Nonparametric extreme conditional expectile estimation

Stéphane Girard, Gilles Stupfler, Antoine Usseglio-Carleve

Nonparametric extreme conditional expectile estimation

Stéphane Girard[a], Gilles Stupfler[b] & Antoine Usseglio-Carleve[a]

[a] Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
[b] University of Nottingham, School of Mathematical Sciences, Nottingham NG7 2RD, UK

Abstract. Expectiles and quantiles can both be defined as the solution of minimization problems. Contrary to quantiles though, expectiles are determined by tail expectations rather than tail probabilities, and define a coherent risk measure. For these two reasons in particular, expectiles have recently started to be considered as serious candidates to become standard tools in actuarial and financial risk management. However, expectiles and their sample versions do not benefit from a simple explicit form, making their analysis significantly harder than that of quantiles and order statistics. This difficulty is compounded when one wishes to integrate auxiliary information about the phenomenon of interest through a finite-dimensional covariate, in which case the problem becomes the estimation of conditional expectiles. In this paper, we exploit the fact that the expectiles of a distribution F are in fact the quantiles of another distribution E explicitly linked to F, in order to construct nonparametric kernel estimators of extreme conditional expectiles. We analyze the asymptotic properties of our estimators in the context of conditional heavy-tailed distributions. Applications to simulated data and real insurance data are provided.

Keywords. Conditional distribution, expectiles, extrapolation, extreme value analysis, kernel estimation, heavy tails.

1 Introduction

The concept of quantile is a fundamental tool for risk measurement. In a financial or actuarial setting, it is commonly known as Value-at-Risk, see Embrechts et al. (1997) and Linsmeier and Pearson (2000) for a basic introduction. For a real-valued random variable Y, having distribution function F, the quantile at a level $\alpha \in (0, 1)$ is defined as the generalized inverse $q(\alpha) = \inf\{y \in \mathbb{R} \,|\, F(y) \geq \alpha\}$. A traditional way of analyzing extreme risk is to estimate high quantiles of a variable Y appropriate to the situation at hand, such as the negative daily log-return of a stock market index in finance, or the magnitude of a claim in insurance. Despite the simplicity and interpretability of quantiles, using them as the single tool for risk assessment is not without disadvantages. For instance, quantiles do not, in general, induce a coherent risk measure in the sense of Artzner et al. (1999), because they fail to be subadditive (see Acerbi, 2002). Besides, quantiles only use information on the frequency of tail events and not on their actual magnitudes; this is an issue in risk management, where the focus is not only what constitutes an extreme level of loss but also what a typical extreme loss will be.

These drawbacks of quantiles motivated the introduction of a number of alternative risk measures, among which expectiles, which are the focus of the present paper. Expectiles were introduced by Newey and Powell (1987), following earlier work of Koenker and Bassett (1978) which characterized quantiles as solutions of an L^1-minimization problem:

$$q(\alpha) \in \operatorname*{arg\,min}_{t \in \mathbb{R}} \mathbb{E}\left(\rho_\alpha(Y - t) - \rho_\alpha(Y)\right) \tag{1.1}$$

where $\rho_\alpha(y) = |\alpha - \mathbb{1}_{\{y \leq 0\}}| \, |y|$ is the quantile check function and $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Note the $\in$ sign, accounting for the fact that the minimizer may not be unique in (1.1); there is actually equality if the distribution function of Y is increasing. The idea of Newey and Powell (1987) was to replace

L^1-minimization by L^2-minimization, giving rise to expectiles as minimizers of an asymmetrically weighted mean squared deviations criterion:

$$e(\alpha) = \arg\min_{t \in \mathbb{R}} \mathbb{E} \left(\eta_\alpha(Y - t) - \eta_\alpha(Y) \right), \tag{1.2}$$

where $\eta_\alpha(y) = |\alpha - \mathbb{1}_{\{y \leq 0\}}| \, y^2$ is the expectile check function. Quantiles and expectiles both belong to the family of L^p-quantiles, introduced in Chen (1996), themselves part of the wider family of $M-$quantiles (see Breckling and Chambers, 1988). However, unlike quantiles, expectiles are determined by tail expectations rather than tail probabilities, and therefore depend on both the probability of tail values and their actual realizations (see Kuan et al., 2009). Expectiles at a level $\alpha > 1/2$ are also the only $M-$quantiles that define a coherent risk measure (see Bellini et al., 2014). In fact, expectiles induce the only coherent law-invariant risk measure that is elicitable (see Ziegel, 2016) and, as such, benefit from the existence of a natural backtesting methodology (see Gneiting, 2011). This makes expectiles a sensible tool for risk management and a number of authors have experimented their use in this context: for instance, Taylor (2008) relates expectiles to the Expected Shortfall, Kuan et al. (2009) work on risk assessment with expectiles in an autoregressive model, Cai and Weng (2016) introduce a reinsurance methodology based on expectiles and Bellini and Di Bernardino (2017) explore financial risk management with expectiles. Recent advanced theoretical developments, on expectiles of a fixed level and therefore staying away from the extremes of the data, have been brought by Holzmann and Klar (2016) and Krätschmer and Zähle (2017).

It often happens in practical applications that Y is recorded along with auxiliary information represented by a random covariate $\boldsymbol{X} \in \mathbb{R}^p$. In this context, one can give a more precise answer to the problem of inferring the extremes of Y by focusing on the conditional extremes of Y given $\boldsymbol{X}$. Some recent applied examples are the development of a general strategy by Chavez-Demoulin et al. (2016) for the construction of stratified models for large operational losses, and modeling large insurance claims conditional on climate variables in Rohrbeck et al. (2018). An interesting question, on which there is a growing body of literature, is then to infer the conditional extremes of Y given $\boldsymbol{X}$ under a general extreme value setting, when one only assumes that the distribution of Y given $\boldsymbol{X}$ belongs to the domain of attraction of a Generalized Extreme Value distribution (see Chapter 1 of de Haan and Ferreira, 2006). This problem has typically been tackled from the perspective of extreme conditional quantile estimation, namely, the estimation of extreme quantiles of the conditional distribution function. Recent contributions in this area include Daouia et al. (2011) who used a fixed number of nonparametric conditional quantile estimators to estimate the conditional tail index, later generalized in Daouia et al. (2013) to a regression context with conditional response distributions belonging to the general max-domain of attraction; the latter situation is also considered in Goegebeur et al. (2017). The context when the response variable Y is randomly right-censored is examined in Stupfler (2016).

By contrast, extreme conditional expectile estimation has been left virtually untouched. The only attempt in the literature so far seems to have been made by Usseglio-Carleve (2018), who estimates a variety of extreme conditional risk measures, including expectiles, under the assumption that the vector $(\boldsymbol{X}, Y)$ has a so-called consistent elliptical distribution. This is a rather strong distributional assumption which does not appear to be easy to check on real data. In fact, the theory on extreme expectiles is still largely unexplored even in the unconditional case. This is in no small part due to the absence of a closed form expression for expectiles, making the analysis of extreme expectiles much harder than that of extreme quantiles. Although asymptotic equivalents and expansions of extreme population expectiles were derived by Bellini et al. (2014), Mao et al. (2015), Mao and Yang (2015) and Bellini and Di Bernardino (2017), the work on extreme expectile estimation is so far restricted to the papers by Daouia et al. (2018, 2019b) that constructed and studied classes of estimators in the unconditional heavy-tailed case. It is the purpose of this paper to introduce a fully nonparametric methodology for the estimation of extreme conditional expectiles, in the case when Y given $\boldsymbol{X}$ has a heavy-tailed, or equivalently Pareto-type, distribution. This technique allows us to avoid the strong modeling condition of Usseglio-Carleve (2018), while still working within a framework of heavy tails which is appropriate to the description of actuarial and financial data; see *e.g.* the discussions on p.9 of Embrechts et al. (1997) and p.1 of Resnick (2007).

Our work plan is the following. We start by recalling that, in the words of Jones (1994), expectiles are quantiles of an explicitly known distribution, so that the problem of estimating conditional expectiles reduces to estimating a suitable conditional distribution function and then inverting this estimator. This is done under conditional analogues of standard extreme value conditions, plus mild conditions on the regularity of the density of $\boldsymbol{X}$ and of the conditional distribution function of Y. Such assumptions are tailored to our construction of nonparametric kernel estimators of extreme conditional expectiles. We start by estimating so-called intermediate conditional expectiles, whose order $\alpha_n \to 1$ as the sample size $n \to \infty$ with α_n not too large (the meaning of which will be explained in the text). Under these assumptions, we obtain the pointwise asymptotic normality of our intermediate conditional expectile estimators. We then exploit the conditional heavy-tailed assumption to construct an extrapolated Weissman-type estimator (see Weissman, 1978) of extreme conditional expectiles, whose order α_n can tend to 1 at any rate. The asymptotic distribution of the extrapolated estimator is obtained under sensible conditions we shall explain in detail, including a second-order condition which is standard in extreme value analysis and that quantifies the gap between the underlying distribution and a purely Pareto distribution. One drawback of the Weissman-type extrapolation methodology that is specific to the estimation of extreme expectiles is that its use incurs bias due not only to the distance between the distribution of interest and a relevant Pareto-type distribution, as would be the case for extreme quantile estimation, but also to the actual heaviness of the tail of the distribution. The finite-sample performance of the Weissman-type estimator typically deteriorates quite strongly due to this latter source of bias. This motivates our final estimators, whose structure is designed precisely to deal with this bias using a very simple multiplicative correction factor.

The paper is organized as follows. Section 2 states our nonparametric framework and the basic building blocks of our estimation procedures. Section 3 introduces our estimators and gives our main results, first in the intermediate case and then in the properly extreme case. Section 4 showcases the performance of our estimators on several simulated situations. We illustrate the applicability of our procedures on a real sample of actuarial data in Section 5. All proofs are relegated to the Appendix.

2   Nonparametric conditional expectile estimation

Let $(\boldsymbol{X}_i, Y_i)$, $i = 1, \ldots, n$, be independent realizations of a random vector $(\boldsymbol{X}, Y) \in \mathbb{R}^p \times \mathbb{R}$. We denote by g the probability density function of $\boldsymbol{X}$ and let, throughout the paper, $\boldsymbol{x}$ be a fixed point in $\mathbb{R}^p$ such that $g(\boldsymbol{x}) > 0$. We write $F(y|\boldsymbol{x}) = \mathbb{P}(Y \leq y | \boldsymbol{X} = \boldsymbol{x})$ for the conditional distribution function of Y given $\boldsymbol{X} = \boldsymbol{x}$. Our central assumption throughout is that $Y | \boldsymbol{X} = \boldsymbol{x}$ has a heavy-tailed distribution. In other words, we assume that there exists $\gamma(\boldsymbol{x}) > 0$, called the conditional tail index, such that the survival function $\overline{F}(\cdot|\boldsymbol{x}) = 1 - F(\cdot|\boldsymbol{x})$ is regularly varying with index $-1/\gamma(\boldsymbol{x})$:

$$\forall y > 0, \ \lim_{t \to \infty} \frac{\overline{F}(ty|\boldsymbol{x})}{\overline{F}(t|\boldsymbol{x})} = y^{-1/\gamma(\boldsymbol{x})}. \tag{2.1}$$

Equivalently, according to Theorem 1.2.1 in de Haan and Ferreira (2006), the conditional distribution function $F(\cdot|\boldsymbol{x})$ belongs to the Fréchet maximum domain of attraction. The parameter $\gamma(\boldsymbol{x})$ tunes the tail heaviness of the conditional distribution. In particular, if $\gamma(\boldsymbol{x}) > a$ then $\mathbb{E}[Y^{1/a} \mathbb{1}_{\{Y>0\}} | \boldsymbol{X} = \boldsymbol{x}] = \infty$ (a precise statement is Exercise 1.16 p.35 in de Haan and Ferreira, 2006). Since the definition of expectiles in (1.2) requires $\mathbb{E}[|Y| \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$, our minimal working assumption throughout will be that $\gamma(\boldsymbol{x}) < 1$ and $\mathbb{E}[Y_- \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$, where $Y_- = \max(-Y, 0)$.

Our work is based on the following observation made by Jones (1994). For any $\alpha \in (0, 1)$, the conditional expectile of level α, that is

$$e(\alpha|\boldsymbol{x}) = \underset{t \in \mathbb{R}}{\arg\min} \, \mathbb{E}\left(\eta_\alpha(Y - t) - \eta_\alpha(Y) \mid \boldsymbol{X} = \boldsymbol{x}\right), \tag{2.2}$$

is actually the quantile of level α associated to the distribution function $E(y|\boldsymbol{x})$ defined by

$$1 - E(y|\boldsymbol{x}) = \frac{\mathbb{E}\left[(Y - y)\mathbb{1}_{\{Y>y\}} \mid \boldsymbol{X} = \boldsymbol{x}\right]}{2\,\mathbb{E}\left[(Y - y)\mathbb{1}_{\{Y>y\}} \mid \boldsymbol{X} = \boldsymbol{x}\right] + (y - \mathbb{E}\left[Y \mid \boldsymbol{X} = \boldsymbol{x}\right])}.$$

For convenience, we define $\overline{E}(y|\boldsymbol{x}) = 1 - E(y|\boldsymbol{x})$, as well as (whenever these definitions make sense)

$$\forall k \in \{0, 1, 2, \ldots\}, \ \psi^{(k)}(y|\boldsymbol{x}) = \mathbb{E}\left[(Y-y)^k \mathbb{1}_{\{Y>y\}} \,|\, \boldsymbol{X} = \boldsymbol{x}\right] g(\boldsymbol{x}) \text{ and } m^{(k)}(\boldsymbol{x}) = \mathbb{E}[Y^k \,|\, \boldsymbol{X} = \boldsymbol{x}].$$

This allows us to write $e(\alpha|\boldsymbol{x}) = \inf\{y \in \mathbb{R} \,|\, \overline{E}(y|\boldsymbol{x}) \leq 1 - \alpha\}$, with

$$\overline{E}(y|\boldsymbol{x}) = \frac{\psi^{(1)}(y|\boldsymbol{x})}{2\psi^{(1)}(y|\boldsymbol{x}) + \left(y - m^{(1)}(\boldsymbol{x})\right) g(\boldsymbol{x})}. \tag{2.3}$$

Our construction of an estimator of $e(\alpha|\boldsymbol{x})$, for $\alpha \to 1$, thus follows the idea of Daouia et al. (2013): we first estimate the conditional survival function $\overline{E}(y|\boldsymbol{x})$ for high values of y, and our expectile estimator is obtained from the inverse of this estimator of the conditional survival function.

To estimate $\overline{E}(y|\boldsymbol{x})$, we note that the unknown quantities appearing in Equation (2.3) are either the density of $\boldsymbol{X}$ or conditional moments of certain functions of Y. An appropriate class of nonparametric estimators for the estimation of density functions and conditional moments is the family of kernel estimators. Let K be a probability density function on $\mathbb{R}^p$, and define the following kernel estimators:

$$\widehat{g}_n(\boldsymbol{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right), \quad \widehat{m}_n^{(1)}(\boldsymbol{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n Y_i \, K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right) \bigg/ \widehat{g}_n(\boldsymbol{x})$$

$$\text{and} \quad \forall k \in \{0, 1, 2, \ldots\}, \ \widehat{\psi}_n^{(k)}(y|\boldsymbol{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n (Y_i - y)^k K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right) \mathbb{1}_{\{Y_i > y\}}$$

where h_n is a positive bandwidth sequence such that $h_n \to 0$ and $nh_n^p \to \infty$ as $n \to \infty$. The estimators $\widehat{g}_n(\boldsymbol{x})$ and $\widehat{m}_n^{(1)}(\boldsymbol{x})$ are respectively known in the literature as the Parzen-Rosenblatt estimator (see Rosenblatt, 1956; Parzen, 1962) and the Nadaraya-Watson estimator (see Nadaraya, 1964; Watson, 1964). The conditional survival function of Y is obtained as $\overline{F}(y|\boldsymbol{x}) = \psi^{(0)}(y|\boldsymbol{x})/g(\boldsymbol{x})$. Its estimator $\widehat{\overline{F}}_n(y|\boldsymbol{x}) = \widehat{\psi}_n^{(0)}(y|\boldsymbol{x})/\widehat{g}_n(\boldsymbol{x})$ plays a central role in the construction of the (extreme) conditional quantile estimator $\widehat{q}_n(\alpha|\boldsymbol{x})$ of Daouia et al. (2013), as the latter is nothing but its generalized inverse:

$$\widehat{q}_n(\alpha|\boldsymbol{x}) = \inf\left\{y \in \mathbb{R} \,\Big|\, \widehat{\overline{F}}_n(y|\boldsymbol{x}) \leq 1 - \alpha\right\}. \tag{2.4}$$

In our context of conditional expectile estimation, we estimate the survival function $\overline{E}(\cdot|\boldsymbol{x})$ by

$$\widehat{\overline{E}}_n(y|\boldsymbol{x}) = \frac{\widehat{\psi}_n^{(1)}(y|\boldsymbol{x})}{2\widehat{\psi}_n^{(1)}(y|\boldsymbol{x}) + \left(y - \widehat{m}_n^{(1)}(\boldsymbol{x})\right) \widehat{g}_n(\boldsymbol{x})}$$

and we obtain a conditional expectile estimator at level $\alpha \in (0, 1)$ by inverting this empirical survival function $\widehat{\overline{E}}_n(\cdot|\boldsymbol{x})$:

$$\widehat{e}_n(\alpha|\boldsymbol{x}) = \inf\left\{y \in \mathbb{R} \,\Big|\, \widehat{\overline{E}}_n(y|\boldsymbol{x}) \leq 1 - \alpha\right\}. \tag{2.5}$$

It is remarkable that this conditional expectile estimator is actually also the expectile of the empirical conditional distribution

$$\widehat{F}_n(y|\boldsymbol{x}) = 1 - \frac{\widehat{\psi}_n^{(0)}(y|\boldsymbol{x})}{\widehat{g}_n(\boldsymbol{x})} = \sum_{i=1}^n \frac{K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_j}{h_n}\right)} \mathbb{1}_{\{Y_i \leq y\}}$$

in the sense that

$$\widehat{e}_n(\alpha|\boldsymbol{x}) = \arg\min_{t \in \mathbb{R}} \sum_{i=1}^n \eta_\alpha(Y_i - t) K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right) = \arg\min_{t \in \mathbb{R}} \int_{\mathbb{R}} \eta_\alpha(y - t) d\widehat{F}_n(y|\boldsymbol{x}).$$

This is readily checked by noting that the above minimization criterion is a strictly convex, continuously differentiable function of t whose derivative cancels at the value of t satisfying

$$1 - \alpha = \frac{\widehat{\psi}_n^{(1)}(t|\boldsymbol{x})}{2\widehat{\psi}_n^{(1)}(t|\boldsymbol{x}) + \left(t - \widehat{m}_n^{(1)}(\boldsymbol{x})\right)\widehat{g}_n(\boldsymbol{x})} = \widehat{\overline{E}}_n(t|\boldsymbol{x}).$$

The estimator $\widehat{e}_n(\alpha|\boldsymbol{x})$ can therefore be seen as a nonparametric version of the expectile estimator introduced originally in Newey and Powell (1987) in the standard linear regression model, and as a kernel smoothed version of the LAWS estimator studied in the unconditional extreme case in Daouia et al. (2018). A similar construction is used in Guo and Härdle (2012), where the asymptotic properties of the estimator are derived for a fixed α. Our contribution in this paper is to consider an expectile level $\alpha = \alpha_n$ such that $\alpha_n \to 1$ as $n \to \infty$, which is the appropriate framework in the management of extreme risk. This is the focus of our next section.

