



**HAL**  
open science

## Estimation of conditional laws given an extreme component

Anne-Laure Fougères, Philippe Soulier

► **To cite this version:**

Anne-Laure Fougères, Philippe Soulier. Estimation of conditional laws given an extreme component. 2009. hal-00288135v3

**HAL Id: hal-00288135**

**<https://hal.science/hal-00288135v3>**

Preprint submitted on 24 Oct 2009 (v3), last revised 13 Sep 2010 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of conditional laws given an extreme component

Anne-Laure Fougères\*

Philippe Soulier†

October 24, 2009

## Abstract

Let  $(X, Y)$  be a bivariate random vector. The estimation of a probability of the form  $P(Y \leq y \mid X > t)$  is challenging when  $t$  is large, and a fruitful approach consists in studying, if it exists, the limiting conditional distribution of the random vector  $(X, Y)$ , suitably normalized, given that  $X$  is large. There already exists a wide literature on bivariate models for which this limiting distribution exists. In this paper, a statistical analysis of this problem is done. Estimators of the limiting distribution (which is assumed to exist) and the normalizing functions are provided, as well as an estimator of the conditional quantile function when the conditioning event is extreme. Consistency of the estimators is proved and a functional central limit theorem for the estimator of the limiting distribution is obtained. The small sample behavior of the estimator of the conditional quantile function is illustrated through simulations.

## 1 Introduction

Let  $(X, Y)$  be a bivariate random vector for which the conditional distribution of  $Y$  given that  $X > t$  is of interest, for values of  $t$  such that the conditioning event is a rare event. This happens for example when the possible contagion between two dependent market returns  $X$  and  $Y$  is investigated, see e.g. [Bradley and Taqqu \(2004\)](#) or [Abdous et al. \(2008\)](#). The estimation of a probability of the form  $P(Y \leq y \mid X > t)$  starts to be challenging as soon as  $t$  is large, since the conditional empirical distribution becomes useless when no observations are available. A fruitful alternative approach consists in studying, if it exists, the limiting distribution of the random vector  $(X, Y)$  conditionally on  $X$  to be large. This corresponds to assuming that there exist functions  $m$ ,  $a$  and  $\psi$ , and a bivariate distribution function (cdf)  $F$  on  $[0, \infty) \times (-\infty, \infty)$  with non degenerate marginal distributions, such that

$$\lim_{t \rightarrow \infty} \mathbb{P}(X \leq t + \psi(t)x ; Y \leq m(t) + a(t)y \mid X > t) = F(x, y) \quad (1)$$

at all point of continuity of  $F$ . This approach was suggested by [Heffernan and Tawn \(2004\)](#) and investigated by [Heffernan and Resnick \(2007\)](#). Models for which condition (1) holds have already been investigated in many references. [Eddy and Gale \(1981\)](#) and [Berman \(1992\)](#)

---

\*Université Claude Bernard - Lyon 1

†Université Paris Ouest-Nanterre

proved that (1) holds for spherical distributions; bivariate elliptic distributions were investigated by [Abdous et al. \(2005\)](#), multivariate elliptic distributions and related distributions by [Hashorva \(2006\)](#); [Hashorva et al. \(2007\)](#). The analysis of the underlying geometric structure (ellipticity of the level sets of the densities) has lead to various generalizations by [Barbe \(2003\)](#) and [Balkema and Embrechts \(2007\)](#). See also [Fougères and Soulier \(2010\)](#) for a recent review on the subject.

An important issue that still has to be addressed is the statistical estimation of the functions  $a$  and  $m$  that appear in (1), as well as the limiting distribution function  $F$ . This is the aim of the present paper. Two problems are considered. The first one is the nonparametric estimation of the limiting distribution and of the normalizing functions. This allows for instance to test for a specific limiting distribution, e.g. the standard Gaussian distribution which appears in many examples. Since we are also interested in the case where the conditioning event is beyond the range of observations, a semiparametric procedure will be defined to allow this extrapolation. This can only be done under more restrictive assumptions, which are satisfied by most models already investigated in the literature.

The paper is organized as follows. In Section 2, we rephrase (1) in terms of vague convergence of measures in order to use the point process techniques and the results of [Heffernan and Resnick \(2007\)](#). We also introduce moment assumptions which are needed to prove the consistency of the non parametric estimators introduced in Section 3. A functional central limit theorem is obtained under a second order condition. A specific analysis of the case of a limiting distribution with product form is done in Section 4. The functional central limit theorem is used to derive a goodness of fit test for the second marginal of the limiting distribution  $F$ . In Section 4.2, semi-parametric estimators that allow extrapolations beyond the range of the observations are studied and applied to the estimation of conditional quantiles when the conditioning event is extreme. A simulation study is given in Section 5, which illustrates the behavior of the goodness of fit test proposed in Section 4.1 and of the estimator of the conditional quantile proposed in Section 4.2. This results are applied in Section 6 to some financial data. Section 7 collects the proofs.

## 2 Assumptions and preliminary results

We first rephrase the convergence (1) in terms of vague convergence of measures, in order to use point process techniques and the results of [Heffernan and Resnick \(2007\)](#). See also [Das and Resnick \(2008, 2009\)](#). Condition (1) implies that the marginal distribution of  $X$  belongs to the domain of attraction of an extreme value distribution with index  $\gamma \in \mathbb{R}$ , i.e. there exist normalizing sequences  $\{a_n\}$  and  $\{b_n\}$  with  $a_n > 0$  such that  $\mathbb{P}(\max_{1 \leq i \leq n} (X_i - b_n)/a_n \leq x)$  converges to  $\exp\{-\bar{P}_\gamma(x)\}$  for each  $x$  such that  $1 + \gamma x > 0$ , where  $\bar{P}_\gamma(x) = (1 + \gamma x)^{-1/\gamma}$  if  $\gamma \neq 0$  and  $\bar{P}_0(x) = e^{-x}$ , and the random variables  $X_i$  are independent copies of  $X$ . For simplicity, we assume that  $\gamma \geq 0$ , and in the case  $\gamma = 0$  we assume that the right endpoint of the marginal distribution of  $X$  is infinite.

Recall that measure defined on the Borel sigma-field of a locally compact separable space  $E$  is called a Radon measure if it is finite on compact sets. A sequence of Radon measures  $\sigma_n$  defined on  $E$  converges vaguely to a Radon measure  $\sigma$  if  $\int_E f(x)\sigma_n(dx)$  converges to  $\int_E f(x)\sigma(dx)$  for all compactly supported function  $f$ . See [Resnick \(1987, Chapter 3\)](#) or

[Heffernan and Resnick \(2007, Appendix A3\)](#). We will consider vague convergence of Radon measures defined on the Borel sigma-fields of  $(-1/\gamma, \infty]$  or  $(-1/\gamma, \infty] \times [-\infty, \infty]$ .

**Assumption 1.** *There exist  $\gamma \geq 0$ , monotone functions  $a, b, m$  and  $\psi$  such that the marginal distribution of  $X$  is in the domain of attraction of the extreme value distribution with extreme value index  $\gamma$  and the sequence of measures  $\nu_n$  defined by*

$$\nu_n(\cdot) = n\mathbb{P}\left(\left\{\frac{X - b(n)}{\psi \circ b(n)}, \frac{Y - m \circ b(n)}{a \circ b(n)}\right\} \in \cdot\right)$$

*converges vaguely on  $(-1/\gamma, \infty] \times [-\infty, \infty]$  to a Radon measure  $\nu$  such that  $\nu([0, \infty) \times (-\infty, \infty)) = 1$ , the distribution function  $y \mapsto \nu([0, \infty) \times (-\infty, y])$  is non degenerate and the application  $(x, y) \mapsto \nu([x, \infty) \times (-\infty, y])$  is continuous on  $(-1/\gamma, \infty] \times [-\infty, \infty]$ .*

The link between Assumption 1 and Equation (1) is that the limiting distribution  $F$  is given, for all positive  $x$  and real  $y$ , by

$$F(x, y) = \nu([0, x] \times (-\infty, y]) .$$

Assumption 1 also implies that  $F$  is continuous and that the sequence of probability distribution functions  $F_n$  defined, for all positive  $x$  and real  $y$ , by

$$F_n(x, y) = \nu_n([0, x] \times (-\infty, y])$$

converges to  $F$  locally uniformly. Assumption 1 can also be interpreted as the weak convergence to  $F$  of the vector  $(X - b(n))/\psi \circ b(n), (Y - m \circ b(n))/a \circ b(n)$  conditionnally on  $X > b(n)$ , i.e. for all bounded continuous function  $h$  on  $[0, \infty) \times (-\infty, \infty)$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[h\left(\frac{X - b(n)}{\psi \circ b(n)}, \frac{Y - m \circ b(n)}{a \circ b(n)}\right) \mid X > b(n)\right] = \int_0^\infty \int_{-\infty}^\infty h(x, y) F(dx, dy) . \quad (2)$$

*Remark 1.* All results concerning only the marginal distribution of  $X$  are obtained by applying the usual extreme value theory. In particular, the functions  $\psi$  and  $b$  are determined by the marginal distribution of  $X$  only. The function  $b$  can and will be chosen as  $b = (1/(1 - F_X))^\leftarrow$  where  $F_X$  is the distribution function of  $X$ . The function  $\psi$  satisfies

$$\lim_{x \rightarrow +\infty} \frac{\psi(x + \psi(x)u)}{\psi(x)} = 1 + \gamma u . \quad (3)$$

See [\(Resnick, 1987, Propositions 1.4 and 1.11\)](#). For any  $x > -1/\gamma$ , it holds that

$$\nu([x, \infty] \times [-\infty, \infty]) = (1 + \gamma x)^{-1/\gamma} ,$$

with the usual convention that this expression must be read as  $e^{-x}$  when  $\gamma = 0$ .

*Remark 2.* Assumption 1 has little implications on the functions  $a$  and  $m$  and on the distribution  $\Psi$  defined by

$$\Psi(z) = \int_0^\infty \int_{-\infty}^z \nu(dx, dy) .$$

If  $Y$  is independent of  $X$ , then  $\Psi$  is the distribution of  $Y$ ,  $a \equiv 1$  and  $m \equiv 0$ . Thus  $\Psi$  can be any probability distribution. In particular, it is not necessarily an extreme value distribution.

*Remark 3.* If the pair  $(X, Y)$  satisfies Assumption 1, then so does any affine transformation of  $(X, Y)$ . For instance, if  $X$  and  $Y$  have finite mean and variance, then  $((X - \mathbb{E}[X])/\text{var}^{1/2}(X), (Y - \mathbb{E}[Y])/\text{var}^{1/2}(Y))$  also satisfies Assumption 1. But non linear transformations of  $(X, Y)$  do not necessarily satisfy the assumption. In particular, the usual (in extreme value theory) transformation of  $X$  and  $Y$  to random variables with prescribed marginal distributions, is not always possible, as investigated in (Heffernan and Resnick, 2007, Section 7). It is never possible in the cases where the joint limiting distribution is a product measure. Consequently, we do not make any specific assumption on the marginal distributions of  $X$  and  $Y$ .

