



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. 2020. hal-03306230v1

HAL Id: hal-03306230

<https://inria.hal.science/hal-03306230v1>

Preprint submitted on 3 Apr 2020 (v1), last revised 16 Jan 2024 (v7)

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

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

^a Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

^b Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

Abstract: 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 10 years. Unlike quantiles, expectiles induce coherent risk measures and are calculated using tail expectations rather than merely tail probabilities; contrary to the popular quantile-based Expected Shortfall, they define elicitable risk measures. The behaviour and estimation of extreme expectiles using independent and identically distributed heavy-tailed observations has been investigated in a recent series of papers. The case of extreme conditional expectile estimation has, however, not been addressed so far in the literature. We build here a general theory for the estimation of extreme conditional expectiles in heteroscedastic regression models with heavy-tailed noise; an important feature of our approach is that it 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 heteroscedastic models, heteroscedastic single-index models and autoregressive time series models. Our estimators are showcased on a numerical simulation study, as well as on real sets of actuarial and financial data.

MSC2010 Classifications: Primary 62G32, Secondary 62G08, 62G20, 62G30.

Keywords: Expectiles, extreme value analysis, heavy-tailed distribution, heteroscedasticity, regression models, residual-based estimators, single-index model.

1 Introduction

1.1 Motivation

The question of how to model extreme events arises in many fields of statistical applications, such as finance and insurance. A traditional way of considering extreme events is to estimate extreme quantiles of a well-chosen random variable Y , such as

the daily log-return of a stock market index in finance, or the magnitude of a claim in insurance. There are many practical applications where $Y \in \mathbb{R}$ is recorded along with valuable information represented by a covariate $\mathbf{X} \in \mathbb{R}^d$, and a better understanding of the extremes of Y can be achieved by inferring the conditional extremes of Y given \mathbf{X} . Recent examples include the analysis of high healthcare costs given patient age and indicators for chronic diseases (Wang and Li, 2013), a general strategy for the development of stratified models for large operational losses (Chavez-Demoulin *et al.*, 2016) and modelling large insurance claims conditional on climate variables (Rohrbeck *et al.*, 2018). We focus on the case when Y given \mathbf{X} is heavy-tailed; the heavy-tailed assumption underpins the aforementioned papers and is generally appropriate to the modelling of actuarial and financial data (see *e.g.* p.1 of Resnick, 2007).

Under no further assumptions on the structure of the pair (\mathbf{X}, Y) , general nonparametric smoothing methods can be used to tackle this problem, such as those of Daouia *et al.* (2011), Gardes and Stupfler (2014), and Goegebeur *et al.* (2015). Since, apart from the conditional heavy-tailed hypothesis, nothing is assumed on (\mathbf{X}, Y) , those techniques suffer from the curse of dimensionality, which is compounded in the context of 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 Davison and Smith (1990) and later Chavez-Demoulin and Davison (2005), were built on parametric models. Later attempts have mostly been based on linear quantile regression: a seminal paper is Chernozhukov (2005), later continued for conditional heavy-tailed distributions by Wang *et al.* (2012), Wang and Li (2013) and He *et al.* (2016). These methods can all be put in the wider framework of quantile regression approaches, pioneered by Koenker and Bassett (1978). An approach based on the notion of a Tail Dimension Reduction subspace was adopted by Gardes (2018).

Yet, all 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 from the point of view of risk assessment, where the focus is not only what constitutes an extreme level of loss but also what the value of a typical extreme loss will be. One way of tackling this problem is to work with expectiles. The τ th regression expectile of Y given $\mathbf{X} = \mathbf{x}$ is obtained from the formulation of the τ th regression quantile of Koenker and Bassett (1978) 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 (Newey and Powell, 1987) and $\mathbb{1}\{\cdot\}$ is the indicator function. Expectiles are well-defined and unique as soon as the underlying distribution has a finite first moment (see Abdous and Remillard, 1995 and Theorem 1 in Newey and Powell, 1987). Unlike quantiles, expectiles are determined by tail expectations rather than tail probabilities, and therefore depend on both the probability of tail values and their actual realisations (see Kuan *et al.*,

2009). In addition, expectiles induce the only coherent law-invariant risk measure that is elicitable (in the sense of Artzner *et al.*, 1999; see Bellini *et al.*, 2014 and Ziegel, 2016) and therefore benefit from the existence of a natural backtesting methodology. For these reasons, expectiles are a sensible risk management tool to use, whether as a complement or an alternative to quantiles.

The literature on expectiles has so far essentially focused on estimating expectiles with a fixed, non-extreme level τ (see *e.g.* Holzmann and Klar, 2016 and Krättschmer and Zähle, 2017). By contrast, the estimation of extreme expectiles, for which $\tau = \tau_n \rightarrow 1$ as the sample size n tends to infinity, is still largely unexplored. Asymptotic expansions of extreme population expectiles were derived in various models by, among others, Bellini *et al.* (2014) and Bellini and Di Bernardino (2017). At the sample level, Daouia *et al.* (2018, 2020) initiated the estimation of extreme expectiles, in the unconditional heavy-tailed case. They did not, however, tackle the question of estimating extreme conditional expectiles. Providing a statistical solution to this question, with the additional objective of handling models that may cope with a large but fixed dimension of the covariate vector \mathbf{X} , is precisely the focus of this paper. Our idea, which we summarise in the next section, is to work in regression models explicitly linking Y to \mathbf{X} and exploit their structure as a way to carry out an analysis of the conditional extremes of the data at hand, based on extreme expectiles.

1.2 Expectiles and regression models

For the sake of simplicity, we describe first our general idea using the popular example of 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 underlying distributions of \mathbf{X} and the noise variable ε . We also suppose that \mathbf{X} is observed, independent of ε , and has a support K such that $1 + \boldsymbol{\theta}^\top \mathbf{x} > 0$ for $\mathbf{x} \in K$. In this model, by location equivariance and positive homogeneity of expectiles (Theorem 1(iii) in Newey and Powell, 1987), we may write $\xi_\tau(Y|\mathbf{x}) = \alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_\tau(\varepsilon)$. This suggests to estimate the extreme conditional expectile $\xi_{\tau_n}(Y|\mathbf{x})$, where $\tau = \tau_n \rightarrow 1$ as $n \rightarrow \infty$, by combining estimators of α , $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\xi_{\tau_n}(\varepsilon)$.

A natural idea for that is to, first, construct estimators $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ of the model parameters using a (possibly iterative) weighted least squares method. We can then construct residuals $\hat{\varepsilon}_i^{(n)} = [Y_i - (\hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i)]/[1 + \hat{\boldsymbol{\theta}}^\top \mathbf{X}_i]$. These can be used to estimate extreme expectiles of ε based on the $\hat{\varepsilon}_i^{(n)}$ rather than on the unobservable ε_i , thus producing an estimator $\hat{\xi}_{\tau_n}(\varepsilon)$. This expectile estimator can be adapted from, for instance, an empirical asymmetric least squares method (see Daouia *et al.*, 2018, 2020). If ε has a finite second moment, then the weighted least squares method produces \sqrt{n} -consistent estimators, making the residuals of the model “close” to the unobservable errors, and it is therefore reasonable to expect that the asymptotic normality properties of the estimators of Daouia *et al.* (2018, 2020) carry over to their versions based on the residuals

of the model. An estimator of the extreme conditional expectile $\xi_{\tau_n}(Y|\mathbf{x})$ is then readily obtained as $\widehat{\xi}_{\tau_n}(Y|\mathbf{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{x})\widehat{\xi}_{\tau_n}(\varepsilon)$.

Here we generalise this construction in heteroscedastic regression models of the form $Y = g(\mathbf{X}) + s(\mathbf{X})\varepsilon$ where g and $s > 0$ are two measurable functions of \mathbf{X} . If ε is centred and has unit variance, this model can essentially be rewritten $\mathbb{E}(Y|\mathbf{X}) = g(\mathbf{X})$ and $\text{Var}(Y|\mathbf{X}) = s^2(\mathbf{X})$, and is for instance called *location-dispersion regression model* in Van Keilegom and Wang (2010). Even though our theory will be valid for arbitrary regression models of this form, one should read this paper having in mind models that are adapted to the consideration of a large dimension d , as in such models the estimation of g and s will not suffer from the curse of dimensionality (and thus reasonable rates of convergence can be achieved). This class of models includes the location-scale shift linear regression model, as well as, among others, single-index regression models, partially linear regression models and additive models, as well as popular time series models such as $\text{AR}(p)$ and $\text{ARCH}(p)$ when \mathbf{X} contains lagged values of the variable of interest Y .

In this general model, $\xi_{\tau_n}(Y|\mathbf{x}) = g(\mathbf{x}) + s(\mathbf{x})\xi_{\tau_n}(\varepsilon)$. The viewpoint we deliberately adopt is that the estimation of g and s 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 semi-parametric techniques. In Section 2, we will thus focus on developing a toolbox for the estimation of $\xi_{\tau_n}(\varepsilon)$ given that estimators of g and s , and therefore residuals of the model, are available. In Section 3, we shall then consider four specific examples for which we provide the full asymptotic analysis of the estimation of the extreme conditional expectile $\xi_{\tau_n}(Y|\mathbf{x})$: the location-scale shift linear regression model in Section 3.1, a heteroscedastic single-index model in Section 3.2, and two autoregressive time series models in Section 3.3. Section 4 examines the behaviour of our estimators on simulated and real data, and Section 5 discusses our findings. Proofs and additional numerical results are deferred to the Supplementary Material document.

2 General theoretical toolbox for extreme expectile estimation in heavy-tailed regression models

Our general theoretical framework is the following. Let (\mathbf{X}_i, Y_i) , $1 \leq i \leq n$ be part of a stationary sequence of copies of the random pair (\mathbf{X}, Y) , and suppose that

$$Y = g(\mathbf{X}) + s(\mathbf{X})\varepsilon \tag{1}$$

where g and $s > 0$ are two measurable functions of \mathbf{X} . The unobserved noise variable ε is assumed to be centred and independent of the observed covariate \mathbf{X} , and in addition, we assume throughout that the $\varepsilon_i = (Y_i - g(\mathbf{X}_i))/s(\mathbf{X}_i)$ are independent.

We assume throughout Section 2 that g and s have been estimated, and we concentrate on the challenging question of estimating the extreme expectile $\xi_{\tau_n}(\varepsilon)$, with the ultimate

goal of combining all these estimators to construct an estimator of $\xi_{\tau_n}(Y|\mathbf{x})$. We work in the context of heavy-tailed distributions, also called Pareto-type distributions, which are found to model the tail structure of many financial and actuarial data examples quite well. More precisely, denoting by $\tau \mapsto q_\tau(\varepsilon)$ the quantile function of ε , we assume that the following first-order extreme value condition is satisfied:

$\mathcal{C}_1(\gamma)$ The tail quantile function of ε , defined by $U(t) = q_{1-t^{-1}}(\varepsilon)$ for $t > 1$, is regularly varying in a neighbourhood of $+\infty$ with index $\gamma > 0$, that is,

$$\lim_{t \rightarrow +\infty} \frac{U(tx)}{U(t)} = x^\gamma \quad \text{for all } x > 0.$$

Condition $\mathcal{C}_1(\gamma)$ is equivalent to assuming that the survival function of ε , denoted hereafter by $\overline{F} : x \mapsto \mathbb{P}(\varepsilon > x)$, is regularly varying in a neighbourhood of $+\infty$ with index $-1/\gamma < 0$. The tail index γ is the key driver of the tail heaviness of \overline{F} : the tail of \overline{F} gets heavier as γ increases. Together with condition $\mathbb{E}|\varepsilon_-| < \infty$, where $\varepsilon_- = \min(\varepsilon, 0)$, the assumption $\gamma < 1$ then ensures that the first moment of ε exists, which entails that expectiles of ε of any order are well-defined. Both of these conditions shall therefore be part of our minimal assumptions throughout this paper.

An additional complication in our setup is that the ε_i are unobserved, and thus $\xi_{\tau_n}(\varepsilon)$ cannot be estimated using standard techniques. However, because g and s have been estimated, by \overline{g} and \overline{s} say, we have access to residuals $\widehat{\varepsilon}_i^{(n)} = (Y_i - \overline{g}(\mathbf{X}_i))/\overline{s}(\mathbf{X}_i)$ constructed from the regression model (1). Our idea throughout this section will be to construct estimators of extreme expectiles based on the observable $\widehat{\varepsilon}_i^{(n)}$, and study their theoretical properties under the assumption that they are in some sense “close” to the true unobserved ε_i .

We shall start by the case of a so-called *intermediate level* τ_n , meaning that $\tau_n \rightarrow 1$ and $n(1 - \tau_n) \rightarrow \infty$. Our intermediate expectile estimators $\xi_{\tau_n}(\varepsilon)$ will then be extrapolated to properly extreme levels τ'_n , satisfying $n(1 - \tau'_n) \rightarrow c$, where c is a finite nonnegative constant, using a semiparametric approach warranted by the heavy-tailed assumption. Section 2.1 below focuses on a residual-based Least Asymmetrically Weighted Squares (LAWS) estimator. Section 2.2 then introduces a competitor to this estimator, based on the connection between (theoretical) extreme expectiles and quantiles. Section 2.3 extrapolates both of these estimators far into the right tail using a Weissman-type construction (Weissman, 1978), and combines these extrapolated devices with the estimators of g and s 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 here that τ_n is an intermediate sequence, *i.e.* $\tau_n \rightarrow 1$ and $n(1 - \tau_n) \rightarrow \infty$. If the true errors ε_i were available, we could estimate $\xi_{\tau_n}(\varepsilon)$ using the LAWS estimator

$$\arg \min_{u \in \mathbb{R}} \sum_{i=1}^n \eta_{\tau_n}(\varepsilon_i - u).$$

This estimator is analysed in Theorem 2 of Daouia *et al.* (2018). We replace here the unobserved ε_i by the observed residuals $\widehat{\varepsilon}_i^{(n)}$, resulting in the estimator

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

The next theorem is a flexible asymptotic result showing that the estimator $\widehat{\xi}_{\varepsilon}(\tau_n)$ is a $\sqrt{n(1-\tau_n)}$ -relatively asymptotically normal estimator of the high, intermediate expectile $\xi_{\varepsilon}(\tau_n)$ provided the gap between the series of residuals $\widehat{\varepsilon}_i^{(n)}$ and their unobservable counterparts ε_i is not too large.

Theorem 1. *Assume that there is $\delta > 0$ such that $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$, that ε 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 $\widehat{\varepsilon}_i^{(n)}$, $1 \leq i \leq n$, satisfies*

$$\forall i \in \{1, \dots, n\}, |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \leq R_{n,i}(1 + |\varepsilon_i|) \quad \text{where} \quad \sqrt{n(1-\tau_n)} \max_{1 \leq i \leq n} R_{n,i} \xrightarrow{\mathbb{P}} 0.$$

Then we have

$$\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1-2\gamma} \right).$$

Theorem 1 can be seen as a generalisation of Theorem 2 in Daouia *et al.* (2018) to the case when the variables ε_i are unobserved. Note that the former is stated under the exact same moment conditions as the latter. The difference lies in the fact that the estimator $\widehat{\xi}_{\tau_n}(\varepsilon)$ is much more difficult to handle directly, which is the reason for the introduction of the extra condition on the distance between the ε_i and the $\widehat{\varepsilon}_i^{(n)}$: this allows for a control of the gap between the estimator $\widehat{\xi}_{\tau_n}(\varepsilon)$ and its counterpart calculated on the unobserved ε_i . Loosely speaking, this condition will be satisfied when the structure of the regression model $Y = g(\mathbf{X}) + s(\mathbf{X})\varepsilon$ is estimated at a faster rate than the standard $\sqrt{n(1-\tau_n)}$ -rate of convergence of intermediate expectile estimators. We also highlight that although one should read this result thinking that the $\widehat{\varepsilon}_i^{(n)}$ are residuals of a regression model, this is not in fact necessary for the result to hold, as the proof does not use any assumption on how the $\widehat{\varepsilon}_i^{(n)}$ are constructed.

The construction of this estimator is purely nonparametric. In particular, it does not use the assumption of a heavy-tailed noise variable. We concentrate now on building an alternative, semiparametric estimator based on an asymptotic connection between extreme expectiles and extreme quantiles, which is warranted by the heavy-tailed assumption.

2.2 Intermediate step, indirect construction

We start by recalling, as shown in Proposition 2.3 of Bellini and Di Bernardino (2017), that the heavy-tailed condition on $t \mapsto U(t) = q_{1-t}(\varepsilon)$ entails $\xi_{\tau}(\varepsilon)/q_{\tau}(\varepsilon) \rightarrow (\gamma^{-1} -$

$1)^{-\gamma}$ as $\tau \uparrow 1$. Therefore, if $\bar{\gamma}$ is a consistent estimator of γ , 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

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

This is essentially the construction of Section 3.1.1 in Daouia *et al.* (2018). An extension of Theorem 1 therein (see Proposition 1 in Appendix A.1) shows that under the following classical second-order refinement of condition $\mathcal{C}_1(\gamma)$, which quantifies the approximation error in the above relationship between high quantiles and expectiles, the asymptotic distribution of the estimator $\tilde{\xi}_{\tau_n}(\varepsilon)$ is determined under high-level conditions on the pair $(\bar{\gamma}, \bar{q}_{\tau_n}(\varepsilon))$.

$\mathcal{C}_2(\gamma, \rho, A)$ 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 $\rho = 0$.

Our purpose is to show how one may construct and study such estimators $\bar{\gamma}$ and $\bar{q}_{\tau_n}(\varepsilon)$, based on residuals. Denote by $z_{1,n} \leq z_{2,n} \leq \dots \leq z_{n,n}$ the ordered n -tuple associated with an n -tuple (z_1, z_2, \dots, z_n) . A number of existing estimators of γ 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 (Hill, 1975):

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

We may also adapt in the same way the moment-type statistics which intervene in the construction of the moment estimator of Dekkers *et al.* (1989), and the general class of estimators studied by Segers (2001). These estimators all depend on the choice of an effective sample size $k = k(n)$, which is a sequence of integers such that $k \rightarrow \infty$ and $k/n \rightarrow 0$; it is useful, in our context, 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 $s \mapsto \hat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}$ 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 a couple of results on the tail empirical quantile process for independent and identically distributed random variables. It is known that under condition $\mathcal{C}_2(\gamma, \rho, A)$, one can approximate the tail empirical quantile process of the ε_i in a weighted uniform sense:

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

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 de Haan and Ferreira, 2006). In addition to

satisfying $k \rightarrow \infty$ and $k/n \rightarrow 0$, the sequence $k = k(n)$ should be such that $\sqrt{k}A(n/k) = O(1)$. For certain results which require the study of the log-spacings $\log \varepsilon_{n-[ks],n} - \log \varepsilon_{n-k,n}$, such as the convergence of the Hill estimator, the following approximation of the log-tail empirical quantile process is sometimes preferred: uniformly in $s \in (0, 1]$,

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

Our next result shows that, if the error made in the construction of the $\widehat{\varepsilon}_i^{(n)}$ is not too large, then these approximation results hold with the process $s \mapsto \varepsilon_{n-[ks],n}$ replaced by its residual-based version $s \mapsto \widehat{\varepsilon}_{n-[ks],n}^{(n)}$. This suggests that residual-based versions of the aforementioned estimators can be constructed and studied.

Theorem 2. *Assume that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds. Let $k = k(n) \rightarrow \infty$ be a sequence of integers such that $k/n \rightarrow 0$ and $\sqrt{k}A(n/k) = O(1)$. Suppose moreover that the array of random variables $\widehat{\varepsilon}_i^{(n)}$, $1 \leq i \leq n$, satisfies*

$$\forall i \in \{1, \dots, n\}, \quad |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \leq R_{n,i}(1 + |\varepsilon_i|) \quad \text{where} \quad \sqrt{k} \max_{1 \leq i \leq n} R_{n,i} \xrightarrow{\mathbb{P}} 0.$$

Then there exists a sequence W_n of standard Brownian motions such that, for any $\delta > 0$ sufficiently small:

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

and

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

uniformly in $s \in (0, 1]$.

As an illustration, we work out the asymptotic properties of the residual-based, Hill-type estimator of the tail index γ of the errors, as well as the asymptotic behaviour of the related indirect expectile estimator of the noise variable in the following corollary.

Corollary 1. *Assume that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds. Let $\tau_n \uparrow 1$ be such that $n(1 - \tau_n) \rightarrow \infty$ and $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \rightarrow \lambda \in \mathbb{R}$. Suppose moreover that the array of random variables $\widehat{\varepsilon}_i^{(n)}$, $1 \leq i \leq n$, satisfies*

$$\forall i \in \{1, \dots, n\}, \quad |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \leq R_{n,i}(1 + |\varepsilon_i|) \quad \text{where} \quad \sqrt{n(1 - \tau_n)} \max_{1 \leq i \leq n} R_{n,i} \xrightarrow{\mathbb{P}} 0.$$

Then, defining $\bar{\gamma} = \widehat{\gamma}_{[n(1-\tau_n)]}$ and $\bar{q}_{\tau_n}(\varepsilon) = \widehat{\varepsilon}_{n-[n(1-\tau_n)],n}^{(n)}$, we have

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

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 $\mathbb{E}|\varepsilon_-| < \infty$, $0 < \gamma < 1$ and $\sqrt{n(1-\tau_n)}/q_{\tau_n}(\varepsilon) = O(1)$, one has

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

with $m(\gamma) = (1-\gamma)^{-1} - \log(\gamma^{-1} - 1)$ and

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

One benefit of using the indirect expectile estimator $\tilde{\xi}_{\tau_n}(\varepsilon)$ is that its asymptotic variance tends to be lower than that of the direct LAWS estimator $\hat{\xi}_{\tau_n}(\varepsilon)$. In the case when $\bar{\gamma}$ is taken to be the Hill estimator, this is evidenced in Figure 1 in Daouia *et al.* (2018), where it is seen that the indirect estimator has lower variance than the direct estimator for $\gamma > 0.3$. However, the indirect estimator is typically biased, as Corollary 1 shows. A more complete comparison of these two estimators in the case of completely observed data is provided in Section 3 of Daouia *et al.* (2018).

Let us conclude by mentioning that one could of course estimate γ using the $Y_i = g(\mathbf{X}_i) + s(\mathbf{X}_i)\varepsilon_i$ directly, because the Y_i are obtained by changing location and scale of the ε_i and so they all have tail index γ . A competitor to the residual-based Hill estimator $\hat{\gamma}_k$ is thus

$$\check{\gamma}_k = \frac{1}{k} \sum_{i=1}^k \log \frac{Y_{n-i+1,n}}{Y_{n-k,n}}.$$

A numerical comparison of the estimators $\hat{\gamma}_k$ and $\check{\gamma}_k$ (which we do not report here for the sake of brevity) 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 location shift $g(\mathbf{X}_i)$ and of the scaling $s(\mathbf{X}_i)$ in the Y_i introduces a large amount of bias in the estimation of γ 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 de Haan and Ferreira (2006, p.83).

2.3 Extrapolation for extreme conditional expectile estimation

Our next and final focus in this theoretical toolbox is to develop high-level results for the estimation of properly extreme conditional expectiles $\xi_{\tau'_n}(Y|\mathbf{x})$, whose level $\tau'_n \rightarrow 1$ satisfies $n(1-\tau'_n) \rightarrow c < \infty$ as $n \rightarrow \infty$. One would typically choose $\tau'_n = 1 - 1/n$, or more generally $\tau'_n = 1 - p_n$ for an exceedance probability p_n not greater than $1/n$, see *e.g.* Chapter 4 of de Haan and Ferreira (2006) in the context of extreme quantile estimation. An important idea, developed by Weissman (1978), is that intermediate quantiles of order τ_n can be extrapolated to the extreme level τ'_n , using the approximate

Pareto shape of the underlying quantile function. This methodology 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

$$\frac{\xi_{\tau'_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} \approx \frac{q_{\tau'_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} = \frac{U((1 - \tau'_n)^{-1})}{U((1 - \tau_n)^{-1})} \approx \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\gamma} \quad \text{as } n \rightarrow \infty.$$

This suggests to consider the following class of plug-in estimators of $\xi_{\tau'_n}$:

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

where $\bar{\gamma}$ and $\bar{\xi}_{\tau_n}(\varepsilon)$ are consistent estimators of γ and of the intermediate expectile $\xi_{\tau_n}(\varepsilon)$, respectively. In our context, these would be based on residuals $\hat{\varepsilon}_i^{(n)} = (Y_i - \bar{g}(\mathbf{X}_i))/\bar{s}(\mathbf{X}_i)$ obtained via estimators $\bar{g}(\mathbf{x})$ and $\bar{s}(\mathbf{x})$ of $g(\mathbf{x})$ and $s(\mathbf{x})$. Once $\xi_{\tau'_n}(\varepsilon)$ is estimated, one can estimate $\xi_{\tau'_n}(Y|\mathbf{x})$ in model (1) by $\bar{\xi}_{\tau'_n}^*(Y|\mathbf{x}) = \bar{g}(\mathbf{x}) + \bar{s}(\mathbf{x})\bar{\xi}_{\tau'_n}^*(\varepsilon)$. Our next main result gives conditions for the convergence of this estimator.

