Ann Inst Stat Math (2013) 65:105–124
DOI 10.1007/s10463-012-0360-2

Estimation of central shapes of error distributions in linear regression problems

P. Y. Lai · Stephen M. S. Lee

Received: 22 April 2011 / Revised: 20 February 2012 / Published online: 17 April 2012

Abstract Consider a linear regression model subject to an error distribution which is symmetric about 0 and varies regularly at 0 with exponent ζ. We propose two estimators of ζ, which characterizes the central shape of the error distribution. Both methods are motivated by the well-known Hill estimator, which has been extensively studied in the related problem of estimating tail indices, but substitute reciprocals of small L_p residuals for the extreme order statistics in its original definition. The first method requires careful choices of p and the number k of smallest residuals employed for calculating the estimator. The second method is based on subsampling and works under less restrictive conditions on p and k. Both estimators are shown to be consistent for ζ and asymptotically normal. A simulation study is conducted to compare our proposed procedures with alternative estimates of ζ constructed using resampling methods designed for convergence rate estimation.

Keywords Centre exponent · L_p estimator · Regression · Subsampling

1 Introduction

Consider a random sample $(Y_1, X_1), \ldots, (Y_n, X_n)$ under a typical linear regression model $Y_i = X_i^{\mathrm{T}}\beta_0 + U_i$, $i = 1, \ldots, n$, where $(U_1, \ldots, U_n)$ and $(X_1, \ldots, X_n)$ denote two independent random samples drawn from the univariate distribution function F_U and the d-variate distribution function F_X, respectively, and β_0 is an unknown d-variate parameter in $\mathbb{R}^d$. Assume that F_U is symmetric about 0 and satisfies

P. Y. Lai · S. M. S. Lee (✉)
Department of Statistics and Actuarial Science,
The University of Hong Kong, Pokfulam Road, Hong Kong, China
e-mail: smslee@hku.hk

P. Y. Lai
e-mail: pylaipy@graduate.hku.hk

$$F_U(u) - F_U(0) = \mathrm{sgn}(u)|u|^{\zeta} L(|u|)/\zeta, \quad \text{for } |u| \leq C, \tag{1}$$

for some $\zeta, C > 0$ and nonnegative function L on $(0, \infty)$ slowly varying near 0. The parameter ζ, which we shall term the "centre exponent" following Rogers (2001), characterizes regular variation of $|F_U(u) - F_U(0)|$ at the origin and thus controls the shape of F_U near 0. The above specification of F_U encompasses a rich class of symmetric error distributions continuous at 0. They may have S- or inverse S-shapes according to $\zeta < 1$ or > 1, respectively, and may or may not possess densities. If F_U has a non-zero derivative at 0, then $\zeta = 1$ necessarily. Special parametric examples include normal and Laplace distributions, for which $\zeta = 1$, as well as reflected gamma and double Weibull distributions, for which ζ is given by the shape parameter. Modelling of the error distribution F_U has very often been engineered to accommodate different types of tail behaviour of the response variable, without paying much regard for the central shape of F_U. By the same token, while nonparametric estimators of tail indices abound in the literature, the problem of estimating centre exponents remains little studied.

We contend, however, that knowledge of ζ is of such practical import that a nonparametric method for its consistent estimation is worth developing. A consistent estimator of ζ has, first of all, a diagnostic value in assessing the suitability of the proposed regression model. While we incline in standard regression problems to the assumption of a finite positive error density at 0 so that $\zeta = 1$, an estimate of ζ significantly different from 1 suggests inadequacy of the assumption and sheds light on ways to refine our modelling strategy qualitatively. For example, a large value of ζ evinces possible bimodality of F_U, suggesting the existence of a mixture of clusters, each espousing a possibly different linear relation, so that some sort of stratification may be necessary for enhancing the interpretive value of the regression model. A small value of ζ, on the other hand, indicates a possible mixture of two error distributions differing in scale, thus calling into question the assumption of homoscedasticity. Such diagnostic value becomes yet more important when the regression model consists of more than one covariates so that standard scatter plots may fail to discern certain meaningful patterns in the observed data.

Accurate estimation of ζ also plays an indirect yet important role in efficient estimation of the regression coefficients β_0. It has been shown that knowledge of ζ, coupled with a judicious choice of p, may yield for the resulting L_p estimator of β_0 a convergence rate faster than $n^{1/2}$, that is the rate achieved by the conventional least squares method L_2: see Arcones (1998, 1999) and Lai and Lee (2005). Consequently, predictions obtained from such L_p estimates also enjoy a higher order of accuracy compared to those derived from the least squares method.

If the random errors U_i were observable, the problem of estimating ζ resembles one of estimating the tail index of the distribution of $|U_i|^{-1}$, which has regularly varying tails, based on the observed random sample $|U_1|^{-1}, \ldots, |U_n|^{-1}$. The pioneering work of Hill (1975) has inspired a rich literature on estimation of tail indices, which supplies a good source of solutions in this case. However, adaptation of the Hill estimator to the present context where the U_i are in fact unobserved poses two major difficulties. First, although the distributions of Y_i and U_i have tail indices closely related to each other so that Hill estimators of their tail indices can be based directly on the

observable Y_i, there is no trivial relationship between the distributions of $|U_i|^{-1}$ and $|Y_i|^{-1}$ which one can exploit to calculate an analogous Hill estimator of ζ based on the observable $|Y_i|^{-1}$ alone. For practical implementation of Hill's method, one should therefore substitute each U_i by an appropriate computable residual $\hat{U}_i = Y_i - X_i^{\mathrm{T}}\hat{\beta}$, for some estimate $\hat{\beta}$ of β_0. The second difficulty concerns the correct choice of $\hat{\beta}$. As we shall show in Sect. 2, consistency of the Hill estimator of ζ derived from $|\hat{U}_1|^{-1}, \ldots, |\hat{U}_n|^{-1}$ requires a reasonably fast, and ζ-specific, convergence rate of $\hat{\beta}$. The traditional L_1 or L_2 estimators may not be adequate for providing the required fast rates.

We propose in this paper two nonparametric methods for consistent estimation of ζ, and establish their asymptotic properties. Both of our proposed estimators apply Hill's procedure to reciprocals of residuals $|\hat{U}_i|^{-1}$, but in two different manners. In each case the residual is defined as $\hat{U}_i = \hat{U}_i(p) = Y_i - X_i^{\mathrm{T}}\hat{\beta}(p)$, where $\hat{\beta}(p)$ denotes the L_p estimate of β_0 for some $p \in (0, \infty)$. The first method calculates the Hill estimate of ζ from the k largest $|\hat{U}_i|^{-1}$, for some $k = o(n)$. We establish consistency and asymptotic normality of this estimate under $\zeta \neq 2$ and some ζ-specific conditions on the two tuning parameters p and k. The latter conditions require, for $\zeta < 2$, p be sufficiently small and k be sufficiently large. Our second method calculates the Hill estimate based on the ℓ largest values of a subset of size m taken from $\{|\hat{U}_1|^{-1}, \ldots, |\hat{U}_n|^{-1}\}$, for some $m = o(n)$ and $\ell = o(m)$. The estimate of ζ is then obtained by averaging the above Hill estimates over the $\binom{n}{m}$ distinct subsets or over some random or nonrandom collection, possibly with repetitions, of these subsets. This method yields consistent estimates under much less stringent conditions on the choices of p, m and ℓ, and is more convenient to implement in practice. In particular, use of L_p estimates $\hat{\beta}(p)$ in the calculation of the residuals $\hat{U}_i(p)$ can be confined to the conventional choices $p = 1$ and 2.

Empirical determination of k in the Hill estimator of a tail index has been a topic of intensive research. For example, Hall and Welsh (1985) estimate adaptively the tail index by selecting k optimally from an initial range of possible values. Hall (1990) and Caers and van Dyck (1999) apply the bootstrap to select k in the sense of minimizing the mean squared error (MSE). Beirlant et al. (1996) derive the best k by a weighted regression in a Pareto quantile plot. De Haan and Peng (1998) suggest to choose k that yields minimal asymptotic MSE. These methods can be adapted without difficulty to our present context to determine either k or ℓ empirically. Practical implementation of our two estimation methods is thus complicated mainly by the need to select an appropriate estimate $\hat{\beta}(p)$ in the calculation of residuals $\hat{U}_i(p)$. We suggest two simple computational algorithms to automatically select p, and examine the performance of the resulting Hill estimates of ζ through a simulation study.

Section 2 establishes consistency and asymptotic normality results relevant to our two proposed estimators. Section 3 describes computational algorithms for calculating the estimators. Section 4 reports simulation results and compares our proposed procedures with four alternative estimators obtained by resampling methods. Section 5 illustrates the applications of the methods to three real data sets. Section 6 concludes our findings. All technical details are deferred to Appendix.

