



**HAL**  
open science

# Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models

Stéphane Girard, Gilles Stupfler, Antoine Usseglio-Carleve

► **To cite this version:**

Stéphane Girard, Gilles Stupfler, Antoine Usseglio-Carleve. Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. CMStatistics 2021 - 14th International Conference of the ERCIM WG on Computational and Methodological Statistics, Dec 2021, London, United Kingdom. 10.1214/21-AOS2087 . hal-03501801

**HAL Id: hal-03501801**

**<https://hal.science/hal-03501801v1>**

Submitted on 12 Jul 2024

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

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

# EXTREME CONDITIONAL EXPECTILE ESTIMATION IN HEAVY-TAILED HETEROSCEDASTIC REGRESSION MODELS

BY STÉPHANE GIRARD<sup>1</sup>, GILLES STUPFLER<sup>2,\*</sup> AND ANTOINE USSEGLIO-CARLEVE<sup>1,2,3,†</sup>

<sup>1</sup>Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France, [Stephane.Girard@inria.fr](mailto:Stephane.Girard@inria.fr)

<sup>2</sup>Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France, \*[gilles.stupfler@ensai.fr](mailto:gilles.stupfler@ensai.fr);  
†[antoine.usseglio-carleve@ensai.fr](mailto:antoine.usseglio-carleve@ensai.fr)

<sup>3</sup>Toulouse School of Economics, University of Toulouse Capitole, France

Expectiles define a least squares analogue of quantiles. They have been the focus of a substantial quantity of research in the context of actuarial and financial risk assessment over the last decade. The behaviour and estimation of unconditional extreme expectiles using independent and identically distributed heavy-tailed observations has been investigated in a recent series of papers. We build here a general theory for the estimation of extreme conditional expectiles in heteroscedastic regression models with heavy-tailed noise; our approach is supported by general results of independent interest on residual-based extreme value estimators in heavy-tailed regression models, and is intended to cope with covariates having a large but fixed dimension. We demonstrate how our results can be applied to a wide class of important examples, among which linear models, single-index models as well as ARMA and GARCH time series models. Our estimators are showcased on a numerical simulation study and on real sets of actuarial and financial data.

## 1. Introduction.

1.1. *Motivation.* A traditional way of considering extreme events is to estimate extreme quantiles of a random variable  $Y \in \mathbb{R}$ , such as the negative daily log-return of a stock market index in finance, so that large values of  $Y$  correspond to extreme losses on the market, or the magnitude of a claim in insurance. A better understanding of the extremes of  $Y$  can often be achieved by inferring the conditional extremes of  $Y$  given a covariate  $\mathbf{X}$ . Recent examples include the analysis of high healthcare costs in [49] and large insurance claims in [41]. We focus on the case when  $Y$  given  $\mathbf{X}$  is heavy-tailed (*i.e.* Paretian-tailed); this assumption underpins the aforementioned papers and is generally appropriate to the modelling of actuarial and financial data. Under no further assumptions on the structure of  $(\mathbf{X}, Y)$ , non-parametric smoothing methods such as those of [7, 18] can be used. Those techniques suffer from the curse of dimensionality, compounded in conditional extreme value statistics by the necessity to select only the few high observations relevant to the analysis. Early attempts at tackling the low-dimensional restriction, such as [12], were built on parametric models. Later attempts have mostly used quantile regression: a seminal paper is [6], developed further by [23, 49, 50]. An approach based on Tail Dimension Reduction was adopted by [17].

These techniques, and more generally the current state of art in conditional extreme value analysis, rely on quantiles, which only use the information on the frequency of tail events and not on their actual magnitudes. This is an issue in risk assessment, where knowing the

---

*MSC 2010 subject classifications:* Primary 62G32; secondary 62G08, 62G20, 62G30

*Keywords and phrases:* Expectiles, extreme value analysis, heavy-tailed distribution, heteroscedasticity, regression models, residual-based estimators, single-index model, tail empirical process of residuals

magnitude of typical extreme losses is important. One way of tackling this problem is to work with expectiles, introduced in [38]. The  $\tau$ th regression expectile of  $Y$  given  $\mathbf{X}$  is obtained from the  $\tau$ th regression quantile by replacing absolute deviations by squared deviations:

$$\xi_\tau(Y|\mathbf{x}) = \arg \min_{\theta \in \mathbb{R}} \mathbb{E}([\eta_\tau(Y - \theta) - \eta_\tau(Y)] | \mathbf{X} = \mathbf{x}),$$

where  $\eta_\tau(y) = |\tau - \mathbb{1}\{y \leq 0\}|y^2$  is the expectile check function and  $\mathbb{1}\{\cdot\}$  is the indicator function. Expectiles are well-defined and unique when the underlying distribution has a finite first moment (see [1] and Theorem 1 in [38]). Unlike quantiles, expectiles depend on both the probability of tail values and their realisations (see [31]). In addition, expectiles induce the only coherent, law-invariant and elicitable risk measure (see [55]) and therefore benefit from the existence of a natural backtesting methodology. Expectiles are thus a sensible risk management tool to use, as a complement or an alternative to quantiles.

The literature has essentially focused on estimating expectiles with a fixed level  $\tau$  (see e.g. [27, 30]). The estimation of extreme expectiles, where  $\tau = \tau_n \rightarrow 1$  as the sample size  $n$  tends to infinity, remains largely unexplored; it was initiated by [9, 11] in the unconditional heavy-tailed case. Our focus is to provide and discuss the theory of estimators of extreme conditional expectiles, in models that may cope with a large but fixed dimension of the covariate  $\mathbf{X}$ . In doing so we shall develop a novel theory of independent interest for the asymptotic analysis of residual-based extreme value estimators in heavy-tailed regression models.

*1.2. Expectiles and regression models.* We outline our general idea in the location-scale shift linear regression model. Let  $(\mathbf{X}_i, Y_i)$ ,  $1 \leq i \leq n$  be a sample from a random pair  $(\mathbf{X}, Y)$  such that  $Y = \alpha + \boldsymbol{\beta}^\top \mathbf{X} + (1 + \boldsymbol{\theta}^\top \mathbf{X})\varepsilon$ . The parameters  $\alpha \in \mathbb{R}$ ,  $\boldsymbol{\beta} \in \mathbb{R}^d$  and  $\boldsymbol{\theta} \in \mathbb{R}^d$  are unknown, and so are the distributions of the covariate  $\mathbf{X} \in \mathbb{R}^d$  and the unobserved noise variable  $\varepsilon \in \mathbb{R}$ . We also suppose that  $\mathbf{X}$  is independent of  $\varepsilon$ , and has a support  $K$  such that  $1 + \boldsymbol{\theta}^\top \mathbf{x} > 0$  for all  $\mathbf{x} \in K$ . In this model, by location equivariance and positive homogeneity of expectiles (Theorem 1(iii) in [38]), we may write  $\xi_\tau(Y|\mathbf{x}) = \alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_\tau(\varepsilon)$ . A natural idea to estimate the extreme conditional expectile  $\xi_{\tau_n}(Y|\mathbf{x})$ , where  $\tau = \tau_n \rightarrow 1$  as  $n \rightarrow \infty$ , is to first construct estimators  $\hat{\alpha}$ ,  $\hat{\boldsymbol{\beta}}$  and  $\hat{\boldsymbol{\theta}}$  of the model parameters using a weighted least squares method and then construct residuals which can be used, instead of the unobservable errors, to estimate extreme expectiles of  $\varepsilon$ . This expectile estimator can be adapted from, for instance, an empirical asymmetric least squares method (see [9, 11]). If  $\varepsilon$  has a finite second moment, the weighted least squares approach produces  $\sqrt{n}$ -consistent estimators, and it is reasonable to expect that the asymptotic normality properties of the estimators of [9, 11] carry over to their residual-based versions. An estimator of the extreme conditional expectile  $\xi_{\tau_n}(Y|\mathbf{x})$  is then readily obtained as  $\hat{\xi}_{\tau_n}(Y|\mathbf{x}) = \hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \hat{\boldsymbol{\theta}}^\top \mathbf{x})\hat{\xi}_{\tau_n}(\varepsilon)$ .

Our main objective in this paper is to generalise this construction in heteroscedastic regression models of the form  $Y = g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$  where  $g$  and  $\sigma > 0$  are two measurable functions of  $\mathbf{X}$ , so that  $\xi_{\tau_n}(Y|\mathbf{x}) = g(\mathbf{x}) + \sigma(\mathbf{x})\xi_{\tau_n}(\varepsilon)$ . If  $\varepsilon$  is centred and has unit variance, this model can essentially be viewed as  $\mathbb{E}(Y|\mathbf{X}) = g(\mathbf{X})$  and  $\text{Var}(Y|\mathbf{X}) = \sigma^2(\mathbf{X})$ , and is called *location-dispersion regression model* in [47]. Even though our theory will be valid for arbitrary regression models of this form, one should keep in mind models adapted to the consideration of a large dimension  $d$ , where the estimation of  $g$  and  $\sigma$  will not suffer from the curse of dimensionality and thus reasonable rates of convergence can be achieved. The viewpoint we deliberately adopt is that the estimation of  $g$  and  $\sigma$  is the “easy” part of the estimation of  $\xi_{\tau_n}(Y|\mathbf{x})$  because, depending on the model, it can be tackled by known parametric or semiparametric techniques that are easy to implement and converge faster than the extreme value procedure for the estimation of  $\xi_{\tau_n}(\varepsilon)$ . This converts the problem of conditional extreme value estimation into the question of being able to carry out extreme value

inference based on residuals rather than the unobserved noise variables, which is nonetheless a difficult question because residuals are neither independent nor identically distributed.

In Section 2, given that residuals of the model are available, we provide high-level, fairly easy to check and reasonable sufficient conditions under which the asymptotics of residual-based estimators of  $\xi_{\tau_n}(\varepsilon)$  are those of their unfeasible, unobserved error-based counterparts. Several of our results are of independent interest: in particular, we prove in Section 2.2 a non-trivial result on Gaussian approximations of the tail empirical process of the residuals, which is an important step in proving asymptotic theory for extreme value estimators in general regression models. The idea of carrying out conditional extreme value estimation using residuals of location-scale regression models is not new: it has been used since at least [37] and more recently in [2, 26, 35] in the context of the estimation of extreme conditional Value-at-Risk and Expected Shortfall. A novel contribution of this paper is to provide a very general theoretical framework for tackling such questions. In Section 3, we shall then consider five fully worked-out examples. We start with the location-scale shift linear regression model in Section 3.1, a heteroscedastic single-index model in Section 3.2, and a heteroscedastic, Tobit-type left-censored model in Section 3.3. The latter example allows us to show how our method adapts to a situation where the model  $Y = g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$  is valid in the right tail rather than globally. Aside from these three examples, we study the two general ARMA and GARCH time series models in Section 3.4 as a way to illustrate how our results may be used to tackle the problem of dynamic extreme conditional expectile estimation. Section 4 examines the behaviour of our estimators on simulated and real data, and Section 5 discusses our findings and research perspectives. All the necessary mathematical proofs, as well as further details and results related to our finite-sample studies, are deferred to the Supplementary Material [19].

**2. General theoretical toolbox for extreme expectile estimation in heavy-tailed regression models.** Our general framework is the following. Let  $(\mathbf{X}_i, Y_i)$ ,  $1 \leq i \leq n$  be part of a (strictly) stationary sequence of copies of the random pair  $(\mathbf{X}, Y)$ , with  $Y \in \mathbb{R}$ , such that

$$(1) \quad Y = g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$$

where  $g$  and  $\sigma > 0$  are two measurable functions of  $\mathbf{X}$ . The unobserved noise variable  $\varepsilon \in \mathbb{R}$  is centred and independent of  $\mathbf{X}$ ; in other words, for each  $i$ ,  $\mathbf{X}_i$  is independent of  $\varepsilon_i$ , although we do not assume independence between the pairs  $(\mathbf{X}_i, \varepsilon_i)$ . In addition, we suppose throughout that the  $\varepsilon_i = (Y_i - g(\mathbf{X}_i))/\sigma(\mathbf{X}_i)$  are independent.

It follows from this assumption that a conditional expectile  $\xi_{\tau_n}(Y|\mathbf{x})$  can be written as  $\xi_{\tau_n}(Y|\mathbf{X} = \mathbf{x}) = g(\mathbf{x}) + \sigma(\mathbf{x})\xi_{\tau_n}(\varepsilon|\mathbf{X} = \mathbf{x}) = g(\mathbf{x}) + \sigma(\mathbf{x})\xi_{\tau_n}(\varepsilon)$ , where we used the location equivariance and positive homogeneity to obtain the first identity, and the independence between  $\mathbf{X}$  and  $\varepsilon$  to get the second identity. We assume throughout Section 2 that  $g$  and  $\sigma$  have been estimated, and we concentrate on estimating the extreme expectile  $\xi_{\tau_n}(\varepsilon)$ , with the objective of ultimately constructing an estimator of  $\xi_{\tau_n}(Y|\mathbf{x})$ . Denoting by  $\tau \mapsto q_\tau(\varepsilon)$  the quantile function of  $\varepsilon$ , we work under the following first-order Pareto-type condition:

$\mathcal{C}_1(\gamma)$  The tail quantile function of  $\varepsilon$ , defined by  $U(t) = q_{1-t^{-1}}(\varepsilon)$  for  $t > 1$ , is regularly varying with index  $\gamma > 0$ :  $U(tx)/U(t) \rightarrow x^\gamma$  as  $t \rightarrow \infty$  for any  $x > 0$ .

Condition  $\mathcal{C}_1(\gamma)$  is equivalent to assuming that the survival function of  $\varepsilon$ , denoted hereafter by  $\bar{F} : x \mapsto \mathbb{P}(\varepsilon > x)$ , is regularly varying with index  $-1/\gamma < 0$  (see [13], Proposition B.1.9). Together with condition  $\mathbb{E}|\varepsilon_-| < \infty$ , where  $\varepsilon_- = \min(\varepsilon, 0)$ , the assumption  $\gamma < 1$  ensures that the first moment of  $\varepsilon$  exists, which entails that expectiles of  $\varepsilon$  of any order are well-defined. Both of these conditions shall be part of our minimal assumptions throughout.

The essential difficulty to overcome in our setup is that the  $\varepsilon_i$  are unobserved. However, because  $g$  and  $\sigma$  have been estimated, by  $\bar{g}$  and  $\bar{\sigma}$  say, we have access to residuals  $\hat{\varepsilon}_i^{(n)} = (Y_i - \bar{g}(\mathbf{X}_i))/\bar{\sigma}(\mathbf{X}_i)$  constructed from the regression model (1). Our idea in this section will be to construct estimators of extreme expectiles based on the observable  $\hat{\varepsilon}_i^{(n)}$ , and study their theoretical properties when they are in some sense “close” to the true, unobserved  $\varepsilon_i$ .

We start by the case of an *intermediate level*  $\tau_n$ , meaning that  $\tau_n \rightarrow 1$  and  $n(1 - \tau_n) \rightarrow \infty$ . Section 2.1 below focuses on a residual-based Least Asymmetrically Weighted Squares (LAWS) estimator. Section 2.2 then introduces a competitor based on the connection between (theoretical) extreme expectiles and quantiles and new general results on tail empirical processes of residuals in heavy-tailed models. Section 2.3 extrapolates these estimators to properly extreme levels  $\tau'_n$  using a Weissman-type construction warranted by the heavy-tailed assumption (see [51]), and combines these extrapolated devices with the estimators of  $g$  and  $\sigma$  to finally obtain an estimator of the extreme conditional expectile  $\xi_{\tau'_n}(Y|\mathbf{x})$ .