Theorem 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 $\tau_n, \tau'_n \uparrow 1$ satisfy*

$$n(1 - \tau_n) \rightarrow \infty, \quad n(1 - \tau'_n) \rightarrow c < \infty, \quad \frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \rightarrow \infty, \quad (2)$$

$$\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \rightarrow \lambda \in \mathbb{R} \quad \text{and} \quad \frac{\sqrt{n(1 - \tau_n)}}{q_{\tau_n}(\varepsilon)} = O(1). \quad (3)$$

Suppose finally that the estimators $\bar{g}(\mathbf{x})$, $\bar{s}(\mathbf{x})$ and $\bar{\xi}_{\tau_n}(\varepsilon)$ are such that $\bar{g}(\mathbf{x}) - g(\mathbf{x}) = O_{\mathbb{P}}(1)$, $\sqrt{n(1 - \tau_n)}(\bar{s}(\mathbf{x}) - s(\mathbf{x})) = O_{\mathbb{P}}(1)$ and $\sqrt{n(1 - \tau_n)}(\bar{\xi}_{\tau_n}(\varepsilon)/\xi_{\tau_n}(\varepsilon) - 1) = O_{\mathbb{P}}(1)$. If there is a nondegenerate limiting random variable Γ such that $\sqrt{n(1 - \tau_n)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$, then

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

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

The goal of the next section is to work out several examples where this condition on the rate of estimation of the structure of the model can be satisfied. In each example, we explain our estimation method, and show how this yields consistent estimators of extreme conditional expectiles whose asymptotic behaviour we can obtain, using the high-level results we have just developed.

3 Applications of our theoretical results

3.1 Location-scale shift linear regression model

We concentrate here on applications of our results 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 ε , 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 $\mathbf{x} \in K$.

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,

$$\xi_{\tau_n}(Y|\mathbf{x}) = \alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon) \quad \text{with} \quad \xi_{\tau_n}(\varepsilon) \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty,$$

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 the two-stage following procedure to estimate α , $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, based on n independent data points $(\mathbf{X}_i, Y_i)_{1 \leq i \leq n}$.

1. (Preliminary step) Compute the ordinary least squares estimators $\tilde{\alpha}$ and $\tilde{\boldsymbol{\beta}}$ of α and $\boldsymbol{\beta}$, and then the ordinary least squares estimator of $\boldsymbol{\theta}$ based on the absolute residuals $\tilde{Z}_i = |Y_i - (\tilde{\alpha} + \tilde{\boldsymbol{\beta}}^\top \mathbf{X}_i)|$:

$$(\tilde{\alpha}, \tilde{\boldsymbol{\beta}}) = \arg \min_{(a, \mathbf{b})} \sum_{i=1}^n (Y_i - a - \mathbf{b}^\top \mathbf{X}_i)^2$$

and $\tilde{\boldsymbol{\theta}} = \frac{\tilde{\boldsymbol{\nu}}}{\tilde{\mu}}$ where $(\tilde{\mu}, \tilde{\boldsymbol{\nu}}) = \arg \min_{(c, \mathbf{d})} \sum_{i=1}^n (\tilde{Z}_i - c - \mathbf{d}^\top \mathbf{X}_i)^2$.

2. (Weighted step) Compute the least squares estimators $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ of α and $\boldsymbol{\beta}$, weighted according to estimated standard deviations obtained via $\tilde{\boldsymbol{\theta}}$, and then the weighted least squares estimator of $\boldsymbol{\theta}$ based on the absolute residuals $\hat{Z}_i = |Y_i - (\hat{\alpha} + \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i)|$:

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

and $\hat{\boldsymbol{\theta}} = \frac{\hat{\boldsymbol{\nu}}}{\hat{\mu}}$ where $(\hat{\mu}, \hat{\boldsymbol{\nu}}) = \arg \min_{(c, \mathbf{d})} \sum_{i=1}^n \left(\frac{\hat{Z}_i - c - \mathbf{d}^\top \mathbf{X}_i}{1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i} \right)^2$.

This procedure can be seen as a one-iteration version of a more general iterative weighted least squares procedure, where estimates obtained at a given step are fed back into the next iteration to update the weights, this procedure being repeated n_0 times, where n_0 has to be chosen. General iterative weighted least squares techniques have been found to work well for linear models featuring replications, but their performance in cases without replications is less clear (see Chen and Shao, 1993); even in models with replications, there are standard examples where the optimal number of iterations is simply $n_0 = 1$ (see Inoue, 2003). Our choice of one iteration is made because we have observed on simulation studies that iterating the procedure further does not seem to improve the accuracy of the estimators in practice.

Once these estimates have been obtained, one can focus on the estimation of $\xi_{\tau_n}(\varepsilon)$. We start by constructing the sample of (weighted) residuals $\widehat{\varepsilon}_i^{(n)} = (Y_i - (\widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i))/(1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{X}_i)$. One can then estimate $\xi_{\tau_n}(\varepsilon)$ by the direct LAWS estimator $\widehat{\xi}_{\tau_n}(\varepsilon)$ described in Section 2.1. The weighted least squares estimators of α , $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are \sqrt{n} -consistent (for the sake of completeness, we state this result as Lemma 4 in Appendix A.2). The consistency and asymptotic normality of $\widehat{\xi}_{\tau_n}(\varepsilon)$ are therefore a corollary of Theorem 1, and this in turn yields the asymptotic behaviour of the estimators $\widehat{\xi}_{\tau_n}(Y|\mathbf{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{x})\widehat{\xi}_{\tau_n}(\varepsilon)$ (at the intermediate level) and $\widehat{\xi}_{\tau'_n}^*(Y|\mathbf{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{x}) \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\bar{\gamma}} \widehat{\xi}_{\tau_n}(\varepsilon)$ (at the extreme level), where $\bar{\gamma}$ is a consistent estimator of γ constructed on the residuals.

Corollary 2. *Assume that the setup is that of the heteroscedastic linear model (M_1) . 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) *Assume here that ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0 < \gamma < 1/2$. Then for any $\mathbf{x} \in K$,*

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y|\mathbf{x})}{\xi_{\tau_n}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

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

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

We may similarly obtain the asymptotic normality of the indirect estimators $\widetilde{\xi}_{\tau_n}(Y|\mathbf{x})$ and $\widetilde{\xi}_{\tau'_n}^*(Y|\mathbf{x})$ of the intermediate and extreme expectiles $\xi_{\tau_n}(Y|\mathbf{x})$ and $\xi_{\tau'_n}(Y|\mathbf{x})$, defined as $\widetilde{\xi}_{\tau_n}(Y|\mathbf{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{x})(\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \widehat{\varepsilon}_{n - \lfloor n(1 - \tau_n) \rfloor, n}^{(n)}$ and $\widetilde{\xi}_{\tau'_n}^*(Y|\mathbf{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \mathbf{x} + (1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{x}) \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\bar{\gamma}} (\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \widehat{\varepsilon}_{n - \lfloor n(1 - \tau_n) \rfloor, n}^{(n)}$. Here $\bar{\gamma}$ can be taken as the Hill estimator

of γ calculated on the residuals, the asymptotic properties of the estimators being then obtained using Corollary 1 and Theorem 3. We omit the straightforward details for the sake of brevity.

3.2 Heteroscedastic single-index model

The linear structure of the location-scale shift linear regression model of Section 3.1 has its limitations in practice in that it fails to model more complex regression relationships between Y and \mathbf{X} . A more flexible model is the heteroscedastic single-index model we recall below.

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 ε , 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° . Besides, the variable ε 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.* Horowitz (2009). In this regression model, the conditional mean $\mathbb{E}(Y|\mathbf{X}) = g(\boldsymbol{\beta}^\top \mathbf{X})$ and variance $\text{Var}(Y|\mathbf{X}) = \text{Var}(\varepsilon)\sigma^2(\boldsymbol{\beta}^\top \mathbf{X})$ have the same single-index structure. There are variants of this model where the direction of projection in σ is a vector $\boldsymbol{\theta}$ possibly different from $\boldsymbol{\beta}$ (see *e.g.* Zhu *et al.*, 2013). In practice, model (M_2) is already very flexible, and for the sake of simplicity we therefore ignore this more general case.

In model (M_2), $\xi_{\tau_n}(Y|\mathbf{x}) = g(\boldsymbol{\beta}^\top \mathbf{x}) + \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \xi_{\tau_n}(\varepsilon)$. To estimate extreme conditional expectiles of Y , we shall estimate the direction $\boldsymbol{\beta}$, then the functions g and σ , and finally the extreme unconditional expectile $\xi_{\tau_n}(\varepsilon)$ of the error ε . There are numerous estimators (see *e.g.* Chapter 2 of Horowitz, 2009) which make it possible to estimate $\boldsymbol{\beta}$ at the rate \sqrt{n} . We thus assume that such an estimator $\widehat{\boldsymbol{\beta}}$ has been constructed, *i.e.* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_{\mathbb{P}}(1)$. It remains to find estimators of the functions g and σ . We do so nonparametrically, and we start by estimating g with

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

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 estimator is inspired by an estimator considered by Härdle and Stoker (1989), with the addition of a truncating sequence to help deal with the heavy tail of the errors. To estimate the function σ , we note that, analogously to what we observed in the location-scale shift linear regression model (M_1), $\sigma(\boldsymbol{\beta}^\top \mathbf{X})$ is in fact the conditional first moment of $|Y - g(\boldsymbol{\beta}^\top \mathbf{X})|$.

Again, we do not have access to the true values of β or g , so we introduce the absolute residuals $\widehat{Z}_{i,h_n,t_n} = \left| Y_i - \widehat{g}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{X}_i) \right|$ and consider the following (truncated) Nadaraya-Watson type estimator of σ :

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

In Proposition 2 (see Appendix A.2) 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 $n^{2/5}/\sqrt{\log n}$. Similar results, mostly on the estimation of the link function g , are available in the literature; see for example Kuchibhotla and Patra (2020) for an estimator based on smoothing splines, as well as references therein.

The residuals are then $\widehat{\varepsilon}_i^{(n)} = (Y_i - \widehat{g}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{X}_i)) / \widehat{\sigma}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{X}_i)$. Translated in terms of these residuals, Proposition 2 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 K_0 makes sense since it is well-known that kernel regression estimators strongly suffer from boundary effects (see, among others in the very large literature on the subject, Kyung-Joon and Schucany, 1998). This restriction is not important in practice for inference about the single-index model (M_2) , since one would in any case only trust the estimates of g and σ 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 number. Define

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

with $\tau_N = \tau_m$ on the event $N = m > 0$. This finally yields the intermediate conditional expectile estimator $\widehat{\xi}_{\tau_N}(Y|\mathbf{x}) = \widehat{g}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{x}) + \widehat{\sigma}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{x}) \widehat{\xi}_{\tau_N}(\varepsilon)$. Extrapolating the intermediate conditional expectile estimator to an extreme level then suggests the extreme conditional expectile estimator $\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x}) = \widehat{g}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{x}) + \widehat{\sigma}_{h_n,t_n}(\widehat{\beta}^\top \mathbf{x}) \left(\frac{1 - \tau'_N}{1 - \tau_N} \right)^{-\bar{\gamma}} \widehat{\xi}_{\tau_N}(\varepsilon)$. Again, the estimator $\bar{\gamma}$ is typically calculated using high order statistics of the residuals $\widehat{\varepsilon}_i^{(n)}$, and to guarantee that these approximate the true unobservable noise variables ε_i reasonably well, we restrict ourselves to using the residuals $\widehat{\varepsilon}_{1,K_0}^{(n)}, \dots, \widehat{\varepsilon}_{N,K_0}^{(n)}$ whose corresponding covariate vectors $\mathbf{X}_i \in K_0$. The next result examines the asymptotic distributions of the estimators $\widehat{\xi}_{\tau_N}(Y|\mathbf{x})$ and $\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x})$.

Theorem 4. Assume that ε satisfies condition $\mathcal{C}_1(\gamma)$ and that the conditions of Proposition 2 in Appendix A.2 hold (including $0 < \gamma < 1/2$). Let $\tau_n = 1 - n^{-a}$ with $a \in (1/5, 1)$, K_0 be a compact subset of K° such that $\mathbb{P}(\mathbf{X} \in K_0) > 0$, and $N = N(K_0, n)$.

(i) We have, for any $\mathbf{x} \in K_0$,

$$\sqrt{N(1 - \tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(Y|\mathbf{x})}{\xi_{\tau_N}(Y|\mathbf{x})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

(ii) Assume moreover that ε satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ with $\rho < 0$. Suppose also that $\tau_n, \tau'_n \uparrow 1$ satisfy (2) and (3). If there is a nondegenerate limiting random variable Γ such that $\sqrt{N(1 - \tau_N)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Gamma$, then for any $\mathbf{x} \in K_0$,

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

Compared to Corollary 2 in the location-scale shift linear regression model, Theorem 4 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 (semi-parametric) estimation of the structure of the model is faster than the common rate of estimation of the corresponding intermediate expectile of the noise variable and of the tail index γ . Another difference is that the order of the conditional expectile to be estimated is actually random and dictated by the quantity $N = N(K_0, n)$ of covariates $\mathbf{X}_i \in K_0$ (where the structure of the model can be estimated at the rate $n^{2/5}/\sqrt{\log n}$). As such, the rates of convergence of our estimators are also random and dictated by this available sample size N . This is in fact natural: the more data points we can use, the further we can go in the estimation of the conditional expectiles. Let us finally mention that, for any $\mathbf{x} \in K_0$, an analogous result is possible for an indirect extreme conditional expectile estimator, of the form

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

Just like $\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x})$, its asymptotic distribution is controlled by that of $\bar{\gamma}$. We omit the details.

3.3 Autoregressive time series models

In addition to their good axiomatic and statistical properties, it was noted by Bellini and Di Bernardino (2017) that expectiles can, in financial contexts, be interpreted in terms of the gain-loss ratio: an expectile defines the amount of money that should be added to a financial position in order to have a prespecified, sufficiently high gain-loss

ratio (calculated in terms of the order τ). The gain-loss ratio is a popular performance measure in portfolio management and is well-known in the literature on no good deal valuation in incomplete markets (see Bellini and Di Bernardino, 2017 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.

3.3.1 The AR(p) model

Let us start by describing what the implications of our results are in the classical autoregressive AR(p) model.

Model (M_3) The time series (Y_t) satisfies $Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + \varepsilon_t$ where $\phi_1, \dots, \phi_p \in \mathbb{R}$ are unknown coefficients. It is assumed that the process (Y_t) is causal, *i.e.* the polynomial $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$ has no root inside the unit disk of the complex plane. Suppose also that (ε_t) is an independent sequence of centred and square-integrable copies of an innovation variable ε with balanced tails, in the sense that $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \rightarrow p \in (0, 1]$ as $x \rightarrow \infty$.

In model (M_3) it is known, for instance following the discussion in Section 3.1 of Brockwell and Davis (1991), that (Y_t) is a stationary random process that can be represented as a linear time series. A conditional expectile of Y here therefore means an expectile of the present observation given p past values, which can be formalised as

$$\xi_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p}) = \sum_{j=1}^p \phi_j y_{-j} + \xi_{\tau_n}(\varepsilon)$$

where $y_{-1}, \dots, y_{-p} \in \mathbb{R}$ represent the p lagged conditioning values. In this context, the problem of extreme conditional expectile estimation can thus be seen as a simple version of dynamic extreme expectile estimation, not to be confused with the altogether different problem of estimating extreme expectile of the marginal distribution of the time series (Y_t) .

To estimate the coefficients ϕ_j so as to be able to filter the time series (Y_t) and then estimate $\xi_{\tau_n}(\varepsilon)$, a wealth of methods are available, such as the Yule-Walker estimator, the maximum likelihood estimator assuming a Gaussian model, and a version of the least squares estimator. All these estimators are \sqrt{n} -asymptotically normal, see Theorem 8.1.1 and Theorem 10.8.2 in Brockwell and Davis (1991). For the sake of simplicity, and to present a unified result, we then assume that the coefficients $\phi_1, \dots, \phi_p \in \mathbb{R}$ are estimated by $\widehat{\phi}_{1,n}, \dots, \widehat{\phi}_{p,n}$ such that for any $j \in \{1, \dots, p\}$, $\widehat{\phi}_{j,n} = \phi_j + O_{\mathbb{P}}(n^{-1/2})$. This allows one to construct the residuals $\widehat{\varepsilon}_t^{(n)} = Y_t - \sum_{j=1}^p \widehat{\phi}_{j,n} Y_{t-j}$, for $p+1 \leq t \leq n$. We

have the following result on the estimators

$$\begin{aligned}\widehat{\xi}_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p}) &= \sum_{j=1}^p \widehat{\phi}_{j,n} y_{-j} + \widehat{\xi}_{\tau_n}(\varepsilon) \quad (\tau_n \text{ intermediate}) \\ \text{and } \widehat{\xi}_{\tau'_n}^*(Y_0 | y_{-1}, \dots, y_{-p}) &= \sum_{j=1}^p \widehat{\phi}_{j,n} y_{-j} + \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}} \widehat{\xi}_{\tau_n}(\varepsilon) \quad (\tau'_n \text{ extreme})\end{aligned}$$

where $\widehat{\xi}_{\tau_n}(\varepsilon)$ is the direct LAWS estimator of $\xi_{\tau_n}(\varepsilon)$ based on the residuals and $\bar{\gamma}$ is a consistent estimator of γ , both constructed on the residuals $\widehat{\varepsilon}_t^{(n)}$, for $p + 1 \leq t \leq n$.

Corollary 3. *Assume that the setup is that of the autoregressive model (M_3) . 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) *Assume that ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0 < \gamma < 1/2$, and that $n^{2\gamma+\delta}(1 - \tau_n) \rightarrow 0$ for some $\delta > 0$. Then for any $y_{-1}, \dots, y_{-p} \in \mathbb{R}$,*

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})}{\xi_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

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

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left(\frac{\widehat{\xi}_{\tau'_n}^*(Y_0 | y_{-1}, \dots, y_{-p})}{\xi_{\tau'_n}(Y_0 | y_{-1}, \dots, y_{-p})} - 1 \right) \xrightarrow{d} \Gamma.$$

Linear autoregressive models such as the AR(p) are simple and widely applicable, but well-known for failing to replicate the time-varying volatility and volatility clustering typically displayed by financial time series. Our next focus is on the class of ARCH models introduced by Engle (1982), which is arguably the simplest and best-known class of heteroscedastic time series models.

3.3.2 The ARCH(p) model

We focus here on the general formulation of the ARCH model with p lags.

Model (M_4) The time series (Y_t) satisfies $Y_t = \sigma_t \varepsilon_t$, with $\sigma_t = \sqrt{\alpha + \sum_{j=1}^p \beta_j Y_{t-j}^2}$, where $\alpha, \beta_1, \dots, \beta_p > 0$ are unknown coefficients and (ε_t) is an independent sequence of copies of a centred innovation variable ε such that $\mathbb{E}(\varepsilon^2) = 1$ and $\mathbb{P}(\varepsilon^2 = 1) < 1$.

Suppose also that the sequence of matrices

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

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

The conditions on the Lyapunov exponent of (A_t) are necessary and sufficient for the existence of a (strictly) stationary ARCH(p) sequence. This solution is in fact also nonanticipative, in the sense that Y_t is measurable with respect to the past σ -field generated by $\{\varepsilon_s, s \leq t\}$, and ergodic. We refer here to Section 2.2 of Francq and Zakoïan (2010). The condition $\mathbb{P}(\varepsilon^2 = 1) < 1$ ensures identifiability of the parameters.

Noting that $Y_t^2 = \alpha + \sum_{j=1}^p \beta_j Y_{t-j}^2 + u_t$, with $u_t = \sigma_t^2(\varepsilon_t^2 - 1)$, we find that Y_t^2 satisfies an AR(p)-type model. A sensible method for the estimation of the unknown coefficients $\alpha, \beta_1, \dots, \beta_p$ is then to use a weighted least squares estimator in the spirit of the estimator used in Section 3.1 for the estimation of the location-scale shift linear regression model:

$$(\widehat{\alpha}_n, \widehat{\beta}_{1,n}, \dots, \widehat{\beta}_{p,n}) = \arg \min_{(a, b_1, \dots, b_p)} \sum_{t=p+1}^n \left(\frac{Y_t^2 - a - b_1 Y_{t-1}^2 - \cdots - b_p Y_{t-p}^2}{\widetilde{\alpha}_n + \widetilde{\beta}_{1,n} Y_{t-1}^2 + \cdots + \widetilde{\beta}_{p,n} Y_{t-p}^2} \right)^2$$

where $(\widetilde{\alpha}_n, \widetilde{\beta}_{1,n}, \dots, \widetilde{\beta}_{p,n}) = \arg \min_{(a, b_1, \dots, b_p)} \sum_{t=p+1}^n (Y_t^2 - a - b_1 Y_{t-1}^2 - \cdots - b_p Y_{t-p}^2)^2$.

It is shown in, among others, Theorem 6.3 p.132 of Francq and Zakoïan (2010) that the estimator $(\widehat{\alpha}_n, \widehat{\beta}_{1,n}, \dots, \widehat{\beta}_{p,n})$ is \sqrt{n} -asymptotically normal if $\mathbb{E}(\varepsilon^4) < \infty$. A constrained version of the estimator forcing the estimated parameters to be strictly positive with probability 1 is possible (see for instance Section 6.3 in Francq and Zakoïan, 2010), but we shall not pursue this here for the sake of simplicity. Constructing the residuals as $\widehat{\varepsilon}_t^{(n)} = Y_t / (\widehat{\alpha}_n + \sum_{j=1}^p \widehat{\beta}_{j,n} Y_{t-j}^2)^{1/2}$, for $p+1 \leq t \leq n$, we now find estimators of extreme conditional expectiles of Y_0 given p past values y_{-1}, \dots, y_{-p} as

$$\widehat{\xi}_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p}) = \left(\widehat{\alpha}_n + \sum_{j=1}^p \widehat{\beta}_{j,n} Y_{t-j}^2 \right)^{1/2} \widehat{\xi}_{\tau_n}(\varepsilon) \quad (\tau_n \text{ intermediate})$$

and $\widehat{\xi}_{\tau'_n}^*(Y_0 | y_{-1}, \dots, y_{-p}) = \left(\widehat{\alpha}_n + \sum_{j=1}^p \widehat{\beta}_{j,n} Y_{t-j}^2 \right)^{1/2} \times \left(\frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\bar{\gamma}} \widehat{\xi}_{\tau_n}(\varepsilon) \quad (\tau'_n \text{ extreme})$

where $\widehat{\xi}_{\tau_n}(\varepsilon)$ is the direct LAWS estimator of $\xi_{\tau_n}(\varepsilon)$ based on residuals and $\bar{\gamma}$ is a consistent estimator of γ , both constructed on the residuals $\widehat{\varepsilon}_t^{(n)}$.

Corollary 4. Assume that the setup is that of the autoregressive conditional heteroscedastic model (M_4). Suppose also that $\mathbb{E}|\varepsilon_-|^4 < \infty$, and that $\tau_n \uparrow 1$ is such that $n(1 - \tau_n) \rightarrow \infty$.

