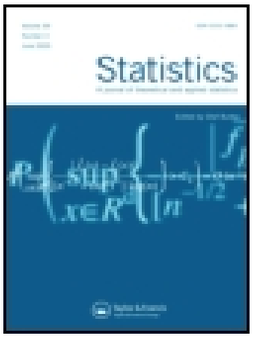

Statistics

A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: https://www.tandfonline.com/loi/gsta20

Hill estimator of projections of functional data on principal components

Mihyun Kim & Piotr Kokoszka

To cite this article: Mihyun Kim & Piotr Kokoszka (2019): Hill estimator of projections of functional data on principal components, Statistics

To link to this article: https://doi.org/10.1080/02331888.2019.1609476

Published online: 24 Apr 2019.

STATISTICS
https://doi.org/10.1080/02331888.2019.1609476

Taylor & Francis
Taylor & Francis Group

Hill estimator of projections of functional data on principal components

Mihyun Kim and Piotr Kokoszka

Department of Statistics, Colorado State University, Fort Collins, CO, USA

ABSTRACT
Functional principal component scores are commonly used to reduce mathematically infinitely dimensional functional data to finite dimensional vectors. In certain applications, most notably in finance, these scores exhibit tail behaviour consistent with the assumption of regular variation. Knowledge of the index of the regular variation, α, is needed to apply methods of extreme value theory. The most commonly used method of the estimation of α is the Hill estimator. We derive conditions under which the Hill estimator computed from the sample scores is consistent for the tail index of the unobservable population scores.

ARTICLE HISTORY
Received 14 June 2018
Accepted 8 April 2019

KEYWORDS
Functional data; Hill estimator; principal components

AMS CLASSIFICATIONS
62G32; 62H25

1. Introduction

A fundamental technique of functional data analysis is to replace infinite dimensional curves by coefficients of their projections onto suitable, fixed or data-driven systems, e.g. Bosq [1], Ramsay and Silverman [2], Horváth and Kokoszka [3], Hsing and Eubank [4]. A finite number of these coefficients encode the shape of the curves and are amenable to various statistical procedures. The best systems are those that lead to low dimensional representations, and so provide the most efficient dimension reduction. Of these, the functional principal components (FPCs) have been most extensively used, with hundreds of papers dedicated to various aspects of their theory and applications.

We assume that the random functions X_i are iid random elements of the Hilbert space $L^2 = L^2([0,1])$ with the inner product $\langle x, y\rangle = \int x(t)y(t)\,\mathrm{d}t$, which generates the norm $\|x\| = \sqrt{\langle x, x\rangle}$. If $E\|X_1\|^2 < \infty$, then

$$X_i(t) = \sum_{j=1}^{\infty} \xi_{ij} v_j(t), \quad E\xi_{ij}^2 = \lambda_j, \tag{1}$$

where v_j are the FPCs, the eigenfunctions of the *covariance operator* C defined by

$$x \ni L^2 \mapsto C(x) = E[\langle X_1, x\rangle X_1] \in L^2. \tag{2}$$

CONTACT Piotr Kokoszka piotr.kokoszka@colostate.edu

The covariance operator C is a positive integral Hilbert–Schmidt operator with the kernel $c(t,s) = E[X_1(t)X_1(s)]$, so v_j are defined explicitly by

$$\int c(t,s)v_j(s)\,\mathrm{d}s = \lambda_j v_j(t), \quad t \in [0,1],\ j = 1,2,\ldots$$

The random variables $\xi_{ij} = \langle X_i, v_j\rangle$ in (1) are called the scores of X_i with respect to v_j. They satisfy $E\xi_{ij} = 0$, $E[\xi_{ij}\xi_{ij'}] = 0$ if $j' \neq j$ and $E\xi_{ij}^2 = \lambda_j$.

The functions v_j and the variances λ_j are unknown parameters, which must be estimated. The sample covariance operator $\widehat{C}$ is defined as a kernel operator with the kernel $\hat{c}(t,s) = N^{-1}\sum_{n=1}^{N} X_n(t)X_n(s)$. The FPCs v_j and the eigenvalues λ_j are estimated by $\hat{v}_j$ and $\hat{\lambda}_j$ which satisfy

$$\int \hat{c}(t,s)\hat{v}_j(s)\,\mathrm{d}s = \hat{\lambda}_j\hat{v}_j(t). \tag{3}$$

The scores ξ_{ij} are then approximated by their sample counterparts $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j\rangle$.