3 Main results

3.1 Estimation of intermediate expectiles

We start by studying the joint asymptotic properties of estimators (2.4) and (2.5), when $\alpha = \alpha_n \to 1$. Choose a norm $\|\cdot\|$ on $\mathbb{R}^p$. We make the following assumption, which is standard in the nonparametric literature.

$(\mathcal{K})$ The density function K is bounded and its support S is the unit ball in $\mathbb{R}^p$ for the norm $\|\cdot\|$.

We assume in this section that $nh_n^p(1 - \alpha_n) \to \infty$. Such quantile (or expectile) levels α_n therefore converge to 1 but cannot do so too quickly; we follow here the conventions of extreme value theory and call such a level α_n *intermediate* throughout the paper. The condition $nh_n^p(1 - \alpha_n) \to \infty$ can also be found in Daouia et al. (2013), and is the analogue of the condition $n(1 - \alpha_n) \to \infty$ used in Daouia et al. (2018, 2019b) for the estimation of unconditional intermediate expectiles. The presence of the multiplicative term h_n^p is due to the use of the kernel smoothing method in our conditional framework.

To estimate intermediate conditional expectiles, we refine assumption (2.1) in the following way:

$\mathcal{C}_1(\gamma(\boldsymbol{x}))$ The survival function $\overline{F}(\cdot|\boldsymbol{x})$ is continuously differentiable and satisfies

$$\lim_{y \to \infty} \frac{y\overline{F}'(y|\boldsymbol{x})}{\overline{F}(y|\boldsymbol{x})} = -\frac{1}{\gamma(\boldsymbol{x})}.$$

It follows from Theorem 1.1.11 in de Haan and Ferreira (2006) that condition (2.1) is indeed satisfied if $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ holds. Another consequence of condition $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ is that the conditional density function $f(\cdot|\boldsymbol{x}) = -\overline{F}'(\cdot|\boldsymbol{x})$ is also regularly varying, with index $-1/\gamma(\boldsymbol{x}) - 1$. Assuming that $\overline{F}(\cdot|\boldsymbol{x})$ is continuously differentiable is unlikely to be very restrictive in practice; Newey and Powell (1987) point out in their Theorem 1 that it is a reasonable sufficient condition for expectiles to characterize the underlying distribution.

Our final assumption in this section is a local Lipschitz condition on the density function g and the conditional moments $m^{(1)}(\boldsymbol{x})$ and $m^{(2)}(\boldsymbol{x})$. Similar conditions are used, for instance, in Krzyzak (1986), or more recently in Daouia et al. (2013) and El Methni et al. (2014) in the context of conditional extreme value analysis. We denote by $B(\boldsymbol{x}, r)$ the ball with center $\boldsymbol{x}$ and radius r and by $\vee$ the maximum operator.

$(\mathcal{L})$ We have $g(\boldsymbol{x}) > 0$, $m^{(2)}(\boldsymbol{x}) < \infty$ and there exist $c, r > 0$ such that

$$\forall \boldsymbol{x}' \in B(\boldsymbol{x}, r), \ \left|g(\boldsymbol{x}) - g(\boldsymbol{x}')\right| \vee \left|m^{(1)}(\boldsymbol{x}) - m^{(1)}(\boldsymbol{x}')\right| \vee \left|m^{(2)}(\boldsymbol{x}) - m^{(2)}(\boldsymbol{x}')\right| \leq c \|\boldsymbol{x} - \boldsymbol{x}'\|.$$

Before we state our first main result, we introduce some useful notation for the oscillation of the conditional survival function $\overline{F}(\cdot|\boldsymbol{x})$ above a high level y_n:

$$\omega_{h_n}(y_n|\boldsymbol{x}) = \sup_{\substack{z \geq y_n \\ \boldsymbol{x}' \in B(\boldsymbol{x},h_n)}} \frac{1}{\log z} \left| \log \frac{\overline{F}(z|\boldsymbol{x}')}{\overline{F}(z|\boldsymbol{x})} \right|. \tag{3.1}$$

The quantity $\omega_{h_n}(y_n|\boldsymbol{x})$ measures the discrepancy between the extremes of the conditional distributions of Y at neighboring points. Similar quantities are introduced in Gardes and Stupfler (2014, 2019) and Stupfler (2013, 2016). In order to get an idea of the typical asymptotic behavior of $\omega_{h_n}(y_n|\boldsymbol{x})$, consider the Karamata representation of $\overline{F}(\cdot|\boldsymbol{x})$ (see Theorem 1.3.1 in Bingham et al., 1989):

$$\forall z \geq 1, \ \overline{F}(z|\boldsymbol{x}) = z^{-1/\gamma(\boldsymbol{x})} \exp\left(\eta(z|\boldsymbol{x}) + \int_1^z \frac{\epsilon(u|\boldsymbol{x})}{u} du \right),$$

where $\eta(\cdot|\boldsymbol{x})$ and $\epsilon(\cdot|\boldsymbol{x})$ are measurable functions converging, respectively, to a constant and 0 at infinity. In this context, it is straightforward to prove that if there are $c, r > 0$ with

$$\forall \boldsymbol{x}' \in B(\boldsymbol{x},r), \ \left|\gamma(\boldsymbol{x}) - \gamma(\boldsymbol{x}')\right| \vee \sup_{z \geq y_n} \left| \frac{\eta(z|\boldsymbol{x}) - \eta(z|\boldsymbol{x}')}{\log z} \right| \vee \sup_{z \geq 1} \left| \epsilon(z|\boldsymbol{x}) - \epsilon(z|\boldsymbol{x}') \right| \leq c \, \|\boldsymbol{x} - \boldsymbol{x}'\|$$

then $\omega_{h_n}(y_n|\boldsymbol{x}) = O(h_n)$. The same kind of discussion may be found in Stupfler (2013, 2016).

We are now ready to write our first result on the asymptotic properties of the estimator $\widehat{e}_n(\alpha_n|\boldsymbol{x})$ in the intermediate case $nh_n^p(1-\alpha_n) \to \infty$. We actually obtain the joint asymptotic normality of a finite number J of empirical conditional intermediate expectiles $\widehat{e}_n(\alpha_{n,j}|\boldsymbol{x})$, with $1 - \alpha_{n,j} = \tau_j(1-\alpha_n)$, $0 < \tau_1, \ldots, \tau_J \leq 1$, together with an empirical conditional intermediate quantile $\widehat{q}_n(a_n|\boldsymbol{x})$. Let $\|K\|_2^2 = \int_S K^2(\boldsymbol{u})d\boldsymbol{u}$ denote the squared L^2-norm of K.

Theorem 1. *Assume that* $(\mathcal{K})$, $(\mathcal{L})$ *and* $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ *hold. Suppose also that* $\gamma(\boldsymbol{x}) < 1/2$ *and that there exists* $\delta \in (0,1)$ *with* $\mathbb{E}[Y_-^{2+\delta} \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. *Let* $\alpha_n \to 1$, $h_n \to 0$, *and* $a_n = 1 - \tau(1-\alpha_n)(1 + o(1))$, *where* $\tau > 0$. *Assume that* $nh_n^p(1-\alpha_n) \to \infty$, $nh_n^{p+2}(1-\alpha_n) \to 0$ *and*

$$\sqrt{nh_n^p(1-\alpha_n)} \log(1-\alpha_n) \times \left[\omega_{h_n}((1-\delta)e(\alpha_n|\boldsymbol{x})|\boldsymbol{x}) \vee \omega_{h_n}((1-\delta)q(a_n|\boldsymbol{x})|\boldsymbol{x}) \right] \to 0. \tag{3.2}$$

Then

$$\sqrt{nh_n^p(1-\alpha_n)} \left\{ \left(\frac{\widehat{e}_n(\alpha_{n,j}|\boldsymbol{x})}{e(\alpha_{n,j}|\boldsymbol{x})} - 1 \right)_{1 \leq j \leq J}, \left(\frac{\widehat{q}_n(a_n|\boldsymbol{x})}{q(a_n|\boldsymbol{x})} - 1 \right) \right\} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}_{J+1}, \frac{\|K\|_2^2}{g(\boldsymbol{x})} \gamma^2(\boldsymbol{x})\boldsymbol{\Sigma}(\boldsymbol{x}) \right),$$

where $1 - \alpha_{n,j} = \tau_j(1-\alpha_n)$ *for some* $0 < \tau_1 < \tau_2 < \cdots < \tau_J \leq 1$ *and* $\boldsymbol{\Sigma}(\boldsymbol{x})$ *is the symmetric matrix having entries*

$$\Sigma_{j,l}(\boldsymbol{x}) = \tau_l^{-1} \left[\frac{1}{1 - 2\gamma(\boldsymbol{x})} \left(\frac{\tau_l}{\tau_j} \right)^{\gamma(\boldsymbol{x})} - 1 \right],$$

$$\Sigma_{j,J+1}(\boldsymbol{x}) = \tau_j^{-1} \left(\frac{\tau_j}{\tau} \right)^{\gamma(\boldsymbol{x})} \left[\left(\frac{1}{\gamma(\boldsymbol{x})^{-1} - 1} \wedge \frac{\tau_j}{\tau} \right)^{1-\gamma(\boldsymbol{x})} + \left(\frac{1}{\gamma(\boldsymbol{x})^{-1} - 1} \wedge \frac{\tau_j}{\tau} \right)^{-\gamma(\boldsymbol{x})} - \left(\frac{\tau_j}{\tau} \right)^{-\gamma(\boldsymbol{x})} \right],$$

$$\Sigma_{J+1,J+1}(\boldsymbol{x}) = \tau^{-1},$$

for $j, l \in \{1, \ldots, J\}$, $j \leq l$.

Theorem 1 requires the conditions $\gamma(\boldsymbol{x}) < 1/2$ and $\mathbb{E}[Y_-^{2+\delta} \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$, which essentially amount to assuming that the conditional variance of Y given $\boldsymbol{X} = \boldsymbol{x}$ is finite. The conditions $nh_n^{p+2}(1-\alpha_n) \to 0$ and $\sqrt{nh_n^p(1-\alpha_n)} \log(1-\alpha_n) \times \left[\omega_{h_n}((1-\delta)e(\alpha_n|\boldsymbol{x})|\boldsymbol{x}) \vee \omega_{h_n}((1-\delta)q(a_n|\boldsymbol{x})|\boldsymbol{x}) \right] \to 0$ ensure that the bias incurred by the use of the kernel smoothing technique is asymptotically negligible.

Theorem 1 is useful for jointly estimating conditional intermediate expectiles at several different levels, and will be used in the next paragraph for the analysis of our extrapolation technique. We conclude this section by stating below a simpler but instructive corollary in the joint distribution of empirical conditional expectile and quantile. This result is a direct consequence of Theorem 1 in the case $J = 1$, $\tau_1 = 1$ and $\tau = 1$.

Corollary 1. *Under the conditions of Theorem 1, with assumption* (3.2) *replaced by the weaker assumption*

$$\sqrt{nh_n^p(1-\alpha_n)} \log(1-\alpha_n) \times \omega_{h_n}((1-\delta)e(\alpha_n|\boldsymbol{x})|\boldsymbol{x}) \to 0,$$

one has

$$\sqrt{nh_n^p(1-\alpha_n)} \left(\frac{\widehat{e}_n(\alpha_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})} - 1, \frac{\widehat{q}_n(\alpha_n|\boldsymbol{x})}{q(\alpha_n|\boldsymbol{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(\boldsymbol{0}_2, \frac{||K||_2^2}{g(\boldsymbol{x})} \gamma^2(\boldsymbol{x}) \boldsymbol{\Lambda}(\boldsymbol{x}) \right),$$

where the matrix $\boldsymbol{\Lambda}(\boldsymbol{x})$ *is defined as*

$$\boldsymbol{\Lambda}(\boldsymbol{x}) = \begin{pmatrix} \dfrac{2\gamma(\boldsymbol{x})}{1-2\gamma(\boldsymbol{x})} & \dfrac{(\gamma(\boldsymbol{x})^{-1}-1)^{\gamma(\boldsymbol{x})}}{1-\gamma(\boldsymbol{x})} - 1 \\ \dfrac{(\gamma(\boldsymbol{x})^{-1}-1)^{\gamma(\boldsymbol{x})}}{1-\gamma(\boldsymbol{x})} - 1 & 1 \end{pmatrix}.$$

This corollary gives in particular the asymptotic normality of $\widehat{e}_n(\alpha_n|\boldsymbol{x})$ as

$$\sqrt{nh_n^p(1-\alpha_n)} \left(\frac{\widehat{e}_n(\alpha_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{||K||_2^2}{g(\boldsymbol{x})} \times \frac{2\gamma^3(\boldsymbol{x})}{1-2\gamma(\boldsymbol{x})} \right),$$

of which an unconditional analogue is Theorem 2 in Daouia et al. (2018). More generally, Corollary 1 can be seen as a conditional analogue of Theorem 3 of Daouia et al. (2019b).

Our main results so far are restricted to intermediate levels α_n, therefore preventing us from estimating the properly extreme conditional expectiles which are of interest in risk assessment. Providing estimators of arbitrarily extreme conditional expectiles, based on an extrapolation procedure warranted by the assumption of conditional heavy tails, is the focus of our next section.

3.2 Estimation of extreme conditional expectiles by extrapolation

In this section, the aim is to estimate conditional expectiles at a level β_n such that $nh_n^p(1-\beta_n) \to c < \infty$ as $n \to \infty$. Contrary to intermediate levels, such levels of conditional expectiles are rarely, if at all, observed in the sample. For that purpose, we can therefore no longer use estimator (2.5). To construct an adapted estimator, we note that Assumption (2.1) can be rewritten (see Theorem 1.2.1 and Corollary 1.2.10 in de Haan and Ferreira, 2006)

$$\forall y > 0, \lim_{t \to \infty} \frac{q(1 - 1/(ty)|\boldsymbol{x})}{q(1 - 1/t|\boldsymbol{x})} = y^{\gamma(\boldsymbol{x})}.$$

Now, by Proposition 1 in Daouia et al. (2019b),

$$\frac{e(\alpha|\boldsymbol{x})}{q(\alpha|\boldsymbol{x})} \to (\gamma(\boldsymbol{x})^{-1} - 1)^{-\gamma(\boldsymbol{x})} \quad \text{as} \quad \alpha \to 1. \tag{3.3}$$

It follows that

$$\forall y > 0, \lim_{t \to \infty} \frac{e(1 - 1/(ty)|\boldsymbol{x})}{e(1 - 1/t|\boldsymbol{x})} = y^{\gamma(\boldsymbol{x})}.$$

This suggests that when α_n and β_n satisfy suitable conditions, we may write an extrapolation formula linking $e(\beta_n|\boldsymbol{x})$ to $e(\alpha_n|\boldsymbol{x})$:

$$e(\beta_n|\boldsymbol{x}) \approx \left(\frac{1-\alpha_n}{1-\beta_n} \right)^{\gamma(\boldsymbol{x})} e(\alpha_n|\boldsymbol{x}). \tag{3.4}$$

When α_n is an intermediate sequence, this approximation can be used to define a plug-in estimator of $e(\beta_n|\boldsymbol{x})$, based on an estimator of $\gamma(\boldsymbol{x})$ and on the estimator $\widehat{e}_n(\alpha_n|\boldsymbol{x})$. Estimators of $\gamma(\boldsymbol{x})$ typically rely on the highest observations in the ball $B(\boldsymbol{x}, h_n)$ only (see *e.g.* Gardes and Stupfler, 2014). It is therefore more convenient, in the extrapolation context, to set $\alpha_n = 1 - k_n/n$, where $k_n/n \to 0$ is the fraction of those observations used in the estimation of $\gamma(\boldsymbol{x})$, and work with this sequence k_n satisfying

$k_n h_n^p = n h_n^p (1 - \alpha_n) \to \infty$. Since $k_n h_n^p$ is asymptotically proportional to the local average number of (high) observations relevant to the estimation of $e(\alpha_n | \boldsymbol{x})$ within the ball $B(\boldsymbol{x}, h_n)$, the condition $k_n h_n^p \to \infty$ is the analogue of the assumption $k_n \to \infty$ classically encountered in the unconditional case for the estimation of the tail index (see Chapter 3 in de Haan and Ferreira, 2006). Using this parametrization, we introduce the following class of extrapolated estimators:

$$\widehat{e}_{n,k_n}^W (\beta_n | \boldsymbol{x}) = \left(\frac{k_n}{n(1 - \beta_n)} \right)^{\widehat{\gamma}_{k_n}(\boldsymbol{x})} \widehat{e}_n(1 - k_n/n | \boldsymbol{x}). \tag{3.5}$$

Here $\widehat{\gamma}_{k_n}(\boldsymbol{x})$ is any consistent estimator of $\gamma(\boldsymbol{x})$. This is a class of Weissman-type estimators (see Weissman, 1978, for the estimation of unconditional extreme quantiles).

To study the asymptotic properties of estimators part of the class (3.5), we have to quantify precisely the bias incurred by the use of the extrapolation formula (3.4). Our key tool for this is the following second-order condition.

$\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot | \boldsymbol{x}))$ Condition $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ holds, and there exist $\rho(\boldsymbol{x}) \leq 0$ and a positive or negative function $A(\cdot | \boldsymbol{x})$ such that:

$$\forall y > 0, \ \lim_{t \to \infty} \frac{1}{A(t|\boldsymbol{x})} \left(\frac{q(1 - 1/(ty)|\boldsymbol{x})}{q(1 - 1/t|\boldsymbol{x})} - y^{\gamma(\boldsymbol{x})} \right) = \begin{cases} y^{\gamma(\boldsymbol{x})} \dfrac{y^{\rho(\boldsymbol{x})} - 1}{\rho(\boldsymbol{x})} & \text{if } \rho(\boldsymbol{x}) < 0, \\[2mm] y^{\gamma(\boldsymbol{x})} \log y & \text{if } \rho(\boldsymbol{x}) = 0. \end{cases} \tag{3.6}$$

According to Theorem 2.3.9 in de Haan and Ferreira (2006), condition (3.6) itself indeed generalizes condition (2.1), since it is equivalent to

$$\forall y > 0, \ \lim_{t \to \infty} \frac{1}{A(1/\overline{F}(t|\boldsymbol{x})|\boldsymbol{x})} \left(\frac{\overline{F}(ty|\boldsymbol{x})}{\overline{F}(t|\boldsymbol{x})} - y^{-1/\gamma(\boldsymbol{x})} \right) = y^{-1/\gamma(\boldsymbol{x})} \frac{y^{\rho(\boldsymbol{x})/\gamma(\boldsymbol{x})} - 1}{\gamma(\boldsymbol{x})\rho(\boldsymbol{x})}.$$

By Theorem 2.3.3 in de Haan and Ferreira (2006), the function $|A(\cdot | \boldsymbol{x})|$ is then regularly varying with index $\rho(\boldsymbol{x})$, and this function is the primary driver of bias in typical extrapolation procedures for the estimation of extreme quantiles of heavy-tailed distributions. More on this interpretation of this condition can be found in Beirlant et al. (2004) and de Haan and Ferreira (2006) along with a number of examples of commonly used continuous distributions satisfying it. Under this condition, it can be shown that (see Lemma 8 in the Appendix):

$$\frac{e(\beta_n|\boldsymbol{x})}{e(1 - k_n/n|\boldsymbol{x})} = \left(\frac{k_n}{n(1 - \beta_n)} \right)^{\gamma(\boldsymbol{x})} \left(1 + O\left(|A(n/k_n|\boldsymbol{x})|\right) + O\left(1/e(1 - k_n/n|\boldsymbol{x})\right) \right). \tag{3.7}$$

We are now in a position to write a general asymptotic result for the class of estimators (3.5).