Obviously, the functions  $a$  and  $m$  are defined up to asymptotic equivalence, i.e. if  $m'$  and  $a'$  satisfy

$$\lim_{x \rightarrow \infty} \frac{a'(x)}{a(x)} = 1, \quad \lim_{x \rightarrow \infty} \frac{m(x) - m'(x)}{a(x)} = 0,$$

then the measure  $\nu'_n$  defined as  $\nu_n$  but with  $a'$  and  $m'$  instead of  $a$  and  $m$  converges vaguely to the same limit measure  $\nu$ . Beyond this trivial remark, the following result summarizes Heffernan and Resnick (2007, Propositions 1 and 2) and contains most of what can be inferred from Assumption 1. Recall that a function  $f$  defined on a neighborhood of infinity is said to be regularly varying if there exists a constant  $\alpha \in \mathbb{R}$  such that

$$\lim_{x \rightarrow \infty} \frac{f(tx)}{f(x)} = t^\alpha$$

for all  $t > 0$ . If  $\alpha = 0$ , the function is called slowly varying.

**Lemma 1.** *Under Assumption 1, there exists  $\zeta \in \mathbb{R}$  such that the function  $a \circ b$  is regularly varying at infinity with index  $\zeta$  and the function  $m$  satisfies*

$$\lim_{t \rightarrow \infty} \frac{m \circ b(tx) - m \circ b(t)}{a \circ b(t)} = J_\zeta(x),$$

with  $J_\zeta(x) = (x^\zeta - 1)/\zeta$  if  $\zeta \neq 0$  and  $J_0(x) = c \log(x)$  for some  $c \in \mathbb{R}$ , and the convergence is locally uniform on  $(0, \infty)$ .

For a sequence  $(X_i, Y_i)$ ,  $1 \leq i \leq n$ , let  $X_{(n:i)}$  denote the  $i$ -th order statistic and  $Y_{[n:i]}$  denote its concomitant, i.e.  $X_{(n:1)}, \dots, X_{(n:n)}$  is the ordering of  $X_1, \dots, X_n$  in increasing order, and  $Y_{[n:i]}$  is the  $Y$ -variable corresponding to  $X_{(n:i)}$ .

Recall that an intermediate sequence is a sequence of integers  $k_n$  such that  $\lim_{n \rightarrow \infty} k_n = \lim_{n \rightarrow \infty} n/k_n = \infty$ . In accordance with common use and for the clarity of notation, the dependence on  $n$  will be implicit in the sequel.

Define the random measure

$$\tilde{\nu}_n = \frac{1}{k} \sum_{i=1}^n \delta_{(\{X_i - b(n/k)\}/\psi \circ b(n/k), \{Y_i - m \circ b(n/k)\}/a \circ b(n/k))}. \quad (4)$$

Applying Resnick (1986, Proposition 5.3) (see also Resnick (1987, Exercise 3.5.7)), we straightforwardly obtain the following result.

**Proposition 2.** *If Assumption 1 holds, then for any intermediate sequence  $k$ ,  $\tilde{\nu}_n$  converges weakly to  $\nu$  locally uniformly on  $(-1/\gamma, \infty] \times [-\infty, \infty]$ .*

Consequently,  $\tilde{\nu}_n([0, x] \times (-\infty, y])$  converges weakly locally uniformly to  $F(x, y)$ . But  $\tilde{\nu}_n$  is not an estimator, since it involves the unknown functions  $a$  and  $m$ . In order to define estimators of these functions, and of the distribution function  $F$ , we will need to prove convergence of integrals of unbounded functions with respect to the random measure  $\tilde{\nu}_n$ . Therefore we need to strengthen Assumption 1.

**Assumption 2.** *There exists  $p^* > 0$ ,  $q^* > 0$  such that for any  $\epsilon \in (0, 1/\gamma)$ ,*

$$\lim_{n \rightarrow \infty} \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} |x|^{p^*} |y|^{q^*} \nu_n(dx, dy) = \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} |x|^{p^*} |y|^{q^*} \nu(dx, dy). \quad (5)$$

Condition (5) can be seen as a strengthening of (1) and (2) in order to obtain the convergence of conditional moments. Under Assumption 2, for all  $0 < p \leq p^*$  and  $0 < q \leq q^*$ , it holds that

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[(X - t)^p | Y - m(t)|^q | X > t]}{\psi^p(t) a^q(t)} = \int_0^{\infty} \int_{-\infty}^{\infty} x^p |y|^q \nu(dx, dy). \quad (6)$$

For the reason mentioned in Remark 1, Assumption 1 implies the convergence (5) with  $q^* = 0$  and any  $p^* < 1/\gamma$ . In applications, it will be assumed that  $q^* \geq 2$ . The function  $a$  and the limiting measure  $\nu$  are defined up to a change of scale, thus, without loss of generality, we assume henceforth that

$$\int_0^{\infty} \int_{-\infty}^{\infty} y^2 \nu(dx, dy) = \int_{-\infty}^{\infty} y^2 \Psi(dy) = 1. \quad (7)$$

**Proposition 3.** *If Assumptions 1 and 2 hold, then for any intermediate sequence  $k$  and any continuous function  $g$  such that  $|g(x, y)| \leq C(|x| \vee 1)^{p^*} (|y| \vee 1)^{q^*}$ , for any  $\epsilon \in (0, 1/\gamma)$ ,*

$$\int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} g(x, y) \tilde{\nu}_n(dx, dy) \rightarrow_P \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} g(x, y) \nu(dx, dy). \quad (8)$$

For historical interest, we can also mention the following consequence of Assumption 1, first stated in Eddy and Gale (1981, Theorem 6.1) in a restricted case of spherical distributions.

**Proposition 4.** *Under Assumption 1,  $\{Y_{[n:n]} - m \circ b(n/k)\} / a \circ b(n/k)$  converges weakly to  $\Psi$ . If moreover  $\nu$  is a product measure, then  $\{Y_{[n:n]} - m \circ b(n/k)\} / a \circ b(n/k)$  is asymptotically independent of  $X_{(n:n)}$ .*

Let us finally mention that Davydov and Egorov (2000) obtained functional limit theorems for sums of concomitants corresponding to a number  $k$  of order statistics such that  $k/n \rightarrow 0$ . Their problem differs from ours. Their assumptions on the joint distribution of the random pairs are much weaker than Assumption 1, but their results are of a very different nature and it does not seem possible to use them to derive Propositions 2-3 for instance.

### 3 Nonparametric estimation of $\psi$ , $a$ , $m$ and $F$

In this section, we introduce nonparametric estimators of the functions  $\psi$ ,  $m$ ,  $a$  and  $F$  based on i.i.d. observations  $(X_1, Y_1), \dots, (X_n, Y_n)$  of a bivariate distribution which satisfies Assumption 2.

#### 3.1 Definitions and consistency

In order to estimate nonparametrically the limiting distribution  $F$ , we first need nonparametric estimators of the quantities  $\psi(X_{(n:n-k)})$ ,  $m(X_{(n:n-k)})$  and  $a(X_{(n:n-k)})$ , with  $k$  an intermediate sequence, i.e. such that  $k \rightarrow \infty$  and  $k/n \rightarrow 0$ . The estimation of  $\psi(X_{(n:n-k)})$  is a well known estimation issue, see e.g. De Haan and Ferreira (2006, Section 4.2). If the extreme value index  $\gamma$  of  $X$  is less than 1, then  $\psi$  can be estimated as the mean residual life. Let  $\hat{\gamma}$  be a consistent estimator of  $\gamma$  (see e.g. De Haan and Ferreira (2006, Chapter 3) or Beirlant et al. (2004, Chapter 5)) and define

$$\hat{\psi}(X_{(n:n-k)}) = \frac{1 - \hat{\gamma}}{k} \sum_{i=1}^k \{X_{(n:n-i+1)} - X_{(n:n-k)}\}. \quad (9)$$

It follows straightforwardly from Proposition 3 that  $\hat{\psi}(X_{(n:n-k)})/\psi \circ b(n/k) \rightarrow_P 1$ . If it is moreover assumed (as in Section 4 below) that  $\gamma = 0$ , then the above estimator can be modified accordingly:

$$\hat{\psi}(X_{(n:n-k)}) = \frac{1}{k} \sum_{i=1}^k \{X_{(n:n-i+1)} - X_{(n:n-k)}\}. \quad (10)$$

In order to estimate  $m$ , define

$$\hat{m}(X_{(n:n-k)}) = \frac{\sum_{i=1}^k Y_{[n:n-i+1]} \{X_{(n:n-i+1)} - X_{(n:n-k)}\}}{\sum_{i=1}^k \{X_{(n:n-i+1)} - X_{(n:n-k)}\}}. \quad (11)$$

**Proposition 5.** *If Assumption 1 holds and Assumption 2 holds with  $p^* \geq 1$  and  $q^* \geq 1$ , then, for any intermediate sequence  $k$ , it holds that*

$$\frac{\hat{m}(X_{(n:n-k)}) - m \circ b(n/k)}{a \circ b(n/k)} \rightarrow_P \mu,$$

where  $\mu = (1 - \gamma) \int_0^\infty \int_{-\infty}^\infty xy \nu(dx, dy)$ . If moreover  $m(x) = \rho x$  and either  $\mu = 0$  and  $a(x) = O(x)$  or  $a(x) = o(x)$  then  $\hat{m}(X_{(n:n-k)})/X_{(n:n-k)}$  is a consistent estimator of  $\rho$ .

*Remark 4.* A sufficient condition for  $\mu = 0$  is the symmetry of the measure  $\nu$  with respect to the second variable. This happens in particular if  $\nu$  is a product measure, and the distribution  $\Psi$  is symmetric.

We now estimate  $a(X_{(n:n-k)})$ . Many estimators can be defined, each needing an ad hoc moment assumption. The one we have chosen needs  $q^* \geq 2$  in Assumption 2. Define

$$\hat{a}(X_{(n:n-k)}) = \left\{ \frac{1}{k} \sum_{i=1}^k \{Y_{[n:n-i+1]} - \hat{m}(X_{(n:n-k)})\}^2 \right\}^{1/2}. \quad (12)$$

**Proposition 6.** *If Assumption 1 holds and Assumption 2 holds with  $p^* \geq 1$  and  $q^* \geq 2$ , and if  $\mu = 0$ , then, for any intermediate sequence  $k$ , it holds that*

$$\hat{a}(X_{(n:n-k)})/a \circ b(n/k) \rightarrow_P 1.$$

*Remark 5.* If  $\mu \neq 0$ , then  $\hat{a}(X_{(n:n-k)})/a \circ b(n/k) \rightarrow_P \tau$ , with

$$\tau^2 = 1 - 2\mu \int_{-\infty}^{\infty} y \Psi(dy) + \mu^2. \quad (13)$$

We can now consider the nonparametric estimator of the limiting joint distribution  $F$ . Define

$$\begin{aligned} \hat{F}(x, y) &= \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{X_{(n:n-i+1)} \leq X_{(n:n-k)} + \hat{\psi}(X_{(n:n-k)})x\}} \times \mathbb{1}_{\{Y_{[n:n-i+1]} \leq \hat{m}(X_{(n:n-k)}) + \hat{a}(X_{(n:n-k)})y\}}. \end{aligned} \quad (14)$$

Denote  $u_n = \hat{\psi}(X_{(n:n-k)})/\psi \circ b(n/k)$  and

$$v_n = \frac{\hat{a}(X_{(n:n-k)})}{a \circ b(n/k)}, \quad \xi_n = \frac{\hat{m}(X_{(n:n-k)}) - m \circ b(n/k)}{a \circ b(n/k)}. \quad (15)$$

Then

$$\hat{F}(x, y) = \tilde{v}_n([\tilde{x}_n, \tilde{x}_n + u_n x] \times (-\infty, \xi_n + v_n y]).$$

Thus Propositions 2, 5 and 6 easily yield the consistency of  $\hat{F}(x, y)$ , as stated in the following theorem.