In most inferential scenarios, replacing v_j by $\hat{v}_j$, and λ_j by $\hat{\lambda}_j$ is asymptotically negligible, see Yao et al. [5], Gabrys and Kokoszka [6], Berkes et al. [7], Horváth et al. [8,9], among dozens of recent papers by other authors. Even though many different inferential problems have been considered, they are all related to some form of inference for mean and covariance structures. In this paper, we study a totally different problem, the consistency of the Hill estimator, which is one of the most widely used tools of extreme value theory, see, e.g. Embrechts et al. [10], Beirlant et al. [11] and Resnick [12]. Its definition is given in Section 2. It is designed to estimate the tail index $\alpha > 0$ of a positive random variable, say Y, which satisfies $P(Y > x) \sim x^{-\alpha}$ (up to a slowly varying function). As argued above, in the context of functional data X_i, one often works with the projections, $\langle X_i, v_j\rangle$, $v_j \in L^2$. The question is whether the Hill estimator based on the estimated projections $\langle X_i, \hat{v}_j\rangle$, the only feasible estimator, can be used to estimate the tail index of the projections $\langle X_i, v_j\rangle$, assuming the latter have regular varying tail probabilities. A priori, there could be a systematic bias due to the effect of the estimation of the v_j by the $\hat{v}_j$. A problem of this type has not been studied. Consistency of the Hill estimator has been established in several settings, but always assuming that the observations (the $\langle X_i, \hat{v}_j\rangle$ in our case) have regularly varying tail probabilities. The projections onto the $\hat{v}_j$ can be expected to be only approximately regularly varying (because $\hat{v}_j$ is close to v_j), so none of the existing results can be used. A self-contained background on regular variation is presented in Appendix 1.

Even for samples of iid positive random variables, the consistency of the Hill estimator is far from trivial. The first proof in the iid setting was developed by Mason [13]. Hsing [14] introduced a general approach to establishing the consistency in case of dependent data, including both stationary times series and triangular arrays. The sample scores do form a triangular array, but we were unable to adapt Hsing's method to accommodate the transition from the sample scores to the unobservable population scores. We developed an approach based on the vague convergence of radon measures [12,15]. The Hill estimator for various stochastic models was studied by Resnick and Stărică [16,17] and Wang and Resnick [18].

The paper is organized as follows. In Section 2, we introduce the framework and state our main result, Theorem 2.1, which is proven in Section 4, after some preparation in

Section 3. To make the paper self-contained, Appendix 1 contains a minimal background on regular variation. Appendix 2 presents a motivating data example.

2. Assumptions and the main result

The most elegant, but in fact unnecessarily strong, assumption is that the function X whose copies X_i, $1 \le i \le n$, we observe is regularly varying in L^2. The space L^2 is infinitely dimensional and not locally compact, so we cannot define regular variation using the framework of Resnick [12,15], but we can use a similar and more general framework of Hult and Lindskog [19] who use M_0 convergence in place of the vague convergence in the Euclidean space with zero removed and compactified at infinity. Since we work with projections onto the real line, any definition of regular variation in L^2 which implies regular variation of these projections would work. According to Hult and Lindskog [19], a function X in L^2 (or any Banach space) is regularly varying with index $\alpha > 0$ if

$$P(\|X\| > u) = u^{-\alpha} L(u) \tag{4}$$

and

$$\frac{P(u^{-1}X \in \cdot)}{P(\|X\| > u)} \xrightarrow{M_0} \mu(\cdot), \quad u \to \infty, \tag{5}$$

where μ is a non-null measure (exponent measure) and L is a slowly varying function. There are several equivalent definitions, see Appendix 1, Chapter 2 of Meiguet [20] contains more details.

Set

$$U(u) = P(|\langle X, v\rangle| > u), \quad \widehat{U}(u) = P(|\langle X, \hat{v}\rangle| > u),$$

where v is one of the FPCs v_j in (1) and $\hat{v}$ its estimated defined by (3). The function U is regularly varying with index α, in the notation of Resnick [15], $U \in \mathrm{RV}_{-\alpha}$. To see this, consider the set $\mathcal{A}_v = \{x : |\langle x, v\rangle| > 1\}$, and observe that $|\langle X, v\rangle| > u$ iff $u^{-1}X \in A_v$. By (4) and (5),

$$\frac{U(tu)}{U(u)} = \frac{P((tu)^{-1}X \in \mathcal{A}_v)}{P(\|X\| > tu)} \frac{P(\|X\| > tu)}{P(\|X\| > u)} \frac{P(\|X\| > u)}{P(u^{-1}X \in \mathcal{A}_v)} \to t^{-\alpha},$$

provided $\mu(\mathcal{A}_v) > 0$. It cannot be expected that $\widehat{U} \in \mathrm{RV}_{-\alpha}$; for a fixed n, $\hat{v}$ is a random function whose distribution will, in general, influence the distribution of $\langle X, \hat{v}\rangle$. Only some form of asymptotic regular variation can be expected because $\widehat{U}$ approaches U, in several ways, as $n \to \infty$.

The same argument shows that if $\mu(\{x : \langle x, v\rangle > 1\}) > 0$, then the function $U_+(u) = P(\langle X, v\rangle > u)$ is in $\mathrm{RV}_{-\alpha}$, and if $\mu(\{x : \langle x, v\rangle < -1\}) > 0$, then $U_-(u) = P(\langle X, v\rangle < -u)$ is in $\mathrm{RV}_{-\alpha}$. To avoid repetitions of almost identical statements, we focus in the following on the estimation of the tail index of the function U. We will work under the following assumption.

Assumption 2.1: The functions $X_1, X_2, \ldots X_n$ are independent and have the same distribution as X. The function v is such that the function $U(u) = P(|\langle X, v\rangle| > u)$ is regularly varying with index $\alpha > 2, \alpha \neq 4$.