*2.1. Intermediate step, direct construction: residual-based LAWS.* Assume that  $\tau_n$  is an intermediate sequence, i.e.  $\tau_n \rightarrow 1$  and  $n(1 - \tau_n) \rightarrow \infty$ . If the errors  $\varepsilon_i$  were available, we could estimate  $\xi_{\tau_n}(\varepsilon)$  by  $\check{\xi}_{\tau_n}(\varepsilon)$  minimising  $\sum_{i=1}^n \eta_{\tau_n}(\varepsilon_i - u)$  with respect to  $u$ . We replace the unobserved  $\varepsilon_i$  by the observed residuals  $\hat{\varepsilon}_i^{(n)}$ , resulting in the LAWS estimator

$$\hat{\xi}_{\tau_n}(\varepsilon) = \arg \min_{u \in \mathbb{R}} \sum_{i=1}^n \eta_{\tau_n}(\hat{\varepsilon}_i^{(n)} - u).$$

Our first main theorem is a flexible result stating that  $\hat{\xi}_{\tau_n}(\varepsilon)$  is a  $\sqrt{n(1 - \tau_n)}$ -relatively asymptotically normal estimator of the high, intermediate expectile  $\xi_{\tau_n}(\varepsilon)$  provided the gap between residuals and unobservable errors is not too large. For technical extensions to the case of a random sample size or independent arrays, see Lemmas C.5 and C.8 of [19].

**THEOREM 2.1.** *Assume that there is  $\delta > 0$  such that  $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$ , that  $\varepsilon$  satisfies condition  $\mathcal{C}_1(\gamma)$  with  $0 < \gamma < 1/2$  and  $\tau_n \uparrow 1$  is such that  $n(1 - \tau_n) \rightarrow \infty$ . Suppose moreover that the array of random variables  $\hat{\varepsilon}_i^{(n)}$ ,  $1 \leq i \leq n$ , satisfies*

$$(2) \quad \sqrt{n(1 - \tau_n)} \max_{1 \leq i \leq n} \frac{|\hat{\varepsilon}_i^{(n)} - \varepsilon_i|}{1 + |\varepsilon_i|} \xrightarrow{\mathbb{P}} 0.$$

*Then we have  $\sqrt{n(1 - \tau_n)} \begin{pmatrix} \hat{\xi}_{\tau_n}(\varepsilon) \\ \xi_{\tau_n}(\varepsilon) \end{pmatrix} - 1 \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1 - 2\gamma}\right)$ .*

**REMARK 1.** Theorem 2.1 is a non-trivial extension of Theorem 2 in [9] to the case when the  $\varepsilon_i$  are unobserved. The difference lies in the fact that the estimator  $\hat{\xi}_{\tau_n}(\varepsilon)$  is much more difficult to handle directly; Condition (2), on the weighted distance between the  $\varepsilon_i$  and the  $\hat{\varepsilon}_i^{(n)}$ , allows for a control of the gap between  $\hat{\xi}_{\tau_n}(\varepsilon)$  and the unfeasible  $\check{\xi}_{\tau_n}(\varepsilon)$ , with the presence of the denominator  $1 + |\varepsilon_i|$  making it possible to deal with heteroscedasticity in practice. We shall use this key condition again in our results in Section 2.2. It will be satisfied when the structure of the model  $Y = g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$  is estimated at a faster rate than the  $\sqrt{n(1 - \tau_n)}$ -rate of convergence of intermediate expectile estimators. The proof is based on rigorously establishing that  $\hat{\xi}_{\tau_n}(\varepsilon)$  and  $\check{\xi}_{\tau_n}(\varepsilon)$  have the same asymptotic distribution; the striking fact is that the theoretical arguments fundamentally only require stationarity of the  $\varepsilon_i$ , with independence only being crucial for concluding that  $\check{\xi}_{\tau_n}(\varepsilon)$  is asymptotically Gaussian by Theorem 2 in [9] and therefore  $\hat{\xi}_{\tau_n}(\varepsilon)$  must be so. Theorem 2.1 can then be

expected to have analogues when the  $\varepsilon_i$  are stationary but weakly dependent, thus covering (for example) regression models with time series errors as in [46], as long as one can prove the  $\sqrt{n(1-\tau_n)}$ -asymptotic normality of  $\check{\xi}_{\tau_n}(\varepsilon)$ . An example of such a result for stationary and mixing  $\varepsilon_i$  has been investigated in [10].

*2.2. Intermediate step, indirect construction.* We start by recalling, as shown in Proposition 2.3 of [3], that the heavy-tailed condition on  $t \mapsto U(t) = q_{1-t^{-1}}(\varepsilon)$  entails

$$\lim_{\tau \uparrow 1} \frac{\xi_\tau(\varepsilon)}{q_\tau(\varepsilon)} = (\gamma^{-1} - 1)^{-\gamma}.$$

Therefore, if  $\bar{\gamma}$  is a consistent estimator of  $\gamma$ , and  $\bar{q}_{\tau_n}(\varepsilon)$  is a consistent estimator of  $q_{\tau_n}(\varepsilon)$ , we can estimate the intermediate expectile  $\xi_{\tau_n}(\varepsilon)$  by the so-called indirect estimator

$$\tilde{\xi}_{\tau_n}(\varepsilon) = (\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \bar{q}_{\tau_n}(\varepsilon).$$

An extension of Theorem 1 in [9] (see Proposition A.1 in [19]) shows that under the following classical second-order refinement of condition  $\mathcal{C}_1(\gamma)$ , the asymptotic distribution of the estimator  $\tilde{\xi}_{\tau_n}(\varepsilon)$  is determined under high-level conditions on  $(\bar{\gamma}, \bar{q}_{\tau_n}(\varepsilon))$ .

$\mathcal{C}_2(\gamma, \rho, A)$  For all  $x > 0$ ,

$$\lim_{t \rightarrow \infty} \frac{1}{A(t)} \left[ \frac{U(tx)}{U(t)} - x^\gamma \right] = x^\gamma \frac{x^\rho - 1}{\rho},$$

where  $A$  is a function converging to 0 at infinity and having constant sign, and  $\rho \leq 0$ . Here and in what follows,  $(x^\rho - 1)/\rho$  is to be read as  $\log x$  when  $\rho = 0$ .

We now explain how one may construct and study residual-based estimators  $\bar{\gamma}$  and  $\bar{q}_{\tau_n}(\varepsilon)$ . Let  $z_{1,n} \leq z_{2,n} \leq \dots \leq z_{n,n}$  be the ordered  $n$ -tuple associated with an  $n$ -tuple  $(z_1, z_2, \dots, z_n)$ . A number of estimators of  $\gamma$  can be adapted to our case and written as a functional of the tail empirical quantile process of the residuals, among which the popular Hill estimator ([24]):

$$\hat{\gamma}_k = \frac{1}{k} \sum_{i=1}^k \log \frac{\hat{\varepsilon}_{n-i+1,n}^{(n)}}{\hat{\varepsilon}_{n-k,n}^{(n)}} = \int_0^1 \log \left( \frac{\hat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\hat{\varepsilon}_{n-k,n}^{(n)}} \right) ds.$$

Here  $\lfloor \cdot \rfloor$  denotes the floor function. We may also adapt in the same way the moment-type statistics which intervene in the construction of the moment estimator of [14], and the general class of estimators studied by [42]. These estimators depend on the choice of an effective sample size  $k = k(n) \rightarrow \infty$  and  $k/n \rightarrow 0$ ; it is useful to think of  $k$  as being  $k = \lfloor n(1 - \tau_n) \rfloor$ .

It is therefore worthwhile to study the asymptotic behaviour of the tail empirical quantile process  $s \mapsto \hat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}$  of residuals, and of its log-counterpart. This is of course a difficult task, because the array of residuals is not made of independent random variables. To tackle this problem, we first recall that under condition  $\mathcal{C}_2(\gamma, \rho, A)$ , one can write a weighted uniform Gaussian approximation of the tail empirical quantile process of the (unobserved)  $\varepsilon_i$ :

$$\frac{\varepsilon_{n-\lfloor ks \rfloor, n}}{q_{1-k/n}(\varepsilon)} = s^{-\gamma} + \frac{1}{\sqrt{k}} \left( \gamma s^{-\gamma-1} W_n(s) + \sqrt{k} A(n/k) s^{-\gamma} \frac{s^{-\rho} - 1}{\rho} + s^{-\gamma-1/2-\delta} o_{\mathbb{P}}(1) \right)$$

uniformly in  $s \in (0, 1]$ , where  $W_n$  is a sequence of standard Brownian motions and  $\delta > 0$  is arbitrarily small (see Theorem 2.4.8 in [13]), provided  $k = k(n) \rightarrow \infty$ ,  $k/n \rightarrow 0$ , and  $\sqrt{k} A(n/k) = O(1)$ . Strictly speaking, this approximation is only valid for appropriate versions of the tail empirical quantile process on an appropriate probability space; since taking such versions has no consequences on weak convergence results, we do not emphasise this in the sequel. For certain results which require the study of the log-spacings

$\log \varepsilon_{n-\lfloor ks \rfloor, n} - \log \varepsilon_{n-k, n}$ , such as the convergence of the Hill estimator, an approximation of the log-tail empirical quantile process is sometimes preferred: uniformly in  $s \in (0, 1]$ ,

$$\frac{1}{\gamma} \log \left( \frac{\varepsilon_{n-\lfloor ks \rfloor, n}}{q_{1-k/n}(\varepsilon)} \right) = \log \frac{1}{s} + \frac{1}{\sqrt{k}} \left( s^{-1} W_n(s) + \sqrt{k} A(n/k) \frac{1}{\gamma} \frac{s^{-\rho} - 1}{\rho} + s^{-1/2-\delta} o_{\mathbb{P}}(1) \right).$$

Our next result is that, if the error made in the construction of the residuals is not too large, then these approximations hold for the tail empirical quantile process of residuals as well.

**THEOREM 2.2.** *Assume that condition  $\mathcal{C}_2(\gamma, \rho, A)$  holds. Let  $k = k(n) = \lfloor n(1 - \tau_n) \rfloor$  where  $\tau_n \uparrow 1$ ,  $n(1 - \tau_n) \rightarrow \infty$  and  $\sqrt{n(1 - \tau_n)} A((1 - \tau_n)^{-1}) = O(1)$ . Suppose that the array of random variables  $\widehat{\varepsilon}_i^{(n)}$ ,  $1 \leq i \leq n$ , satisfies (2). Then there exists a sequence  $W_n$  of standard Brownian motions such that, for any  $\delta > 0$  sufficiently small: uniformly in  $s \in (0, 1]$ ,*

$$\frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{q_{1-k/n}(\varepsilon)} = s^{-\gamma} + \frac{1}{\sqrt{k}} \left( \gamma s^{-\gamma-1} W_n(s) + \sqrt{k} A(n/k) s^{-\gamma} \frac{s^{-\rho} - 1}{\rho} + s^{-\gamma-1/2-\delta} o_{\mathbb{P}}(1) \right)$$

and

$$\frac{1}{\gamma} \log \left( \frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{q_{1-k/n}(\varepsilon)} \right) = \log \frac{1}{s} + \frac{1}{\sqrt{k}} \left( s^{-1} W_n(s) + \sqrt{k} A(n/k) \frac{1}{\gamma} \frac{s^{-\rho} - 1}{\rho} + s^{-1/2-\delta} o_{\mathbb{P}}(1) \right).$$

Theorem 2.2 is the second main contribution of this paper. It is a non-trivial asymptotic result, because there is no guarantee that ranks of the original error sequence are preserved in the residual sequence, and it therefore is not obvious at first sight that Condition (2) on the gap between errors and their corresponding residuals is in fact sufficient to ensure that the tail empirical quantile process based on residuals has similar properties to its unobserved errors-based analogue. As an illustration, we work out the asymptotic properties of the residual-based, Hill-type estimator of the extreme value index  $\gamma$  of the errors, as well as the asymptotic behaviour of the related indirect intermediate expectile estimator in Corollary 2.1 below.

**COROLLARY 2.1.** *Assume that condition  $\mathcal{C}_2(\gamma, \rho, A)$  holds. Let  $\tau_n \uparrow 1$  satisfy  $n(1 - \tau_n) \rightarrow \infty$  and  $\sqrt{n(1 - \tau_n)} A((1 - \tau_n)^{-1}) \rightarrow \lambda \in \mathbb{R}$ . Suppose that the array of random variables  $\widehat{\varepsilon}_i^{(n)}$ ,  $1 \leq i \leq n$ , satisfies (2). If  $\bar{\gamma} = \widehat{\gamma}_{\lfloor n(1 - \tau_n) \rfloor}$  and  $\bar{q}_{\tau_n}(\varepsilon) = \widehat{\varepsilon}_{n-\lfloor n(1 - \tau_n) \rfloor, n}^{(n)}$ , then*

$$\sqrt{n(1 - \tau_n)} \left( \bar{\gamma} - \gamma, \frac{\bar{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} \left( \frac{\lambda}{1 - \rho} + \gamma \int_0^1 \left[ \frac{W(s)}{s} - W(1) \right] ds, \gamma W(1) \right)$$

where  $W$  is a standard Brownian motion. In particular,

$$\sqrt{n(1 - \tau_n)} \left( \bar{\gamma} - \gamma, \frac{\bar{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} (\Gamma, \Theta),$$

where  $\Gamma \sim \mathcal{N}(\lambda/(1 - \rho), \gamma^2)$  and  $\Theta \sim \mathcal{N}(0, \gamma^2)$  are independent. As a consequence, if moreover  $\mathbb{E}|\varepsilon_-| < \infty$ ,  $0 < \gamma < 1$ ,  $\mathbb{E}(\varepsilon) = 0$  and  $\sqrt{n(1 - \tau_n)}/q_{\tau_n}(\varepsilon) = O(1)$ , one has

$$\sqrt{n(1 - \tau_n)} \left( \frac{\widetilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} \mathcal{N} \left( \lambda \left[ \frac{m(\gamma)}{1 - \rho} - b(\gamma, \rho) \right], \gamma^2 [1 + [m(\gamma)]^2] \right),$$

with  $m(\gamma) = (1 - \gamma)^{-1} - \log(\gamma^{-1} - 1)$  and  $b(\gamma, \rho) = \frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho}$ .

This result is our third main contribution. Such results on residual-based extreme value estimators appear to be quite scarce in the literature: see Section 2 in [50] and Section 3 in [49] in linear quantile regression models, Proposition 2 in Appendix A of [26] in ARMA-GARCH models, and Section 3 in [48] in a nonparametric homoscedastic quantile regression model. Our result relaxes these strong modelling assumptions, and provides a reasonable general theoretical framework for the estimation of the extreme value index and intermediate quantile via residuals of a regression model. Similarly to Theorem 2.2, this result is of wider interest in general extreme value regression problems with heavy-tailed random errors.

REMARK 2. When, with probability 1,  $g(\mathbf{X})$  is bounded and  $\sigma(\mathbf{X})$  is positive and bounded (this is the setup of our simulation study for linear and single-index models, see Section 4.1), one could estimate  $\gamma$  using the  $Y_i = g(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\varepsilon_i$  directly, because then the  $Y_i$  all have extreme value index  $\gamma$  (see Lemma A.4 in [19]). A competitor to the estimator  $\hat{\gamma}_k$  is thus  $\check{\gamma}_k = k^{-1} \sum_{i=1}^k \log(Y_{n-i+1,n}/Y_{n-k,n})$ . A numerical comparison of the estimators  $\hat{\gamma}_k$  and  $\check{\gamma}_k$  (which we do not report to save space) shows, however, that the residual-based estimator  $\hat{\gamma}_k$  has by far the best finite-sample performance. The idea is that the presence of the shift  $g(\mathbf{X}_i)$  and scaling  $\sigma(\mathbf{X}_i)$  in the  $Y_i$  introduces a large amount of bias in the estimation of  $\gamma$  by  $\check{\gamma}_k$ ; removing these two components in the calculation of the residuals substantially improves finite-sample results. A related point is made in [13] (p.83).