Theorem 2. *Assume that $(\mathcal{K})$, $(\mathcal{L})$ and $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ hold. Suppose also that $\gamma(\boldsymbol{x}) < 1/2$, $\rho(\boldsymbol{x}) < 0$ and that there exists $\delta \in (0, 1)$ with $\mathbb{E}[Y_-^{2+\delta} \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. Let $\beta_n \to 1$, $h_n \to 0$, $k_n \to \infty$ be such that $n h_n^p (1 - \beta_n) \to c < \infty$, $k_n/n \to 0$ and $k_n h_n^p \to \infty$. Assume further that $k_n h_n^{p+2} \to 0$ and*

(i) $\sqrt{k_n h_n^p}\, A(n/k_n|\boldsymbol{x}) = O(1)$ and $\sqrt{k_n h_n^p}/e(1 - k_n/n|\boldsymbol{x}) = O(1)$,

(ii) $\sqrt{k_n h_n^p}\, \log(k_n/n) \times \omega_{h_n}((1 - \delta)e(1 - k_n/n|\boldsymbol{x})|\boldsymbol{x}) \to 0$,

(iii) $\sqrt{k_n h_n^p}/\log(k_n/[n(1 - \beta_n)]) \to \infty$.

If in addition $\sqrt{k_n h_n^p}\left(\widehat{\gamma}_{k_n}(\boldsymbol{x}) - \gamma(\boldsymbol{x}) \right) \xrightarrow{d} \Gamma$, where Γ is a nondegenerate distribution, then

$$\frac{\sqrt{k_n h_n^p}}{\log(k_n/[n(1 - \beta_n)])} \left(\frac{\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

We observe that the limiting distribution of $\widehat{e}^W_{n,k_n}(\beta_n|\boldsymbol{x})$ is controlled by the asymptotic distribution of $\widehat{\gamma}_{k_n}(\boldsymbol{x})$. This is typical of Weissman-type estimators, and is here a consequence of the fact that the convergence of $\widehat{e}^W_{n,k_n}(\beta_n|\boldsymbol{x})$ is governed by that of the extrapolation factor $(k_n/[n(1-\beta_n)])^{\widehat{\gamma}_{k_n}(\boldsymbol{x})}$. The latter approximates the theoretical factor $(k_n/[n(1-\beta_n)])^{\gamma(\boldsymbol{x})}$ at a slower rate than both the rate of convergence of $\widehat{e}_n(1-k_n/n|\boldsymbol{x})$ to $e(1-k_n/n|\boldsymbol{x})$, given by Corollary 1, and the speed of convergence to 0 of the bias term that is incurred by the use of (3.4) and which can be controlled by (3.7). Controlling this bias term is precisely the purpose of the bias condition (i) in Theorem 2.

In view of Equation (3.3), another class of estimators may be introduced, exploiting the asymptotic relationship between quantiles and expectiles. Instead of noting that the conditional tail expectile function is regularly varying with index $\gamma(\boldsymbol{x})$, one may simply write

$$e(\beta_n|\boldsymbol{x}) \approx (\gamma(\boldsymbol{x})^{-1} - 1)^{-\gamma(\boldsymbol{x})} q(\beta_n|\boldsymbol{x}).$$

This approximation suggests to estimate the extreme conditional expectile $e(\beta_n|\boldsymbol{x})$ by plugging in a Weissman-type estimator of the extreme conditional quantile $q(\beta_n|\boldsymbol{x})$ and an estimator of the conditional tail index $\gamma(\boldsymbol{x})$, thus resulting in the following class of extrapolated estimators:

$$\widetilde{e}^W_{n,k_n}(\beta_n|\boldsymbol{x}) = \left(\frac{k_n}{n(1-\beta_n)}\right)^{\widehat{\gamma}_{k_n}(\boldsymbol{x})} \widehat{q}_n(1-k_n/n|\boldsymbol{x}) \left(\widehat{\gamma}_{k_n}(\boldsymbol{x})^{-1} - 1\right)^{-\widehat{\gamma}_{k_n}(\boldsymbol{x})}. \tag{3.8}$$

The following high-level result states the rate of convergence of the estimators from the class (3.8).

Theorem 3. *Assume that* $(\mathcal{K})$, $(\mathcal{L})$ *and* $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ *hold. Suppose also that* $\gamma(\boldsymbol{x}) < 1$, $\rho(\boldsymbol{x}) < 0$ *and that* $\mathbb{E}[Y_- \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. *Let* $\beta_n \to 1$, $h_n \to 0$, $k_n \to \infty$ *be such that* $nh_n^p(1-\beta_n) \to c < \infty$, $k_n/n \to 0$ *and* $k_n h_n^p \to \infty$. *Assume further that* $k_n h_n^{p+2} \to 0$ *and*

(i) $\sqrt{k_n h_n^p}\, A(n/k_n|\boldsymbol{x}) = O(1)$ *and* $\sqrt{k_n h_n^p}/e(1-k_n/n|\boldsymbol{x}) = O(1)$,

(ii) $\sqrt{k_n h_n^p}\, \log(k_n/n) \times \omega_{h_n}((1-\delta)q(1-k_n/n|\boldsymbol{x})|\boldsymbol{x}) \to 0$ *for some* $\delta \in (0,1)$,

(iii) $\sqrt{k_n h_n^p}/\log(k_n/[n(1-\beta_n)]) \to \infty$.

If in addition $\sqrt{k_n h_n^p}\,(\widehat{\gamma}_{k_n}(\boldsymbol{x}) - \gamma(\boldsymbol{x})) \xrightarrow{d} \Gamma$, *then*

$$\frac{\sqrt{k_n h_n^p}}{\log(k_n/[n(1-\beta_n)])} \left(\frac{\widetilde{e}^W_{n,k_n}(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} - 1\right) \xrightarrow{d} \Gamma.$$

It should be noted that Theorem 3 holds under integrability conditions that are weaker than those of Theorem 2; this is due to the fact that estimator (3.8) does not rely on the estimator $\widehat{e}_n(1-k_n/n|\boldsymbol{x})$. The finite-sample performances of the two classes of estimators $\widehat{e}^W_{n,k_n}(\beta_n|\boldsymbol{x})$ and $\widetilde{e}^W_{n,k_n}(\beta_n|\boldsymbol{x})$ are compared in a simulation study in Section 4.

In the previous definitions, $\widehat{\gamma}_{k_n}(\boldsymbol{x})$ is a generic estimator of $\gamma(\boldsymbol{x})$ converging at an appropriate rate. Some examples of such conditional tail index estimators may be found for instance in Daouia et al. (2013), El Methni et al. (2014) and Gardes and Stupfler (2014). We now study a handful of other alternatives exclusively based on the use of high conditional expectiles.

3.3 Estimation of the conditional tail index

In Daouia et al. (2013), the following class of estimators is proposed, based on empirical quantiles and inspired by the Pickands estimator introduced in Pickands (1975):

$$\widehat{\gamma}^P_{k_n}(\boldsymbol{x}) = \frac{1}{\log(2)} \log\left(\frac{\widehat{q}_n(1-k_n/(4n)|\boldsymbol{x}) - \widehat{q}_n(1-k_n/(2n)|\boldsymbol{x})}{\widehat{q}_n(1-k_n/(2n)|\boldsymbol{x}) - \widehat{q}_n(1-k_n/n|\boldsymbol{x})}\right).$$

It is proved therein that, under suitable conditions, $\sqrt{k_n h_n^p}\,(\widehat{\gamma}^P_{k_n}(\boldsymbol{x})/\gamma(\boldsymbol{x}) - 1)$ is asymptotically Gaussian with asymptotic variance

$$\frac{\|K\|_2^2}{g(\boldsymbol{x})} V(\boldsymbol{x}) \quad \text{where} \quad V(\boldsymbol{x}) = \frac{\left(2^{2\gamma(\boldsymbol{x})+1} + 1\right)}{\left(2^{\gamma(\boldsymbol{x})} - 1\right)^2 \log^2(2)}.$$

9

The asymptotic proportionality relationship in (3.3) motivates our first alternative estimator of $\gamma(\boldsymbol{x})$: we replace empirical conditional quantiles by empirical conditional expectiles, thus producing the estimator

$$\widehat{\gamma}_{k_n}^{(1)}(\boldsymbol{x}) = \frac{1}{\log(2)} \log \left(\frac{\widehat{e}_n(1 - k_n/(4n)|\boldsymbol{x}) - \widehat{e}_n(1 - k_n/(2n)|\boldsymbol{x})}{\widehat{e}_n(1 - k_n/(2n)|\boldsymbol{x}) - \widehat{e}_n(1 - k_n/n|\boldsymbol{x})} \right).$$

Due to the presence of the estimator $\widehat{e}_n(1 - k_n/(4n)|\boldsymbol{x})$, which has significantly larger variance than the other two quantities $\widehat{e}_n(1 - k_n/(2n)|\boldsymbol{x})$ and $\widehat{e}_n(1 - k_n/n|\boldsymbol{x})$, this estimator has a large asymptotic variance (see Theorem 4). We then propose the simpler, modified Pickands-type estimator

$$\widehat{\gamma}_{k_n}^{(2)}(\boldsymbol{x}) = \frac{1}{\log(2)} \log \left(\frac{\widehat{e}_n(1 - k_n/(2n)|\boldsymbol{x})}{\widehat{e}_n(1 - k_n/n|\boldsymbol{x})} \right).$$

This estimator has, as we will see, the advantage of having a substantially lower asymptotic variance. We finally introduce a different, arguably simpler estimator outside of the Pickands-type framework, which has the advantage of depending upon a single empirical conditional expectile. This estimator is motivated by rewriting the asymptotic relationship (3.3) as

$$\frac{\overline{F}(e(\alpha|\boldsymbol{x})|\boldsymbol{x})}{1 - \alpha} \to \gamma(\boldsymbol{x})^{-1} - 1 \quad \text{as} \quad \alpha \to 1.$$

We deduce from this relationship the estimator

$$\widehat{\gamma}_{k_n}^{(3)}(\boldsymbol{x}) = \left(1 + \frac{\widehat{\overline{F}}_n \left(\widehat{e}_n(1 - k_n/n|\boldsymbol{x})|\boldsymbol{x} \right)}{k_n/n} \right)^{-1}.$$

Our next theorem provides the asymptotic distributions of each of these three conditional tail index estimators.

Theorem 4. *Assume that* $(\mathcal{K})$ *and* $(\mathcal{L})$ *hold. Suppose also that* $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ *is fulfilled with* $\gamma(\boldsymbol{x}) < 1/2$, *and that there exists* $\delta \in (0, 1)$ *with* $\mathbb{E}[Y_-^{2+\delta} \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. *Let* (k_n) *and* (h_n) *be two sequences such that* $h_n \to 0$, $k_n/n \to 0$ *and* $k_n h_n^p \to \infty$. *Assume further that* $k_n h_n^{p+2} \to 0$ *and*

(i) $\sqrt{k_n h_n^p}/e(1 - k_n/n|\boldsymbol{x}) \to \lambda \in \mathbb{R}$,

(ii) $\sqrt{k_n h_n^p} A(n/k_n|\boldsymbol{x}) \to 0$,

(iii) $\sqrt{k_n h_n^p} \log(k_n/n) \times \omega_{h_n} ((1 - \delta)e(1 - k_n/n|\boldsymbol{x})|\boldsymbol{x}) \to 0$.

Then for $q = 1, 2, 3$:

$$\sqrt{k_n h_n^p} \left(\frac{\widehat{\gamma}_{k_n}^{(q)}(\boldsymbol{x})}{\gamma(\boldsymbol{x})} - 1, \frac{\widehat{e}_n(1 - k_n/n|\boldsymbol{x})}{e(1 - k_n/n|\boldsymbol{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(\begin{pmatrix} \lambda m^{(1)}(\boldsymbol{x}) b_q(\boldsymbol{x}) \\ 0 \end{pmatrix}, \frac{\|K\|_2^2}{g(\boldsymbol{x})} \begin{pmatrix} v_q(\boldsymbol{x}) & c_q(\boldsymbol{x}) \\ c_q(\boldsymbol{x}) & \frac{2\gamma^3(\boldsymbol{x})}{1 - 2\gamma(\boldsymbol{x})} \end{pmatrix} \right),$$

where $b_1(\boldsymbol{x}) = 0$, $b_2(\boldsymbol{x}) = \left(2^{-\gamma(\boldsymbol{x})} - 1 \right) / \log(2)$, $b_3(\boldsymbol{x}) = 1 - \gamma(\boldsymbol{x})$ *and*

$$c_1(\boldsymbol{x}) = \frac{\gamma(\boldsymbol{x}) \left(2^{\gamma(\boldsymbol{x})} - 1 \right)}{\log(2)(1 - 2\gamma(\boldsymbol{x}))}, \quad v_1(\boldsymbol{x}) = \frac{2^{\gamma(\boldsymbol{x})+1} + 2(1 + \gamma(\boldsymbol{x})) - 2^{3\gamma(\boldsymbol{x})+1}}{\log^2(2) \left(2^{\gamma(\boldsymbol{x})} - 1 \right)^2 (1 - 2\gamma(\boldsymbol{x}))} - \frac{2^{2\gamma(\boldsymbol{x})+1}}{\log^2(2) \left(2^{\gamma(\boldsymbol{x})} - 1 \right)^2},$$

$$c_2(\boldsymbol{x}) = \frac{\gamma(\boldsymbol{x}) \left(2^{\gamma(\boldsymbol{x})} - 1 \right)}{\log(2)(1 - 2\gamma(\boldsymbol{x}))}, \quad v_2(\boldsymbol{x}) = \frac{2}{\log^2(2)} \frac{1 + \gamma(\boldsymbol{x}) - 2^{\gamma(\boldsymbol{x})}}{1 - 2\gamma(\boldsymbol{x})},$$

$$c_3(\boldsymbol{x}) = \frac{\gamma^2(\boldsymbol{x})}{1 - 2\gamma(\boldsymbol{x})}, \quad v_3(\boldsymbol{x}) = \frac{\gamma(\boldsymbol{x})(1 - \gamma(\boldsymbol{x}))}{1 - 2\gamma(\boldsymbol{x})}.$$

Even though our result only highlights that the estimators $\widehat{\gamma}_{k_n}^{(q)}(\boldsymbol{x})$, for $q = 2, 3$, suffer from one specific source of bias due to the tail heaviness of the conditional distribution through the magnitude of $1/e(1 - k_n/n|\boldsymbol{x})$ (see condition (i) in Theorem 4 above), their use may actually incur further
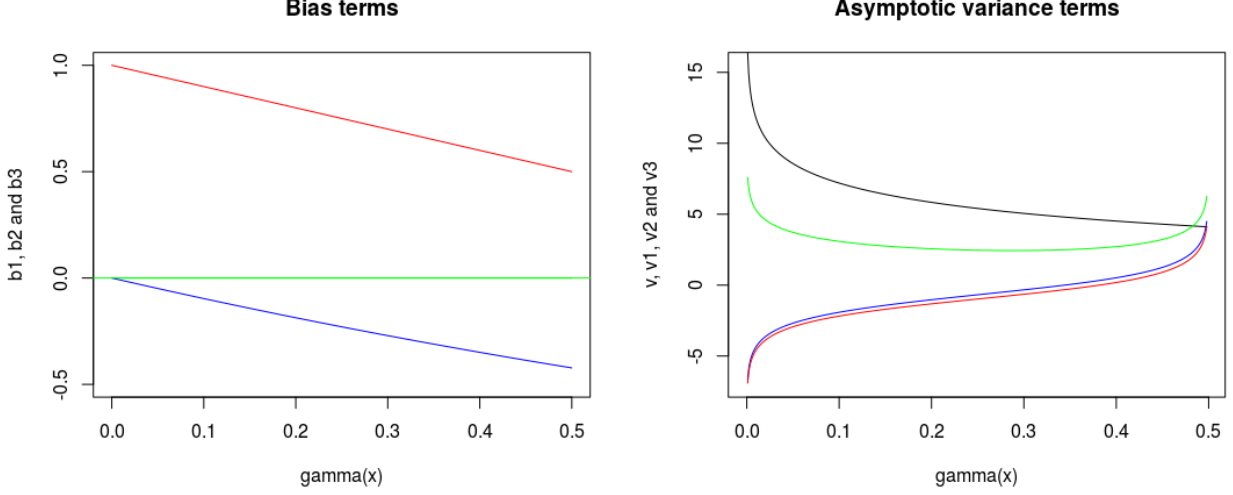
Figure 1: Left panel: bias terms $b_1(\boldsymbol{x})$ (green), $b_2(\boldsymbol{x})$ (blue) and $b_3(\boldsymbol{x})$ (red). Right panel: Asymptotic variance terms $V(\boldsymbol{x})$ (black), $v_1(\boldsymbol{x})$ (green), $v_2(\boldsymbol{x})$ (blue) and $v_3(\boldsymbol{x})$ (red), on the log-scale. x−axis: value of $\gamma(\boldsymbol{x}) \in (0, 1/2)$.

bias due to the second-order framework (see condition (ii)) and the local regression context (see condition (iii)). The latter two biases are, however, typically difficult to correct; in particular, modern correction methods for the bias due to the second-order framework involve the estimation of the (conditional) second-order parameter $\rho(\boldsymbol{x})$, which even in the unconditional case is a notoriously difficult problem (see *e.g.* the Introduction of Cai et al., 2013). By contrast, the simple expression of the bias component $\lambda m^{(1)}(\boldsymbol{x}) b_q(\boldsymbol{x})$ makes its elimination a very simple task. In addition, due to the fact that the tail conditional expectile function $t \mapsto 1/e(1 - t^{-1}|\boldsymbol{x})$ and the second-order function $t \mapsto |A(t|\boldsymbol{x})|$ are respectively regularly varying with indices $-\gamma(\boldsymbol{x}) > -1/2$ and $\rho(\boldsymbol{x})$, the source of bias due to the second-order framework can only dominate if $|\rho(\boldsymbol{x})| < \gamma(\boldsymbol{x}) < 1/2$ (and then of course $\lambda = 0$). Such cases of a second-order parameter close to 0 are situations where a Pareto tail tends to be an unreliable representation of the tail of the underlying distribution, and this is known to be very difficult to handle in the extreme value theory setup (see *e.g.* Resnick, 2007, and in particular the discussion about "Hill horror plots"). This is our rationale for specifically emphasising the bias purely due to the expectile methodology. In terms of variance, $\widehat{\gamma}_{k_n}^{(2)}(\boldsymbol{x})$ and $\widehat{\gamma}_{k_n}^{(3)}(\boldsymbol{x})$ seem to be very close, and much better than $\widehat{\gamma}_{k_n}^{(1)}(\boldsymbol{x})$ across the range $0 < \gamma(\boldsymbol{x}) < 1/2$, as Figure 1 illustrates. In other words, $\widehat{\gamma}_{k_n}^{(2)}(\boldsymbol{x})$ and $\widehat{\gamma}_{k_n}^{(3)}(\boldsymbol{x})$ both appear to be much less variable than both $\widehat{\gamma}_{k_n}^{P}(\boldsymbol{x})$ and $\widehat{\gamma}_{k_n}^{(1)}(\boldsymbol{x})$, but they do suffer from finite-sample bias.