The assumption $\alpha > 2$ is needed because if $X \in \mathrm{RV}_{-\alpha}$ with $0 < \alpha < 2$, then, by (4), $E\|X\|^2 = \infty$, so the FPCs are not defined. If $\alpha = 2$, then either $E\|X\|^2 = \infty$ or $E\|X\|^2 < \infty$ are possible, and complex assumptions on the slowly varying functions L are needed to derive various rather technical results. We therefore assume $\alpha > 2$. Another phase transition occurs at $\alpha = 4$ separating, in a similar way, the cases with $E\|X\|^4 = \infty$ and $E\|X\|^4 < \infty$.

In our theory, the index α can depend on the direction v, but we do not emphasize it in our notation. We also note that even though the observed functions $X_1, X_2, \ldots X_n$ are iid, the sample scores $\langle X_i, \hat{v}\rangle$ are no longer independent because $\hat{v}$ depends on all $X_1, X_2, \ldots, X_n$. They form a triangular array of dependent random variables, which are identically distributed for each fixed n. The Hill estimator must be based on the projections $\langle X_i, \hat{v}\rangle$. Before recalling its definition, we introduce the following random variables:

$$Y = |\langle X, v\rangle|, \quad \widehat{Y} = |\langle X, \hat{v}\rangle|,$$
$$Y_i = |\langle X_i, v\rangle|, \quad \widehat{Y}_i = |\langle X_i, \hat{v}\rangle|.$$

This allows us to define

$$H_{k,n} = \frac{1}{k}\sum_{i=1}^{k-1} \ln Y_{(i)} - \ln Y_{(k)}, \quad \widehat{H}_{k,n} = \frac{1}{k}\sum_{i=1}^{k-1} \ln \widehat{Y}_{(i)} - \ln \widehat{Y}_{(k)},$$

with the convention that $Y_{(1)}$ is the largest order statistic. In the functional data context, $H_{k,n}$ is an infeasible Hill estimator because the FPC v is not observed; $\widehat{H}_{k,n}$ is the Hill estimator that can be actually computed. We want to establish condition under which it converges in probability to α^{-1}, where α is the index of regular variation of Y.

We further define

$$1 - F(u) = P(Y > u) = U(u), \quad b(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right).$$

We will use the representation

$$b(t) = t^{1/\alpha} L_b(t), \tag{6}$$

where L_b is a slowly varying function.

The approach in Chapter 4 of Resnick [12] is based on vague convergence to the measure on the positive half-line, which is defined by

$$\nu_\alpha(x, \infty] = x^{-\alpha}, \quad x > 0.$$

Our approach involves a sequence of 'increasingly empirical' measures, with only the last one being observable. We set

$$\nu_n = \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/b(n/k)}, \quad \nu_n^\star = \frac{1}{k}\sum_{i=1}^{n} I_{Y_i/Y_{(k)}}, \quad \nu_n^\dagger = \frac{1}{k}\sum_{i=1}^{n} I_{\widehat{Y}_i/b(n/k)}, \quad \hat{\nu}_n = \frac{1}{k}\sum_{i=1}^{n} I_{\widehat{Y}_i/\widehat{Y}_{(k)}}.$$

Any argument must involve some bounds on a suitable distance between $\widehat{Y}_i$ and Y_i. These will involve the covariance operator C and the sample covariance operator $\widehat{C}$. If v is the jth

eigenfunction of C and $\hat{v}$ is the jth eigenfunction of $\widehat{C}$, then (see, e.g. Lemma 2.3 in [3]),

$$\|\hat{v} - v\| \leq A_v \|\widehat{C} - C\|_{\mathcal{L}}, \tag{7}$$

where $\|\cdot\|_{\mathcal{L}}$ is the usual operator norm, $A_v = 2d_j^{-1}\sqrt{2}$ with

$$d_1 = \lambda_1 - \lambda_2, \quad d_j = \min\left\{\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}\right\},$$

assuming that the eigenvalue λ_j of C is such that $d_j > 0$. Since

$$|\widehat{Y}_i - Y_i| \leq |\left\langle X_i, \hat{v}_j - v_j\right\rangle| \leq \|X_i\| \|\hat{v}_j - v_j\|,$$

we conclude from (7) that

$$|\widehat{Y}_i - Y_i| \leq A_v \|X_i\| \|\widehat{C} - C\|_{\mathcal{L}}. \tag{8}$$

If $\alpha > 4$, then, see e.g. Theorem 2.5 in Horváth and Kokoszka [3],

$$E\|\widehat{C} - C\|_{\mathcal{L}}^2 = O(n^{-1}). \tag{9}$$

The case of regularly varying X with tail index $\alpha \in (2, 4)$ is studied in Kokoszka et al. [21]. Under weak conditions, Relation (9) must be replaced by

$$E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta} \leq L_\beta(n) n^{-\beta(1-2/\alpha)}, \quad \forall\, \beta \in (0, \alpha/2), \tag{10}$$

where L_β is a slowly varying function. For a fixed α, the strongest bound is obtained as $\beta \nearrow \alpha/2$, in which case $\beta(1 - 2/\alpha) \nearrow \alpha/2 - 1$. As $\alpha \nearrow 4$ and $\beta \nearrow \alpha/2$, relation (10) thus approaches, in a heuristic sense, relation (9). It is enough to impose a slightly weaker, but more convenient, condition:

$$E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta} = O\left(n^{-\kappa}\right), \quad \forall\, \beta \in \left(1, \frac{\alpha}{2}\right), \forall\, \kappa \in \left(0, \beta\left(1 - \frac{2}{\alpha}\right)\right). \tag{11}$$

The above discussion shows that the following Assumption 2.2 basically always holds as long as $d_j > 0$ in (7). We formulated it for ease of reference and to emphasize that only certain properties of the sample covariance operator $\widehat{C}$ are used; $\widehat{C}$ could, in principle, be a different estimator of C, which has those properties.