2 Theory

Recall that the L_p estimator $\hat{\beta}(p)$ of β_0 is conventionally defined as the value of β which minimizes the criterion function $n^{-1}\sum_{i=1}^{n}|Y_i - X_i^{\mathrm{T}}\beta|^p$. Lai and Lee (2005) show under the general class of error distributions (1) that the precise mode of calculation of $\hat{\beta}(p)$ in fact depends on the values of ζ and p. For example, if $\zeta > 1$ and $p < 1$, $\hat{\beta}(p)$ should be found by locally maximizing, instead of globally minimizing, the criterion function. The following lemma recapitulates the findings of Lai and Lee (2005) which are useful for establishing the theory of our proposed estimators of ζ.

Lemma 1 *Assume that* $\mathbb{E}\,|U_1|^{\max\{2p-2,0\}} < \infty$, X_1 *has finite moments of all orders and* $\mathbb{P}(X_1^{\mathrm{T}}\beta = 0) < 1$ *for any nonzero* $\beta \in \mathbb{R}^d$. *Then, under the class of error distributions* (1), $n^{\gamma(p,\zeta)}l(n)(\hat{\beta}(p) - \beta_0) = O_p(1)$ *for some* $\gamma(p,\zeta) > 0$ *and some positive function* l *slowly varying at* ∞ *whenever*

(i) $p + \zeta \geq 2$ *and* $p \neq 1$, *in which case* $\gamma = 1/2$;
(ii) $p + \zeta < 2$, $2p + \zeta > 2$ *and* $\zeta \neq 1$, *in which case* $\gamma = 2^{-1}(p + \zeta - 1)^{-1}$;
(iii) $2p + \zeta \leq 2$ *and* $\zeta \neq 1$, *in which case* $\gamma = 1/\zeta$;
(iv) $p = 1$, $\zeta \geq 1$, *in which case* $\gamma = 1/(2\zeta)$;

provided that

- *under case* (i), $\hat{\beta}(p)$ *globally minimizes or locally maximizes* $n^{-1}\sum_{i=1}^{n}|Y_i - X_i^{\mathrm{T}}\beta|^p$ *according as* $p > 1$ *or* $p < 1$, *respectively;*
- *under cases* (ii) *and* (iii), $\hat{\beta}(p)$ *globally minimizes or locally maximizes* $n^{-1}\sum_{i=1}^{n}|Y_i - X_i^{\mathrm{T}}\beta|^p$ *according as* $\zeta < 1$ *or* $\zeta > 1$, *respectively;*
- *under case* (iv), $\hat{\beta}(p)$ *globally minimizes* $n^{-1}\sum_{i=1}^{n}|Y_i - X_i^{\mathrm{T}}\beta|^p$.

We see from Lemma 1 that for $\zeta > 2$, the fastest convergence rate $n^{1/2}$ of $\hat{\beta}(p)$ can be attained by any $p \neq 1$; whilst for $\zeta < 2$ and $\zeta \neq 1$, the conventional $n^{1/2}$ rate can be improved upon by choosing $p < 2 - \zeta$ and the best rate, of order $n^{1/\zeta}$ up to a slowly varying factor, is attained by $p \leq 1 - \zeta/2$. The latter result is particularly important to ensure reasonable convergence rates for $\hat{\beta}(p)$ in our first estimation method when $1 \neq \zeta < 2$. For in this case we require the residual $\hat{U}_i(p)$ and the true error U_i be equal up to an order smaller than the conventional $O(n^{-1/2})$. The case $p < 1 = \zeta$ is more intricate in that the asymptotic behaviour of $\hat{\beta}(p)$ depends further on the second-order properties of F_U in the neighbourhood of 0. Details can be found in Lai and Lee (2005) Theorem 2 and will not be presented here.

Lai and Lee (2008) develop a computational algorithm for automating the correct mode of calculation of $\hat{\beta}(p)$ as required by Lemma 1, without prior knowledge of the value of ζ. Suppose now that $\hat{\beta}(p)$ has been computed from the regression data $(Y_1, X_1), \ldots, (Y_n, X_n)$. Define $\hat{U}_i(p) = Y_i - X_i^{\mathrm{T}}\hat{\beta}(p)$. For each nonempty subset $\mathcal{M}$ of size m in $\mathcal{N} = \{1, \ldots, n\}$, denote by $\hat{U}_{\mathcal{M}}^{(1)}(p) \leq \cdots \leq \hat{U}_{\mathcal{M}}^{(m)}(p)$ the ordered sequence of the residuals $\{|\hat{U}_i(p)| : i \in \mathcal{M}\}$. Define, for $\ell \leq m$, the Hill estimator of ζ to be

$$\hat{\zeta}_{\mathcal{M}}(p,\ell) = \ell\left\{\sum_{i=1}^{\ell}\log\left[\hat{U}_{\mathcal{M}}^{(\ell+1)}(p)/\hat{U}_{\mathcal{M}}^{(i)}(p)\right]\right\}^{-1}.$$

We prove in Appendix our main theorems, which establish consistency and asymptotic normality of $\hat{\zeta}_{\mathcal{M}}(p, \ell)$ under two different scenarios: (a) $\mathcal{M} = \mathcal{N}$ and (b) $m = o(n)$, respectively. In addition to regular variation of $|F_U(u) - F_U(0)|$ at 0, a second-order variation condition is required for proving asymptotic normality:

$$\lim_{t\to\infty} \{L(x/t)/L(1/t) - 1\}/g(t) = K(1 - x^{\kappa})/\kappa, \quad \text{for all } x > 0, \tag{2}$$

for some $K \in \mathbb{R}$, $\kappa \geq 0$ and function g varying regularly at ∞ with exponent $-\kappa$ and satisfying $\lim_{t\to\infty} g(t) = 0$. Note that the condition (2) is equivalent to requiring that the function $x \mapsto F(1/x) - F(0)$ be second-order $(-\zeta, -\kappa)$ regularly varying at ∞. Second-order regular variation conditions on the tails of distributions have been considered by, for example, Geluk and de Haan (1987); Goldie and Smith (1987), De Haan and Stadtmüller (1996) and Cheng and Pan (1998). Asymptotic normality of the Hill estimator of the tail index has been proved by many authors: see, for example, Hall (1982) under a special form of second-order regular variation, and Haeusler and Teugels (1985) under more general conditions. Cheng and Pan (1998) establish Edgeworth expansions for the Hill estimator under second-order regular variation conditions as well as under the stronger condition that the variation occurs at a rapid rate. Translated to our present context, their latter condition has the form:

$$\lim_{t\to\infty} \{L(x/t)/L(1/t) - 1\}/g(t) = 0, \quad \text{for all } x > 0, \tag{3}$$

for every function g varying regularly at ∞ with nonpositive exponent.

Theorem 1 *Assume the conditions of Lemma 1, (2) and that $k \propto n^{\Delta}$ for some $\Delta > 0$. Then*

$$k^{1/2}\left\{\hat{\zeta}_{\mathcal{N}}(p, k) - \zeta\right\} \text{ converges in distribution to } N(0, \zeta^2),$$

provided that

(i) $0 < \zeta < 2$, $0 < p < 2 - \zeta$ *and*

$$\Delta \in \begin{cases} \left(\dfrac{1 - 2^{-1}(p + \zeta - 1)^{-1}\zeta}{1 - \max(\zeta, 1)/2},\ 2\kappa/(\zeta + 2\kappa)\right) & \text{if } p > 1 - \zeta/2; \\ (0,\ 2\kappa/(\zeta + 2\kappa)) & \text{if } p \leq 1 - \zeta/2; \end{cases}$$

or

(ii) $\zeta > 2$, $0 < p \neq 1$ *and* $0 < \Delta < 2\kappa/(\zeta + 2\kappa)$.

In particular, if (3) holds, then the upper bound $2\kappa/(\zeta + 2\kappa)$ for Δ in (i) *and* (ii) *above can be replaced by* 1.

Theorem 1 implies consistency of $\hat{\zeta}_{\mathcal{N}}(p, k)$ at a convergence rate of order $k^{1/2}$ when $\zeta \neq 2$. The rate can be made as fast as $n^{\kappa/(\zeta+2\kappa)-\epsilon}$, for any $\epsilon > 0$, by choosing Δ sufficiently close to $2\kappa/(\zeta + 2\kappa)$ and p sufficiently small. The asymptotic normality result provides a convenient means to inference about ζ. In a different context Hall

(1982) establishes a similar result for Hill's estimation of tail indices when the distribution function F satisfies $1 - F(x) = Cx^{-\zeta}\{1 + Dx^{-\kappa} + o(x^{-\kappa})\}$ as $x \to \infty$ for some $C > 0$ and $D \in \mathbb{R}$, and proves that the Hill estimator has an optimal convergence rate of order $n^{\kappa/(\zeta+2\kappa)}$ if we set $k \propto n^{2\kappa/(\zeta+2\kappa)}$.