REMARK 3. The earlier work of [25] provides general tools to obtain the asymptotic normality of the Hill estimator based on a filtered process. The essential difference with our approach is that we put our assumptions directly on the gap between the residuals and the unobserved noise variables; by contrast, the methodology of [25] essentially assumes that the residuals are obtained through a parametric filter, and makes technical assumptions on the regularity of the parametric model and the gap between the estimated parameter and its true value. The latter approach is very powerful when working with time series models, as typical such models (ARMA, GARCH, ARMA-GARCH) have a parametric formulation. By contrast, we avoid the parametric specification and therefore can handle a large class of possibly semiparametric regression models (such as heteroscedastic single-index models, see Section 3.2), while still providing useful results for time series models (see Section 3.4).

The theory in [25] allows for non-independent errors in autoregressive time series, see Section 3.2 therein. This corresponds to when the filter does not correctly describe the underlying structure of the time series, and can be used in misspecified models. Our results use the independence of the errors, but may also be extended to the stationary weakly dependent case: our argument for the proof of Theorem 2.2 (and hence for Corollary 2.1) relies on, first, quantifying the gap between the tail empirical quantile process based on the unobserved errors and its version based on the residuals (see Lemma A.3 of [19]), and then on a Gaussian approximation of the tail empirical quantile process for independent heavy-tailed variables. Inspecting the proofs reveals that both of these steps can in fact be carried out when the  $\varepsilon_i$  are only stationary,  $\beta$ -mixing and satisfy certain anti-clustering conditions, because a Gaussian approximation of the tail empirical quantile process also holds then, see for instance Theorem 2.1 in [15].

2.3. *Extrapolation for extreme conditional expectile estimation.* We finally develop high-level results for the estimation of properly extreme conditional expectiles whose level  $\tau'_n \rightarrow 1$  can converge to 1 at an arbitrarily fast rate. One would typically choose  $\tau'_n = 1 - p_n$  for an exceedance probability  $p_n$  not greater than  $1/n$ , see *e.g.* Chapter 4 of [13] in the context of extreme quantile estimation. Following [51], intermediate quantiles of order  $\tau_n$  can be extrapolated to the extreme level  $\tau'_n$ , using the heavy-tailed assumption. This idea successfully



carries over to expectile estimation because of the asymptotic proportionality relationship  $\xi_\tau(\varepsilon)/q_\tau(\varepsilon) \rightarrow (\gamma^{-1} - 1)^{-\gamma}$  as  $\tau \uparrow 1$ , resulting in the following class of estimators of  $\xi_{\tau'_n}(\varepsilon)$ :

$$\bar{\xi}_{\tau'_n}^*(\varepsilon) = \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}} \bar{\xi}_{\tau_n}(\varepsilon),$$

where  $\bar{\gamma}$  and  $\bar{\xi}_{\tau_n}(\varepsilon)$  are consistent estimators of  $\gamma$  and of the intermediate expectile  $\xi_{\tau_n}(\varepsilon)$ . In the context of a regression model of the form (1), these would be based on the residuals obtained via estimators  $\bar{g}(\mathbf{x})$  and  $\bar{\sigma}(\mathbf{x})$  of  $g(\mathbf{x})$  and  $\sigma(\mathbf{x})$ . One can then estimate  $\xi_{\tau'_n}(Y|\mathbf{x})$  in model (1) by  $\bar{\xi}_{\tau'_n}^*(Y|\mathbf{x}) = \bar{g}(\mathbf{x}) + \bar{\sigma}(\mathbf{x})\bar{\xi}_{\tau'_n}^*(\varepsilon)$ . We examine the convergence of this estimator.

**THEOREM 2.3.** *Assume that  $\mathbb{E}|\varepsilon_-| < \infty$  and condition  $\mathcal{C}_2(\gamma, \rho, A)$  holds with  $0 < \gamma < 1$  and  $\rho < 0$ . Assume further that  $\mathbb{E}(\varepsilon) = 0$  and  $\tau_n, \tau'_n \uparrow 1$  satisfy*

$$(3) \quad n(1 - \tau_n) \rightarrow \infty, \quad \frac{1 - \tau'_n}{1 - \tau_n} \rightarrow 0, \quad \frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \rightarrow \infty,$$

$$(4) \quad \sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \rightarrow \lambda \in \mathbb{R} \quad \text{and} \quad \frac{\sqrt{n(1 - \tau_n)}}{q_{\tau_n}(\varepsilon)} = O(1).$$

Suppose also that  $\sqrt{n(1 - \tau_n)}(\bar{\xi}_{\tau_n}(\varepsilon)/\xi_{\tau_n}(\varepsilon) - 1) = O_{\mathbb{P}}(1)$  and  $\sqrt{n(1 - \tau_n)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$ , where  $\Gamma$  is nondegenerate. Then

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left( \frac{\bar{\xi}_{\tau'_n}^*(\varepsilon)}{\xi_{\tau'_n}(\varepsilon)} - 1 \right) \xrightarrow{d} \Gamma.$$

Finally, if model (1) holds (with  $\mathbf{X}$  independent of  $\varepsilon$ ) and, at a given point  $\mathbf{x}$ , the estimators  $\bar{g}(\mathbf{x})$  and  $\bar{\sigma}(\mathbf{x})$  satisfy  $\bar{g}(\mathbf{x}) - g(\mathbf{x}) = O_{\mathbb{P}}(1)$  and  $\sqrt{n(1 - \tau_n)}(\bar{\sigma}(\mathbf{x}) - \sigma(\mathbf{x})) = O_{\mathbb{P}}(1)$ , then

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left( \frac{\bar{\xi}_{\tau'_n}^*(Y|\mathbf{x})}{\xi_{\tau'_n}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

**REMARK 4.** This result applies to the residual-based direct LAWS estimator and indirect quantile-based estimator under the conditions that ensure their  $\sqrt{n(1 - \tau_n)}$ -consistency. These conditions essentially amount to assuming that the structure of the model is estimated at a rate faster than  $\sqrt{n(1 - \tau_n)}$ , see Theorem 2.1, the related Remark 1, and Corollary 2.1.

### 3. Applications of our theoretical results.

3.1. *Location-scale shift linear regression model.* We concentrate here on applications in the popular example of location-scale shift linear regression model, which we recall below.

**Model ( $M_1$ )** The random pair  $(\mathbf{X}, Y)$  is such that  $Y = \alpha + \boldsymbol{\beta}^\top \mathbf{X} + (1 + \boldsymbol{\theta}^\top \mathbf{X})\varepsilon$ . Here the random covariate  $\mathbf{X}$  is independent of the centred noise variable  $\varepsilon$ , and has a density function  $f_{\mathbf{X}}$  on  $\mathbb{R}^d$  whose support is a compact set  $K$  such that  $1 + \boldsymbol{\theta}^\top \mathbf{x} > 0$  for all  $\mathbf{x} \in K$ .

Model ( $M_1$ ) features heteroscedasticity. It is well-known that in this model, traditional methods such as ordinary least squares are consistent but inefficient. A particular concern in our case is also to find accurate estimators of the heteroscedasticity parameter  $\boldsymbol{\theta}$ ; indeed,  $\xi_{\tau_n}(Y|\mathbf{x}) = \alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon)$  with  $\xi_{\tau_n}(\varepsilon) \rightarrow \infty$  as  $n \rightarrow \infty$ , so that, when  $n$  is large, even a moderately large error in the estimation of  $\boldsymbol{\theta}$  can result in a substantial error in the estimation of the extreme conditional expectile  $\xi_{\tau_n}(Y|\mathbf{x})$ . We suggest a two-stage procedure to estimate  $(\alpha, \boldsymbol{\beta}, \boldsymbol{\theta})$ , based on independent data points  $(\mathbf{X}_i, Y_i)_{1 \leq i \leq n}$ .

- (Preliminary step) Compute the ordinary least squares estimators  $\tilde{\alpha}$  and  $\tilde{\beta}$  of  $\alpha$  and  $\beta$ , and then the ordinary least squares estimator of  $\theta$  based on the absolute residuals  $\tilde{Z}_i = |Y_i - (\tilde{\alpha} + \tilde{\beta}^\top \mathbf{X}_i)|$ , that is,  $\tilde{\theta} = \tilde{\nu}/\tilde{\mu}$  and

$$(\tilde{\alpha}, \tilde{\beta}) = \arg \min_{(a, \mathbf{b})} \sum_{i=1}^n (Y_i - a - \mathbf{b}^\top \mathbf{X}_i)^2, \quad (\tilde{\mu}, \tilde{\nu}) = \arg \min_{(c, \mathbf{d})} \sum_{i=1}^n (\tilde{Z}_i - c - \mathbf{d}^\top \mathbf{X}_i)^2.$$

- (Weighted step) Compute the least squares estimators  $\hat{\alpha}$  and  $\hat{\beta}$  of  $\alpha$  and  $\beta$ , weighted using estimated standard deviations obtained via  $\tilde{\theta}$ , and then the weighted least squares estimator of  $\theta$  based on the absolute residuals  $\hat{Z}_i = |Y_i - (\hat{\alpha} + \hat{\beta}^\top \mathbf{X}_i)|$ , i.e.  $\hat{\theta} = \hat{\nu}/\hat{\mu}$  and

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(a, \mathbf{b})} \sum_{i=1}^n \left( \frac{Y_i - a - \mathbf{b}^\top \mathbf{X}_i}{1 + \tilde{\theta}^\top \mathbf{X}_i} \right)^2, \quad (\hat{\mu}, \hat{\nu}) = \arg \min_{(c, \mathbf{d})} \sum_{i=1}^n \left( \frac{\hat{Z}_i - c - \mathbf{d}^\top \mathbf{X}_i}{1 + \tilde{\theta}^\top \mathbf{X}_i} \right)^2.$$

REMARK 5. This is a one-iteration version of a general weighted least squares procedure where estimates obtained at a given step are fed back into the next iteration to update weights, this procedure being repeated  $n_0$  times. Simulation results seem to indicate that iterating the procedure further does not improve the accuracy of the estimators in practice.

Once these estimates have been obtained, we can construct the sample of (weighted) residuals  $\hat{\varepsilon}_i^{(n)} = (Y_i - (\hat{\alpha} + \hat{\beta}^\top \mathbf{X}_i))/(1 + \hat{\theta}^\top \mathbf{X}_i)$  satisfying Condition (2) since the weighted least squares estimators are  $\sqrt{n}$ -consistent (see Lemma C.1 in [19] and also (52) in the proof of Corollary 3.1). One can then estimate  $\xi_{\tau_n}(\varepsilon)$  by the direct LAWS estimator  $\hat{\xi}_{\tau_n}(\varepsilon)$  described in Section 2.1. The consistency and asymptotic normality of  $\hat{\xi}_{\tau_n}(\varepsilon)$  are therefore a corollary of Theorem 2.1, and this in turn yields the asymptotic behaviour of the estimators

$$\begin{aligned} \hat{\xi}_{\tau_n}(Y|\mathbf{x}) &= \hat{\alpha} + \hat{\beta}^\top \mathbf{x} + (1 + \hat{\theta}^\top \mathbf{x}) \hat{\xi}_{\tau_n}(\varepsilon) \quad (\text{intermediate level}) \\ \text{and } \hat{\xi}_{\tau'_n}^*(Y|\mathbf{x}) &= \hat{\alpha} + \hat{\beta}^\top \mathbf{x} + (1 + \hat{\theta}^\top \mathbf{x}) \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}} \hat{\xi}_{\tau_n}(\varepsilon) \quad (\text{extreme level}) \end{aligned}$$

where  $\bar{\gamma}$  is a consistent estimator of  $\gamma$  constructed on the residuals.

COROLLARY 3.1. *Assume that the setup is that of the heteroscedastic linear model ( $M_1$ ). Assume that  $\varepsilon$  satisfies condition  $\mathcal{C}_1(\gamma)$  with  $0 < \gamma < 1/2$ . Suppose also that  $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$  for some  $\delta > 0$ , and that  $\tau_n \uparrow 1$  with  $n(1 - \tau_n) \rightarrow \infty$ .*

$$(i) \text{ Then for any } \mathbf{x} \in K, \sqrt{n(1 - \tau_n)} \left( \frac{\hat{\xi}_{\tau_n}(Y|\mathbf{x})}{\hat{\xi}_{\tau_n}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

(ii) *Assume further that  $\varepsilon$  satisfies condition  $\mathcal{C}_2(\gamma, \rho, A)$  with  $\rho < 0$ . Suppose also that  $\tau_n, \tau'_n \uparrow 1$  satisfy (3) and (4). If there is a nondegenerate limiting random variable  $\Gamma$  such that  $\sqrt{n(1 - \tau_n)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$ , then*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left( \frac{\hat{\xi}_{\tau'_n}^*(Y|\mathbf{x})}{\hat{\xi}_{\tau'_n}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

We may similarly obtain the asymptotic normality of the indirect estimators  $\tilde{\xi}_{\tau_n}(Y|\mathbf{x})$  and  $\tilde{\xi}_{\tau'_n}^*(Y|\mathbf{x})$  of the intermediate and extreme expectiles  $\xi_{\tau_n}(Y|\mathbf{x})$  and  $\xi_{\tau'_n}(Y|\mathbf{x})$ , defined as

$$\tilde{\xi}_{\tau_n}(Y|\mathbf{x}) = \hat{\alpha} + \hat{\beta}^\top \mathbf{x} + (1 + \hat{\theta}^\top \mathbf{x})(\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \hat{\varepsilon}_{n - \lfloor n(1 - \tau_n) \rfloor, n}^{(n)}$$

$$\text{and } \tilde{\xi}_{\tau_n}^*(Y|\mathbf{x}) = \hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \hat{\boldsymbol{\theta}}^\top \mathbf{x}) \left( \frac{1 - \tau_n'}{1 - \tau_n} \right)^{-\bar{\gamma}} (\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \varepsilon_{n - \lfloor n(1 - \tau_n) \rfloor, n}^{(n)}.$$

Here  $\bar{\gamma}$  is the residual-based Hill estimator of  $\gamma$ ; the asymptotic properties of the estimators are obtained using Corollary 2.1 and Theorem 2.3. See Corollary E.1 in [19].

**REMARK 6.** Corollary 3.1 requires a second moment of the noise variable  $\varepsilon$  because of the use of the weighted least squares method and the residual-based LAWS estimator of intermediate expectiles. The R package `CASdatasets` contains numerous examples of real actuarial data sets for which the assumption of a finite variance is perfectly sensible. When this assumption is violated, the alternative is to use a more robust method for the estimation of the model structure and then use the indirect expectile estimator of Section 2.2. A more robust method for the estimation of  $\alpha$  and  $\beta$  is, for instance, the one-step estimator of [39]. Such methods typically require some regularity on the joint distribution of  $(\mathbf{X}, \varepsilon)$ , but avoid moment assumptions. The convergence of the indirect expectile-based estimator built on the residuals will then only require a finite first moment, see Corollary 2.1 and Theorem 2.3.

**3.2. Heteroscedastic single-index model.** A model with greater flexibility is the heteroscedastic single-index model; the single-index structure allows to handle complicated regression equations in a satisfactory way, including when the dimension  $d$  is large.

**Model ( $M_2$ )** The random pair  $(\mathbf{X}, Y)$  is such that  $Y = g(\boldsymbol{\beta}^\top \mathbf{X}) + \sigma(\boldsymbol{\beta}^\top \mathbf{X})\varepsilon$ . Here  $g$  and  $\sigma > 0$  are measurable functions. The random covariate  $\mathbf{X}$  is independent of the noise variable  $\varepsilon$ , and has a density function  $f_{\mathbf{X}}$  on  $\mathbb{R}^d$  whose support is a compact and convex set  $K$  with nonempty interior  $K^\circ$ . Besides, the variable  $\varepsilon$  is centred and such that  $\mathbb{E}|\varepsilon| = 1$ .