Assumption 2.2: Relation (7) holds. The estimator $\widehat{C}$ satisfies (9) if $\alpha > 4$ and (11) if $\alpha \in (2, 4)$.

Since Y_i are iid and in $\mathrm{RV}_{-\alpha}$, the only conditions needed to ensure that $H_{k,n} \xrightarrow{P} \alpha^{-1}$ are $k = k(n) \to \infty$ and $k/n \to 0$, as $n \to \infty$. In our setting, we want to estimate the index α of unobservable random variables Y_i based on their observed approximations $\widehat{Y}_i$. It can be expected that the rate of the approximation will impose additional conditions on the rate at which k tends to infinity with n. A sufficient condition is formulated in Assumption 2.3 below.

Define the function

$$\gamma(\alpha) = \begin{cases} \dfrac{\alpha-2}{2\alpha-2}, & \alpha \in (4,\infty), \\ \dfrac{1}{\alpha-1}, & \alpha \in (3,4], \\ 2-\dfrac{\alpha}{2}, & \alpha \in (2,3]. \end{cases} \tag{12}$$

Observe that $\gamma(\cdot)$ is continuous at $\alpha = 4$ with $\gamma(4) = 1/3$, and at $\alpha = 3$ with $\gamma(3) = 1/2$. It is increasing on $(4,\infty)$ with $\lim_{\alpha\nearrow\infty}\gamma(\alpha) = \frac{1}{2}$ and decreasing on $(2,4)$ with $\lim_{\alpha\searrow 2}\gamma(\alpha) = 1$.

We will write $k >> n^{\gamma}$, for some $\gamma \in (0,1)$, if $k/n^{\gamma} \to \infty$.

Assumption 2.3: We assume that $k >> n^{\gamma}$ for some $\gamma \in (\gamma(\alpha), 1)$, with $\gamma(\alpha)$ defined in (12).

According to Assumption 2.3, as $\alpha \searrow 2$, the order of k approaches n. One can say that as the value of α approaches the smallest possible value for which the functional principal components exit, almost all observations must be used to ensure the consistency of the Hill estimator. The theory thus begins to break down because this intuitively contradicts the assumption $k/n \to 0$.

Theorem 2.1: *Suppose Assumptions* 2.1, 2.2 *and* 2.3 *hold. Then* $\widehat{H}_{k,n} \xrightarrow{P} \alpha^{-1}$.

While Theorem 2.1 is formulated in the specific setting of projections of functional data onto population and estimated FPCs, it is hoped that the approach we develop will be, in general outlines, applicable to other contexts where the tail indexed must be inferred from approximations to unobserved data. For example, only $Y_i + \varepsilon_i$ with correlated errors ε_i may be observed. It is also hoped that the theory developed for the most commonly used Hill estimator may be used to guide similar developments for other estimators of the tail index.

3. Preliminary results

We collect in this section several results, none of which is particularly profound or difficult to prove, but put together they play an important role in the proof of Theorem 2.1. By placing them in a preparatory section, we will also avoid repeatedly distracting from the main flow of the argument in Section 4.

Following Resnick [12], denote by $M_+ = M_+(0,\infty]$ the space of Radon measures on $(0,\infty]$ with the topology of vague convergence.

Lemma 3.1: *The function h on M_+ defined by $h(\mu) = \mu(z,\infty]$ is continuous at ν_α.*

Proof: Suppose $\mu_n \to \nu_\alpha$. This implies that for any relatively compact B with $\nu_\alpha(\partial B) = 0$, $\mu_n(B) \to \nu_\alpha(B)$. Taking $B = (z,\infty]$, we obtain $h(\mu_n) = \mu_n(B) \to \nu_\alpha(B) = h(\nu_\alpha)$. ■

Lemma 3.2: *The function h on M_+ defined by*

$$h(\mu) = \int_z^M \mu(x, \infty] x^{-1} \, \mathrm{d}x$$

is continuous at ν_α.

Proof: Suppose $\mu_n \to \nu_\alpha$. By Lemma 3.1, for every $x > 0$, $\mu_n(x, \infty]x^{-1} \to \nu_\alpha(x, \infty]x^{-1}$. The convergence

$$\int_z^M \mu_n(x, \infty] x^{-1} \, \mathrm{d}x \to \int_z^M \nu_\alpha(x, \infty] x^{-1} \, \mathrm{d}x$$

follows from the dominated convergence theorem because for $x > z$ and sufficiently large n,

$$\mu_n(x, \infty] \le \mu_n(z, \infty] \le 2\nu_\alpha(z, \infty] = 2z^{-\alpha}. \qquad \blacksquare$$

The measure ν_n is a random element of M_+, ν_α its deterministic (constant) element. The following lemma follows from Theorem 4.1 and relation (4.21) in Resnick [12].