The upper bound $2\kappa/(\zeta + 2\kappa)$ on Δ is typically required for proving asymptotic normality even if the Hill estimator of ζ were calculated from the true errors U_i: see, for example, Hall (1982) and Haeusler and Teugels (1985). The other conditions are needed to close the gaps between the residuals $\hat{U}_i(p)$ and the U_i. When $\zeta = 2$, $\hat{\beta}(p)$ has convergence rates at most of order $n^{1/2}$ for any choice of p, which is not adequate to provide the order of proximity required between $\hat{U}_i(p)$ and U_i for establishing asymptotic normality. It is clear from Theorem 1(i) that choice of $p \leq 1 - \zeta/2$ broadens the feasible range of Δ to that specified in (ii), which suggests that a small p is recommendable in practice to permit more liberal choices of k. We describe in Sect. 3 a recursive algorithm for empirically determining valid values of p and k as stipulated by Theorem 1.

Unavailability of asymptotic properties at an isolated case $\zeta = 2$ and the need to search for a valid choice of p within some ζ-specific range pose two drawbacks in the computation of the estimate $\hat{\zeta}_{\mathcal{N}}(p, k)$. Indeed, if ζ is less than and close to 2, the conditions in Theorem 1(i) restrict the choice of p to a very narrow range $(0, 2-\zeta)$ and, if we wish to avoid an infeasible condition on k, to an even narrower range $(0, 1-\zeta/2]$. A simple trick to relax the constraints on p and k is to calculate an analogous estimate $\hat{\zeta}_{\mathcal{M}}(p, \ell)$ based on only m of the n residuals $\hat{U}_i(p)$ so that $\hat{U}_i(p)$ and U_i would be sufficiently close to each other relative to a sample of size m. The loss in efficiency due to use of only m residuals in the calculation of the estimate can be recovered to some extent by averaging the estimates over the $\binom{n}{m}$ distinct choices of the m residuals, or over a large number of randomly selected subsets of m residuals. Our next theorem states conditions under which $\ell^{1/2}(\hat{\zeta}_{\mathcal{M}}(p, \ell) - \zeta)$ is asymptotically normal.

Theorem 2 *Assume* (2) *and the conditions of Lemma* 1 *so that* $n^{\gamma(p,\zeta)}l(n)(\hat{\beta}(p) - \beta_0) = O_p(1)$ *for some constant* $\gamma(p, \zeta) > 0$ *and some positive function* l *slowly varying at infinity. Let* $\mathcal{M}$ *be a subset of size* m *arbitrarily selected from* $\{1, \ldots, n\}$. *Assume that* $m \propto n^{\rho}$ *and* $\ell \propto m^{\delta}$ *for some* $\rho, \delta \in (0, 1)$. *Then*

$$\ell^{1/2}\left(\hat{\zeta}_{\mathcal{M}}(p, \ell) - \zeta\right) \text{ converges in distribution to } N(0, \zeta^2),$$

provided that

(i) $0 < \zeta < 1$, $0 < \delta < 2\kappa/(\zeta + 2\kappa)$ *and* $\rho(1 - \delta/2) < \gamma(p, \zeta)\zeta$; *or*
(ii) $\zeta \geq 1$, $0 < \delta < 2\kappa/(\zeta + 2\kappa)$ *and* $\rho(1 - \delta + \delta\zeta/2) < \gamma(p, \zeta)\zeta$.

In particular, if (3) *holds, then the upper bound* $2\kappa/(\zeta + 2\kappa)$ *for* δ *in* (i) *and* (ii) *above can be replaced by* 1.

The corollary below follows immediately from Theorem 2.

Corollary 1 *Under the conditions of Theorem* 2 *and the conditions of part* (i) *or* (ii) *specified therein, we have that*

$$B^{-1}\sum_{b=1}^{B}\hat{\zeta}_{\mathcal{M}_b}(p,\ell)=\zeta+O_p(\ell^{-1/2}),$$

for any B subsets $\mathcal{M}_1,\ldots,\mathcal{M}_B$, each of size m, drawn either randomly or nonrandomly from $\{1,\ldots,n\}$.

We see from Theorem 2(ii) that validity of the estimate $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ for the case $\zeta=2$ is ensured by simply setting $\delta<\kappa/(1+\kappa)$ and $\rho<1$ if a $n^{1/2}$-consistent estimator, such as $\hat{\beta}(p)$ with $p\neq 1$, is used to calculate the residuals. Thus, $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ does not have the first drawback of $\hat{\zeta}_{\mathcal{N}}(p,k)$. In general, we can easily find combinations of ρ and δ which meet the conditions specified in (i) or (ii) of Theorem 2. Suppose, for example, that we fix ρ and δ such that $\rho(1-\delta/2)<1/2$ and $\delta<2\kappa/(\zeta+2\kappa)$. Then if $\zeta<1$, the conditions in (i) are satisfied trivially by taking $p=1$ so that $\gamma(1,\zeta)=1/(2\zeta)$: see Lemma 1(ii). If $\zeta\geq 1$, then the conditions in (ii) hold for $p=2$ so that $\gamma(2,\zeta)=1/2$: see Lemma 1(i). One can therefore obtain a consistent estimate of ζ by calculating only L_1 or L_2 residuals, depending on whether $\zeta<1$ or $\zeta\geq 1$, respectively. The convergence rate of $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ is of order $n^{\rho\delta/2}$, which is generally slower than the best rate of $\hat{\zeta}_{\mathcal{N}}(p,k)$ calculated from the full set of residuals. For example, if we adopt the aforementioned scheme for setting p, ρ and δ, then the rate can be made as fast as $n^{\kappa/(2\zeta+2\kappa)-\epsilon}$, for any $\epsilon>0$, by choosing δ sufficiently close to $2\kappa/(\zeta+2\kappa)$ and ρ sufficiently close to $1/(2-\delta)$. Higher efficiency may result from averaging $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ over more subsets $\mathcal{M}$ drawn from $\mathcal{N}$, although we do not attempt to derive the optimal theoretical convergence rate in this case and are content to give a somewhat trivial rate $\ell^{1/2}$ in Corollary 1 which holds for averages over any number of subsets $\mathcal{M}$. We note also from the last assertions of Theorems 1 and 2 that under the more stringent condition (3), both estimates $\hat{\zeta}_{\mathcal{N}}(p,k)$ and $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ can be constructed to yield the best convergence rates of order $n^{1/2-\epsilon}$ for any $\epsilon>0$, since $\lim_{\kappa\to\infty}\kappa/(\zeta+2\kappa)=\lim_{\kappa\to\infty}\kappa/(2\zeta+2\kappa)=1/2$.

As in our first method, the choice of p in the construction of $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ still hinges upon the unknown ζ, but the dependence is, as has been illustrated above, much less delicate in this case and a judicious choice between L_1 and L_2 estimators suffices. The second method is therefore expected to enjoy a more stable performance in practice. An algorithm is described in Sect. 3 for implementing the method.

3 Computational algorithms

Motivated by Hill (1975) estimator of a tail index and our Theorems 1 and 2, we propose two computational algorithms for estimating the centre exponent ζ. Method I, described in Sect. 3.1, updates the choices of p and k recursively until the estimate $\hat{\zeta}_{\mathcal{N}}(p,k)$ satisfies conditions determined by the current values of p and k. Section 3.2 describes Method II which averages the estimates $\hat{\zeta}_{\mathcal{M}}(p,\ell)$ over a large number of subsamples randomly generated from the original sample, for $p=1$ or 2 and for some pre-determined values of m and ℓ.

We do not intend to provide a construction of an optimal estimator having the fastest convergence rate, which is outside the scope of this paper but clearly constitutes an

important problem for further investigation. We have seen that the convergence rates of $\hat{\zeta}_{\mathcal{N}}(p, k)$ and $\hat{\zeta}_{\mathcal{M}}(p, \ell)$ depend crucially on k and ℓ, respectively. Many procedures are available in the literature for optimal selection of the sample fraction in Hill estimation and can be adapted to our present context, although the mutual dependence between ζ, p and the sample fractions calls for a more sophisticated procedure. We emphasize instead here the importance of selecting an appropriate L_p estimate in the calculation of the residuals $\hat{U}_i(p)$. We see from Theorems 1 and 2 that the value of p generally determines a lower bound on the valid choice of k or ℓ. Both Methods I and II pay due regard to the choice of p, and provide an efficient means to identify feasible values of p and k (or ℓ) in order to yield a valid Hill estimator of ζ. On the other hand, the upper bounds on k and ℓ are controlled by both ζ and κ in the form of the conditions $\Delta, \delta < 2\kappa/(\zeta + 2\kappa)$. In special situations, the conditions can be quite trivial. For example, under (3), we can replace $2\kappa/(\zeta + 2\kappa)$ by 1. Hall (1982) describes common special cases, such as powers of smooth distributions and extreme value distributions, for which $\kappa = \zeta$ so that $2\kappa/(\zeta + 2\kappa) = 2/3$. In general, we can calculate some preliminary consistent estimates of ζ and κ from the sample as suggested by, for example, Beirlant et al. (1996), and incorporate the conditions $\Delta, \delta < 2\kappa/(\zeta + 2\kappa)$ into the algorithms of the two methods.