For identifiability purposes, we will assume that  $g$  is continuously differentiable,  $\|\boldsymbol{\beta}\| = 1$  (where  $\|\cdot\|$  denotes the Euclidean norm) and that the first non-zero component of  $\boldsymbol{\beta}$  is positive. This guarantees that  $\boldsymbol{\beta}$  is identifiable. Other sets of identifiability conditions are possible, see *e.g.* [28]. In this regression model, the conditional mean and variance have the same single-index structure. There are analogue models where the direction of projection in  $\sigma$  is a vector  $\boldsymbol{\theta}$  possibly different from  $\boldsymbol{\beta}$  (see *e.g.* [54]). In practice, model ( $M_2$ ) is already very flexible, and for the sake of simplicity we therefore ignore this more general case; in the latter, the direction in the variance component can be estimated at the  $\sqrt{n}$ -rate (see Theorem 1 in [54]), and it is readily checked that our methodology below extends to this case.

In model ( $M_2$ ),  $\xi_{\tau_n}(Y|\mathbf{x}) = g(\boldsymbol{\beta}^\top \mathbf{x}) + \sigma(\boldsymbol{\beta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon)$ . There are numerous  $\sqrt{n}$ -consistent estimators of  $\boldsymbol{\beta}$  (see *e.g.* Chapter 2 of [28]). We thus assume that such an estimator  $\hat{\boldsymbol{\beta}}$  has been constructed, *i.e.*  $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_{\mathbb{P}}(1)$ . Estimate now  $g$  with

$$\hat{g}_{h_n, t_n}(z) = \frac{\sum_{i=1}^n Y_i \mathbb{1}\{|Y_i| \leq t_n\} L\left(\frac{z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right)}{\sum_{i=1}^n L\left(\frac{z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right)}.$$

Here  $L$  is a probability density function on  $\mathbb{R}$ ,  $h_n \rightarrow 0$  is a bandwidth sequence and  $t_n \rightarrow \infty$  is a positive truncating sequence. This is inspired by an estimator of [22]; truncating helps in dealing with heavy tails. Besides, analogously to what we observed in model ( $M_1$ ),  $\sigma(\boldsymbol{\beta}^\top \mathbf{X})$  is the conditional first moment of  $|Y - g(\boldsymbol{\beta}^\top \mathbf{X})|$ . Introduce then absolute residuals  $\hat{Z}_{i, h_n, t_n} = |Y_i - \hat{g}_{h_n, t_n}(\hat{\boldsymbol{\beta}}^\top \mathbf{X}_i)|$  and consider a Nadaraya-Watson-type estimator:

$$\hat{\sigma}_{h_n, t_n}(z) = \frac{\sum_{i=1}^n \hat{Z}_{i, h_n, t_n} \mathbb{1}\{\hat{Z}_{i, h_n, t_n} \leq t_n\} L\left(\frac{z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right)}{\sum_{i=1}^n L\left(\frac{z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right)}.$$

In Proposition C.1 (see [19]) we show that, under conditions tailored to our framework, both of these estimators converge uniformly on any compact subset  $K_0$  of the interior of the support of  $\mathbf{X}$  at the rate  $n^{2/5}/\sqrt{\log n}$  under the condition  $nh_n^5 \rightarrow c \in (0, \infty)$ . Similar results, mostly on the estimation of the link function  $g$ , are available in the literature; see for example [32] for an estimator based on smoothing splines, as well as references therein.

The residuals are then  $\widehat{\varepsilon}_i^{(n)} = (Y_i - \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i)) / \widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i)$ . Translated in terms of these residuals, Proposition C.1 of [19] reads

$$\frac{n^{2/5}}{\sqrt{\log n}} \max_{1 \leq i \leq n} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{1 + |\varepsilon_i|} \mathbb{1}\{\mathbf{X}_i \in K_0\} = O_{\mathbb{P}}(1),$$

for any compact subset  $K_0$  of the interior of the support of  $\mathbf{X}$ . The restriction to such a compact subset makes sense since kernel regression estimators strongly suffer from boundary effects (see, among many others, [33]). This restriction is not important in practice since one would only trust the estimates of  $g$  and  $\sigma$  on a sub-domain of the support where sufficiently many observations from  $\mathbf{X}$  have been recorded. It implies, however, that the residuals  $\widehat{\varepsilon}_i^{(n)}$  that can be used for the estimation of the high conditional expectile are those for which  $\mathbf{X}_i \in K_0$ . More precisely, let  $\widehat{\varepsilon}_{1, K_0}^{(n)}, \dots, \widehat{\varepsilon}_{N, K_0}^{(n)}$  be those residuals whose corresponding covariate vectors  $\mathbf{X}_i \in K_0$  and  $N = N(K_0, n) = \sum_{i=1}^n \mathbb{1}\{\mathbf{X}_i \in K_0\}$  be their total number. Define

$$\widehat{\xi}_{\tau_N}(\varepsilon) = \arg \min_{u \in \mathbb{R}} \sum_{i=1}^N \eta_{\tau_N}(\widehat{\varepsilon}_{i, K_0}^{(n)} - u),$$

with  $\tau_N = \tau_m$  when  $N = m > 0$ . This yields the estimators

$$\widehat{\xi}_{\tau_N}(Y|\mathbf{x}) = \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) + \widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) \widehat{\xi}_{\tau_N}(\varepsilon) \text{ (intermediate level)}$$

$$\text{and } \widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x}) = \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) + \widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) \left( \frac{1 - \tau'_N}{1 - \tau_N} \right)^{-\bar{\gamma}} \widehat{\xi}_{\tau_N}(\varepsilon) \text{ (extreme level).}$$

Again, the estimator  $\bar{\gamma}$  is typically calculated using high order statistics of the residuals  $\widehat{\varepsilon}_{i, K_0}^{(n)}$ ; for example, this can be the Hill estimator taking into account the top  $\lfloor N(1 - \tau_N) \rfloor$  order statistics of these residuals (see Lemma C.6(ii) in [19] for the asymptotic properties of this estimator). The next result focuses on the estimators  $\widehat{\xi}_{\tau_N}(Y|\mathbf{x})$  and  $\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x})$ .

**THEOREM 3.1.** *Work in model  $(M_2)$ . Assume that  $\varepsilon$  satisfies condition  $\mathcal{C}_1(\gamma)$  with  $0 < \gamma < 1/2$  and the conditions of Proposition C.1 in [19] hold. Let  $\tau_n = 1 - n^{-a}$  with  $a \in (1/5, 1)$ ,  $K_0$  be a compact subset of  $K^\circ$  such that  $\mathbb{P}(\mathbf{X} \in K_0) > 0$ , and  $N = N(K_0, n)$ .*

$$(i) \text{ We have, for any } \mathbf{x} \in K_0, \sqrt{N(1 - \tau_N)} \left( \frac{\widehat{\xi}_{\tau_N}(Y|\mathbf{x})}{\widehat{\xi}_{\tau_N}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

(ii) *Assume moreover that  $\varepsilon$  satisfies condition  $\mathcal{C}_2(\gamma, \rho, A)$  with  $\rho < 0$ . Suppose also that  $\tau_n, \tau'_n \uparrow 1$  satisfy (3) and (4). If there is a nondegenerate limiting random variable  $\Gamma$  such that  $\sqrt{N(1 - \tau_N)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$ , then for any  $\mathbf{x} \in K_0$ ,*

$$\frac{\sqrt{N(1 - \tau_N)}}{\log[(1 - \tau_N)/(1 - \tau'_N)]} \left( \frac{\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x})}{\widehat{\xi}_{\tau'_N}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \Gamma.$$

**REMARK 7.** Compared to Corollary 3.1, Theorem 3.1 features the additional restriction  $\tau_n = 1 - n^{-a}$  with  $a \in (1/5, 1)$ . This means that the intermediate expectile to be estimated has to be high enough so that the rate of (semiparametric) estimation of the structure of the model is faster than that of the intermediate expectile and the extreme value index  $\gamma$ .

REMARK 8. In Theorem 3.1, the order of the conditional expectile to be estimated and rates of convergence are random and dictated by the number  $N = N(K_0, n)$  of covariates  $\mathbf{X}_i \in K_0$  (where model structure can be estimated at the rate  $n^{2/5}/\sqrt{\log n}$ ). Random convergence rates are not unusual in situations where the effective sample size is random: see, for example, Corollary 1.1 in [43] and Theorem 3 in [52] in the context of randomly truncated observations. The random rate of convergence  $\sqrt{N(1 - \tau_N)}$  in Theorem 3.1 can nonetheless be replaced by a nonrandom rate because, with the notation of Theorem 3.1 and if  $p_0 = \mathbb{P}(\mathbf{X} \in K_0)$ ,  $\sqrt{N(1 - \tau_N)} = [np_0]^{(1-a)/2}(1 + o_{\mathbb{P}}(1))$  by the law of large numbers. Similarly, in convergence (ii) and if  $\tau'_n = 1 - n^{-b}$  with  $b > a$ , one can replace  $1 - \tau'_N$  by the nonrandom sequence  $1 - \tau'_{np_0} = (np_0)^{-b}$  and the rate of convergence in (ii) can be substituted with the nonrandom rate of convergence  $[np_0]^{(1-a)/2}/[(b-a)\log(np_0)]$ .

Let us finally mention that, if  $\bar{\gamma}$  is the residual-based Hill estimator, an analogous result (Theorem E.1 in [19]) holds for the indirect extreme conditional expectile estimator

$$\tilde{\xi}_{\tau'_N}^*(Y|\mathbf{x}) = \hat{g}_{h_n, t_n}(\hat{\beta}^\top \mathbf{x}) + \hat{\sigma}_{h_n, t_n}(\hat{\beta}^\top \mathbf{x}) \left( \frac{1 - \tau'_N}{1 - \tau_N} \right)^{-\bar{\gamma}} (\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \varepsilon_{N - \lfloor N(1 - \tau_N) \rfloor, N, K_0}^{(n)}.$$

Again, its asymptotic distribution is controlled by that of  $\bar{\gamma}$ .

3.3. *Heteroscedastic left-censored (Tobit) regression model.* We briefly discuss how the assumption that our model describes globally the structure of  $(\mathbf{X}, Y)$  can be relaxed, through the example of the left-censored regression model below.

**Model** ( $M_3$ ) The random pair  $(\mathbf{X}, Y)$  satisfies  $Y = g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon$  when  $g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon > y_0$ , and  $Y = y_0$  otherwise. Here  $y_0$  is known and  $g$  and  $\sigma > 0$  are measurable functions. The random covariate  $\mathbf{X} \in \mathbb{R}^d$  is independent of the centred noise variable  $\varepsilon$  such that  $\mathbb{E}|\varepsilon| = 1$ . On the support of  $\mathbf{X}$ , the functions  $g$  and  $\sigma$  are bounded and  $\sigma$  is bounded away from 0.

When  $g$  is linear and  $\sigma$  is constant, this is the Tobit model of [45] with non-Gaussian errors. The heteroscedastic case is considered in *e.g.* [34, 40], where it is shown how a linear  $g$  can be estimated at the  $\sqrt{n}$ -rate, with standard nonparametric rates obtained under no assumption on  $g$ . Such models are important in economics (see [45]) and insurance (to model a net loss, *i.e.* claim amount minus deductible when the former exceeds the latter, and 0 otherwise).

Here, if  $\varepsilon$  is heavy-tailed, there is  $\tau_c \in (0, 1)$  such that for  $\tau \in [\tau_c, 1]$ , the conditional quantile function of  $Y$  given  $\mathbf{X}$  satisfies  $q_\tau(Y|\mathbf{x}) = g(\mathbf{x}) + \sigma(\mathbf{x})q_\tau(\varepsilon)$  (see Lemma C.7(i) of [19]), linking model ( $M_3$ ) to the tail regression models of [48, 50]. We do not have an analogue formula for expectiles because they are not equivariant by taking increasing transformations, but

$$\xi_\tau(Y|\mathbf{x}) \approx \xi_\tau(g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon|\mathbf{X} = \mathbf{x}) = g(\mathbf{x}) + \sigma(\mathbf{x})\xi_\tau(\varepsilon) \text{ (see Lemma C.7(ii) of [19])}$$

as  $\tau \uparrow 1$ , which is much weaker than the relationship  $\xi_\tau(Y|\mathbf{x}) = g(\mathbf{x}) + \sigma(\mathbf{x})\xi_\tau(\varepsilon)$  true when the regression model is valid globally. It is also weaker than a specification of the form  $\xi_\tau(Y|\mathbf{x}) = r(\mathbf{x}) + \xi_\tau(\varepsilon)$  for  $\tau \in [\tau_c, 1]$ , which would be an expectile-based version of the model of [48].

Assume that there are estimators  $\hat{g}$  of  $g$  and  $\hat{\sigma}$  of  $\sigma$  which are  $v_n$ -uniformly consistent (for some  $v_n \rightarrow \infty$ ) on a measurable subset  $K_0$  of the support of  $\mathbf{X}$  such that  $\mathbb{P}(\mathbf{X} \in K_0) > 0$ . Let  $(\mathcal{X}_i, \mathcal{Y}_i, e_i)$  stand for all those  $N$  vectors (where  $N$  is random) relative to noncensored observations with covariate vectors in  $K_0$ , *i.e.*  $\mathcal{Y}_i = g(\mathcal{X}_i) + \sigma(\mathcal{X}_i)e_i$  and  $\mathcal{X}_i \in K_0$  for  $1 \leq i \leq N$ . Construct residuals as  $\hat{e}_i^{(N)} = (\mathcal{Y}_i - \hat{g}(\mathcal{X}_i))/\hat{\sigma}(\mathcal{X}_i)$ . These approximate unobservable  $e_i$  that, given  $N = m > 0$ , are  $m$  i.i.d. copies of a random variable  $e$  such that  $\mathbb{P}(e > t) =$

$p^{-1}\mathbb{P}(\varepsilon > t)$  for  $t$  large enough, where  $p = \mathbb{P}(\varepsilon > (y_0 - g(\mathbf{X}))/\sigma(\mathbf{X}) \mid \mathbf{X} \in K_0) > 0$  (see Lemma C.7(iii) of [19]). In particular one easily shows that if  $\varepsilon$  has extreme value index  $\gamma$ , then  $e$  has too, and  $\xi_\tau(\varepsilon)/\xi_\tau(e) \rightarrow p^\gamma$  as  $\tau \uparrow 1$  (see Lemma C.7(iv) of [19]). Let  $N_0 = \sum_{i=1}^n \mathbb{1}\{\mathbf{X}_i \in K_0\}$ . The fact that  $N/N_0$  is a  $\sqrt{n}$ -consistent estimator of  $p$  motivates the estimators

$$\widehat{\xi}_{\tau_N}(Y|\mathbf{x}) = \widehat{g}(\mathbf{x}) + \widehat{\sigma}(\mathbf{x}) \left( \frac{N}{N_0} \right)^{\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor}} \widehat{\xi}_{\tau_N}(e) \text{ (intermediate level)}$$

$$\text{and } \widehat{\xi}_{\tau_N}^*(Y|\mathbf{x}) = \widehat{g}(\mathbf{x}) + \widehat{\sigma}(\mathbf{x}) \left( \frac{N}{N_0} \right)^{\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor}} \left( \frac{1 - \tau'_N}{1 - \tau_N} \right)^{-\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor}} \widehat{\xi}_{\tau_N}(e) \text{ (extreme level)}$$

where  $\widehat{\xi}_{\tau_N}(e)$  is the LAWS estimator of the expectile of  $e$  at level  $\tau_N$ , based on the residuals  $\widehat{e}_i^{(N)}$ , and  $\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor}$  is the Hill estimator based on the top  $\lfloor N(1-\tau_N) \rfloor$  elements of these same residuals. We examine the convergence of the above estimators next.