This discussion motivates the introduction of a simple technique to deal with what is generally the source of the most substantial part of the bias in the estimators $\widehat{\gamma}_{k_n}^{(q)}(\boldsymbol{x})$ for $q = 2, 3$. We propose to correct them as follows:

$$\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x}) = \widehat{\gamma}_{k_n}^{(q)}(\boldsymbol{x}) \left(1 - \widehat{m}_n^{(1)}(\boldsymbol{x}) \frac{\widehat{b}_q(\boldsymbol{x})}{\widehat{e}_n(1 - k_n/n|\boldsymbol{x})}\right),$$

where $\widehat{b}_q(\boldsymbol{x})$ is an estimator of $b_q(\boldsymbol{x})$ obtained by plugging in the estimator $\widehat{\gamma}_{k_n}^{(q)}(\boldsymbol{x})$ in place of $\gamma(\boldsymbol{x})$. We then have the following result stating that the bias specific to the high expectile methodology has been successfully eliminated thanks to this multiplicative correction.

Corollary 2. *Under the conditions of Theorem 4, one has for $q = 2, 3$ that*

$$\sqrt{k_n h_n^p} \left(\frac{\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})}{\gamma(\boldsymbol{x})} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{||K||_2^2}{g(\boldsymbol{x})} v_q(\boldsymbol{x})\right).$$

11

We now examine the finite-sample performance of our extrapolation methodology combined with our bias reduction device in the simulation study below.

4   Simulation study

In this section, we propose to showcase our estimators on simulated samples of conditionally heavy-tailed data. For that purpose, we consider a one-dimensional covariate ($p = 1$) which is uniformly distributed on the unit interval $[0, 1]$ along with two conditional distributions for Y given $X = x \in [0, 1]$:

- A Pareto distribution with tail index $\gamma(x)$, namely

$$\forall y > 1, \ \overline{F}(y|x) = y^{-1/\gamma(x)}.$$

- A Burr type XII distribution with parameters $\gamma(x)$ and $\rho(x) < 0$, that is

$$\forall y > 0, \ \overline{F}(y|x) = \left(1 + y^{-\rho(x)/\gamma(x)}\right)^{1/\rho(x)}.$$

The Pareto distribution is the "ideal" case in the statistical analysis of heavy tails, as it defines a homogeneous function of degree $-1/\gamma(x)$ rather than merely a regularly varying function as in our basic heavy-tailed assumption (2.1). As such it does not, strictly speaking, satisfy condition $\mathcal{C}_2(\gamma(x), \rho(x), A(\cdot|x))$, although it is straightforward to see that our theorems also hold for this distribution with the convention that $A(\cdot|x) = 0$ (a similar remark is made on p.74 of de Haan and Ferreira, 2006). Our Burr type XII distribution, meanwhile, satisfies condition $\mathcal{C}_2(\gamma(x), \rho(x), A(\cdot|x))$ with $A(y|x) = \gamma(x)y^{\rho(x)}$, for $y > 0$. We specify the functions γ and ρ as

$$\forall x \in [0, 1], \ \gamma(x) = \frac{1}{4} + \frac{\sin(2\pi x)}{20} \ \text{ and } \ \rho(x) \equiv -1.$$

We simulate $N = 500$ replications of a sample of size $n = 1{,}000$ independent copies from the distribution of (X, Y). Our aim will be to estimate the conditional expectiles of level $\beta_n = 1 - 1/n = 0.999$. These conditional expectiles do not have a simple closed form, so we approximate them with high accuracy by calculating numerically the derivative of the cost function in the right-hand side of Equation (2.2), for $\alpha = \beta_n$, and then by finding the unique root of this derivative using the standard R routine uniroot.

Our general extrapolation technique described in Equation (3.5) requires the use of threshold and bandwidth sequences k_n and h_n. We choose these sequences by an adapted cross-validation algorithm about which we give details below.

4.1   Cross-validation procedure

The first step will be to select a bandwidth h_n within a data-driven grid $\mathcal{H}$. We let throughout K be the Epanechnikov kernel

$$K(t) = \frac{3}{4}\left(1 - t^2\right) \mathbb{1}_{\{|t| < 1\}}.$$

Our grid $\mathcal{H}$ of values of h is the regular mesh of size 30 of the interval $(h_{\min}, h_{\max}]$, where

$$h_{\min} := 4 \max_{x \in \mathcal{G}} \min_{1 \leq i \leq n} |x - X_i| \ \text{ and } \ h_{\max} := 1/2, \ \text{ with } \ \mathcal{G} := \{0, 0.01, 0.02, \ldots, 1\}.$$

Our selected value h_n^* of h_n is

$$h_n^* = \arg\min_{h_n \in \mathcal{H}} \sum_{\ell \in \mathcal{H}_n} \sum_{x_j \in \mathcal{G}} \left| \log\left(\frac{\widehat{e}_n^{(-j)}(0.95|x_j)}{\overline{e}_\ell(0.95|x_j)} \right) \right|,$$

where $\widehat{e}_n^{(-j)}(0.95|x_j)$ is the conditional expectile estimator (2.5) calculated at level 0.95 and based on those observations (X_i, Y_i) such that $h_n/4 < |X_i - x_j| < h_n$, while $\overline{e}_\ell(0.95|x_j)$ is the empirical

$0.95-$expectile based on the observations (X_i, Y_i) such that $|X_i - x_j| \leq \ell/4$. The idea behind this cross-validation criterion is to choose a bandwidth h_n which allows a reasonably accurate estimation of high (but not too extreme) expectiles across the whole of the interval $[0,1]$; the true value of $e(0.95|x)$ is of course unknown and is here estimated using the quantity $\overline{e}_\ell(0.95|x)$. A similar criterion may be found in Durrieu et al. (2015) in the context of extreme quantile estimation. Once h_n has been chosen as h_n^*, we select our value k_n^* of k_n to be the first local minimum of the cross-validation score

$$k \mapsto \sum_{x_j \in \mathcal{G}} \left(\overline{\gamma}_k^{(-j)}(x_j) - \check{\gamma}_H(x_j) \right)^2,$$

over k, where $\overline{\gamma}_k^{(-j)}(x_j)$ is the conditional tail index estimator $\widetilde{\gamma}_k^{(2)}(x_j)$ or $\widetilde{\gamma}_k^{(3)}(x_j)$ based on observations (X_i, Y_i) such that $h_n^*/2 < |X_i - x_j| < h_n^*$, and $\check{\gamma}_H(x_j)$ is the Hill estimator (see for instance de Haan and Ferreira, 2006) of $\gamma(x_j)$ based on those n_j observations Y_i such that $|X_i - x_j| \leq h_n^*/2$ (calculated with an effective sample size equal to $\lfloor n_j^{0.6} \rfloor$). A similar approach is developed in El Methni et al. (2014). Those choices of h_n and k_n are expected to provide fairly accurate estimates of both the intermediate conditional expectile and conditional tail index that are the building blocks for our extrapolation procedures.

4.2 Bias correction

In practice, our extrapolated estimators (3.5) and (3.8) are biased even if the estimator $\widehat{\gamma}_{k_n}(\boldsymbol{x})$ therein is chosen to be one of our bias-reduced estimators $\widetilde{\gamma}_{k_n}(\boldsymbol{x})$. This is obvious from reading Equation (3.7). A more careful analysis of the remainder term of this expansion reveals that

$$
\begin{aligned}
\left(\frac{k_n}{n(1-\beta_n)} \right)^{\gamma(\boldsymbol{x})} \frac{e(1-k_n/n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} &= 1 + m^{(1)}(\boldsymbol{x})\gamma(\boldsymbol{x})\left(\gamma(\boldsymbol{x})^{-1}-1\right)^{\gamma(\boldsymbol{x})}\left(\frac{1}{q(1-k_n/n|\boldsymbol{x})} - \frac{1}{q(\beta_n|\boldsymbol{x})} \right) \\
&\quad + O(|A(n/k_n|\boldsymbol{x})|).
\end{aligned}
$$

See Equation (6.1) in the proof of Lemma 8. Just as in Section 3.3, we propose to correct the bias due to the remainder term $1/q(1-k_n/n|\boldsymbol{x})$. In addition, and although it is asymptotically negligible compared to $1/q(1-k_n/n|\boldsymbol{x})$, we introduce a bias correction for the remainder term due to $1/q(\beta_n|\boldsymbol{x})$ (this offers, in our experience, a substantial improvement in finite-sample performance). Applied to the class of estimators $\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$, this results in the following class of estimators:

$$
\begin{aligned}
&\widehat{e}_{n,k_n}^{W,BR}(\beta_n|\boldsymbol{x}) \\
&= \widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})\left(1 + \widehat{m}_n^{(1)}(\boldsymbol{x})\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})\left(\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})^{-1}-1\right)^{\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})}\left(\frac{1}{\widehat{q}_n^W(\beta_n|\boldsymbol{x})} - \frac{1}{\widehat{q}_n(1-k_n/n|\boldsymbol{x})} \right) \right),
\end{aligned}
$$

where $\widehat{q}_n^W(\beta_n|\boldsymbol{x})$ is the Weissman-type estimator of $q(\beta_n|\boldsymbol{x})$ deduced by extrapolating $\widehat{q}_n(1-k_n/n|\boldsymbol{x})$ and estimating $\gamma(\boldsymbol{x})$ with $\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})$. It is straightforward to show that $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|\boldsymbol{x})$ and $\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$ have the same asymptotic properties. This principle can also be applied to the construction of a bias-reduced version of $\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$ as follows:

$$
\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|\boldsymbol{x}) = \widetilde{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})\left(1 + \widehat{m}_n^{(1)}(\boldsymbol{x})\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})\frac{\left(\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})^{-1}-1\right)^{\widetilde{\gamma}_{k_n}^{(q)}(\boldsymbol{x})}}{\widehat{q}_n^W(\beta_n|\boldsymbol{x})} \right).
$$

These two versions of our estimators are therefore used in the sequel in place of $\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$ and $\widetilde{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$. To illustrate the performance of our methodologies, we compare them with the naive extrapolated versions $\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$ and $\widetilde{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})$ where the estimator $\widehat{\gamma}_{k_n}(\boldsymbol{x})$ is the local Hill estimator calculated using the k_n top observations Y_i whose covariates X_i have a distance from $\boldsymbol{x}$ which is not greater than h_n, with the pair (h_n, k_n) chosen using the cross-validation procedure explained in Section 4.1.

13

4.3 Results

Figures 2 and 3 give an overview of the performances of our estimators. A general comment is that the simpler estimators $\widehat{e}_{n,k_n}^{W}(\beta_n|x)$ and $\widetilde{e}_{n,k_n}^{W}(\beta_n|x)$ which do not feature any bias reduction perform overall quite poorly, especially for large values of $\gamma(x)$. Another general remark is that the variability of each estimator increases as $\gamma(x)$ increases, as expected in view of the monotonicity of the asymptotic variances v_2 and v_3 in Theorem 4.

Our estimators $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ and $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$ seem to have fairly good performance overall. There is not clearly a best estimator among the four versions that were tested; in the Pareto case, the estimator $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$ using $\widetilde{\gamma}_{k_n}^{(2)}(x)$ seems to perform well but has a fairly high variance, while its counterpart $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ has a lower variance but appears to underestimate the true extreme conditional expectile for the highest values of $\gamma(x)$ (around $x = 1/4$). The higher finite-sample variability of $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$ comes from the fact that it is a bias-reduced version of $\widetilde{e}_{n,k_n}^{W}(\beta_n|x)$, which is constructed using the multiplicative factor $(\widetilde{\gamma}_{k_n}^{(q)}(x)^{-1} - 1)^{\widetilde{\gamma}_{k_n}^{(q)}(x)}$. This quantity becomes very large as the estimator $\widetilde{\gamma}_{k_n}^{(q)}(x)$ gets closer to 1. Making an error in the estimation of $\gamma(x)$ is therefore more detrimental to the stability of $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$ than it is to that of $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$. In addition, the versions of these estimators which use $\widetilde{\gamma}_{k_n}^{(3)}(x)$ instead seem to perform well for the highest values of $\gamma(x)$ but still overestimate the true value for the lowest values of $\gamma(x)$. In the case of the Burr distribution, the estimator $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ with $\widetilde{\gamma}_{k_n}^{(2)}(x)$ performs very well and better than its counterpart $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$. The versions of these estimators based on $\widetilde{\gamma}_{k_n}^{(3)}(x)$ maintain a respectable level of accuracy.

5 Real data example

We study here a data set on motorcycle insurance. The data was collected from the former Swedish insurance provider Wasa, and comprises data on motorcycle insurance policies and claims over the period 1994-1998. The data set, available from `www.math.su.se/GLMbook` and the R package `insuranceData` and analyzed in Ohlsson and Johansson (2010), contains among others the claim severity Y (defined as the ratio of claim cost by number of claims for each given policyholder) in SEK, and the age X of the policyholder in years. This results in $n = 670$ pairs (X_i, Y_i). Our goal is to infer some of the features of the conditional extremes of this data set.