3.1 Method I

We have seen from Theorem 1 that consistency and asymptotic normality of $\hat{\zeta}_{\mathcal{N}}(p, k)$ follow from a set of conditions as summarized below:

(a) if $\zeta \leq 1$, then we need $p < 2 - \zeta$ and

$$\Delta \in \left(\max \left\{ 2 - \frac{\zeta}{p + \zeta - 1}, 0 \right\}, \frac{2\kappa}{\zeta + 2\kappa} \right);$$

(b) if $1 < \zeta < 2$, then we need $p < 2 - \zeta$ and

$$\Delta \in \left(\max \left\{ \frac{2 - \zeta/(p + \zeta - 1)}{2 - \zeta}, 0 \right\}, \frac{2\kappa}{\zeta + 2\kappa} \right);$$

(c) if $\zeta > 2$, then we need $p \neq 1$ and $\Delta \in (0, 2\kappa/(\zeta + 2\kappa))$.

Denote by $[\cdot]$ the integer part function. Our Method I identifies a feasible triplet $(p, \Delta, \hat{\zeta}_{\mathcal{N}}(p, [n^{\Delta}]))$ which satisfies one of the above conditions. To start the algorithm, initial choices of p and Δ are required and determined by an initial consistent estimate $\hat{\zeta}_A$ of ζ calculated using any valid method such as those described in Sect. 4. Then we update the values of $(p, \Delta, \hat{\zeta}_{\mathcal{N}}(p, [n^{\Delta}]))$ until one of the conditions (a)–(c) is satisfied. The algorithm goes as follows:

Step 1. Calculate an initial consistent estimate $\hat{\zeta}_A$. Set $i = 0$, $\hat{\zeta}_0 = \hat{\zeta}_A$ and fix some $\omega_1, \omega_2 \in (0, 1)$.

Step 2. Increment i. Update p_i and Δ_i according to the following rules:

(i) if $\hat{\zeta}_{i-1} < 2$, calculate $p_i = \omega_1(2 - \hat{\zeta}_{i-1})$ and $k_i = [n^{\Delta_i}]$, where

$$\Delta_i = \begin{cases} \omega_2 \max\left\{2 - \hat{\zeta}_{i-1}/(p_i + \hat{\zeta}_{i-1} - 1), 0\right\} + (1 - \omega_2), & \hat{\zeta}_{i-1} \le 1, \\ \omega_2 \max\left\{\dfrac{2 - \hat{\zeta}_{i-1}/(p_i + \hat{\zeta}_{i-1} - 1)}{2 - \hat{\zeta}_{i-1}}, 0\right\} + (1 - \omega_2), & \hat{\zeta}_{i-1} > 1; \end{cases}$$

(ii) if $\hat{\zeta}_{i-1} > 2$, fix $p_i = 2$ and $k_i = [n^{1-\omega_2}]$.

Step 3. Compute $\hat{\zeta}_i = \hat{\zeta}_{\mathcal{N}}(p_i, k_i)$. Terminate and return $\hat{\zeta}_i$ as the estimate of ζ if $(p, k, \zeta) = (p_i, k_i, \hat{\zeta}_i)$ satisfies any one of the conditions (a)–(c); go to *Step 2* otherwise.

Determination of p_i in the above procedure differs under two different cases: (i) $\hat{\zeta}_{i-1} < 2$, and (ii) $\hat{\zeta}_{i-1} > 2$. For case (i), we recommend that p_i be restricted to small values by taking a small $\omega_1 < 0.5$ to capitalize on the accelerated convergence rate of $\hat{\beta}(p_i)$. For case (ii), we fix $p_i = 2$ and calculate the convenient least squares estimate $\hat{\beta}(2)$. In fact, any value of $p_i \neq 1$, which yields a convergence rate of order $n^{1/2}$, is appropriate here: see Lemma 1(i). We should, however, avoid the use of $p_i = 1$ due to the very slow rate $n^{1/(2\zeta)}$ of $\hat{\beta}(1)$: see Lemma 1(iv). The choice of Δ_i determines the number of small residuals used for calculating the Hill estimate, and controls the amount of information extracted from the centre of F_U relative to that from the tails. In general, a large Δ_i reduces variance but increases bias, in a way depending on the values of ζ and κ. For higher efficiency we are inclined to include more residuals in our calculations by fixing some small $\omega_2 < 0.5$, but are aware that ω_2 might need to be chosen adaptively according to the second-order behaviour of the slowly varying function L to account for the upper bound $2\kappa/(\zeta + 2\kappa)$ on Δ.

Our empirical experience finds that the algorithm converges after only a few, less than 6 on average, iterations, although it might occasionally oscillate between different regimes and fail to produce a stable estimate.

3.2 Method II

Our Method II essentially calculates the Hill estimator $\hat{\zeta}_{\mathcal{M}}(p, \ell)$ based on a subset of size $m = [n^{\rho}]$ of the original sample, for some $\rho \in (0, 1)$. Relative to the smaller sample size m, the estimator $\hat{\beta}(p)$, which is calculated from the full sample of size n, has a fast convergence rate. Thus the unobserved random errors are closer to the residuals derived from $\hat{\beta}(p)$ than to the L_p residuals which would have normally been obtained from a sample of size m. This results in substantially weakened conditions on the choice of p, making Method II much more convenient to implement than Method I. That the Hill estimates are calculated from subsamples generally leads to some loss in efficiency compared to Method I which calculates the estimate from the full sample. However, as we shall see in a simulation study, both methods yield quite accurate results and its practical convenience makes Method II a competitive alternative to Method I.

We have briefly mentioned a possible procedure for Method II in Sect. 2. The following algorithm formalizes the procedure further.

Step 1. Fix some ρ, δ satisfying $\rho(1-\delta/2) < 1/2$. Set $m = [n^\rho]$ and $\ell = [m^\delta]$.
Step 2. Calculate an initial consistent estimate $\hat{\zeta}_A$. If $\hat{\zeta}_A < 1$, set $p = 1$; otherwise set $p = 2$.
Step 3. Generate a large number B of subsets $\mathcal{M}_1, \ldots, \mathcal{M}_B$, each of size m, randomly from $\mathcal{N}$. For each $b = 1, \ldots, B$, calculate $\hat{\zeta}_{\mathcal{M}_b}(p, \ell)$.
Step 4. Calculate $\hat{\zeta} = B^{-1} \sum_{b=1}^{B} \hat{\zeta}_{\mathcal{M}_b}(p, \ell)$ and return $\hat{\zeta}$ as the estimate of ζ.

We note that in *Step 3*, if computer resources permit, we may set $B = \binom{n}{m}$ and calculate $\hat{\zeta}_{\mathcal{M}_b}(p, \ell)$ for all the $\binom{n}{m}$ distinct subsets $\mathcal{M}_b$ of size m selected from $\mathcal{N}$.

4 Simulation study

We conducted a simulation study under a location model $Y_i = \beta_0 + U_i$, where U_i has the density function

$$f_U(u) \propto \begin{cases} |u|^{\zeta-1}\mathbf{1}\{|u| \le 1\} & \text{with probability } 0.75, \\ |u|^{-3.01}\mathbf{1}\{|u| > 1\} & \text{with probability } 0.25, \end{cases}$$

and $\mathbf{1}\{\cdot\}$ denotes the indicator function. The true value β_0 was fixed at 0. Note that the specification of f_U above allows for a heavy-tailed component which accounts for 25 % of the complete distribution. A broad spectrum of symmetric error densities was chosen to exemplify a variety of shapes at the origin: (a) $\zeta = 0.3$, (b) $\zeta = 0.8$, (c) $\zeta = 1.3$, (d) $\zeta = 1.8$, (e) $\zeta = 2.3$, and (f) $\zeta = 3.8$. We approximated the MSE of each estimator by averaging over 1,000 random samples of size $n = 100$ drawn from each distribution considered above.