(i) Assume that ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0 < \gamma < 1/4$. Then for any $y_{-1}, \dots, y_{-p} \in \mathbb{R}$,

$$\sqrt{n(1 - \tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})}{\xi_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1 - 2\gamma} \right).$$

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

$$\frac{\sqrt{n(1 - \tau_n)}}{\log[(1 - \tau_n)/(1 - \tau'_n)]} \left(\frac{\widehat{\xi}_{\tau'_n}^*(Y_0 | y_{-1}, \dots, y_{-p})}{\xi_{\tau'_n}(Y_0 | y_{-1}, \dots, y_{-p})} - 1 \right) \xrightarrow{d} \Gamma.$$

4 Finite-sample study

In this section, we showcase our estimators both on simulated and real data. Sections 4.1 and 4.2 contain simulation studies, first in the case of regression models in the spirit of those of Sections 3.1 and 3.2, and then in the time series contexts examined in Section 3.3. Sections 4.3 and 4.4 apply our methods to, respectively, a set of insurance data and a financial time series. Throughout this section, and in order to improve finite-sample results, the direct and indirect versions of our estimators $\widehat{\xi}_{\tau_n}^*$ and $\widetilde{\xi}_{\tau_n}^*$ and the Hill estimator $\widehat{\gamma}_k$ feature a bias reduction method inspired by results of Gomes *et al.* (2016) and Daouia *et al.* (2018), and are thus denoted by $\widehat{\xi}_{\tau_n}^{*,\text{RB}}$, $\widetilde{\xi}_{\tau_n}^{*,\text{RB}}$ and $\widehat{\gamma}_k^{\text{RB}}$. To save space, we only give a general idea of our setup and present the main conclusions we can draw from the simulation studies and data analyses; full implementation details and numerical results can be found in Appendix C.

4.1 Simulation study: linear and single-index models

We simulate here $N = 500$ replications of data sets made of $n = 1,000$ observations (\mathbf{X}_i, Y_i) , $1 \leq i \leq n$, with $\mathbf{X} \in \mathbb{R}^4$. We simulate from two different models describing the relationship between Y and \mathbf{X} :

$$(G1) \quad Y = 1 + \boldsymbol{\beta}^\top \mathbf{X} + \left(\frac{1}{2} + \boldsymbol{\beta}^\top \mathbf{X} \right) \varepsilon,$$

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

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 $\boldsymbol{\beta} = (1, 1, 1, 1)$ and ε is a noise variable, independent of \mathbf{X} , having a normalised symmetric Burr distribution with tail index $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$ and 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 strategies:

- (S1) Estimate directly $\xi_{\tau_n}(Y|\mathbf{x})$ by the unconditional estimator $\xi_{\tau_n}(Y)$ using the direct LAWS estimator of Daouia *et al.* (2018).
- (S2) We assume that Y is linked to \mathbf{X} by the linear heteroscedastic model of Section 3.1 and use the strategy outlined there (with the direct estimator only).
- (S3) We assume that Y is linked to \mathbf{X} by the heteroscedastic single index model of Section 3.2 and use the strategy outlined there with the direct estimator.
- (S4) Identical to (S3), but the indirect estimator is used instead.

We also compare these procedures with the following two benchmarks:

- (B1) We assume that Y is linked to \mathbf{X} by a homoscedastic single index model with $\boldsymbol{\beta}$ known and equal to $(1, 1, 1, 1)$. The extreme expectile is estimated using the direct estimator.
- (B2) Identical to (S3) above, although $\boldsymbol{\beta}$ is assumed to be known and equal to $(1, 1, 1, 1)$.

In each of these procedures, the intermediate expectile level used in the tail index and extreme expectile estimators is fixed at 0.9, corresponding to $k_n = n(1 - \tau_n) = 100$. We estimate extreme conditional expectiles at multiple orders, each larger than τ_n , and compare our estimators in terms of relative Root Mean Squared Error. These errors are given in Tables C.1 and C.2 in Appendix C.1.

It may be seen that in model (M1), which is exactly linear, method (S2) is clearly the best and very close to the benchmark (B2), and single-index based methods (S3) and (S4) perform fairly well. In model (M2), methods (S3) and (S4) are the best, with a slight advantage for the direct estimator used in method (S3), although method (M2) remains surprisingly competitive even though its performance markedly deteriorates when the tail of the noise variable gets heavier, since then incorrectly estimating the heteroscedasticity factor becomes more aproblematic. Methods (S2), (S3) and (S4) clearly outperform the naive method (S1) in every tested setup. Note also that not accounting for heteroscedasticity is very detrimental to the finite-sample performance of the estimator, as benchmark (B1) shows (despite the unrealistic advantage of a correct pre-specification of the direction $\boldsymbol{\beta}$).

4.2 Simulation study: autoregressive time series models

In this section, we simulate $N = 1,000$ time series (Y_0, \dots, Y_n) of size $n + 1 = 1,001$. We consider two time series models:

(T1) An AR(1) model $Y_t = \phi Y_{t-1} + \varepsilon_t$.

(T2) An ARCH(1) model $Y_t = \sqrt{\alpha + \beta Y_{t-1}^2} \varepsilon_t$.

In both models, the series ε_t is an independent sequence of innovations having a symmetric Burr distribution with tail index $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$ and second-order parameter $\rho = -1$.

In each sample $(Y_0^{(j)}, \dots, Y_n^{(j)})$, the aim is to estimate a one-step ahead extreme expectile on observation time n ; letting $Y_n^{(j)} = y^{(j)}$, this translates, by stationarity, into estimating an extreme expectile $\xi_{\tau'_n}(Y_0 | Y_{-1} = y^{(j)})$. Following the methodology introduced in Section 3.3, we estimate model coefficients using a least squares estimator, and we then compute the N direct and indirect conditional expectile estimates at level $\tau'_n = 1 - 5/n = 0.995$. In order to quantify the accuracy of these estimators, we calculate their relative Root Mean Squared Error (RMSE). In the AR(1) model, we take $\phi \in \{0.1, 0.3, 0.5\}$; in the ARCH(1) model, we fix either $\alpha = 1/2$ and take $\beta \in \{0.1, 0.3, 0.5\}$, or we fix $\beta = 0.5$ and take $\alpha \in \{0.1, 0.3, 0.5\}$.

The obtained RMSE are given in Tables C.3 and C.4 in Appendix C. It can be seen there that, in the AR(1) model, the RMSE tends to increase with ϕ , that is, when the data (Y_t) get more dependent. In the ARCH(1) model, there is (perhaps unsurprisingly) no dependence on α , but the dependence on β is less clear. In both models, the performance of the estimators decreases as the tail of the noise variable gets heavier, and the indirect estimator has an advantage over the direct estimator.

4.3 Real data analysis: vehicle insurance data

We consider here the “Vehicle Insurance Customer Data” (available at <https://www.kaggle.com/ranja7/vehicle-insurance-customer-data> and upon request with the authors), made of $n = 9,134$ total (in the sense of 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. Our aim is to estimate an expectile $\xi_{\tau'_n}(Y|\mathbf{x})$. We propose to follow the methodology of Section 3.2 and use the heteroscedastic single-index model in conjunction with kernel regression and the direct extreme expectile estimator. A cross-validation procedure using the R function `npindexbw` (from the package `np`) gives a selected value of the bandwidth $h^* \approx 0.1$ (all covariates having been standardised by their respective maxima). We also choose $t^* = \infty$ and with these choices, we obtain $\hat{\beta} \simeq (-0.923, 0.386, -0.001, -0.002)$, which seems to indicate that only lifetime value X_1 and income X_2 appear to play a role in the prediction of Y . The

empirical functions \hat{g} and $\hat{\sigma}$ are depicted in the top left panel of Figure 1 (the kernel function L is taken to be the Epanechnikov kernel).

Equipped with these estimates, we now estimate extreme conditional expectiles $\xi_{\tau'_n}(Y|\mathbf{x})$ at levels $\tau'_n \in \{1 - 5/(nh^*), 1 - 2.5/(nh^*), 1 - 1/(nh^*)\} = \{0.995, 0.997, 0.999\}$. The top right panel of Figure 1 then shows the behaviour of the direct extreme conditional expectile estimator for $k^* = 200$ and $\tau^* = 1 - k^*/n$ (the right panel of Figure C.1 in Appendix C.3 shows that the heavy-tailed assumption is indeed reasonable). It can be seen there that the heteroscedasticity in the single-index model specification captures the variation in the shape of the data cloud fairly well, and the extreme conditional expectile curves give a reasonable idea of the conditional extremes of the data. The question of interpreting the expectile curves, meanwhile, is not always straightforward (see *e.g.* Chen, 2018). However, in this insurance example, the expectile $\xi_{\tau'_n}(Y|\mathbf{x})$ satisfies the following gain-loss ratio criterion (see Bellini and Di Bernardino, 2017):

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

In other words, the quantity $\xi_{\tau'_n}(Y|\mathbf{x})$ can be thought of as 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 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.

4.4 Real data analysis: AUD/CHF exchange rate

We study here the daily log-returns of the AUD/CHF exchange rate from 1st March 2015 to 28th February 2019, represented in the bottom left panel of Figure 1 and giving a sample of size $n = 1,043$. McKenzie (1997) has shown that ARCH models were fitting Australian exchange rates very well, and fitted the AUD/CHF rate with an ARCH(2) model. We therefore apply the methodology introduced in Section 3.3.2 with $p = 2$; the weighted least squares estimators we discussed return $\hat{\alpha}_n = 0.0000306$, $\hat{\beta}_{1,n} = 0.141$ and $\hat{\beta}_{2,n} = 0.126$. It has been suggested several times in the literature that the good axiomatic and sample properties of expectiles can be used to estimate quantiles (see *e.g.* Taylor, 2008 and Bellini and Di Bernardino, 2017). We follow this idea here and we estimate the extreme conditional quantile $q_{\tau'_n}(Y_t | Y_{t-1}, Y_{t-2})$ of level $\tau'_n = 0.995 \approx 1 - 5/n$, for each $t \in \{3, \dots, n\}$. For that purpose, we invert the construction of Section 2.2: from the residual-based LAWS estimator $\hat{\xi}_{\tau'_n}^{\text{*,RB}}(\varepsilon)$, we construct the extreme quantile estimator $\hat{q}_{\tau'_n}^{\text{*,RB}}(\varepsilon) = ((\hat{\gamma}_k^{\text{RB}})^{-1} - 1) \hat{\gamma}_k^{\text{RB}} \hat{\xi}_{\tau'_n}^{\text{*,RB}}(\varepsilon)$. We fix $k^* = 50$, $\tau_n = 1 - k^*/(n - p)$ and get estimates $\hat{\gamma}_k^{\text{RB}} = 0.214$ and $\hat{q}_{\tau'_n}^{\text{*,RB}}(\varepsilon) = 2.54$, with

corresponding estimates $\widehat{q}_{\tau'_n}^{\star,\text{RB}}(Y_t | Y_{t-1}, Y_{t-2})$ represented in the bottom right panel of Figure 1 (evidence of a heavy right tail is given on the right panel of Figure C.2 in Appendix C.3). These estimates appear to account well for the sharp decrease in volatility happening around March 2016. To check further that our estimates make sense, we recall the characterisation of $q_{\tau'_n}(\varepsilon)$ as $0.995 = \tau'_n = \mathbb{P}(\varepsilon \leq q_{\tau'_n}(\varepsilon)) = \mathbb{E}(\mathbb{1}\{\varepsilon \leq q_{\tau'_n}(\varepsilon)\})$ and compare that with

$$\frac{1}{n-p} \sum_{i=1}^{n-p} \mathbb{1}\{\widehat{\varepsilon}_i < \widehat{q}_{\tau'_n}^{\star,\text{RB}}(\varepsilon)\} \approx 0.99520.$$

This is indeed very close to the expected value $\tau'_n = 0.995$. Our estimate can be compared with a (bias-reduced version of the) classical extrapolated estimate $\widetilde{q}_{\tau'_n}^{\star,\text{RB}}(\varepsilon)$ of Weissman (1978), which is $\widetilde{q}_{\tau'_n}^{\star,\text{RB}}(\varepsilon) = 2.65$. Our estimate $\widehat{q}_{\tau'_n}^{\star,\text{RB}}(\varepsilon) = 2.54$ is slightly lower; this makes sense, as the estimated value of γ is lower than $1/4$, and thus extreme expectile-based estimates can be thought to reflect this rather light tail by producing lower point estimates than their quantile counterparts.

5 Discussion

Expectile regression is a way of generalising regression for the mean. Consequently, regression models provide a sensible and natural framework for the development of extreme conditional expectile estimation methods. This is the contribution of this paper, in regression models that can cope with a large but fixed dimension of the explanatory covariate. 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 estimators of extreme expectiles of the noise variable, based on these residuals. In practice this readily yields asymptotically normal estimators of extreme conditional expectiles, as we show in our four worked-out examples covering some important cases of regression models. The methods appear to work reasonably well in our applications to insurance and financial data.

In doing so, we also contribute to the broader question of how to model extremes in the presence of a large number of covariates. The works of Wang *et al.* (2012), Wang and Li (2013), He *et al.* (2016) and Gardes (2018) introduce dedicated modelling assumptions on the tail conditional quantiles of Y . The general form of the model of Wang *et al.* (2012) assumes that the quantile regression function $\mathbf{x} \mapsto q_\tau(Y|\mathbf{x})$ is linear, for τ sufficiently close to 1, with varying slope and intercept terms. This model is not straightforward to interpret in terms of its relationship between Y and \mathbf{x} : even in the simpler case when the quantile regression function $\mathbf{x} \mapsto q_\tau(Y|\mathbf{x})$ is assumed to be linear (for any value of τ), this model is the arguably complicated linear model linking Y to \mathbf{X} with random coefficients (see p.808 of Chernozhukov, 2005). By comparison, our generic model provides a straightforward way of seeing the effect \mathbf{X} has on Y . It also has the merit of avoiding any kind of crossing problem (unlike the method of Wang *et*

al., 2012), since the structure of the model is estimated only once. The nonparametric model of Gardes (2018), meanwhile, rests upon the assumption that there exists a Tail Dimension Reduction subspace, which is essentially a central subspace (see *e.g.* Cook and Li, 2002) specific to the tail of Y . Estimating this subspace can therefore only be done using the pairs (\mathbf{X}_i, Y_i) such that Y_i is large. This entails a potentially substantial loss of interpretability and modelling strength compared to when the model was actually valid for, and estimated using, the whole sample, with its extremes inferred afterwards, which is our approach here. Besides, the aforementioned papers focus on the case of independent and identically distributed data (\mathbf{X}_i, Y_i) ; our method, meanwhile, only assumes that the noise variables are independent and identically distributed, which allows us to consider, in addition to classical regression models such as heteroscedastic linear or single-index models, time series examples such as the popular AR and ARCH models.

There are of course numerous perspectives for future work, of which we mention a few here. We have not considered estimators of the structure of the model that would be robust to the heavy-tailed structure of the noise. This is partly because they are more difficult to implement than standard solutions such as weighted least squares or Nadaraya-Watson estimators, but also because the performance of such estimators is actually reasonably good in practice (a related point is made by Davis *et al.*, 1992 for autoregressive model estimation). One fruitful avenue for further research will be to investigate how alternative estimators for the structure of the model, putting less weight on the tails but having equivalent rates of convergence, can be used within our toolbox. It will also be important to generalise our construction to wider classes of time series models. One would certainly want our method to be able to be used for the analysis of GARCH models, for instance. Finally, and even though we focus here on expectiles, the basic principle of our approach relies on the fact that the quantity to be estimated is location equivariant and positive homogeneous. These two requirements are satisfied by numerous other functionals, such as the wide class of coherent spectral risk measures (Acerbi, 2002) containing the very recently introduced concept of *extremiles* (Daouia *et al.*, 2019). The latter concept defines a new least squares analogue of quantiles, represents a novel way of generalising mean regression, and has several appealing properties. The adaptation of our general approach to the estimation of extreme conditional versions of extremiles therefore constitutes an interesting perspective for further work.

Acknowledgements

This research was supported by the French National Research Agency under the grant ANR-19-CE40-0013-01/ExtremReg project. Stéphane Girard gratefully acknowledges the support of the Chair Stress Test, Risk Management and Financial Steering, led by the French Ecole Polytechnique and its Foundation and sponsored by BNP Paribas, and the support of the French National Research Agency in the framework of the Investissements d’Avenir programme (ANR-15-IDEX-02).

References

- Abdous, B. and Remillard, B. (1995). Relating quantiles and expectiles under weighted-symmetry, *Annals of the Institute of Statistical Mathematics* **47**: 371–384.
- Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance* **26**: 1505–1518.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk, *Mathematical Finance* **9**: 203–228.
- Bellini, F., Klar, B., Müller, A. and Gianin, E.R. (2014). Generalized quantiles as risk measures, *Insurance: Mathematics and Economics* **54**: 41–48.
- Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles, *The European Journal of Finance* **23**: 487–506.
- Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods* (second edition), Springer, New York.
- Chavez-Demoulin, V. and Davison, A.C. (2005). Generalized additive modelling of sample extremes, *Journal of the Royal Statistical Society: Series C* **54**: 207–222.
- Chavez-Demoulin, V., Embrechts, P. and Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates, *Journal of Risk and Insurance* **83**: 735–776.
- Chen, J.M. (2018). On exactitude in financial regulation: Value-at-Risk, Expected Shortfall, and expectiles, *Risks* **6**: 61.
- Chen, J. and Shao, J. (1993). Iterative weighted least squares estimators, *Annals of Statistics* **21**: 1071–1092.
- Chernozhukov, V. (2005). Extremal quantile regression, *Annals of Statistics* **33**: 806–839.
- Cook, R.D. and Li, B. (2002). Dimension reduction for conditional mean in regression, *Annals of Statistics* **30**: 455–474.
- Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves, *TEST* **20**: 311–333.
- 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.
- 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.
- Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment, *Bernoulli* **26**: 531–556.
- Davis, R.A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with

- infinite variance, *Stochastic Processes and their Applications* **40**: 145–180.
- Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society: Series B* **52**: 393–442.
- 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.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**: 987–1007.
- Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*, John Wiley & Sons.
- Gardes, L. (2018). Tail dimension reduction for extreme quantile estimation, *Extremes* **21**: 57–95.
- Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator, *Extremes* **17**: 45–75.
- Goegebeur, Y., Guillou, A. and Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **51**: 1190–1213.
- Gomes, M.I., Brillhante, 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.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer.
- 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.
- 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.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics* **3**: 1163–1174.
- Holzmann, H. and Klar, B. (2016). Expectile asymptotics, *Electronic Journal of Statistics* **10**: 2355–2371.
- Horowitz, J.L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*, Springer.
- Inoue, K. (2003). Iterative weighted least-squares estimates in a heteroscedastic linear regression model, *Journal of Statistical Planning and Inference* **110**: 133–146.
- Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica* **46**: 33–50.
- Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures, *Scandinavian Journal of Statistics* **44**: 425–454.

- 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.
- Kuchibhotla, A.K. and Patra, R.K. (2020). Efficient estimation in single index models through smoothing splines, *Bernoulli* **26**: 1587–1618.
- Kyung-Joon, C. and Schucany, W.R. (1998). Nonparametric kernel regression estimation near endpoints, *Journal of Statistical Planning and Inference* **66**: 289–304.
- McKenzie, M.D. (1997). ARCH modelling of Australian bilateral exchange rate data, *Applied Financial Economics* **7**: 147–164.
- Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing, *Econometrica* **55**: 819–847.
- Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer.
- 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.
- Segers, J. (2001). Residual estimators, *Journal of Statistical Planning and Inference* **98**: 15–27.
- Taylor, J.W. (2008). Estimating Value at Risk and Expected Shortfall using expectiles, *Journal of Financial Econometrics* **6**: 231–252.
- 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.
- Wang, H.J. and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation, *Journal of American Statistical Association* **108**: 1062–1074.
- 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.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association* **73**: 812–815.
- Zhu, L., Dong, Y. and Li, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling, *Statistica Sinica* **23**: 1235–1255.
- Ziegel, J.F. (2016). Coherence and elicibility, *Mathematical Finance* **26**: 901–918.

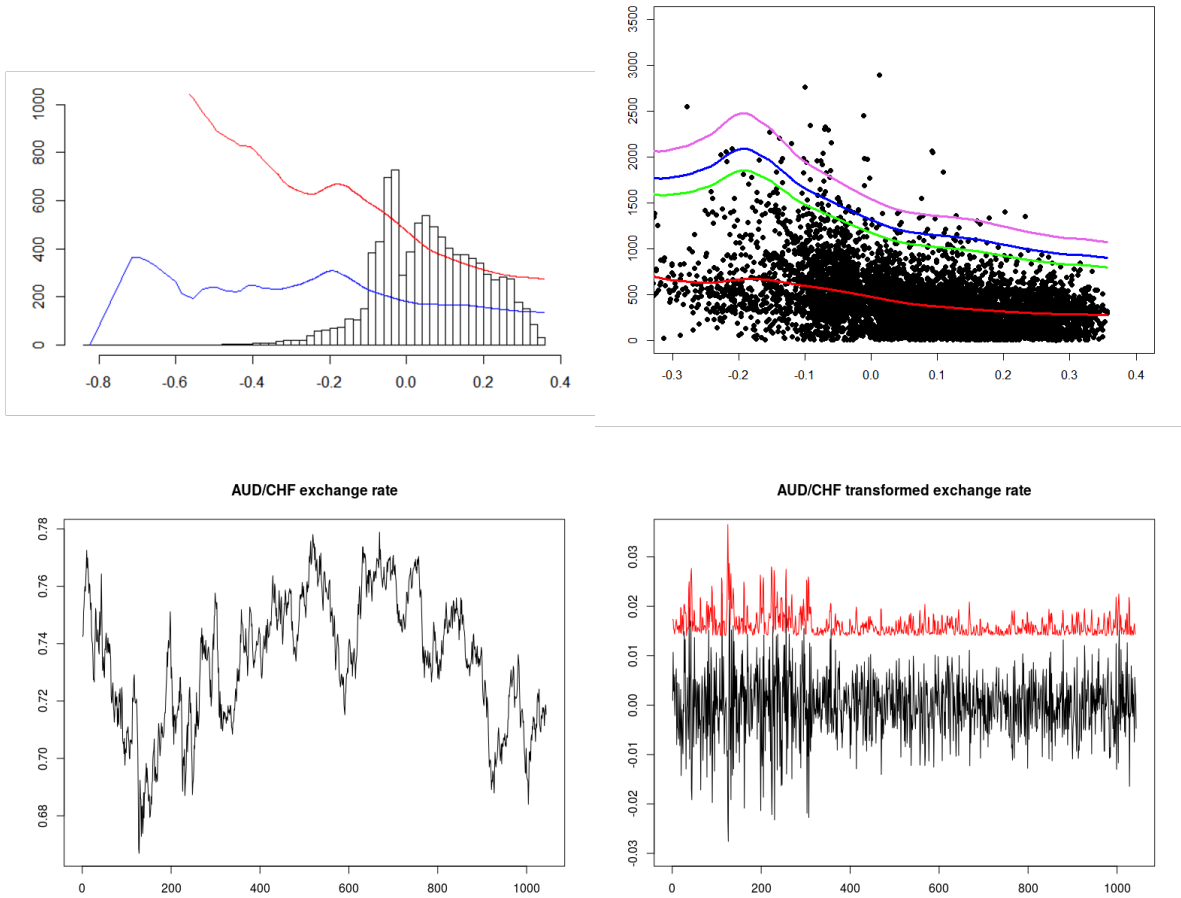


Figure 1: Top: Vehicle Insurance Customer data, left: estimates of g and σ superimposed on a histogram of the $\hat{\beta}^T \mathbf{X}_i$, right: estimates of the regression mean (red curve) and of the estimated conditional expectile estimators at levels $\tau'_n = 1 - 5/(nh^*)$ (green curve), $\tau'_n = 1 - 2.5/(nh^*)$ (blue curve) and $\tau'_n = 1 - 1/(nh^*)$ (purple curve) in the $(\hat{\beta}^T \mathbf{x}, y)$ plane. Bottom: AUD/CHF exchange rate data, left: daily AUD/CHF exchange rate from 1st March 2015 to 28th February 2019, right: daily log-returns (black curve) and estimated conditional quantile curve (red curve) at level $\tau'_n = 0.995$ over the observation period.

Supplementary Material for Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models

Stéphane Girard, Gilles Stupfler and Antoine Usseglio-Carleve

This supplementary material document contains the proofs of all theoretical results in the main paper, preceded by auxiliary results and their proofs (Sections A and B). It also provides further simulation results and complementary figures related to our data analyses (Section C).