We first carry out, in Figure 4, visual checks of whether the heavy-tailed assumption makes sense for this sample of data. The boxplot and histogram of the Y_i both give descriptive evidence that Y has a heavy right tail. To further confirm that the heavy-tailed framework is appropriate, we drew a quantile-quantile plot of the weighted log-spacings within the top of the data against the quantiles of the unit exponential distribution. Formally, let $Y_{1,n} \leq Y_{2,n} \leq \cdots \leq Y_{n,n}$ denote the order statistics of the sample $(Y_1, \ldots, Y_n)$. Let $Z_{i,n} = i \log(Y_{n-i+1,n}/Y_{n-i,n})$, $1 \leq i \leq n-1$, denote the weighted log-spacings computed from the consecutive top order statistics. It is known that, if Y is heavy-tailed with tail index γ then, for low i, the $Z_{i,n}$ are approximately independent copies of an exponential random variable with mean γ (see *e.g.* Beirlant et al., 2004, pp.109–110). The bottom panel of Figure 4 therefore gives a quantile-quantile plot of the $Z_{i,n}$ for $1 \leq i \leq \lfloor n/5 \rfloor$ ($\lfloor \cdot \rfloor$ denotes the floor function) versus the exponential distribution. The relationship in this quantile-quantile plot is approximately linear, which constitutes further evidence that the heavy tail assumption on Y makes sense.
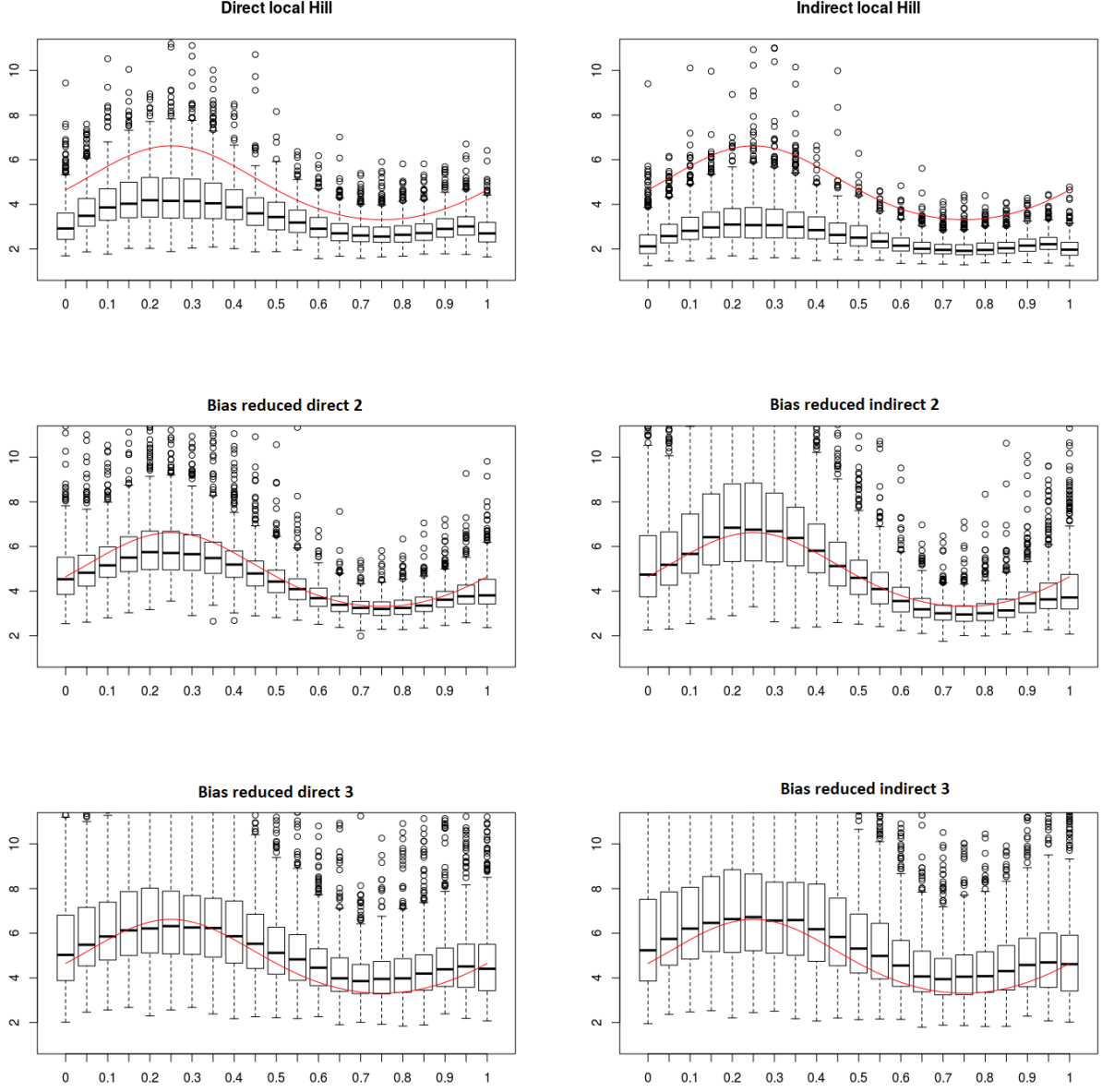
Figure 2: Simulation results in the case of the Pareto distribution, with $\beta_n = 1 - 1/n$. Top panels: Naive extrapolated estimators $\widehat{e}_{n,k_n}^{W}(\beta_n|x)$ (left) and $\widetilde{e}_{n,k_n}^{W}(\beta_n|x)$ (right) based on local Hill conditional tail index estimation. Middle panels: Bias-reduced extrapolated estimators $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ (left) and $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$ (right) based on the conditional tail index estimator $\widetilde{\gamma}_{k_n}^{(2)}(x)$. Bottom panels: Bias-reduced extrapolated estimators $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ (left) and $\widetilde{e}_{n,k_n}^{W,BR}(\beta_n|x)$ (right) based on the conditional tail index estimator $\widetilde{\gamma}_{k_n}^{(3)}(x)$. The red curve is the true extreme conditional expectile curve $x \mapsto e(\beta_n|x)$ to be estimated.
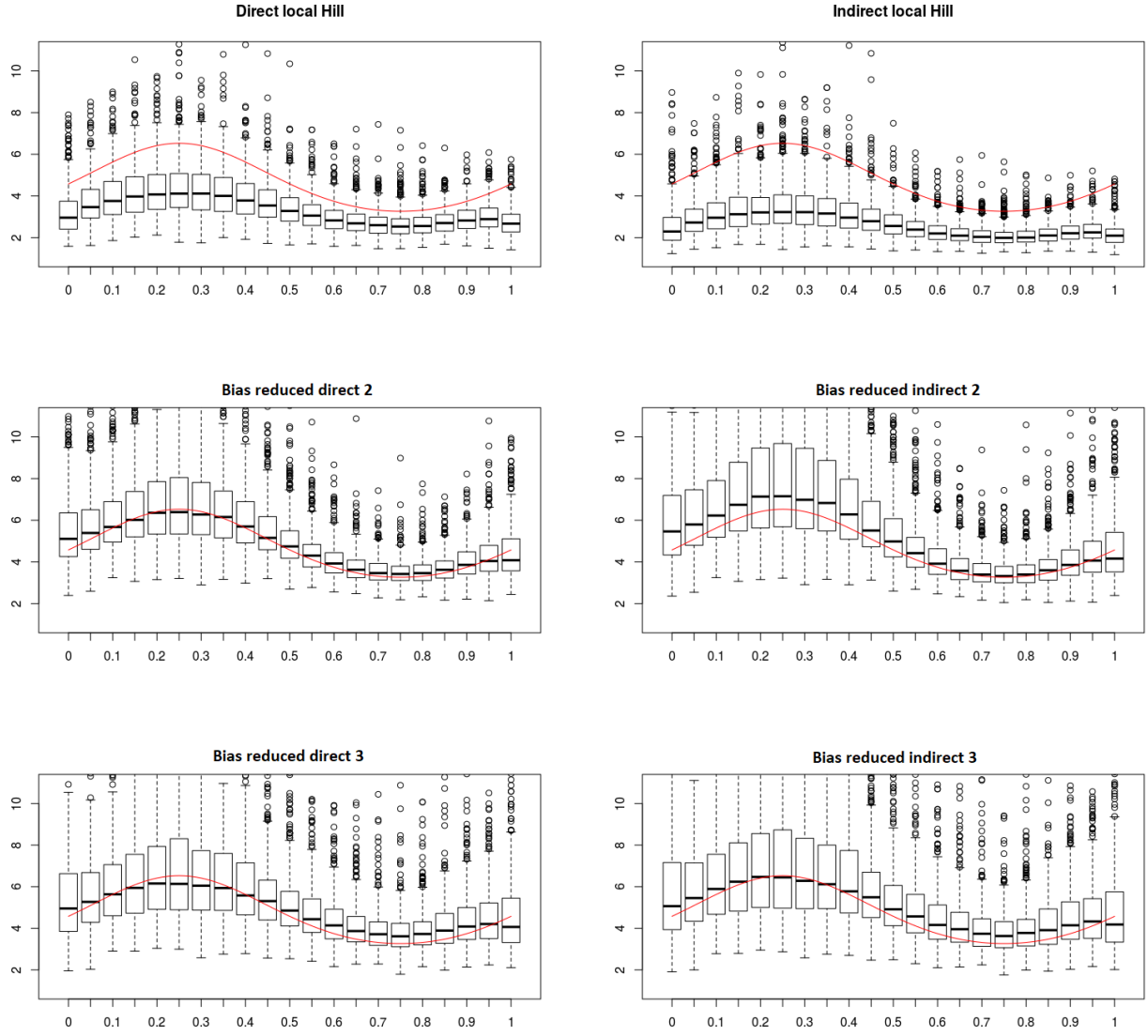
Figure 3: As in Figure 2, but with a Burr distribution where $\rho(x) \equiv -1$.
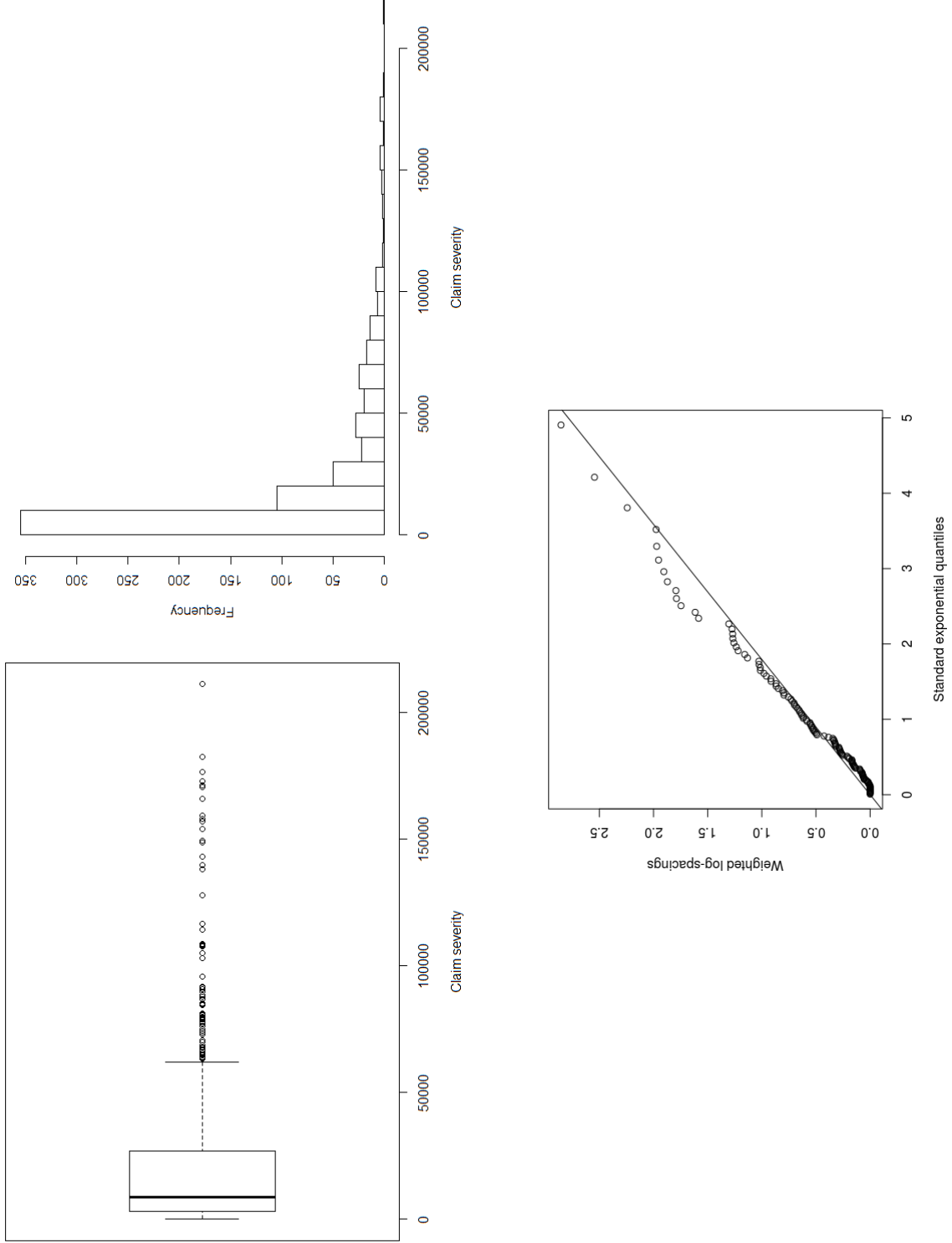
Figure 4: Swedish motorcycle insurance data analysis. Top row, left panel: boxplot of the claim severities Y_i. Top row, right panel: histogram of the Y_i. Bottom row: quantile-quantile plot of weighted log-spacings $Z_{i,n} = i \log(Y_{n-i+1,n}/Y_{n-i,n})$ (for $1 \leq i \leq \lfloor n/5 \rfloor$) versus standard exponential quantiles, where the solid line has a slope equal to the Hill estimator of the Y_i computed on the top $\lfloor n/5 \rfloor$ data points.

We therefore carry out our analysis of conditional extremes using the methodology described in Sections 4.1 and 4.2. Our cross-validation procedure ($\mathcal{H}$ being in this case a finer, regular grid of size 100 of the interval $(h_{\min}, 40]$) yields $h_n^* = 7.82 \approx (X_{\max} - X_{\min})/8$, which motivates us to estimate an extreme conditional expectile at level $\beta_n = 1 - 8/n \approx 0.988$. To simplify the presentation, we only calculate and represent our estimate $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ of the extreme conditional expectile using the estimator $\widetilde{\gamma}_{k_n}^{(2)}(x)$. We compare this estimate to the extreme conditional quantile estimate $\widehat{q}_n^W(\beta_n|x)$ based on the estimator $\widetilde{\gamma}_{k_n}^{(2)}(x)$. The cross-validation procedure on k_n then selects $k_n^* = 190$. The results are represented in Figure 5.

It can be seen in this Figure that the extreme conditional quantile estimate shows a slight increase on the interval $x \in [20, 60]$ but arguably stays quite flat and does not appear to flag any difference in conditional tail behaviour as x varies. By contrast, the extreme conditional expectile estimate is markedly different for young policyholders and middle-aged policyholders. This reflects the change in the conditional behavior of extremes in the data: it can be seen in Figure 5 that claim severity for older policyholders is typically much more right-skewed than for younger policyholders. In this sense, the extreme conditional expectile methodology captures information about conditional tail behavior that the extreme conditional quantile cannot identify.
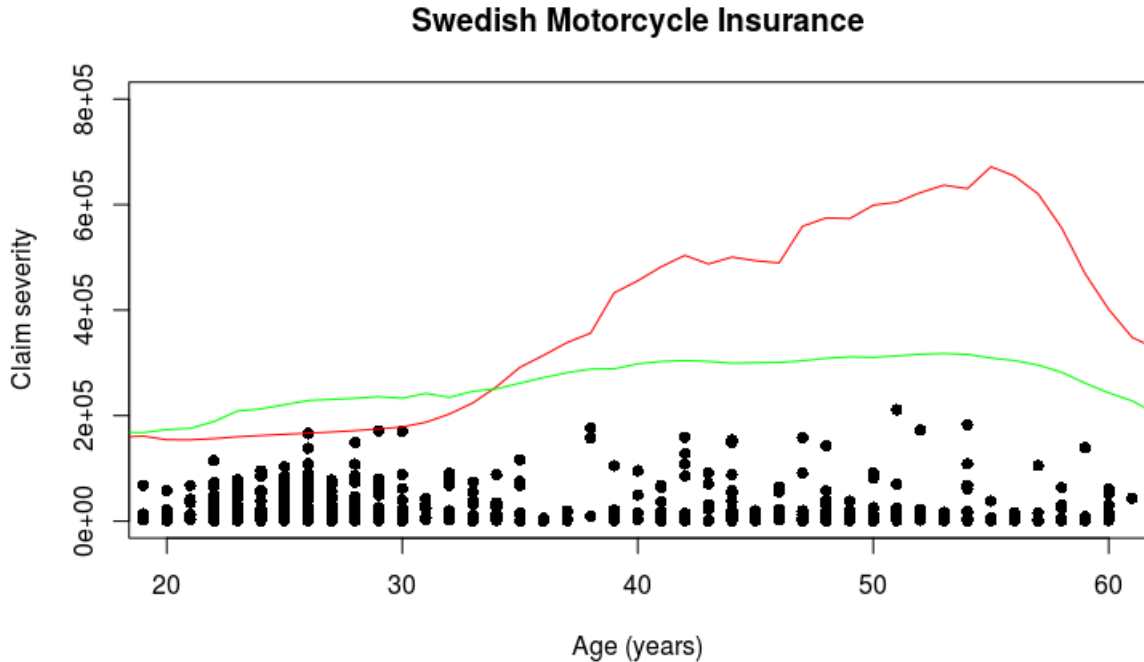

Figure 5: Swedish motorcycle insurance data analysis. x−axis: age of policyholder, y−axis: claim severity. Green curve: extreme conditional quantile estimate $x \mapsto \widehat{q}_n^W(\beta_n|x)$ using $\widetilde{\gamma}_{k_n}^{(2)}(x)$, red curve: extreme conditional expectile estimate $\widehat{e}_{n,k_n}^{W,BR}(\beta_n|x)$ based on $\widetilde{\gamma}_{k_n}^{(2)}(x)$.

6 Appendix

6.1 Preliminary results

The first lemma controls the relative oscillation of $\overline{F}(y_n|\cdot)$ in a neighborhood of $\boldsymbol{x}$. It is a straightforward consequence of a combination of the definition of $\omega_{h_n}(y|\boldsymbol{x})$ in (3.1) with a first-order Taylor expansion of the exponential function.

Lemma 1. *Let $y_n \to \infty$, $h_n \to 0$ be such that $\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n) \to 0$. Then uniformly in $\boldsymbol{x}' \in B(\boldsymbol{x}, h_n)$,*

$$\frac{\overline{F}(y_n|\boldsymbol{x}')}{\overline{F}(y_n|\boldsymbol{x})} - 1 = O\left(\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n)\right).$$

In the next lemma and throughout, $\mathcal{B}(\cdot, \cdot)$ denotes the Beta function (see p.258 of Abramowitz and Stegun, 1966, for details).

Lemma 2. *Assume that $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ holds.*

  i) For all $a \in [0, 1/\gamma(\boldsymbol{x}))$,

$$\psi^{(a)}(y|\boldsymbol{x}) = \frac{\mathcal{B}\left(a+1, \gamma(\boldsymbol{x})^{-1} - a\right)}{\gamma(\boldsymbol{x})} g(\boldsymbol{x})y^a\overline{F}(y|\boldsymbol{x})(1 + o(1)) \text{ as } y \to \infty.$$

  ii) If $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ holds, then for all $a \in [0, 1/\gamma(\boldsymbol{x}))$,

$$\psi^{(a)}(y|\boldsymbol{x}) = g(\boldsymbol{x})y^a\overline{F}(y|\boldsymbol{x})\left\{ \frac{\mathcal{B}\left(a+1, \gamma(\boldsymbol{x})^{-1} - a\right)}{\gamma(\boldsymbol{x})} \right.$$
$$\left. + A(1/\overline{F}(y|\boldsymbol{x}))\, a \int_0^1 (1-u)^{a-1}u^{1/\gamma(\boldsymbol{x})-a-1}\frac{u^{-\rho(\boldsymbol{x})/\gamma(\boldsymbol{x})} - 1}{\gamma(\boldsymbol{x})\rho(\boldsymbol{x})}du(1 + o(1)) \right\}$$

  as $y \to \infty$, with the convention $(u^t - 1)/t = \log u$ when $t = 0$.

Proof. In the case $a = 0$, $\psi^{(0)}(y|\boldsymbol{x}) = \overline{F}(y|\boldsymbol{x})g(\boldsymbol{x})$ and there is nothing to prove. We focus on the case $a > 0$. Note that $H(t) = (t-1)^a \mathbb{1}_{\{t\geq 1\}}$ defines an absolutely continuous function with nonnegative derivative $H'(t) = a(t-1)^{a-1}\mathbb{1}_{\{t\geq 1\}}$ almost everywhere and we have, for any b with $a < b$,

$$\int_1^\infty H'(t)t^{-b}dt = a \int_1^\infty (t-1)^{a-1}t^{-b}dt = a\,\mathcal{B}(a, b-a) = b\,\mathcal{B}(a+1, b-a)$$

by using the change of variables $u = 1 - t^{-1}$. The result then follows by applying Lemma 1(i) and (ii) in Daouia et al. (2019a) with this choice of H. $\qquad\square$

The corollary below follows from Lemma 2. Statement *iii*) is obtained from Karamata's theorem (see Theorem B.1.5 in de Haan and Ferreira, 2006).

Corollary 3. *Suppose $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ holds with $\gamma(\boldsymbol{x}) < 1$ and $\mathbb{E}[Y_- \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. Then:*

  i) We have the asymptotic proportionality relationship

$$\lim_{y\to\infty} \frac{\overline{E}(y|\boldsymbol{x})}{\overline{F}(y|\boldsymbol{x})} = \frac{\gamma(\boldsymbol{x})}{1 - \gamma(\boldsymbol{x})}.$$

  ii) The function $\overline{E}(\cdot|\boldsymbol{x})$ is regularly varying, i.e.

$$\forall y > 0, \ \lim_{t\to\infty} \frac{\overline{E}(ty|\boldsymbol{x})}{\overline{E}(t|\boldsymbol{x})} = y^{-1/\gamma(\boldsymbol{x})} \text{ and thus } \lim_{t\to\infty} \frac{e(1 - 1/(ty)|\boldsymbol{x})}{e(1 - 1/t|\boldsymbol{x})} = y^{\gamma(\boldsymbol{x})}.$$

iii) The function $\overline{E}(\cdot|\boldsymbol{x})$ is continuously differentiable with

$$\lim_{y\to\infty}\frac{y\overline{E}'(y|\boldsymbol{x})}{\overline{E}(y|\boldsymbol{x})} = \lim_{y\to\infty}\frac{y\overline{F}'(y|\boldsymbol{x})}{\overline{F}(y|\boldsymbol{x})} = -\frac{1}{\gamma(\boldsymbol{x})}.$$

Lemma 3. *Assume that* $(\mathcal{L})$ *and* $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ *hold. Let* $y_n \to \infty$, $h_n \to 0$ *be such that* $\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n) \to 0$. *Then, uniformly in* $\boldsymbol{x}' \in B(\boldsymbol{x},h_n)$ *and for all* $a \in [0,1/\gamma(\boldsymbol{x}))$,

$$\frac{\psi^{(a)}(y_n|\boldsymbol{x}')}{\psi^{(a)}(y_n|\boldsymbol{x})} - 1 = O(h_n) + O(\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n)).$$

Proof. For $a = 0$ we have, in view of Lemma 1 and condition $(\mathcal{L})$,

$$\frac{\psi^{(0)}(y_n|\boldsymbol{x}')}{\psi^{(0)}(y_n|\boldsymbol{x})} = \frac{\overline{F}(y_n|\boldsymbol{x}')g(\boldsymbol{x}')}{\overline{F}(y_n|\boldsymbol{x})g(\boldsymbol{x})} = 1 + O(h_n) + O\left(\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n)\right)$$

uniformly in $\boldsymbol{x}' \in B(\boldsymbol{x},h_n)$, hence the first result. We turn to the second result. For any $a \in (0,1/\gamma(\boldsymbol{x}))$, we have, by an integration by parts,

$$\frac{\psi^{(a)}(y_n|\boldsymbol{x})}{g(\boldsymbol{x})} = a\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})dz.$$

From Equation (3.1), we find

$$a\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})z^{-\omega_{h_n}(y_n|\boldsymbol{x})}dz \leq \frac{\psi^{(a)}(y_n|\boldsymbol{x}')}{g(\boldsymbol{x}')} \leq a\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})z^{\omega_{h_n}(y_n|\boldsymbol{x})}dz.$$