For comparison with our Methods I and II, we included in the study four other methods, all of which are motivated by the subsampling approach, introduced by Politis et al. (1999, Chapter 8.2), to estimating convergence rates under very general conditions. Lemma 1(ii) and (iv) shows that the L_1 estimator $\hat{\beta}(1)$ has a convergence rate of order $n^{1/(2\zeta)}$, up to a slowly varying factor, for $\zeta > 0$. We can therefore derive a consistent estimate of ζ via consistent estimation of this rate. Lai and Lee (2005) establish m out of n bootstrap consistency for L_p regression in the present context, which suggests an alternative, sampling-with-replacement, scheme to replace subsampling in the above approach. Arguing as in Politis et al. (1999, Chapter 8.2), the above estimates of ζ suffer from an error of order $o_p(1/\log n)$, which is clearly inferior to the error rates of our Methods I and II. We describe below the four methods briefly.

(i) Subsampling method based on estimated quantiles
Fix S subsample sizes $m_1 < \cdots < m_S$. For each $s = 1, \ldots, S$, draw B' without-replacement subsamples, each of size m_s, from $(Y_1, \ldots, Y_n)$ and calculate the L_1 estimate $\hat{\beta}_s^{\dagger(b)}(1)$ from the bth subsample, $b = 1, \ldots, B'$. Fix J arbitrary points $t_1, \ldots, t_J$. Calculate, for $s = 1, \ldots, S$ and $j = 1, \ldots, J$,

$$V_{s,j} = \inf \left\{ x : B'^{-1} \sum_{b=1}^{B'} 1\{|\hat{\beta}_s^{\dagger(b)}(1) - \hat{\beta}(1)| \le e^x\} \ge t_j \right\}.$$

An estimate of ζ is then given by

$$\hat{\zeta}_{SQ} = -\frac{\sum_{s=1}^{I}\left(\log m_s - S^{-1}\sum_{s=1}^{S}\log m_s\right)^2}{2\sum_{s=1}^{S}\left(V_{s,\cdot} - \bar{V}\right)\left(\log m_s - S^{-1}\sum_{s=1}^{S}\log m_s\right)},$$

where $\bar{V} = (SJ)^{-1}\sum_{s=1}^{S}\sum_{j=1}^{J} V_{s,j}$ and $V_{s,\cdot} = J^{-1}\sum_{j=1}^{J} V_{s,j}$. Details of the above procedure can be found in Politis et al. (1999, Chapter 8.2).

(ii) m out of n bootstrap method based on estimated quantiles
Proceed exactly as in Method (i) except that subsamples are replaced by with-replacement bootstrap samples of sizes $m_1, \ldots, m_S$. Denote by $\hat{\zeta}_{BQ}$ the resulting estimate.

(iii) Subsampling method based on estimated MSE
The procedure is similar to Method (i) except that the estimated quantiles $V_{s,j}$ are replaced by estimated MSEs. Following the notations in (i), an estimate of ζ is given by

$$\hat{\zeta}_{SM} = \frac{\sum_{s=1}^{S}(W_s - \bar{W})^2}{\sum_{s=1}^{S}(W_s - \bar{W})\log\left\{B'^{-1}\sum_{b=1}^{B'}(\hat{\beta}_s^{\dagger(b)}(1) - \hat{\beta}(1))^2\right\}},$$

where $W_s = \log(n/m_s)$ and $\bar{W} = S^{-1}\sum_{s=1}^{S} W_s$.

(iv) m out of n bootstrap method based on estimated MSE
Replace subsamples by m out of n bootstrap samples in Method (iii). This yields an alternative estimate, denoted by $\hat{\zeta}_{BM}$, of ζ. Lai and Lee (2008) have suggested this approach to adaptively select an optimal p for L_p regression.

In the study we set $S = 3$, $B' = 3{,}000$, $J = 30$, $t_j = [0.75 + 0.2j/J]$ for $j = 1, \ldots, J$, and the subsample sizes m_s were set to be 17, 31 and 56.

In both Methods I and II, we set the initial estimate $\hat{\zeta}_A = \hat{\zeta}_{BM}$. For Method I, we fixed $\omega_1 = 0.2$ and $\omega_2 = 0.4$. In cases where the algorithm did not converge to a stable estimate, we increased ω_1 and ω_2 in *Step 2* and repeated the algorithm until a stable estimate resulted. Throughout the study, we found that the algorithm eventually converged in approximately 95 % of the total number of trials. For Method II, we took $B = 3{,}000$, $\rho = 0.5113$ and $\delta = 0.8265$. The values of ρ and δ were chosen such that $\rho(1 - \delta/2) \approx 0.3$, $m = 10$ and $\ell = 6$.

Table 1 reports the MSEs of the various estimators constructed in the study. In general, the performances of the four resampling approaches (i)–(iv) were satisfactory for $\zeta < 1$ but deteriorated rapidly as ζ increased. For $\zeta > 2$, Method (iii) performed very poorly with extremely large MSEs. Our Methods I and II succeeded in substantially reducing the estimation error for large ζ. In contrary to asymptotic predictions, Method II outperformed Method I in all cases and was clearly the most accurate among all the methods under study.

Table 1 MSEs of various estimates of ζ under error densites (a)–(f)

Density	ζ	Method I	Method II	(i) $\hat{\zeta}_{SQ}$	(ii) $\hat{\zeta}_{BQ}$	(iii) $\hat{\zeta}_{SM}$	(iv) $\hat{\zeta}_{BM}$
(a)	0.3	3.903 [1.96]	0.007 [0.37]	0.008	0.017	0.010	0.004
(b)	0.8	0.381 [1.30]	0.018 [0.88]	0.039	0.042	0.069	0.034
(c)	1.3	0.134 [1.42]	0.068 [1.31]	0.165	0.165	0.281	0.166
(d)	1.8	0.264 [1.80]	0.191 [1.85]	0.580	1.112	1.578	0.410
(e)	2.3	0.494 [2.30]	0.333 [2.36]	1.435	3.456	755.722	0.847
(f)	2.8	0.702 [2.87]	0.527 [2.84]	2.665	6.853	399.604	2.447

Estimated mean values of the estimates calculated by Methods I and II are given in brackets

5 Applications to real data

We illustrate the applications of the two estimation methods with three real data examples, all of which concern simple linear regression of Y_i on a single covariate x_i. The results obtained for the examples exemplify, respectively, three distinct inferences drawn from the estimates of ζ, highlighting at the same time the practical relevance of centre exponents to regression problems.

We set as initial estimates for both procedures $\hat{\zeta}_A = \hat{\zeta}_{SM}$, which was calculated using $S = 3, m_1 = [n^{0.5}], m_2 = [n^{0.7}]$ and $m_3 = [n^{0.9}]$. The same number of subsamples was used for calculating both $\hat{\zeta}_{SM}$ and the Method II estimate, such that $B' = B$ was set to be 50,000 in example 2 and 5,000 in examples 1 and 3. Settings of the parameters ω_1, ω_2, ρ and δ followed those chosen in the simulation study of Sect. 4.

Example 1 The data set consists of $n = 39$ observations of two variables, the average hourly wage in dollars (Y_i) and the average highest grade of school completed (x_i), obtained from 39 demographic subgroups of households with a male head earning less than \$15,000 annually in 1966. The data are displayed in Example 1 of Fig. 1. Method I estimated ζ to be 0.40, based on the choices $p = 0.66$ and $k = 9$. Drawing on the result of Theorem 1, a p value can be calculated using normal approximation for testing $\zeta = 1$ against $\zeta < 1$, and found to be 2.5×10^{-6}, which implies a very significant departure of the central shape from the standard assumption of a finite positive error density at 0. Method II also returns a small estimate, 0.54, of ζ. It follows from Lemma 1 that L_p estimation of β_0 may be made more efficient than L_2 by choosing a small $p \leq 1 - \zeta/2$. For illustration we include two regression lines, obtained, respectively by L_2 and $L_{0.7}$ estimation, in Fig. 1. Our small estimates of ζ can be explained by the cluster of data points lying close to the regression lines, indicating a possibly infinite peak of the error distribution at 0, as would have been the case should ζ be smaller than 1.