**Theorem 7.** *Under Assumptions 1 and 2 with  $\gamma < 1$ ,  $p^* \geq 1$  and  $q^* \geq 2$ , if  $\mu = 0$ , then for any intermediate sequence  $k$ ,  $\hat{F}(x, y)$  converges weakly to  $F(x, y)$ .*

We can also define an estimator of the second marginal  $\Psi$  of  $F$ . Denote

$$\hat{\Psi}(y) = \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{Y_{[n:n-i+1]} \leq \hat{m}(X_{(n:n-k)}) + \hat{a}(X_{(n:n-k)})y\}}. \quad (16)$$

Then, under the assumptions of Theorem 7,  $\hat{\Psi}$  also converges to  $\Psi$ . Note that if  $\mu \neq 0$ , then  $\hat{\Psi}(z)$  converges weakly to  $\Psi(\mu + \tau z)$ , with  $\tau$  defined in (13).

### 3.2 Central limit theorems

In order to obtain central limit theorems, we need to strengthen Assumptions 1 and 2.

**Assumption 3.** *There exist positive real numbers  $p^\dagger$  and  $q^\dagger$ , a function  $c$  such that  $\lim_{t \rightarrow \infty} c(t) = 0$  and a Radon measure  $\mu^\dagger$  on  $(-1/\gamma, \infty) \times (-\infty, \infty)$  such that for any  $\epsilon \in (0, 1/\gamma)$ , and any measurable function  $h$  such that  $|h(x, y)| \leq (|x| \vee 1)^{p^\dagger} (|y| \vee 1)^{q^\dagger}$ , it holds that*

$$\int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| \mu^\dagger(dx, dy) < \infty,$$



and

$$\left| \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} h(x, y) \nu_n(dx, dy) - \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} h(x, y) \nu(dx, dy) \right| \leq c \circ b(n) \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| \mu^\dagger(dx, dy). \quad (17)$$

*Remark 6.* Taking  $h = \mathbb{1}_{[0, x] \times (-\infty, y]}$ , (17) yields

$$|F_n(x, y) - F(x, y)| \leq c \circ b(n) \mu^\dagger([0, x] \times (-\infty, y]) \quad (18)$$

where  $F_n(x, y) = \nu_n([0, x] \times (-\infty, y])$ . This is a classical second order condition (see e.g. [de Haan and Resnick \(1993, Condition 4.1\)](#)), which gives a non uniform rate of convergence in Condition (1). The condition (17) is stronger than (18) in the sense that it moreover gives a rate of convergence for conditional moments.

For a sequence  $k$  depending on  $n$ , define the random measure  $\tilde{\mu}_n$  by

$$\tilde{\mu}_n = k^{1/2} (\tilde{\nu}_n - \nu)$$

and denote

$$W_n(x, y) = \tilde{\mu}_n((x, \infty) \times (-\infty, y]).$$

The next results states the functional convergence of  $W_n$  in the space  $\mathcal{D}((-1/\gamma, \infty) \times (-\infty, \infty))$  of right-continuous and left-limited functions, endowed with Skorohod's  $J_1$  topology.

**Proposition 8.** *If Assumption 3 holds with  $p^\dagger \geq 2$  and  $q^\dagger \geq 4$  and if the sequence  $k$  is chosen such that*

$$\lim_{n \rightarrow \infty} k^{1/2} c \circ b(n/k) = 0, \quad (19)$$

*then  $k$  is an intermediate sequence and the sequence of processes  $W_n$  converges weakly in  $\mathcal{D}((-1/\gamma, \infty) \times (-\infty, \infty))$  to a Gaussian process  $W$  with autocovariance function*

$$\text{cov}(W(x, y), W(x', y')) = \nu([x \vee x', +\infty) \times [-\infty, y \wedge y']). \quad (20)$$

*Moreover, the sequence of random measures  $\tilde{\mu}_n$  converges weakly (in the sense of finite dimensional distributions) to an independently scattered Gaussian random measure  $W$  with control measure  $\nu$  on the space of measurable functions  $g$  such that  $|g(x, y)|^2 \leq C(x \vee 1)^{p^\dagger} (|y| \vee 1)^{q^\dagger}$ , i.e.  $W(g)$  is a centered Gaussian random variable with variance*

$$\int_{-1/\gamma}^{\infty} \int_{-\infty}^{\infty} g^2(s, t) \nu(ds, dt)$$

*and  $W(g), W(h)$  are independent if  $\int gh d\nu = 0$ .*

The proof is in section 7. Applying Proposition 8, we easily obtain the following corollary. For  $i, j \geq 0$ , denote  $g_{i,j}(x, y) = x^i y^j \mathbb{1}_{\{x > 0\}}$ .

**Corollary 9.** *Under the assumptions of Proposition 8 and if moreover  $\mu = 0$ , then*

$$k^{1/2} \left\{ \frac{X_{(n:n-k)} - b(n/k)}{\psi \circ b(n/k)}, \frac{\hat{m}(X_{(n:n-k)}) - m \circ b(n/k)}{a \circ b(n/k)}, \frac{\hat{a}(X_{(n:n-k)})}{a \circ b(n/k)} - 1 \right\}$$

*converges jointly with  $k^{1/2}(\tilde{\nu}_n - \nu)$  to a Gaussian vector which can be expressed as*

$$(W(g_{0,0}), (1 - \gamma)W(g_{1,1}), \frac{1}{2}W(g_{0,2})) .$$

Proposition 8 and Corollary 9 straightforwardly yield a functional central limit theorem for the estimator  $\hat{\Psi}$  of  $\Psi$  defined in (16). Recall that  $F(x, y) = \nu([0, x] \times (-\infty, y])$ .

**Theorem 10.** *If Assumption 3 holds with  $p^\dagger \geq 2$  and  $q^\dagger \geq 4$ , if  $\mu = 0$ , if  $F$  (and hence  $\Psi$ ) is differentiable and if the intermediate sequence  $k$  satisfies (19), then  $k^{1/2}(\hat{\Psi} - \Psi)$  converges in  $\mathcal{D}((-\infty, +\infty))$  to the process  $M$  defined by*

$$M(y) = W(0, y) - \frac{\partial F}{\partial x}(0, y)W(g_{0,0}) + \Psi'(y)\{(1 - \gamma)W(g_{1,1}) + \frac{1}{2}W(g_{0,2})y\} . \quad (21)$$

We prove Theorem 10 here in order to explain the last two terms in the right hand side of (21).

*Proof of Theorem 10.* Recall the definitions of  $v_n$  and  $\xi_n$  in (15) and define

$$\tilde{x}_n = \frac{X_{(n:n-k)} - b(n/k)}{\psi \circ b(n/k)} . \quad (22)$$

Then

$$\begin{aligned} k^{1/2}\{\hat{\Psi}(y) - \Psi(y)\} &= k^{1/2}\{\tilde{\nu}_n([\tilde{x}_n, \infty) \times (-\infty, \xi_n + v_n y]) - \Psi(y)\} \\ &= \tilde{\mu}_n([\tilde{x}_n, \infty) \times (-\infty, \xi_n + v_n y]) \end{aligned} \quad (23)$$

$$+ k^{1/2}\{\nu([\tilde{x}_n, \infty) \times (-\infty, \xi_n + v_n y]) - \Psi(y)\} . \quad (24)$$

By Proposition 8, the term in (23) converges weakly to  $W(0, y)$ . By Corollary 9 and the delta method, the term in (24) converges weakly to

$$-\frac{\partial F}{\partial x}(0, y)W(g_{0,0}) + \Psi'(y)\{(1 - \gamma)W(g_{1,1}) + \frac{1}{2}W(g_{0,2})y\} .$$

□

## 4 Case of a product measure

In this section, guided by examples (see e.g. [Fougères and Soulier \(2010\)](#)), we make the following additional assumption.

**Assumption 4.** *The function  $\psi$  is an auxiliary function satisfying  $\lim_{x \rightarrow \infty} \psi(x)/x = 0$ , there exists  $\rho \in \mathbb{R}$  such that  $m(x) = \rho x$  and the measure  $\nu$  is of the form*

$$\nu([x, \infty) \times (-\infty, y]) = e^{-x}\Psi(y) , \quad (25)$$

*where  $\Psi$  is a distribution function on  $\mathbb{R}$ .*

The assumption  $m(x) = \rho x$  is satisfied by most known examples. Cf. [Fougères and Soulier \(2010\)](#) for a review of models satisfying these assumptions. It is possible to have  $m(x) = \rho x$  even when  $\nu$  is not a product measure, as in the case of elliptical distributions with regularly varying tails, cf. [Abdous et al. \(2005\)](#).

The condition  $\lim_{x \rightarrow \infty} \psi(x)/x = 0$  implies that the extreme value index of  $X$  is 0 (cf. [Resnick \(1987, Lemma 1.2\)](#)). We now recall the necessary and sufficient condition for  $\nu$  to be a product measure proved by [Heffernan and Resnick \(2007, Proposition 2\)](#).

**Lemma 11.** *The measure  $\nu$  is a product measure if and only if  $a \circ b$  is slowly varying at infinity and*

$$\lim_{t \rightarrow \infty} \frac{b(tx) - b(t)}{a \circ b(t)} = 0. \quad (26)$$

The main consequence of Assumption 4 and of Lemma 11 is that

$$\psi(x) = o(a(x)),$$

(by application of [De Haan and Ferreira \(2006, Theorem B.2.21\)](#)) and this implies that given  $X > t$ ,  $(X - t)/a(t)$  converges in probability to zero. We thus have the following Corollary.

**Corollary 12.** *If Assumptions 1 and 4 hold then, for all  $x \geq 0$  and  $y \in (-\infty, \infty)$ ,*

$$\lim_{t \rightarrow \infty} \mathbb{P}(X \leq t + \psi(t)x, Y - \rho X \leq a(t)y \mid X > t) = (1 - e^{-x})\Psi(y).$$

Define the measure  $\nu_n^\ddagger$  on  $(-1/\gamma, +\infty) \times [-\infty, +\infty]$  by

$$\nu_n^\ddagger(\cdot) = n\mathbb{P}\left(\left\{\frac{X - b(n)}{\psi \circ b(n)}, \frac{Y - \rho X}{a \circ b(n)}\right\} \in \cdot\right). \quad (27)$$

Then  $\nu_n^\ddagger$  converges vaguely on  $(-\infty, +\infty) \times [-\infty, +\infty]$  to  $\nu$ .