A Appendix: auxiliary results and their proofs

A.1 Theoretical toolbox

Lemma 1 below is a result on the mean excess function of a sample of heavy-tailed random variables, used in the proof of Theorem 1.

Lemma 1. *Assume that there is $\delta > 0$ such that $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$, that ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0 < \gamma < 1/2$ and $\tau_n \uparrow 1$ is such that $n(1 - \tau_n) \rightarrow \infty$. Let moreover $t_n \rightarrow \infty$ be a nonrandom sequence such that $\bar{F}(t_n)/(1 - \tau_n) \rightarrow c \in (0, \infty)$. Then*

$$\frac{1}{nt_n(1 - \tau_n)} \sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i > t_n\} \xrightarrow{\mathbb{P}} \frac{c}{1 - \gamma}.$$

Proof of Lemma 1. Write first

$$\frac{1}{nt_n(1 - \tau_n)} \sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i > t_n\} = \frac{c + o(1)}{nt_n \bar{F}(t_n)} \sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i > t_n\}.$$

The idea is now to split the sum on the right-hand side as follows:

$$\frac{1}{nt_n \bar{F}(t_n)} \sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i > t_n\} = \frac{1}{n \bar{F}(t_n)} \sum_{i=1}^n \mathbb{1}\{\varepsilon_i > t_n\} + \frac{1}{nt_n \bar{F}(t_n)} \sum_{i=1}^n (\varepsilon_i - t_n) \mathbb{1}\{\varepsilon_i > t_n\}.$$

Straightforward expectation and variance calculations yield

$$\mathbb{E} \left(\frac{1}{n\bar{F}(t_n)} \sum_{i=1}^n \mathbb{1}\{\varepsilon_i > t_n\} \right) = 1,$$

$$\text{Var} \left(\frac{1}{n\bar{F}(t_n)} \sum_{i=1}^n \mathbb{1}\{\varepsilon_i > t_n\} \right) = O \left(\frac{1}{n\bar{F}(t_n)} \right) = O \left(\frac{1}{n(1-\tau_n)} \right) \rightarrow 0,$$

$$\mathbb{E} \left(\frac{1}{nt_n\bar{F}(t_n)} \sum_{i=1}^n (\varepsilon_i - t_n) \mathbb{1}\{\varepsilon_i > t_n\} \right) = \frac{1}{t_n} \int_{t_n}^{\infty} \frac{\bar{F}(x)}{\bar{F}(t_n)} dx \rightarrow \frac{\gamma}{1-\gamma},$$

and
$$\text{Var} \left(\frac{1}{nt_n\bar{F}(t_n)} \sum_{i=1}^n (\varepsilon_i - t_n) \mathbb{1}\{\varepsilon_i > t_n\} \right) = O \left(\frac{1}{n\bar{F}(t_n)} \right) = O \left(\frac{1}{n(1-\tau_n)} \right) \rightarrow 0.$$

Therefore

$$\frac{1}{nt_n(1-\tau_n)} \sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i > t_n\} \xrightarrow{\mathbb{P}} c \left(1 + \frac{\gamma}{1-\gamma} \right) = \frac{c}{1-\gamma}$$

as announced. \square

The next auxiliary result is an extension of Theorem 1 in Daouia *et al.* (2018). It drops the assumption of an independent sequence and of an increasing underlying distribution function. We note that the bias term $b(\gamma, \rho)$ of our result below is simpler than the corresponding bias term of Theorem 1 in Daouia *et al.* (2018), due to the assumption of a centred noise variable.

Proposition 1. *Assume that $\mathbb{E}|\varepsilon_-| < \infty$, and that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $0 < \gamma < 1$. Let $\tau_n \uparrow 1$ be such that $n(1-\tau_n) \rightarrow \infty$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \rightarrow \lambda \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}/q_{\tau_n}(\varepsilon) = O(1)$. Then, if*

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

we have

$$\sqrt{n(1-\tau_n)} \left(\frac{\tilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} m(\gamma)\Gamma + \Theta - \lambda b(\gamma, \rho)$$

with $m(\gamma) = (1-\gamma)^{-1} - \log(\gamma^{-1} - 1)$ and

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

Proof of Proposition 1. Note that $(\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}} \xrightarrow{\mathbb{P}} (\gamma^{-1} - 1)^{-\gamma}$ and $\bar{q}_{\tau_n}(\varepsilon)/q_{\tau_n}(\varepsilon) - 1 \xrightarrow{\mathbb{P}} 0$, so that linearising leads to

$$\begin{aligned} \frac{\tilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 &= \left(\frac{(\bar{\gamma}^{-1} - 1)^{-\bar{\gamma}}}{(\gamma^{-1} - 1)^{-\gamma}} - 1 \right) + \left(\frac{\bar{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1 \right) (1 + o_{\mathbb{P}}(1)) \\ &+ \left(\frac{(\gamma^{-1} - 1)^{-\gamma} q_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

To control the bias term, use Proposition 1 in Daouia *et al.* (2020), of which a consequence is, for the centred variable ε ,

$$\sqrt{n(1-\tau_n)} \left(\frac{(\gamma^{-1}-1)^{-\gamma} q_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \rightarrow -\lambda \left[\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} + \frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho} \right] = -\lambda b(\gamma, \rho).$$

Reporting this in (4) and using the delta-method, we obtain

$$\sqrt{n(1-\tau_n)} \left(\frac{\tilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} m(\gamma)\Gamma + \Theta - \lambda b(\gamma, \rho).$$

This is precisely the required result. \square

The following rearrangement lemma is an extension of Lemma 1 in Gardes and Stupfler (2014), which we use in the proof of Lemma 3 below.

Lemma 2. *Let $n \geq 2$ and (a_1, \dots, a_n) and (b_1, \dots, b_n) be two n -tuples of real numbers such that for all $i \in \{1, \dots, n\}$, $a_i \leq b_i$. Then for all $i \in \{1, \dots, n\}$, $a_{i,n} \leq b_{i,n}$.*

Proof of Lemma 2. See the proof of Lemma 1 in Gardes and Stupfler (2014), which, although the original result was stated for n -tuples featuring no ties, carries over to this more general case with no modification. \square

The following lemma is the key to the proof of Theorem 2. In our context, it will be used to give an approximation of the tail empirical quantile process of the errors by the tail empirical quantile process of the residuals.

Lemma 3. *Let $k = k(n) \rightarrow \infty$ be a sequence of integers with $k/n \rightarrow 0$. Assume that ε has an infinite right endpoint. Suppose further that the ε_i are independent copies of ε and that the array of random variables $\widehat{\varepsilon}_i^{(n)}$, $1 \leq i \leq n$, satisfies*

$$\forall i \in \{1, \dots, n\}, |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \leq R_{n,i}(1 + |\varepsilon_i|) \quad \text{where} \quad \max_{1 \leq i \leq n} R_{n,i} \xrightarrow{\mathbb{P}} 0.$$

Then we have both

$$\sup_{0 < s \leq 1} \left| \frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}} - 1 \right| = O_{\mathbb{P}} \left(\max_{1 \leq i \leq n} R_{n,i} \right) \quad \text{and} \quad \sup_{0 < s \leq 1} \left| \log \left(\frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}} \right) \right| = O_{\mathbb{P}} \left(\max_{1 \leq i \leq n} R_{n,i} \right).$$

Proof of Lemma 3. Define $R_n = \max_{1 \leq i \leq n} R_{n,i}$. Clearly:

$$\forall i \in \{1, \dots, n\}, \varepsilon_i - R_n(1 + |\varepsilon_i|) =: \xi_i \leq \widehat{\varepsilon}_i^{(n)} \leq \zeta_i := \varepsilon_i + R_n(1 + |\varepsilon_i|).$$

It then follows from Lemma 2 that

$$\forall i \in \{1, \dots, n\}, \xi_{i,n} \leq \widehat{\varepsilon}_{i,n}^{(n)} \leq \zeta_{i,n}.$$

Note that for any $r \in (-1, 1)$, the function $x \mapsto x + r(1 + |x|)$ is increasing. Therefore, on the event $\{R_n \leq 1/4\}$, whose probability gets arbitrarily high as n increases, we have:

$$\forall i \in \{1, \dots, n\}, \varepsilon_{i,n} - R_n(1 + |\varepsilon_{i,n}|) = \xi_{i,n} \leq \widehat{\varepsilon}_{i,n}^{(n)} \leq \zeta_{i,n} = \varepsilon_{i,n} + R_n(1 + |\varepsilon_{i,n}|).$$

Now, by Lemma 3.2.1 in de Haan and Ferreira (2006) together with the equality $\varepsilon \stackrel{d}{=} U(Z)$ where Z has a unit Pareto distribution, we get $\varepsilon_{n-k,n} \rightarrow +\infty$ almost surely. On the event $A_n := \{R_n \leq 1/4\} \cap \{\varepsilon_{n-k,n} \geq 1\}$, which likewise has probability arbitrarily large, we obtain

$$\forall i \geq n - k, (1 - R_n)\varepsilon_{i,n} - R_n \leq \widehat{\varepsilon}_{i,n}^{(n)} \leq (1 + R_n)\varepsilon_{i,n} + R_n.$$

In other words, on A_n ,

$$\forall s \in (0, 1], -2R_n \leq -R_n \left(1 + \frac{1}{\varepsilon_{n-\lfloor ks \rfloor, n}}\right) \leq \frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}} - 1 \leq R_n \left(1 + \frac{1}{\varepsilon_{n-\lfloor ks \rfloor, n}}\right) \leq 2R_n.$$

This shows that

$$\sup_{0 < s \leq 1} \left| \frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}} - 1 \right| = O_{\mathbb{P}}(R_n).$$

Note further that, on A_n ,

$$\forall s \in (0, 1], \log(1 - 2R_n) \leq \log\left(\frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}}\right) \leq \log(1 + 2R_n).$$

Since $\log(1 + x) \leq x$ and $\log(1 - x) \geq -2x$ for all $x \in [0, 1/2]$, this yields, on A_n ,

$$\forall s \in (0, 1], \left| \log\left(\frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}}\right) \right| \leq 4R_n.$$

As a consequence,

$$\sup_{0 < s \leq 1} \left| \log\left(\frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\varepsilon_{n-\lfloor ks \rfloor, n}}\right) \right| = O_{\mathbb{P}}(R_n).$$

This concludes the proof. □

A.2 Worked-out examples

Lemma 4 gives the rate of convergence of the weighted least squares estimators in model (M_1) . Here and throughout all $O_{\mathbb{P}}(1)$ statements are meant componentwise.

Lemma 4. *Assume that (\mathbf{X}_i, Y_i) are independent random pairs generated from model (M_1) . Suppose further that $\mathbb{E}(\varepsilon^2) < \infty$. Then we have*

$$\sqrt{n}(\widehat{\alpha} - \alpha) = O_{\mathbb{P}}(1), \quad \sqrt{n}(\widehat{\beta} - \beta) = O_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{n}(\widehat{\theta} - \theta) = O_{\mathbb{P}}(1).$$

Proof of Lemma 4. We introduce the notation

$$\mathbf{x} = \begin{pmatrix} 1 & \mathbf{X}_1^\top \\ \vdots & \vdots \\ 1 & \mathbf{X}_n^\top \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } \mathbf{\Omega} = \text{diag}([1 + \boldsymbol{\theta}^\top \mathbf{X}_1]^2, \dots, [1 + \boldsymbol{\theta}^\top \mathbf{X}_n]^2).$$

A preliminary step is to remark that for any $\mathbf{a} = (a_0, a_1, \dots, a_d)^\top \in \mathbb{R}^{d+1}$,

$$\begin{aligned} \mathbf{a}^\top \mathbf{x}^\top \mathbf{x} \mathbf{a} &= \sum_{i=1}^n [a_0 + (a_1, \dots, a_d) \mathbf{X}_i]^2 > 0 \\ \text{and } \mathbf{a}^\top \mathbf{x}^\top \mathbf{\Omega}^{-1} \mathbf{x} \mathbf{a} &= \sum_{i=1}^n [1 + \boldsymbol{\theta}^\top \mathbf{X}_i]^{-2} [a_0 + (a_1, \dots, a_d) \mathbf{X}_i]^2 > 0 \end{aligned}$$

with probability 1, because \mathbf{X} has a continuous distribution (and as such, does not put mass on affine hyperplanes of \mathbb{R}^d). The symmetric matrices $\mathbf{x}^\top \mathbf{x}$ and $\mathbf{x}^\top \mathbf{\Omega}^{-1} \mathbf{x}$ therefore have full rank with probability 1. Since, by the law of large numbers,

$$\frac{1}{n} [\mathbf{x}^\top \mathbf{x}]_{i+1, j+1} \xrightarrow{\mathbb{P}} \mathbb{E}(X_i X_j) \text{ and } \frac{1}{n} [\mathbf{x}^\top \mathbf{\Omega}^{-1} \mathbf{x}]_{i+1, j+1} \xrightarrow{\mathbb{P}} \mathbb{E}([1 + \boldsymbol{\theta}^\top \mathbf{X}]^{-2} X_i X_j)$$

(where $X_0 = 1$ for notational convenience), the same argument shows that $\mathbf{x}^\top \mathbf{x}/n$ and $\mathbf{x}^\top \mathbf{\Omega}^{-1} \mathbf{x}/n$ converge in probability to symmetric positive definite matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ say.

Our first step is to show that the preliminary estimators $\tilde{\alpha}$, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\theta}}$ are \sqrt{n} -consistent. Rewrite model (M_1) for the available data as

$$\mathbf{Y} = \mathbf{x} \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}^\top = (\varepsilon_1, \dots, \varepsilon_n)$ and \circ denotes the Hadamard (entrywise) product of matrices. By standard least squares theory,

$$\begin{pmatrix} \tilde{\alpha} \\ \tilde{\boldsymbol{\beta}} \end{pmatrix} = (\mathbf{x}^\top \mathbf{x})^{-1} \mathbf{x}^\top \mathbf{Y}.$$

A direct calculation then yields

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\tilde{\alpha} - \alpha) \\ \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix} &= n (\mathbf{x}^\top \mathbf{x})^{-1} \times \frac{1}{\sqrt{n}} \mathbf{x}^\top \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} \\ &= n (\mathbf{x}^\top \mathbf{x})^{-1} \times \begin{pmatrix} n^{-1/2} \sum_{i=1}^n [1 + \boldsymbol{\theta}^\top \mathbf{X}_i] \varepsilon_i \\ n^{-1/2} \sum_{i=1}^n [1 + \boldsymbol{\theta}^\top \mathbf{X}_i] X_{i1} \varepsilon_i \\ \vdots \\ n^{-1/2} \sum_{i=1}^n [1 + \boldsymbol{\theta}^\top \mathbf{X}_i] X_{id} \varepsilon_i \end{pmatrix}. \end{aligned}$$

Set for notational convenience $X_{i0} = 1$. Since, for any $m \in \{0, 1, \dots, d\}$, the random variables $[1 + \boldsymbol{\theta}^\top \mathbf{X}_i] X_{im} \varepsilon_i$, $1 \leq i \leq n$, are independent, centred and square-integrable, the standard multivariate central limit theorem combined with the convergence $n (\mathbf{x}^\top \mathbf{x})^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_1^{-1}$ yields

$$\sqrt{n}(\tilde{\alpha} - \alpha) = O_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_{\mathbb{P}}(1). \quad (4)$$

We then prove that $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_{\mathbb{P}}(1)$. Recalling that

$$\tilde{\boldsymbol{\theta}} = \frac{\tilde{\boldsymbol{\nu}}}{\tilde{\mu}} \quad \text{and} \quad \boldsymbol{\theta} = \frac{\boldsymbol{\nu}}{\mu}$$

where $\mu = \mathbb{E}|\varepsilon| > 0$ and $\boldsymbol{\nu} = \mu \boldsymbol{\theta}$, it is enough to show that $\sqrt{n}(\tilde{\mu} - \mu) = O_{\mathbb{P}}(1)$ and $\sqrt{n}(\tilde{\boldsymbol{\nu}} - \boldsymbol{\nu}) = O_{\mathbb{P}}(1)$. Defining

$$\mathbf{Z} = \begin{pmatrix} |Y_1 - (\alpha + \boldsymbol{\beta}^\top \mathbf{X}_1)| \\ \vdots \\ |Y_n - (\alpha + \boldsymbol{\beta}^\top \mathbf{X}_n)| \end{pmatrix} = \mathbf{x} \begin{pmatrix} \mu \\ \boldsymbol{\nu} \end{pmatrix} + \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \mathbf{e},$$

where $\mathbf{e}^\top = (|\varepsilon_1| - \mathbb{E}|\varepsilon|, \dots, |\varepsilon_n| - \mathbb{E}|\varepsilon|)$, and defining then $\tilde{\mathbf{Z}}$ in the obvious way, we have

$$\begin{pmatrix} \tilde{\mu} \\ \tilde{\boldsymbol{\nu}} \end{pmatrix} = (\mathbf{x}^\top \mathbf{x})^{-1} \mathbf{x}^\top \tilde{\mathbf{Z}}.$$

We therefore obtain

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\tilde{\mu} - \mu) \\ \sqrt{n}(\tilde{\boldsymbol{\nu}} - \boldsymbol{\nu}) \end{pmatrix} &= n (\mathbf{x}^\top \mathbf{x})^{-1} \times \frac{1}{\sqrt{n}} \mathbf{x}^\top \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \mathbf{e} \right\} \\ &+ n (\mathbf{x}^\top \mathbf{x})^{-1} \times \mathbf{x}^\top \left(\frac{1}{\sqrt{n}} [\tilde{\mathbf{Z}} - \mathbf{Z}] \right). \end{aligned} \quad (5)$$

Since $e = |\varepsilon| - \mathbb{E}|\varepsilon|$ is independent of \mathbf{X} and has a finite variance, repeating the proof of (4) gives

$$n (\mathbf{x}^\top \mathbf{x})^{-1} \times \frac{1}{\sqrt{n}} \mathbf{x}^\top \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \mathbf{e} \right\} = O_{\mathbb{P}}(1). \quad (6)$$

Furthermore,

$$\mathbf{x}^\top \left(\frac{1}{\sqrt{n}} [\tilde{\mathbf{Z}} - \mathbf{Z}] \right) = \begin{pmatrix} n^{-1/2} \sum_{i=1}^n [\tilde{Z}_i - Z_i] \\ n^{-1/2} \sum_{i=1}^n X_{i1} [\tilde{Z}_i - Z_i] \\ \vdots \\ n^{-1/2} \sum_{i=1}^n X_{id} [\tilde{Z}_i - Z_i] \end{pmatrix}.$$

Recalling that \mathbf{X} lies in a compact set, we find that for any $m \in \{0, 1, \dots, d\}$,

$$\left| n^{-1/2} \sum_{i=1}^n X_{im} [\tilde{Z}_i - Z_i] \right| = O_{\mathbb{P}} \left(\sqrt{n} \max_{1 \leq i \leq n} |(\tilde{\alpha} - \alpha) + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}_i| \right) = O_{\mathbb{P}}(1)$$

by (4). Combining this with (5), (6) and the convergence $n(\mathbf{x}^\top \mathbf{x})^{-1} \xrightarrow{\mathbb{P}} \Sigma_1^{-1}$, we get indeed $\sqrt{n}(\tilde{\mu} - \mu) = O_{\mathbb{P}}(1)$ and $\sqrt{n}(\tilde{\nu} - \nu) = O_{\mathbb{P}}(1)$ and thus

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_{\mathbb{P}}(1). \quad (7)$$

We are now ready to prove the convergence of the weighted estimators $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$. By standard weighted least squares theory,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} \end{pmatrix} = (\mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \mathbf{x})^{-1} \mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \mathbf{Y}.$$

It follows that

$$\begin{pmatrix} \sqrt{n}(\hat{\alpha} - \alpha) \\ \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix} = n(\mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \mathbf{x})^{-1} \times \frac{1}{\sqrt{n}} \mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} \quad (8)$$

where $\tilde{\boldsymbol{\Omega}}$ is obtained from $\boldsymbol{\Omega}$ in the obvious manner. Note that for any $i, j \in \{0, \dots, d\}$,

$$\frac{1}{n} [\mathbf{x}^\top \boldsymbol{\Omega}^{-1} \mathbf{x}]_{i+1, j+1} = \frac{1}{n} \sum_{k=1}^n \frac{X_{ki} X_{kj}}{[1 + \boldsymbol{\theta}^\top \mathbf{X}_k]^2} \text{ and } \frac{1}{n} [\mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \mathbf{x}]_{i+1, j+1} = \frac{1}{n} \sum_{k=1}^n \frac{X_{ki} X_{kj}}{[1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_k]^2}.$$

Recalling once again that \mathbf{X} lies in a compact set, that $1 + \boldsymbol{\theta}^\top \mathbf{X}$ is bounded from below by a positive constant, and (7), we find, by the law of large numbers,

$$\frac{1}{n} [\mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \mathbf{x}] - \frac{1}{n} [\mathbf{x}^\top \boldsymbol{\Omega}^{-1} \mathbf{x}] \xrightarrow{\mathbb{P}} 0 \text{ and thus } n(\mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \mathbf{x})^{-1} \xrightarrow{\mathbb{P}} \Sigma_2^{-1}. \quad (9)$$

Besides, for any $m \in \{0, 1, \dots, d\}$,

$$\begin{aligned} & \left[\frac{1}{\sqrt{n}} \mathbf{x}^\top \tilde{\boldsymbol{\Omega}}^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} - \frac{1}{\sqrt{n}} \mathbf{x}^\top \boldsymbol{\Omega}^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} \right]_{m+1} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [1 + \boldsymbol{\theta}^\top \mathbf{X}_i] X_{im} \varepsilon_i \left[\frac{1}{[1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^2} - \frac{1}{[1 + \boldsymbol{\theta}^\top \mathbf{X}_i]^2} \right] \\ &= -\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \left\{ \frac{1}{n} \sum_{i=1}^n X_{im} \varepsilon_i \frac{2 + \boldsymbol{\theta}^\top \mathbf{X}_i + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i}{[1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^2 [1 + \boldsymbol{\theta}^\top \mathbf{X}_i]} \mathbf{X}_i \right\}. \end{aligned}$$

Using again the properties of \mathbf{X} and (7), some straightforward algebra yields that

$$R_n := \sqrt{n} \max_{1 \leq i \leq n} \left| \frac{2 + \boldsymbol{\theta}^\top \mathbf{X}_i + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i}{[1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^2} - \frac{2}{1 + \boldsymbol{\theta}^\top \mathbf{X}_i} \right| = O_{\mathbb{P}}(1).$$

Conclude that

$$\begin{aligned} & \left[\frac{1}{\sqrt{n}} \mathbf{x}^\top \tilde{\Omega}^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} - \frac{1}{\sqrt{n}} \mathbf{x}^\top \Omega^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} \right]_{m+1} \\ &= -2\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \left\{ \frac{1}{n} \sum_{i=1}^n X_{im} \varepsilon_i \left[\frac{2}{[1 + \boldsymbol{\theta}^\top \mathbf{X}_i]^2} + O_{\mathbb{P}}\left(\frac{R_n}{\sqrt{n}}\right) \right] \mathbf{X}_i \right\}. \end{aligned}$$

Since $\boldsymbol{\varepsilon}$ is centred and independent of \mathbf{X} , we may combine the properties of \mathbf{X} and (7) with the law of large numbers to get

$$\frac{1}{\sqrt{n}} \mathbf{x}^\top \tilde{\Omega}^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} - \frac{1}{\sqrt{n}} \mathbf{x}^\top \Omega^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} = o_{\mathbb{P}}(1). \quad (10)$$

Now clearly

$$\left[\frac{1}{\sqrt{n}} \mathbf{x}^\top \Omega^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} \right]_{m+1} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_{im}}{1 + \boldsymbol{\theta}^\top \mathbf{X}_i} \varepsilon_i$$

so that, by the standard multivariate central limit theorem,

$$\frac{1}{\sqrt{n}} \mathbf{x}^\top \Omega^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \boldsymbol{\varepsilon} \right\} = O_{\mathbb{P}}(1). \quad (11)$$

Combining (8), (9), (10) and (11) results in

$$\sqrt{n}(\hat{\alpha} - \alpha) = O_{\mathbb{P}}(1) \quad \text{and} \quad \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_{\mathbb{P}}(1).$$

We complete the proof by showing that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_{\mathbb{P}}(1)$. It is again enough to show that $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = O_{\mathbb{P}}(1)$ and $\sqrt{n}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) = O_{\mathbb{P}}(1)$. Write

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\ \sqrt{n}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) \end{pmatrix} &= n \left(\mathbf{x}^\top \tilde{\Omega}^{-1} \mathbf{x} \right)^{-1} \times \frac{1}{\sqrt{n}} \mathbf{x}^\top \tilde{\Omega}^{-1} \left\{ \left[\mathbf{x} \begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix} \right] \circ \mathbf{e} \right\} \\ &\quad + n \left(\mathbf{x}^\top \tilde{\Omega}^{-1} \mathbf{x} \right)^{-1} \times \mathbf{x}^\top \tilde{\Omega}^{-1} \left(\frac{1}{\sqrt{n}} [\hat{\mathbf{Z}} - \mathbf{Z}] \right). \end{aligned}$$

Furthermore,

$$\mathbf{x}^\top \tilde{\Omega}^{-1} \left(\frac{1}{\sqrt{n}} [\hat{\mathbf{Z}} - \mathbf{Z}] \right) = \begin{pmatrix} n^{-1/2} \sum_{i=1}^n [1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^{-2} [\hat{Z}_i - Z_i] \\ n^{-1/2} \sum_{i=1}^n X_{i1} [1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^{-2} [\hat{Z}_i - Z_i] \\ \vdots \\ n^{-1/2} \sum_{i=1}^n X_{id} [1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^{-2} [\hat{Z}_i - Z_i] \end{pmatrix}.$$

Recalling the properties of \mathbf{X} and the \sqrt{n} -convergence of $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\theta}}$, we find that for any $m \in \{0, 1, \dots, d\}$,

$$\left| n^{-1/2} \sum_{i=1}^n X_{im} [1 + \tilde{\boldsymbol{\theta}}^\top \mathbf{X}_i]^{-2} [\hat{Z}_i - Z_i] \right| = O_{\mathbb{P}} \left(\sqrt{n} \max_{1 \leq i \leq n} |(\hat{\alpha} - \alpha) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}_i| \right) = O_{\mathbb{P}}(1).$$

Combining this with (9) and straightforward adaptations of (10) and (11) with \mathbf{e} in place of $\boldsymbol{\varepsilon}$, we find $\sqrt{n}(\widehat{\mu} - \mu) = O_{\mathbb{P}}(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\nu}} - \boldsymbol{\nu}) = O_{\mathbb{P}}(1)$ as required. \square

Lemma 5 is a general uniform consistency result which is useful for the analysis of the single-index model (M_2) .