Example 2 The second data set contains $n = 20$ observations on the tool life in hours (Y_i) and the lathe speed in revolutions per minute (x_i): see Example 2 of Fig. 1 for a scatter plot of the data. The centre exponent ζ was estimated to be 3.83, based on $p = 0.15$ and $k = 6$, using Method I, and 2.23 using Method II. Subjecting the first estimate to a test of $\zeta = 1$ against $\zeta > 1$ results in a small p value 0.035. According to Lemma 1, all L_p estimators of β_0 except L_1 have the same convergence rate of order

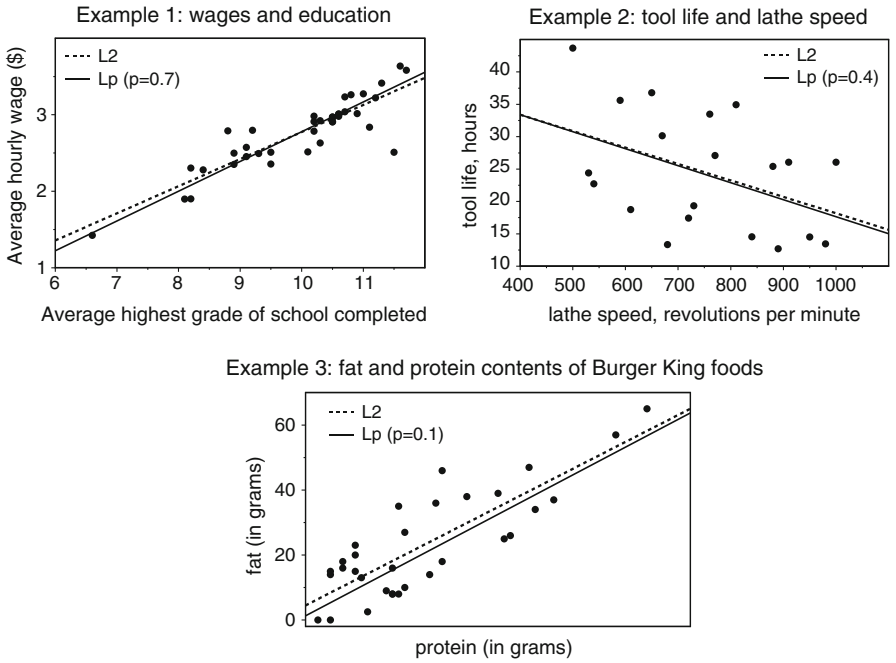

Fig. 1 Real data examples: *scatter plots* of regression data, fitted with L_2 (*dashed*) and L_p (*solid*) regression lines, where $p = 0.7$, 0.4 and 0.1 in examples 1, 2 and 3, respectively

$n^{1/2}$ if $\zeta > 2$, as can be illustrated by the proximity of the L_2 and $L_{0.4}$ regression lines shown in Fig. 1. It is evident that the regression lines classify the 20 observations into two distinct subgroups, suggesting some sparsity of data along the regression lines, as befits a centre exponent $\zeta > 1$.

Example 3 The third example concerns a simple linear regression model fitted to the amounts of fat (Y_i) and protein (x_i), both measured in grams, found in 32 foods sold at Burger King: see Example 3 of Fig. 1. Results obtained by the two methods differ somewhat in magnitude, with ζ estimated to be 0.91 using Method I and 1.77 using Method II. The two-sided p value associated with the first estimate for testing $\zeta = 1$ is found by normal approximation to be 0.80, based on the choices $p = 0.23$ and $k = 7$. Thus, we find no strong evidence against the standard assumption of a finite positive error density at 0. As in Example 1, the case $1 \neq \zeta < 2$ favours the use of a small $p \leq 1 - \zeta/2$ in L_p estimation of β_0 in order to achieve the optimal convergence rate. Figure 1 reveals a noticeable discrepancy between the L_2 and $L_{0.1}$ regression lines for this example, with the latter deemed more accurate by asymptotic theory. A comparison of the three examples in Fig. 1 shows that the relationship between the data pattern and the fitted regression lines in Example 3 occupies a somewhat middle ground between the two more extreme cases found for Examples 1 and 2, an observation corroborated by the numerical estimates obtained of ζ for the three examples.

The regression lines drawn in Fig. 1 help exemplify the numerical difference between the L_2 fit and the less conventional L_p fit based on an informed choice of p. Lai and Lee (2008) compare the two fits in a thorough simulation study, providing empirical evidence for the benefit of using an optimal p in L_p regression.

6 Conclusion and discussion

We have established consistency and asymptotic normality of two new versions of Hill estimators of the centre exponent ζ under a regression setup. Both estimators require careful choices of L_p estimators for calculating the residuals. Two methods for constructing the Hill estimators in practice have been described and investigated empirically. Method I automatically updates p and the sample fraction until they satisfy some ζ-specific conditions and a valid estimate of ζ is obtained. Success of this method hinges upon broadening of the range of p to include $p < 1$ on top of the more conventional choices within the interval $[1, 2]$. The broadened range equips L_p estimators with sufficiently fast convergence rates for yielding a reliable Hill estimator of ζ. Method II computes the Hill estimator from the ℓ smallest values of a subset of size m taken from the n absolute L_p residuals. It allows us to restrict the choice of p to either 1 or 2, and is therefore more convenient to implement in practice. Despite the efficiency loss due to the use of a smaller sample in the derivation of the Hill estimate, we have found from our simulation study that Method II actually gave the best performance among all the competing methods. We remark, however, that the number ℓ of extreme residuals involved in Method II is of the order $n^{\rho\delta}$, for some $\rho, \delta < 1$, which may turn out to be very small for even a moderate sample size n. We have found empirically that constraining $\ell \geq 6$ in practical applications yields quite satisfactory results in general. Three real data examples have been presented to further illustrate the practical relevance of our proposed estimators to standard linear regression problems.

We note lastly that the asymptotic properties of the two proposed estimators depend crucially on the assumption (1) as well as the validity of the regression function. Misspecification of the latter may strip the estimates of their intended meanings, rendering them almost irrelevant to the central shapes of the true error distribution.

7 Appendix

We first state a result which is an easy consequence of Proposition 2.1 of Resnick and Stărică (1997a). Define $\tilde{F}_U(x) = \mathbb{P}(|U_1|^{-1} \leq x)$ and $b(t) = 1/\tilde{F}_U^{-1}(1/2 + 1/(2t))$, where $\tilde{F}_U^{-1}(y) = \inf\{x : \tilde{F}_U(x) \geq y\}$. Note that $b(t)$ is regularly varying with exponent $1/\zeta$ at ∞. Define the process $\mathbb{W}_{k,n}(y) = k^{1/2}(k^{-1}\sum_{i=1}^{n} \mathbf{1}\{b(n/k)|U_i| \leq y^{1/\zeta}\} - y)$ for $y > 0$.

Lemma 2 *Assume* (1) *and that* $k \propto n^{\Delta}$ *with* $0 < \Delta < 2\kappa/(\zeta + 2\kappa)$. *Then* $\{\mathbb{W}_{k,n}(y) : y > 0\}$ *converges weakly to a standard Brownian motion* $\{W(y) : y > 0\}$ *in* $D[0, \infty)$.

Denote in what follows by $\|\cdot\|$ the Euclidean norm, and by C a generic positive constant which may differ from occurrence to occurrence.

7.1 Proof of Theorem 1

We start by proving that the process

$$\left\{\hat{\mathbb{W}}_{k,n}(y) : y > 0\right\} = \left\{k^{1/2}\left(k^{-1}\sum_{i=1}^{n}\mathbf{1}\{b(n/k)|\hat{U}_i(p)| \le y^{1/\zeta}\} - y\right) : y > 0\right\}$$

converges weakly to the standard Brownian motion $W(y)$, following the arguments used by Resnick and Stărică (1997b) for proving their Proposition 3.2. First we claim that for any interval $[c, d] \subset (0, \infty]$ and any $\epsilon > 0$,

$$\mathbb{P}\left(k^{-1/2}\sup_{x\in[c,d]}\left|\sum_{i=1}^{n}\mathbf{1}\{b(n/k)|\hat{U}_i(p)| < 1/x\} - \sum_{i=1}^{n}\mathbf{1}\{b(n/k)|U_i| < 1/x\}\right| > \epsilon\right) \to 0. \tag{4}$$

Set $\epsilon_n = \alpha k^{-1/2}$ for some sufficiently small $\alpha > 0$. That $Ia_1 \to 0$ in the aforementioned Proposition implies that

$$\mathbb{P}\left(k^{-1/2}\sup_{x\in[c,d]}\sum_{i=1}^{n}\mathbf{1}\{x^{-1}(1+\epsilon_n)^{-1} \le b(n/k)|U_i| < 1/x\} > \epsilon\right) \to 0. \tag{5}$$

Suppose that $r_n(\hat{\beta}(p) - \beta_0) = O_p(1)$ and $r_n = n^{\gamma(p,\zeta)}l(n)$ for some $\gamma(p, \zeta) > 0$ and slowly varying function l. Then, for any $M > 0$,

$$\begin{aligned}
&\mathbb{P}\left(k^{-1/2}\sup_{x\in[c,d]}\sum_{i=1}^{n}\mathbf{1}\{b(n/k)|U_i| < x^{-1}(1+\epsilon_n)^{-1},\ b(n/k)|\hat{U}_i(p)| \ge 1/x\} > \epsilon\right)\\
&\le \mathbb{P}(r_n\|\hat{\beta}(p) - \beta_0\| \ge M)\\
&\quad + \mathbb{P}\left(k^{-1/2}\sum_{i=1}^{n}\mathbf{1}\{\|X_i\| > r_n\epsilon_n(1+\epsilon_n)^{-1}(dM)^{-1}b(n/k)^{-1}\} > \epsilon\right)\\
&= I + II, \quad \text{say.}
\end{aligned}$$