## 4.1 Nonparametric estimation

Under Assumption 4, we can define new estimators of  $\rho$ ,  $a$  and the marginal distribution  $\Psi$  as follows:

$$\hat{\rho} = \frac{\sum_{i=1}^k Y_{[n:n-i+1]} \{X_{(n:n-i+1)} - X_{(n:n-k)}\}}{\sum_{i=1}^k X_{(n:n-i+1)} \{X_{(n:n-i+1)} - X_{(n:n-k)}\}}, \quad (28)$$

$$\check{a}(X_{(n:n-k)}) = \left[ \frac{1}{k} \sum_{i=1}^k \{Y_{[n:n-i+1]} - \hat{\rho} X_{(n:n-i+1)}\}^2 \right]^{1/2}, \quad (29)$$

$$\check{\Psi}(z) = \frac{1}{k} \sum_{i=1}^k \mathbb{1}_{\{Y_{[n:n-i+1]} \leq \hat{\rho} X_{(n:n-i+1)} + \check{a}(X_{(n:n-k)})z\}}. \quad (30)$$

**Theorem 13.** *If Assumptions 1, 2 (with  $p^* = 1$  and  $q^* = 2$ ) and 4 hold and if  $\mu = 0$ , then for any intermediate sequence  $k$ ,  $b(n/k)(\hat{\rho} - \rho)/a \circ b(n/k)$  converges weakly to 0,  $\check{a}(X_{(n:n-k)})/a \circ b(n/k)$  converges weakly to 1 and  $\check{\Psi}$  is a consistent estimator of  $\Psi$ . If moreover  $a(x) = o(x)$  then  $\hat{\rho}$  converges weakly to  $\rho$ .*

The proof of Theorem 13 is along the lines of the proof of Propositions 5, 6 and Theorem 7. The only difference is that instead of the random measure  $\tilde{\nu}_n$  defined in (4) we use the measure  $\check{\nu}_n$ , defined by

$$\check{\nu}_n = \frac{1}{k} \sum_{i=1}^n \delta_{(\{X_i - b(n/k)\}/\psi \circ b(n/k), \{Y_i - \rho X_i\}/a \circ b(n/k))}, \quad (31)$$

which converges weakly to the measure  $\nu$  for any intermediate sequence  $k$ , as a consequence of Corollary 12 and Resnick (1986, Proposition 5.3). The details are omitted.

In order to prove central limit theorems, we now introduce a second order assumption which is a modification of Assumption 3 that accounts for the random centering. Recall the measure  $\nu_n^\ddagger$  defined in (27).

**Assumption 5.** *There exist positive real numbers  $p^\ddagger$  and  $q^\ddagger$ , a function  $\tilde{c}$  such that  $\lim_{t \rightarrow \infty} \tilde{c}(t) = 0$  and a Radon measure  $\mu^\ddagger$  on  $(-1/\gamma, \infty) \times (-\infty, \infty)$  such that for any  $\epsilon \in (0, 1/\gamma)$ , and any measurable function  $h$  such that  $|h(x, y)| \leq (|x| \vee 1)^{p^\ddagger} (|y| \vee 1)^{q^\ddagger}$ , it holds that*

$$\int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| \mu^\ddagger(dx, dy) < \infty,$$

and

$$\left| \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} h(x, y) \nu_n^\ddagger(dx, dy) - \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} h(x, y) \nu(dx, dy) \right| \leq \tilde{c} \circ b(n) \int_{-\epsilon}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| \mu^\ddagger(dx, dy). \quad (32)$$

The difference with Assumption 3 is the presence of measure  $\nu_n^\ddagger$  instead of  $\nu_n$ . It can be shown that Assumptions 3 and 4 with a smoothness assumption on  $\Psi$  imply Assumption 5, but with the same rate function  $c$  as in Assumption 3, whereas in some cases Assumption 5 can be proved directly with a function  $\tilde{c}$  which goes to zero at infinity faster than  $c$ . The following results could be stated under Assumption 3, but the interest of Assumption 5 is to take into account the possibility of faster rates of convergence of the estimators than those allowed by Assumption 3.

As an example, consider the case of a bivariate Gaussian vector with standard marginals and correlation  $\rho$ . Abdous et al. (2005) have shown that  $\lim_{x \rightarrow \infty} \mathbb{P}(Y \leq \rho x + \sqrt{1 - \rho^2} y \mid X > x) = \Phi(y)$  (where  $\Phi$  is the distribution function of the standard Gaussian law), and a rate of convergence of order  $x^{-1}$  has been proved in Abdous et al. (2008). But of course, since  $(Y - \rho X)/\sqrt{1 - \rho^2}$  is standard Gaussian and independent of  $X$ , for all  $x$  it holds that  $\mathbb{P}(Y \leq \rho X + \sqrt{1 - \rho^2} y \mid X > x) = \Phi(y)$ . For general elliptical bivariate random vectors, it is also proved in Abdous et al. (2008) that the rate of convergence with random centering can be the square of the rate with deterministic centering. Assumption 5 can also be checked for the generalized elliptical distributions studied in Fougères and Soulier (2010).

We can now state central limit theorems for  $\check{a}(X_{(n:n-k)})$ ,  $\hat{\rho}$  and  $\hat{\Psi}$  which parallels Corollary 9 and Theorem 10. The proof is also omitted.

**Theorem 14.** *If Assumptions 1, 4 and 5 hold with  $p^\dagger \geq 2$  and  $q^\dagger \geq 4$ , if  $\Psi$  is differentiable and if  $\mu = 0$  and if the intermediate sequence  $k$  is chosen such that*

$$\lim_{n \rightarrow \infty} k^{1/2} \tilde{c} \circ b(n/k) = 0, \quad (33)$$

then  $k^{1/2}\{\check{\Psi} - \Psi\}$  converges weakly in  $\mathcal{D}((-\infty, \infty))$  to the process  $M$  defined in (21) and

$$k^{1/2} \left( \frac{b(n/k)(\hat{\rho} - \rho)}{a \circ b(n/k)}, \frac{\check{a}(X_{(n:n-k)})}{a \circ b(n/k)} - 1 \right)$$

converges jointly with  $k^{1/2}(\check{\Psi} - \Psi)$  to the Gaussian vector  $(W(g_{1,1}), W(g_{0,2}))$ .

*Remark 7.* As mentioned above, if we only assume Assumption 3 instead of Assumption 5 and (33) with  $c$  instead of  $\tilde{c}$  then the conclusion of the theorem still holds.

### Kolmogorov-Smirnov Test

In the case  $\gamma = 0$  and when the limiting measure  $\nu$  has product form, then  $\frac{\partial}{\partial x} F(0, y) = \Psi(y)$ . Define  $B(t) = W(0, \Psi^{-1}(t))$ . Then  $B$  is a standard Brownian motion on  $[0, 1]$  and

$$W(0, y) - \frac{\partial}{\partial x} F(0, y) W(g_{0,0}) = B \circ \Psi(y) - \Psi(y) B(1) = \mathcal{B} \circ \Psi(y)$$

where  $\mathcal{B}$  is a standard Brownian bridge. By the same change of variable,  $W(g_{0,2})$  can be represented as

$$V = \int_0^1 \{\Psi^{-1}(t)\}^2 dB(t).$$

Since  $\mu = 0$  and  $\int_{-\infty}^{\infty} y^2 \Psi(dy) = 1$ , it is easily seen that

$$\begin{aligned} \text{var}(W(g_{1,1})) &= 2, \quad \text{cov}(W(g_{0,0}), W(g_{0,1})) = 0, \\ \text{cov}(W(0, y), W(g_{1,1})) &= \int_{-\infty}^y z \Psi(dz) = \int_0^{\Psi(y)} \Psi^{-1}(u) du, \\ \text{cov}(W(g(1, 1)), W(g_{0,2})) &= \int_{-\infty}^{\infty} z^3 \Psi(dz) = \int_0^{\infty} \{\Psi^{-1}(u)\}^3 du. \end{aligned}$$

Thus,  $W(g_{1,1})$  can be represented as

$$U = \int_0^1 \Psi^{-1}(s) dB(s) + N,$$

where  $N$  is a standard Gaussian random variable independent of the Brownian motion  $B$ . Since all random variables involved are jointly Gaussian, this shows that  $M(y)$  has the same distribution as

$$\mathcal{B} \circ \Psi(y) + \Psi'(y) \left\{ U + \frac{1}{2} y V \right\}.$$

Finally, since  $\Psi$  is continuous,  $\sup_{y \in \mathbb{R}} |M(y)|$  has the same distribution as

$$Z = \sup_{t \in [0, 1]} \left| \mathcal{B}(t) + \Psi' \circ \Psi^{-1}(t) \left\{ U + \frac{1}{2} \Psi^{-1}(t) V \right\} \right|. \quad (34)$$

The extra terms come from the estimation of the functions  $a$  and  $m$ . If they were known, the limiting distribution would be the Brownian bridge as expected. Nevertheless, this distribution depends only on  $\Psi$ , so it can be used for a goodness-of-fit test. See Section 5.1 for a numerical illustration.

## 4.2 Semi-parametric estimation

Two problems arise in practice: the estimation of the conditional probability  $\theta(x, y) = \mathbb{P}(Y \leq y \mid X > x)$  and of the conditional quantile  $y = \theta^{\leftarrow}(x, p)$  for some fixed  $p \in (0, 1)$  and for some extreme  $x$ , i.e. beyond the range of the observations.

If  $x$  lies *within* the range of the observations, then  $\theta(x, y)$  can be estimated empirically by

$$\hat{\theta}_{\text{emp}}(x, y) = \frac{1}{k} \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq y\}} \mathbb{1}_{\{X_i > x\}},$$

for  $x = X_{(n:n-k)}$ . The most interesting situation for using the limit distributions that arise in Assumption 1 is when  $x$  is outside the range of the observations, so that an empirical estimate is no longer available. In such a situation, a semi-parametric approach will be needed to extrapolate the functions  $a(x)$ ,  $m(x)$  and  $\psi(x)$  for values  $x$  beyond  $X_{(n:n)}$ . This requires some modeling restrictions. We still assume that Assumption 4 holds and we assume moreover that there exists  $\sigma > 0$  such that

$$a(x) = \sigma \sqrt{x\psi(x)}. \quad (35)$$

We will also assume that the limiting distribution function  $\Psi$  in (25) is known. These assumptions hold in particular for bivariate elliptical distribution, see Abdous et al. (2008). There, and in many other examples,  $\Psi$  is the distribution function of the standard Gaussian law. See also Fougères and Soulier (2010). Assumption 4 and (35) imply that

$$\lim_{x \rightarrow \infty} \theta(x, \rho x + \sigma \sqrt{x\psi(x)}z) = \Psi(z), \quad (36)$$

so that  $\theta(x, y)$  can be approximated for  $x$  large enough by

$$\Psi \left( \frac{y - \rho x}{\sigma \sqrt{x\psi(x)}} \right).$$

Thus, in order to estimate  $\theta$ , we need a semi-parametric estimator of  $\psi$ . For this purpose, we make the following assumption on the marginal distribution of  $X$ .

**Assumption 6.** *The distribution function  $H$  of  $X$  satisfies*

$$1 - H(x) = e^{-x^\beta \{c + O(x^{\beta\eta})\}}$$

with  $\beta > 0$  and  $\eta < 0$ .

Under Assumption 6, an admissible auxiliary function is given by

$$\psi(x) = \frac{1}{c\beta} x^{1-\beta}. \quad (37)$$

Under (35), the normalizing function  $a$  is then

$$a(x) = \frac{\sigma}{\sqrt{c\beta}} x^{1-\beta/2}.$$

Let  $k$  and  $k_1$  be intermediate sequences. For the sake of clarity, in the sequel, we make explicit the dependence of the estimators with respect to  $k$  or  $k_1$ . Semi-parametric estimators of  $\beta$  and  $a(x)$  are given by

$$\hat{\beta}_k = \frac{\sum_{i=1}^k \log \log(n/i) - \log \log(n/k)}{\sum_{i=1}^k \log(X_{(n:n-i+1)}) - \log(X_{(n:n-k)})}, \quad (38)$$

$$\check{a}_{k_1}(x) = \check{a}_{k_1}(X_{(n:n-k_1)}) \left( \frac{x}{X_{(n:n-k_1)}} \right)^{1-\hat{\beta}_k/2}, \quad (39)$$

where  $\check{a}_{k_1}(X_{(n:n-k_1)})$  is the nonparametric estimator defined in (29).