This implies

$$\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})(z^{-\omega_{h_n}(y_n|\boldsymbol{x})}-1)dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})dz} \leq \frac{\psi^{(a)}(y_n|\boldsymbol{x}')}{\psi^{(a)}(y_n|\boldsymbol{x})}\frac{g(\boldsymbol{x})}{g(\boldsymbol{x}')} - 1$$

$$\leq \frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})(z^{\omega_{h_n}(y_n|\boldsymbol{x})}-1)dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})dz}$$

and thus, by the inequality $|e^t - 1| \leq |t|e^{|t|}$,

$$\left|\frac{\psi^{(a)}(y_n|\boldsymbol{x}')}{\psi^{(a)}(y_n|\boldsymbol{x})}\frac{g(\boldsymbol{x})}{g(\boldsymbol{x}')} - 1\right| \leq \omega_{h_n}(y_n|\boldsymbol{x})\frac{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})z^{\omega_{h_n}(y_n|\boldsymbol{x})}\log(z)dz}{\int_{y_n}^{\infty}(z-y_n)^{a-1}\overline{F}(z|\boldsymbol{x})dz}.$$

With the change of variables $z = ty_n$, we find

$$\left|\frac{\psi^{(a)}(y_n|\boldsymbol{x}')}{\psi^{(a)}(y_n|\boldsymbol{x})}\frac{g(\boldsymbol{x})}{g(\boldsymbol{x}')} - 1\right| \leq \omega_{h_n}(y_n|\boldsymbol{x})y_n^{\omega_{h_n}(y_n|\boldsymbol{x})}\frac{\int_1^{\infty}(t-1)^{a-1}\dfrac{\overline{F}(ty_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}t^{\omega_{h_n}(y_n|\boldsymbol{x})}\log(ty_n)dt}{\int_1^{\infty}(t-1)^{a-1}\dfrac{\overline{F}(ty_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}dt}.$$

Using Potter bounds (see Theorem 1.5.6 p.25 in Bingham et al., 1989) we find, for n large enough and $\delta > 0$ so small that $a + \delta < 1/\gamma(\boldsymbol{x})$,

$$\frac{\int_1^{\infty}(t-1)^{a-1}\dfrac{\overline{F}(ty_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}t^{\omega_{h_n}(y_n|\boldsymbol{x})}\log(ty_n)dt}{\int_1^{\infty}(t-1)^{a-1}\dfrac{\overline{F}(ty_n|\boldsymbol{x})}{\overline{F}(y_n|\boldsymbol{x})}dt} \leq (1+\log(y_n))\frac{\int_1^{\infty}(t-1)^{a-1}t^{\delta-1/\gamma(\boldsymbol{x})}dt}{\int_1^{\infty}(t-1)^{a-1}t^{-\delta-1/\gamma(\boldsymbol{x})}dt}$$

and the right-hand side is a $O(\log(y_n))$. Since $\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n) \to 0$, we get

$$\left|\frac{\psi^{(a)}(y_n|\boldsymbol{x}')}{\psi^{(a)}(y_n|\boldsymbol{x})}\frac{g(\boldsymbol{x})}{g(\boldsymbol{x}')} - 1\right| = O(\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n)).$$

Using condition $(\mathcal{L})$ concludes the proof. $\qquad\square$

Lemma 4. *Assume that $(\mathcal{K})$ and $(\mathcal{L})$ hold. If $h_n \to 0$ is such that $nh_n^p \to \infty$, then*

$$\mathbb{E}\left[\widehat{g}_n(\boldsymbol{x})\right] = g(\boldsymbol{x}) + O(h_n) \ \text{ and } \ \mathbb{V}ar\left[\widehat{g}_n(\boldsymbol{x})\right] = \frac{g(\boldsymbol{x})\|K\|_2^2}{nh_n^p}(1 + o(1)).$$

Moreover,

$$\text{if } \ \mathbb{E}[|Y| \,|\, \boldsymbol{X} = \boldsymbol{x}] < \infty \ \text{ then } \ \mathbb{E}\left[\widehat{m}_n^{(1)}(\boldsymbol{x})\widehat{g}_n(\boldsymbol{x})\right] = m^{(1)}(\boldsymbol{x})g(\boldsymbol{x}) + O(h_n)$$

$$\text{and if } \ \mathbb{E}[Y^2 \,|\, \boldsymbol{X} = \boldsymbol{x}] < \infty \ \text{ then } \ \mathbb{V}ar\left[\widehat{m}_n^{(1)}(\boldsymbol{x})\widehat{g}_n(\boldsymbol{x})\right] = \frac{m^{(2)}(\boldsymbol{x})g(\boldsymbol{x})\|K\|_2^2}{nh_n^p}(1 + o(1)).$$

Proof. The first two results are classical (see Parzen, 1962, for a proof). For n large enough, the expectation $\mathbb{E}[\widehat{m}_n^{(1)}(\boldsymbol{x})\widehat{g}_n(\boldsymbol{x})]$ may be written as

$$\frac{1}{h_n^p}\mathbb{E}\left[YK\left(\frac{\boldsymbol{x} - \boldsymbol{X}}{h_n}\right)\right] = \frac{1}{h_n^p}\int_{\mathbb{R}^p} m^{(1)}(\boldsymbol{t})g(\boldsymbol{t})K\left(\frac{\boldsymbol{x} - \boldsymbol{t}}{h_n}\right)d\boldsymbol{t} = \int_S m^{(1)}(\boldsymbol{x} - \boldsymbol{u}h_n)g(\boldsymbol{x} - \boldsymbol{u}h_n)K(\boldsymbol{u})d\boldsymbol{u}.$$

Assumption $(\mathcal{L})$ entails $m^{(1)}(\boldsymbol{x} - \boldsymbol{u}h_n) = m^{(1)}(\boldsymbol{x}) + O(h_n)$ and $g(\boldsymbol{x} - \boldsymbol{u}h_n) = g(\boldsymbol{x}) + O(h_n)$ uniformly on $\boldsymbol{u} \in S$, hence the second identity. The final asymptotic equivalent is obtained through similar calculations. $\qquad\square$

Lemma 5. *Assume that $(\mathcal{K})$, $(\mathcal{L})$ and $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ hold. Let $y_n \to \infty$, $h_n \to 0$ be such that $nh_n^p \to \infty$ and $\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n) \to 0$ as $n \to \infty$. Then, for any $a \in [0, 1/\gamma(\boldsymbol{x}))$ and $b \geq 1$,*

$$\mathbb{E}\left[(Y - y_n)^a K^b\left(\frac{\boldsymbol{x} - \boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y > y_n\}}\right]$$
$$= h_n^p \psi^{(a)}(y_n|\boldsymbol{x})\int_S K^b(\boldsymbol{u})d\boldsymbol{u}(1 + O(h_n) + O(\omega_{h_n}(y_n|\boldsymbol{x})\log(y_n))).$$

Proof. The result is immediate by noting that

$$\mathbb{E}\left[(Y - y_n)^a K^b\left(\frac{\boldsymbol{x} - \boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y > y_n\}}\right] = \int_{\mathbb{R}^p}\int_{y_n}^{\infty}(z - y_n)^a K^b\left(\frac{\boldsymbol{x} - \boldsymbol{t}}{h_n}\right)f(z|\boldsymbol{t})g(\boldsymbol{t})\,d\boldsymbol{t}\,dz$$
$$= h_n^p\int_S \psi^{(a)}(y_n|\boldsymbol{x} - h_n\boldsymbol{u})K^b(\boldsymbol{u})d\boldsymbol{u}$$

(with the change of variables $\boldsymbol{t} = \boldsymbol{x} - h_n\boldsymbol{u}$) and using Lemma 3. $\qquad\square$

Lemma 6. *Assume that $(\mathcal{K})$, $(\mathcal{L})$ and $\mathcal{C}_1(\gamma(\boldsymbol{x}))$ hold. Suppose also that $\gamma(\boldsymbol{x}) < 1/2$ and that there exists $\delta \in (0, 1)$ with $\mathbb{E}[Y_-^{2+\delta} \,|\, \boldsymbol{X} = \boldsymbol{x}] < \infty$. Let $y_n \to \infty$, $h_n \to 0$ and $b_n = \theta y_n(1 + o(1))$, where $\theta > 0$. Assume further that $nh_n^p\overline{F}(y_n|\boldsymbol{x}) \to \infty$, $nh_n^{p+2}\overline{F}(y_n|\boldsymbol{x}) \to \infty$ and*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\ \omega_{h_n}((1 - \delta)(\theta \wedge 1)y_n|\boldsymbol{x})\log(y_n) \to 0.$$

Then if, for all $j \in \{1, \ldots, J\}$, $y_{n,j} = \tau_j^{-\gamma(\boldsymbol{x})}y_n(1 + o(1))$ with $0 < \tau_1 < \tau_2 < \ldots < \tau_J \leq 1$, one has

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left\{\left(\frac{\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} - 1\right)_{1 \leq j \leq J}, \left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})}{\psi^{(0)}(b_n|\boldsymbol{x})} - 1\right)\right\} \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}_{J+1}, \frac{\|K\|_2^2}{g(\boldsymbol{x})}\boldsymbol{V}(\boldsymbol{x})\right)$$

where $\boldsymbol{V}(\boldsymbol{x})$ is the symmetric matrix with entries

$$\begin{cases}
V_{j,l}(\boldsymbol{x}) = \dfrac{1 - \gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}\tau_l^{-1}\left[\dfrac{1}{1 - 2\gamma(\boldsymbol{x})}\left(\dfrac{\tau_j}{\tau_l}\right)^{-\gamma(\boldsymbol{x})} - 1\right], \ j, l \in \{1, \ldots, J\}^2, \ j \leq l, \\[4mm]
V_{j,J+1}(\boldsymbol{x}) = \dfrac{\gamma(\boldsymbol{x})\theta^{1/\gamma(\boldsymbol{x})}\left(\theta \vee \tau_j^{-\gamma(\boldsymbol{x})}\right)^{1-1/\gamma(\boldsymbol{x})} + (1 - \gamma(\boldsymbol{x}))\left(\theta \vee \tau_j^{-\gamma(\boldsymbol{x})} - \tau_j^{-\gamma(\boldsymbol{x})}\right)}{\gamma(\boldsymbol{x})\tau_j^{1-\gamma(\boldsymbol{x})}}, \ j \in \{1, \ldots, J\}, \\[4mm]
V_{J+1,J+1}(\boldsymbol{x}) = \theta^{1/\gamma(\boldsymbol{x})}.
\end{cases}$$

Proof. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J, \beta_{J+1}) \in \mathbb{R}^{J+1}$. One has:

$$\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \sum_{j=1}^{J} \beta_j \left(\frac{\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} - 1 \right) + \beta_{J+1} \left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})}{\psi^{(0)}(b_n|\boldsymbol{x})} - 1 \right)$$

$$= \sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left\{ \sum_{j=1}^{J} \beta_j \left(\frac{\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x}) - \mathbb{E}\left[\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})\right]}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} \right) + \beta_{J+1} \left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x}) - \mathbb{E}\left[\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})\right]}{\psi^{(0)}(b_n|\boldsymbol{x})} \right) \right\}$$

$$+ \sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left\{ \sum_{j=1}^{J} \beta_j \left(\frac{\mathbb{E}\left[\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})\right]}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} - 1 \right) + \beta_{J+1} \left(\frac{\mathbb{E}\left[\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})\right]}{\psi^{(0)}(b_n|\boldsymbol{x})} - 1 \right) \right\}.$$

According to Lemma 5,

$$\begin{cases} \mathbb{E}\left[\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})\right] = \psi^{(0)}(b_n|\boldsymbol{x})(1 + O(\omega_{h_n}(b_n|\boldsymbol{x})\log(b_n)) + O(h_n)), \\ \mathbb{E}\left[\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})\right] = \psi^{(1)}(y_{n,j}|\boldsymbol{x})(1 + O(\omega_{h_n}(y_{n,j}|\boldsymbol{x})\log(y_{n,j})) + O(h_n)). \end{cases}$$

Noticing that for n large enough, $y_{n,j} > y_n(1-\delta)$, we obtain $\omega_{h_n}(y_{n,j}|\boldsymbol{x}) \leq \omega_{h_n}((1-\delta)y_n|\boldsymbol{x})$, $\forall j \in \{1, \ldots, J\}$. Similarly $b_n > \theta y_n(1-\delta)$ and thus $\omega_{h_n}(b_n|\boldsymbol{x}) \leq \omega_{h_n}((1-\delta)\theta y_n|\boldsymbol{x})$. Moreover, $\log(y_{n,j}) = O(\log(y_n))$ for any $j \in \{1, \ldots, J\}$ and $\log(b_n) = O(\log(y_n))$. Therefore

$$\sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left\{ \sum_{j=1}^{J} \beta_j \left(\frac{\mathbb{E}\left[\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})\right]}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} - 1 \right) + \beta_{J+1} \left(\frac{\mathbb{E}\left[\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})\right]}{\psi^{(0)}(b_n|\boldsymbol{x})} - 1 \right) \right\} = o(1).$$

We now focus on the asymptotic distribution of

$$Z_n = \sqrt{nh_n^p \overline{F}(y_n|\boldsymbol{x})} \left\{ \sum_{j=1}^{J} \beta_j \left(\frac{\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x}) - \mathbb{E}\left[\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})\right]}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} \right) + \beta_{J+1} \left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x}) - \mathbb{E}\left[\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})\right]}{\psi^{(0)}(b_n|\boldsymbol{x})} \right) \right\}.$$

We clearly have $\mathbb{E}[Z_n] = 0$. In addition, $\mathbb{V}\text{ar}[Z_n] = \overline{F}(y_n|\boldsymbol{x})\boldsymbol{\beta}^\top \boldsymbol{B}^{(n)} \boldsymbol{\beta}$, where $\boldsymbol{B}^{(n)}$ is the symmetric matrix-valued sequence having entries

$$\begin{cases} B_{j,l}^{(n)} = \dfrac{\text{cov}\left((Y-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}, (Y-y_{n,l})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,l}\}}\right)}{h_n^p \psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(1)}(y_{n,l}|\boldsymbol{x})}, \quad j,l \in \{1, \ldots, J\},\ j \leq l, \\[3mm] B_{j,J+1}^{(n)} = \dfrac{\text{cov}\left((Y-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}, K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>b_n\}}\right)}{h_n^p \psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(0)}(b_n|\boldsymbol{x})}, \quad j \in \{1, \ldots, J\}, \\[3mm] B_{J+1,J+1}^{(n)} = \dfrac{\mathbb{V}\text{ar}\left[K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>b_n\}}\right]}{h_n^p [\psi^{(0)}(b_n|\boldsymbol{x})]^2}. \end{cases}$$

Let us first focus, for $j \leq l$, on the term $B_{j,l}^{(n)} = A_{j,l}^{(n)}/[\psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(1)}(y_{n,l}|\boldsymbol{x})]$. Since $y_{n,j} > y_{n,l}$ for n large enough, we find:

$$\begin{aligned} A_{j,l}^{(n)} &= \frac{1}{h_n^p}\mathbb{E}\left[(Y-y_{n,j})(Y-y_{n,l})K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right] \\ &\quad - \frac{1}{h_n^p}\mathbb{E}\left[(Y-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right]\mathbb{E}\left[(Y-y_{n,l})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,l}\}}\right]. \end{aligned}$$

According to Lemma 5, the second term is equal to $h_n^p \psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(1)}(y_{n,l}|\boldsymbol{x})(1+o(1))$. It thus remains to focus on the first term of $A_{j,l}^{(n)}$ which we rewrite as

$$\frac{1}{h_n^p}\mathbb{E}\left[(Y-y_{n,j})^2 K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right] + (y_{n,j}-y_{n,l})\frac{1}{h_n^p}\mathbb{E}\left[(Y-y_{n,j})K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right].$$

22

Using Lemma 5, we get

$$\frac{1}{h_n^p}\mathbb{E}\left[(Y-y_{n,j})^2 K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right] = \psi^{(2)}(y_{n,j}|\boldsymbol{x})\|K\|_2^2(1+o(1))$$

and

$$\frac{(y_{n,j}-y_{n,l})}{h_n^p}\mathbb{E}\left[(Y-y_{n,j})K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right] = \left(\tau_j^{-\gamma(\boldsymbol{x})}-\tau_l^{-\gamma(\boldsymbol{x})}\right)\|K\|_2^2\psi^{(1)}(y_{n,j}|\boldsymbol{x})y_n(1+o(1)).$$

Besides, Lemma 2 and Corollary 3 $i)$ provide, for any j,

$$\begin{cases} \psi^{(1)}(y_{n,j}|\boldsymbol{x}) = \dfrac{\gamma(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}g(\boldsymbol{x})\tau_j^{-\gamma(\boldsymbol{x})}y_n\tau_j\overline{F}(y_n|\boldsymbol{x})(1+o(1)), \\[3mm] \psi^{(2)}(y_{n,j}|\boldsymbol{x}) = \dfrac{2\gamma(\boldsymbol{x})^2}{(1-2\gamma(\boldsymbol{x}))(1-\gamma(\boldsymbol{x}))}g(\boldsymbol{x})\tau_j^{-2\gamma(\boldsymbol{x})}y_n^2\tau_j\overline{F}(y_n|\boldsymbol{x})(1+o(1)). \end{cases}$$

Straightforward calculations then yield:

$$B_{j,l}^{(n)} = \frac{\|K\|_2^2}{g(\boldsymbol{x})}\frac{1-\gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}\tau_l^{-1}\left[\frac{1}{1-2\gamma(\boldsymbol{x})}\frac{\tau_j^{-\gamma(\boldsymbol{x})}}{\tau_l^{-\gamma(\boldsymbol{x})}}-1\right]\frac{1}{\overline{F}(y_n|\boldsymbol{x})}(1+o(1)).$$

We now deal with $B_{j,J+1}^{(n)}$ for $j\in\{1,\ldots,J\}$, which can be rewritten as

$$\frac{\mathbb{E}\left[(Y-y_{n,j})K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\vee b_n\}}\right] - \mathbb{E}\left[(Y-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}\right]\mathbb{E}\left[K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>b_n\}}\right]}{h_n^p\psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(0)}(b_n|\boldsymbol{x})}.$$

Using Lemma 5, the second term in the numerator equals $h_n^{2p}\psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(0)}(b_n|\boldsymbol{x})(1+o(1))$ and the first term can be rewritten

$$\mathbb{E}\left[(Y-y_{n,j})K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\vee b_n\}}\right]$$
$$= h_n^p\|K\|_2^2\left[\psi^{(1)}(y_{n,j}\vee b_n|\boldsymbol{x}) + (y_{n,j}\vee b_n - y_{n,j})\psi^{(0)}(b_n|\boldsymbol{x})\right](1+o(1)).$$

Combining Lemma 2 with Corollary 3 $i)$ and the asymptotic equivalent $\overline{F}(y_{n,j}\vee b_n|\boldsymbol{x})/\overline{F}(y_n|\boldsymbol{x}) = (\theta\vee\tau_j^{-\gamma(\boldsymbol{x})})^{-1/\gamma(\boldsymbol{x})}(1+o(1))$, we get

$$B_{j,J+1}^{(n)}$$
$$= \frac{\psi^{(1)}(y_{n,j}\vee b_n|\boldsymbol{x}) + (y_{n,j}\vee b_n - y_{n,j})\psi^{(0)}(b_n|\boldsymbol{x})}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})\psi^{(0)}(b_n|\boldsymbol{x})}\|K\|_2^2(1+o(1))$$
$$= \frac{\|K\|_2^2}{g(\boldsymbol{x})}\frac{\gamma(\boldsymbol{x})\left(\theta\vee\tau_j^{-\gamma(\boldsymbol{x})}\right)^{1-1/\gamma(\boldsymbol{x})} + (1-\gamma(\boldsymbol{x}))\left(\theta\vee\tau_j^{-\gamma(\boldsymbol{x})}-\tau_j^{-\gamma(\boldsymbol{x})}\right)\theta^{-1/\gamma(\boldsymbol{x})}}{\gamma(\boldsymbol{x})\tau_j^{1-\gamma(\boldsymbol{x})}\theta^{-1/\gamma(\boldsymbol{x})}}\frac{1}{\overline{F}(y_n|\boldsymbol{x})}(1+o(1)).$$