It follows from Markov's inequality and existence of $\mathbb{E}\|X_1\|^\lambda$ for all $\lambda > 0$ that

$$II \le \epsilon^{-1}nk^{-1/2}\mathbb{P}(\|X_1\| > Cr_nk^{-1/2}b(n/k)^{-1}) \le Cnk^{-1/2}\{k^{1/2}b(n/k)/r_n\}^\lambda,$$

which converges to 0 provided that λ is sufficiently large and

$$\Delta/2 - \gamma(p, \zeta) + (1 - \Delta)/\zeta < 0. \tag{6}$$

Under (6), on letting $n \to \infty$ followed by $M \to \infty$, we have that both I and II converge to 0. This, together with (5), imply the claim (4). Under the further condition that

$$\Delta < 2\kappa/(\zeta + 2\kappa), \tag{7}$$

Lemma 2 and (4) imply the required weak convergence of $\hat{\mathbb{W}}_{k,n}$.

In what follows we argue as in the proof of Resnick and Stărică (1997a) Proposition 3.1 to establish asymptotic normality of $\hat{\xi}_{\mathcal{N}}(p,k)$. Write $\lceil y \rceil$ for the smallest integer $\geq y$. Replacing y by $\hat{U}_{\mathcal{N}}^{(\lceil ky \rceil)}(p)^{\zeta} b(n/k)^{\zeta}$ and noting weak convergence of $\hat{\mathbb{W}}_{k,n}(y)$, we obtain that

$$k^{1/2}\left(\hat{U}_{\mathcal{N}}^{(\lceil ky \rceil)}(p)^{\zeta} b(n/k)^{\zeta} - y\right) \xrightarrow{\mathcal{D}} -W(y). \tag{8}$$

Thus, for any $a > 1$,

$$\begin{aligned}
&-\int_{a^{-\zeta}}^{1} k^{-1/2} \sum_{i=1}^{n} \mathbf{1}\left\{\hat{U}_{\mathcal{N}}^{(k)}(p)/|\hat{U}_i(p)| \geq x^{-1/\zeta}\right\} x^{1/\zeta}\, \mathrm{d}x^{-1/\zeta} - k^{1/2}(1-a^{-\zeta})/\zeta \\
&= -\int_{a^{-\zeta}}^{1} \hat{\mathbb{W}}_{k,n}\left(x\, b(n/k)^{\zeta} \hat{U}_{\mathcal{N}}^{(k)}(p)^{\zeta}\right) x^{1/\zeta}\, \mathrm{d}x^{-1/\zeta} \\
&\quad + k^{1/2}\zeta^{-1}(1-a^{-\zeta})\left[b(n/k)^{\zeta} \hat{U}_{\mathcal{N}}^{(k)}(p)^{\zeta} - 1\right] \\
&\xrightarrow{\mathcal{D}} -\int_{a^{-\zeta}}^{1} W(x) x^{1/\zeta}\, \mathrm{d}x^{-1/\zeta} - \zeta^{-1}(1-a^{-\zeta}) W(1).
\end{aligned} \tag{9}$$

The last convergence follows from (8) and weak convergence of $\hat{\mathbb{W}}_{k,n}$. Next we claim that, for any $\epsilon > 0$,

$$\lim_{a\to\infty} \limsup_{n} \mathbb{P}\left(\int_{0}^{a^{-\zeta}} \left|\hat{\mathbb{W}}_{k,n}\left(x\, b(n/k)^{\zeta} \hat{U}_{\mathcal{N}}^{(k)}(p)^{\zeta}\right)\right| x^{1/\zeta}\, \mathrm{d}x^{-1/\zeta} > \epsilon\right) = 0. \tag{10}$$

The probability in (10) is bounded above by

$$\mathbb{P}\left(\int_{0}^{(1+\epsilon')a^{-\zeta}} \left|\hat{\mathbb{W}}_{k,n}(u)\right| u^{1/\zeta}\, \mathrm{d}u^{-1/\zeta} > \epsilon\right) + \mathbb{P}\left(\left|b(n/k)^{\zeta} \hat{U}_{\mathcal{N}}^{(k)}(p)^{\zeta} - 1\right| > \epsilon'\right)$$

$= III + IV$, say, for any $\epsilon' > 0$. That IV converges to 0 follows directly from (8). The probability III is bounded above by

$$\begin{aligned}
&\mathbb{P}\left(k^{-1/2} \int_{0}^{(1+\epsilon')a^{-\zeta}} \left|\sum_{i=1}^{n}\left[\mathbf{1}\{b(n/k)|\hat{U}_i(p)| \leq u^{1/\zeta}\} - \mathbf{1}\{b(n/k)|U_i| \leq u^{1/\zeta}\}\right]\right| \right. \\
&\qquad \left. \times\, u^{1/\zeta}\, \mathrm{d}u^{-1/\zeta} > \epsilon/2\right) + \mathbb{P}\left(\int_{0}^{(1+\epsilon')a^{-\zeta}} \left|\mathbb{W}_{k,n}(u)\right| u^{1/\zeta}\, \mathrm{d}u^{-1/\zeta} > \epsilon/2\right) \\
&= V + VI, \quad \text{say.}
\end{aligned}$$

The proof of Resnick and Stărică (1997a) Proposition 3.1 establishes immediately that

$$\lim_{a\to\infty} \limsup_n (VI) = 0. \tag{11}$$

Consider now

$$\begin{aligned}
&\mathbb{P}\left(k^{-1/2}\int_{(1+\epsilon')^{-1/\zeta}a}^{\infty}\sum_{i=1}^{n}\mathbf{1}\{b(n/k)|\hat{U}_i(p)|>1/x,\ b(n/k)|U_i|\le 1/x\}\,\mathrm{d}x/x>\epsilon\right)\\
&\le \mathbb{P}\left(k^{-1/2}\int_{(1+\epsilon')^{-1/\zeta}a}^{\infty}\sum_{i=1}^{n}\mathbf{1}\{x^{-1}(1+\epsilon_n)^{-1}\le b(n/k)|U_i|\le 1/x\}\,\mathrm{d}x/x>\epsilon/2\right)\\
&\quad+\mathbb{P}\left(k^{-1/2}\int_{(1+\epsilon')^{-1/\zeta}a}^{\infty}\sum_{i=1}^{n}\mathbf{1}\{b(n/k)|U_i|<x^{-1}(1+\epsilon_n)^{-1},\right.\\
&\qquad\left.b(n/k)|\hat{U}_i(p)|>1/x\}\mathrm{d}x/x>\epsilon/2\right)=VII+VIII \quad \text{say},
\end{aligned}$$

where $\epsilon_n=\alpha k^{-1/2}$. With α sufficiently small and a sufficiently large, we can ensure $|(1+\epsilon')a^{-\zeta}\zeta^{-1}k^{1/2}\{1-(1+\epsilon_n)^{-\zeta}\}|<\epsilon/4$, so that

$$\begin{aligned}
\lim_{a\to\infty}\limsup_n (VII) &\le \lim_{a\to\infty}\limsup_n \mathbb{P}\left(\int_0^{(1+\epsilon')a^{-\zeta}}\left|\mathbb{W}_{k,n}(u)\right.\right.\\
&\qquad\left.\left.-\mathbb{W}_{k,n}(u(1+\epsilon_n)^{-\zeta})\right|u^{1/\zeta}\mathrm{d}u^{-1/\zeta}>\epsilon/4\right)\\
&=0,
\end{aligned}\tag{12}$$

using arguments similar to those establishing (11). Writing $\epsilon''=(1+\epsilon')^{1/\zeta}a^{-1}$, we have, by integrating over x, Markov's inequality and for $M>0$,

$$VIII \le I + 2\epsilon^{-1}(J_1+J_2),$$

where $I=\mathbb{P}(r_n\|\hat{\beta}(p)-\beta_0\|\ge M)$ as before, which converges to 0 as $n\to\infty$ followed by $M\to\infty$,

$$\begin{aligned}
J_1 = nk^{-1/2}\mathbb{E}\left[\log\left\{\frac{|U_1|+\|X_1\|M/r_n}{(1+\epsilon_n)|U_1|}\right\};\right.\\
\left.|U_1|<\frac{\|X_1\|M}{r_n\epsilon_n},\ |U_1|+\frac{\|X_1\|M}{r_n}\le\frac{\epsilon''}{b(n/k)}\right]
\end{aligned}$$

and

$$J_2 = nk^{-1/2}\mathbb{E}\left[\log\left\{\frac{\epsilon''/b(n/k)}{(1+\epsilon_n)|U_1|}\right\};\right.$$
$$\left.(1+\epsilon_n)|U_1| < \frac{\epsilon''}{b(n/k)},\ |U_1| + \frac{\|X_1\|M}{r_n} > \frac{\epsilon''}{b(n/k)}\right].$$