**Proposition 15.** *If Assumption 6 holds, and if  $k$  is an intermediate sequence such that*

$$\lim_{n \rightarrow \infty} \log(k)/\log(n) = \lim_{n \rightarrow \infty} k \log^{2\eta}(n) = 0, \quad (40)$$

*then  $k^{1/2}(\hat{\beta}_k - \beta)$  converges weakly to the centered Gaussian distribution with variance  $\beta^{-2}$ . Suppose moreover that Assumptions 1, 4 and 5 hold with  $p^\dagger = 2$  and  $q^\dagger = 4$  and that  $\mu = 0$  and (35) holds. Let  $(x_n)$  be a sequence and  $k_1$  be an intermediate sequence such that*

$$\lim_{n \rightarrow \infty} k_1^{1/2} \tilde{c} \circ b(n/k_1) = 0 \quad (41)$$

$$\lim_{n \rightarrow \infty} k/k_1 = 0, \quad (42)$$

$$\lim_{n \rightarrow \infty} \log(b(n/k_1))/\log(x_n) = 1, \quad (43)$$

$$\lim_{n \rightarrow \infty} k^{-1/2} \log(x_n) = 0. \quad (44)$$

*Then  $\frac{k^{1/2}}{\log(x_n)} \left\{ \frac{\check{a}_{k_1}(x_n)}{a(x_n)} - 1 \right\}$  converges weakly to the centered Gaussian distribution with variance  $\beta^{-2}$ .*

*Remark 8.* By the arguments following Assumption 5, it can be seen that the conclusion of Proposition 15 still holds if Assumption 5 is replaced by Assumption 3 and  $\tilde{c}$  is replaced by  $c$  in (41).

The previous results lead to natural estimators of the conditional probability  $\theta(x, y) = \mathbb{P}(Y \leq y \mid X > x)$  and of the conditional quantile  $y = \theta^{\leftarrow}(x, p)$ . Define

$$\hat{\theta}(x, y) = \Psi \left( \frac{y - \hat{\rho}x}{\check{a}_{k_1}(x)} \right). \quad (45)$$

Under Assumptions 1, 2, 4 and (35), Theorem 13 implies that for fixed  $x$  and  $y$ ,  $\hat{\theta}(x, y)$  is a consistent estimator of  $\Psi((y - \rho x)/a(x))$ , but a biased estimator of  $\theta(x, y)$ . The remaining bias, which is an approximation error due to the asymptotic nature of equation (36), can be

bounded thanks to the second order Assumption 5. For more details, see Abdous et al. (2008, Section 3.2) for a treatment in the elliptical case.

We now investigate more thoroughly the estimation of the conditional quantile  $y_n = \theta^-(x_n, p)$  for some fixed  $p \in (0, 1)$  and some extreme sequence  $x_n$ , i.e. beyond the range of the observations, or equivalently,  $x_n > b(n)$ . An estimator  $\hat{y}_n$  is defined by

$$\hat{y}_n = \hat{\rho}_{k_1} x_n + \check{a}_{k_1}(x_n) \Psi^{-1}(p), \quad (46)$$

where  $\hat{\rho}_{k_1}$  is the nonparametric estimator defined in (28).

**Corollary 16.** *Let the assumptions of Proposition 15 hold with Assumption 3 instead of Assumption 5 and  $c$  instead of  $\tilde{c}$  in (41),  $\Psi' \circ \Psi^{-1}(p) > 0$  and*

$$\lim_{n \rightarrow \infty} \frac{b(n/k_1)}{b(n)} = \lim_{n \rightarrow \infty} \frac{b(n/k_1)}{x_n} = 1.$$

(i) *If  $\Psi^{-1}(p) \neq 0$ , then*

$$\frac{k_1^{1/2} x_n}{\log(x_n) a(x_n)} \left\{ \frac{\hat{y}_n}{y_n} - 1 \right\}$$

*converges weakly to a centered Gaussian law with variance  $\{\Psi^{-1}(p)/\rho\beta\}^2$ .*

(ii) *If  $\Psi^{-1}(p) = 0$ , then*

$$\frac{k_1^{1/2} x_n}{a(x_n)} \left\{ \frac{\hat{y}_n}{y_n} - 1 \right\}$$

*converges weakly to a centered Gaussian law with variance 2.*

## 5 Numerical Illustration

In this section, we perform a small sample simulation study with two purposes. We analyze the behavior of the Kolmogorov-Smirnov test proposed in Section 4.1 and we illustrate the behavior of the estimator of the conditional quantile proposed in Section 4.2.

### 5.1 Goodness-of-fit test for the distribution $\Psi$

Assume that the hypotheses of Section 4 hold, so that the nonparametric estimation procedure described in Section 4.1 can be used. Three types of distributions are considered, each of them restricted to the positive quadrant for convenience. These distributions are:

(a) the elliptical distribution with radial survival function  $P(R > t) = e^{-t}$ , and Pearson correlation coefficient  $\rho = 0.5$  ;

(b) the distribution with radial representation  $R(\cos[(\pi/2 + \arcsin \rho)T - \arcsin \rho], \sin[(\pi/2 + \arcsin \rho)T])$ , where  $P(R > t) = e^{-t^2/2}$ ,  $T$  has a non uniform concave density function  $f_T(t) = 4/\{\pi + \pi(2t - 1)^2\}$  on  $[0, 1]$ , and  $\rho = 0.5$ ;

(c) the distribution with radial representation  $R(\cos[(\pi/2 + \arcsin \rho)T - \arcsin \rho], \sin[(\pi/2 + \arcsin \rho)T])$ , where  $P(R > t) = e^{-t^2/2}$ ,  $T$  has a non uniform convex density function  $f_T(t) = 2 - 4/\{\pi(1 + (2t - 1)^2)\}$  on  $[0, 1]$ , and  $\rho = 0.5$ .



Case **(a)** is an example of the standard elliptical case, for which estimation results already exist (see [Abdous et al. \(2008\)](#)), whereas **(b)** and **(c)** illustrate the situation where the density level lines are “asymptotically elliptic” (see [Fougères and Soulier \(2010\)](#)). In these three cases,  $\Psi$  is the Normal distribution function (denoted by  $\Phi$ ), and Assumption 6 is fulfilled with  $\beta = 2$ . Figure 1 illustrates the estimation of  $\Psi$  via the nonparametric estimator  $\check{\Psi}$  defined by (30) for one sample ( $n = 1000, k = 100$ ) of distribution **(b)**.

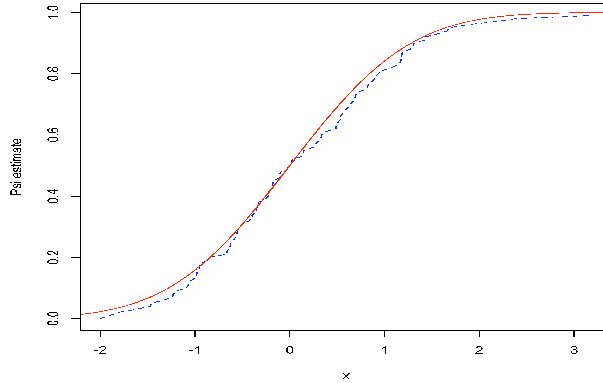


Figure 1: Estimation of  $\Psi$  via the nonparametric estimator  $\check{\Psi}_n$  for one sample ( $n = 1000, k = 100$ ) of distribution **(b)**.

The Kolmogorov-Smirnov goodness-of-fit test performed here admits therefore as test statistic

$$T_{KS} = \sup_{y \in \mathbb{R}} \sqrt{k} |\check{\Psi}(y) - \Phi(y)|. \quad (47)$$

As shown in Section 4.1,  $T_{KS}$  has asymptotically the same distribution as the random variable  $Z$  defined in (34). Quantiles of this distribution have been obtained numerically and are listed in Table 1.

Table 1: Quantiles  $q_\alpha$  of order  $1 - \alpha$  of  $Z$ .

$\alpha$	0.01	0.05	0.10	0.15	0.20	0.25
$q_\alpha$	1.598	1.297	1.174	1.076	1.029	0.980

We have compared these theoretical levels to the empirical levels obtained by simulation. In the three cases **(a)** to **(c)**, 1000 samples of size  $n = 10^3, 10^4$  and  $10^5$ , are simulated. The  $k$  observations having the largest first component are kept, for three different values of  $k$ , and the nonparametric estimate  $\check{\Psi}$  given in (30) is computed with this reduced sample. The observed values of the test statistic  $T_{KS}$  are compared to the quantiles listed in Table 1. For brevity, we present only the results corresponding to the two theoretical levels  $\alpha = (0.05, 0.1)$ . These empirical levels are shown in Table 2.

A common feature for the three distributions is that the results are rather sensitive to the reduced number of observations  $k$ . However, the value of  $k$  leading to the best adequation

Table 2: Empirical levels ( $\hat{\alpha}_{0.05}, \hat{\alpha}_{0.1}$ ) associated to theoretical levels (0.05, 0.1) for the goodness-of-fit test with statistic  $T_{KS}$ . The original sample size is denoted by  $n$ , and the number of observations used for the estimation is denoted by  $k$ . Notation **(a)**–**(c)** refers to the three bivariate distributions listed above. The boldface characters point out the best result in each case.

$n$	$k$	<b>(a)</b>	<b>(b)</b>	<b>(c)</b>
1000	50	<b>(0.053, 0.095)</b>	(0.031, 0.066)	(0.027, 0.050)
	100	(0.140, 0.231)	<b>(0.055, 0.102)</b>	<b>(0.04, 0.085)</b>
	150	(0.327, 0.453)	(0.071, 0.147)	(0.077, 0.153)
10000	50	(0.059, 0.095)	(0.03, 0.061)	(0.028, 0.045)
	100	<b>(0.052, 0.099)</b>	(0.038, 0.07)	<b>(0.038, 0.088)</b>
	200	(0.101, 0.183)	<b>(0.054, 0.096)</b>	(0.065, 0.125)
100000	100	<b>(0.051, 0.082)</b>	(0.037, 0.075)	<b>(0.044, 0.071)</b>
	200	(0.080, 0.133)	(0.041, 0.087)	(0.0795, 0.128)
	500	(0.140, 0.257)	<b>(0.05, 0.103)</b>	(0.20, 0.298)

between empirical and theoretical levels is rather stable in most cases studied ( $k = 100$  in two thirds of the cases).