Finally, using Lemmas 2, 5 and Corollary 3 $i)$, the variance term $B_{J+1,J+1}^{(n)}$ is clearly

$$\frac{\mathbb{E}\left[K^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>b_n\}}\right] - \left\{\mathbb{E}\left[K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>b_n\}}\right]\right\}^2}{h_n^p[\psi^{(0)}(b_n|\boldsymbol{x})]^2} = \frac{\|K\|_2^2}{g(\boldsymbol{x})}\theta^{1/\gamma(\boldsymbol{x})}\frac{1}{\overline{F}(y_n|\boldsymbol{x})}(1+o(1)).$$

Therefore, $\mathbb{V}\mathrm{ar}[Z_n]\to\|K\|_2^2\boldsymbol{\beta}^\top\boldsymbol{V}(\boldsymbol{x})\boldsymbol{\beta}/g(\boldsymbol{x})$, where $\boldsymbol{V}(\boldsymbol{x})$ is given in the statement of the lemma. It only remains to prove the asymptotic normality of Z_n. We thus write $Z_n = \sum_{i=1}^n Z_{i,n}$, where

$$Z_{i,n} = \frac{\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}}{nh_n^p}\sum_{j=1}^J\beta_j\frac{(Y_i-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\mathbb{1}_{\{Y_i>y_{n,j}\}} - \mathbb{E}\left[(Y_i-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\mathbb{1}_{\{Y_i>y_{n,j}\}}\right]}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})}$$

$$+ \frac{\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}}{nh_n^p}\beta_{J+1}\frac{K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\mathbb{1}_{\{Y_i>b_n\}} - \mathbb{E}\left[K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\mathbb{1}_{\{Y_i>b_n\}}\right]}{\psi^{(0)}(b_n|\boldsymbol{x})}.$$

We prove that there is $\delta > 0$ such that $n\mathbb{E}|Z_{1,n}|^{2+\delta} \to 0$ as $n \to \infty$; the result then follows by the Lyapunov central limit theorem (see Theorem 27.3 p.362 of Billingsley, 1995). For that purpose, note that, if $T_1, \ldots, T_q$ have finite $(2+\delta)$-moments,

$$\left\{\mathbb{E}\left|\sum_{i=1}^{q}[T_i - \mathbb{E}(T_i)]\right|^{2+\delta}\right\}^{1/(2+\delta)} \leq \sum_{i=1}^{q}\{\mathbb{E}|T_i|^{2+\delta}\}^{1/(2+\delta)} + \mathbb{E}|T_i| \leq 2q\max_{1\leq i\leq q}\{\mathbb{E}|T_i|^{2+\delta}\}^{1/(2+\delta)}$$

by the triangle inequality for the standard norm in the space of random variables with finite $(2+\delta)$-moment. Therefore,

$$\mathbb{E}|Z_{1,n}|^{2+\delta} = \mathrm{O}\left(\left(\frac{\overline{F}(y_n|\boldsymbol{x})}{nh_n^p}\right)^{1+\delta/2}\left\{\max_{1\leq j\leq J}\mathbb{E}\left|\frac{(Y-y_{n,j})K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>y_{n,j}\}}}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})}\right|^{2+\delta} + \mathbb{E}\left|\frac{K\left(\frac{\boldsymbol{x}-\boldsymbol{X}}{h_n}\right)\mathbb{1}_{\{Y>b_n\}}}{\psi^{(0)}(b_n|\boldsymbol{x})}\right|^{2+\delta}\right\}\right).$$

Combine now, for δ small enough, Lemmas 2 and 5 with the asymptotic equivalents $\overline{F}(b_n|\boldsymbol{x}) = \theta^{-1/\gamma(\boldsymbol{x})}\overline{F}(y_n|\boldsymbol{x})(1+o(1))$ and $\overline{F}(y_{n,j}|\boldsymbol{x}) = \tau_j\overline{F}(y_n|\boldsymbol{x})(1+o(1))$ to get

$$n\mathbb{E}|Z_{1,n}|^{2+\delta} = \mathrm{O}\left(\left[nh_n^p\overline{F}(y_n|\boldsymbol{x})\right]^{-\delta/2}\right),$$

which clearly tends to 0 as $n \to \infty$. The result is proved. $\qquad\square$

Lemma 7. *Under the conditions of Lemma 6 and with $\boldsymbol{V}(\boldsymbol{x})$ given in Lemma 6,*

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left\{\left(\frac{\widehat{\overline{E}}_n(y_{n,j}|\boldsymbol{x})}{\overline{E}(y_{n,j}|\boldsymbol{x})} - 1\right)_{1\leq j\leq J}, \left(\frac{\widehat{\overline{F}}_n(b_n|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1\right)\right\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}_{J+1}, \frac{\|K\|_2^2}{g(\boldsymbol{x})}\boldsymbol{V}(\boldsymbol{x})\right).$$

Proof. Set

$$R_{n,j} := \frac{\widehat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x}) - \dfrac{\widehat{m}_n^{(1)}(\boldsymbol{x})\widehat{g}_n(\boldsymbol{x}) - m^{(1)}(\boldsymbol{x})g(\boldsymbol{x})}{y_{n,j}} + 2\dfrac{\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x}) - \psi^{(1)}(y_{n,j}|\boldsymbol{x})}{y_{n,j}}}{g(\boldsymbol{x}) - \dfrac{m^{(1)}(\boldsymbol{x})g(\boldsymbol{x})}{y_{n,j}} + \dfrac{2\psi^{(1)}(y_{n,j}|\boldsymbol{x})}{y_{n,j}}}.$$

We may write

$$\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\sum_{j=1}^{J}\beta_j\left(\frac{\widehat{\overline{E}}_n(y_{n,j}|\boldsymbol{x})}{\overline{E}(y_{n,j}|\boldsymbol{x})} - 1\right) + \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\beta_{J+1}\left(\frac{\widehat{\overline{F}}_n(b_n|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1\right)$$

$$= \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\sum_{j=1}^{J}\beta_j\left(\frac{\widehat{\psi}_n^{(1)}(y_{n,j}|\boldsymbol{x})}{\psi^{(1)}(y_{n,j}|\boldsymbol{x})} - 1\right)(1+R_{n,j})^{-1} + \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\beta_{J+1}\left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})}{\psi^{(0)}(b_n|\boldsymbol{x})} - 1\right)\frac{g(\boldsymbol{x})}{\widehat{g}_n(\boldsymbol{x})}$$

$$+ \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\sum_{j=1}^{J}\beta_j\left((1+R_{n,j})^{-1} - 1\right) + \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\beta_{J+1}\left(\frac{g(\boldsymbol{x})}{\widehat{g}_n(\boldsymbol{x})} - 1\right).$$

A combination of Lemmas 4 and 6 directly entails

$$\forall j \in \{1, \ldots, J\}, \ R_{n,j} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_n^p}}\right) \text{ and } \frac{\widehat{g}_n(\boldsymbol{x})}{g(\boldsymbol{x})} - 1 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_n^p}}\right)$$

and the result then follows by applying Lemma 6. $\qquad\square$

Lemma 8. *Assume that $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ holds. Suppose also that $\gamma(\boldsymbol{x}) < 1$, $\rho(\boldsymbol{x}) < 0$ and $\mathbb{E}[Y_- \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. Let $\beta_n \to 1$, $k_n \to \infty$ be such that $k_n/n \to 0$ and $n(1-\beta_n)/k_n \to 0$. Then:*

$$\left(\frac{k_n}{n(1-\beta_n)}\right)^{\gamma(\boldsymbol{x})}\frac{e(1-k_n/n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} = 1 + B(1-k_n/n|\boldsymbol{x}) + O(|A(n/k_n|\boldsymbol{x})|),$$

where $B(\tau|\boldsymbol{x}) = \gamma(\boldsymbol{x})\dfrac{(\gamma(\boldsymbol{x})^{-1} - 1)^{\gamma(\boldsymbol{x})}}{q(\tau|\boldsymbol{x})}(m^{(1)}(\boldsymbol{x}) + o(1)) = \dfrac{\gamma(\boldsymbol{x})}{e(\tau|\boldsymbol{x})}(m^{(1)}(\boldsymbol{x}) + o(1))$ *as* $\tau \to 1$.

Proof. Write

$$\left(\frac{k_n}{n(1-\beta_n)}\right)^{\gamma(\boldsymbol{x})} \frac{e(1-k_n/n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} = \left(\frac{k_n}{n(1-\beta_n)}\right)^{\gamma(\boldsymbol{x})} \frac{q(1-k_n/n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})} \times \frac{e(1-k_n/n|\boldsymbol{x})}{q(1-k_n/n|\boldsymbol{x})} \times \frac{q(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})}.$$

Following the proof of Theorem 4.3.8 in de Haan and Ferreira (2006), we get

$$\left(\frac{k_n}{n(1-\beta_n)}\right)^{\gamma(\boldsymbol{x})} \frac{q(1-k_n/n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})} = 1 + O(|A(n/k_n|\boldsymbol{x})|).$$

Moreover, according to Proposition 1 in Daouia et al. (2019b),

$$\begin{cases} \dfrac{q(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} = (\gamma(\boldsymbol{x})^{-1} - 1)^{\gamma(\boldsymbol{x})} \left(1 - B(\beta_n|\boldsymbol{x}) + O(|A((1-\beta_n)^{-1}|\boldsymbol{x})|)\right), \\[2mm] \dfrac{e(1-k_n/n|\boldsymbol{x})}{q(1-k_n/n|\boldsymbol{x})} = (\gamma(\boldsymbol{x})^{-1} - 1)^{-\gamma(\boldsymbol{x})} \left(1 + B(1-k_n/n|\boldsymbol{x}) + O(|A(n/k_n|\boldsymbol{x})|)\right). \end{cases} \tag{6.1}$$

Since $n(1-\beta_n) = o(k_n)$ and $\rho(\boldsymbol{x}) < 0$, we find $B(\beta_n|\boldsymbol{x}) = o(|B(1-k_n/n|\boldsymbol{x})|)$ and $A((1-\beta_n)^{-1}|\boldsymbol{x}) = o(|A(n/k_n|\boldsymbol{x})|)$, hence the result. $\square$

Lemma 9. *Assume that $\mathcal{C}_2(\gamma(\boldsymbol{x}), \rho(\boldsymbol{x}), A(\cdot|\boldsymbol{x}))$ holds. Suppose also that $\gamma(\boldsymbol{x}) < 1$ and $\mathbb{E}[Y_- \mid \boldsymbol{X} = \boldsymbol{x}] < \infty$. If $k_n \to \infty$ and $k_n/n \to 0$, then:*

$$\frac{e(1-k_n/(4n)|\boldsymbol{x}) - e(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x}) - e(1-k_n/n|\boldsymbol{x})} = 2^{\gamma(\boldsymbol{x})} \left(1 + o(1/e(1-k_n/n|\boldsymbol{x})) + O(|A(n/k_n|\boldsymbol{x})|)\right).$$

Proof. With the notation of Lemma 8, Proposition 1 in Daouia et al. (2019b), applied to the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$, provides:

$$(\gamma(\boldsymbol{x})^{-1} - 1)^{\gamma(\boldsymbol{x})} \frac{e(\tau|\boldsymbol{x})}{q(\tau|\boldsymbol{x})} = 1 + B(\tau|\boldsymbol{x}) + O(|A((1-\tau)^{-1}|\boldsymbol{x})|).$$

The result follows by using this asymptotic expansion and noticing/recalling that $t \mapsto |B(1-t^{-1}|\boldsymbol{x})| \propto 1/e(1-t^{-1}|\boldsymbol{x})$ and $t \mapsto |A(t|\boldsymbol{x})|$ are regularly varying with respective indices $-\gamma(\boldsymbol{x})$ and $\rho(\boldsymbol{x})$. $\square$

6.2 Proofs of main results

6.2.1 Proof of Theorem 1

Let us denote $\sigma_n = 1/\sqrt{nh_n^p(1-\alpha_n)}$, $\boldsymbol{z} = (z_1, \ldots, z_J, t)$ and focus on the probability

$$\begin{aligned} \Phi_n(\boldsymbol{z}) &= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\sigma_n^{-1}\left(\frac{\widehat{e}_n(\alpha_{n,j}|\boldsymbol{x})}{e(\alpha_{n,j}|\boldsymbol{x})} - 1\right) \le z_j\right\} \cap \left\{\sigma_n^{-1}\left(\frac{\widehat{q}_n(a_n|\boldsymbol{x})}{q(a_n|\boldsymbol{x})} - 1\right) \le t\right\}\right) \\ &= \mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\widehat{e}_n(\alpha_{n,j}|\boldsymbol{x}) \le e(\alpha_{n,j}|\boldsymbol{x})(1 + z_j\sigma_n)\right\} \cap \left\{\widehat{q}_n(a_n|\boldsymbol{x}) \le q(a_n|\boldsymbol{x})(1 + t\sigma_n)\right\}\right). \end{aligned}$$

By noticing that $1 - \alpha_{n,j} = \overline{E}(e(\alpha_{n,j}|\boldsymbol{x})|\boldsymbol{x})$ and $1 - a_n = \overline{F}(q(a_n|\boldsymbol{x})|\boldsymbol{x})$, and using that for all y, α, $\widehat{e}_n(\alpha|\boldsymbol{x}) \le y \Leftrightarrow \widehat{\overline{E}}_n(y|\boldsymbol{x}) \le 1 - \alpha$ and $\widehat{q}_n(\alpha|\boldsymbol{x}) \le y \Leftrightarrow \widehat{\overline{F}}_n(y|\boldsymbol{x}) \le 1 - \alpha$, we find that $\Phi_n(\boldsymbol{z})$ is equal to

$$\mathbb{P}\left(\bigcap_{j=1}^{J}\left\{\widehat{\overline{E}}_n(e(\alpha_{n,j}|\boldsymbol{x})(1 + z_j\sigma_n)|\boldsymbol{x}) \le \overline{E}(e(\alpha_{n,j}|\boldsymbol{x})|\boldsymbol{x})\right\} \cap \left\{\widehat{\overline{F}}_n(q(a_n|\boldsymbol{x})(1 + t\sigma_n)|\boldsymbol{x}) \le \overline{F}(q(a_n|\boldsymbol{x})|\boldsymbol{x})\right\}\right).$$

Letting $y_n = e(\alpha_n|\boldsymbol{x})$, $y_{n,j} = e(\alpha_{n,j}|\boldsymbol{x})(1 + z_j\sigma_n)$ and $b_n = q(a_n|\boldsymbol{x})(1 + t\sigma_n)$, we have

$$\begin{aligned} \Phi_n(\boldsymbol{z}) = \mathbb{P}\Bigg(&\bigcap_{j=1}^{J}\left\{\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\overline{E}}_n(y_{n,j}|\boldsymbol{x})}{\overline{E}(y_{n,j}|\boldsymbol{x})} - 1\right) \le \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\overline{E}(e(\alpha_{n,j}|\boldsymbol{x})|\boldsymbol{x})}{\overline{E}(y_{n,j}|\boldsymbol{x})} - 1\right)\right\} \\ &\cap \left\{\sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\widehat{\overline{F}}_n(b_n|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1\right) \le \sqrt{nh_n^p\overline{F}(y_n|\boldsymbol{x})}\left(\frac{\overline{F}(q(a_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1\right)\right\}\Bigg). \end{aligned}$$

25

By Corollary 3 *iii*), we may, for any $j \in \{1, \ldots, J\}$, use a Taylor expansion to obtain that there exists $\theta_{n,j} \in [0,1]$ with:

$$\overline{E}(y_{n,j}|\boldsymbol{x}) = 1 - \alpha_{n,j} + z_j \sigma_n e(\alpha_{n,j}|\boldsymbol{x}) \overline{E}'(e(\alpha_{n,j}|\boldsymbol{x}) + \theta_{n,j} z_j \sigma_n e(\alpha_{n,j}|\boldsymbol{x})|\boldsymbol{x}).$$

According to Corollary 3 *ii*),

$$\overline{E}(y_{n,j}|\boldsymbol{x}) = 1 - \alpha_{n,j} + z_j \sigma_n e(\alpha_{n,j}|\boldsymbol{x}) \overline{E}'(e(\alpha_{n,j}|\boldsymbol{x})|\boldsymbol{x})(1 + o(1)) = (1 - \alpha_{n,j}) \left(1 - \frac{z_j \sigma_n}{\gamma(\boldsymbol{x})}(1 + o(1)) \right).$$

Since $\sqrt{n h_n^p \overline{F}(y_n|\boldsymbol{x})} = \sigma_n^{-1} \sqrt{\gamma(\boldsymbol{x})^{-1} - 1}(1 + o(1))$ (see Corollary 3 *i*)), we have

$$\forall j \in \{1, \ldots, J\}, \ \sqrt{n h_n^p \overline{F}(y_n|\boldsymbol{x})} \left(\frac{\overline{E}(e(\alpha_{n,j}|\boldsymbol{x})|\boldsymbol{x})}{\overline{E}(y_{n,j}|\boldsymbol{x})} - 1 \right) = \frac{z_j}{\gamma(\boldsymbol{x})} \sqrt{\frac{1 - \gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}}(1 + o(1)).$$

By using the same arguments, we also get

$$\sqrt{n h_n^p \overline{F}(y_n|\boldsymbol{x})} \left(\frac{\overline{F}(q(a_n|\boldsymbol{x})|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1 \right) = \frac{t}{\gamma(\boldsymbol{x})} \sqrt{\frac{1 - \gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}}(1 + o(1)).$$

Finally, $y_{n,j} = e(\alpha_{n,j}|\boldsymbol{x})(1 + z_j \sigma_n) = \tau_j^{-\gamma(\boldsymbol{x})} y_n(1 + o(1))$ (see Corollary 3 *i*)) and $b_n = q(a_n|\boldsymbol{x})(1 + t\sigma_n) = \left(\gamma(\boldsymbol{x})^{-1} - 1 \right)^{\gamma(\boldsymbol{x})} \tau^{-\gamma(\boldsymbol{x})} y_n(1 + o(1))$ (by combining Corollary 3 *ii*) and Proposition 1 in Daouia et al. (2019b)). Thus, for n large enough, $\omega_{h_n}(y_{n,j}|\boldsymbol{x}) \leq \omega_{h_n}((1 - \delta)e(1 - k_n/n|\boldsymbol{x})|\boldsymbol{x})$ and $\omega_{h_n}(b_n|\boldsymbol{x}) \leq \omega_{h_n}((1 - \delta)q(a_n|\boldsymbol{x})|\boldsymbol{x})$. Moreover, $\log(y_{n,j}) = O(\log(y_n))$ and $\log(b_n) = O(\log(y_n))$. Noting that $\log(y_n) = O(|\log(1 - \alpha_n)|)$ by the regular variation property of the tail expectile function $t \mapsto e(1 - t^{-1}|\boldsymbol{x})$, Lemma 7 can then be applied (with $\theta = \tau^{-\gamma(\boldsymbol{x})} \left(\gamma(\boldsymbol{x})^{-1} - 1 \right)^{\gamma(\boldsymbol{x})}$) to conclude the proof. $\square$

