivariate tail regions which are not necessarily joint tails, and Heffernan and Resnick (2007) provided a complementary mathematical framework in the bivariate case.

In these papers, the key assumption is that there exists a limit for the conditional distribution of Y suitably centered and renormalized, given that X tends to infinity. This assumption was first checked for bivariate spherical distributions by Eddy and Gale (1981) and Berman (1992, Theorem 12.4.1). Abdous et al. (2005) obtained it for bivariate elliptical distributions; Hashorva (2006) for multivariate elliptical distributions, Balkema and Embrechts (2007) for generalized multivariate elliptical distributions, and Hashorva et al. (2007) for Dirichlet multivariate distributions.

Elliptical distributions form a large family of multivariate laws, which have received considerable attention, especially in the financial risk context. See Artzner et al. (1999), Embrechts et al. (2003), Hult and Lindskog (2002), among others. Assume from now on that (X, Y) is elliptically distributed; Theorem 1 of Abdous et al. (2005) exhibits the asymptotic behavior of $\theta(x, y)$ when $x \to \infty$ for such an elliptical pair. The appropriate rate $y = y(x)$ is explicited to get a non degenerate behavior of $\lim_{x\to\infty} \theta(x, y)$. This rate depends on the tail behavior of the radial random variable R defined by the relation $R^2 = (X^2 - 2\rho XY + Y^2)/(1 - \rho^2)$, where ρ is the Pearson correlation coefficient between X and Y. The only parameters involved in $y(x)$ and in the limiting distribution are: the Pearson correlation coefficient ρ, and the index of regular variation of R (say α) or an auxiliary function of R (say ψ), depending on wether R has a regularly or rapidly varying upper tail.

In financial applications, the regularly varying behavior of the upper tails is commonly encountered. Abdous et al. (2005) provided a simulation study in the specific case where R has a regularly varying tail. Existing estimators of ρ and α were used therein to obtain a practical way to estimate excess probabilities. However, this assumption fails to hold in some situations, as shown by Levy and Duchin (2004), who compared the fit of 11 distributions on a wide range of stock returns and investment horizons. They concluded that the logistic distribution, which has a rapidly varying tail, gives the best fit in most of the cases for weekly and monthly returns.

The aim of this paper is to focus on the case where the radial component R associated with the elliptical pair (X, Y) has a rapidly varying upper tail. A second-order approximation result is obtained, which refines Theorem 1 of Abdous *et al.* in the case of rapid variation of R. As we mentioned earlier, under the rapidly varying upper tail assumption, the obtained approximation involves an auxiliary function ψ which has to be estimated. To the best of our knowledge, estimating the auxiliary function has never been considered in the literature. We propose three statistical procedures for the estimation of $\theta(x, y)$, for fixed x, y, with x large. They respectively make use of Abdous *et al.*'s approximation result, its second-order refinement, or an alternative method which does not rely on an analytic approximation result.

The conditional quantiles – also called regression quantiles – of a response Y given a covariate X have received an important attention, see for example Koenker and Bassett (1978), or Koenker (2005). More specifically, extremal conditional quantiles have proven to be useful in economics and financial applications, see e.g. Chernozhukov (2005). In the same spirit, one might be interested by estimating the conditional quantile function $\theta(x, \cdot)^{\leftarrow}$ when x is fixed and large. This problem is also addressed in detail in the present paper, and the simulation study performed is presented in both terms of estimation of θ and $\theta(x, \cdot)^{\leftarrow}$.

The paper is organized as follows: The second-order approximation result is presented in Section 2, as well as some remarks and examples illustrating the theorem. The statistical procedures are described in Section 3, where a semi-parametric estimator of ψ is proposed. Section 4 deals with a comparative simulation study, while Section 5 provides an application to the financial context, revisiting data studied by Levy and Duchin (2004). Some concluding comments are given in Section 6. Proofs are relegated to the Appendix.

2 Asymptotic approximation

Consider a bivariate elliptical random vector (X, Y). General background on elliptical distributions can be found e.g. in Fang et al. (1990). One can focus without loss of theoretical generality on the standard case where $\mathbb{E}\,X = \mathbb{E}\,Y = 0$ and $\operatorname{Var} X = \operatorname{Var} Y = 1$. A convenient representation is then the following (see for example Hult and Lindskog (2002)): (X, Y) has a standard elliptical distribution with radial positive distribution function H and Pearson correlation coefficient ρ if it can be expressed as

$$(X, Y) = R\left(\cos U, \rho \cos U + \sqrt{1-\rho^2}\sin U\right) \ ,$$

where R and U are independent, R has distribution function H with $\mathbb{E}\,R^2 = 2$ and U is uniformly distributed on $[-\pi/2, 3\pi/2]$. Hereafter, to avoid trivialities we assume $|\rho| < 1$.

Let Φ denote the normal distribution function and φ its density, i.e.

$$\varphi(t) = \frac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}} \ , \quad \Phi(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t \ .$$

This paper deals with elliptical distributions with rapidly varying marginal upper tails, or equivalently with a rapidly varying radial component. More precisely, the radial component R associated with (X, Y) is assumed to be such that there exists an *auxiliary function* ψ for which one gets, for any positive t

$$\lim_{x\to\infty} \frac{P\{R > x + t\psi(x)\}}{P(R > x)} = e^{-t}. \tag{1}$$

Such a condition implies that R belongs to the max-domain of attraction of the Gumbel distribution, see Resnick (1987, p.26) for more details. de Haan (1970) introduced this class of distributions as the $\Gamma-$*varying class*. The function ψ is positive, absolutely continuous and satisfies $\lim_{t\to\infty} \psi'(t) = 0$, $\lim_{t\to\infty} \psi(t)/t = 0$ and $\lim_{t\to\infty} \psi\{t + x\psi(t)\}/\psi(t) = 1$ for each positive x. It is only unique up to asymptotic equivalence.

Let us recall Abdous et al. (2005)'s result in this rapidly varying context (see Theorem 1, (ii)):

Theorem 1 *Let (X, Y) be a bivariate standardized elliptical random variable with Pearson correlation coefficient ρ and radial component satisfying (1). Then for each $z \in \mathbb{R}$ one has*

$$\lim_{x\to\infty} P(Y \le \rho x + z\sqrt{1-\rho^2}\sqrt{x\psi(x)} \mid X > x) \ = \ \Phi(z).$$

In the following subsection, a rate of convergence is provided for the approximation result stated in Theorem 1, as well as a second order correction. Note that if (X, Y) has an elliptic distribution with correlation coefficient ρ, then the couple $(X, -Y)$ has an elliptic distribution with correlation coefficient $-\rho$. Therefore, one can focus on non-negative ρ. From now on, assume that $\rho \geq 0$. As a consequence, one can assume that both $x > 0$ and $y > 0$.

2.1 Main result

For each distribution function H, denote by $\bar{H}$ the survival function $\bar{H} = 1 - H$. The following assumption will be sufficient to obtain the main result. It is a strengthening of (1).

Hypothesis 2 *Let H be a rapidly varying distribution function such that*

$$\left| \frac{\bar{H}\{x + t\psi(x)\}}{\bar{H}(x)} - \mathrm{e}^{-t} \right| \leq \chi(x)\Theta(t) \; , \tag{2}$$

for all $t \geq 0$ and x large enough, where $\lim_{x\to\infty} \chi(x) = 0$, ψ satisfies

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 0 \; ; \tag{3}$$

and Θ is locally bounded and integrable over $[0, \infty)$.

The following result is a second-order approximation for conditional excess probabilities in the elliptical case with rapidly varying radial component.

Theorem 3 *Let (X, Y) be a bivariate elliptical vector with Pearson correlation coefficient $\rho \in [0, 1)$ and radial distribution H that satisfies Hypothesis 2. Then for all $x > 0$ and $z \in \mathbb{R}$,*

$$\mathbb{P}(Y \leq \rho x + z\sqrt{1-\rho^2}\sqrt{x\psi(x)} \mid X > x)$$
$$= \Phi(z) - \sqrt{\frac{\psi(x)}{x}} \frac{\rho\,\varphi(z)}{\sqrt{1-\rho^2}} + O\left(\chi(x) + \frac{\psi(x)}{x}\right) \; , \tag{4}$$

$$\mathbb{P}(Y \leq \rho x + z\sqrt{1-\rho^2}\sqrt{x\psi(x)} + \rho\psi(x) \mid X > x) \tag{5}$$
$$= \Phi(z) + O\left(\chi(x) + \frac{\psi(x)}{x}\right) \; . \tag{6}$$

All the terms $O(\;)$ are locally uniform with respect to z.

Remark 4 *This result provides a rate of convergence in the approximation result of Abdous et al. (2005, Theorem 1) and a second-order correction. This correction is useful only if $\chi(x) = o(\sqrt{\psi(x)/x})$.*

Remark 5 *Theorem 3 and the formula*

$$\mathbb{P}(X \leq x' \; ; \; Y \leq y \mid X > x) \\ = \mathbb{P}(Y \leq y \mid X > x) - \mathbb{P}(X > x' \mid X > x)\mathbb{P}(Y \leq y \mid X > x')$$

yield (with some extra calculations) the asymptotic joint distribution (and the rate of convergence) of (X, Y) *given that* $X > x$ *when* x *is large: for all* $x > 0$ *and* $z \in \mathbb{R}$,

$$\mathbb{P}(X \leq x + t\psi(x) \; ; \; Y \leq \rho x + z\sqrt{1-\rho^2}\sqrt{x\psi(x)} \mid X > x) \\ = (1 - \mathrm{e}^{-t})\Phi(z) + O\left(\chi(x) + \sqrt{\frac{\psi(x)}{x}}\right) , \quad (7)$$

where all the terms $O(\;)$ *are locally uniform with respect to* z.