## 5.2 Semi-parametric estimation of the conditional quantile function

Assume that Assumptions 1, 4, 6 and equation (35) hold and that the limiting distribution  $\Psi$  is the standard Gaussian distribution  $\Phi$  and  $\cdot$ . The small sample behavior of the semi-parametric estimator  $\hat{y}_n(p)$  of the quantile function  $\theta^\leftarrow(x_n, p)$  defined by Equation (46) is illustrated in Figure 2 for the three distributions presented in Section 5.1. In each case, 100 samples of size 10000 are simulated. A proportion of 1% of the observations is used, which are the 100 observations with largest first component. For each sample, the conditional quantile function  $\theta^\leftarrow(x, p)$  is estimated for two values of  $x$  corresponding to the theoretical  $X$ -quantiles of order  $1 - \epsilon$ , where  $\epsilon = 10^{-4}$  and  $\epsilon = 10^{-5}$ . Figure 2 summarizes the quality of these estimations by showing the median, and the 2.5%- and 97.5%-quantiles of  $\hat{y}_n(p)$  for the two fixed values of  $x$  specified above.

The estimation results are globally good, and the best ones are obtained for cases **(a)** and **(c)**, see rows 1 and 3 of Figure 2. Besides, one can observe a slight improvement as the conditioning event becomes more extreme.

These empirical interval confidence compare well with those obtained by applying the central limit theorem of Corollary 16. We do not show them on Figure 2 for the sake of clarity.

## 6 Data analysis

To illustrate the use of the new procedures, and more specifically the Kolmogorov-Smirnov goodness-of-fit test proposed in Section 4.1, the hypothesis of  $\Psi = \Phi$ , where  $\Phi$  is the standard

Table 3: Observed values  $t_{KS}$  of the test statistic  $T_{KS}$  defined by (47) in terms of the proportion  $r$  or number  $k$  of observations used.

$r$	0.05	0.10	0.15	0.20
$k$	22	45	68	91
$t_{KS}$	0.842	0.847	0.777	0.948

Gaussian cdf, is tested using the series of monthly returns for the 3M stock and the Dow Jones Industrial Average from January 1970 to January 2008 ( $n = 457$  values). These data were used by Levy and Duchin (2004) and revisited by Abdous et al. (2008). In the latter paper, the hypothesis of bivariate ellipticity was accepted through a test of elliptical symmetry proposed by Huffer and Park (2007) and the contagion from the Dow Jones to the 3M stock was tested. As shown in Abdous et al. (2005), ellipticity implies that Condition (1) holds and that the limiting distribution is the Gaussian law. The present procedure allows to test for the Gaussian conditional limit law without assuming ellipticity, but the weaker assumption (1). The observed values of the test statistic  $T_{KS}$  defined by (47) in terms of different choices of threshold  $k$  (or equivalently in terms of the proportion  $r$  of observations used,  $k = nr$ ) are summarized in Table 3. According to Table 1, all these observed values correspond to a  $p$ -value greater than 0.25, which leads to accept the hypothesis  $\Psi = \Phi$ .

## 7 Proofs

*Proof of Proposition 3.* By Proposition 2, the weak convergence of  $\tilde{\nu}_n$  to  $\nu$  implies that for any compact set  $K$  of  $(-1/\gamma, \infty) \times (-\infty, \infty)$  such that  $\nu(\partial K) = 0$  and any function  $h$ , it holds that

$$\lim_{n \rightarrow \infty} \iint_K h(x, y) \tilde{\nu}_n(dx, dy) = \iint_K h(x, y) \nu(dx, dy) \quad \text{in probability.}$$

For  $\epsilon, M > 0$ ,  $\epsilon < 1/\gamma$ , define  $K = [-\epsilon, M] \times [-M, M]$  and  $K^c = [-\epsilon, \infty) \times (-\infty, \infty) \setminus K$ . Let  $h$  be a nonnegative function on  $[-\epsilon, \infty) \times (-\infty, \infty)$  such that  $h(x, y) \leq C(|x| \vee 1)^{q^* - 1} (|y| \vee 1)^{p^* - 1}$ . We must prove that

$$\limsup_{M \rightarrow \infty} \lim_{n \rightarrow \infty} \iint_{K^c} h(x, y) \tilde{\nu}_n(dx, dy) = 0, \quad (48)$$

in probability. Since

$$\mathbb{E} \left[ \iint_{K^c} h(x, y) \tilde{\nu}_n(dx, dy) \right] = \iint_{K^c} h(x, y) \nu_{n/k}(x, y),$$

Assumption 2 implies that

$$\begin{aligned} & \lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E} \left[ \iint_{K^c} h(x, y) \tilde{\nu}_n(dx, dy) \right] \\ &= \lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \iint_{K^c} h(x, y) \nu_{n/k}(dx, dy) = \lim_{M \rightarrow \infty} \iint_{K^c} h(x, y) \nu(dx, dy). \end{aligned}$$

This yields (48) and concludes the proof of Proposition 3.  $\square$

*Proof of Proposition 5.* Write

$$\frac{\hat{m}(X_{(n:n-k)}) - m \circ b(n/k)}{a \circ b(n/k)} = \frac{S_n}{T_n},$$

with

$$S_n = \frac{1}{k} \sum_{i=1}^k \frac{Y_{[n:n-i+1]} - m \circ b(n/k)}{a \circ b(n/k)} \frac{X_{(n:n-i+1)} - X_{(n:n-k)}}{\psi \circ b(n/k)},$$

$$T_n = \frac{1}{k} \sum_{i=1}^k \frac{X_{(n:n-i+1)} - X_{(n:n-k)}}{\psi \circ b(n/k)}.$$

We have already seen that  $T_n$  converges weakly to  $1/(1-\gamma)$ . Recall that we have defined

$$\tilde{x}_n = \frac{X_{(n:n-k)} - b(n/k)}{\psi \circ b(n/k)}.$$

By definition of  $\tilde{\nu}_n$ , we have, (with  $x_+ = \sup(x, 0)$  for any real number  $x$ )

$$\begin{aligned} S_n &= \frac{1}{k} \sum_{i=1}^n \frac{Y_i - m \circ b(n/k)}{a(X_{(n:n-k)})} \left\{ \frac{X_i - b(n/k)}{\psi \circ b(n/k)} - \tilde{x}_n \right\}_+ = \int_{\tilde{x}_n}^{\infty} \int_{-\infty}^{\infty} (x - \tilde{x}_n) y \tilde{\nu}_n(dx, dy) \\ &= \int_0^{\infty} \int_{-\infty}^{\infty} xy \tilde{\nu}_n(dx, dy) - \int_0^{\tilde{x}_n} \int_{-\infty}^{\infty} xy \tilde{\nu}_n(dx, dy) - \tilde{x}_n \int_{\tilde{x}_n}^{\infty} \int_{-\infty}^{\infty} y \tilde{\nu}_n(dx, dy). \end{aligned} \quad (49)$$

By Proposition 3, the first term in (49) converges to  $\mu/(1-\gamma)$ . Under Assumption 1, it is well known that  $\tilde{x}_n = o_P(1)$ . Cf. De Haan and Ferreira (2006, Theorem 2.2.1). This and Assumption 2 imply that the last two terms in (49) are  $o_P(1)$ . Thus  $S_n$  converges weakly to  $\mu/(1-\gamma)$  by Proposition 3. If  $m(x) = \rho x$ , then  $\hat{\rho} - \rho \sim X_{(n:n-k)}^{-1} a(X_{(n:n-k)}) \mu$  which converges to 0 if  $a(x) = o(x)$  or if  $\mu = 0$  and  $a(x) = O(x)$ .  $\square$

*Proof of Proposition 6.* We show that  $\hat{a}^2(X_{(n:n-k)})/a^2 \circ b(n/k)$  converges weakly to 1. Recall that  $\xi_n = \{\hat{m}(X_{(n:n-k)}) - m \circ b(n/k)\}/a \circ b(n/k)$ . By Proposition 5,  $\xi_n = o_P(1)$ , and noting that  $\tilde{\nu}_n\{\tilde{x}_n, \infty\} \times [-\infty, \infty] = 1$ , where  $\tilde{\nu}_n$  and  $\tilde{x}_n$  are respectively defined by (4) and (22), we have

$$\begin{aligned} \frac{\hat{a}^2(X_{(n:n-k)})}{a^2 \circ b(n/k)} &= \frac{1}{k} \sum_{i=1}^n \left\{ \frac{Y_i - m \circ b(n/k)}{a \circ b(n/k)} - \xi_n \right\}^2 \mathbb{1}_{\left\{ \frac{X_i - b(n/k)}{\psi \circ b(n/k)} \geq \tilde{x}_n \right\}} \\ &= \int_{\tilde{x}_n}^{\infty} \int_{-\infty}^{\infty} (y - \xi_n)^2 \tilde{\nu}_n(dx, dy) \\ &= \int_{\tilde{x}_n}^{\infty} \int_{-\infty}^{\infty} y^2 \tilde{\nu}_n(dx, dy) - 2\xi_n \int_{\tilde{x}_n}^{\infty} \int_{-\infty}^{\infty} y \tilde{\nu}_n(dx, dy) + \xi_n^2 \\ &= \int_0^{\infty} \int_{-\infty}^{\infty} y^2 \tilde{\nu}_n(dx, dy) + o_P(1). \end{aligned}$$

Thus  $\hat{a}(X_{(n:n-k)})/a \circ b(n/k)$  converges weakly to 1 by Proposition 3 and equation (7).  $\square$

*Proof of Proposition 8.* We start by proving the convergence of the finite dimensional distributions of  $W_n$ . Denote  $G_n(x, y) = \nu_n((x, \infty) \times (-\infty, y])$ ,  $G(x, y) = \nu((x, \infty) \times (-\infty, y])$ ,  $\tilde{X}_i = \{X_i - b(n/k)\}/\psi \circ b(n/k)$ ,  $\tilde{Y}_i = \{Y_i - m \circ b(n/k)\}/a \circ b(n/k)$  and

$$\begin{aligned}\xi_{n,i}(x, y) &= k^{-1/2} \{ \mathbb{1}_{\{\tilde{X}_i > x, \tilde{Y}_i \leq y\}} - \mathbb{P}(\tilde{X}_i > x, \tilde{Y}_i \leq y) \} \\ &= k^{-1/2} \{ \mathbb{1}_{\{\tilde{X}_i > x, \tilde{Y}_i \leq y\}} - kn^{-1} G_{n/k}(x, y) \} .\end{aligned}$$

Then for each  $n$ , the random variables  $\xi_{n,i}$ ,  $1 \leq i \leq n$  are i.i.d.,

$$\text{cov}(\xi_{n,i}(x, y), \xi_{n,i}(x', y')) = \frac{1}{n} G_{n/k}(x \vee x', y \wedge y') - \frac{k}{n^2} G_{n/k}(x, y) G_{n/k}(x', y') ,$$

and

$$W_n(x, y) = \sum_{i=1}^n \xi_{n,i}(x, y) + k^{1/2} \{ G_{n/k}(x, y) - G(x, y) \} .$$

Assumption 3 and (19) imply that  $k^{1/2}(G_{n/k} - G)$  converges to zero locally uniformly. The Lindeberg central limit theorem (cf. Araujo and Giné (1980)) and (19) yield the convergence of finite dimensional distributions of  $\sum_{i=1}^n \xi_{n,i}(x, y)$  to the Gaussian process with autocovariance defined by (20). Tightness can be obtained as in Einmahl et al. (1993) by using an exponential inequality such as Inequality 1 in the aforementioned reference.