Lemma 5. *Assume that $(\mathcal{X}_i, \mathcal{Y}_i)_{i \geq 1}$ are independent copies of a bivariate random pair $(\mathcal{X}, \mathcal{Y})$ such that:*

- \mathcal{X} has support $[a, b]$, with $a < b$, and a density function $f_{\mathcal{X}}$ which is uniformly bounded on compact sub-intervals of (a, b) .
- There exists $\delta > 0$ such that $\mathbb{E}|\mathcal{Y}|^{2+\delta} < \infty$ and the conditional moment function $z \mapsto \mathbb{E}[|\mathcal{Y}|^{2+\delta} | \mathcal{X} = z]$ is uniformly bounded on compact sub-intervals of (a, b) .

Let further:

- (\mathcal{V}_i) be a sequence of independent copies of a bounded random variable \mathcal{V} .
- L be a Lipschitz continuous function with support contained in $[-1, 1]$.

Assume finally that $nh_n^5 \rightarrow c \in (0, \infty)$, and $t_n = n^t$ with $2/(5+\delta) < t < 2/5$. Then for any $a_1, b_1 \in [a, b]$ with $a < a_1 < b_1 < b$,

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathbb{1}\{|\mathcal{Y}_i| \leq t_n\} \mathcal{V}_i L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y} \mathcal{V} L\left(\frac{z - \mathcal{X}}{h_n}\right)\right] \right| = O_{\mathbb{P}}(1).$$

We note that, as a consequence, we have a similar uniform consistency result for the non-truncated version of the smoothed empirical moment, that is

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathcal{V}_i L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y} \mathcal{V} L\left(\frac{z - \mathcal{X}}{h_n}\right)\right] \right| = O_{\mathbb{P}}(1)$$

under the further assumption $\mathbb{E}|\mathcal{Y}|^{5/2+\delta} < \infty$. This follows from noting that

$$\mathbb{P}\left(\bigcup_{i=1}^n \{|\mathcal{Y}_i| > t_n\}\right) \leq n\mathbb{P}(|\mathcal{Y}| > t_n) = O\left(\frac{n}{t_n^{5/2+\delta}}\right) = O\left(n^{1-(5+2\delta)/(5+\delta)}\right) = o(1)$$

by Markov's inequality. The stronger moment assumption $\mathbb{E}|\mathcal{Y}|^{5/2+\delta} < \infty$ already appears in Masry (1996) in the context of local polynomial estimation.

Proof of Lemma 5. The basic idea is to control the oscillation of the random function

$$z \mapsto \frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathbb{1}\{|\mathcal{Y}_i| \leq t_n\} \mathcal{V}_i L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y} \mathcal{V} L\left(\frac{z - \mathcal{X}}{h_n}\right)\right] \right|$$

and then use this control to prove that it is sufficient to show uniform consistency over a fine grid instead, which can be done by using Bernstein's exponential inequality. Our proof adapts the method of Hansen (2008, proof of Theorem 2).

Define $Y_i^{(n)} := \mathcal{Y}_i \mathcal{V}_i \mathbb{1}\{|\mathcal{Y}_i| \leq t_n\}$ and $Y^{(n)} := \mathcal{Y} \mathcal{V} \mathbb{1}\{|\mathcal{Y}| \leq t_n\}$. Then

$$\begin{aligned} & \frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathbb{1}\{|\mathcal{Y}_i| \leq t_n\} \mathcal{V}_i L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[\mathcal{Y} \mathcal{V} L\left(\frac{z - \mathcal{X}}{h_n}\right) \right] \right| \\ & \leq \frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z - \mathcal{X}}{h_n}\right) \right] \right| \\ & + \frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \frac{1}{h_n} \mathbb{E} \left[|\mathcal{Y}| |\mathcal{V}| \mathbb{1}\{|\mathcal{Y}| > t_n\} \left| L\left(\frac{z - \mathcal{X}}{h_n}\right) \right| \right]. \end{aligned} \quad (12)$$

The second term on the right-hand side of (12) is controlled by noting that, thanks to a change of variables,

$$\begin{aligned} & \frac{1}{h_n} \mathbb{E} \left[|\mathcal{Y}| |\mathcal{V}| \mathbb{1}\{|\mathcal{Y}| > t_n\} \left| L\left(\frac{z - \mathcal{X}}{h_n}\right) \right| \right] \\ & = \mathcal{O} \left(\int_{-1}^1 \mathbb{E} [|\mathcal{Y}| \mathbb{1}\{|\mathcal{Y}| > t_n\} | \mathcal{X} = z - h_n u] |L(u)| f_{\mathcal{X}}(z - h_n u) du \right) \\ & = \mathcal{O} \left(t_n^{-1-\delta} \int_{-1}^1 \mathbb{E} [|\mathcal{Y}|^{2+\delta} | \mathcal{X} = z - h_n u] |L(u)| f_{\mathcal{X}}(z - h_n u) du \right) = \mathcal{O}(t_n^{-1-\delta}) \end{aligned}$$

uniformly in $z \in [a_1, b_1]$. Here the boundedness of \mathcal{V} , the integrability of $|L|$ and the assumption that the $(2 + \delta)$ -conditional moment of \mathcal{Y} and the density function $f_{\mathcal{X}}$ are uniformly bounded on compact sub-intervals of (a, b) were all used. Finally

$$t_n^{-1-\delta} = n^{-(1+\delta)t} = o(n^{-2/5}) = o\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)$$

so that

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \frac{1}{h_n} \mathbb{E} \left[|\mathcal{Y}| |\mathcal{V}| \mathbb{1}\{|\mathcal{Y}| > t_n\} \left| L\left(\frac{z - \mathcal{X}}{h_n}\right) \right| \right] = o(1). \quad (13)$$

Combining (12) and (13), we find that it is sufficient to show that

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z - \mathcal{X}}{h_n}\right) \right] \right| = \mathcal{O}_{\mathbb{P}}(1). \quad (14)$$

We now replace the supremum in (14) by a supremum over a grid by focusing on the oscillation of the left-hand side. For a given $z \in \mathbb{R}$, let

$$A_n(z) := \left\{ z' \in [a_1, b_1] \left| |z' - z| \leq h_n \frac{\sqrt{\log n}}{n^{2/5}} \right. \right\}.$$

Then $[a_1, b_1]$ is covered by the $A_n(z_{n,j})$, with

$$z_{n,j} = a_1 + j h_n \frac{\sqrt{\log n}}{n^{2/5}}, \quad j = 1, \dots, \left\lfloor \frac{b_1 - a_1}{h_n \frac{\sqrt{\log n}}{n^{2/5}}} \right\rfloor =: N_n,$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Besides, writing $|L(z') - L(z)| \leq C_L |z' - z|$ by Lipschitz continuity of L , we also find

$$|z' - z| \leq 1 \Rightarrow |L(z') - L(z)| \leq |z' - z| \mathcal{L}(z) \quad \text{with } \mathcal{L}(z) := C_L \mathbb{1}\{|z| \leq 2\}.$$

Let now $z_{n,j}$ be a grid point and $z \in A_n(z_{n,j})$. By construction $|z - z_{n,j}|/h_n \leq \sqrt{\log(n)}/n^{2/5}$ which converges to 0, so that, for n large enough,

$$\forall i \in \{1, \dots, n\}, \quad \left| L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - L\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) \right| \leq \frac{\sqrt{\log n}}{n^{2/5}} \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right).$$

Then

$$\begin{aligned} & \frac{n^{2/5}}{\sqrt{\log n}} \sup_{z \in A_n(z_{n,j})} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z - \mathcal{X}}{h_n}\right) \right] \right| \\ & \leq \frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right] \right| \\ & + \frac{1}{nh_n} \sum_{i=1}^n |Y_i^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) + \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right] \\ & \leq \frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right] \right| \\ & + \left| \frac{1}{nh_n} \sum_{i=1}^n |Y_i^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right] \right| \\ & + 2 \times \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right]. \end{aligned}$$

By the boundedness of \mathcal{V} , of $f_{\mathcal{X}}$ and of $z \mapsto \mathbb{E}[|Y| | \mathcal{X} = z]$ over compact sub-intervals of (a, b) , we find, for n large enough,

$$\sup_{a_1 \leq z \leq b_1} \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L}\left(\frac{z - \mathcal{X}}{h_n}\right) \right] \leq C_0$$

where C_0 is a finite constant. Consequently, for any constant $C > 2C_0$,

$$\begin{aligned} & \frac{n^{2/5}}{\sqrt{\log n}} \sup_{z \in A_n(z_{n,j})} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z - \mathcal{X}}{h_n}\right) \right] \right| \\ & \leq \frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right] \right| \\ & + \frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n |Y_i^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L}\left(\frac{z_{n,j} - \mathcal{X}}{h_n}\right) \right] \right| + C \end{aligned}$$

where the (crude) inequality $n^{2/5}/\sqrt{\log n} \geq 1$, for n large enough, was used. Conclude, by writing $[a_1, b_1] \subset \cup_{1 \leq j \leq N_n} A_n(z_{n,j})$, that

$$\begin{aligned} & \mathbb{P} \left(\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right| > 3C \right) \\ & \leq N_n \max_{1 \leq j \leq N_n} \mathbb{P} \left(\frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L \left(\frac{z_{n,j} - \mathcal{X}_i}{h_n} \right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L \left(\frac{z_{n,j} - \mathcal{X}}{h_n} \right) \right] \right| > C \right) \\ & + N_n \max_{1 \leq j \leq N_n} \mathbb{P} \left(\frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n |Y_i^{(n)}| \mathcal{L} \left(\frac{z_{n,j} - \mathcal{X}_i}{h_n} \right) - \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L} \left(\frac{z_{n,j} - \mathcal{X}}{h_n} \right) \right] \right| > C \right). \end{aligned}$$

We finish the proof by showing

$$\mathbb{P} \left(\frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right| > C \right) = o\left(\frac{1}{n}\right) \quad (15)$$

and

$$\mathbb{P} \left(\frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n |Y_i^{(n)}| \mathcal{L} \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \frac{1}{h_n} \mathbb{E} \left[|Y^{(n)}| \mathcal{L} \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right| > C \right) = o\left(\frac{1}{n}\right) \quad (16)$$

for C large enough, uniformly in $z \in [a_1, b_1]$. Since N_n is of order $n^{2/5}/(h_n \sqrt{\log(n)}) \approx n^{3/5}/\sqrt{\log(n)} = o(n)$, this will entail

$$\mathbb{P} \left(\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \leq z \leq b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \frac{1}{h_n} \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right| > 3C \right) = o(1)$$

for C large enough, which is sufficient for our purposes. We only show (15) uniformly in $z \in [a_1, b_1]$; the proof of (16) is identical. Rewrite the left-hand side of (15) as

$$\mathbb{P} \left(\left| \sum_{i=1}^n \left\{ Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right\} \right| > C u_n \right), \quad \text{with } u_n := n^{3/5} h_n \sqrt{\log n}.$$

Let v be a constant such that $|\mathcal{V}| \leq v$ almost surely. Note that for any i we have the crude bound

$$\left| Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right| \leq 2v t_n \max_{-1 \leq u \leq 1} |L(u)|.$$

Remark also that, for n large enough,

$$\text{Var} \left(Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right) \leq v^2 \mathbb{E} \left[\mathcal{Y}^2 L^2 \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \leq D h_n$$

for some finite constant D , by uniform boundedness of $f_{\mathcal{X}}$ and $z \mapsto \mathbb{E}[\mathcal{Y}^2 | \mathcal{X} = z]$ over compact sub-intervals of (a, b) . By the Bernstein exponential inequality we get

$$\begin{aligned} & \mathbb{P} \left(\left| \sum_{i=1}^n \left\{ Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right\} \right| > C u_n \right) \\ & \leq 2 \exp \left(- \frac{C^2 u_n^2 / 2}{D n h_n + 2 C v t_n u_n \max_{[-1,1]} |L| / 3} \right). \end{aligned}$$

Recalling that $t_n = n^t$ with $2/(5 + \delta) < t < 2/5$, $u_n = n^{3/5} h_n \sqrt{\log n}$ and $n h_n^5 \rightarrow c \in (0, \infty)$, one finds

$$\frac{1}{\log n} \times \frac{C^2 u_n^2 / 2}{D n h_n + 2 C v t_n u_n \max_{[-1,1]} |L| / 3} \rightarrow \frac{c^{1/5} C^2}{2D} \text{ as } n \rightarrow \infty$$

and therefore there is a constant $C' > 0$, independent of C , such that for n large enough

$$\mathbb{P} \left(\left| \sum_{i=1}^n \left\{ Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right\} \right| > C u_n \right) \leq 2 \exp(-C' C^2 \log n)$$

uniformly in $z \in [a_1, b_1]$. For C large enough, this yields

$$\mathbb{P} \left(\left| \sum_{i=1}^n \left\{ Y_i^{(n)} L \left(\frac{z - \mathcal{X}_i}{h_n} \right) - \mathbb{E} \left[Y^{(n)} L \left(\frac{z - \mathcal{X}}{h_n} \right) \right] \right\} \right| > C u_n \right) = \mathcal{O} \left(\frac{1}{n} \right)$$

which is equivalent to (15). This completes the proof. \square

Lemma 6 provides a uniform control, tailored to the assumptions of Proposition 2, of the gap between smoothed moments and their asymptotic equivalents.

Lemma 6. *Assume that the bivariate random pair $(\mathcal{X}, \mathcal{Y})$ is such that:*

- \mathcal{X} has support $[a, b]$, with $a < b$, and a density function $f_{\mathcal{X}}$ which has a continuous derivative on (a, b) .
- The conditional moment function $m_{\mathcal{Y}|\mathcal{X}} : z \mapsto \mathbb{E}(\mathcal{Y} | \mathcal{X} = z)$ is well-defined and has a continuous derivative on (a, b) .
- L is a bounded measurable function with support contained in $[-1, 1]$.

Then, as $h \rightarrow 0$:

(i) For any $a_1, b_1 \in [a, b]$ with $a < a_1 < b_1 < b$, we have, uniformly in $z \in [a_1, b_1]$,

$$\begin{aligned} & \frac{1}{h} \mathbb{E} \left[\mathcal{Y} L \left(\frac{z - \mathcal{X}}{h} \right) \right] \\ & = m_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) \int_{-1}^1 L(u) du - h \{ m'_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) + m_{\mathcal{Y}|\mathcal{X}}(z) f'_{\mathcal{X}}(z) \} \int_{-1}^1 u L(u) du + o(h). \end{aligned}$$

(ii) If moreover $f_{\mathcal{X}}$ and $m_{\mathcal{Y}|\mathcal{X}}$ are twice continuously differentiable on (a, b) then, uniformly in $z \in [a_1, b_1]$,

$$\begin{aligned} & \frac{1}{h} \mathbb{E} \left[\mathcal{Y} L \left(\frac{z - \mathcal{X}}{h} \right) \right] \\ &= m_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) \int_{-1}^1 L(u) du - h \{ m'_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) + m_{\mathcal{Y}|\mathcal{X}}(z) f'_{\mathcal{X}}(z) \} \int_{-1}^1 u L(u) du \\ &+ \frac{h^2}{2} \{ m''_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) + 2m'_{\mathcal{Y}|\mathcal{X}}(z) f'_{\mathcal{X}}(z) + m_{\mathcal{Y}|\mathcal{X}}(z) f''_{\mathcal{X}}(z) \} \int_{-1}^1 u^2 L(u) du + o(h^2). \end{aligned}$$

Proof of Lemma 6. Note that

$$\frac{1}{h} \mathbb{E} \left[\mathcal{Y} L \left(\frac{z - \mathcal{X}}{h} \right) \right] = \int_{-1}^1 m_{\mathcal{Y}|\mathcal{X}}(z - hu) f_{\mathcal{X}}(z - hu) L(u) du.$$

Parts (i) and (ii) are obtained by using the following Taylor formulae with integral remainder:

$$\varphi(z + \delta) = \varphi(z) + \delta \varphi'(z) + \int_z^{z+\delta} [\varphi'(t) - \varphi'(z)] dt$$

and

$$\varphi(z + \delta) = \varphi(z) + \delta \varphi'(z) + \frac{\delta^2}{2} \varphi''(z) + \int_z^{z+\delta} (z + \delta - t) [\varphi''(t) - \varphi''(z)] dt$$

applied to the function $\varphi : z \mapsto m_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z)$. To get a uniform control of the remainders, use the fact that this function has uniformly continuous derivatives on any compact sub-interval of $[a, b]$, by Heine's theorem. \square

Our final auxiliary result is the uniform consistency (with rate) of the estimators of g and σ in the heteroscedastic single-index model of Section 3.2.

Proposition 2. *Assume that $(\mathbf{X}_i, Y_i)_{i \geq 1}$ are independent random pairs generated from the single-index model (M_2) . Assume further that:*

- *The functions g and $\sigma > 0$ are continuous on K_{β} and twice continuously differentiable on the interior K_{β}° of K_{β} .*
- *The projection $\beta^{\top} \mathbf{X}$ has a density function $f_{\beta^{\top} \mathbf{X}}$ which is twice continuously differentiable and positive on K_{β}° .*
- *Each of the conditional moment functions $z \mapsto \mathbb{E}(X_j | \beta^{\top} \mathbf{X} = z)$, $j \in \{1, \dots, d\}$ is continuously differentiable on K_{β}° .*
- *There is $\delta > 0$ such that $\mathbb{E}|\varepsilon|^{2+\delta} < \infty$.*
- *L is a twice continuously differentiable and symmetric probability density function with support contained in $[-1, 1]$.*

Assume also that $nh_n^5 \rightarrow c \in (0, \infty)$, and $t_n = n^t$ with $2/(5 + \delta) < t < 2/5$. Then, for any compact subset K_0 of K° and any estimator $\widehat{\boldsymbol{\beta}}$ such that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_{\mathbb{P}}(1)$, we have

$$\begin{aligned} \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x}) \right| &= O_{\mathbb{P}}(1) \\ \text{and } \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \right| &= O_{\mathbb{P}}(1). \end{aligned}$$

Before proving this result, note that when K is convex, its projection $K_{\boldsymbol{\beta}} := \{\boldsymbol{\beta}^\top \mathbf{x}, \mathbf{x} \in K\}$, which is also the support of $\boldsymbol{\beta}^\top \mathbf{X}$, is a compact interval containing at least two points (because K has a nonempty interior). Note also that Proposition 2 is tailored to our framework in the sense that the assumption $\mathbb{E}|\varepsilon|^{2+\delta} < \infty$, which puts a constraint on the tail heaviness of the noise variable, is intuitively close to minimal for the estimation of g and σ by estimators of Nadaraya-Watson type. An inspection of the proof reveals that a similar theorem holds if \widehat{g}_{h_n, t_n} and $\widehat{\sigma}_{h_n, t_n}$ are replaced by non-truncated versions, under the stronger moment assumption $\mathbb{E}|\varepsilon|^{5/2+\delta} < \infty$; see the comment below the statement of Lemma 5. The regularity assumption on $z \mapsto \mathbb{E}(X_j | \boldsymbol{\beta}^\top \mathbf{X} = z)$ is a technical requirement, which is for instance satisfied if the density function $f_{\mathbf{X}}$ is continuously differentiable and positive on K° .