Remark 6 *Hashorva (2006) obtained that the conditional limit distribution of* Y *given that* $X = x$ *for multivariate elliptical vectors with rapidly varying radial component is also the Gaussian distribution. The equality of these two asymptotic distributions is not true in general. In the elliptical context, this is only true if the radial variable is rapidly varying. The conditional distribution of* Y *given that* $X = x$ *is of course related to the joint distribution of* (X, Y) *given that* $X > x$ *via the formula*

$$\mathbb{P}(X \leq x' \; ; \; Y \leq y \mid X > x) = \int_x^{x'} \mathbb{P}(Y \leq y \mid X = u) P_X(du),$$

where P_X *is the distribution of* X. *However, the limiting behavior of the integrand is not sufficient to obtain the limit of the integral, so that* (7) *is not a straightforward consequence of Hashorva's result.*

2.2 Remarks and examples

Hypothesis 2 gives a rate of convergence in the conditional excess probability approximation. To the best of our knowledge, the literature deals more classically with second order conditions providing limits (see for instance Beirlant et al. (2004, Section 3.3)) or with pointwise or uniform rates of convergence (cf. e.g. de Haan and Stadtmüller (1996); Beirlant et al. (2003)). The need here is to have a non uniform bound that can be used for dominated convergence arguments. However, in the examples given below, the non uniform rates $\chi(x)$ that we exhibit are the same as the optimal uniform rates provided by de Haan and Stadtmüller (1996).

One can however check that Hypothesis 2 holds for usual rapidly varying functions, in particular for most of the so-called *Von Mises distribution functions*, which satisfy (see e.g. Resnick (1987, p.40))

$$\bar{H}(x) = d \exp\left\{ - \int_{x_0}^{x} \frac{\mathrm{d}s}{\psi_H(s)} \right\} , \qquad (8)$$

for x greater than some $x_0 \geq 0$, where $d > 0$ and ψ_H is positive, absolutely continuous and $\lim_{x\to\infty} \psi_H'(x) = 0$. Note that under this assumption, $\psi_H = \bar{H}/H'$, and ψ_H is an auxiliary function in the sense of (1). In the sequel, auxiliary functions ψ_H satisfying (8) will be called Von Mises auxiliary functions.

The following lemma provides sufficient conditions for Hypothesis 2, which could be weakened at the price of additional technicalities.

Lemma 7 *Let H be a Von Mises distribution function, and assume that,*

(i) ψ_H is ultimately monotone, differentiable and $|\psi_H'|$ is ultimately decreasing;
(ii) if ψ_H is decreasing, then either $\lim_{x\to\infty} \psi_H(x) > 0$ or there exist positive constants c_1 and c_2 such that for all $x \geq 0$ and $u \geq 0$,

$$\frac{\psi_H(x)}{\psi_H(x+u)} \leq c_1 \mathrm{e}^{c_2 u} . \qquad (9)$$

Then Hypothesis 2 holds for ψ_H:

$$\left| \frac{\bar{H}\{x + t\psi_H(x)\}}{\bar{H}(x)} - \mathrm{e}^{-t} \right| \leq \chi(x)\Theta(t) ,$$

with $\chi(x) = O(|\psi_H'(x)|)$ and $\Theta(t) = O((1+t)^{-\kappa})$ for an arbitrary $\kappa > 0$.

Remark 8 *Assumption (9) holds if ψ_H is regularly varying with index $\gamma < 0$. It also holds if $\psi_H(x) = \mathrm{e}^{-cx}$ for some $c > 0$.*

Remark 9 *If ψ_H is regularly varying with index $\gamma < 1$, and ψ_H' ultimately decreasing, then $\psi_H'(x) = o\left(\sqrt{\psi_H(x)/x}\right)$ and hence the second-order correction is useful.*

Remark 10 *This bound corresponds to a worst case scenario. In many particular cases, a much faster rate of convergence can be obtained. For instance, if $\bar{H}(x) = \mathrm{e}^{-t}$, then (2) holds with $\psi_H \equiv 1$, whence $\chi(x) = 1/x$, but for any positive x and t, $\bar{H}\{x + t\psi_H(x)\}/\bar{H}(x) = \mathrm{e}^{-t}$. The rate of convergence is infinite here. If $\bar{H}$ is the Gumbel distribution, then $\psi_H \equiv 1$ and the rate of convergence in (2) is exponential: $|\bar{H}(x+t)/\bar{H}(x) - \mathrm{e}^{-t}| \leq 2\mathrm{e}^{-x}\mathrm{e}^{-t}$.*

Some examples of continuous distributions satisfying Hypothesis 2 are given for illustration in Section 4, Table 1. The following example illustrates a discrete situation.

Example 11 *(Discrete distribution in the domain of attraction of the Gumbel law). Let ψ be a concave increasing function such that $\lim_{x\to\infty}\psi(x) = +\infty$ and $\lim_{x\to\infty}\psi'(x) = 0$ and define*

$$\bar{H}_{\#}(x) = \exp\left\{-\int_{x_0}^{x}\frac{\mathrm{d}s}{\psi(s)}\right\} , \quad \bar{H}(x) = \bar{H}_{\#}([x]) ,$$

where $[x]$ is the integer part of x. Then H is a discrete distribution function, $\bar{H}$ belongs to the domain of attraction of the Gumbel distribution, but does not satisfy Condition (8). Nevertheless, following the lines of the proof of Lemma 7, one can check that Condition (2) holds with

$$\chi(x) = O(\psi'([x]) + 1/\psi(x)) , \qquad \Theta(t) = O((1+t)^{-\kappa}) , \tag{10}$$

for any arbitrary $\kappa > 0$.

The proof of (10) is in the Appendix.

One can deduce from Lemma 7 that Theorem 3 holds for $\psi = \psi_H$. The following lemma says what happens if one uses an asymptotically equivalent auxiliary function ψ instead of ψ_H in Theorem 3.

Lemma 12 *Under the assumption of Lemma 7, let ψ be equivalent to ψ_H at infinity and define $\xi(x) = |\psi(x) - \psi_H(x)|/\psi_H(x)$. Then*

$$\left|\frac{\bar{H}(x+t\psi(x))}{\bar{H}(x)} - \mathrm{e}^{-t}\right| \le O\left\{|\psi_H'(x)| + \xi(x)\right\} \Theta(t) . \tag{11}$$

Remark 13 *A consequence of Lemma 12 is that if an auxiliary function ψ is used instead of ψ_H in Theorem 3, then the second-order correction is relevant provided that $\xi(x) = o\left(\sqrt{\psi_H(x)/x}\right)$. This is the case in the examples given on line 1 and line 3 of Table 1 ('Normal' and 'Logis') if one takes $\psi(x) = 1/x$ or $\psi(x) = 1/(2x)$ respectively; and in the example given on line 5 of Table 1 ('Lognor') when using $\psi(x) = x/\log(x)$.*

3 Statistical procedure

For *given* large positive x and y, consider the problems of estimating $\theta(x,y) = \mathbb{P}(Y \le y \mid X > x)$ and the conditional quantile function $\theta(x,\cdot)^{\leftarrow}$. Note in passing that in a practical situation, x is not a threshold nor a parameter of the statistical procedure, but a value imposed by the practical problem (e.g. a high quantile of the marginal distribution of X). The empirical distribution function is useless, since there might be no observations in the considered range. We suggest to estimate these quantities by means of Theorem 3.

Assume that a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is available, drawn from an elliptical distribution with radial component satisfying Hypothesis 2. Note that in this section we do not assume that the distribution is standardized, so that a preliminary standardization is required.

3.1 Definition of the estimators

The estimation of $\theta(x, y)$ and the conditional quantile function requires estimates of μ_X, μ_Y, σ_X, σ_Y, ρ and ψ. Let $\hat{\mu}_X$, $\hat{\mu}_Y$, $\hat{\sigma}_X$, $\hat{\sigma}_Y$, $\hat{\rho}_n$ and $\hat{\psi}_n$ denote such estimates. For fixed $x, y > 0$, define

$$\hat{\theta}_{n,1}(x, y) = \Phi\left(\frac{\hat{y} - \hat{\rho}_n \hat{x}}{\sqrt{1 - \hat{\rho}_n^2}\sqrt{\hat{x}\hat{\psi}_n(\hat{x})}}\right) , \tag{12}$$

$$\hat{\theta}_{n,2}(x, y) = \Phi\left(\frac{\hat{y} - \hat{\rho}_n \hat{x} - \hat{\rho}_n \hat{\psi}_n(\hat{x})}{\sqrt{1 - \hat{\rho}_n^2}\sqrt{\hat{x}\hat{\psi}_n(\hat{x})}}\right) , \tag{13}$$

where $\hat{x} = (x - \hat{\mu}_X)/\hat{\sigma}_X$ and $\hat{y} = (y - \hat{\mu}_Y)/\hat{\sigma}_Y$.

In order to estimate the conditional quantile function $\theta(x, \cdot)^{\leftarrow}$, define, for fixed $\theta \in (0, 1)$,

$$\hat{y}_{n,1} = \hat{\mu}_Y + \hat{\sigma}_Y \left\{\hat{\rho}_n \hat{x} + \sqrt{1 - \hat{\rho}_n^2}\sqrt{\hat{x}\hat{\psi}_n(\hat{x})}\Phi^{-1}(\theta)\right\} , \tag{14}$$

$$\hat{y}_{n,2} = \hat{\mu}_Y + \hat{\sigma}_Y \left\{\hat{\rho}_n \hat{x} + \hat{\rho}_n \hat{\psi}_n(\hat{x}) + \sqrt{1 - \hat{\rho}_n^2}\sqrt{\hat{x}\hat{\psi}_n(\hat{x})}\Phi^{-1}(\theta)\right\} . \tag{15}$$

Estimating μ_X, μ_Y, σ_X, σ_Y and ρ is a classical topic, and the empirical version of each quantity can easily be used. Under the assumption of elliptical distributions, however, one can observe a better stability when the Pearson correlation coefficient is estimated by $\hat{\rho}_n = \sin(\pi\hat{\tau}_n/2)$, where $\hat{\tau}_n$ is the empirical Kendall's tau. See e.g. Hult and Lindskog (2002) for more details. This estimator is $\sqrt{n}$-consistent and asymptotically normal.