6.2.2 Proof of Theorem 2

The key is to write

$$
\begin{aligned}
\log \left(\frac{\widehat{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} \right) &= \log \left(\frac{k_n}{n(1 - \beta_n)} \right) (\widehat{\gamma}_{k_n}(\boldsymbol{x}) - \gamma(\boldsymbol{x})) + \log \left(\frac{\widehat{e}_n(1 - k_n/n|\boldsymbol{x})}{e(1 - k_n/n|\boldsymbol{x})} \right) \\
&\quad + \log \left(\left[\frac{k_n}{n(1 - \beta_n)} \right]^{\gamma(\boldsymbol{x})} \frac{e(1 - k_n/n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})} \right).
\end{aligned}
$$

The conclusion now follows from our assumptions on $\widehat{\gamma}_{k_n}(\boldsymbol{x})$, Theorem 1 and Lemma 8, together with a straightforward application of the delta-method. $\square$

6.2.3 Proof of Theorem 3

We first briefly explain why $\widehat{q}_n(1 - k_n/n|\boldsymbol{x})$ is still $\sqrt{k_n h_n^p}$–consistent under our assumptions which are weaker than those of Theorem 1. Inspecting the proof of Theorem 1 reveals that

$$\mathbb{P} \left(\sqrt{k_n h_n^p} \left(\frac{\widehat{q}_n(1 - k_n/n|\boldsymbol{x})}{q(1 - k_n/n|\boldsymbol{x})} - 1 \right) \leq t \right) = \mathbb{P} \left(\sqrt{k_n h_n^p} \left(\frac{\widehat{\overline{F}}_n(b_n|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1 \right) \leq \frac{t}{\gamma(\boldsymbol{x})}(1 + o(1)) \right)$$

with $b_n = q(1 - k_n/n|\boldsymbol{x})(1 + t/\sqrt{k_n h_n^p})$. Write then

$$\sqrt{k_n h_n^p} \left(\frac{\widehat{\overline{F}}_n(b_n|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})} - 1 \right) = \sqrt{k_n h_n^p} \left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})}{\psi^{(0)}(b_n|\boldsymbol{x})} - 1 \right) \frac{g(\boldsymbol{x})}{\widehat{g}_n(\boldsymbol{x})} + \sqrt{k_n h_n^p} \left(\frac{g(\boldsymbol{x})}{\widehat{g}_n(\boldsymbol{x})} - 1 \right).$$

Inspecting the proof of Lemma 6 shows that under our assumptions,

$$\sqrt{k_n h_n^p}\left(\frac{\widehat{\psi}_n^{(0)}(b_n|\boldsymbol{x})}{\psi^{(0)}(b_n|\boldsymbol{x})}-1\right)=O_{\mathbb{P}}(1)+O\left(\sqrt{k_n h_n^p}\left[h_n+\omega_{h_n}(b_n|\boldsymbol{x})\log(b_n)\right]\right)$$

$$=O_{\mathbb{P}}(1)+O\left(\sqrt{k_n h_n^p}\left[h_n+|\log(k_n/n)|\times\omega_{h_n}((1-\delta)q(1-k_n/n|\boldsymbol{x})|\boldsymbol{x})\right]\right)$$

$$=O_{\mathbb{P}}(1).$$

Inspecting then the proof of Lemma 7 and applying Lemma 4 shows that

$$\sqrt{k_n h_n^p}\left(\frac{\widehat{\overline{F}}_n(b_n|\boldsymbol{x})}{\overline{F}(b_n|\boldsymbol{x})}-1\right)=O_{\mathbb{P}}(1)\quad\text{and so}\quad\sqrt{k_n h_n^p}\left(\frac{\widehat{q}_n(1-k_n/n|\boldsymbol{x})}{q(1-k_n/n|\boldsymbol{x})}-1\right)=O_{\mathbb{P}}(1).$$

The final key point is to write

$$\log\left(\frac{\widetilde{e}_{n,k_n}^W(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})}\right)=\log\left(\frac{k_n}{n(1-\beta_n)}\right)(\widehat{\gamma}_{k_n}(\boldsymbol{x})-\gamma(\boldsymbol{x}))+\log\left(\frac{\widehat{q}_n(1-k_n/n|\boldsymbol{x})}{q(1-k_n/n|\boldsymbol{x})}\right)$$

$$+\log\left(\frac{(\widehat{\gamma}_{k_n}(\boldsymbol{x})^{-1}-1)^{-\widehat{\gamma}_{k_n}(\boldsymbol{x})}}{(\gamma(\boldsymbol{x})^{-1}-1)^{-\gamma(\boldsymbol{x})}}\right)+\log\left(\left[\frac{k_n}{n(1-\beta_n)}\right]^{\gamma(\boldsymbol{x})}\frac{q(1-k_n/n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}\right)$$

$$+\log\left((\gamma(\boldsymbol{x})^{-1}-1)^{-\gamma(\boldsymbol{x})}\frac{q(\beta_n|\boldsymbol{x})}{e(\beta_n|\boldsymbol{x})}\right).$$

The conclusion now follows from our assumptions on $\widehat{\gamma}_{k_n}(\boldsymbol{x})$, the convergence of $\widehat{q}_n(1-k_n/n|\boldsymbol{x})$, condition $\mathcal{C}_2(\gamma(\boldsymbol{x}),\rho(\boldsymbol{x}),A(\cdot|\boldsymbol{x}))$ and Proposition 1 in Daouia et al. (2019b), together with a straightforward application of the delta-method. □

6.2.4 Proof of Theorem 4

Asymptotic distribution of $\widehat{\gamma}_{k_n}^{(1)}(\boldsymbol{x})$. We start by focusing on the asymptotic distribution of

$$\sqrt{k_n h_n^p}\left(\frac{\widehat{e}_n(1-k_n/(4n)|\boldsymbol{x})-\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})}{\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})-\widehat{e}_n(1-k_n/n|\boldsymbol{x})}-2^{\gamma(\boldsymbol{x})}\right)=A_n+B_n,$$

where

$$A_n=\sqrt{k_n h_n^p}\left(\frac{\widehat{e}_n(1-k_n/(4n)|\boldsymbol{x})-\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(4n)|\boldsymbol{x})-e(1-k_n/(2n)|\boldsymbol{x})}\times\frac{e(1-k_n/(2n)|\boldsymbol{x})-e(1-k_n/n|\boldsymbol{x})}{\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})-\widehat{e}_n(1-k_n/n|\boldsymbol{x})}-1\right)$$

$$\times\frac{e(1-k_n/(4n)|\boldsymbol{x})-e(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x})-e(1-k_n/n|\boldsymbol{x})},$$

$$B_n=\sqrt{k_n h_n^p}\left(\frac{e(1-k_n/(4n)|\boldsymbol{x})-e(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x})-e(1-k_n/n|\boldsymbol{x})}-2^{\gamma(\boldsymbol{x})}\right).$$

According to Lemma 9, $B_n\to 0$. We then focus on A_n and we note that, by Theorem 1 and some straightforward algebra,

$$\frac{e(1-k_n/(2n)|\boldsymbol{x})-e(1-k_n/n|\boldsymbol{x})}{\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})-\widehat{e}_n(1-k_n/n|\boldsymbol{x})}=1+O_{\mathbb{P}}\left(1/\sqrt{k_n h_n^p}\right).$$

Applying Lemma 9 again, we get

$$A_n=2^{\gamma(\boldsymbol{x})}\sqrt{k_n h_n^p}\frac{e(1-k_n/(4n)|\boldsymbol{x})}{e(1-k_n/(4n)|\boldsymbol{x})-e(1-k_n/(2n)|\boldsymbol{x})}\left(\frac{\widehat{e}_n(1-k_n/(4n)|\boldsymbol{x})}{e(1-k_n/(4n)|\boldsymbol{x})}-1\right)(1+o_{\mathbb{P}}(1))$$

$$-2^{\gamma(\boldsymbol{x})}\sqrt{k_n h_n^p}\frac{e(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(4n)|\boldsymbol{x})-e(1-k_n/(2n)|\boldsymbol{x})}\left(\frac{\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x})}-1\right)(1+o_{\mathbb{P}}(1))$$

$$-2^{\gamma(\boldsymbol{x})}\sqrt{k_n h_n^p}\frac{e(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x})-e(1-k_n/n|\boldsymbol{x})}\left(\frac{\widehat{e}_n(1-k_n/(2n)|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x})}-1\right)(1+o_{\mathbb{P}}(1))$$

$$+2^{\gamma(\boldsymbol{x})}\sqrt{k_n h_n^p}\frac{e(1-k_n/n|\boldsymbol{x})}{e(1-k_n/(2n)|\boldsymbol{x})-e(1-k_n/n|\boldsymbol{x})}\left(\frac{\widehat{e}_n(1-k_n/n|\boldsymbol{x})}{e(1-k_n/n|\boldsymbol{x})}-1\right)(1+o_{\mathbb{P}}(1)).$$

By noticing that $e\left(1-j\lambda k_n/n|\boldsymbol{x}\right)/\left(e\left(1-\lambda k_n/n|\boldsymbol{x}\right)-e\left(1-2\lambda k_n/n|\boldsymbol{x}\right)\right)\to j^{-\gamma(\boldsymbol{x})}(1-2^{-\gamma(\boldsymbol{x})})^{-1}$ for any $\lambda>0$, $j=1,2$, we deduce that

$$A_n=\sqrt{k_n h_n^p}\left(\frac{2^{\gamma(\boldsymbol{x})}}{1-2^{-\gamma(\boldsymbol{x})}},-\frac{1+2^{\gamma(\boldsymbol{x})}}{1-2^{-\gamma(\boldsymbol{x})}},\frac{1}{1-2^{-\gamma(\boldsymbol{x})}}\right)\begin{pmatrix}\dfrac{\widehat{e}_n\left(1-k_n/(4n)|\boldsymbol{x}\right)}{e\left(1-k_n/(4n)|\boldsymbol{x}\right)}-1\\[2mm]\dfrac{\widehat{e}_n\left(1-k_n/(2n)|\boldsymbol{x}\right)}{e\left(1-k_n/(2n)|\boldsymbol{x}\right)}-1\\[2mm]\dfrac{\widehat{e}_n\left(1-k_n/n|\boldsymbol{x}\right)}{e\left(1-k_n/n|\boldsymbol{x}\right)}-1\end{pmatrix}+o_{\mathbb{P}}(1).$$

The conclusion now follows from a straightforward application of Theorem 1 and of the delta-method.

Asymptotic distribution of $\widehat{\gamma}_{k_n}^{(2)}(\boldsymbol{x})$. The proof is entirely similar to the proof for $\widehat{\gamma}_{k_n}^{(1)}(\boldsymbol{x})$ and is therefore omitted.

Asymptotic distribution of $\widehat{\gamma}_{k_n}^{(3)}(\boldsymbol{x})$. Let us denote for convenience $\alpha_n=1-k_n/n$, $\sigma_n=1/\sqrt{k_n h_n^p}$, $\theta=\gamma(\boldsymbol{x})^{-1}-1$, and focus on the probability

$$\Phi_n(z)=\mathbb{P}\left(\sigma_n^{-1}\left(\frac{\widehat{\overline{F}}_n\left(\widehat{e}_n(\alpha_n|\boldsymbol{x})|\boldsymbol{x}\right)}{1-\alpha_n}-\theta\right)\leq z\right)=\mathbb{P}\left(\widehat{\overline{F}}_n\left(\widehat{e}_n(\alpha_n|\boldsymbol{x})|\boldsymbol{x}\right)\leq(1-\alpha_n)\left(\theta+z\sigma_n\right)\right).$$

Equivalently $\Phi_n(z)=\mathbb{P}\left(\widehat{e}_n(\alpha_n|\boldsymbol{x})\geq\widehat{q}_n\left(\beta_n|\boldsymbol{x}\right)\right)$, where $\beta_n=1-\left(1-\alpha_n\right)\left(\theta+z\sigma_n\right)$, and so

$$\Phi_n(z)=\mathbb{P}\left(\sigma_n^{-1}\left(\frac{\widehat{e}_n(\alpha_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}-1\right)\geq\sigma_n^{-1}\left(\frac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\right)\frac{q(\beta_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}+\sigma_n^{-1}\left(\frac{q(\beta_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}-1\right)\right).$$

According to Theorem 2.3.9 in de Haan and Ferreira (2006), Equation (3.6) provides:

$$\frac{q(\beta_n|\boldsymbol{x})}{q(\alpha_n|\boldsymbol{x})}=(\theta+z\sigma_n)^{-\gamma(\boldsymbol{x})}\left(1+O(|A((1-\alpha_n)^{-1}|\boldsymbol{x})|)\right).$$

Apply then Proposition 1 in Daouia et al. (2019b) to the conditional distribution of Y given $\boldsymbol{X}=\boldsymbol{x}$ to get

$$\frac{q(\beta_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}=\left(1-\gamma(\boldsymbol{x})\frac{z\sigma_n}{\theta}(1+o(1))\right)\left(1-\frac{\gamma(\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}(m^{(1)}(\boldsymbol{x})+o(1))+O(|A((1-\alpha_n)^{-1}|\boldsymbol{x})|)\right).$$

Recall that $\sigma_n=1/\sqrt{k_n h_n^p}$ and $\theta=\gamma(\boldsymbol{x})^{-1}-1$ to find

$$\sigma_n^{-1}\left(\frac{q(\beta_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}-1\right)\to-\frac{\gamma^2(\boldsymbol{x})}{1-\gamma(\boldsymbol{x})}z-\lambda\gamma(\boldsymbol{x})m^{(1)}(\boldsymbol{x})\ \text{ as }n\to\infty.$$

We can therefore rewrite $\Phi_n(z)$ as the probability

$$\mathbb{P}\left(\frac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}\left[-\sigma_n^{-1}\left(\frac{\widehat{e}_n(\alpha_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}-1\right)+\sigma_n^{-1}\left(\frac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\right)\right](1+o(1))-\lambda\frac{1-\gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}m^{(1)}(\boldsymbol{x})\leq z\right).$$

In other words, the asymptotic distribution of $\sigma_n^{-1}\left(\dfrac{\widehat{\overline{F}}_n\left(\widehat{e}_n(\alpha_n|\boldsymbol{x})|\boldsymbol{x}\right)}{1-\alpha_n}-\theta,\dfrac{\widehat{e}_n(\alpha_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}-1\right)$ is that of

$$\begin{pmatrix}-\dfrac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}&\dfrac{1-\gamma(\boldsymbol{x})}{\gamma^2(\boldsymbol{x})}\\[2mm]1&0\end{pmatrix}\sigma_n^{-1}\begin{pmatrix}\dfrac{\widehat{e}_n(\alpha_n|\boldsymbol{x})}{e(\alpha_n|\boldsymbol{x})}-1\\[2mm]\dfrac{\widehat{q}_n(\beta_n|\boldsymbol{x})}{q(\beta_n|\boldsymbol{x})}-1\end{pmatrix}-\begin{pmatrix}\lambda\dfrac{1-\gamma(\boldsymbol{x})}{\gamma(\boldsymbol{x})}m^{(1)}(\boldsymbol{x})\\[2mm]0\end{pmatrix}.$$

The conclusion follows from an application of Theorem 1 with $J=1$, $\tau_1=1$, $\tau=\gamma(\boldsymbol{x})^{-1}-1$ and a use of the delta-method. $\square$

References

Abramowitz, M. and Stegun, I. A. (1966). *Handbook of Mathematical Functions*. U.S. National Bureau of Standards.

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley.

Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506.

Bellini, F., Klar, B., Muller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41–48.

Billingsley, P. (1995). *Probability and Measure (Third edition)*. John Wiley & Sons.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*. Cambridge University Press.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4):761–772.

Cai, J. and Weng, C. (2016). Optimal reinsurance with expectile. *Scandinavian Actuarial Journal*, 2016(7):624–645.

Cai, J.-J., de Haan, L., and Zhou, C. (2013). Bias correction in extreme value statistics with index around zero. *Extremes*, 16(2):173–201.

Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, 83(3):735–776.

Chen, Z. (1996). Conditional L_p-quantiles and their application to the testing of symmetry in nonparametric regression. *Statistics & Probability Letters*, 29(2):107–115.

Daouia, A., Gardes, L., and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19(5B):2557–2589.

Daouia, A., Gardes, L., Girard, S., and Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST*, 20(2):311–333.

Daouia, A., Girard, S., and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 80(2):263–292.

Daouia, A., Girard, S., and Stupfler, G. (2019a). Extreme M-quantiles as risk measures: from L^1 to L^p optimization. *Bernoulli*, 25(1):264–309.

Daouia, A., Girard, S., and Stupfler, G. (2019b). Tail expectile process and risk assessment, preprint. `https://hal.archives-ouvertes.fr/hal-01744505`.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer-Verlag New York.

Durrieu, G., Grama, I., Pham, Q.-K., and Tricot, J.-M. (2015). Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles. *Extremes*, 18(3):437–478.

El Methni, J., Gardes, L., and Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41(4):988–1012.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer.

Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes*, 17(1):45–75.

Gardes, L. and Stupfler, G. (2019). An integrated functional Weissman estimator for conditional extreme quantiles. *REVSTAT: Statistical Journal*, 17(1):109–144.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Goegebeur, Y., Guillou, A., and Osmann, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates. *Communications in Statistics – Theory and Methods*, 46(1):319–343.

Guo, M. and Härdle, W. K. (2012). Simultaneous confidence bands for expectile functions. *AStA Advances in Statistical Analysis*, 96(4):517–541.

Holzmann, H. and Klar, B. (2016). Expectile asymptotics. *Electronic of Journal of Statistics*, 10:2355–2371.

Jones, M. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters*, 20:149–153.

Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics*, 44(2):425–454.

Krzyzak, A. (1986). The rates of convergence of kernel regression estimates and classification rules. *IEEE Transactions on Information Theory*, 32(5):668–679.

Kuan, C.-M., Yeh, J.-H., and Hsu, Y.-C. (2009). Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, 150(2):261–270.

Linsmeier, T. J. and Pearson, N. D. (2000). Value at risk. *Financial Analysts Journal*, 56(2):47–67.

Mao, T., Ng, K., and Hu, T. (2015). Asymptotic expansions of generalized quantiles and expectiles for extreme risks. *Probability in the Engineering and Informational Sciences*, 29(3):309–327.

Mao, T. and Yang, F. (2015). Risk concentration based on expectiles for extreme risks under FGM copula. *Insurance: Mathematics and Economics*, 64:429–439.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.

Ohlsson, E. and Johansson, B. (2010). *Non-Life Insurance Pricing with Generalized Linear Models.* Springer.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer.

Rohrbeck, C., Eastoe, E., Frigessi, A., and Tawn, J. (2018). Extreme value modelling of water-related insurance claims. *Annals of Applied Statistics*, 12(1):246–282.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

Stupfler, G. (2013). A moment estimator for the conditional extreme-value index. *Electronic Journal of Statistics*, 7:2298–2343.

Stupfler, G. (2016). Estimating the conditional extreme-value index under random right-censoring. *Journal of Multivariate Analysis*, 144:1–24.

Taylor, J. (2008). Estimating Value at Risk and Expected Shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252.

Usseglio-Carleve, A. (2018). Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electronic Journal of Statistics*, 12(2):4057–4093.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815.

Ziegel, J. (2016). Coherence and elicitability. *Mathematical Finance*, 26(4):901–918.