Lemma 3.3: *In the space $M_+(0, \infty]$, $\nu_n \xrightarrow{P} \nu_\alpha$ and $\nu_n^\star \xrightarrow{P} \nu_\alpha$.*

The next lemma follows from relation (4.17) in the proof of Theorem 4.2 in Resnick [12].

Lemma 3.4: *If Y_i are iid and in* $\mathrm{RV}_{-\alpha}$, *then*

$$\frac{Y_{(k)}}{b\left(\frac{n}{k}\right)} \xrightarrow{P} 1.$$

Lemma 3.5: *For any $a, b \ge 0$,*

$$|[a \wedge 1] - [b \wedge 1]| \le |a - b|.$$

Proof: There are four cases:

(1) $a > 1, b > 1, \ |1 - 1| = 0 \le |a - b|$;
(2) $a > 1, b \le 1, |1 - b| = 1 - b < a - b = |a - b|$;
(3) $a \le 1, b > 1, |a - 1| = 1 - a < b - a = |a - b|$;
(4) $a \le 1, b \le 1, |a - b| \le |a - b|$.

$\blacksquare$

The following statements are proven is Section 3.4 of Resnick [15]. The metric ρ which compactifies $(0, \infty]$ at ∞ is

$$\rho(u, v) = \left| \frac{1}{u} - \frac{1}{v} \right|.$$

The distance between measures $\mu_1, \mu_2 \in M_+(0, \infty]$ is defined by

$$d(\mu_1, \mu_2) = \sum_{m=1}^{\infty} 2^{-m} \left\{ |\mu_1(f_m) - \mu_2(f_m)| \wedge 1 \right\}. \tag{13}$$

The functions $f_m \in C_K(0, \infty]$ are of the form

$$f(x) = 1 - [c\rho(x, B) \wedge 1], \tag{14}$$

for some $c > 0$, relatively compact $B \subset (0, \infty]$, and the metric ρ defined above.

Lemma 3.6: *For any metric ρ and any set B,*

$$|\rho(a_1, B) - \rho(a_2, B)| \leq \rho(a_1, a_2).$$

Proof: Recall that $\rho(a, B) = \inf_{b \in B} \rho(a, b)$. For any $b \in B$,

$$\rho(a_1, b) \leq \rho(a_1, a_2) + \rho(a_2, b).$$

Taking the infimum of the left-hand side, we obtain

$$\rho(a_1, B) \leq \rho(a_1, a_2) + \rho(a_2, b).$$

Taking the infimum of the right-hand side, we obtain

$$\rho(a_1, B) \leq \rho(a_1, a_2) + \rho(a_2, B).$$

Consequently,

$$\rho(a_1, B) - \rho(a_2, B) \leq \rho(a_1, a_2).$$

Switching a_1 and a_2, we obtain the claim. ■

Lemma 3.7: *Suppose random variables $H_m(n)$, $m, n \geq 1$, satisfy $0 \leq H_m(n) \leq 1$ and*

$$\forall\, m \geq 1, \quad H_m(n) \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

Then,

$$\sum_{m=1}^{\infty} 2^{-m} H_m(n) \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

Proof: Define

$$\begin{aligned} S(n) &= \sum_{m=1}^{\infty} 2^{-m} H_m(n) \\ &= \sum_{m \le M} 2^{-m} H_m(n) + \sum_{m > M} 2^{-m} H_m(n) \\ &=: S_M(n) + S_M^{\star}(n). \end{aligned}$$

Fix $\varepsilon > 0$ and observe that

$$P(S(n) > \varepsilon) \le P\left(S_M(n) > \frac{\varepsilon}{2}\right) + P\left(S_M^{\star}(n) > \frac{\varepsilon}{2}\right).$$

Since $S_M^{\star}(n) \le 2^{-M}$, we can choose M so large that $P\left(S_M^{\star}(n) > \varepsilon/2\right) = 0$. For such a (fixed) M,

$$P(S(n) > \varepsilon) \le P\left(S_M(n) > \frac{\varepsilon}{2}\right) \to 0. \qquad \blacksquare$$

4. Proof of Theorem 2.1

The proof of Theorem 2.1 is constructed from a series of results, of which Proposition 4.1 is the most prominent. To facilitate the understanding of the proofs of Proposition 4.1 and Theorem 2.1, we note that

If $\alpha \in (3, 4)$, then $1/(\alpha - 1) > 2 - \alpha/2$,
If $\alpha \in (2, 3)$, then $1/(\alpha - 1) < 2 - \alpha/2$.

We may thus write

$$\gamma(\alpha) = \max\left\{\frac{1}{\alpha - 1}, 2 - \frac{\alpha}{2}\right\}, \quad \alpha \in (2, 4]. \tag{15}$$

Proposition 4.1: *Under the assumptions of Theorem* 2.1, $d(\nu_n^{\dagger}, \nu_n) \xrightarrow{P} 0$.