Consider now the problem of the estimation of ψ. Since auxiliary functions are defined up to asymptotic equivalence, a particular representant must be a priori chosen in order to define the estimator. We assume that an admissible auxiliary function is

$$\psi(x) = \frac{1}{c\beta} x^{1-\beta} , \tag{16}$$

for some constants $c > 0$ and $\beta > 0$. Under this assumption, estimation of ψ boils down to estimating c and β.

A wide literature exists on estimators of β, see e.g. Beirlant et al. (1999), Gardes and Girard (2006), Dierckx et al. (2007), among others. The method chosen here is the one proposed in Beirlant et al. (1996). Let k_n be a user chosen threshold and $R_{j,n}$, $1 \le j \le n$ be the order statistics of the sample $R_1, \dots, R_n$. The estimator of β is obtained as the slope of the Weibull quantile plot at the point $(\log\log(n/k_n), \log(R_{n-k_n,n}))$:

$$\hat{\beta}_n = \frac{k_n^{-1} \sum_{i=1}^{k_n} \log\log(n/i) - \log\log(n/k_n)}{k_n^{-1} \sum_{i=1}^{k_n} \log(R_{n-i+1,n}) - \log(R_{n-k_n,n})} \; . \tag{17}$$

An estimator of c is then naturally given by

$$\hat{c}_n = \frac{1}{k} \sum_{i=1}^{k_n} \frac{\log(n/i)}{R_{n-i+1,n}^{\hat{\beta}_n}} \; . \tag{18}$$

Actually, in our context, the radial component is not observed. We estimate the R_i's by

$$\hat{R}_i^2 = \hat{X}_i^2 + (\hat{Y}_i - \hat{\rho}_n \hat{X}_i)^2/(1 - \hat{\rho}_n^2) \; ,$$

where $\hat{X}_i = (X_i - \hat{\mu}_X)/\hat{\sigma}_X$ and $\hat{Y}_i = (Y_i - \hat{\mu}_Y)/\hat{\sigma}_Y$, and we plug these values in (17) and (18). Define then $\hat{\psi}_n(x) = x^{1-\hat{\beta}_n}/(\hat{c}_n \hat{\beta}_n)$.

3.2 Discussion of the estimation error versus approximation error

Denote

$$\hat{z}_{n,1} = \frac{y - \hat{\rho}_n x}{\sqrt{1-\hat{\rho}_n^2}\sqrt{x\hat{\psi}_n(x)}} \; , \quad \hat{z}_{n,2} = \frac{y - \hat{\rho}_n x - \hat{\rho}_n \hat{\psi}_n(x)}{\sqrt{1-\hat{\rho}_n^2}\sqrt{x\hat{\psi}_n(x)}} \; ,$$
$$z_1 = \frac{y - \rho x}{\sqrt{(1-\rho^2)}\sqrt{x^{2-\beta}/(c\beta)}} \; , \quad z_2 = \frac{y - \rho x - \rho x^{1-\beta}/(c\beta)}{\sqrt{(1-\rho^2)}\sqrt{x^{2-\beta}/(c\beta)}} \; .$$

Then, for $i = 1, 2$,

$$\hat{\theta}_{n,i}(x,y) - \theta(x,y) = \Phi(\hat{z}_{n,i}) - \Phi(z_i) + \Phi(z_i) - \theta(x,y) \; .$$

This shows that the estimators defined in (12) and (13) have two sources of error: the first one, $\Phi(\hat{z}_{n,i}) - \Phi(z_i)$, comes from the estimation of ρ, μ, σ and ψ, and the second one, $\Phi(z_i) - \theta(x,y)$, from the asymptotic nature of the approximations (4) and (6).

The order of magnitude of the estimation error can be measured by the rate of convergence of the estimators. In order to obtain a rate of convergence for the estimators $\hat{\beta}_n$ and $\hat{c}_n$, we assume that H is a Von Mises distribution function with

$$\psi_H(x) = \frac{1}{c\beta} x^{1-\beta}\{1 + t(x)\} , \tag{19}$$

where t is a regularly varying function [1] with index $\eta\beta$ for some $\eta < 0$. This implies that $\bar{H}(x) = \exp\{-cx^\beta[1 + s(x)]\}$, where s is also regularly varying with index $\eta\beta$. Under this assumption, the function ψ defined in (16) is an admissible auxiliary function and Girard (2004) has shown that $\hat{\beta}_n$ is $k_n^{1/2}$-consistent, for any sequence k_n such that

$$k_n \to \infty , \quad k_n^{1/2} \log^{-1}(n/k_n) \to 0 , \quad k_n^{1/2} b(\log(n/k_n)) \to 0 ,$$

where b is regularly varying with index η, see (Girard, 2004, Theorem 2) for details. Similarly, it can be shown that $k_n^{1/2}(\hat{c}_n - c) = O_P(1)$ under the same assumptions on the sequence k_n. Thus, for any x, $\hat{\psi}_n(x)$ is a $k_n^{1/2}$-consistent estimator of $\psi(x)$. Besides, $\hat{\rho}_n$, $\hat{\mu}_n$ and $\hat{\sigma}_n$ are $\sqrt{n}$-consistent, so the estimation error $\Phi(\hat{z}_{n,i}) - \Phi(z_i)$ is of order $k_n^{-1/2}$ in probability.

If ψ_H satisfies (19), then Hypothesis 2 holds, and Theorem 3 and Lemma 12 provide a bound for the deterministic approximation error $\Phi(z_i) - \theta(x, y)$. Some easy computation shows that the second order correction is useful only if $\eta < -1/2$.

3.3 Discussion of an alternative method

The method described previously makes use of the asymptotic approximations of Theorem 3. It could be thought that a direct method, making use of Formula (A.1) and of an estimator of $\bar{H}$, would yield a better estimate of $\theta(x, y)$, since it would avoid this approximation step. Recall however that we specifically need to estimate the tail of the radial distribution, so that a nonparametric estimator of H cannot be considered. The traditional solution given by extreme value theory consists in fitting a parametric model for the tail. This will always induce an approximation error.

Nevertheless, there exists a situation in which the approximation error can be canceled: If the radial component is exactly Weibull distributed, i.e. $\bar{H}(x) =$

[1] A function f is regularly varying at infinity with index α if for all $t > 0$, $\lim_{x\to\infty} f(tx)/f(x) = t^\alpha$.

$\exp\{-cx^\beta\}$, then ψ_H satisfies (16), so for any $x, u > 0$, (8) implies that

$$\frac{\bar{H}(xu)}{\bar{H}(x)} = \exp\left\{\int_x^{xu} \frac{\mathrm{d}s}{\psi(s)}\right\} .$$

Therefore, in this specific case, a consistent estimator of $\theta(x, y)$, say $\hat{\theta}_{n,3}(x, y)$, can be introduced, which does not make use of any asymptotic expansion as in Theorem 3. Through formula (A.1), we get explicitly:

$$\hat{\theta}_{n,3}(x, y) = 1 - \frac{\int_{\arctan(\hat{t}_0)}^{\pi/2} \hat{K}(\hat{x}, \hat{x}, \cos(u))\,\mathrm{d}u + \int_{-\hat{U}_0}^{\arctan(\hat{t}_0)} \hat{K}(\hat{x}, \hat{y}, \sin(u + \hat{U}_0))\,\mathrm{d}u}{2\int_0^{\pi/2} \hat{K}(\hat{x}, \hat{x}, \cos(u))\,\mathrm{d}u} ,$$

where

$$\hat{K}(x, y, v) = \exp\left\{\int_x^{y/v} \frac{\mathrm{d}s}{\hat{\psi}_n(s)}\right\} ,$$
$$\hat{t}_0 = (\hat{y}/\hat{x} - \hat{\rho}_n)/\sqrt{1-\hat{\rho}_n^2} , \quad \hat{U}_0 = \arctan(\hat{\rho}_n/\sqrt{1-\hat{\rho}_n^2}) ,$$
$$\hat{x} = (x - \hat{\mu}_X)/\hat{\sigma}_X , \quad \hat{y} = (y - \hat{\mu}_Y)/\hat{\sigma}_Y .$$

We included this estimator in the simulation study as a benchmark when looking at elliptical Kotz distributed observations (see Table 1).

4 Simulation study

To assess the performance of the proposed estimators, we simulated several families of bivariate standard elliptical distributions. Recall that a standardized bivariate elliptical density function can be written as $f(x, y) = Cg\{(x^2 - 2\rho xy + y^2)/(1-\rho^2)\}$, where g is called the generator, ρ is the Pearson correlation coefficient and C is a normalizing constant. The density of the radial component R is given by $H'(r) = Krg(r^2)$, where K is a normalizing constant (see, e.g., Fang et al. (1990)).

The distributions used are presented in Table 1. The Pearson correlation coefficient will be either $\rho = 0.5$ or $\rho = 0.9$. Three of them (Normal, Kotz and Logis) are Von Mises distributions which satisfy both Hypothesis 2 and (19); in addition, the Von Mises auxiliary function of the Kotz distribution satisfies (16). The Lognor and the modified Kotz distributions satisfy Hypothesis 2 but not (16), and finally the bivariate Student distribution has a regularly varying radial component, so it does not satisfy any of the assumptions. These three distributions are used to explore the robustness of the proposed estimation method.

Table 1
Bivariate elliptical distributions used for the simulations. The elliptical generator g is given, the Von Mises auxiliary function ψ_H (or an equivalent), the function χ defined in Hypothesis 2, and the values of the parameters used (in addition to $\rho \in \{0.5, 0.9\}$).