Proof of Proposition 2. We start by proving the assertion on \widehat{g}_{h_n, t_n} . Define a truncated pseudo-Nadaraya-Watson estimator by

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

The idea is to write

$$\begin{aligned} \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x}) \right| &\leq \left| g(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x}) \right| \\ &\quad + \left| \widetilde{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - g(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) \right| \\ &\quad + \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \widetilde{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) \right| \end{aligned} \quad (17)$$

and control each term on the right-hand side of (17) separately. To control the first term, we first apply the mean value theorem:

$$\left| g(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x}) \right| \leq \left| (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{x} \right| \times \sup_{\lambda \in [0, 1]} \left| g'(\boldsymbol{\beta}^\top \mathbf{x} + \lambda (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{x}) \right|.$$

Since $K_0 \subset K^\circ$, the distance between the compact set K_0 and the (compact) topological boundary of K is positive, *i.e.* $\rho := \inf\{\|x - y\|, x \in K_0, y \in K \setminus K^\circ\} > 0$. It is then straightforward to show that, letting $K_{\boldsymbol{\beta}} = [u, v]$, we have $\boldsymbol{\beta}^\top \mathbf{x} \in [u + \rho/2, v - \rho/2]$ for any $\mathbf{x} \in K_0$. Since $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, we obtain that, with arbitrarily large probability as $n \rightarrow \infty$,

$$\forall \lambda \in [0, 1], \forall \mathbf{x} \in K_0, \boldsymbol{\beta}^\top \mathbf{x} + \lambda (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{x} \in [u + \rho/4, v - \rho/4]. \quad (18)$$

Because g' is continuous and therefore bounded on compact intervals contained in (u, v) , this gives

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| g\left(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}\right) - g\left(\boldsymbol{\beta}^\top \mathbf{x}\right) \right| = O_{\mathbb{P}}\left(\frac{n^{2/5}}{\sqrt{\log n}} \times \frac{1}{\sqrt{n}}\right) = o_{\mathbb{P}}(1). \quad (19)$$

To control the second term, we show the uniform consistency of the regression pseudo-estimator \widetilde{g}_{h_n, t_n} . The assumptions of Lemma 5 are fulfilled for $(\mathcal{X}, \mathcal{Y}, \mathcal{V}) = (\boldsymbol{\beta}^\top \mathbf{X}, Y, 1) = (\boldsymbol{\beta}^\top \mathbf{X}, g(\boldsymbol{\beta}^\top \mathbf{X}) + \sigma(\boldsymbol{\beta}^\top \mathbf{X})\varepsilon, 1)$ and $(\mathcal{X}, \mathcal{Y}, \mathcal{V}) = (\boldsymbol{\beta}^\top \mathbf{X}, 1, 1)$. Recalling that ε is independent of \mathbf{X} and centred, Lemma 5 then provides

$$\widetilde{g}_{h_n, t_n}(z) = \frac{\frac{1}{h_n} \mathbb{E}\left[Y L\left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}}{h_n}\right) \right] + O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)}{\frac{1}{h_n} \mathbb{E}\left[L\left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}}{h_n}\right) \right] + O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)}$$

uniformly on any (fixed) compact subset of $K_{\boldsymbol{\beta}}^o = (u, v)$. Noting that $h_n \sim (c/n)^{1/5}$ and $\int_{-1}^1 uL(u)du = 0$ (because L is symmetric), Lemma 6(ii) therefore entails

$$\widetilde{g}_{h_n, t_n}(z) = \frac{f_{\boldsymbol{\beta}^\top \mathbf{X}}(z)g(z) + O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)}{f_{\boldsymbol{\beta}^\top \mathbf{X}}(z) + O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)} = g(z) + O_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)$$

uniformly on any compact subset of (u, v) , the last equality being correct because $f_{\boldsymbol{\beta}^\top \mathbf{X}}$ is bounded from below by a positive constant on such sets. Together with (18) for $\lambda = 1$, this yields

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \widetilde{g}_{h_n, t_n}\left(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}\right) - g\left(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}\right) \right| = O_{\mathbb{P}}(1). \quad (20)$$

We conclude by controlling the third term in the right-hand side of (17). The idea is to define $\mathfrak{Y}_i^{(n)} := Y_i \mathbb{1}\{|Y_i| \leq t_n\}$ and, for any z and $p = 0, 1$,

$$\begin{aligned} \widehat{m}_n^{(p)}(z) &:= \frac{1}{nh_n} \sum_{i=1}^n \left[\mathfrak{Y}_i^{(n)} \right]^p L\left(\frac{z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right) \\ \text{and } \widetilde{m}_n^{(p)}(z) &:= \frac{1}{nh_n} \sum_{i=1}^n \left[\mathfrak{Y}_i^{(n)} \right]^p L\left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n}\right). \end{aligned}$$

With this notation,

$$\begin{aligned} \widehat{g}_{h_n, t_n}(z) - \widetilde{g}_{h_n, t_n}(z) &= \frac{\widehat{m}_n^{(1)}(z)}{\widehat{m}_n^{(0)}(z)} - \frac{\widetilde{m}_n^{(1)}(z)}{\widetilde{m}_n^{(0)}(z)} \\ &= \frac{[\widehat{m}_n^{(1)}(z) - \widetilde{m}_n^{(1)}(z)]\widetilde{m}_n^{(0)}(z) - [\widehat{m}_n^{(0)}(z) - \widetilde{m}_n^{(0)}(z)]\widetilde{m}_n^{(1)}(z)}{(\widetilde{m}_n^{(0)}(z) + [\widehat{m}_n^{(0)}(z) - \widetilde{m}_n^{(0)}(z)]\widetilde{m}_n^{(0)}(z))}. \end{aligned} \quad (21)$$

Since

$$|\tilde{m}_n^{(0)}(z) - f_{\mathcal{X}}(z)| = o_{\mathbb{P}}(1) \text{ and } |\tilde{m}_n^{(1)}(z) - f_{\mathcal{X}}(z)g(z)| = o_{\mathbb{P}}(1) \quad (22)$$

uniformly on any compact subset of (u, v) by Lemmas 5 and 6(ii), we concentrate on differences of the form

$$\hat{m}_n^{(p)}(z) - \tilde{m}_n^{(p)}(z) = \frac{1}{nh_n} \sum_{i=1}^n [\mathfrak{Y}_i^{(n)}]^p \left\{ L\left(\frac{z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right) - L\left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n}\right) \right\}.$$

By Taylor's theorem with integral remainder applied to the function L , we find

$$\begin{aligned} & \hat{m}_n^{(p)}(z) - \tilde{m}_n^{(p)}(z) \\ &= -\frac{1}{nh_n} \sum_{i=1}^n [\mathfrak{Y}_i^{(n)}]^p \times \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}_i}{h_n} L' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \\ &+ \frac{1}{nh_n} \sum_{i=1}^n [\mathfrak{Y}_i^{(n)}]^p \times \frac{1}{2} \left\{ \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}_i}{h_n} \right\}^2 L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \\ &+ \frac{1}{nh_n} \sum_{i=1}^n [\mathfrak{Y}_i^{(n)}]^p \times \int_{(z - \boldsymbol{\beta}^\top \mathbf{X}_i)/h_n}^{(z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i)/h_n} \left(\frac{z - \hat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n} - s \right) \left\{ L''(s) - L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \right\} ds \\ &=: T_{1,n}(z) + T_{2,n}(z) + T_{3,n}(z). \end{aligned} \quad (23)$$

We handle these three terms separately.

Control of $T_{1,n}(z)$: Note that

$$T_{1,n}(z) = -\frac{1}{h_n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \left[\frac{1}{nh_n} \sum_{i=1}^n [\mathfrak{Y}_i^{(n)}]^p L' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \mathbf{X}_i \right].$$

Recall that \mathbf{X} has compact support; Lemma 5 (choosing $\mathcal{V} = X_j$, $1 \leq j \leq d$) then yields

$$T_{1,n}(z) = -\frac{1}{h_n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \left[\frac{1}{h_n} \mathbb{E} \left(Y^p L' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}}{h_n} \right) \mathbf{X} \right) + O_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right) \right]$$

uniformly on any compact subset of (u, v) . Because for any $j \in \{1, \dots, d\}$,

$$\mathbb{E}(Y^p X_j | \boldsymbol{\beta}^\top \mathbf{X} = z) = [\mathbb{1}\{p = 0\} + g(z)\mathbb{1}\{p = 1\}] \mathbb{E}(X_j | \boldsymbol{\beta}^\top \mathbf{X} = z),$$

the conditional moment function $z \mapsto \mathbb{E}(Y^p X_j | \boldsymbol{\beta}^\top \mathbf{X} = z)$ satisfies the regularity requirements of Lemma 6(i). By Lemma 6(i) and the symmetry of L ,

$$\frac{1}{h_n} \mathbb{E} \left(Y^p L' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}}{h_n} \right) \mathbf{X} \right) = O(h_n)$$

uniformly on any compact subset of (u, v) . Since $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_{\mathbb{P}}(1/\sqrt{n})$, this yields

$$T_{1,n}(z) = O_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right) = o_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right) \text{ uniformly on any compact subset of } (u, v). \quad (24)$$

Control of $T_{2,n}(z)$: Recall that \mathbf{X} has compact support, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_{\mathbb{P}}(1/\sqrt{n})$, and L'' is bounded to obtain, using the law of large numbers,

$$\sup_{z \in \mathbb{R}} |T_{2,n}(z)| = O_{\mathbb{P}} \left(\frac{1}{nh_n^3} \times \frac{1}{n} \sum_{i=1}^n |Y_i|^p \right) = O_{\mathbb{P}} \left(\frac{1}{nh_n^3} \right) = O_{\mathbb{P}} \left(\frac{1}{n^{2/5}} \right) = o_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right). \quad (25)$$

Control of $T_{3,n}(z)$: Use a change of variables to rewrite the integral term in $T_{3,n}(z)$ as

$$\begin{aligned} & \int_{(z-\boldsymbol{\beta}^\top \mathbf{X}_i)/h_n}^{(z-\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i)/h_n} \left(\frac{z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n} - s \right) \left\{ L''(s) - L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \right\} ds \\ &= \int_0^{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i/h_n} \left(\frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i}{h_n} - u \right) \left\{ L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} + u \right) - L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \right\} du. \end{aligned}$$

Since \mathbf{X} has compact support and $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_{\mathbb{P}}(1/\sqrt{n})$ we have

$$\max_{1 \leq i \leq n} \left| \frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i}{h_n} \right| = O_{\mathbb{P}} \left(\frac{1}{h_n \sqrt{n}} \right) = o_{\mathbb{P}}(1).$$

By uniform continuity of the continuous and compactly supported function L'' , it follows that

$$\max_{1 \leq i \leq n} \sup_{z \in \mathbb{R}} \sup_{|u| \leq |(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i|/h_n} \left| L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} + u \right) - L'' \left(\frac{z - \boldsymbol{\beta}^\top \mathbf{X}_i}{h_n} \right) \right| = o_{\mathbb{P}}(1).$$

We then get

$$\begin{aligned} \sup_{z \in \mathbb{R}} |T_{3,n}(z)| &= O_{\mathbb{P}} \left(\frac{1}{nh_n} \sum_{i=1}^n |Y_i|^p \left| \int_0^{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i/h_n} \left| \frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i}{h_n} - u \right| du \right| \right) \\ &= O_{\mathbb{P}} \left(\frac{1}{nh_n} \sum_{i=1}^n |Y_i|^p \left[\frac{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i}{h_n} \right]^2 \right) \\ &= O_{\mathbb{P}} \left(\frac{1}{nh_n^3} \times \frac{1}{n} \sum_{i=1}^n |Y_i|^p \right) = O_{\mathbb{P}} \left(\frac{1}{nh_n^3} \right) = o_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right). \end{aligned} \quad (26)$$

Combine (21), (22), (23), (24), (25) and (26) to obtain

$$\widehat{g}_{h_n, t_n}(z) - \widetilde{g}_{h_n, t_n}(z) = \frac{o_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right)}{f_{\boldsymbol{\beta}^\top \mathbf{X}}(z) + o_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right)} = o_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right)$$

uniformly on any compact subset of (u, v) . Using (18) again with $\lambda = 1$, we get

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \widehat{g}_{h_n, t_n} \left(\widehat{\boldsymbol{\beta}}^\top \mathbf{x} \right) - \widetilde{g}_{h_n, t_n} \left(\widehat{\boldsymbol{\beta}}^\top \mathbf{x} \right) \right| = o_{\mathbb{P}}(1). \quad (27)$$

Combining (17), (19), (20) and (27) concludes the proof of the assertion on \widehat{g}_{h_n, t_n} .

We turn to the control of $\widehat{\sigma}_{h_n, t_n}$, where the added difficulty is that the computation of the estimator is based on the absolute residuals $\widehat{Z}_{i, h_n, t_n} = \left| Y_i - \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i) \right|$ rather than on the “true values” $Z_i := |Y_i - g(\boldsymbol{\beta}^\top \mathbf{X}_i)|$. We thus introduce its pseudo-estimator analogue based on the Z_i ,

$$\bar{\sigma}_{h_n, t_n}(z) := \sum_{i=1}^n Z_i \mathbb{1}\{Z_i \leq t_n\} L\left(\frac{z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right) \bigg/ \sum_{i=1}^n L\left(\frac{z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right)$$

and we seek to control $|\widehat{\sigma}_{h_n, t_n}(z) - \bar{\sigma}_{h_n, t_n}(z)|$, for $z = \widehat{\boldsymbol{\beta}}^\top \mathbf{x}$, uniformly in $\mathbf{x} \in K_0$. Write

$$\begin{aligned} & \widehat{\sigma}_{h_n, t_n}(z) - \bar{\sigma}_{h_n, t_n}(z) \\ &= \sum_{i=1}^n \left[\widehat{Z}_{i, h_n, t_n} \mathbb{1}\{\widehat{Z}_{i, h_n, t_n} \leq t_n\} - Z_i \mathbb{1}\{Z_i \leq t_n\} \right] L\left(\frac{z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right) \bigg/ \sum_{i=1}^n L\left(\frac{z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i}{h_n}\right). \end{aligned}$$

Note that the only pairs (\mathbf{X}_i, Y_i) making a nonzero contribution to this difference are those for which $|z - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i| \leq h_n$. For $\mathbf{x} \in K_0$, we thus focus on controlling

$$\sup_{\mathbf{x} \in K_0} \left| \widehat{Z}_{i, h_n, t_n} \mathbb{1}\{\widehat{Z}_{i, h_n, t_n} \leq t_n\} - Z_i \mathbb{1}\{Z_i \leq t_n\} \right| \mathbb{1}\left\{ \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \right\}.$$

Since $\left| \widehat{Z}_{i, h_n, t_n} - Z_i \right| \leq \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i) - g(\boldsymbol{\beta}^\top \mathbf{X}_i) \right|$, the triangle inequality yields

$$\begin{aligned} & \sup_{\mathbf{x} \in K_0} \left| \widehat{Z}_{i, h_n, t_n} \mathbb{1}\{\widehat{Z}_{i, h_n, t_n} \leq t_n\} - Z_i \mathbb{1}\{Z_i \leq t_n\} \right| \mathbb{1}\left\{ \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \right\} \\ & \leq \sup_{\mathbf{x} \in K_0} \max_{i: |\widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i| \leq h_n} \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i) - g(\boldsymbol{\beta}^\top \mathbf{X}_i) \right| \end{aligned} \quad (28)$$

$$+ \sup_{\mathbf{x} \in K_0} Z_i \left| \mathbb{1}\{\widehat{Z}_{i, h_n, t_n} \leq t_n\} - \mathbb{1}\{Z_i \leq t_n\} \right| \mathbb{1}\left\{ \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \right\}. \quad (29)$$

We focus on (28) first, where the idea is to use our uniform convergence result on \widehat{g}_{h_n, t_n} . Write

$$\left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \Rightarrow \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \boldsymbol{\beta}^\top \mathbf{X}_i \right| \leq h_n + |(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}_i| = h_n + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$$

irrespective of the index i and $\mathbf{x} \in K_0$, so that, with arbitrarily large probability as $n \rightarrow \infty$,

$$\forall i \in \{1, \dots, n\}, \forall \mathbf{x} \in K_0, \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \Rightarrow \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \boldsymbol{\beta}^\top \mathbf{X}_i \right| \leq 2h_n.$$

Recall that, by (18), $\widehat{\boldsymbol{\beta}}^\top \mathbf{x} \in [u + \rho/4, v - \rho/4]$ with arbitrarily large probability as $n \rightarrow \infty$, irrespective of $\mathbf{x} \in K_0$. Since $h_n \rightarrow 0$, this yields, with arbitrarily large probability as $n \rightarrow \infty$,

$$\forall i \in \{1, \dots, n\}, \forall \mathbf{x} \in K_0, \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \Rightarrow \boldsymbol{\beta}^\top \mathbf{X}_i \in [u + \rho/8, v - \rho/8].$$

In other words, for such indices i , \mathbf{X}_i belongs to the intersection of K and the inverse image of the closed interval $[u + \rho/8, v - \rho/8]$ by the (continuous) projection mapping $\mathbf{x} \mapsto \boldsymbol{\beta}^\top \mathbf{x}$. This intersection is itself a compact set K_1 , say, and therefore, with arbitrarily large probability as $n \rightarrow \infty$,

$$\forall i \in \{1, \dots, n\}, \forall \mathbf{x} \in K_0, |\widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i| \leq h_n \Rightarrow \mathbf{X}_i \in K_1.$$

Note also that $K_1 \subset K^\circ$ since K_1 is contained in the (open) inverse image of the open interval $(u + \rho/16, v - \rho/16)$ by the same projection mapping. It then follows from our uniform convergence result on \widehat{g}_{h_n, t_n} that

$$\sup_{\mathbf{x} \in K_0} \max_{i: |\widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i| \leq h_n} \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i) - g(\boldsymbol{\beta}^\top \mathbf{X}_i) \right| = O_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right). \quad (30)$$

We can now control (29). Clearly

$$\left| \mathbb{1} \left\{ \widehat{Z}_{i, h_n, t_n} \leq t_n \right\} - \mathbb{1} \left\{ Z_i \leq t_n \right\} \right| = \mathbb{1} \left\{ \widehat{Z}_{i, h_n, t_n} \leq t_n, Z_i > t_n \right\} + \mathbb{1} \left\{ \widehat{Z}_{i, h_n, t_n} > t_n, Z_i \leq t_n \right\}.$$

Recall that $\left| \widehat{Z}_{i, h_n, t_n} - Z_i \right| \leq \left| \widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i) - g(\boldsymbol{\beta}^\top \mathbf{X}_i) \right|$ and use (30) together with the assumption $t_n \rightarrow \infty$ to find that, with arbitrarily large probability as $n \rightarrow \infty$,

$$\begin{aligned} \forall i \in \{1, \dots, n\}, \quad & \sup_{\mathbf{x} \in K_0} Z_i \left| \mathbb{1} \left\{ \widehat{Z}_{i, h_n, t_n} \leq t_n \right\} - \mathbb{1} \left\{ Z_i \leq t_n \right\} \right| \mathbb{1} \left\{ \left| \widehat{\boldsymbol{\beta}}^\top \mathbf{x} - \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i \right| \leq h_n \right\} \\ & \leq Z_i \mathbb{1} \left\{ Z_i \leq 2t_n, Z_i > t_n \right\} + Z_i \mathbb{1} \left\{ Z_i > t_n/2, Z_i \leq t_n \right\} \\ & \leq Z_i \mathbb{1} \left\{ t_n/2 < Z_i \leq 2t_n \right\}. \end{aligned} \quad (31)$$

Combine (30) and (31) to obtain, with arbitrarily large probability as $n \rightarrow \infty$,

$$\begin{aligned} & \sup_{\mathbf{x} \in K_0} \left| \widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \bar{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) \right| \\ & \leq \sup_{\mathbf{x} \in K_0} \left[\bar{\sigma}_{h_n, 2t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \bar{\sigma}_{h_n, t_n/2}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) \right] + O_{\mathbb{P}} \left(\frac{\sqrt{\log n}}{n^{2/5}} \right). \end{aligned} \quad (32)$$

To conclude, note that since $\mathbb{E}|\varepsilon| = 1$,

$$Z := |Y - g(\boldsymbol{\beta}^\top \mathbf{X})| = \sigma(\boldsymbol{\beta}^\top \mathbf{X}) + \sigma(\boldsymbol{\beta}^\top \mathbf{X})(|\varepsilon| - \mathbb{E}|\varepsilon|).$$

This single-index model linking Z to \mathbf{X} has the same structure as model (M_2) and satisfies our assumptions, with g replaced by σ and ε replaced by $|\varepsilon| - \mathbb{E}|\varepsilon|$. Since for this model $\bar{\sigma}_{h_n, t_n}$ plays the role of \widehat{g}_{h_n, t_n} , we can use the first part of the Proposition to get

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \bar{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \right| = O_{\mathbb{P}}(1). \quad (33)$$

The result then follows by using (32) to write

$$\begin{aligned} \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \right| & \leq \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \bar{\sigma}_{h_n, 2t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \right| \\ & \quad + \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \bar{\sigma}_{h_n, t_n/2}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \right| \\ & \quad + \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\mathbf{x} \in K_0} \left| \bar{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \right| + O_{\mathbb{P}}(1) \end{aligned}$$

and then by using (33) as well as its analogues with t_n replaced by $t_n/2$ and $2t_n$. \square

B Appendix: proofs of the main results

Proof of Theorem 1. Note that

$$\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) = \arg \min_{u \in \mathbb{R}} \chi_n(u)$$

$$\text{with } \chi_n(u) := \frac{1}{2\xi_{\tau_n}^2(\varepsilon)} \sum_{i=1}^n \left[\eta_{\tau_n} \left(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - \frac{u\xi_{\tau_n}(\varepsilon)}{\sqrt{n(1-\tau_n)}} \right) - \eta_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon)) \right].$$

Define

$$\psi_n(u) := \frac{1}{2\xi_{\tau_n}^2(\varepsilon)} \sum_{i=1}^n \left[\eta_{\tau_n} \left(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - \frac{u\xi_{\tau_n}(\varepsilon)}{\sqrt{n(1-\tau_n)}} \right) - \eta_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon)) \right].$$

In other words, $\psi_n(u)$ is the empirical criterion based on the true, unobservable errors ε_i . Note that for any n , $u \mapsto \psi_n(u)$ is a continuously differentiable convex function. We shall prove that, pointwise in u , $\chi_n(u) - \psi_n(u) \xrightarrow{\mathbb{P}} 0$. The result will then be a straightforward consequence of the convexity lemma of Geyer (1996) (stated as Theorem 5 in Knight, 1999) together with the convergence

$$\psi_n(u) \xrightarrow{d} -uZ\sqrt{\frac{2\gamma}{1-2\gamma}} + \frac{u^2}{2\gamma} \text{ as } n \rightarrow \infty$$

(in the finite-dimensional sense, with Z being standard Gaussian) shown in the proof of Theorem 2 in Daouia *et al.* (2018).

We start by recalling that

$$\frac{1}{2}(\eta_{\tau}(x-y) - \eta_{\tau}(x)) = -\int_0^y \varphi_{\tau}(x-t)dt$$

where $\varphi_{\tau}(y) = |\tau - \mathbb{1}\{y \leq 0\}|y$ (see Lemma 2 in Daouia *et al.*, 2018). Therefore

$$\chi_n(u) - \psi_n(u) = -\frac{1}{\xi_{\tau_n}^2(\varepsilon)} \sum_{i=1}^n \int_0^{u\xi_{\tau_n}(\varepsilon)/\sqrt{n(1-\tau_n)}} [\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)]dt.$$

Set $I_n(u) = [0, |u|\xi_{\tau_n}(\varepsilon)/\sqrt{n(1-\tau_n)}]$. Since

$$|\chi_n(u) - \psi_n(u)| \leq \frac{|u|}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \sup_{|t| \in I_n(u)} |\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)|,$$

it is enough to show that

$$T_n(u) := \frac{1}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \sup_{|t| \in I_n(u)} |\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)| \xrightarrow{\mathbb{P}} 0. \quad (34)$$

We now apply Lemma 3 in Daouia *et al.* (2018), which gives, for any $x, h \in \mathbb{R}$,

$$|\varphi_\tau(x - h) - \varphi_\tau(x)| \leq |h|(1 - \tau + 2\mathbb{1}\{x > \min(h, 0)\}).$$

This translates into

$$\begin{aligned} & |\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)| \\ & \leq |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|(1 - \tau_n + 2\mathbb{1}\{\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0)\}). \end{aligned}$$

Hence the inequality

$$T_n(u) \leq T_{1,n} + T_{2,n}(u) \tag{35}$$

$$\text{with } T_{1,n} := \frac{\sqrt{1 - \tau_n}}{\xi_{\tau_n}(\varepsilon)\sqrt{n}} \sum_{i=1}^n |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|$$

$$\text{and } T_{2,n}(u) := \frac{2}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1 - \tau_n)}} \sum_{i=1}^n \sup_{|t| \in I_n(u)} |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \mathbb{1}\{\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0)\}.$$

We first focus on $T_{1,n}$. We have

$$T_{1,n} \leq \left[\frac{\sqrt{n(1 - \tau_n)}}{\xi_{\tau_n}(\varepsilon)} \max_{1 \leq i \leq n} R_{n,i} \right] \times \frac{1}{n} \sum_{i=1}^n (1 + |\varepsilon_i|) = \text{O}_{\mathbb{P}} \left(\frac{\sqrt{n(1 - \tau_n)}}{\xi_{\tau_n}(\varepsilon)} \max_{1 \leq i \leq n} R_{n,i} \right)$$

by the law of large numbers. Note now that $\xi_{\tau_n}(\varepsilon) \rightarrow \infty$ and thus

$$T_{1,n} = \text{O}_{\mathbb{P}} \left(\frac{\sqrt{n(1 - \tau_n)}}{q_{\tau_n}(\varepsilon)} \max_{1 \leq i \leq n} R_{n,i} \right) = \text{o}_{\mathbb{P}} \left(\sqrt{n(1 - \tau_n)} \max_{1 \leq i \leq n} R_{n,i} \right) \xrightarrow{\mathbb{P}} 0 \tag{36}$$

by assumption. We now turn to the control of $T_{2,n}(u)$, for which we write, for any t ,

$$\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0) \Rightarrow \varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > 0 \text{ or } \widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t > 0.$$

It follows that, for n large enough, we have, for any t such that $|t| \in I_n(u)$,

$$\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0) \Rightarrow \varepsilon_i > \frac{\xi_{\tau_n}(\varepsilon)}{2} \text{ or } \widehat{\varepsilon}_i^{(n)} > \frac{\xi_{\tau_n}(\varepsilon)}{2}. \tag{37}$$