Proof: Since each function f_m in (13) has compact support in $(0, \infty]$,

$$s_m := \inf\left\{\operatorname{supp}(f_m)\right\} > 0.$$

Therefore

$$\begin{aligned} |\nu_n^{\dagger}(f_m) - \nu_n(f_m)| &= \left|\int f_m \, \mathrm{d}\nu_m^{\dagger} - \int f_m \, \mathrm{d}\nu_n\right| \\ &\le \frac{1}{k} \sum_{i=1}^{n} \left|f_m\left(\frac{\widehat{Y}_i}{b(n/k)}\right) - f_m\left(\frac{Y_i}{b(n/k)}\right)\right| \\ &= \frac{1}{k} \sum_{i \in \mathcal{I}_m} \left|f_m\left(\frac{\widehat{Y}_i}{b(n/k)}\right) - f_m\left(\frac{Y_i}{b(n/k)}\right)\right|, \end{aligned}$$

where

$$\mathcal{I}_m = \left\{ i \geq 1 : \widehat{Y}_i > s_m b(n/k) \text{ or } Y_i > s_m b(n/k) \right\}.$$

Since each f_m is of the form (14), by Lemmas 3.5 and 3.6,

$$\begin{aligned} |\nu_n^\dagger(f_m) - \nu_n(f_m)| &\leq \frac{c_m}{k} \sum_{i \in \mathcal{I}_m} \left| \rho\left(\frac{\widehat{Y}_i}{b(n/k)}, B_m \right) - \rho\left(\frac{Y_i}{b(n/k)}, B_m \right) \right| \\ &\leq \frac{c_m}{k} \sum_{i \in \mathcal{I}_m} \left| \rho\left(\frac{\widehat{Y}_i}{b(n/k)}, \frac{Y_i}{b(n/k)} \right) \right|. \end{aligned}$$

The claim will thus follow from the convergence

$$\sum_{m=1}^{\infty} 2^{-m} \left\{ \left[\frac{c_m}{k} \sum_{i \in \mathcal{I}_m} \left| \frac{b(n/k)}{\widehat{Y}_i} - \frac{b(n/k)}{Y_i} \right| \right] \wedge 1 \right\} \xrightarrow{P} 0,$$

which, in turn, by Lemma 3.7, will follow from

$$\frac{1}{k} \sum_{i \in \mathcal{I}(n)} \left| \frac{b(n/k)}{\widehat{Y}_i} - \frac{b(n/k)}{Y_i} \right| \xrightarrow{P} 0, \tag{16}$$

where, for some $s^\star > 0$,

$$\mathcal{I}(n) = \left\{ i \geq 1 : \widehat{Y}_i > s^\star b(n/k) \text{ or } Y_i > s^\star b(n/k) \right\}.$$

Set

$$\mathcal{I}^{(1)}(n) = \left\{ i \geq 1 : Y_i > s^\star b(n/k) \right\}, \quad \mathcal{I}^{(2)}(n) = \left\{ i \geq 1 : \widehat{Y}_i > s^\star b(n/k) \right\}.$$

Relation (16) will follow once we have shown that for $g = 1$ and $g = 2$,

$$\frac{b(n/k)}{k} \sum_{i \in \mathcal{I}^{(g)}(n)} \frac{|\widehat{Y}_i - Y_i|}{\widehat{Y}_i Y_i} \xrightarrow{P} 0. \tag{17}$$

We verify (17) for $g = 1$. The argument for $g = 2$ is basically the same; the roles of $\widehat{Y}_i$ and Y_i must be interchanged.

Fix $\varepsilon > 0$. First observe that

$$P\left(\frac{b(n/k)}{k} \sum_{i \in \mathcal{I}^{(1)}(n)} \frac{|\widehat{Y}_i - Y_i|}{\widehat{Y}_i Y_i} > \varepsilon \right) \leq P\left(G(n) > \varepsilon \right),$$

where

$$G(n) = \frac{1}{s^\star k} \sum_{i \in \mathcal{I}^{(1)}(n)} \frac{|\widehat{Y}_i - Y_i|}{\widehat{Y}_i}.$$

Next, use the decomposition

$$P\left(G(n) > \varepsilon \right) = P_1(n) + P_2(n),$$

with

$$P_1(n) = P\left(G(n) > \varepsilon,\ \exists\, i \in \mathcal{I}^{(1)}(n) :\ \widehat{Y}_i \le \tfrac{1}{2}s^\star b(n/k)\right);$$
$$P_2(n) = P\left(G(n) > \varepsilon,\ \forall\, i \in \mathcal{I}^{(1)}(n) : \widehat{Y}_i > \tfrac{1}{2}s^\star b(n/k)\right).$$

Observe that

$$\begin{aligned} P_1(n) &\le P\left(\exists\, i \in \mathcal{I}^{(1)}(n) : \widehat{Y}_i \le \frac{1}{2}s^\star b(n/k)\right) \\ &\le P\left(\exists\, i \le n : Y_i > s^\star b(n/k) \text{ and } \widehat{Y}_i \le \frac{1}{2}s^\star b(n/k)\right) \\ &\le P\left(\exists\, i \le n : |\widehat{Y}_i - Y_i| > \frac{1}{2}s^\star b(n/k)\right) \\ &= P\left(\max_{1\le i\le n} |\widehat{Y}_i - Y_i| > \frac{1}{2}s^\star b(n/k)\right). \end{aligned} \tag{18}$$