Bivariate law	Generator $g(u)$	$\psi_H(x)$	$\chi(x)$	Parameters
Normal	$\mathrm{e}^{-u/2}$	$\frac{1}{x} + O\left(\frac{1}{x^3}\right)$	$O(x^{-2})$	
Kotz	$u^{\beta/2-1}\mathrm{e}^{-u^{\beta/2}}$	$x^{1-\beta}/\beta$	$O(x^{-\beta})$	$\beta \in \{1, 4\}$
Logis†	$\frac{\mathrm{e}^{-u}}{(1+e^{-u})^2}$	$\frac{1+\mathrm{e}^{-x^2}}{2x}$	$O(x^{-2})$	
Modified Kotz	$g_\star(u)$‡	$\frac{x^{1-\beta}}{1+\beta \log x}$	$O\left(\frac{x^{-\beta}}{\log x}\right)$	$\beta = 3/2$
Lognor†	$\frac{1}{u}\mathrm{e}^{-(\log^2 u)/8}$	$\frac{x}{\log x} + O\left(\frac{x}{\log^3 x}\right)$	$O\left(\frac{1}{\log x}\right)$	
Student	$\left(1+\frac{u}{\nu}\right)^{-(\nu+2)/2}$	–	–	$\nu \in \{3, 20\}$

† "Logis" and "Lognor" refer to the elliptical distributions with logistic and lognormal generator, respectively.
‡ $g_\star(u) = \{(3/8)\log u + 1/2\}u^{-1/4}\exp\{-(1/2)u^{3/4}\log u\}$.

In each case, 200 samples of size 500 were simulated. Several values of x were chosen, corresponding to the $(1-p)$-quantile of the marginal distribution of X, with $p = 10^{-3}$, $p = 10^{-4}$ and $p = 10^{-5}$. For each value of x, we computed (by numerical integration) the theoretical values of y corresponding to the probability $\theta(x, y) = .05, .1, .2, \ldots, .8, .9, .95$. Then we estimated $\theta(x, y)$ via the three proposed methods (cf. Section 3). For the estimation of the auxiliary function ψ, the threshold chosen corresponds to the highest 10% of the estimated $\hat{R}_i$s. It must be noted that this choice is independent of x. For each fixed x we also estimated the conditional quantile function $\theta(x, \cdot)^{\leftarrow}$ by both methods (14) and (15). We did not compute the estimated quantile function for Method 3 since it would involve the numerical inversion of the integrals wich appear in (A.1). This actually is one advantage of Methods 1 and 2 over Method 3.

Some general features can be observed, which comfort the theoretical expectations. (i) First of all, in the given range of x and y, there were hardly any observations, so that the empirical conditional distribution function is useless. (ii) For a given probability θ, the variability of the estimators slightly increases with x for all underlying distributions. For a given x, the variability of the estimators is greater for medium values of θ. (iii) The results for the Student distribution are as expected: if the degree of freedom ν is large, the estimation shows a high variability but moderate bias, while if ν is small, then the estimation is clearly inconsistent. (iv) The results for the Logis and modified

Kotz distributions are similar to the Gaussian distribution. (v) As described in Section 3.3, Method 3 is markedly better for the Kotz distribution only.

Hence we have chosen to report only the results for the largest value of x (corresponding to the 10^{-5}-quantile of the marginal distribution of X) and the Normal, Kotz (with parameter $\beta = 1$ and $\beta = 4$) and Lognor distributions, for $\rho = .5$ and $\rho = .9$. Figures A.1-A.4 illustrate the behavior of the estimators of the probability θ: Median, 2.5% and 97.5% quantiles of the estimation error $\hat{\theta}_{n,i}(x,y) - \theta(x,y)$ $(i = 1,2,3)$ are shown as a function of the estimated probability. Figure A.5 shows the estimated conditional quantile functions $\hat{y}_{n,i}(x,y)$ $(i = 1,2)$ and the theoretical conditional quantile function $y = \theta(x,\cdot)^{\leftarrow}$ for only three distributions and $\rho = .9$, because the results are much more stable as the correlation or the distribution vary. Median, 2.5% and 97.5% quantiles of the estimated conditional quantile function $\hat{y}_{n,i}(x,y)$ $(i = 1,2)$ are given as a function of the probability.

From these simulation results, one can see that the estimator of θ by Method 1 presents a systematic positive bias, which of course induces an underestimation of the conditional quantile function. As expected, Method 2 corrects this systematic bias; the correction is better when ρ is large. This is also true for the Lognormal generator, though in a lesser extent.

As already mentioned, the Lognor and modified Kotz distributions do not satisfy Assumption (19). In both cases the radial component belongs to an extended Weibull type family, with auxiliary function ψ of the form

$$\psi(x) = cx^{1-\beta}(\log x)^{-\delta}\{1 + o(1)\} ,$$

with $c > 0$, $\beta \geq 0$ and $\delta > 0$ if $\beta = 0$. The modified Kotz distribution corresponds to $\beta = 3/2$ and $\delta = 1$ and Lognor corresponds to $\beta = 0$ and $\delta = 1$. The simulation results are much better for the modified Kotz than for the Lognor distribution. This tends to prove that the method is not severely affected by the logarithmic factor, as long as $\beta > 0$.

5 Financial application

In this section, the practical usefulness of our estimation procedure is illustrated in the context of financial contagion, for which an estimation of the conditional excess probability is needed. Data used by Levy and Duchin (2004) are here revisited. More precisely, we consider the series of monthly returns for the 3M stock and the Dow Jones Industrial Average, for the period ranging from January 1970 to January 2008. In the sequel, we arbitrarily investigate the conditional behavior of the 3M stock monthly returns, given some extreme values of the Dow Jones Industrial Average.

According to Levy and Duchin (2004), these two series can be marginally fitted by a logistic distribution. Indeed, a Kolmogorov-Smirnov goodness-of-fit test of the logistic distribution gave us a P-value of 0.48 for the 3M returns and 0.49 for the Dow Jones Industrial Average returns. These P-values were obtained via a Monte-carlo simulation, following the procedure outlined by Stephens (1979). Moreover, the test of elliptical symmetry of Huffer and Park (2007) was used to show that the data fit the bivariate elliptical model (P-value= 0.61). Finally, we checked that both marginal upper tails exhibit rapid variation; for this, we performed a Generalized Pareto Distribution fit to the 15%-largest values, and checked via a test based on the profile likelihood 95%-confidence interval that the shape parameter could be considered as equal to 0 at level 5% (see e.g. Coles (2001) for details on these classical procedures).

Consequently, the estimation procedures presented in the previous sections can be applied to these data. As an illustration, we depict in Figure A.6 the three estimates of $y \mapsto 1 - \theta(x, y) = \mathbb{P}(Y > y \mid X > x)$, for different values of x corresponding to the .975, .99, .999 and .9999 quantiles of the fitted logistic distribution, together with the estimated marginal survival function $\mathbb{P}(Y > y)$. This last probability was estimated via the logistic distribution fitted to the Y_i's. It clearly stems from these graphics that $1 - \hat{\theta}_{n,2}$ and $1 - \hat{\theta}_{n,3}$ provide very similar estimates, whereas $1 - \hat{\theta}_{n,1}$ gives uniformly smaller values. All these estimates are uniformly greater than the marginal survival function of Y. This allows to conclude that the data exhibit contagion from the Dow Jones Industrial Average to the 3M stock.

6 Concluding remarks

In this paper we restricted attention to elliptically distributed random pairs (X, Y) having a rapidly varying radial component. Three methods have been proposed to estimate the conditional excess probability $\theta(x, y)$ for large x. Under this specific assumption of elliptical distributions, Methods 2 and 3 revealed comparable results and outperformed Method 1.

Methods 1 and 2 make use of an asymptotic approximation of the conditional excess distribution function by the Gaussian distribution function. As shown by Balkema and Embrechts (2007), this approximation remains valid outside the family of elliptical distributions, under geometric assumptions on the level curves of the joint density function of (X, Y). This suggests that these methods may be robust to departure from the elliptical family. This is an undergoing research project.

Acknowledgement

The first author gratefully acknowledges the support of the Natural Sciences and Engineering Research Council of Canada. We also acknowledge R-project and programs.

References

Abdous, B., Fougères, A.-L., Ghoudi, K., 2005. Extreme behaviour for bivariate elliptical distributions. Canad. J. Statist. 33 (3), 317–334.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Math. Finance 9 (3), 203–228.

Balkema, G., Embrechts, P., 2007. High Risk Scenarios and Extremes. A Geometric Approach. European Mathematical Society. Publishing House.

Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. Extremes 2 (2), 177–200.

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., 2004. Statistics of extremes. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.

Beirlant, J., Raoult, J.-P., Worms, R., 2003. On the relative approximation error of the generalized pareto approximation for a high quantile. Extremes 6 (4), 335–360.

Beirlant, J., Vynckier, P., Teugels, J., 1996. Practical analysis of extreme values. Leuwen University Press.

Berman, S. M., 1992. Sojourns and extremes of stochastic processes. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

Bradley, B. O., Taqqu, M., 2004. Framework for analyzing spatial contagion between financial markets. Finance Letters 2 (6), 8–16.

Bradley, B. O., Taqqu, M., 2005. Empirical evidence on spatial contagion between financial markets. Finance Letters 3 (1), 77–86.

Chernozhukov, V., 2005. Extremal quantile regression. Ann. Statist. 33 (2), 806–839.

Coles, S., 2001. An introduction to statistical modeling of extreme values. Springer Series in Statistics. Springer-Verlag London Ltd., London.

de Haan, L., 1970. On regular variation and its application to the weak convergence of sample extremes. Vol. 32 of Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam.

de Haan, L., 1985. Extremes in higher dimensions: the model and some statistics. Bull. Inst. Internat. Statist. 51 (4), No. 6.3, 15 pp.; Vol. V: pp. 185–192.

de Haan, L., Resnick, S. I., 1977. Limit theory for multivariate sample extremes. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 40 (4), 317–337.

de Haan, L., Stadtmüller, U., 1996. Generalized regular variation of second order. J. Austral. Math. Soc. Ser. A 61 (3), 381–395.

Dierckx, G., Beirlant, J., de Waal, D. J., Guillou, A., 2007. A new estimation method for weibull-type tails, preprint.

Draisma, G., Drees, H., Ferreira, A., de Haan, L., 2004. Bivariate tail estimation: dependence in asymptotic independence. Bernoulli 10 (2), 251–280.

Eddy, W. F., Gale, J. D., 1981. The convex hull of a spherically symmetric sample. Adv. in Appl. Probab. 13 (4), 751–763.