Now, for n large enough and with arbitrarily large probability as $n \rightarrow \infty$, $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \leq (1 + |\varepsilon_i|)/2$ for any $i \in \{1, \dots, n\}$, so that after some algebra,

$$\widehat{\varepsilon}_i^{(n)} > \frac{\xi_{\tau_n}(\varepsilon)}{2} \Rightarrow \varepsilon_i + \frac{1}{2}|\varepsilon_i| > \frac{1}{2}(\xi_{\tau_n}(\varepsilon) - 1) \Rightarrow \varepsilon_i + \frac{1}{2}|\varepsilon_i| > \frac{1}{4}\xi_{\tau_n}(\varepsilon)$$

because $\xi_{\tau_n}(\varepsilon) \rightarrow \infty$. Since the quantity $x + |x|/2$ can only be positive if $x > 0$, it follows that, with arbitrarily large probability,

$$\widehat{\varepsilon}_i^{(n)} > \frac{\xi_{\tau_n}(\varepsilon)}{2} \Rightarrow \varepsilon_i > \frac{1}{6}\xi_{\tau_n}(\varepsilon). \tag{38}$$

Combining (37) and (38) results in the following bound, valid with arbitrarily large probability as $n \rightarrow \infty$:

$$T_{2,n}(u) \leq \frac{2}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^n |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \mathbb{1} \left\{ \varepsilon_i > \frac{1}{6} \xi_{\tau_n}(\varepsilon) \right\}.$$

By assumption on $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|$, this leads to

$$T_{2,n}(u) \leq 4 \left[\frac{\sqrt{n(1-\tau_n)}}{\xi_{\tau_n}(\varepsilon)} \max_{1 \leq i \leq n} R_{n,i} \right] \times \frac{1}{n(1-\tau_n)} \sum_{i=1}^n \varepsilon_i \mathbb{1} \left\{ \varepsilon_i > \frac{1}{6} \xi_{\tau_n}(\varepsilon) \right\}.$$

Finally, the regular variation property of \bar{F} and the asymptotic proportionality relationship between $\xi_{\tau_n}(\varepsilon)$ and $q_{\tau_n}(\varepsilon)$ ensure that

$$\lim_{n \rightarrow \infty} \frac{\bar{F}(\xi_{\tau_n}(\varepsilon)/6)}{1-\tau_n} \text{ exists, is positive and finite.}$$

Lemma 1 then entails

$$T_{2,n}(u) = O_{\mathbb{P}} \left(\sqrt{n(1-\tau_n)} \max_{1 \leq i \leq n} R_{n,i} \right) \xrightarrow{\mathbb{P}} 0 \quad (39)$$

by assumption. Combining (34), (35), (36) and (39) completes the proof. \square

Proof of Theorem 2. To prove the first expansion, write

$$\frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{q_{1-k/n}(\varepsilon)} - s^{-\gamma} = \frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{\varepsilon_{n-[ks],n}} \left(\frac{\varepsilon_{n-[ks],n}}{q_{1-k/n}(\varepsilon)} - s^{-\gamma} \right) + s^{-\gamma} \left(\frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{\varepsilon_{n-[ks],n}} - 1 \right).$$

Use Lemma 3 and Theorem 2.4.8 in de Haan and Ferreira (2006) to get

$$\begin{aligned} & \frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{\varepsilon_{n-[ks],n}} \left(\frac{\varepsilon_{n-[ks],n}}{q_{1-k/n}(\varepsilon)} - s^{-\gamma} \right) \\ &= \frac{1}{\sqrt{k}} \left[\gamma s^{-\gamma-1} W_n(s) + \sqrt{k} A(n/k) s^{-\gamma} \frac{s^{-\rho} - 1}{\rho} + s^{-\gamma-1/2-\delta} o_{\mathbb{P}}(1) \right] \end{aligned} \quad (40)$$

uniformly in $s \in (0, 1]$. Applying Lemma 3 again gives

$$s^{-\gamma} \left| \frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{\varepsilon_{n-[ks],n}} - 1 \right| \leq s^{-\gamma-1/2-\delta} \left| \frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{\varepsilon_{n-[ks],n}} - 1 \right| = \frac{s^{-\gamma-1/2-\delta}}{\sqrt{k}} o_{\mathbb{P}}(1) \quad (41)$$

uniformly in $s \in (0, 1]$. Combine (40) and (41) to complete the proof of the first expansion. The proof of the second expansion is based on the equality

$$\log \left(\frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{q_{1-k/n}(\varepsilon)} \right) = \log \left(\frac{\varepsilon_{n-[ks],n}}{q_{1-k/n}(\varepsilon)} \right) + \log \left(\frac{\widehat{\varepsilon}_{n-[ks],n}^{(n)}}{\varepsilon_{n-[ks],n}} \right)$$

and follows exactly the same ideas. \square

Proof of Corollary 1. Notice that, by Theorem 2, there is a sequence W_n of standard Brownian motions such that, for any $\delta > 0$ sufficiently small:

$$\begin{aligned}\widehat{\gamma}_k &= \int_0^1 \log \left(\frac{\widehat{\varepsilon}_{n-\lfloor ks \rfloor, n}^{(n)}}{\widehat{\varepsilon}_{n-k, n}^{(n)}} \right) ds \\ &= \int_0^1 \left\{ -\gamma \log s + \frac{\gamma}{\sqrt{k}} [s^{-1}W_n(s) - W_n(1)] + A \left(\frac{n}{k} \right) \left[\frac{s^{-\rho} - 1}{\rho} + s^{-1/2-\delta} o_{\mathbb{P}}(1) \right] \right\} ds.\end{aligned}$$

We then obtain that $\widehat{\gamma}_k$ can be written

$$\sqrt{k}(\widehat{\gamma}_k - \gamma) = \frac{\lambda}{1 - \rho} + \gamma \int_0^1 [s^{-1}W_n(s) - W_n(1)] + o_{\mathbb{P}}(1).$$

Similarly,

$$\sqrt{k} \left(\frac{\widehat{\varepsilon}_{n-k, n}^{(n)}}{q_{1-k/n}(\varepsilon)} - 1 \right) = \gamma W_n(1) + o_{\mathbb{P}}(1).$$

Noting that the Gaussian terms in these two asymptotic expansions are independent completes the proof. \square

Proof of Theorem 3. The key is to note that

$$\begin{aligned}\frac{\overline{\xi}_{\tau'_n}^*(Y|\mathbf{x})}{\xi_{\tau'_n}(Y|\mathbf{x})} - 1 &= \left(1 + \frac{g(\mathbf{x})}{s(\mathbf{x})\xi_{\tau'_n}(\varepsilon)} \right)^{-1} \left(\frac{\overline{\xi}_{\tau'_n}^*(\varepsilon)}{\xi_{\tau'_n}(\varepsilon)} - 1 \right) \\ &+ \frac{\overline{g}(\mathbf{x}) - g(\mathbf{x})}{g(\mathbf{x}) + s(\mathbf{x})\xi_{\tau'_n}(\varepsilon)} + \left(1 + \frac{g(\mathbf{x})}{s(\mathbf{x})\xi_{\tau'_n}(\varepsilon)} \right)^{-1} \frac{\overline{s}(\mathbf{x}) - s(\mathbf{x})}{s(\mathbf{x})} \frac{\overline{\xi}_{\tau'_n}^*(\varepsilon)}{\xi_{\tau'_n}(\varepsilon)}.\end{aligned}$$

Since $1/\xi_{\tau'_n}(\varepsilon) = o(1/\xi_{\tau_n}(\varepsilon)) = o(1/q_{\tau_n}(\varepsilon)) = o(1/\sqrt{n(1-\tau_n)})$, our assumptions entail

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \left(\frac{\overline{\xi}_{\tau'_n}^*(Y|\mathbf{x})}{\xi_{\tau'_n}(Y|\mathbf{x})} - 1 \right) = \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \left(\frac{\overline{\xi}_{\tau'_n}^*(\varepsilon)}{\xi_{\tau'_n}(\varepsilon)} - 1 \right) (1 + o_{\mathbb{P}}(1)) + o_{\mathbb{P}}(1).$$

Our result is then shown by adapting the proof of Theorem 5 of Daouia *et al.* (2020). We omit the details. \square

Proof of Corollary 2. (i) The key is to write

$$\begin{aligned}\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y|\mathbf{x})}{\xi_{\tau_n}(Y|\mathbf{x})} - 1 \right) &= \frac{(1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon)}{\alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon)} \times \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \\ &+ \frac{\sqrt{1-\tau_n}}{\alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon)} \times \sqrt{n} \left[\widehat{\alpha} - \alpha + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{x} \right] \\ &+ \frac{\sqrt{1-\tau_n} \widehat{\xi}_{\tau_n}(\varepsilon)}{\alpha + \boldsymbol{\beta}^\top \mathbf{x} + (1 + \boldsymbol{\theta}^\top \mathbf{x})\xi_{\tau_n}(\varepsilon)} \times \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \mathbf{x}.\end{aligned}$$

Now

$$\widehat{\varepsilon}_i^{(n)} - \varepsilon_i = \frac{\alpha - \widehat{\alpha} + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \mathbf{X}_i}{1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{X}_i} + \frac{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^\top \mathbf{X}_i}{1 + \widehat{\boldsymbol{\theta}}^\top \mathbf{X}_i} \varepsilon_i.$$

Then clearly, by Lemma 4 and since \mathbf{X} has a compact support,

$$|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \leq R_{n,i}(1 + |\varepsilon_i|) \quad \text{where} \quad \max_{1 \leq i \leq n} R_{n,i} = O_{\mathbb{P}}(1/\sqrt{n}).$$

We conclude by combining Lemma 4, Theorem 1 and the convergence $\xi_{\tau_n}(\varepsilon) \rightarrow \infty$.

(ii) Combine (i) with Theorem 3. □

Proof of Theorem 4. (i) We first show

$$\sqrt{N(1 - \tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 \right) \xrightarrow{d} \mathcal{N} \left(0, \frac{2\gamma^3}{1 - 2\gamma} \right). \quad (42)$$

The key here, as in the proof of Theorem 1, is to show the convergence of the pseudo-optimisation criterion based on the unobserved errors. Let $\varepsilon_{1,K_0}, \dots, \varepsilon_{N,K_0}$ be those noise variables whose corresponding covariate vectors $\mathbf{X}_i \in K_0$, and set

$$\psi_N(u) := \frac{1}{2\xi_{\tau_N}^2(\varepsilon)} \sum_{i=1}^N \left[\eta_{\tau_N} \left(\varepsilon_{i,K_0} - \xi_{\tau_N}(\varepsilon) - \frac{u\xi_{\tau_N}(\varepsilon)}{\sqrt{N(1 - \tau_N)}} \right) - \eta_{\tau_N}(\varepsilon_{i,K_0} - \xi_{\tau_N}(\varepsilon)) \right].$$

Given $N = m > 0$, $(\varepsilon_{1,K_0}, \dots, \varepsilon_{N,K_0}) \stackrel{d}{=} (\varepsilon_1, \dots, \varepsilon_m)$. Besides, $N = N(K_0, n)$ is binomial with parameters n and $\mathbb{P}(\mathbf{X} \in K_0) > 0$, so that $N/n \xrightarrow{\mathbb{P}} \mathbb{P}(\mathbf{X} \in K_0)$. A combination of Theorem 2 in Daouia *et al.* (2018) and the de-conditioning lemma of Stupfler (2019, Lemma 8) yields

$$\psi_N(u) \xrightarrow{d} -uZ \sqrt{\frac{2\gamma}{1 - 2\gamma}} + \frac{u^2}{2\gamma} \quad \text{as } n \rightarrow \infty$$

in the finite-dimensional sense, with Z being standard Gaussian.

Define now

$$\chi_N(u) := \frac{1}{2\xi_{\tau_N}^2(\varepsilon)} \sum_{i=1}^N \left[\eta_{\tau_N} \left(\widehat{\varepsilon}_{i,K_0}^{(n)} - \xi_{\tau_N}(\varepsilon) - \frac{u\xi_{\tau_N}(\varepsilon)}{\sqrt{N(1 - \tau_N)}} \right) - \eta_{\tau_N}(\widehat{\varepsilon}_{i,K_0}^{(n)} - \xi_{\tau_N}(\varepsilon)) \right].$$

Our goal is to show that $\chi_N(u) - \psi_N(u) \xrightarrow{\mathbb{P}} 0$. Following the ideas of the proof of Theorem 1, it is enough to prove that

$$T_{1,N} = \frac{\sqrt{1 - \tau_N}}{\xi_{\tau_N}(\varepsilon)\sqrt{N}} \sum_{i=1}^N |\widehat{\varepsilon}_{i,K_0}^{(n)} - \varepsilon_{i,K_0}| \xrightarrow{\mathbb{P}} 0 \quad (43)$$

and that, if $I_N(u) = [0, |u|\xi_{\tau_N}(\varepsilon)/\sqrt{N(1 - \tau_N)}]$,

$$T_{2,N}(u) = \frac{2}{\xi_{\tau_N}(\varepsilon)\sqrt{N(1 - \tau_N)}} \sum_{i=1}^N \sup_{|t| \in I_N(u)} |\widehat{\varepsilon}_{i,K_0}^{(n)} - \varepsilon_{i,K_0}| \mathbb{1}\{\varepsilon_{i,K_0} - \xi_{\tau_N}(\varepsilon) - t > \min(\varepsilon_{i,K_0} - \widehat{\varepsilon}_{i,K_0}^{(n)}, 0)\} \xrightarrow{\mathbb{P}} 0. \quad (44)$$

Since

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

we get

$$T_{1,N} = o_{\mathbb{P}} \left(\sqrt{\log n} \frac{N^{(1-a)/2}}{n^{2/5}} \times \frac{1}{N} \sum_{i=1}^N (1 + |\varepsilon_{i,K_0}|) \right) = o_{\mathbb{P}} \left(n^{(1-5a)/10} \sqrt{\log n} \right) \rightarrow 0 \quad (45)$$

(here the law of large numbers was combined with Lemma 8 of Stupfler, 2019, to control the average of the $1 + |\varepsilon_{i,K_0}|$). We now turn to the control of $T_{2,N}(u)$. Recall that $N \xrightarrow{\mathbb{P}} \infty$ and follow the ideas leading to (38) in the proof of Theorem 1 to find, for n large enough,

$$\varepsilon_{i,K_0} - \xi_{\tau_N}(\varepsilon) - t > \min(\varepsilon_{i,K_0} - \widehat{\varepsilon}_{i,K_0}^{(n)}, 0) \Rightarrow \varepsilon_{i,K_0} > \frac{1}{6} \xi_{\tau_N}(\varepsilon)$$

with arbitrarily large probability, irrespective of $i \in \{1, \dots, N\}$ and t such that $|t| \in I_N(u)$. Therefore, with arbitrarily large probability as $n \rightarrow \infty$:

$$\begin{aligned} T_{2,N}(u) &\leq \frac{2}{\xi_{\tau_N}(\varepsilon) \sqrt{N(1-\tau_N)}} \sum_{i=1}^N |\widehat{\varepsilon}_{i,K_0}^{(n)} - \varepsilon_{i,K_0}| \mathbb{1} \left\{ \varepsilon_{i,K_0} > \frac{1}{6} \xi_{\tau_N}(\varepsilon) \right\} \\ &= O_{\mathbb{P}} \left(\sqrt{N(1-\tau_N)} \frac{\sqrt{\log n}}{n^{2/5}} \times \frac{1}{N \xi_{\tau_N}(\varepsilon) (1-\tau_N)} \sum_{i=1}^N \varepsilon_{i,K_0} \mathbb{1} \left\{ \varepsilon_{i,K_0} > \frac{1}{6} \xi_{\tau_N}(\varepsilon) \right\} \right). \end{aligned}$$

Combine Lemma 1 with Lemma 8 of Stupfler (2019) to get

$$T_{2,N}(u) = O_{\mathbb{P}} \left(\sqrt{N(1-\tau_N)} \frac{\sqrt{\log n}}{n^{2/5}} \right) = O_{\mathbb{P}} \left(n^{(1-5a)/10} \sqrt{\log n} \right) \xrightarrow{\mathbb{P}} 0. \quad (46)$$

Combine (43), (44), (45) and (46) to get $\chi_N(u) - \psi_N(u) \xrightarrow{\mathbb{P}} 0$. Conclude by noting that

$$\sqrt{N(1-\tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 \right) = \arg \min_{u \in \mathbb{R}} \chi_N(u).$$

On the event $\{N > 0\}$, whose probability tends to 1, the function $u \mapsto \chi_N(u)$ is a continuously differentiable convex function, and since $\chi_N(u) - \psi_N(u) \xrightarrow{\mathbb{P}} 0$, we have

$$\chi_N(u) \xrightarrow{d} -uZ \sqrt{\frac{2\gamma}{1-2\gamma}} + \frac{u^2}{2\gamma} \quad \text{as } n \rightarrow \infty$$

in the finite-dimensional sense, with Z being standard Gaussian. Applying Theorem 5 in Knight (1999) completes the proof of (42). Statement (i) then follows in a straightforward way from Proposition 2 and the representation

$$\begin{aligned} \frac{\widehat{\xi}_{\tau_N}(Y|\mathbf{x})}{\xi_{\tau_N}(Y|\mathbf{x})} - 1 &= \frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 + \frac{\widehat{g}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x})}{g(\boldsymbol{\beta}^\top \mathbf{x}) + \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \xi_{\tau_N}(\varepsilon)} + \frac{\widehat{\sigma}_{h_n, t_n}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}) - \sigma(\boldsymbol{\beta}^\top \mathbf{x})}{g(\boldsymbol{\beta}^\top \mathbf{x}) + \sigma(\boldsymbol{\beta}^\top \mathbf{x}) \xi_{\tau_N}(\varepsilon)} \widehat{\xi}_{\tau_N}(\varepsilon) \\ &\quad - \frac{g(\boldsymbol{\beta}^\top \mathbf{x}) / \xi_{\tau_N}(\varepsilon)}{\sigma(\boldsymbol{\beta}^\top \mathbf{x}) + g(\boldsymbol{\beta}^\top \mathbf{x}) / \xi_{\tau_N}(\varepsilon)} \left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 \right). \end{aligned}$$

(ii) Set $\widehat{\xi}_{\tau'_N}^*(\varepsilon) = \left(\frac{1-\tau'_N}{1-\tau_N}\right)^{-\bar{\gamma}} \widehat{\xi}_{\tau_N}(\varepsilon)$. Use the ideas of the proof of Theorem 3 to find that

$$\frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau'_N)]} \left(\frac{\widehat{\xi}_{\tau'_N}^*(Y|\mathbf{x})}{\xi_{\tau'_N}(Y|\mathbf{x})} - 1 \right) \quad \text{and} \quad \frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau'_N)]} \left(\frac{\widehat{\xi}_{\tau'_N}^*(\varepsilon)}{\xi_{\tau'_N}(\varepsilon)} - 1 \right)$$

have the same asymptotic distribution. Our result is then shown by using the assumption $\sqrt{N(1-\tau_N)}(\bar{\gamma}-\gamma) \xrightarrow{d} \Gamma$, as well as convergence (42) and by adapting the proof of Theorem 5 of Daouia *et al.* (2020). We omit the details. \square

Proof of Corollary 3. (i) Write first

$$\begin{aligned} \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})}{\xi_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})} - 1 \right) &= \frac{\xi_{\tau_n}(\varepsilon)}{\sum_{j=1}^p \phi_j y_{-j} + \xi_{\tau_n}(\varepsilon)} \times \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \\ &+ \sqrt{n(1-\tau_n)} \frac{\sum_{j=1}^p (\widehat{\phi}_{j,n} - \phi_j) y_{-j}}{\sum_{j=1}^p \phi_j y_{-j} + \xi_{\tau_n}(\varepsilon)}. \end{aligned}$$

The residuals $\widehat{\varepsilon}_t^{(n)} = Y_t - \sum_{j=1}^p \widehat{\phi}_{j,n} Y_{t-j}$, for $p+1 \leq t \leq n$, clearly satisfy

$$\max_{p+1 \leq t \leq n} \frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1 + |\varepsilon_t|} = O_{\mathbb{P}} \left(n^{-1/2} \max_{1 \leq t \leq n} |Y_t| \right) = O_{\mathbb{P}} \left(n^{-1/2} \left[1 + \max_{p+1 \leq t \leq n} |Y_t| \right] \right).$$

Observe now that $\max_{p+1 \leq t \leq n} |Y_t| = O_{\mathbb{P}}(\max_{p+1 \leq t \leq n} |\varepsilon_t|)$. If $p = 1$, this follows immediately from the fact that the unique autoregression coefficient ϕ lies in $(-1, 1)$. Otherwise, rewrite the autoregression equation as

$$\begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \cdots & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In other words and with obvious notation, one has $\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$ where the matrix \mathbf{A} is

$$\mathbf{A} = \begin{pmatrix} \phi_1 & \cdots & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

This matrix is essentially the companion matrix of the polynomial $P(z) = 1 - \sum_{j=1}^p \phi_j z^j$. It is a standard exercise in linear algebra to show that \mathbf{A} has characteristic polynomial

$$\det(\lambda \mathbf{I}_p - \mathbf{A}) = \lambda^p - \sum_{j=1}^p \phi_j \lambda^{p-j} = \lambda^p P(1/\lambda).$$

Since P has no root z such that $|z| \leq 1$, all eigenvalues of \mathbf{A} must have a modulus smaller than 1, and thus (\mathbf{A}^n) converges to the zero matrix geometrically fast as $n \rightarrow \infty$. One then concludes similarly by induction. Since the ε_t are independent and satisfy $\mathcal{C}_1(\gamma)$, we find

$$\max_{p+1 \leq t \leq n} |Y_t| = O_{\mathbb{P}} \left(\max_{p+1 \leq t \leq n} |\varepsilon_t| \right) = O_{\mathbb{P}}(n^{\gamma+\delta}) \text{ for all } \delta > 0,$$

by the balanced tails condition, combined with Theorem 1.1.6 and Lemma 1.2.9 in de Haan and Ferreira (2006), and Potter bounds (see *e.g.* Proposition B.1.9.5 in de Haan and Ferreira, 2006). Therefore

$$\max_{p+1 \leq t \leq n} \frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1 + |\varepsilon_t|} = O_{\mathbb{P}}(n^{\gamma-1/2+\delta}).$$

Conclude by combining the \sqrt{n} -consistency of the estimators $\widehat{\phi}_{j,n}$, Theorem 1 and the convergence $\xi_{\tau_n}(\varepsilon) \rightarrow \infty$.

(ii) This is shown by combining (i) with Theorem 3. □

Proof of Corollary 4. (i) Recall that $\widehat{\alpha}_n$ and the $\widehat{\beta}_{j,n}$ are consistent estimators of (strictly) positive parameters, and thus are positive with arbitrarily high probability as $n \rightarrow \infty$. Let $\sigma^2 = \alpha + \sum_{j=1}^p \beta_j y_{-j}^2$, $\widehat{\sigma}_n^2 = \widehat{\alpha}_n + \sum_{j=1}^p \widehat{\beta}_{j,n} y_{-j}^2$ and write

$$\begin{aligned} \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})}{\xi_{\tau_n}(Y_0 | y_{-1}, \dots, y_{-p})} - 1 \right) &= \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \\ &\quad + \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\sigma}_n}{\sigma} - 1 \right) \frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}. \end{aligned}$$

Since, for any choice of y_{-1}, \dots, y_{-p} ,

$$\begin{aligned} &\left| \frac{(\widehat{\alpha}_n + \sum_{j=1}^p \widehat{\beta}_{j,n} y_{-j}^2)^{1/2}}{(\alpha + \sum_{j=1}^p \beta_j y_{-j}^2)^{1/2}} - 1 \right| \\ &= \frac{|\widehat{\alpha}_n - \alpha| + \sum_{j=1}^p |\widehat{\beta}_{j,n} - \beta_j| y_{-j}^2}{(\alpha + \sum_{j=1}^p \beta_j y_{-j}^2)^{1/2} [(\widehat{\alpha}_n + \sum_{j=1}^p \widehat{\beta}_{j,n} y_{-j}^2)^{1/2} + (\alpha + \sum_{j=1}^p \beta_j y_{-j}^2)^{1/2}]} \\ &\leq \frac{|\widehat{\alpha}_n - \alpha|}{\alpha} + \frac{|\widehat{\beta}_{1,n} - \beta_1|}{\beta_1} + \dots + \frac{|\widehat{\beta}_{p,n} - \beta_p|}{\beta_p} = O_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

we get $\widehat{\sigma}_n/\sigma - 1 = O_{\mathbb{P}}(n^{-1/2})$, and we also find

$$\max_{p+1 \leq t \leq n} \frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1 + |\varepsilon_t|} = O_{\mathbb{P}}(n^{-1/2}).$$

Apply Theorem 1 to complete the proof of (i).