**THEOREM 3.2.** *Work in model  $(M_3)$ . Assume that  $\varepsilon$  satisfies condition  $\mathcal{C}_2(\gamma, \rho, A)$  with  $0 < \gamma < 1/2$ . Suppose also that  $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$  for some  $\delta > 0$ , and suppose that  $\widehat{g}$  and  $\widehat{\sigma}$  are  $v_n$ -uniformly consistent estimators (here  $v_n \rightarrow \infty$ ) of  $g$  and  $\sigma$  on  $K_0$  with  $\mathbb{P}(\mathbf{X} \in K_0) > 0$ . Let  $\tau_n = 1 - n^{-a}$  with  $a \in (0, 1)$  and assume that  $n^{1-a}/v_n^2 \rightarrow 0$ .*

(i) *If  $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \rightarrow \lambda \in \mathbb{R}$  and  $\sqrt{n(1-\tau_n)}/q_{\tau_n}(\varepsilon) \rightarrow \mu \in \mathbb{R}$  then, for any  $\mathbf{x} \in K_0$ ,*

$$\sqrt{N(1-\tau_N)} \left( \frac{\widehat{\xi}_{\tau_N}(Y|\mathbf{x})}{\xi_{\tau_N}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N}(\mathbf{b}(\gamma, \rho, p, \mathbf{x}), \mathbf{v}(\gamma, p)) \text{ with}$$

$$\mathbf{b}(\gamma, \rho, p, \mathbf{x})$$

$$= \gamma(\gamma^{-1} - 1)^\gamma \left( p^\gamma \mathbb{E} \left[ \varepsilon \mid \varepsilon > \frac{y_0 - g(\mathbf{X})}{\sigma(\mathbf{X})}, \mathbf{X} \in K_0 \right] - \mathbb{E} \left[ \max \left( \varepsilon, \frac{y_0 - g(\mathbf{x})}{\sigma(\mathbf{x})} \right) \right] \right) \mu$$

$$+ \left\{ \frac{p^{-\rho} \log p}{1 - \rho} + \frac{p^{-\rho} - 1}{\rho} \left( 1 + \rho \left[ \frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} \right] \right) \right\} \lambda \text{ and}$$

$$\mathbf{v}(\gamma, p) = \frac{2\gamma^3}{1 - 2\gamma} + 2 \log p \frac{\gamma^3(\gamma^{-1} - 1)^\gamma}{(1 - \gamma)^2} + (\log p)^2 \gamma^2.$$

(ii) *Assume moreover that  $\rho < 0$  and  $\tau_n, \tau'_n \uparrow 1$  satisfy (3) and (4). Then, for any  $\mathbf{x} \in K_0$ ,*

$$\frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau'_N)]} \left( \frac{\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x})}{\xi_{\tau'_N}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left( p^{-\rho} \frac{\lambda}{1 - \rho}, \gamma^2 \right).$$

Note that when observations with  $\mathbf{X} \in K_0$  are never censored, we find  $p = 1$ ,  $N = \sum_{i=1}^n \mathbb{1}\{\mathbf{X}_i \in K_0\}$ ,  $\mathbf{b}(\gamma, \rho, 1, \mathbf{x}) = 0$  (because  $\mathbb{E}(\varepsilon) = 0$ ) and  $\mathbf{v}(\gamma, 1) = 2\gamma^3/(1 - 2\gamma)$ , which then makes convergence (i) above analogous to Theorem 3.1(i). As expected, the asymptotic distribution in (ii) is identical to that of the classical Weissman-Hill estimator when  $p = 1$ .

**3.4. Time series models.** Expectiles can be interpreted in terms of the gain-loss ratio. This is a popular performance measure in portfolio management, well-known in the literature on no good deal valuation in incomplete markets (see [3] and references therein). Financial applications typically require working with stationary but dependent time series data. We present here, in two such time series contexts, applications of our results to the dynamic prediction of extreme expectiles given past observations. We only focus on LAWS estimators; extensions of our theory to indirect expectile estimation can be found in Appendix E of [19].

3.4.1. *The ARMA model.* We start with the following general ARMA( $p, q$ ) model.

**Model** ( $T_1$ ) The stationary time series  $(Y_t)_{t \in \mathbb{Z}}$  satisfies  $Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$  where  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q \in \mathbb{R}$  are unknown coefficients. The polynomials  $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$  and  $Q(z) = 1 + \sum_{j=1}^q \theta_j z^j$  have no common root, and no root inside the unit disk of the complex plane. Finally,  $(\varepsilon_t)$  is an i.i.d. sequence of copies of  $\varepsilon$  such that  $\mathbb{E}(\varepsilon) = 0$ ,  $\mathbb{E}(\varepsilon^2) < \infty$ , and  $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \rightarrow \ell \in (0, 1]$  as  $x \rightarrow \infty$ .

In model ( $T_1$ ), the process  $(Y_t)$  is causal and invertible, and so can be represented as a linear time series in the  $\varepsilon_{t-j}$ ,  $j \geq 0$ , by Theorem 3.1.1 in [4]. A conditional one-step ahead expectile based on data up to time  $n$  is then  $\xi_{\tau_n}(Y_{n+1} | \mathcal{F}_n) = \sum_{j=1}^p \phi_j Y_{n+1-j} + \sum_{j=1}^q \theta_j \varepsilon_{n+1-j} + \xi_{\tau_n}(\varepsilon)$  where  $\mathcal{F}_n = \sigma(Y_n, Y_{n-1}, \dots)$  is the past  $\sigma$ -field at time  $n$ . In general,  $\sum_{j=1}^p \phi_j Y_{n+1-j} + \sum_{j=1}^q \theta_j \varepsilon_{n+1-j}$  depends on the unobservable  $\varepsilon_n, \dots, \varepsilon_{n+1-q}$ , which are all linear functions of  $(Y_{n+1-j})_{j \geq 1}$  since  $(Y_t)$  is an invertible ARMA process. This is why the dynamic expectile  $\xi_{\tau_n}(Y_{n+1} | \mathcal{F}_n)$  to be estimated is conditional upon the whole past  $\mathcal{F}_n$  of the process; in the AR( $p$ ) case when  $q = 0$ , this becomes the simpler conditional expectile  $\xi_{\tau_n}(Y_{n+1} | Y_n, Y_{n-1}, \dots, Y_{n-p+1})$ , determined by the past  $p$  values only.

Among others, the Gaussian maximum likelihood estimator and the ordinary least squares estimator of the  $\phi_j$  and  $\theta_j$  are  $\sqrt{n}$ -asymptotically normal because  $\mathbb{E}(\varepsilon^2) < \infty$  (see Theorem 10.8.2 in [4]). We then assume that the estimators  $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}, \hat{\theta}_{1,n}, \dots, \hat{\theta}_{q,n}$  are such that  $\hat{\phi}_{j,n} = \phi_j + O_{\mathbb{P}}(n^{-1/2})$  and  $\hat{\theta}_{j,n} = \theta_j + O_{\mathbb{P}}(n^{-1/2})$ . To construct residuals, set  $\hat{\varepsilon}_{\max(p,q)-q+1}^{(n)} = \dots = \hat{\varepsilon}_{\max(p,q)}^{(n)} = 0$  and define  $\hat{\varepsilon}_t^{(n)} = Y_t - \sum_{j=1}^p \hat{\phi}_{j,n} Y_{t-j} - \sum_{j=1}^q \hat{\theta}_{j,n} \hat{\varepsilon}_{t-j}^{(n)}$ , for  $\max(p, q) + 1 \leq t \leq n$ . We consider the asymptotic behaviour of the estimators

$$\begin{aligned} \hat{\xi}_{\tau_n}(Y_{n+1} | \mathcal{F}_n) &= \sum_{j=1}^p \hat{\phi}_{j,n} Y_{n+1-j} + \sum_{j=1}^q \hat{\theta}_{j,n} \hat{\varepsilon}_{n+1-j}^{(n)} + \hat{\xi}_{\tau_n}(\varepsilon) \quad (\tau_n \text{ intermediate}), \\ \hat{\xi}_{\tau'_n}^*(Y_{n+1} | \mathcal{F}_n) &= \sum_{j=1}^p \hat{\phi}_{j,n} Y_{n+1-j} + \sum_{j=1}^q \hat{\theta}_{j,n} \hat{\varepsilon}_{n+1-j}^{(n)} + \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}} \hat{\xi}_{\tau_n}(\varepsilon) \quad (\tau'_n \text{ extreme}), \end{aligned}$$

where  $\hat{\xi}_{\tau_n}(\varepsilon)$  is the LAWS estimator of  $\xi_{\tau_n}(\varepsilon)$  and  $\bar{\gamma}$  is a consistent estimator of  $\gamma$ , both constructed on the residuals  $\hat{\varepsilon}_t^{(n)}$  for  $t_n \leq t \leq n$  only, where  $t_n / \log n \rightarrow \infty$  and  $t_n / n \rightarrow 0$ . This condition on  $t_n$  ensures that the influence of the incorrect starting values for the residuals has vanished; in the autoregressive case  $q = 0$ , one can use all the  $\hat{\varepsilon}_t^{(n)}$  for  $p + 1 \leq t \leq n$ .

**THEOREM 3.3.** *Work in the ARMA model ( $T_1$ ). Assume that  $\varepsilon$  satisfies condition  $\mathcal{C}_1(\gamma)$  with  $0 < \gamma < 1/2$ . Suppose also that there is  $\delta > 0$  such that  $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$ , and that  $\tau_n \uparrow 1$  is such that  $n(1 - \tau_n) \rightarrow \infty$ .*

(i) *If  $n^{2\gamma+\iota}(1 - \tau_n) \rightarrow 0$  for some  $\iota > 0$ , then*

$$\sqrt{n(1 - \tau_n)} \left( \begin{array}{c} \hat{\xi}_{\tau_n}(Y_{n+1} | \mathcal{F}_n) \\ \xi_{\tau_n}(Y_{n+1} | \mathcal{F}_n) \end{array} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

(ii) *Assume further that  $\varepsilon$  satisfies condition  $\mathcal{C}_2(\gamma, \rho, A)$  with  $\rho < 0$ . Suppose also that  $\tau_n, \tau'_n \uparrow 1$  satisfy (3) and (4) (in addition to  $n^{2\gamma+\iota}(1 - \tau_n) \rightarrow 0$ ). If there is a nondegenerate limiting random variable  $\Gamma$  such that  $\sqrt{n(1 - \tau_n)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$ , then*

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left( \begin{array}{c} \hat{\xi}_{\tau'_n}^*(Y_{n+1} | \mathcal{F}_n) \\ \xi_{\tau'_n}(Y_{n+1} | \mathcal{F}_n) \end{array} - 1 \right) \xrightarrow{d} \Gamma.$$

3.4.2. *The GARCH model.* ARMA models are widely applicable but well-known for failing to replicate the time-varying volatility typically displayed by financial time series. Our next focus is on general GARCH( $p, q$ ) models, which arguably constitute the best-known and most employed class of heteroscedastic time series models.

**Model ( $T_2$ )** The stationary time series  $(Y_t)_{t \in \mathbb{Z}}$  satisfies  $Y_t = \sigma_t \varepsilon_t$ , with  $\sigma_t > 0$  such that  $\sigma_t^2 = \omega + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{j=1}^q \alpha_j Y_{t-j}^2$  and  $\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p > 0$  are unknown coefficients, and  $(\varepsilon_t)$  is an i.i.d. sequence of copies of  $\varepsilon$  such that  $\mathbb{E}(\varepsilon) = 0$ ,  $\mathbb{E}(\varepsilon^2) = 1$  and  $\mathbb{P}(\varepsilon^2 = 1) < 1$ . Suppose also that the sequence of matrices

$$A_t = \begin{pmatrix} \alpha_1 \varepsilon_t^2 & \cdots & \cdots & \cdots & \alpha_q \varepsilon_t^2 & \beta_1 \varepsilon_t^2 & \cdots & \cdots & \cdots & \beta_p \varepsilon_t^2 \\ 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \alpha_1 & \cdots & \cdots & \cdots & \alpha_q & \beta_1 & \cdots & \cdots & \cdots & \beta_p \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

has a negative top Lyapunov exponent, *i.e.*  $\lim_{t \rightarrow \infty} t^{-1} \mathbb{E}(\log \|A_t A_{t-1} \cdots A_1\|) < 0$  with probability 1 (where  $\|\cdot\|$  is an arbitrary matrix norm).

The above condition on  $(A_t)$  is necessary and sufficient for the existence of a stationary, nonanticipative solution, see Theorem 2.4 p.30 of [16]. Condition  $\mathbb{P}(\varepsilon^2 = 1) < 1$  ensures identifiability. In pure ARCH models ( $p = 0$ ), one can estimate the model with weighted least squares regression of  $Y_t^2$  on its past. This estimator is  $\sqrt{n}$ -asymptotically normal if  $\mathbb{E}(Y_t^4) < \infty$  (see Theorem 6.3 p.132 in [16]). Under further conditions on model coefficients (see p.41 of [16]), this may reduce to  $\mathbb{E}(\varepsilon^4) < \infty$ , but this is still a substantial restriction in our context of heavy-tailed  $\varepsilon$ . An alternative is the weighted  $L^1$ -regression estimator of [29], whose  $\sqrt{n}$ -asymptotic normality requires some regularity on the distribution of  $\varepsilon$  rather than finite moments. In GARCH models, the self-weighted quasi-maximum exponential likelihood estimator of [53] is  $\sqrt{n}$ -asymptotically normal for square-integrable innovations.

Take  $\sqrt{n}$ -consistent estimators  $\hat{\omega}_n, \hat{\alpha}_{j,n}$  and  $\hat{\beta}_{j,n}$ . To construct residuals, set  $\hat{\sigma}_{\max(p,q)-p+1}^{(n)} = \cdots = \hat{\sigma}_{\max(p,q)}^{(n)} = \hat{\omega}_n$ , and then define  $(\hat{\sigma}_t^{(n)})^2 = \hat{\omega}_n + \sum_{j=1}^p \hat{\beta}_{j,n} (\hat{\sigma}_{t-j}^{(n)})^2 + \sum_{j=1}^q \hat{\alpha}_{j,n} Y_{t-j}^2$  and  $\hat{\varepsilon}_t^{(n)} = Y_t / \hat{\sigma}_t^{(n)}$ , for  $\max(p, q) + 1 \leq t \leq n$ . Denoting again by  $\mathcal{F}_n$  the past  $\sigma$ -field and letting  $\hat{\sigma}_{n+1}^2 = \hat{\omega}_n + \sum_{j=1}^p \hat{\beta}_{j,n} \hat{\sigma}_{n+1-j}^2 + \sum_{j=1}^q \hat{\alpha}_{j,n} Y_{n+1-j}^2$  be the predicted volatility at time  $n + 1$ , one-step ahead estimators of intermediate and extreme conditional expectiles are

$$\hat{\xi}_{\tau_n}(Y_{n+1} | \mathcal{F}_n) = \hat{\sigma}_{n+1} \hat{\xi}_{\tau_n}(\varepsilon) \quad \text{and} \quad \hat{\xi}_{\tau_n}^*(Y_{n+1} | \mathcal{F}_n) = \hat{\sigma}_{n+1} \times \left( \frac{1 - \tau_n'}{1 - \tau_n} \right)^{-\bar{\gamma}} \hat{\xi}_{\tau_n}(\varepsilon)$$

respectively, where  $\hat{\xi}_{\tau_n}(\varepsilon)$  is the LAWS estimator of  $\xi_{\tau_n}(\varepsilon)$  and  $\bar{\gamma}$  is a consistent estimator of  $\gamma$ , both constructed on the residuals  $\hat{\varepsilon}_t^{(n)}$  for  $t_n \leq t \leq n$  only, where  $t_n / \log n \rightarrow \infty$  and  $t_n / n \rightarrow 0$  (for pure ARCH models when  $p = 0$ , all residuals for  $q + 1 \leq t \leq n$  may be used).