Embrechts, P., Lindskog, F., McNeil, A., 2003. Modelling Dependence with Copulas and Applications to Risk Management, In: Handbook of Heavy Tailed Distributions in Finance, ed. S. Rachev. Chapter 8, pp. 329-384. Elsevier.

Fang, K. T., Kotz, S., Ng, K. W., 1990. Symmetric multivariate and related distributions. Vol. 36 of Monographs on Statistics and Applied Probability. Chapman and Hall Ltd., London.

Frahm, G., Junker, M., Schmidt, R., 2005. Estimating the tail-dependence coefficient: properties and pitfalls. Insurance Math. Econom. 37 (1), 80–100.

Gardes, L., Girard, S., 2006. Comparison of Weibull tail-coefficient estimators. REVSTAT 4 (2), 163–188.

Girard, S., 2004. A Hill type estimator of the Weibull tail-coefficient. Comm. Statist. Theory Methods 33 (2), 205–234.

Hashorva, E., 2006. Gaussian approximation of elliptical random vectors. Stoch. Models 22 (3), 441–457.

Hashorva, E., Kotz, S., Kume, A., 2007. L_p-norm generalised symmetrised Dirichlet distributions. Albanian J. Math. 1 (1), 31–56 (electronic).

Heffernan, J. E., Resnick, S. I., 2007. Limit laws for random vectors with an extreme component. Ann. Appl. Probab. 17 (2).

Heffernan, J. E., Tawn, J. A., 2004. A conditional approach for multivariate extreme values. Journal of the Royal Statistical Society. Series B 66 (3), 497–546.

Huffer, F., Park, C., 2007. A test for elliptical symmetry. Journal of Multivariate Analysis 98, 256–281.

Hult, H., Lindskog, F., 2002. Multivariate extremes, aggregation and dependence in elliptical distributions. Adv. in Appl. Probab. 34 (3), 587–608.

Koenker, R., 2005. Quantile regression. Vol. 38 of Econometric Society Monographs. Cambridge University Press, Cambridge.

Koenker, R., Bassett, Jr., G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.

Ledford, A. W., Tawn, J. A., 1996. Statistics for near independence in multivariate extreme values. Biometrika 83 (1), 169–187.

Ledford, A. W., Tawn, J. A., 1997. Modelling dependence within joint tail regions. J. Roy. Statist. Soc. Ser. B 59 (2), 475–499.

Levy, H., Duchin, R., 2004. Asset return distributions and the investment horizon. J. Portfolio Management 30 (3), 47–62.

Maulik, K., Resnick, S. I., 2004. Characterizations and examples of hidden regular variation. Extremes 7 (1), 31–67.
Pickands, III, J., 1981. Multivariate extreme value distributions. Bull. Inst. Internat. Statist. 49 (2), 859–878, 894–902.
Resnick, S. I., 1987. Extreme values, regular variation, and point processes. Vol. 4 of Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York.
Resnick, S. I., 2002. Hidden regular variation, second order regular variation and asymptotic independence. Extremes 5 (4), 303–336.
Stephens, M., 1979. Tests of fit for the logistic distribution based on the empirical distribution function. Biometrika 66 (3), 591–595.

A Appendix

Proof of Theorem 3:
Define $U_0 = \arctan(\rho/\sqrt{1-\rho^2})$. For $x > 0$ and $y \in (0, x)$, we have

$$\mathbb{P}(X > x\ ,\ Y > y) = \mathbb{P}\left(R > \frac{x}{\cos U} \vee \frac{y}{\rho\cos U + \sqrt{1-\rho^2}\sin U}\ ;\ -U_0 \leq U \leq \frac{\pi}{2}\right)\ .$$

Set $t_0 = (y/x - \rho)/\sqrt{1-\rho^2}$. Then $-U_0 < \arctan(t_0)$, and $x/\cos u > y/(\rho\cos u + \sqrt{1-\rho^2}\sin u)$ if and only if $u > \arctan(t_0)$. Hence,

$$\begin{aligned}\mathbb{P}(X > x\ ,\ Y > y) = &\int_{\arctan(t_0)}^{\pi/2} \bar{H}\left\{\frac{x}{\cos u}\right\}\frac{\mathrm{d}u}{2\pi} \\ &+ \int_{-U_0}^{\arctan(t_0)} \bar{H}\left\{\frac{y}{\sin(u+U_0)}\right\}\frac{\mathrm{d}u}{2\pi}\ ,\end{aligned}$$

$$\begin{aligned}&\mathbb{P}(Y > y \mid X > x) \\ &\quad = \frac{\int_{\arctan(t_0)}^{\pi/2} \bar{H}(x/\cos(u))\mathrm{d}u + \int_{-U_0}^{\arctan(t_0)} \bar{H}(y/\sin(u+U_0))\mathrm{d}u}{2\int_0^{\pi/2} \bar{H}(x/\cos(u))\mathrm{d}u}\ . \qquad \text{(A.1)}\end{aligned}$$

If $t_0 \geq 0$, i.e. $y - \rho x \geq 0$, the changes of variables $v = 1/\cos(u)$ and $v = 1/\sin(u+U_0)$ yield:

$$\mathbb{P}(Y > y \mid X > x) = \frac{I_1 + I_2}{I_3}\ , \qquad \text{(A.2)}$$

with

$$
\begin{aligned}
I_1 &= \int_{w_1}^{\infty} \frac{\bar H(vx)}{\bar H(x)} \frac{\mathrm{d}v}{v\sqrt{v^2-1}} \;, \\
I_2 &= \int_{w_2}^{\infty} \frac{\bar H(vy)}{\bar H(x)} \frac{\mathrm{d}v}{v\sqrt{v^2-1}} \;, \\
I_3 &= 2\int_{1}^{\infty} \frac{\bar H(vx)}{\bar H(x)} \frac{\mathrm{d}v}{v\sqrt{v^2-1}} \;, \\
w_1 &= \sqrt{1+t_0^2} = \sqrt{1+(y/x-\rho)^2/(1-\rho^2)} \;, \\
w_2 &= x w_1/y \;.
\end{aligned}
$$

If $t_0 < 0$, then

$$
\mathbb{P}(Y > y \mid X > x) = \frac{I_3 - I_1 + I_2}{I_3} \;. \tag{A.3}
$$

Denote $w_0 = x(w_1-1)/\psi(x)$. In I_1 and I_3, the change of variable $v = 1 + \frac{\psi(x)}{x} t$ yields

$$
\begin{aligned}
I_1 &= \sqrt{\frac{\psi(x)}{x}} \int_{w_0}^{\infty} \frac{\bar H(x+t\psi(x))}{\bar H(x)} \frac{\mathrm{d}t}{(1+t\frac{\psi(x)}{x})\sqrt{1+\frac{t}{2}\frac{\psi(x)}{x}}\sqrt{2t}} \;, \\
I_3 &= 2\sqrt{\frac{\psi(x)}{x}} \int_{0}^{\infty} \frac{\bar H(x+t\psi(x))}{\bar H(x)} \frac{\mathrm{d}t}{(1+t\frac{\psi(x)}{x})\sqrt{1+\frac{t}{2}\frac{\psi(x)}{x}}\sqrt{2t}} \;.
\end{aligned}
$$

In I_2, the change of variable $vy = x + t\psi(x)$ yields

$$
I_2 = \frac{\psi(x)}{x} \int_{w_0}^{\infty} \frac{\bar H(x+t\psi(x))}{\bar H(x)} \frac{(y/x)\,\mathrm{d}t}{(1+t\frac{\psi(x)}{x})\sqrt{1-(y/x)^2+2t\frac{\psi(x)}{x}+\frac{\psi^2(x)}{x^2}t^2}} \;.
$$

Set $J_i = \sqrt{x/\psi(x)} I_i$, $i = 1, 3$ and $J_2 = (x/\psi(x)) I_2$. We start with I_1 and I_3. We will use the following bound, valid for all $B, C > 0$,

$$
0 \le 1 - \frac{1}{(1+C)\sqrt{1+B}} \le B/2 + C \;, \tag{A.4}
$$

which follows from straightforward algebra and the concavity of the function $x \mapsto \sqrt{1+x}$. Applying this bound with $B = \frac{\psi(x)}{x}\frac{t}{2}$ and $C = \frac{\psi(x)}{x} t$ yields

$$
0 \le 1 - \frac{1}{(1+t\frac{\psi(x)}{x})\sqrt{1+\frac{t}{2}\frac{\psi(x)}{x}}} \le \frac{5}{4}\frac{\psi(x)}{x} t \;.
$$

We thus have

$$
\left|J_1 - \sqrt{2\pi}\bar\Phi(\sqrt{2w_0})\right| + \left|J_3 - \sqrt{2\pi}\right| \le 3\chi(x)\int_0^{\infty} \Theta(t)\frac{\mathrm{d}t}{\sqrt{2t}} + \frac{15}{16}\sqrt{2\pi}\,\frac{\psi(x)}{x} \;.
$$

Hence

$$\frac{I_1}{I_3} = \bar{\Phi}(\sqrt{2w_0}) + O\left(\chi(x) + \frac{\psi(x)}{x}\right) . \tag{A.5}$$

Consider now J_2. Applying the bound (A.4) with $C = t\psi(x)/x$ and

$$B = \{2t\frac{\psi(x)}{x} + \frac{\psi^2(x)}{x^2}t^2\}/\sqrt{1-(y/x)^2} \,,$$

and making use of Hypothesis 2, we obtain

$$\left| J_2 - \frac{(y/x)\sqrt{2\pi}\varphi(\sqrt{2w_0})}{\sqrt{1-(y/x)^2}} \right|$$
$$\leq \frac{y/x}{\sqrt{1-(y/x)^2}}\,\chi(x)\int_0^\infty \Theta(t)\,\mathrm{d}t + \frac{y/x}{1-(y/x)^2}\frac{\psi(x)}{x}\left(\sqrt{1-\left(\frac{y}{x}\right)^2} + 1 + \frac{\psi(x)}{x}\right) .$$