We now prove the second part of Proposition 8. Let  $h$  be a  $C^\infty$  function with compact support in  $(-1/\gamma, \infty) \times (-\infty, \infty)$ . The weak convergence of  $W_n$  in  $\mathcal{D}((-1/\gamma, \infty) \times (-\infty, \infty))$  implies that  $\iint h(x, y) W_n(x, y) dx dy$  converges weakly to  $\iint h(x, y) W(x, y) dx dy$ . Thus, by integration by parts, it also holds that  $\iint h(x, y) W_n(dx, dy)$  converges weakly to  $\iint h(x, y) W(dx, dy)$ . Let  $\epsilon \in (0, 1/\gamma)$  and define  $A = [-\epsilon, \infty) \times (-\infty, \infty)$ . Let  $g$  be a measurable function defined on  $A$  such that  $|g(x, y)|^2 \leq C(|x| \vee 1)^{p^\dagger} (|y| \vee 1)^{q^\dagger}$ . Then, for all  $\epsilon > 0$ , there exists a  $C^\infty$  function  $h$  with compact support in  $A$  such that  $\int_A (g - h)^2 d\nu \leq \epsilon$ . Then,

$$\int_A g d\tilde{\mu}_n = \int_A h d\tilde{\mu}_n + \int_A (g - h) d\tilde{\mu}_n .$$

The first term in the right hand side converges weakly to  $W(h)$  and we prove now that the second one converges in probability to 0. Denote  $u = g - h$  and

$$\mu_n = k^{1/2} \{ \nu_{n/k} - \nu \} .$$

Then,

$$\int_A u d\tilde{\mu}_n = k^{-1/2} \sum_{i=1}^n \{ u(\tilde{X}_i, \tilde{Y}_i) - \mathbb{E}[u(\tilde{X}_i, \tilde{Y}_i)] \} + \int_A u d\mu_n .$$

By definition, for any function  $v$ ,  $\mathbb{E}[v(\tilde{X}_1)] = kn^{-1} \int v d\nu_{n/k}$ , thus

$$\mathbb{E} \left[ \left( \int_A u d\tilde{\mu}_n \right)^2 \right] \leq \int_A u^2 d\nu_{n/k} + \left\{ \int_A u d\mu_n \right\}^2 .$$

By assumption on  $g$ , and since  $h$  has compact support, it also holds that  $u^2(x, y) \leq C(|x| \vee 1)^{p^\dagger}(|y| \vee 1)^{q^\dagger}$ . Thus, by Assumption 3 and (19), it holds that  $\lim_{n \rightarrow \infty} \int_A u \, d\mu_n = 0$  and  $\lim_{n \rightarrow \infty} \int_A u^2 \, d\nu_n = \int_A u^2 \, d\nu$ . Thus

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[ \left( \int_A u \, d\tilde{\mu}_n \right)^2 \right] \leq \int_A u^2 \, d\nu \leq \epsilon .$$

Taking into account that  $\text{var}(W(g) - W(h)) = \text{var}(W(g-h)) = \int_A (g-h)^2 \, d\nu \leq \epsilon$ , we conclude that  $W_n(g)$  converges weakly to  $W(g)$ .  $\square$

*Proof of Corollary 9.* We prove separately the claimed limit distributions. The joint convergence is obvious. We start with  $\tilde{x}_n$ , defined in (22). Denote  $\mathbb{G}_n(x) = \tilde{\nu}_n((x, \infty) \times (-\infty, +\infty))$ . By Proposition 8,  $k^{1/2}(\mathbb{G}_n - \bar{P}_\gamma)$  converges weakly in  $\mathcal{D}$  to the process  $B \circ \bar{P}_\gamma$ , where  $B$  is a standard Brownian motion on  $[0, 1]$ . By Vervaat's Lemma (De Haan and Ferreira, 2006, Lemma A.0.2),  $k^{1/2}\{\mathbb{G}_n^\leftarrow - \bar{P}_\gamma^\leftarrow\}$  jointly converges weakly in  $\mathcal{D}$  to  $-(\bar{P}_\gamma^\leftarrow)'B$ . Since  $\mathbb{G}_n^\leftarrow(1) = \tilde{x}_n$ ,  $\bar{P}_\gamma^\leftarrow(1) = 0$  and  $(\bar{P}_\gamma^\leftarrow)'(1) = -1$ , we get the claimed limit distribution for  $k^{1/2}\tilde{x}_n$ .

We now consider  $\xi_n$ , defined in (15). By definition,

$$\begin{aligned} \xi_n &= \frac{\sum_{i=1}^k \{X_{(n:n-i+1)} - X_{(n:n-k)}\} \{Y_{[n:n-i+1]} - m \circ b(n/k)\}}{k\psi \circ b(n/k) a \circ b(n/k)} \\ &\quad \div \frac{\sum_{i=1}^k \{X_{(n:n-i+1)} - X_{(n:n-k)}\}}{k\psi \circ b(n/k)} \\ &= \frac{\int_{\tilde{x}_n}^\infty \int_{-\infty}^\infty (x - \tilde{x}_n) y \tilde{\nu}_n(dx, dy)}{\int_{\tilde{x}_n}^\infty \int_{-\infty}^\infty (x - \tilde{x}_n) \tilde{\nu}_n(dx, dy)} . \end{aligned}$$

Since  $\mu = 0$  by assumption, we obtain

$$k^{1/2}\xi_n = \frac{\int_{\tilde{x}_n}^\infty \int_{-\infty}^\infty (x - \tilde{x}_n) y \tilde{\mu}_n(dx, dy)}{\int_{\tilde{x}_n}^\infty \int_{-\infty}^\infty (x - \tilde{x}_n) \tilde{\nu}_n(dx, dy)} .$$

Since  $\tilde{x}_n = O_P(k^{-1/2})$ , it is easily seen that

$$k^{1/2}\xi_n = \frac{\int_0^\infty \int_{-\infty}^\infty xy \tilde{\mu}_n(dx, dy) + o_P(1)}{\int_0^\infty \int_{-\infty}^\infty x \tilde{\nu}_n(dx, dy) + o_P(1)} .$$

Applying Propositions 3 and 8, we obtain that  $k^{1/2}\xi_n$  converges weakly to  $(1 - \gamma)W(g_{1,1})$ . Consider now  $\hat{a}(X_{(n:n-k)})$ . As in the proof of Proposition 6, we write

$$\frac{\hat{a}^2(X_{(n:n-k)})}{a^2 \circ b(n/k)} = \int_{\tilde{x}_n}^\infty \int_{-\infty}^\infty y^2 \tilde{\nu}_n(dx, dy) - 2\xi_n \int_{\tilde{x}_n}^\infty \int_{-\infty}^\infty y \tilde{\nu}_n(dx, dy) + \xi_n^2 ,$$

and since  $\tilde{x}_n = O_P(k^{-1/2})$  and  $\xi_n = O_P(k^{-1/2})$ , we get

$$k^{1/2} \left\{ \frac{\hat{a}^2(X_{(n:n-k)})}{a^2 \circ b(n/k)} - 1 \right\} = \int_0^\infty \int_{-\infty}^\infty y^2 \tilde{\mu}_n(dx, dy) + o_P(1) .$$

Proposition 8 and the delta method yield that  $k^{1/2}\{\hat{a}(X_{(n:n-k)})/a \circ b(n/k) - 1\}$  converges weakly to  $\frac{1}{2}W(g_{0,2})$ .  $\square$

*Proof of Proposition 15.* The asymptotic normality of  $\hat{\beta}_k$  is proved (under more general conditions) in [Gardes and Girard \(2006, Corollary 1\)](#). Consider now  $\check{a}_{k_1}(x_n)$ . By (35) and (37),

$$a(x) = a(X_{(n:n-k_1)}) \left( \frac{x}{X_{(n:n-k_1)}} \right)^{1-\beta/2},$$

thus, by (39), we obtain

$$\frac{\check{a}_{k_1}(x_n)}{a(x_n)} = \frac{\check{a}_{k_1}(X_{(n:n-k_1)})}{a(X_{(n:n-k_1)})} X_{(n:n-k_1)}^{(\hat{\beta}_k - \beta)/2} x_n^{(\beta - \hat{\beta}_k)/2}.$$

Decomposing further, we get

$$\frac{\check{a}_{k_1}(x_n)}{a(x_n)} - 1 = \left\{ \frac{\check{a}_{k_1}(X_{(n:n-k_1)})}{a(X_{(n:n-k_1)})} - 1 \right\} X_{(n:n-k_1)}^{(\hat{\beta}_k - \beta)/2} x_n^{(\beta - \hat{\beta}_k)/2} \quad (50)$$

$$+ \left\{ X_{(n:n-k_1)}^{(\hat{\beta}_k - \beta)/2} - 1 \right\} \left\{ x_n^{(\beta - \hat{\beta}_k)/2} - 1 \right\} \quad (51)$$

$$+ X_{(n:n-k_1)}^{(\hat{\beta}_k - \beta)/2} - 1 + x_n^{(\beta - \hat{\beta}_k)/2} - 1. \quad (52)$$

Since  $\hat{\beta}_k - \beta = O_P(k^{-1/2})$ ,  $\log(x_n) = o(k^{1/2})$  and  $k/k_1 \rightarrow 0$ , we obtain

$$\begin{aligned} x_n^{(\beta - \hat{\beta}_k)/2} - 1 &\sim (\beta - \hat{\beta}_k) \log(x_n)/2, \\ X_{(n:n-k_1)}^{(\beta - \hat{\beta}_k)/2} - 1 &\sim (\beta - \hat{\beta}_k) \log(X_{(n:n-k_1)})/2 \sim (\beta - \hat{\beta}_k) \log(b(n/k_1))/2, \end{aligned}$$

where the equivalence relations above hold in probability. Thus, by the first part of Proposition 15 and (43) the product in (51) is  $O_P(k^{-1} \log^2(x_n)) = o_P(k^{-1/2} \log(x_n))$  by (44). By Theorem 14,  $\check{a}_{k_1}(X_{(n:n-k_1)})/a(X_{(n:n-k_1)}) - 1 = O_P(k_1^{-1/2})$ , thus the term in the right hand side of (50) is  $O_P(k_1^{-1/2}) = o_P(k^{-1/2} \log(x_n))$  since  $k/k_1 \rightarrow 0$ . Altogether, these bounds yields,

$$\frac{k^{1/2}}{\log(x_n)} \left\{ \frac{\check{a}_{k_1}(x_n)}{a(x_n)} - 1 \right\} = k^{1/2}(\beta - \hat{\beta}_k) + o_P(1),$$

and the proof follows from the asymptotic normality of  $k^{1/2}(\beta - \hat{\beta}_k)$ .  $\square$

*Proof of Corollary 16.* Define  $\tilde{y}_n = \rho x_n + a(x_n)\Psi^{-1}(p)$ . Then

$$\begin{aligned} \hat{y}_n - y_n &= \hat{y}_n - \tilde{y}_n + \tilde{y}_n - y_n \\ &= (\hat{\rho}_{k_1} - \rho)x_n + (\check{a}_{k_1}(x_n) - a(x_n))\Psi^{-1}(p) + \tilde{y}_n - y_n. \end{aligned}$$

In order to study  $\tilde{y}_n - y_n$ , denote  $z_n = (y_n - \rho x_n)/a(x_n)$ . Then  $\lim_{n \rightarrow \infty} z_n = \Psi^{-1}(p)$ . Indeed, if the sequence  $z_n$  is unbounded, then it tends to infinity at least along a subsequence. Choose  $z > \Psi^{-1}(p)$ . Then, for large enough  $n$ ,

$$\begin{aligned} p &= \mathbb{P}(Y \leq \rho x_n + a(x_n)z_n \mid X > x_n) \geq \mathbb{P}(Y \leq \rho x_n + a(x_n)z \mid X > x_n) \\ &\rightarrow \Psi(z) > p. \end{aligned}$$