By (8),

$$\begin{aligned} P_1(n) &\le P\left(A_v \|\widehat{C} - C\|_{\mathcal{L}} \max_{1\le i\le n} \|X_i\| > \frac{1}{2}s^\star b(n/k)\right) \\ &\le \frac{2A_v}{s^\star b(n/k)} E\left[\|\widehat{C} - C\|_{\mathcal{L}} \max_{1\le i\le n} \|X_i\|\right]. \end{aligned} \tag{19}$$

We first consider the case of $\alpha > 4$. By (19) and (9),

$$\begin{aligned} P_1(n) &\le \frac{2A_v}{s^\star b(n/k)} \left\{E\|\widehat{C} - C\|_{\mathcal{L}}^2\right\}^{1/2} \left\{E \max_{1\le i\le n} \|X_i\|^2\right\}^{1/2} \\ &= O\left(\frac{1}{b(n/k)} n^{-1/2} n^{1/2}\right) = O\left(\frac{1}{b(n/k)}\right) = o(1), \end{aligned}$$

where we used

$$E \max_{1\le i\le n} \|X_i\|^2 \le E \sum_{1\le i\le n} \|X_i\|^2 = O(n).$$

By Markov's inequality,

$$\begin{aligned} P_2(n) &\le P\left(\frac{2}{s^{\star 2} k b(n/k)} \sum_{i=1}^n |\widehat{Y}_i - Y_i| > \varepsilon\right) \\ &\le \frac{2}{\varepsilon s^{\star 2} k b(n/k)} \sum_{i=1}^n E|\widehat{Y}_i - Y_i|. \end{aligned} \tag{20}$$

By (8) and (9),

$$E|\widehat{Y}_i - Y_i| \le A_v \left\{E\|X_i\|^2\right\}^{1/2} \left\{E\|\widehat{C} - C\|_{\mathcal{L}}^2\right\}^{1/2} = O(n^{-1/2}).$$

Therefore,

$$P_2(n) = O\left(\frac{n^{1/2}}{kb(n/k)}\right) = o(1).$$

The last equality follows from the assumption $k >> n^{\gamma(\alpha)}$ and (6).

Now consider the case of $\alpha \in (2,4)$. We first show that $P_1(n) = o(1)$. By (18), (8) and Markov's inequality

$$P_1(n) = O\left(\frac{1}{\sqrt{b(n/k)}}\right) E\left[\|\widehat{C} - C\|_{\mathcal{L}}^{1/2} \max_{1\le i\le n} \|X_i\|^{1/2}\right].$$

We apply Hölder's inequality with $p = 2\beta$ and $q = 2\beta/(2\beta - 1)$ to get

$$P_1(n) = O\left(\frac{1}{\sqrt{b(n/k)}}\right) \left\{E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta}\right\}^{1/2\beta} \left\{E \max_{1\le i\le n} \|X_i\|^{\beta/(2\beta-1)}\right\}^{(2\beta-1)/2\beta}.$$

For the above bound to be effective, we need $E\|X_i\|^{\beta/(2\beta-1)} < \infty$, which is implied by $\beta/(2\beta - 1) < \alpha$. Since $2\beta - 1 > 1$ and $\beta < \alpha/2$, this condition always holds. It therefore follows from (11) that

$$P_1(n) = O\left(\frac{1}{\sqrt{b(n/k)}} n^{-\kappa/\beta} n^{(2\beta-1)/2\beta}\right).$$

We can thus conclude that $P_1(n) = o(1)$, if there are β and κ such that $-\kappa + 2\beta - 1 < 0$. This is possible if

$$2\beta - 1 < \beta\left(1 - \frac{2}{\alpha}\right).$$

The above condition can be equivalently stated as $\beta(1 - 2/\alpha) < 1$. Since $\beta < \alpha/2$, $\beta(1 - 2/\alpha) < \alpha/2 - 1 < 1$ because $\alpha < 4$. This completes the verification of $P_1(n) = o(1)$ for $\alpha \in (2,4)$.

To show that $P_2(n) = o(1)$, observe that by (20), Markov's inequality with $0 < r \le 1$, and (8),

$$\begin{aligned} P_2(n) &\le P\left(\frac{2}{s^{\star 2} kb(n/k)} \sum_{i=1}^{n} |\widehat{Y}_i - Y_i| > \varepsilon\right) \\ &= O\left(\frac{1}{k^r b^r(n/k)}\right) E\left(\sum_{i=1}^{n} |\widehat{Y}_i - Y_i|\right)^r \\ &= O\left(\frac{n^r}{k^r b^r(n/k)}\right) E\left[\|\widehat{C} - C\|_{\mathcal{L}} \frac{1}{n} \sum_{i=1}^{n} \|X_i\|\right]^r. \end{aligned}$$

Applying Hölder's inequality with $p = \beta/r$ and $q = \beta/(\beta - r)$, we obtain

$$E\left[\|\widehat{C} - C\|_{\mathcal{L}}^{r} \left(\frac{1}{n}\sum_{i=1}^{n} \|X_i\|\right)^r\right] \le \left\{E\|\widehat{C} - C\|_{\mathcal{L}}^{\beta}\right\}^{r/\beta} \left\{E\left(\frac{1}{n}\sum_{i=1}^{n} \|X_i\|\right)^{rq}\right\}^{1/q}.$$