Choose $y = \rho x + \sqrt{1-\rho^2}\sqrt{x\psi(x)}z$ for some fixed $z \in \mathbb{R}$. Then for large enough x it does hold that $0 < y < x$ and

$$\frac{y/x}{\sqrt{1-(y/x)^2}} = \frac{\rho}{\sqrt{1-\rho^2}} + O\left(\sqrt{\psi(x)/x}\right) .$$

Thus,

$$\frac{I_2}{I_3} = \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\frac{\psi(x)}{x}}\,\varphi(\sqrt{2w_0}) + \,O\left(\chi(x) + \frac{\psi(x)}{x}\right) . \tag{A.6}$$

For $z \geq 0$ and large enough x, plugging (A.5) and (A.6) into (A.2) yields

$$\theta(x,y) = \Phi(\sqrt{2w_0}) - \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\frac{\psi(x)}{x}}\,\varphi(\sqrt{2w_0}) + \,O\left(\chi(x) + \frac{\psi(x)}{x}\right) .$$

For $z < 0$ and large enough x, plugging (A.5) and (A.6) into (A.3) yields

$$\theta(x,y) = \bar{\Phi}(\sqrt{2w_0}) - \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\frac{\psi(x)}{x}}\,\varphi(\sqrt{2w_0}) + \,O\left(\chi(x) + \frac{\psi(x)}{x}\right) .$$

Note now that $w_0 = z^2/2 + O(\psi(x)/x)$, hence $\sqrt{2w_0} = |z| + O(\psi(x)/x)$. Thus, in both cases $z \geq 0$ and $z < 0$, (4) holds. Set $z = z' + \rho\sqrt{\psi(x)/x}/\sqrt{1-\rho^2}$. A Taylor expansion of Φ and φ around z' yields (6). □

Proof of Lemma 7:

$$
\begin{aligned}
\frac{\bar{H}\{x+t\psi(x)\}}{\bar{H}(x)} - \mathrm{e}^{-t} &= \exp\left\{-\int_x^{x+t\psi(x)} \frac{\mathrm{d}s}{\psi(s)}\right\} - \mathrm{e}^{-t} \\
&= \exp\left[-\int_0^t \frac{\psi(x)}{\psi\{x+s\psi(x)\}}\mathrm{d}s\right] - \mathrm{e}^{-t} .
\end{aligned}
$$

Applying the inequality $|\mathrm{e}^{-a} - \mathrm{e}^{-b}| \le |a-b|\mathrm{e}^{-a\wedge b}$ valid for all $a, b \ge 0$ yields

$$
\begin{aligned}
&\left|\frac{\bar{H}\{x+t\psi(x)\}}{\bar{H}(x)} - \mathrm{e}^{-t}\right| \\
&\qquad \le \left|\int_0^t \frac{\psi\{x+s\psi(x)\}-\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s\right| \exp\left[-t \wedge \int_0^t \frac{\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s\right] .
\end{aligned}
$$

Denote

$$
\int_0^t \frac{|\psi\{x+s\psi(x)\}-\psi(x)|}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s = I(x,t)
$$

and

$$
\exp\left[-t \wedge \int_0^t \frac{\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s\right] = E(x,t) .
$$

Case ψ increasing. If ψ is nondecreasing, then $\psi' \ge 0$ and ψ' is decreasing. Thus, for any $\delta > 0$ and large enough x, $\psi'(x) \le \delta$ and

$$
\begin{aligned}
\int_0^t \frac{\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s &\ge \int_0^t \frac{\psi(x)}{\psi(x)+s\psi'(x)\psi(x)}\,\mathrm{d}s \\
&= \int_0^t \frac{\mathrm{d}s}{1+s\psi'(x)} \ge \int_0^t \frac{\mathrm{d}s}{1+s\delta} = \frac{1}{\delta}\log(1+\delta t) .
\end{aligned}
$$

This implies that $E(x,t) \le (1+\delta t)^{-1/\delta}$ for large enough x. Since ψ is increasing and ψ' is decreasing, we also have

$$
I(x,t) \le \psi'(x) \int_0^t \frac{s\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s \le \psi'(x)\,\frac{t^2}{2} .
$$

Thus, for any $\delta > 0$, $I(x,t)E(x,t) = O(|\psi'(x)|t^2(1+\delta t)^{-1/\delta})$.

Case ψ decreasing. If ψ is monotone non increasing, then

$$
\int_0^t \frac{\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s \ge t
$$

and $E(x,t) \le \mathrm{e}^{-t}$. Also, since $|\psi'|$ is decreasing,

$$
I(x,t) \le |\psi'(x)| \int_0^t \frac{s\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s .
$$

If ψ has a positive limit at infinity, then $I(x,t) = O(\psi'(x)t^2)$. Otherwise, $\lim_{x\to\infty}\psi(x) = 0$ and (9) holds. This yields, for large enough x,

$$\begin{aligned} I(x,t) &\le |\psi'(x)| \int_0^t \frac{s\psi(x)}{\psi\{x+s\psi(x)\}}\,\mathrm{d}s \\ &\le c_1|\psi'(x)| \int_0^t s\mathrm{e}^{c_2 s\psi(x)}\,\mathrm{d}s \le c_1|\psi'(x)| \int_0^t s\mathrm{e}^{s/2}\,\mathrm{d}s \,. \end{aligned}$$

Thus $I(x,t) = O(|\psi'(x)|t^2\mathrm{e}^{t/2})$ and $I(x,t)E(x,t) = O(|\psi'(x)|t^2\mathrm{e}^{-t/2})$. This concludes the proof. □

Proof of Lemma 12: Write

$$\left|\frac{\bar H\{x+t\psi(x)\}}{\bar H(x)} - \mathrm{e}^{-t}\right| \le \left|\frac{\bar H\{x+t\psi(x)\}}{\bar H\{x+t\psi_H(x)\}} - 1\right| \mathrm{e}^{-t} \tag{A.7}$$

$$+ \frac{\bar H\{x+t\psi(x)\}}{\bar H\{x+t\psi_H(x)\}} \left|\frac{\bar H\{x+t\psi_H(x)\}}{\bar H(x)} - \mathrm{e}^{-t}\right| \,, \tag{A.8}$$

$$\frac{\bar H\{x+t\psi(x)\}}{\bar H\{x+t\psi_H(x)\}} = \exp\left[-\int_t^{t\psi(x)/\psi_H(x)} \frac{\psi_H(x)}{\psi\{x+s\psi_H(x)\}}\,\mathrm{d}s\right] \,.$$

If ψ_H is increasing, then

$$\left|\int_t^{t\psi(x)/\psi_H(x)} \frac{\psi_H(x)}{\psi_H\{x+s\psi_H(x)\}}\,\mathrm{d}s\right| \le t\xi(x) \,, \tag{A.9}$$

$$\exp\left[-\int_t^{t\psi(x)/\psi_H(x)} \frac{\psi_H(x)}{\psi_H\{x+s\psi_H(x)\}}\,\mathrm{d}s\right] \le \mathrm{e}^{t\xi(x)} \,. \tag{A.10}$$

Since $\xi(x) \to 0$, gathering (A.9) and (A.10) yields, for large enough x,

$$\left|\frac{\bar H\{x+t\psi(x)\}}{\bar H\{x+t\psi_H(x)\}} - 1\right| \mathrm{e}^{-t} \le t\xi(x)\mathrm{e}^{-t/2} \,. \tag{A.11}$$

If ψ_H is decreasing and $\lim_{x\to\infty}\psi_H(x) > 0$, the ratio $\psi_H(x)/\psi_H\{x+s\psi(x)\}$ is bounded above and away from 0, so that

$$\left|\int_t^{t\psi(x)/\psi_H(x)} \frac{\psi_H(x)}{\psi\{x+s\psi_H(x)\}}\,\mathrm{d}s\right| \le Ct\xi(x),$$

and (A.11) still holds.

If ψ_H is decreasing and $\lim_{x\to\infty}\psi_H(x) = 0$, then applying (9) gives that the left-hand side of the previous equation is bounded by $Ct\xi(x)\exp\{2c_2\psi(x)t\}$. Thus, for large enough x, (A.11) still holds.

This provides a bound for the right-hand side of (A.7). The term (A.8) is bounded by Lemma 7. □

Proof of (10): We proceed as in the proof of Lemma 7.

$$\begin{aligned}\frac{\bar H\{x+t\psi(x)\}}{\bar H(x)} &= \exp\left\{-\int_{[x]}^{[x+t\psi(x)]}\frac{\mathrm{d}s}{\psi(s)}\right\}\\ &= \exp\left[-\int_0^{([x+t\psi(x)]-[x])/\psi(x)}\frac{\psi(x)}{\psi([x]+s\psi(x))}\,\mathrm{d}s\right]\,,\end{aligned}$$

$$\begin{aligned}&\int_0^{([x+t\psi(x)]-[x])/\psi(x)}\frac{\psi(x)}{\psi([x]+s\psi(x))}\,\mathrm{d}s - t\\ &\quad= \int_0^{([x+t\psi(x)]-[x])/\psi(x)}\frac{\psi(x)-\psi([x]+s\psi(x))}{\psi([x]+s\psi(x))}\,\mathrm{d}s + \frac{[x+t\psi(x)]-[x]}{\psi(x)} - t\,.\end{aligned}$$

By concavity of ψ,

$$\frac{|\psi(x)-\psi([x]+s\psi(x))|}{\psi([x]+s\psi(x))} = O((s\vee 1)\psi'([x]))\,.$$

By definition of the integral part,

$$t-1/\psi(x) \le \frac{[x+t\psi(x)]-[x]}{\psi(x)} \le t+1/\psi(x)\,.$$

The previous bounds yield

$$\left|\int_0^{([x+t\psi(x)]-[x])/\psi(x)}\frac{\psi(x)}{\psi([x]+s\psi(x))}\,\mathrm{d}s - t\right| = O\Big((t\vee 1)^2\psi'([x]) + 1/\psi(x)\Big)\,.$$