(ii) This is shown by combining (i) with Theorem 3. □

C Appendix: methodology of the simulation section and additional results on real data

In this section, we give all necessary details about our methodology in the simulation studies and real data analyses. Throughout our numerical experiments, we use, as a way to estimate the tail index, the bias-reduced version of the Hill estimator of Gomes *et al.* (2016): this estimator $\widehat{\gamma}_k^{\text{RB}}$ is obtained from the Hill estimator $\widehat{\gamma}_k$ by setting

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

where \widehat{b} and $\widehat{\rho}$ are consistent estimators of the quantities b and ρ 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 Gomes *et al.* (2016); of course, we shall use here their versions built on residuals following the general ideas described in Sections 2 and 3. Our practical implementation uses the R function `mop` from the R package `evt0`. We also consider the following bias-reduced version of the family of direct extreme expectile estimators of ε :

$$\widehat{\xi}_{\tau'_n}^{\star, \text{RB}}(\varepsilon) = \widehat{\xi}_{\tau'_n}^{\star}(\varepsilon) \left(1 - \widehat{\gamma}_k^{\text{RB}} \frac{((\widehat{\gamma}_k^{\text{RB}})^{-1} - 1)^{-\widehat{\rho}} (1 - \widehat{\gamma}_k^{\text{RB}}) \widehat{b}}{\widehat{\rho} (1 - \widehat{\gamma}_k^{\text{RB}} - \widehat{\rho})} \left(\frac{n}{k} \right)^{\widehat{\rho}} + 2\widehat{\gamma}_k^{\text{RB}} \frac{k}{n} \right).$$

This is motivated by the proof of Proposition 1 and Corollary 1 in Daouia *et al.* (2018), which quantify the error made in the asymptotic proportionality relationship $\xi_\tau(\varepsilon)/q_\tau(\varepsilon) \rightarrow (\gamma^{-1} - 1)^{-\gamma}$. We similarly consider the following bias-reduced version of the family of indirect estimators:

$$\widetilde{\xi}_{\tau'_n}^{\star, \text{RB}}(\varepsilon) = \widetilde{\xi}_{\tau'_n}^{\star}(\varepsilon) \left(1 - \widehat{\gamma}_k^{\text{RB}} \frac{\widehat{b}}{\widehat{\rho}} \left(\frac{n}{k} \right)^{\widehat{\rho}} \right).$$

These procedures improve the accuracy of our estimators, without affecting their asymptotic properties (see Gomes *et al.*, 2016). They naturally give rise to estimators $\widehat{\xi}_{\tau'_n}^{\star, \text{RB}}(Y|\mathbf{x})$ and $\widetilde{\xi}_{\tau'_n}^{\star, \text{RB}}(Y|\mathbf{x})$, to which we refer in the present section.

C.1 Simulation study: linear and single-index models

We simulate here $N = 500$ replications of data sets made of $n = 1,000$ observations (\mathbf{X}_i, Y_i) , $1 \leq i \leq n$. We consider throughout this section that $\mathbf{X} \in \mathbb{R}^4$, 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 describing the relationship between Y and \mathbf{X} :

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

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

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 $\boldsymbol{\beta} = (1, 1, 1, 1)$ and ε is a noise variable, independent of \mathbf{X} , with a normalised symmetric Burr distribution, that is, $\varepsilon = \varepsilon_0 / [\Gamma(1 - \gamma)\Gamma(1 + \gamma)]$, where Γ is Euler's Gamma function and ε_0 has density

$$f_0(x) = \frac{1}{2} (1 + |x|^{-\rho/\gamma})^{1/\rho} \quad (x \in \mathbb{R}). \quad (47)$$

We consider $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$ and $\rho = -1$.

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 strategies:

- (S1) We assume that \mathbf{X} brings no information on Y , which is equivalent to assuming the null regression model $Y = \varepsilon$. In this null model, $\xi_{\tau_n}(Y|\mathbf{x}) = \xi_{\tau_n}(Y)$, estimated using the direct LAWS estimator of Daouia *et al.* (2018).
- (S2) We assume that Y is linked to \mathbf{X} by a location-scale shift linear regression model, *i.e.* $Y = \alpha + \boldsymbol{\beta}^\top \mathbf{X} + (1 + \boldsymbol{\theta}^\top \mathbf{X}) \varepsilon$. The methodology used for the estimation of $\xi_{\tau_n}(Y|\mathbf{x})$ is outlined in Section 3.1 (we only use the estimator $\widehat{\xi}_{\tau_n}^{\star, \text{RB}}(Y|\mathbf{x})$ here).
- (S3) We assume that Y is linked to \mathbf{X} by a heteroscedastic single index model in the spirit of Section 3.2, *i.e.* $Y = g(\boldsymbol{\beta}^\top \mathbf{X}) + \sigma(\boldsymbol{\beta}^\top \mathbf{X}) \varepsilon$. The vector $\boldsymbol{\beta}$ is estimated using the algorithm described in Zhu *et al.* (2013, see 1.(a)–(c) on page 1240 therein). The functions g and σ are estimated using the procedure described in Section 3.2, with $h_n = 0.3$ and $t_n = n^{2/5} \approx 15.85$ for the sake of computational efficiency. The extreme expectile is estimated using the bias-reduced direct estimator $\widehat{\xi}_{\tau_n}^{\star, \text{RB}}(Y|\mathbf{x})$.
- (S4) Identical to (S3), but the bias-reduced indirect estimator $\widetilde{\xi}_{\tau_n}^{\star, \text{RB}}(Y|\mathbf{x})$ is used instead.

We also compare these procedures with the following two benchmarks:

- (B1) We assume that Y is linked to \mathbf{X} by a homoscedastic single index model, *i.e.* $Y = g(\boldsymbol{\beta}^\top \mathbf{X}) + \varepsilon$. We consider here that $\boldsymbol{\beta}$ is known and equal to $(1, 1, 1, 1)$. The function g is estimated through a classical Nadaraya-Watson estimator, with a bandwidth chosen using the R package `np`, based on the methodology introduced in Härdle *et al.* (1993). The extreme expectile is then estimated using the direct estimator.
- (B2) Identical to (S3) above, although $\boldsymbol{\beta}$ is assumed to be known and equal to $(1, 1, 1, 1)$.

In each of these procedures, the intermediate expectile level used as an anchor in the tail index and extreme expectile estimators is fixed at 0.9, corresponding to $k_n = n(1 - \tau_n) = 100$. Moreover, in (S3), (S4), (B1) and (B2), we use the Epanechnikov kernel in the estimation of the link functions g and σ . In order to get a global idea of how our estimation methods behave, both over a wide range of extreme levels

and across the sample space, we compute, on each replication m and for a given estimator $\bar{\xi}_{\tau_n}^*(Y|\mathbf{x})$, the following relative mean-squared error:

$$\overline{\text{RMSE}}_m = \sqrt{\frac{1}{k_n n_p} \sum_{j=1}^{k_n} \sum_{i=1}^{n_p} \left(\frac{\bar{\xi}_{1-j/n}^{*,(m)}(Y|\mathbf{x}_i)}{\xi_{1-j/n}(Y|\mathbf{x}_i)} - 1 \right)^2}.$$

Here the points $\mathbf{x}_1, \dots, \mathbf{x}_{n_p}$ are (fixed across replications) independent and identically distributed realizations of a four-variate random vector $\mathbf{Z} \stackrel{d}{=} R\mathbf{D}$, where R is a uniform distribution on $(1, 3)$ and \mathbf{D} is a Dirichlet distribution of order 4 with all parameters equal to $1/4$. The rationale for this choice is that the $\beta^\top \mathbf{x}_i$, $1 \leq i \leq n_p$ are then uniformly distributed on $[1, 3]$, and so choosing a large n_p will provide a good representation of the global performance of our estimator when $\beta^\top \mathbf{x} \in [1, 3]$. We take $n_p = 1,000$ in what follows. The quantity $\bar{\xi}_{1-j/n}^{*,(m)}(Y|\mathbf{x}_i)$ denotes the estimator calculated on the m th replication, at the level $\tau_n = 1 - j/n$ and $\mathbf{x} = \mathbf{x}_i$. The quantity $\overline{\text{RMSE}}_m$ may thus be seen as a uniform error on both the central points of the random projection $\beta^\top \mathbf{X}$ and the tail of the conditional distribution $Y|\mathbf{X}$. Note also that for all $\alpha \in (0, 1)$, the true expectiles $\xi_\alpha(Y|\mathbf{x}_i)$ are directly 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 $\psi(y) = \int_y^\infty f_0(t)dt$ is computed with the R function `integrate`. The quantities $\overline{\text{RMSE}}_m$ are then averaged and listed as the RMSE based on the N replications in Table C.1 and Table C.2.

Procedure	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
(S1)	$3.58 \cdot 10^{-1}$	$3.52 \cdot 10^{-1}$	$3.42 \cdot 10^{-1}$	$3.28 \cdot 10^{-1}$
(S2)	$5.20 \cdot 10^{-2}$	$5.84 \cdot 10^{-2}$	$7.86 \cdot 10^{-2}$	$1.47 \cdot 10^{-1}$
(S3)	$9.04 \cdot 10^{-2}$	$9.43 \cdot 10^{-2}$	$1.13 \cdot 10^{-1}$	$1.47 \cdot 10^{-1}$
(S4)	$1.03 \cdot 10^{-1}$	$1.01 \cdot 10^{-1}$	$1.16 \cdot 10^{-1}$	$1.46 \cdot 10^{-1}$
(B1)	$1.53 \cdot 10^{-1}$	$1.62 \cdot 10^{-1}$	$1.77 \cdot 10^{-1}$	$1.96 \cdot 10^{-1}$
(B2)	$4.23 \cdot 10^{-2}$	$4.81 \cdot 10^{-2}$	$6.46 \cdot 10^{-2}$	$9.63 \cdot 10^{-2}$

Table C.1: Location-scale shift linear regression model (G1). Comparison of the average RMSE of estimation methods (S1)–(S4) and benchmarks (B1)–(B2).

Procedure	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
(S1)	$6.18 \cdot 10^{-1}$	$5.39 \cdot 10^{-1}$	$4.78 \cdot 10^{-1}$	$4.27 \cdot 10^{-1}$
(S2)	$1.09 \cdot 10^{-1}$	$1.05 \cdot 10^{-1}$	$1.35 \cdot 10^{-1}$	$2.28 \cdot 10^{-1}$
(S3)	$9.89 \cdot 10^{-2}$	$1.05 \cdot 10^{-1}$	$1.22 \cdot 10^{-1}$	$1.57 \cdot 10^{-1}$
(S4)	$1.15 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$	$1.27 \cdot 10^{-1}$	$1.58 \cdot 10^{-1}$
(B1)	$2.25 \cdot 10^{-1}$	$2.16 \cdot 10^{-1}$	$2.18 \cdot 10^{-1}$	$2.35 \cdot 10^{-1}$
(B2)	$5.23 \cdot 10^{-2}$	$5.86 \cdot 10^{-2}$	$7.74 \cdot 10^{-2}$	$1.12 \cdot 10^{-1}$

Table C.2: Single-index heteroscedastic model (G2). Comparison of the average RMSE of estimation methods (S1)–(S4) and benchmarks (B1)–(B2).

C.2 Simulation study: autoregressive time series models

In this section, we simulate $N = 1,000$ time series (Y_0, \dots, Y_n) of size $n + 1 = 1,001$. We consider two time series models:

- (T1) An AR(1) model $Y_t = \phi Y_{t-1} + \varepsilon_t$, where the parameter ϕ is estimated with the least squares estimator $\hat{\phi}_n = \sum_{t=1}^n Y_t Y_{t-1} / \sum_{t=1}^n Y_{t-1}^2$.
- (T2) An ARCH(1) model $Y_t = \sqrt{\alpha + \beta Y_{t-1}^2} \varepsilon_t$, where α and β are estimated with the weighted least squares estimators of Section 3.3.

In both models, the series ε_t is an independent sequence of innovations having the common density function f_0 as in (47), with $\rho = -1$; in the ARCH(1) model, these innovations are rescaled by $\sqrt{\Gamma(1 - 2\gamma)\Gamma(1 + 2\gamma)}$ to guarantee that $\mathbb{E}[\varepsilon^2] = 1$.

In each sample $(Y_0^{(j)}, \dots, Y_n^{(j)})$, the aim is to estimate a one-step ahead extreme expectile on observation time n ; letting $Y_n^{(j)} = y^{(j)}$, this translates, by stationarity, into estimating an extreme expectile $\xi_{\tau'_n}(Y_0 | Y_{-1} = y^{(j)})$. Following the methodology introduced in Section 3.3, we then compute the N direct expectile estimates $\hat{\xi}_{\tau'_n}^{\star, \text{RB}}(Y_0 | Y_{-1} = y^{(j)})$ and their indirect counterparts $\tilde{\xi}_{\tau'_n}^{\star, \text{RB}}(Y_0 | Y_{-1} = y^{(j)})$, where $\tau'_n = 1 - 5/n = 0.995$. In order to quantify the accuracy of these estimators, we calculate their Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N \left(\frac{\bar{\xi}_{\tau'_n}(Y_0 | Y_{-1} = y^{(j)})}{\hat{\xi}_{\tau'_n}^{\star, \text{RB}}(Y_0 | Y_{-1} = y^{(j)})} - 1 \right)^2}.$$

Here $\bar{\xi}_{\tau'_n} = \hat{\xi}_{\tau'_n}^{\star, \text{RB}}$ or $\tilde{\xi}_{\tau'_n}^{\star, \text{RB}}$. In the AR(1) model, we take $\phi \in \{0.1, 0.3, 0.5\}$; in the ARCH(1) model, we fix either $\alpha = 1/2$ and take $\beta \in \{0.1, 0.3, 0.5\}$, or we fix $\beta = 0.5$ and take $\alpha \in \{0.1, 0.3, 0.5\}$. In each model, we take $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$. Note that the ARCH(1) model is indeed strictly stationary in such cases since $\beta < 1$: see Remark 2.2 p.24 in Francq and Zakoïan (2010). The obtained RMSE are given in Tables C.3 and C.4.

ϕ	Method	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
0.1	Direct estimator	$4.45 \cdot 10^{-2}$	$8.89 \cdot 10^{-2}$	$1.70 \cdot 10^{-1}$	$3.47 \cdot 10^{-1}$
	Indirect estimator	$3.66 \cdot 10^{-2}$	$7.46 \cdot 10^{-2}$	$1.56 \cdot 10^{-1}$	$3.01 \cdot 10^{-1}$
0.3	Direct estimator	$5.05 \cdot 10^{-2}$	$9.63 \cdot 10^{-2}$	$1.83 \cdot 10^{-1}$	$3.73 \cdot 10^{-1}$
	Indirect estimator	$4.11 \cdot 10^{-2}$	$7.94 \cdot 10^{-2}$	$1.66 \cdot 10^{-1}$	$3.23 \cdot 10^{-1}$
0.5	Direct estimator	$1.26 \cdot 10^{-1}$	$1.18 \cdot 10^{-1}$	$1.88 \cdot 10^{-1}$	$3.60 \cdot 10^{-1}$
	Indirect estimator	$1.10 \cdot 10^{-1}$	$9.88 \cdot 10^{-2}$	$1.73 \cdot 10^{-1}$	$3.15 \cdot 10^{-1}$

Table C.3: RMSE of the direct and indirect expectile estimators in the AR(1) model for $\phi \in \{0.1, 0.3, 0.5\}$ and $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$.

β	Method	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
0.1	Direct estimator	$3.62 \cdot 10^{-2}$	$8.67 \cdot 10^{-2}$	$1.79 \cdot 10^{-1}$	$3.50 \cdot 10^{-1}$
	Indirect estimator	$2.44 \cdot 10^{-2}$	$7.21 \cdot 10^{-2}$	$1.63 \cdot 10^{-1}$	$3.04 \cdot 10^{-1}$
0.3	Direct estimator	$3.64 \cdot 10^{-2}$	$8.71 \cdot 10^{-2}$	$1.75 \cdot 10^{-1}$	$5.12 \cdot 10^{-1}$
	Indirect estimator	$2.44 \cdot 10^{-2}$	$7.27 \cdot 10^{-2}$	$1.61 \cdot 10^{-1}$	$5.57 \cdot 10^{-1}$
0.5	Direct estimator	$3.66 \cdot 10^{-2}$	$8.77 \cdot 10^{-2}$	$1.81 \cdot 10^{-1}$	$4.29 \cdot 10^{-1}$
	Indirect estimator	$2.43 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$1.68 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$

α	Method	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
0.1	Direct estimator	$3.66 \cdot 10^{-2}$	$8.77 \cdot 10^{-2}$	$1.81 \cdot 10^{-1}$	$4.29 \cdot 10^{-1}$
	Indirect estimator	$2.43 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$1.68 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$
0.3	Direct estimator	$3.66 \cdot 10^{-2}$	$8.77 \cdot 10^{-2}$	$1.81 \cdot 10^{-1}$	$4.29 \cdot 10^{-1}$
	Indirect estimator	$2.43 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$1.68 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$
0.5	Direct estimator	$3.66 \cdot 10^{-2}$	$8.77 \cdot 10^{-2}$	$1.81 \cdot 10^{-1}$	$4.29 \cdot 10^{-1}$
	Indirect estimator	$2.43 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$1.68 \cdot 10^{-1}$	$4.62 \cdot 10^{-1}$

Table C.4: RMSE of the direct and indirect expectile estimators in the ARCH(1) model for $\gamma \in \{0.1, 0.2, 0.3, 0.4\}$. Top table: case $\alpha = 0.5$ and $\beta \in \{0.1, 0.3, 0.5\}$. Bottom table: case $\beta = 0.5$ and $\alpha \in \{0.1, 0.3, 0.5\}$

C.3 Further results on real data

This section contains two figures, one on the Vehicle Customer Insurance data, the other on the AUD/CHF exchange rate. Both figures consist of a plot of the bias-reduced Hill estimator and a diagnostic QQ-plot to check that the heavy-tailed assumption makes sense for these sets of data. It should also be noted that the formula for the bias-reduced Weissman quantile estimator used in the application to the AUD/CHF exchange rate is

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

where $\tilde{q}_{\tau'_n}^*(\varepsilon)$ is the standard Weissman quantile estimator (Weissman, 1978) based on residuals and using $\hat{\gamma}_k^{\text{RB}}$ in its estimation of the tail index.

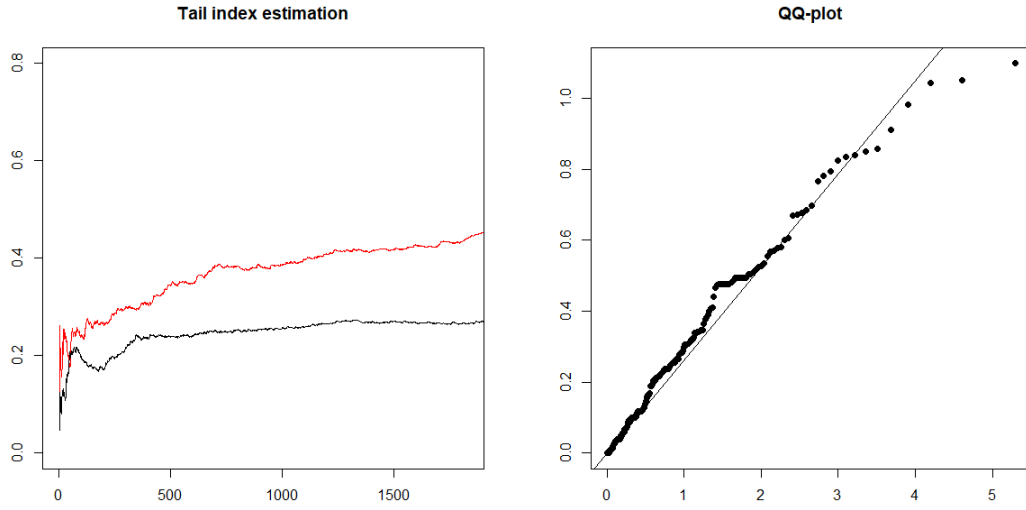


Figure C.1: Vehicle Customer Insurance data. Left panel: Bias-reduced Hill curves $k \mapsto \hat{\gamma}_k^{\text{RB}}$ on the non-filtered data Y_i (black curve) and residuals (red curve). Right panel: Exponential QQ-plot of the log-spacings $\log \left(\hat{\varepsilon}_{n-i+1,n}^{(n)} / \hat{\varepsilon}_{n-k^*+1,n}^{(n)} \right)$ with $k^* = 200$. The straight line has slope $\hat{\gamma}_{k^*}^{\text{RB}} = 0.263$.

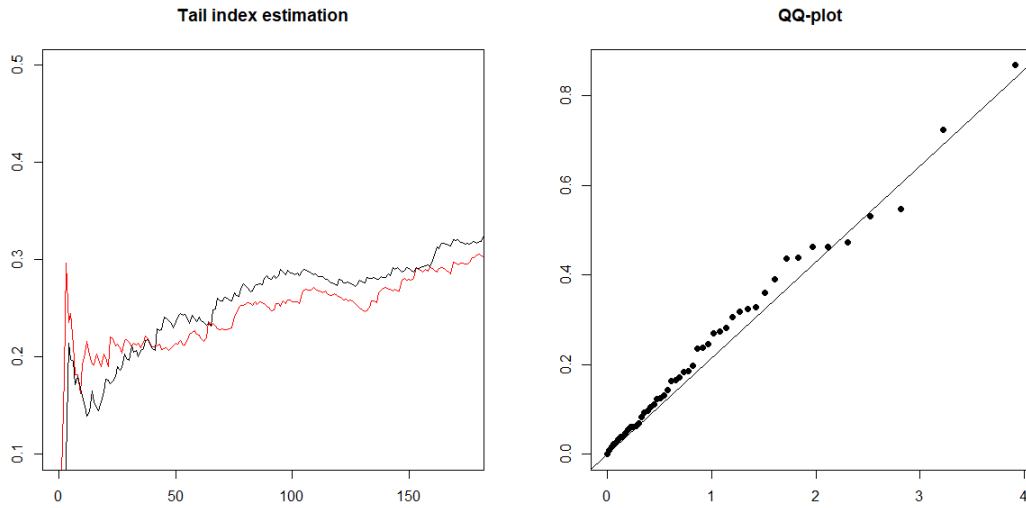


Figure C.2: AUD/CHF exchange rate data. Left panel: Bias-reduced Hill curves $k \mapsto \hat{\gamma}_k^{\text{RB}}$ on the non-filtered data Y_i (black curve) and residuals (red curve). Right panel: Exponential QQ-plot of the log-spacings $\log \left(\hat{\varepsilon}_{n-i+1,n}^{(n)} / \hat{\varepsilon}_{n-k^*+1,n}^{(n)} \right)$ with $k^* = 50$. The straight line has slope $\hat{\gamma}_{k^*}^{\text{RB}} = 0.214$.

References

- 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.
- Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment, *Bernoulli* **26**: 531–556.
- Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*, John Wiley & Sons.
- Gardes, L. and Stupfler, G. (2014). Estimation of the conditional tail index using a smoothed local Hill estimator, *Extremes* **17**: 45–75.
- Geyer, C.J. (1996). On the asymptotics of convex stochastic optimization, unpublished manuscript.
- 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.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer.
- Hansen, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory* **24**: 726–748.
- Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models, *Annals of Statistics* **21**: 157–178.
- Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity, technical report, University of Toronto.
- Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates, *Journal of Time Series Analysis* **17**: 571–599.
- Stupfler, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails, *Extremes* **22**: 749–769.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association* **73**: 812–815.
- Zhu, L., Dong, Y. and Li, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling, *Statistica Sinica* **23**: 1235–1255.