**THEOREM 3.4.** *Work in the GARCH model ( $T_2$ ). Assume that  $\varepsilon$  satisfies condition  $\mathcal{C}_1(\gamma)$  with  $0 < \gamma < 1/2$ . Suppose also that there is  $\delta > 0$  such that  $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$ , and that  $\tau_n = 1 - n^{-a}$  for some  $a \in (0, 1)$ .*



- (i) Then  $\sqrt{n(1-\tau_n)} \left( \frac{\widehat{\xi}_{\tau_n}(Y_{n+1} | \mathcal{F}_n)}{\xi_{\tau_n}(Y_{n+1} | \mathcal{F}_n)} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\gamma^3}{1-2\gamma} \right)$ .
- (ii) Assume further that  $\varepsilon$  satisfies condition  $\mathcal{C}_2(\gamma, \rho, A)$  with  $\rho < 0$ . Suppose also that  $\tau_n, \tau'_n \uparrow 1$  satisfy (3) and (4). If there is a nondegenerate limiting random variable  $\Gamma$  such that  $\sqrt{n(1-\tau_n)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$ , then

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \left( \frac{\widehat{\xi}_{\tau'_n}^*(Y_{n+1} | \mathcal{F}_n)}{\xi_{\tau'_n}^*(Y_{n+1} | \mathcal{F}_n)} - 1 \right) \xrightarrow{d} \Gamma.$$

**4. Finite-sample study.** We showcase our estimators on simulated data (Sections 4.1 and 4.2) and real data (Sections 4.3 and 4.4). Here we use, to estimate the extreme value index, the following bias-reduced version of the Hill estimator  $\widehat{\gamma}_k$ , see [21]:

$$\widehat{\gamma}_k^{\text{RB}} = \widehat{\gamma}_k \left( 1 - \frac{\widehat{b}}{1-\widehat{\rho}} \left( \frac{n}{k} \right)^{\widehat{\rho}} \right),$$

where throughout,  $k = \lfloor n(1-\tau_n) \rfloor$ , and  $\widehat{b}$  and  $\widehat{\rho}$  are consistent estimators of the quantities  $b$  and  $\rho$  under condition  $\mathcal{C}_2(\gamma, \rho, A)$  and the additional assumption that  $A(t) = b\gamma t^\rho$ . The estimators  $\widehat{b}$  and  $\widehat{\rho}$  may be found in [21] and are available from the R function `mop` in the R package `evt0`; of course, we shall use here their residual-based versions. We also consider the following bias-reduced version of the family of direct extreme expectile estimators of  $\varepsilon$ :

$$\begin{aligned} \widehat{\xi}_{\tau'_n}^{\text{RB}}(\varepsilon) &= \widehat{\xi}_{\tau'_n}^*(\varepsilon) \left( 1 + \frac{[n(1-\tau'_n)/k]^{-\widehat{\rho}} - 1}{\widehat{\rho}} \widehat{b} \widehat{\gamma}_k^{\text{RB}} \left( \frac{n}{k} \right)^{\widehat{\rho}} \right) \times \left( \frac{1 + \bar{r}^*(\tau'_n)}{1 + \bar{r}(1-k/n)} \right)^{-\widehat{\gamma}_k^{\text{RB}}} \\ &\quad \times \frac{\widehat{\rho} + [(1/\widehat{\gamma}_k^{\text{RB}} - 1)^{-\widehat{\rho}} [1 + \bar{r}^*(\tau'_n)]^{-\widehat{\rho}} - 1] \widehat{b} \widehat{\gamma}_k^{\text{RB}} (1-\tau'_n)^{-\widehat{\rho}}}{\widehat{\rho} + [(1/\widehat{\gamma}_k^{\text{RB}} - 1)^{-\widehat{\rho}} [1 + \bar{r}(1-k/n)]^{-\widehat{\rho}} - 1] \widehat{b} \widehat{\gamma}_k^{\text{RB}} (n/k)^{\widehat{\rho}}}, \text{ with} \\ \bar{r}(1-k/n) &= \left( 1 - \frac{\widehat{\xi}_{1/2}(\varepsilon)}{\widehat{\xi}_{1-k/n}(\varepsilon)} \right) \frac{1}{1-2k/n} \left( 1 + \frac{\widehat{b} [\widehat{F}_n(\widehat{\xi}_{1-k/n}(\varepsilon))]^{-\widehat{\rho}}}{1 - \widehat{\gamma}_k^{\text{RB}} - \widehat{\rho}} \right)^{-1} - 1, \text{ and} \\ \bar{r}^*(\tau'_n) &= \left( 1 - \frac{\widehat{\xi}_{1/2}(\varepsilon)}{\widehat{\xi}_{\tau'_n}^*(\varepsilon)} \right) \frac{1}{2\tau'_n - 1} \left( 1 + \frac{\widehat{b} (1/\widehat{\gamma}_k^{\text{RB}} - 1)^{-\widehat{\rho}}}{1 - \widehat{\gamma}_k^{\text{RB}} - \widehat{\rho}} (1-\tau'_n)^{-\widehat{\rho}} \right)^{-1} - 1. \end{aligned}$$

This expression is motivated by the proof of Proposition 1 and Corollary 1 in [9]; here  $\widehat{F}_n$  is the empirical survival function (*i.e.* complementary distribution function) of the residuals. We similarly consider the following bias-reduced version of the family of indirect estimators:

$$\begin{aligned} \widetilde{\xi}_{\tau'_n}^{\text{RB}}(\varepsilon) &= \widetilde{\xi}_{\tau'_n}^*(\varepsilon) \left( 1 + \frac{[n(1-\tau'_n)/k]^{-\widehat{\rho}} - 1}{\widehat{\rho}} \widehat{b} \widehat{\gamma}_k^{\text{RB}} \left( \frac{n}{k} \right)^{\widehat{\rho}} \right) \times [1 + \bar{r}^*(\tau'_n)]^{-\widehat{\gamma}_k^{\text{RB}}} \\ &\quad \times \left( 1 + \frac{(1/\widehat{\gamma}_k^{\text{RB}} - 1)^{-\widehat{\rho}} [1 + \bar{r}^*(\tau'_n)]^{-\widehat{\rho}} - 1}{\widehat{\rho}} \widehat{b} \widehat{\gamma}_k^{\text{RB}} (1-\tau'_n)^{-\widehat{\rho}} \right). \end{aligned}$$

These procedures improve the accuracy of our estimators, without affecting their asymptotic properties (see [20, 21]). They naturally give rise to extreme conditional expectile estimators  $\widehat{\xi}_{\tau'_n}^{\text{RB}}(Y|\mathbf{x})$  and  $\widetilde{\xi}_{\tau'_n}^{\text{RB}}(Y|\mathbf{x})$ , to which we refer in the present section.

4.1. *Simulation study: linear and single-index models.* We simulate  $N = 1,000$  samples of  $n = 1,000$  observations  $(\mathbf{X}_i, Y_i)$ ,  $1 \leq i \leq n$ . Here  $\mathbf{X} \in \mathbb{R}^4$ , with independent components, the first three being uniformly distributed on  $(0, 1)$ , and the fourth following a Beta(2, 1) distribution. We then simulate from two different models on  $(\mathbf{X}, Y)$ :

$$(G1) \quad Y = 1 + \beta^\top \mathbf{X} + (1/2 + \beta^\top \mathbf{X}) \varepsilon.$$

$$(G2) \quad Y = 1 + \exp(\beta^\top \mathbf{X} - 2) + (3/2 + \exp(\beta^\top \mathbf{X} - 2)) \varepsilon.$$

Model (G1) is a location-scale shift linear regression model, while model (G2) is a heteroscedastic single-index model. In both cases, the coefficient vector  $\beta = (1, 1, 1, 1)$  and  $\varepsilon$  is a noise variable, independent of  $\mathbf{X}$ , with a normalised symmetric Burr distribution, that is,  $\varepsilon = -\rho\varepsilon_0/B((\gamma - 1)/\rho, (\rho - \gamma)/\rho)$ , where  $B$  is the Beta function and  $\varepsilon_0$  has density

$$(5) \quad f_0(x) = (2\gamma)^{-1}|x|^{-\rho/\gamma-1}(1 + |x|^{-\rho/\gamma})^{1/\rho-1} \quad (x \in \mathbb{R}).$$

We consider the cases  $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$  and the second-order parameter  $\rho = -1$ .

Our aim is to estimate extreme expectiles  $\xi_{\tau'_n}(Y|\mathbf{x})$ , in both of these models. We compare the performances of several procedures, constructed using the following four strategies:

(S1) We assume that  $Y$  is linked to  $\mathbf{X}$  by a location-scale shift linear regression model, *i.e.*  $Y = \alpha + \beta^\top \mathbf{X} + (1 + \theta^\top \mathbf{X}) \varepsilon$ . The methodology used for the estimation of  $\xi_{\tau'_n}(Y|\mathbf{x})$  is outlined in Section 3.1, and the bias-reduced direct estimator is used.

(S1i) Identical to (S1), but the bias-reduced indirect estimator is used instead.

(S2) We assume that  $Y$  is linked to  $\mathbf{X}$  by the heteroscedastic single-index model  $Y = g(\beta^\top \mathbf{X}) + \sigma(\beta^\top \mathbf{X}) \varepsilon$ . The vector  $\beta$  is estimated using the algorithm of [54] (see 1.(a)–(c) on page 1240 therein), with  $g$  and  $\sigma$  estimated using the procedure described in Section 3.2, with  $h_n = 0.3$  and  $t_n = n^{2/5} \approx 15.85$ . The bias-reduced direct estimator is used.

(S2i) Identical to (S2), but the bias-reduced indirect estimator is used instead.

These procedures are compared with the following eight benchmarks:

(B1) We assume no specific structure on  $(\mathbf{X}, Y)$  and, at  $\mathbf{X} = \mathbf{x}$ , we use a local bias-reduced direct estimator relying on those  $Y_i$  whose  $\mathbf{X}_i$  are the 100 nearest neighbours of  $\mathbf{x}$ . In this procedure we use  $k = 20$ , *i.e.*  $\tau_n = 0.8$  for the extrapolation step.

(B1i) Identical to (B1), but the bias-reduced indirect estimator is used instead.

(B2) We assume the homoscedastic single-index model  $Y = g(\beta^\top \mathbf{X}) + \varepsilon$  with known  $\beta = (1, 1, 1, 1)$ . The function  $g$  is estimated through the Nadaraya-Watson estimator, with a bandwidth chosen using the R package `np`. The bias-reduced direct estimator is used.

(B3) Identical to (S2), although  $\beta$  is assumed to be known and equal to  $(1, 1, 1, 1)$ .

(B4) We assume that the structure of the model linking  $Y$  to  $\mathbf{X}$  is fully known, *i.e.* we know  $\beta$  and the location and scale functions, and we use the direct estimator (no bias reduction).

(B4i) Identical to (B4), although the indirect estimator is used instead (no bias reduction).

(B5) Identical to (B4), although the bias-reduced direct estimator is used instead.

(B5i) Identical to (B5), although the bias-reduced indirect estimator is used instead.

In each procedure except (B1) and (B1i), the intermediate expectile level used as an anchor in the extreme value index and extreme expectile estimators is fixed at  $\tau_n = 0.9$ , corresponding to  $k = \lfloor n(1 - \tau_n) \rfloor = 100$ ; in (S2), (S2i), (B2) and (B3), we use the Epanechnikov kernel in the estimation of the link functions  $g$  and  $\sigma$ . To assess the performance of our methods, we compute, for a given estimator  $\bar{\xi}_{\tau'_n}^*(Y|\mathbf{x})$ , the Relative Mean Absolute Deviation (RMAD)

$$\overline{\text{RMAD}} = \text{median}_{1 \leq m \leq N} \left| \frac{\bar{\xi}_{\tau'_n}^{*(m)}(Y|\mathbf{x})}{\xi_{\tau'_n}(Y|\mathbf{x})} - 1 \right|,$$

where  $\mathbf{x}^\top = (1/2, 1/2, 1/2, 1/3)$ . The quantity  $\bar{\xi}_{\tau'_n}^{\mathbf{x},(m)}(Y|\mathbf{x})$  denotes the estimator calculated on the  $m$ th replication, at the level  $\tau'_n = 1 - 5/n = 0.995$ . The error  $\overline{\text{RMAD}}$  gives an idea of the uncertainty on extreme conditional expectiles at a typical data point in the centre of the data cloud. Finally, for all  $\alpha \in (0, 1)$ , the true expectiles  $\xi_\alpha(Y|\mathbf{x})$  are deduced from  $\xi_\alpha(\varepsilon_0)$ , obtained by solving the equation  $\psi(y)/(2\psi(y) + y) = 1 - \alpha$  via the R function `uniroot`, where  $\psi(y) = \int_y^\infty \mathbb{P}(\varepsilon_0 > t) dt$  is computed with the R function `integrate`.

Results are reported in Table F.1 of [19]. In the linear model (G1), methods (S1) and (S1i) are clearly the best, and single-index based methods (S2) and (S2i) perform reasonably well. In fact, for the heaviest tail, methods (S2) and (S2i) slightly outperform (S1) and (S1i) because they are more robust to the highest values in the sample. In the single-index model (G2), methods (S2) and (S2i) perform best, and method (S2) is quite close to the unrealistic benchmark (B3); methods (S1) and (S1i) are heavily penalised by the misspecification of the conditional mean and variance. The nonparametric benchmarks (B1) and (B1i) are surprisingly competitive, perhaps because they benefit from a degree of robustness against heteroscedasticity. Not accounting for heteroscedasticity is indeed very detrimental, as a comparison of method (S2) and benchmarks (B2), (B3) shows, even with the unrealistic advantage of a correct pre-specification of the direction  $\beta$ . Finally, a comparison of benchmarks (B4) and (B5) shows that even though an unrealistic correct pre-specification of the model structure is obviously beneficial, getting the extreme value step right is very important: in the linear model (G1), method (S1) outperforms benchmark (B4) for  $\gamma \in \{0.1, 0.2\}$ , and is competitive otherwise, because it features a bias-reduction scheme at the extreme value step.

It appears that while knowing model structure is an advantage for lighter-tailed models, this advantage disappears when the noise variable has a heavier tail, thus illustrating that the extreme value step, rather than model estimation, is indeed the major contributor to estimation error. For instance, when  $\gamma = 0.2$ , the RMAD of benchmark (B5) is only 5% smaller than the RMAD of method (S2) in the single-index model (G2), and method (S2) is even slightly more accurate when  $\gamma$  is larger. The difference when  $\gamma = 0.1$  makes sense: in this setup where extreme expectiles are comparatively smaller, an error on the conditional mean or variance will have more consequences. Let us conclude that while we used the intermediate level  $k_n = 100$  for the sake of computational efficiency, in practice one may want to use a data-driven criterion for the choice of  $k_n$ . In Appendix F.1 of [19], we suggest an adaptation of an Asymptotic Mean-Squared Error (AMSE) minimisation criterion; we repeated this simulation exercise with this choice of  $k_n$  and observed that there is no obvious advantage in the data-driven choice although results are competitive. Full results are reported in Table F.2 of [19].

4.2. *Simulation study: time series models.* We simulate  $N = 1,000$  replications of time series of size  $n + 1 = 1,001$  from two different models:

- (T1) An ARMA(1, 1) model  $Y_t = \phi Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$ , where the parameters  $\phi$  and  $\theta$  are estimated using default settings of the R function `arma` from package `tseries`.
- (T2) A GARCH(1, 1) model  $Y_t = (\omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2)^{1/2} \varepsilon_t$ , where  $\omega$ ,  $\alpha$  and  $\beta$  are estimated using default settings of the R function `garch` from package `tseries`.