We must also give a lower bound for the integral. Since ψ is concave increasing, $\psi([x]+s\psi(x)) \le \psi(x)+s\psi'(x)\psi(x)$, hence

$$\begin{aligned}\int_0^{([x+t\psi(x)]-[x])/\psi(x)}\frac{\psi(x)}{\psi([x]+s\psi(x))}\,\mathrm{d}s &\ge \int_0^{t-1/\psi(x)}\frac{\mathrm{d}s}{1+s\psi'(x)}\\ &= \frac{1}{\psi'(x)}\log(1+t\psi'(x)-\psi'(x)/\psi(x))\,.\end{aligned}$$

Since $\psi'(x)\to 0$ and $\psi(x)\to\infty$, the arguments of the proof of Lemma 7 can be applied again to conclude the proof of (10). □

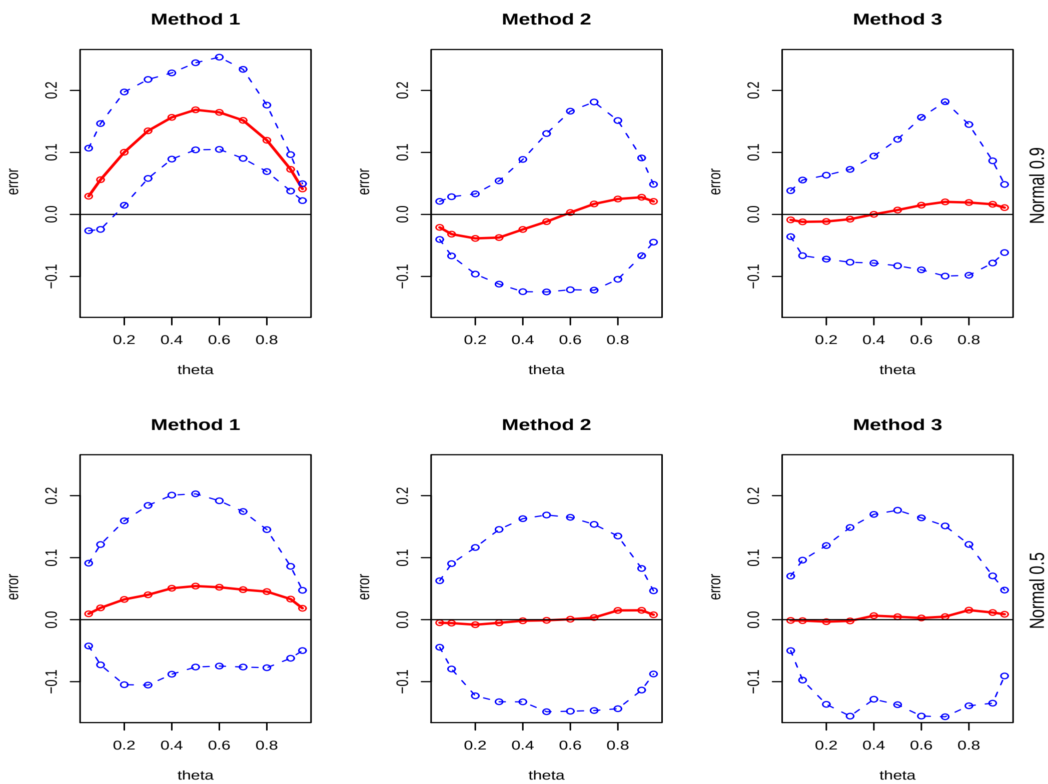

Fig. A.1. *Median, 2.5% and 97.5% quantiles of the estimation error* $\hat{\theta}_{n,i}(x,y) - \theta(x,y)$ $(i = 1, 2, 3)$ *as a function of the estimated probability. Gaussian distribution. First row:* $\rho = .9$*; second row:* $\rho = .5$.

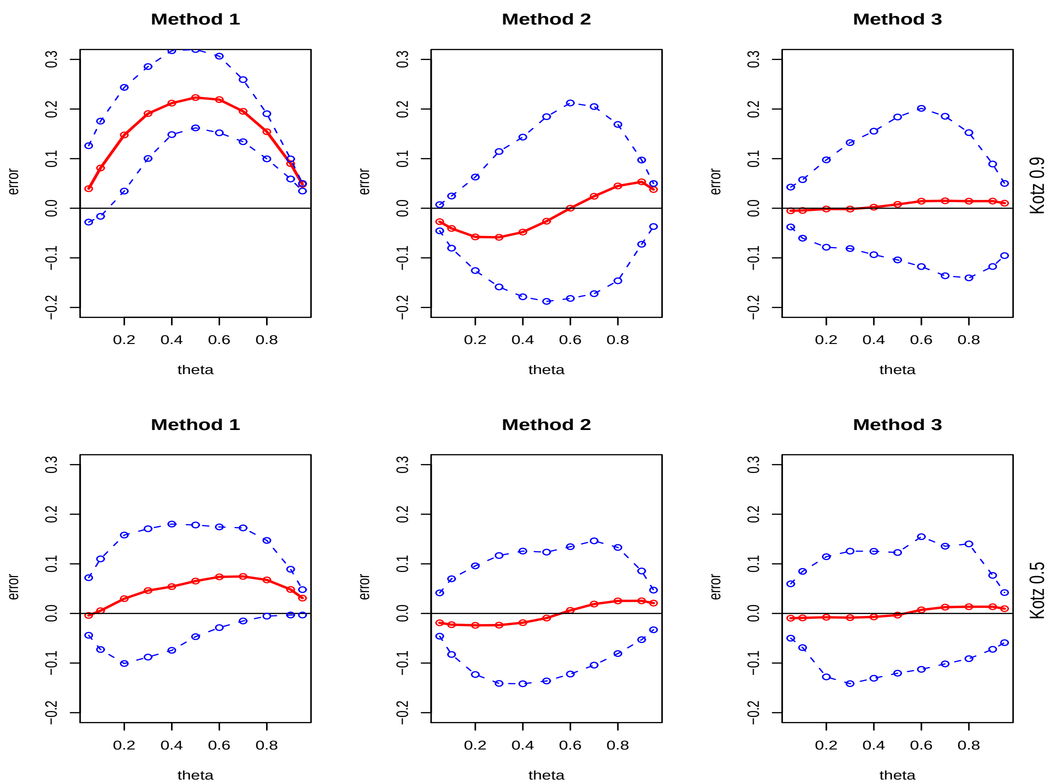

Fig. A.2. *Median, 2.5% and 97.5% quantiles of the estimation error* $\hat{\theta}_{n,i}(x,y) - \theta(x,y)$ $(i = 1, 2, 3)$ *as a function of the estimated probability. Kotz distribution,* $\beta = 1$*. First row:* $\rho = .9$*; second row:* $\rho = .5$*.*

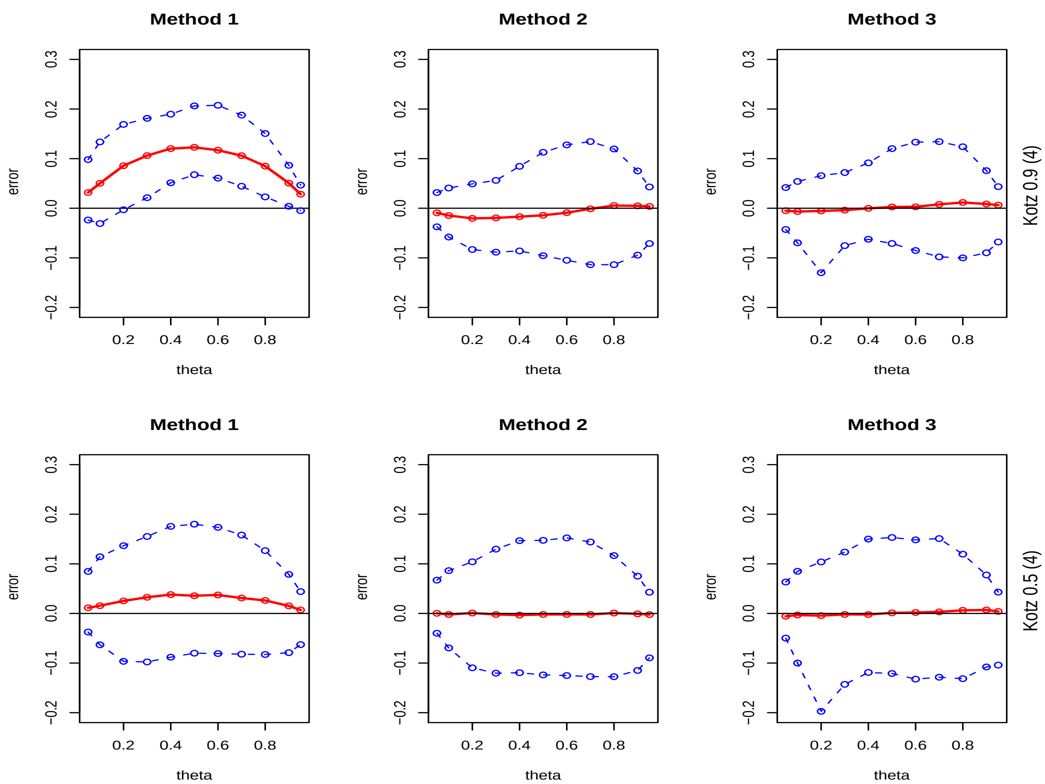

Fig. A.3. *Median, 2.5% and 97.5% quantiles of the estimation error* $\hat{\theta}_{n,i}(x,y) - \theta(x,y)$ $(i = 1,2,3)$ *as a function of the estimated probability. Kotz distribution,* $\beta = 4$*. First row:* $\rho = .9$*; second row:* $\rho = .5$*.*