Thus the sequence  $z_n$  is bounded, and if it converges to  $z$  (along a subsequence), it necessarily holds that  $\Psi(z) = p$ , thus  $z_n$  converges to  $\Psi^{-1}(p)$ . Since we have assumed that  $a(x) = o(x)$ , this implies that  $y_n \sim \rho x_n$  and

$$\frac{\tilde{y}_n - y_n}{y_n} \sim \frac{a(x_n)\{\Psi^{-1}(p) - z_n\}}{\rho x_n} \rightarrow 0.$$

Moreover, since  $\Psi' \circ \Psi^{-1}(p) > 0$ , by a first order Taylor expansion, we have

$$\Psi^{-1}(p) - z_n = \frac{1}{\Psi'(\xi_n)} \{\theta(x_n, y_n) - \Psi(z_n)\},$$

where  $\xi_n = \Psi^{-1}(p) + u\{z_n - \Psi^{-1}(p)\}$  for some  $u \in (0, 1)$ . By Assumption 3,  $\|\theta(x_n, \rho x_n + a(x_n)\cdot) - \Psi\|_\infty = O(c \circ b(n))$ . Since we have already shown that  $z_n$  converges to  $\Psi^{-1}(p)$ ,  $1/\Psi'(\xi_n)$  is bounded for large enough  $n$ , so  $\Psi^{-1}(p) - z_n = O(c \circ b(n))$ . Thus, by (41) (with  $c$  instead of  $\tilde{c}$ ), we get

$$\frac{k^{1/2}x_n}{\log(x_n)a(x_n)} \frac{\tilde{y}_n - y_n}{y_n} = O\left(\frac{k^{1/2}c \circ b(n)}{\log(x_n)}\right) = o\left(\frac{k_1^{1/2}c \circ b(n)}{\log(x_n)}\right) = o(1).$$

Next, by definition, and since  $y_n \sim \rho x_n$  and  $a(x_n) = o(x_n)$ , we have

$$\frac{\hat{y}_n - \tilde{y}_n}{y_n} \sim \frac{\hat{\rho}_{k_1} - \rho}{\rho} + \frac{a(x_n)\Psi^{-1}(p)}{\rho x_n} \left\{ \frac{\check{a}_{k_1}(x_n)}{a(x_n)} - 1 \right\}.$$

Thus,

$$\frac{k^{1/2}}{\log(x_n)} \frac{x_n}{a(x_n)} \frac{\hat{y}_n - \tilde{y}_n}{y_n} \sim \frac{k^{1/2}x_n(\hat{\rho}_{k_1} - \rho)}{\rho a(x_n) \log(x_n)} + \frac{\Psi^{-1}(p)}{\rho} \frac{k^{1/2}}{\log(x_n)} \left\{ \frac{\check{a}_{k_1}(x_n)}{a(x_n)} - 1 \right\}.$$

The first term in the right-hand side tends to zero by Theorem 14 and the assumptions on the sequences  $k_1$ ,  $k$  and  $x_n$ . The second term converges weakly to a centered Gaussian law with variance  $\{\Psi^{-1}(p)/(\rho\beta)\}^2$  by Proposition 15. In the case  $\Psi^{-1}(p) = 0$ , the main term is the first one in the right-hand side of the last display, and we conclude by applying Theorem 14.  $\square$

## References

- Belkacem Abdous, Anne-Laure Fougères, and Kilani Ghoudi. Extreme behaviour for bivariate elliptical distributions. *Revue Canadienne de Statistique*, 33(2):1095–1107, 2005.
- Belkacem Abdous, Anne-Laure Fougères, Kilani Ghoudi, and Philippe Soulier. Estimation of bivariate excess probabilities for elliptical models. *Bernoulli*, 14(4):1065–1088, 2008.
- Aloisio Araujo and Evarist Giné. *The central limit theorem for real and Banach valued random variables*. John Wiley & Sons, New York-Chichester-Brisbane, 1980. Wiley Series in Probability and Mathematical Statistics.
- Guus Balkema and Paul Embrechts. *High risk scenarios and extremes. A geometric approach*. Zurich Lectures in Advanced Mathematics. Zürich: European Mathematical Society, 2007.



- Philippe Barbe. Approximation of integrals over asymptotic sets with applications to probability and statistics. <http://arxiv.org/abs/math/0312132>, 2003.
- Jan Beirlant, Yuri Goegebeur, Jozef Teugels, and Johan Segers. *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004.
- Simeon M. Berman. *Sojourns and extremes of stochastic processes*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1992.
- B. O. Bradley and M. Taqqu. Framework for analyzing spatial contagion between financial markets. *Finance Letters*, 2(6):8–16, 2004.
- Bikramjit Das and Sidney I. Resnick. Conditioning on an extreme component: Model consistency and regular variation on cones. <http://arxiv.org/abs/0805.4373>, 2008.
- Bikramjit Das and Sidney I. Resnick. Detecting a conditional extreme value model. <http://arxiv.org/abs/0902.2996>, 2009.
- Yu. Davydov and V. Egorov. Functional limit theorems for induced order statistics. *Mathematical Methods of Statistics*, 9(3):297–313, 2000.
- Laurens De Haan and Ana Ferreira. *Extreme value theory. An introduction*. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer, 2006.
- Laurens de Haan and Sidney I. Resnick. Estimating the limit distribution of multivariate extremes. *Communications in Statistics. Stochastic Models*, 9(2):275–309, 1993.
- William F. Eddy and James D. Gale. The convex hull of a spherically symmetric sample. *Advances in Applied Probability*, 13(4):751–763, 1981.
- John H. J. Einmahl, Laurens de Haan, and Xin Huang. Estimating a multidimensional extreme-value distribution. *Journal of Multivariate Analysis*, 47(1):35–47, 1993.
- Anne-Laure Fougères and Philippe Soulier. Estimation of conditional laws given an extreme component. To appear in *Stochastic models*, 26(1), 2010.
- Laurent Gardes and Stéphane Girard. Comparison of Weibull tail-coefficient estimators. *REVSTAT*, 4(2):163–188, 2006.
- Enkelejd Hashorva. Gaussian approximation of conditional elliptic random vectors. *Stoch. Models*, 22(3):441–457, 2006.
- Enkelejd Hashorva, Samuel Kotz, and Alfred Kume.  $L_p$ -norm generalised symmetrised Dirichlet distributions. *Albanian Journal of Mathematics*, 1(1):31–56 (electronic), 2007.
- Janet E. Heffernan and Sidney I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571, 2007.
- Janet E. Heffernan and Jonathan A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B*, 66(3):497–546, 2004.

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

Haim Levy and Ran Duchin. Asset return distributions and the investment horizon. *Journal of Portfolio Management*, 30(3):47–62, 2004.

Sidney I. Resnick. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, 18(1):66–138, 1986.

Sidney I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York, 1987.

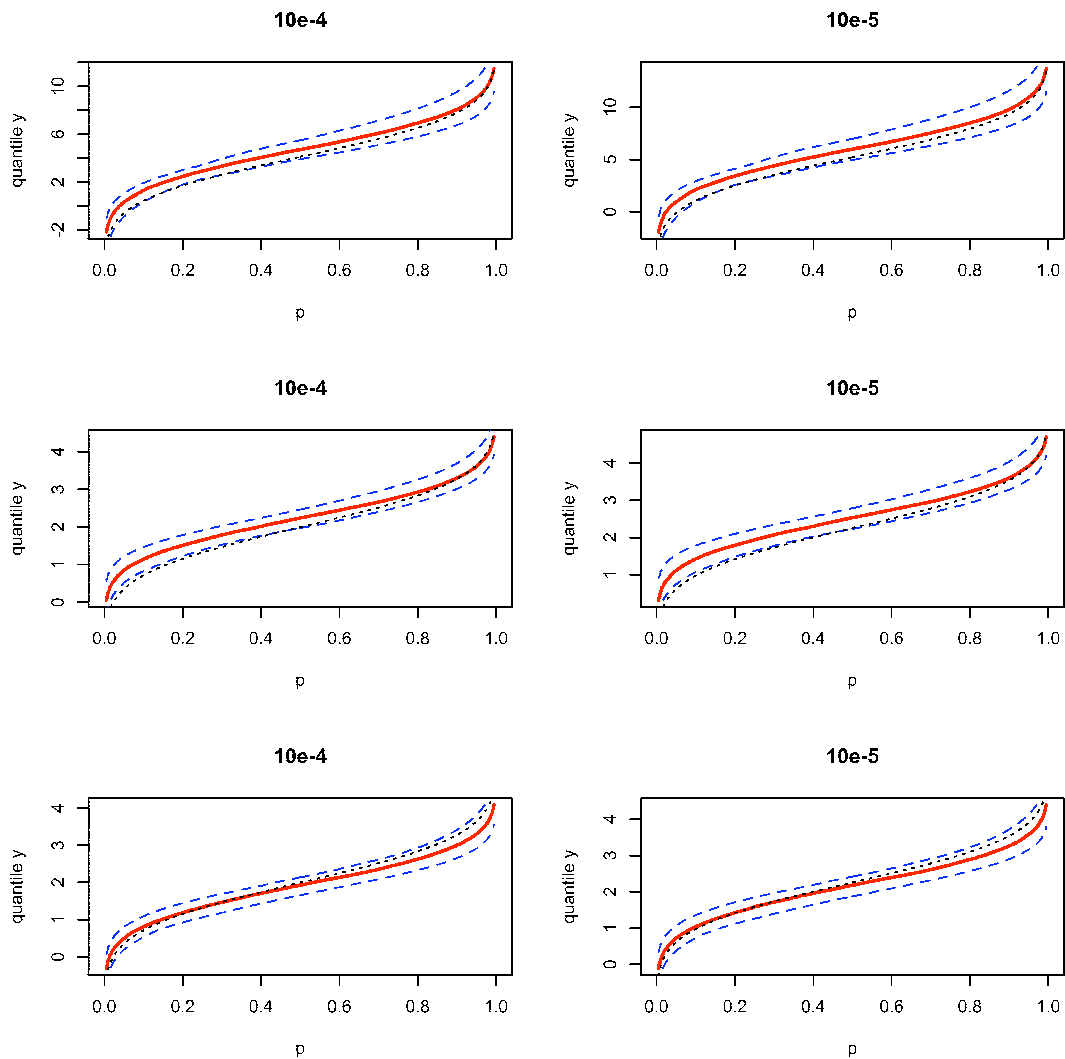


Figure 2: Median (solid line), 2.5%- and 97.5%-quantiles (dashed lines) of the estimated conditional quantile function  $\hat{y} = \theta^{\leftarrow}(x, p)$  defined in (46) and theoretical conditional quantile function  $y$  (dotted line) as a function of the probability  $p \in (0, 1)$ . Each row (from 1 to 3) corresponds to a distribution (from (a) to (c)) as described in Section 5.1. Each column refers to a different value of  $x$ , respectively corresponding to the theoretical  $X$ -quantiles of order  $1 - \epsilon$ , where  $\epsilon = 10^{-4}$  and  $p = 10^{-5}$ .