The  $\varepsilon_t$  are i.i.d. with common density  $f_0$  as in (5) and  $\rho = -1$ ; in the GARCH(1, 1) model, these innovations are rescaled by  $\sqrt{\Gamma(1 - 2\gamma)\Gamma(1 + 2\gamma)}$  to guarantee that  $\mathbb{E}[\varepsilon^2] = 1$ .

We estimate a one-step ahead extreme expectile  $\xi_{\tau'_n}(Y_{n+1}|\mathcal{F}_n)$ , where  $\mathcal{F}_n$  denotes the past  $\sigma$ -field at time  $n$ . We then compute, on the  $m$ th sample, the target value  $\xi_{\tau'_n}^{(m)}(Y_{n+1}|\mathcal{F}_n)$ , its

direct estimate  $\widehat{\xi}_{\tau'_n}^{\star, \text{RB}, (m)}(Y_{n+1} | \mathcal{F}_n)$  and its indirect counterpart  $\widetilde{\xi}_{\tau'_n}^{\star, \text{RB}, (m)}(Y_{n+1} | \mathcal{F}_n)$ , where  $\tau'_n = 1 - 5/n = 0.995$  and  $k_n = n(1 - \tau_n) = 100$ . We calculate their RMAD

$$\text{RMAD} = \text{median}_{1 \leq m \leq N} \left| \frac{\overline{\xi}_{\tau'_n}^{\star, (m)}(Y_{n+1} | \mathcal{F}_n)}{\xi_{\tau'_n}^{(m)}(Y_{n+1} | \mathcal{F}_n)} - 1 \right|, \quad \text{with } \overline{\xi}_{\tau'_n}^{\star, (m)} = \widehat{\xi}_{\tau'_n}^{\star, \text{RB}, (m)} \text{ or } \widetilde{\xi}_{\tau'_n}^{\star, \text{RB}, (m)}.$$

In the ARMA model, we take  $\phi, \theta \in \{0.1, 0.5\}$ ; in the GARCH model, we fix  $\omega = 0.1$  and take  $(\alpha, \beta) \in \{(0.1, 0.1), (0.1, 0.45), (0.45, 0.1), (0.1, 0.85)\}$ . In each model, we take  $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$ . Note that the GARCH model is second-order stationary only if  $\alpha + \beta < 1$  (see Theorem 2.5 in [16]). Our methods are compared with the (unrealistic) benchmarks generated from knowing model coefficients (and therefore observing the innovations).

Results are reported in Table F.3 of [19]. In the ARMA model, the RMAD does not seem overly sensitive to the parameters  $\phi$  and  $\theta$ , but increases with the extreme value index  $\gamma$ . In the GARCH model, errors seem to be sensitive to whether the model is close to second-order stationarity (note the slightly different errors in the case  $(\alpha, \beta) = (0.1, 0.85)$  and  $\gamma \in \{0.1, 0.2\}$ ). In both models, the indirect estimator has an advantage over the direct estimator, which gets smaller as the tail gets heavier. Knowing the true values of the coefficients does not bring a large improvement, except maybe for the lightest tails; this again underlines that most of the estimation error, and hence of the uncertainty on the estimates, originates from the extreme value step, rather than model estimation. With our data-driven choice of  $k_n$ , the indirect estimator typically stays the best.

*4.3. Real data analysis: Vehicle insurance data.* We consider the Vehicle Insurance Customer Data<sup>1</sup>, made of  $n = 9,134$  total (*i.e.* cumulative over the duration of the contract) claim amounts  $Y$  of insurance policyholders according to their lifetime value  $X_1$  (in USD), income  $X_2$  (in USD), number  $X_3$  of months since last claim and number  $X_4$  of months since policy inception. We follow the methodology of Section 3.2. A cross-validation procedure using the R function `npindexbw` (from the package `np`) gives a selected bandwidth  $h^* \approx 0.1$  (for covariates standardised by their respective maxima). We also choose  $t^* = \infty$ . We obtain  $\widehat{\beta} \simeq (-0.923, 0.386, -0.001, -0.002)$ , which seems to indicate that only lifetime value  $X_1$  and income  $X_2$  play a role in the prediction of  $Y$ . The estimated functions  $\widehat{g}$  and  $\widehat{\sigma}$  are depicted in the top left panel of Figure 1 (the kernel function  $L$  is the Epanechnikov kernel).

We now estimate an extreme conditional expectile  $\xi_{\tau'_n}(Y | \mathbf{x})$  at level  $\tau'_n = 1 - 1/(nh^*) \approx 0.999$ . The top right panel of Figure 1 shows the direct extreme conditional expectile estimator for  $k^* = 200$  and  $\tau^* = 1 - k^*/n$  (the bottom right panel of Figure 1 shows that the heavy-tailed assumption on the noise is reasonable). The heteroscedastic single-index model captures the variation in the shape of the data cloud fairly well, and the extreme conditional expectile curve gives a reasonable idea of the conditional extremes of the data. Interpreting an expectile curve, meanwhile, is not always straightforward. However, in this insurance example, the expectile  $\xi_{\tau'_n}(Y | \mathbf{x})$  satisfies the following gain-loss ratio criterion (see [3]):

$$\begin{aligned} 1 - \tau'_n &\approx \frac{1 - \tau'_n}{\tau'_n} = \frac{\mathbb{E}((Y - \xi_{\tau'_n}(Y | \mathbf{x})) \mathbb{1}\{Y > \xi_{\tau'_n}(Y | \mathbf{x})\} | \mathbf{X} = \mathbf{x})}{\mathbb{E}((\xi_{\tau'_n}(Y | \mathbf{x}) - Y) \mathbb{1}\{Y < \xi_{\tau'_n}(Y | \mathbf{x})\} | \mathbf{X} = \mathbf{x})} \\ &\approx \frac{\mathbb{E}((Y - \xi_{\tau'_n}(Y | \mathbf{x})) \mathbb{1}\{Y > \xi_{\tau'_n}(Y | \mathbf{x})\} | \mathbf{X} = \mathbf{x})}{\xi_{\tau'_n}(Y | \mathbf{x}) - \mathbb{E}(Y | \mathbf{X} = \mathbf{x})}. \end{aligned}$$

<sup>1</sup>Available at <https://www.kaggle.com/ranja7/vehicle-insurance-customer-data> and from the authors upon request.

In other words,  $\xi_{\tau'_n}(Y|\mathbf{x})$  is the aggregate premium to be collected over the lifetime of the contract so that, for customers having the list of characteristics  $\mathbf{x}$ , the ratio between average losses exclusively incurred by claims made by such customers above that level and net average profit is approximately the small quantity  $1 - \tau'_n$ . This value  $\xi_{\tau'_n}(Y|\mathbf{x})$  can be thus interpreted as a high safety margin for the insurer, and has an even clearer meaning to reinsurers, who only face a loss when the claim exceeds a certain high threshold.

We compare extreme conditional expectile and quantile estimates at the same level  $\tau'_n$ , the latter being obtained by combining the standard Weissman-type estimate of an extreme quantile of the noise with our estimates  $\hat{g}$  and  $\hat{\sigma}$ . It can be seen in Figure 1 that the extreme conditional quantile estimate is outside a pointwise 95% bootstrap confidence interval for the extreme conditional expectile (constructed using an adapted methodology called *semi-parametric Pareto tail bootstrap*, see Appendix F.2 of [19]). This may be relevant to insurance companies, for whom lower (*i.e.* more optimistic) assessments of risk translate into marketable contracts with lower premiums and hence improved competitiveness, while policy-makers and regulators would favour the higher (*i.e.* more pessimistic) quantile estimates to hedge better against systemic risk. Interestingly, the regression median is below the regression mean, so there is a qualitative difference between central and extreme assessments of risk using expectiles and quantiles: a risk assessment based on the regression mean (*i.e.* a central conditional expectile) is more conservative than if it were based on the regression median (*i.e.* a central conditional quantile), but extreme conditional expectile risk measurements are less conservative than those made with extreme conditional quantiles.

**4.4. Real data analysis: Australian dollar exchange rates.** The analysis of exchange rate risk is a key question in economics. An accurate analysis of exchange rate risk informs strategic decisions made by firms, such as the extent to which they import and export and whether they should invest in foreign markets, which have consequences on their competitiveness on the global marketplace. We study the daily log-returns of the Australian Dollar/Swiss Franc (AUD/CHF) and Australian Dollar/Swedish Krona (AUD/SEK) exchange rates from 1st March 2015 to 28th February 2019, represented in the left panels of Figure 2 (sample size  $n = 1,043$ ). The literature has suggested that expectiles can be fruitfully used to estimate quantiles (see *e.g.* [3, 44]). Our goal is to estimate the (dynamic) extreme conditional quantile  $q_{\tau'_n}(Y_{n+1} | \mathcal{F}_n)$  of level  $\tau'_n = 0.995 \approx 1 - 5/n$  on the final day. We consider a GARCH(1, 1) model, motivated by the finding of [36] that GARCH models fit past Australian exchange rates well; the R function `garch` (in the package `tseries`) returns, with the notation of Section 3.4.2,  $(\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n) = (4.20 \times 10^{-7}, 0.943, 0.0465)$  for AUD/CHF and  $(1.21 \times 10^{-5}, 0.576, 0.119)$  for AUD/SEK. We construct the quantile estimator

$$\hat{q}_{\tau'_n}^{*,\text{RB}}(\varepsilon) = ((\hat{\gamma}_k^{\text{RB}})^{-1} - 1) \hat{\gamma}_k^{\text{RB}} \hat{\xi}_{\tau'_n}^{*,\text{RB}}(\varepsilon).$$

With  $k^* = 50$  and  $\tau_n = 1 - k^*/(n - 1)$ , we get  $\hat{\gamma}_k^{\text{RB}} = 0.189$  for AUD/CHF (resp. 0.211 for AUD/SEK) and  $\hat{q}_{\tau'_n}^{*,\text{RB}}(\varepsilon) = 2.40$  (resp. 2.58) (graphical evidence of a heavy right tail of  $\varepsilon$  is given on the right panels of Figure 2). To check that our estimates make sense, we recall the characterisation of  $q_{\tau'_n}(\varepsilon)$  as  $0.995 = \tau'_n = \mathbb{E}(\mathbb{1}\{\varepsilon \leq q_{\tau'_n}(\varepsilon)\})$  and compare that with

$$\frac{1}{n-1} \sum_{i=2}^n \mathbb{1}\left\{\hat{\varepsilon}_i^{(n)} < \hat{q}_{\tau'_n}^{*,\text{RB}}(\varepsilon)\right\} \approx 0.99424 \text{ for AUD/CHF (resp. } 0.99520 \text{ for AUD/SEK)}.$$

This is indeed very close to the expected value  $\tau'_n = 0.995$ . Our estimate can be compared with a bias-reduced version  $\hat{q}_{\tau'_n}^{*,\text{RB}}(\varepsilon)$  of the classical extrapolated estimate of [51]:

$$\hat{q}_{\tau'_n}^{*,\text{RB}}(\varepsilon) = \hat{q}_{\tau'_n}^*(\varepsilon) \left(1 - \hat{\gamma}_k^{\text{RB}} \frac{\hat{b}}{\hat{\rho}} \left(\frac{n}{k}\right)^{\hat{\rho}}\right),$$

where  $\hat{q}_{\tau_n}^*(\varepsilon)$  is the residual-based Weissman quantile estimator using  $\hat{\gamma}_k^{\text{RB}}$  in its extrapolation step. This estimate is 2.48 for AUD/CHF (resp. 2.64 for AUD/SEK). Our expectile-based estimate of 2.40 (resp. 2.58) is slightly lower; this makes sense, as the estimated value of  $\gamma$  is lower than  $1/4$ , and extreme expectile-based estimates can be thought to reflect this rather light tail by producing lower point estimates than their quantile counterparts (and when  $\gamma > 1/4$ , expectile-based quantile estimates seem to be higher than traditional estimates, see *e.g.* Section 7.1 in [9]). This lower assessment of risk may be interesting to financial companies, as opposed to regulators who may prefer quantile-based estimates. Finally, the predicted estimate of  $q_{\tau_n}(Y_{n+1} | \mathcal{F}_n)$  on 1st March 2019 is 0.0138 with Gaussian and semi-parametric Pareto tail bootstrap 95% confidence intervals (see Appendix F.2 of [19]) being  $[0.0122, 0.0154]$  and  $[0.0116, 0.0160]$  for AUD/CHF (resp. 0.0156, Gaussian confidence interval  $[0.0136, 0.0178]$  and bootstrap confidence interval  $[0.0127, 0.0191]$  for AUD/SEK). This amounts to a daily variation of 1.4% of the AUD/CHF exchange rate (resp. 1.6% for AUD/SEK).

**5. Discussion and perspectives.** We provide a general toolbox for the estimation of extreme conditional expectiles, by showing how a simple assumption on the residuals of the model makes it possible to obtain the convergence of residual-based estimators of the extremes of the noise. By applying our results in examples not limited to low dimensions, we contribute to the broader question of how to model extremes with a large number of covariates. The works of [17, 23, 49, 50] introduce dedicated modelling assumptions on the tail conditional quantiles of  $Y$ . The tail linear quantile regression model of [50] is not straightforward to interpret: even when the conditional quantile is in fact linear in  $\boldsymbol{x}$  (for any  $\tau$ ), this model is the arguably complicated linear model linking  $Y$  to  $\boldsymbol{X}$  with random coefficients (see p.808 of [6]). Our generic model provides a straightforward way of seeing the effect  $\boldsymbol{X}$  has on  $Y$  and avoids the crossing problem (unlike the method of [50]), since the structure of the model is estimated only once. The nonparametric model of [17], meanwhile, rests upon the estimation of a Tail Dimension Reduction subspace, which can only be done using the pairs  $(\boldsymbol{X}_i, Y_i)$  such that  $Y_i$  is large. This entails a potentially substantial loss of modelling strength compared to our approach. Besides, the aforementioned papers focus on the case of i.i.d. data  $(\boldsymbol{X}_i, Y_i)$ ; our method allows us to consider popular time series examples.

Among future research perspectives, it would be nice to extend our results for ARMA and GARCH models in the ARMA-GARCH model, to allow for heteroscedasticity in time series not having mean 0. Besides, the basic principle of our approach relies on location equivariance and positive homogeneity, which are true for numerous interesting functionals, *e.g.* coherent spectral risk measures, including the very recent concept of extremiles ([8]). Adapting our approach to other risk measures constitutes an interesting avenue for further work. Another perspective is to relax the heavy-tailed assumption, to extend the applicability of our method. As far as we know, even in the simple unconditional i.i.d. case, there are currently no estimation procedures available for extreme expectiles of either light-tailed or short-tailed distributions, which are the other setups one would consider in an extreme value framework. Finally, an approach that fully accounts for joint uncertainty between model estimation and extreme value estimation would be an important next step in order to handle the strongest possible forms of heteroscedasticity. This will at least require uniform weighted Gaussian approximations of the tail empirical residual-based quantile process; this very difficult question needs to be solved on a case-by-case basis, because the structure of residuals is completely controlled by the structure of the model. In linear regression, the current state of the art seems to be uniform non-weighted approximations on the real line (see [5], especially Section 6 therein). The absence of weighting makes it impossible to use such results for extreme value inference. We are not aware of such results in single-index models, not even

non-weighted and in the homoscedastic case. This is a very substantial research project in itself which we defer to future work.

**Acknowledgements.** The authors acknowledge an anonymous Associate Editor and three anonymous reviewers for their very helpful comments that led to a greatly improved version of this paper. This research was supported by the French National Research Agency under the grants ANR-15-IDEX-02 and ANR-19-CE40-0013. S. Girard gratefully acknowledges the support of the Chair Stress Test, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas. G. Stupfler also acknowledges support from an AXA Research Fund Award on “Mitigating risk in the wake of the COVID-19 pandemic”.