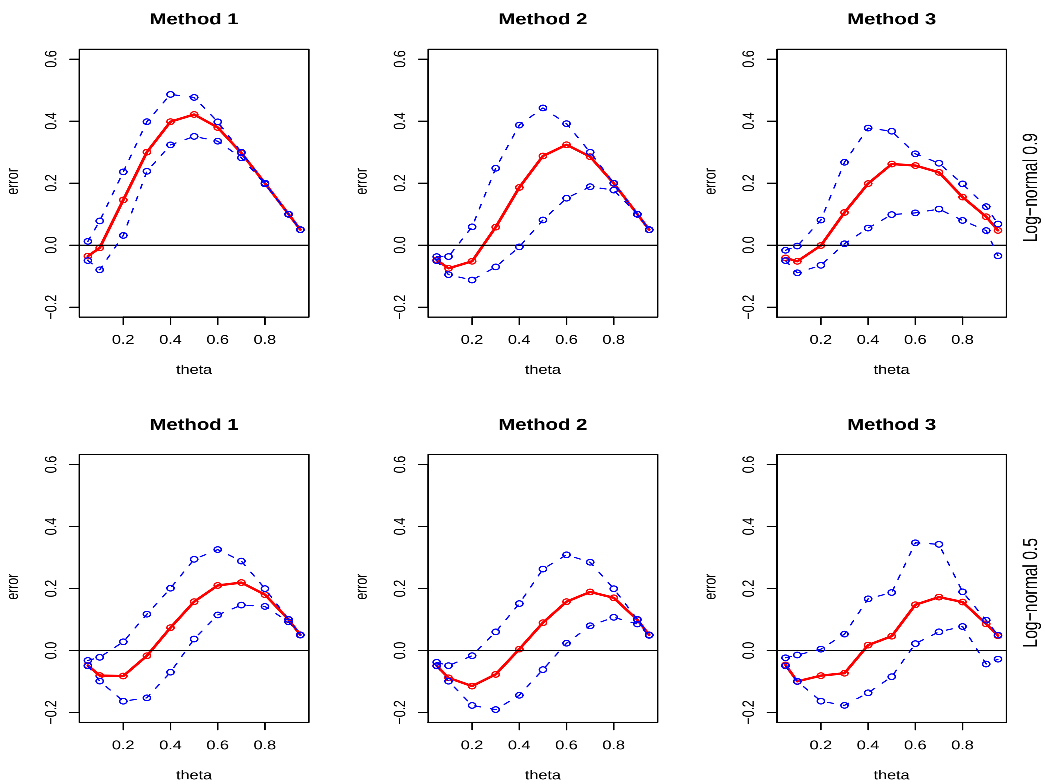

Fig. A.4. *Median, 2.5% and 97.5% quantiles of the estimation error* $\hat{\theta}_{n,i}(x,y) - \theta(x,y)$ $(i = 1,2,3)$ *as a function of the estimated probability. Lognor distribution. First row:* $\rho = .9$*; second row:* $\rho = .5$.

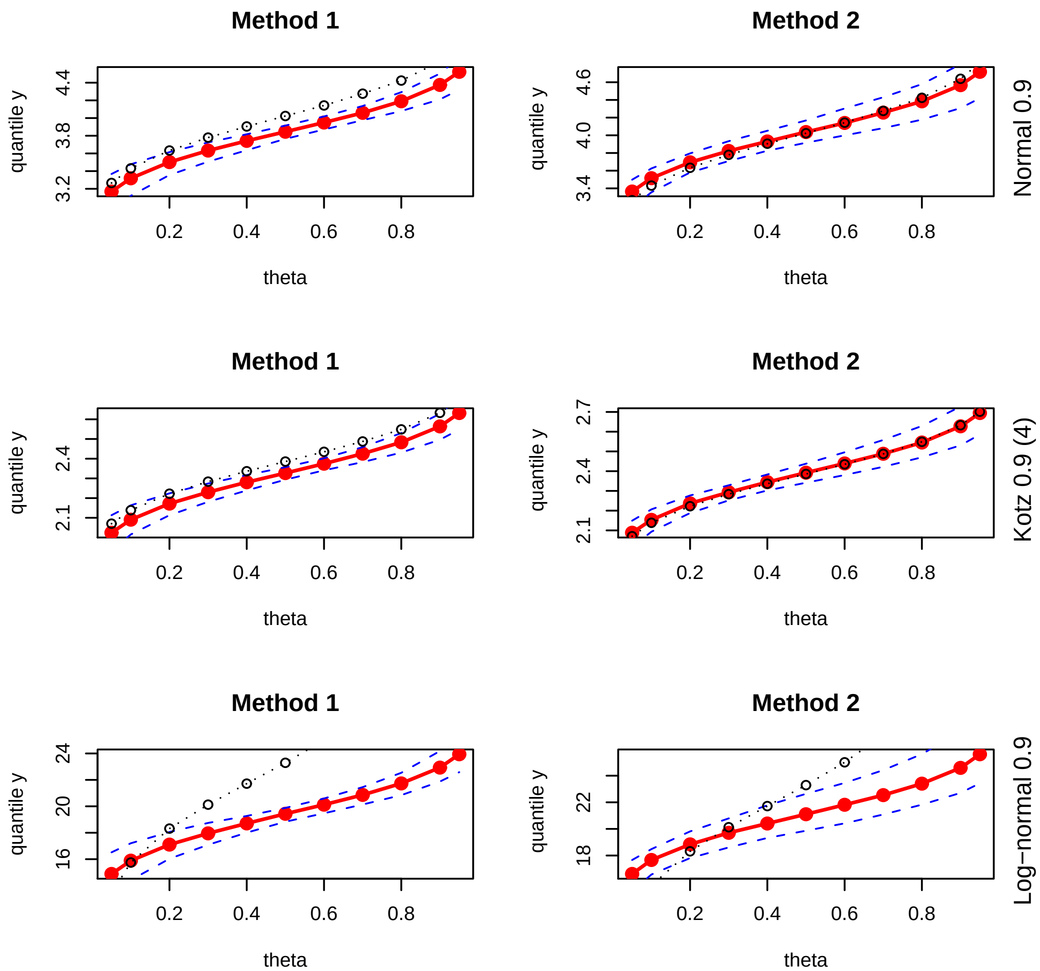

Fig. A.5. *Median (solid line), 2.5% and 97.5% quantiles (dashed lines) of the estimated conditional quantile function* $\hat{y}_{n,i}$ $(i = 1, 2)$ *defined in (14) and (15) and theoretical conditional quantile function* y *(dotted line) as a function of the probability* θ*. First row: Normal distribution; second row: Kotz distribution,* $\beta = 4$*; third row: Lognor distribution. For each of them* $\rho = .9$*.*

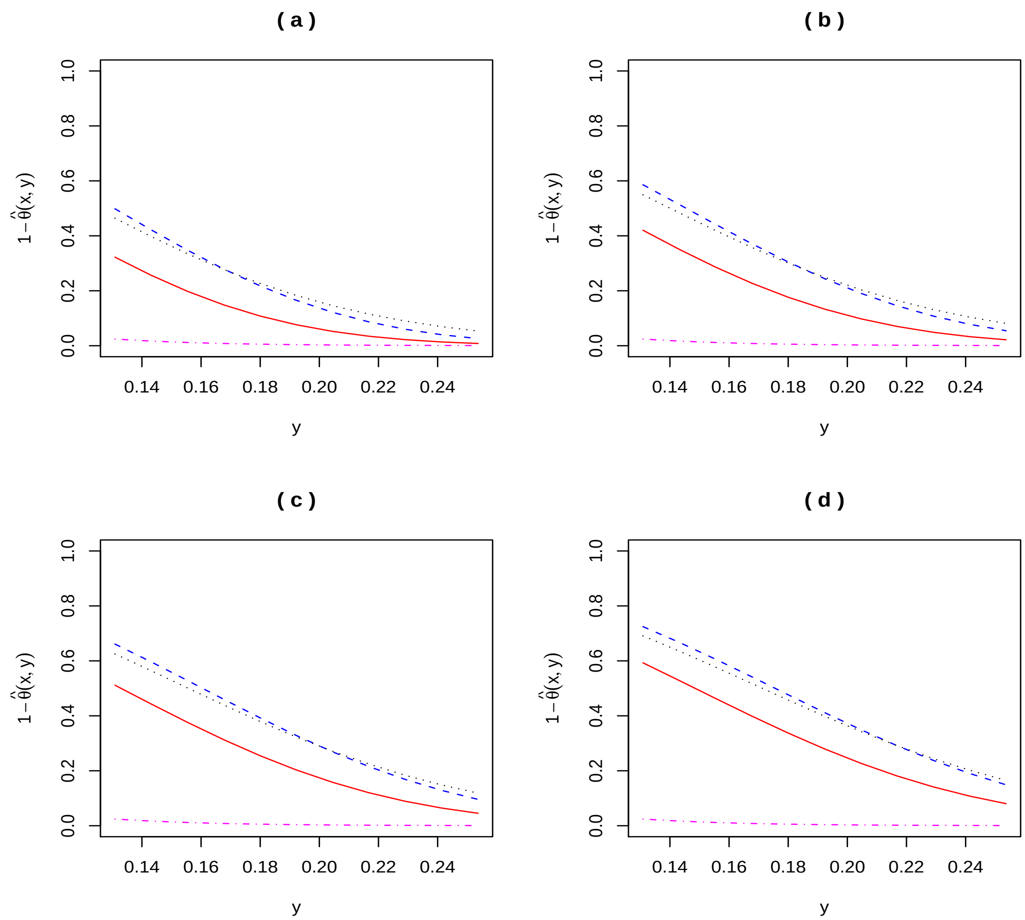

Fig. A.6. Estimates $y \mapsto 1-\hat{\theta}_{n,i}(x,y)$, $(i=1,2,3)$ of the conditional excess distribution of the 3M stock monthly return, given that the Dow Jones Industrial Average monthly return exceeds four extreme values. The subplots (a), (b), (c) and (d) are for the values x corresponding to the .975, .99, .999 and .9999 quantiles of the fitted logistic distribution respectively. The solid line is for $\hat{\theta}_{n,1}$, the dashed line for $\hat{\theta}_{n,2}$ and the dotted line for $\hat{\theta}_{n,3}$. The estimate of the marginal survival function $P(Y > y)$ is in dotted-dashed line.