Using the fact that $b(n/k) = o(\epsilon_n r_n)$, which follows from (6), we have, for sufficiently large n and any $\epsilon''' > 0$,

$$\begin{aligned} J_1 &\le Cnk^{-1/2}\mathbb{E}\left[\int_0^{M\|X_1\|/(r_n\epsilon_n)} \log\left(\frac{|u| + \|X_1\|M/r_n}{(1+\epsilon_n)|u|}\right) \mathrm{d}F(u)\right] \\ &\le Cnk^{-1/2}\mathbb{E}\left[(M\|X_1\|/r_n)^{\zeta+\epsilon'''}\int_{\epsilon_n}^{\infty} \log\left(\frac{1+v}{1+\epsilon_n}\right) v^{-\zeta-1-\epsilon'''}\,\mathrm{d}v\right] \\ &\le Cnk^{-1/2}r_n^{-\zeta-\epsilon'''} \times \begin{cases} 1, & \zeta+\epsilon''' < 1, \\ \log\epsilon_n, & \zeta+\epsilon''' = 1, \\ \epsilon_n^{-\zeta+1-\epsilon'''}, & \zeta+\epsilon''' > 1. \end{cases} \end{aligned}$$

It follows that $J_1 \to 0$ provided that

$$1 - \Delta/2 - \zeta\gamma(p,\zeta) < 0 \quad \text{if } \zeta < 1, \tag{13}$$

or that (6) holds if $\zeta \ge 1$. Noting that, under (6), the set $\{u : \epsilon''/b(n/k) - M\|x\|/r_n < |u| < (1+\epsilon_n)^{-1}\epsilon''/b(n/k)\}$ is empty for sufficiently large n and any fixed $x \in \mathbb{R}^d$, we have $J_2 = 0$ for sufficiently large n. Thus $VIII \to 0$ under conditions (6) and (13). This, together with (11) and (12), imply that $\lim_{a\to\infty}\limsup_n (V) = 0$ and that $\lim_{a\to\infty}\limsup_n (III) = 0$. This establishes our claim (10). It then follows from (9) and (10) that, on letting $a \to \infty$,

$$\begin{aligned} &-\int_0^1 k^{-1/2}\sum_{i=1}^n \mathbf{1}\left\{\hat{U}_{\mathcal{N}}^{(k)}(p)/|\hat{U}_i(p)| \ge x^{-1/\zeta}\right\} x^{1/\zeta}\,\mathrm{d}x^{-1/\zeta} - k^{1/2}/\zeta \\ &\xrightarrow{\mathcal{D}} -\int_0^1 W(x)x^{1/\zeta}\,\mathrm{d}x^{-1/\zeta} - \zeta^{-1}W(1) \overset{\mathcal{D}}{=} N(0,\zeta^{-2}). \end{aligned} \tag{14}$$

That $k^{1/2}\{\hat{\zeta}_{\mathcal{N}}(p,k) - \zeta\} \xrightarrow{\mathcal{D}} N(0,\zeta^2)$ now follows from (14), using the delta method and the fact that

$$\hat{\zeta}_{\mathcal{N}}(p,k)^{-1} = -\int_0^1 k^{-1/2}\sum_{i=1}^n \mathbf{1}\left\{\hat{U}_{\mathcal{N}}^{(k)}(p)/|\hat{U}_i(p)| \ge x^{-1/\zeta}\right\} x^{1/\zeta}\,\mathrm{d}x^{-1/\zeta}.$$

We now examine in detail the conditions (6) and (13). For $\zeta > 2$, we have $\gamma(p,\zeta) = 1/2$ for any $p \ne 1$: see Lemma 1(i). Then (6) reduces to $\Delta < 1$, which is trivial. For $\zeta < 2$, we have, using Lemma 1(ii) and (iii), that $\gamma(p,\zeta) = 2^{-1}(p+\zeta-1)^{-1}$ and

$1/\zeta$ for $1-\zeta/2 < p < 2-\zeta$ and $p \le 1-\zeta/2$, respectively. We then deduce from (6) and (13) the following lower bounds on Δ:

$$\Delta > \begin{cases} 0, & p \le 1-\zeta/2, \\ (1-\max(\zeta,1)/2)^{-1}\left\{1-2^{-1}(p+\zeta-1)^{-1}\zeta\right\}, & 1-\zeta/2 < p < 2-\zeta. \end{cases}$$

The above bounds and (7) give the conditions specified in Theorem 1(i) and (ii).

Under (3), the exponent κ in (2) can be made arbitrarily large, so that the upper bound $2\kappa/(\zeta+2\kappa)$ can be replaced by 1. This completes the proof of Theorem 1.

7.2 Proof of Theorem 2

The proof repeats the arguments for proving Theorem 1, with n, k replaced by m, ℓ, respectively. The convergence rate r_n of $\hat{\beta}(p)$ remains unchanged and has the form $n^{\gamma(p,\zeta)}l(n)$. Noting that $m \propto n^{\rho}$ and $\ell \propto n^{\rho\delta}$, the bounds (6) and (13) become

$$\rho\delta/2 - \gamma(p,\zeta) + \rho(1-\delta)/\zeta < 0 \tag{15}$$

and

$$\rho(1-\delta/2) - \zeta\gamma(p,\zeta) < 0 \quad \text{if } \zeta < 1, \tag{16}$$

respectively. Arguing as in the proof of Theorem 1, $\ell^{1/2}(\hat{\zeta}_{\mathcal{M}}(p,\ell)-\zeta) \xrightarrow{\mathcal{D}} N(0,\zeta^2)$ under conditions (7), (15) and (16), which implies parts (i) and (ii) of Theorem 2. The last assertion follows using the same arguments as given for proving the last part of Theorem 1.

References

Arcones, M. A. (1998). L_p-estimators as estimates of a parameter of location for a sharp-pointed symmetric density. *Scandinavian Journal of Statistics, 25*, 693–715.

Arcones, M. A. (1999). Rates of convergence of L_p-estimators for a density with an infinite cusp. *Journal of Statistical Planning and Inference, 81*, 33–50.

Beirlant, J., Vynckier, P., Teugels, J. L. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association, 91*, 1659–1667.

Caers, J., van Dyck, J. (1999). Nonparametric tail estimation using a double bootstrap method. *Computational Statistics & Data Analysis, 29*, 191–211.

Cheng, S., Pan, J. (1998). Asymptotic expansions of estimators for the tail index with applications. *Scandinavian Journal of Statistics, 25*, 717–728.

De Haan, L., Peng, L. (1998). Comparison of tail index estimators. *Statistica Neerlandica, 52*, 60–70.

De Haan, L., Stadtmüller, U. (1996). Generalized regular variation of second order. *Journal of the Australian Mathematical Society Series A, 61*, 381–395.

Geluk, J. L., de Haan, L. (1987). Regular variation, extensions and Tauberian theorems. *CWI Tract* (Vol. 40). Amsterdam: Centre for Mathematics and Computer Science.

Goldie, C. M., Smith, R. L. (1987). Slow variation with remainder: theory and applications. *Quarterly Journal of Mathematics, 38*, 45–71.

Haeusler, E., Teugels, J. L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics, 13*, 743–756.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society Series B, 44*, 37–42.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameters in nonparametric problems. *Journal of Multivariate Analysis, 32*, 177–203.

Hall, P., Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics, 13*, 331–341.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics, 3*, 1163–1174.

Lai, P. Y., Lee, S. M. S. (2005). An overview of asymptotic properties of L_p regression under general classes of error distributions. *Journal of the American Statistical Association, 100*, 446–458.

Lai, P. Y., Lee, S. M. S. (2008). Ratewise efficient estimation of regression coefficients based on L_p procedures. *Statistica Sinica, 18*, 1619–1640.

Politis, D. N., Romano, J. P., Wolf, M. (1999). *Subsampling*. New York: Springer.

Resnick, S., Stărică, C. (1997a). Smoothing the Hill estimator. *Advances in Applied Probability, 29*, 271–293.

Resnick, S., Stărică, C. (1997b). Asymptotic behaviour of Hill's estimator for autoregressive data. *Communications in Statistics: Stochastic Models, 13*, 703–721.

Rogers, A. J. (2001). Least absolute deviations regression under nonstandard conditions. *Econometric Theory, 17*, 820–852.