## REFERENCES

- [1] ABDOUS, B. and REMILLARD, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics* **47** 371–384.
- [2] AHMAD, A., DEME, E., DIOP, A., GIRARD, S. and USSEGLIO-CARLEVE, A. (2020). Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model. *Electronic Journal of Statistics* **14** 4421–4456.
- [3] BELLINI, F. and DI BERNARDINO, E. (2017). Risk management with expectiles. *The European Journal of Finance* **23** 487–506.
- [4] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods (second edition)*. Springer.
- [5] BÜCHER, A., SEGERS, J. and VOLGUSHEV, S. (2014). When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs. *Annals of Statistics* **42** 1598–1634.
- [6] CHERNOZHUKOV, V. (2005). Extremal quantile regression. *Annals of Statistics* **33** 806–839.
- [7] DAOUIA, A., GARDES, L., GIRARD, S. and LEKINA, A. (2011). Kernel estimators of extreme level curves. *TEST* **20** 311–333.
- [8] DAOUIA, A., GIJBELS, I. and STUPFLER, G. (2019). Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association* **114** 1366–1381.
- [9] 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** 263–292.
- [10] DAOUIA, A., GIRARD, S. and STUPFLER, G. (2019). Extreme M-quantiles as risk measures: From  $L^1$  to  $L^p$  optimization. *Bernoulli* **25** 264–309.
- [11] DAOUIA, A., GIRARD, S. and STUPFLER, G. (2020). Tail expectile process and risk assessment. *Bernoulli* **26** 531–556.
- [12] DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B* **52** 393–442.
- [13] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer.
- [14] DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17** 1833–1855.
- [15] DREES, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9** 617–657.
- [16] FRANCO, C. and ZAKOÏAN, J. M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons.
- [17] GARDES, L. (2018). Tail dimension reduction for extreme quantile estimation. *Extremes* **21** 57–95.
- [18] GARDES, L. and STUPFLER, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator. *Extremes* **17** 45–75.
- [19] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2021). Supplement to "Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models". <https://doi.org/10.1007/s00180-021-01000-0>.
- [20] GIRARD, S., STUPFLER, G. and USSEGLIO-CARLEVE, A. (2021). On automatic bias reduction for extreme expectile estimation. <https://hal.archives-ouvertes.fr/hal-03086048>.
- [21] GOMES, M. I., BRILHANTE, M. F. and PESTANA, D. (2016). New reduced-bias estimators of a positive extreme value index. *Communications in Statistics - Simulation and Computation* **45** 833–862.
- [22] HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* **84** 986–995.
- [23] HE, F., CHENG, Y. and TONG, T. (2016). Estimation of high conditional quantiles using the Hill estimator of the tail index. *Journal of Statistical Planning and Inference* **176** 64–77.

- [24] HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3** 1163–1174.
- [25] HILL, J. B. (2015). Tail index estimation for a filtered dependent time series. *Statistica Sinica* **25** 609–629.
- [26] HOGA, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics* **37** 613–624.
- [27] HOLZMANN, H. and KLAR, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics* **10** 2355–2371.
- [28] HOROWITZ, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Springer.
- [29] HORVÁTH, L. and LIESE, F. (2004).  $L_p$ -estimators in ARCH models. *Journal of Statistical Planning and Inference* **119** 277–309.
- [30] KRÄTSCHMER, V. and ZÄHLE, H. (2017). Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics* **44** 425–454.
- [31] 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** 261–270.
- [32] KUCHIBHOTLA, A. K. and PATRA, R. K. (2020). Efficient estimation in single index models through smoothing splines. *Bernoulli* **26** 1587–1618.
- [33] KYUNG-JOON, C. and SCHUCANY, W. R. (1998). Nonparametric kernel regression estimation near endpoints. *Journal of Statistical Planning and Inference* **66** 289–304.
- [34] LEWBEL, A. and LINTON, O. (2002). Nonparametric censored and truncated regression. *Econometrica* **70** 765–779.
- [35] MARTINS-FILHO, C., YAO, F. and TORERO, M. (2018). Nonparametric estimation of Conditional Value-at-Risk and Expected Shortfall based on extreme value theory. *Econometric Theory* **34** 23–67.
- [36] MCKENZIE, M. D. (1997). ARCH modelling of Australian bilateral exchange rate data. *Applied Financial Economics* **7** 147–164.
- [37] MCNEIL, A. J. and FREY, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* **7** 271–300.
- [38] NEWBY, W. K. and POWELL, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847.
- [39] NEWBY, W. K. and POWELL, J. L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* **6** 295–317.
- [40] POWELL, J. L. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* **54** 1435–1460.
- [41] ROHRBECK, C., EASTOE, E. F., FRIGESSI, A. and TAWN, J. A. (2018). Extreme value modelling of water-related insurance claims. *Annals of Applied Statistics* **12** 246–282.
- [42] SEGERS, J. (2001). Residual estimators. *Journal of Statistical Planning and Inference* **98** 15–27.
- [43] STUTE, W. and WANG, J. L. (2008). The central limit theorem under random truncation. *Bernoulli* **14** 604–622.
- [44] TAYLOR, J. W. (2008). Estimating Value at Risk and Expected Shortfall using expectiles. *Journal of Financial Econometrics* **6** 231–252.
- [45] TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26** 24–36.
- [46] TSAY, R. S. (1984). Regression models with time series errors. *Journal of the American Statistical Association* **79** 118–124.
- [47] VAN KEILEGOM, I. and WANG, L. (2010). Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data. *Electronic Journal of Statistics* **4** 133–160.
- [48] VELTHOEN, J., CAI, J. J., JONGBLOED, G. and SCHMEITS, M. (2019). Improving precipitation forecasts using extreme quantile regression. *Extremes* **22** 599–622.
- [49] WANG, H. J. and LI, D. (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of American Statistical Association* **108** 1062–1074.
- [50] WANG, H. J., LI, D. and HE, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of American Statistical Association* **107** 1453–1464.
- [51] WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the  $k$  largest observations. *Journal of the American Statistical Association* **73** 812–815.
- [52] WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *Annals of Statistics* **13** 163–177.
- [53] ZHU, K. and LING, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *Annals of Statistics* **39** 2131–2163.
- [54] ZHU, L., DONG, Y. and LI, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling. *Statistica Sinica* **23** 1235–1255.
- [55] ZIEGEL, J. F. (2016). Coherence and elicibility. *Mathematical Finance* **26** 901–918.



## SUPPLEMENTARY MATERIAL

The supplementary material document () contains the proofs of all theoretical results. It also provides further theoretical results related to indirect estimators, and further details about our finite-sample procedures and studies.

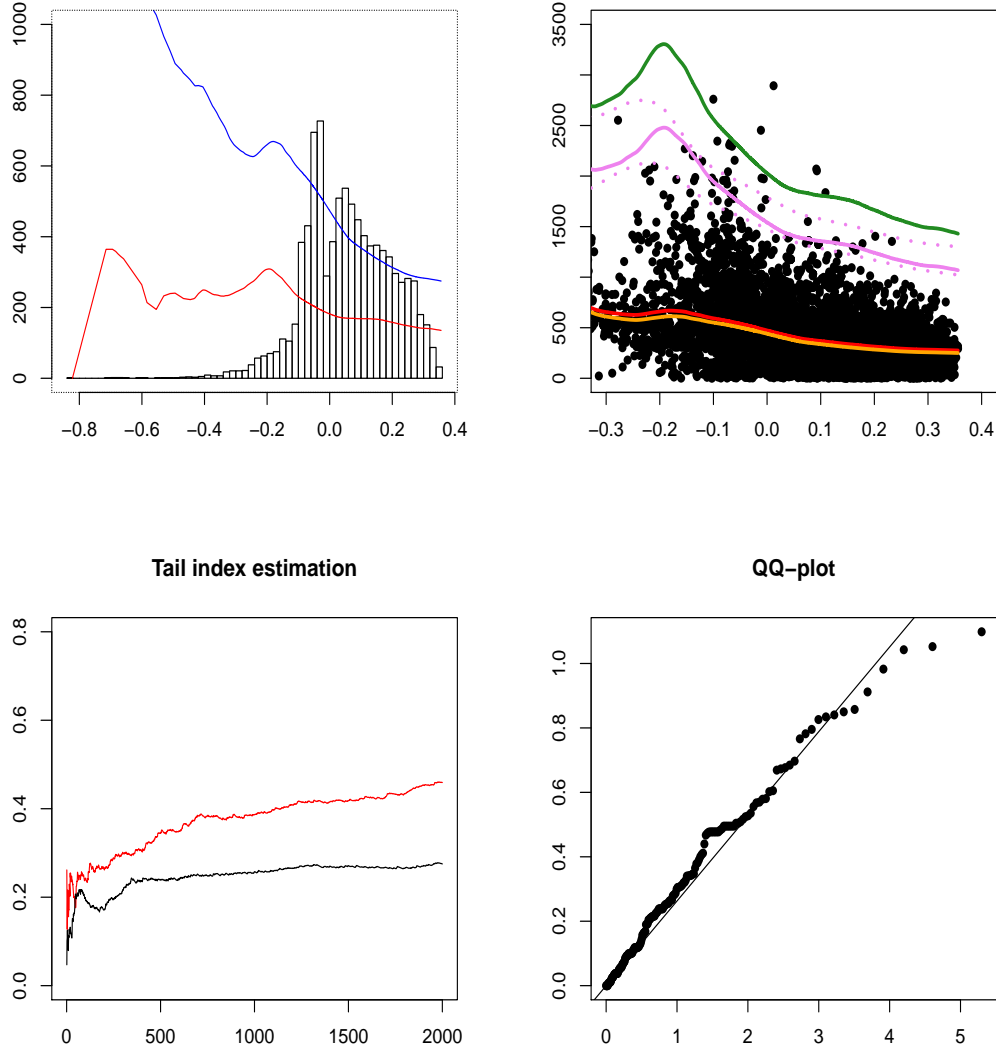


FIG 1. *Vehicle Insurance Customer data. Top left: estimates of  $g$  (red curve) and  $\sigma$  (blue curve) with a histogram of the  $\hat{\beta}^\top \mathbf{X}_i$ . Top right: estimates of the regression mean (red line) and median (orange line) and of the estimated conditional expectile (solid purple line; dotted lines represent bootstrap pointwise 95% confidence intervals) and quantile (green line) at level  $\tau_n = 1 - 1/(nh^*) \approx 0.999$  in the  $(\hat{\beta}^\top \mathbf{x}, y)$  plane. Bottom left: curves  $k \mapsto \hat{\gamma}_k^{\text{RB}}$  on the non-filtered data  $Y_i$  (black curve) and residuals (red curve). Bottom right: Exponential QQ-plot of the log-spacings  $\log(\hat{\varepsilon}_{n-i+1,n}^{(n)}/\hat{\varepsilon}_{n-k^*,n}^{(n)})$ ,  $1 \leq i \leq k^* = 200$ . The straight line has slope  $\hat{\gamma}_{k^*}^{\text{RB}} = 0.263$ .*

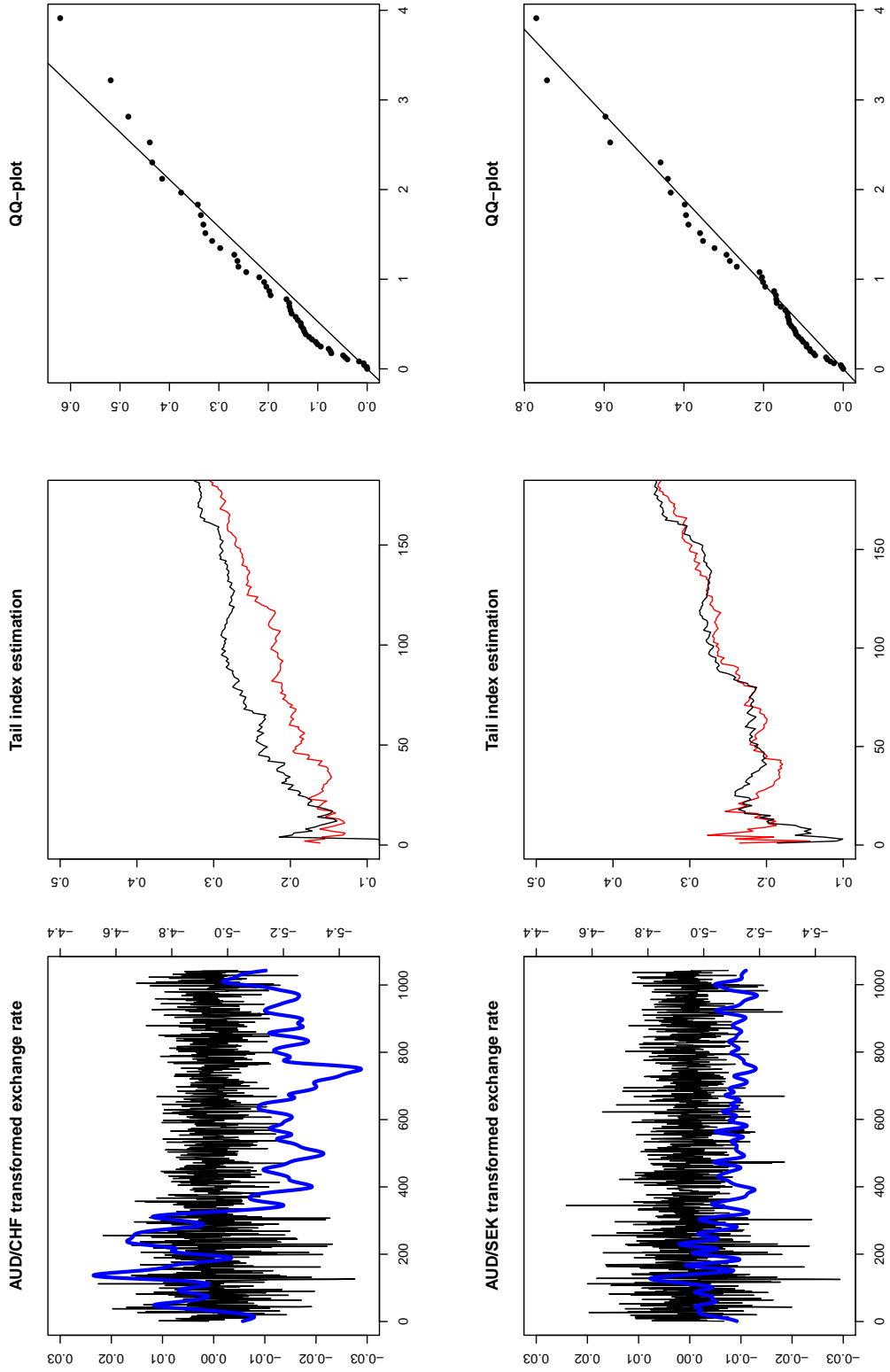


FIG 2. Australian Dollar exchange rate data. Left panels: daily log-returns (black line, and left-hand y-axis) and estimated daily volatility (blue bold line, smoothed using the R `smooth.spline` function with parameter 0.35, and right-hand y-axis) from 1st March 2015 to 28th February 2019. Middle panels: Bias-reduced Hill curves  $k \mapsto \hat{\gamma}_k^{\text{RB}}$  on the non-filtered data  $Y_i$  (black curve) and residuals (red curve). Right panels: Exponential QQ-plot of the log-spacings  $\log(\hat{\varepsilon}_{n-i+1,n}^{(n)} / \hat{\varepsilon}_{n-k^*,n}^{(n)})$ ,  $1 \leq i \leq k^* = 50$ . The straight line has slope  $\hat{\gamma}_{k^*}^{\text{RB}}$ . Top panels: AUD/CHF data, bottom panels: AUD/SEK data.