For $E\|X_i\|^{rq}$ to be finite, we need

$$rq = \frac{r\beta}{\beta - r} < \alpha. \tag{21}$$

Choosing

$$r = \frac{\beta}{\beta + 1} \tag{22}$$

implies $rq = 1$. We thus obtain, with r specified in (22),

$$\begin{aligned} P_2(n) &= O\left(\frac{n^r}{k^r b^r(n/k)}\right) \{E\|X\|\}^{1/q}\, n^{-\kappa r/\beta} \\ &= O\left(\frac{n^{r-r\kappa/\beta}}{k^r b^r(n/k)}\right). \end{aligned}$$

By (6), the claim $P_2(n) = o(1)$ will thus follow if $k >> n^{\gamma}$, where

$$\gamma = \frac{r - \frac{\kappa r}{\beta} - \frac{r}{\alpha}}{r - \frac{r}{\alpha}} = \frac{1 - \frac{1}{\alpha} - \frac{\kappa}{\beta}}{1 - \frac{1}{\alpha}}.$$

The exponent is smaller than 1 and attains its smallest value as κ/β approaches its largest possible value, i.e. $1 - 2/\alpha$. It remains to observe that

$$\frac{1 - \frac{1}{\alpha} - \frac{\kappa}{\beta}}{1 - \frac{1}{\alpha}} = \frac{1}{\alpha - 1}, \quad \text{if } \frac{\kappa}{\beta} = 1 - \frac{2}{\alpha}.$$

■

Remark 4.1: The proof of Proposition 4.1, in the case $\alpha \in (2, 4)$, is valid in (15) is replaced by $\gamma(\alpha) = (\alpha - 1)^{-1}$. Only the latter bound was used. The bound $2 - \alpha/2$ is needed in the proof of Theorem 2.1.

Using Lemma 3.3, we obtain the following corollary.

Corollary 4.1: *Under the assumptions of Theorem* 2.1, $\nu_n^{\dagger} \xrightarrow{P} \nu_\alpha$.

The arguments used in the proofs of Propositions 4.2 and 4.3 are similar to those developed in Sections 4.3. and 4.4 of Resnick [12].

Proposition 4.2: *Under the assumptions of Theorem* 2.1,

$$\frac{\widehat{Y}_{(k)}}{b(n/k)} \xrightarrow{P} 1.$$

Proof: Fix $\varepsilon > 0$ and set

$$P_+(n) = P\left(\frac{\widehat{Y}_{(k)}}{b(n/k)} > 1+\varepsilon\right), \quad P_-(n) = P\left(\frac{\widehat{Y}_{(k)}}{b(n/k)} < 1-\varepsilon\right).$$

Observe that

$$\begin{aligned} P_+(n) &= P\left(I_{\widehat{Y}_{(k)}/b(n/k)}(1+\varepsilon,\infty] = 1\right) \\ &\le P\left(\sum_{i=1}^{n} I_{\widehat{Y}_{(i)}/b(n/k)}(1+\varepsilon,\infty] \ge k\right) \\ &= P\left(\frac{1}{k}\sum_{i=1}^{n} I_{\widehat{Y}_{(i)}/b(n/k)}(1+\varepsilon,\infty] \ge 1\right) \\ &= P\left(\nu_n^{\dagger}(1+\varepsilon,\infty] \ge 1\right). \end{aligned}$$

A similar argument shows that

$$P_-(n) \le P\left(\nu_n^{\dagger}(1-\varepsilon,\infty] < 1\right).$$

The claim follows because by Corollary 4.1 and Lemma 3.1,

$$\begin{aligned} \nu_n^{\dagger}(1+\varepsilon,\infty] &\xrightarrow{P} \nu_\alpha(1+\varepsilon,\infty] = (1+\varepsilon)^{-\alpha} < 1; \\ \nu_n^{\dagger}(1-\varepsilon,\infty] &\xrightarrow{P} \nu_\alpha(1-\varepsilon,\infty] = (1-\varepsilon)^{-\alpha} > 1. \end{aligned}$$

■

Proposition 4.3: *Under the assumptions of Theorem* 2.1, $\hat{\nu}_n \xrightarrow{P} \nu_\alpha$.

Proof: Consider the map $T : M_+ \times (0,\infty) \to M_+$ defined by

$$T(\mu, x)(A) = \mu(xA), \quad \text{for Borel } A \subset (0,\infty].$$

Resnick [12, pp. 83–84] shows that T is continuous. Observe that

$$T\left(\nu_n^{\dagger}, \frac{\widehat{Y}_{(k)}}{b(n/k)}\right) = \hat{\nu}_n, \quad T(\nu_\alpha, 1) = \nu_\alpha.$$

The claim thus follows because by Corollary 4.1 and Proposition 4.2,

$$\left(\nu_n^{\dagger}, \frac{\widehat{Y}_{(k)}}{b(n/k)}\right) \xrightarrow{P} (\nu_\alpha, 1) \quad \text{in } M_+ \times (0,\infty).$$

■

The following lemma may be of independent interest and more general utility.

Lemma 4.1: *Suppose* $y \mapsto P(Y > y) \in \mathrm{RV}_{-\alpha}$ *for some* $\alpha > 0$. *Then,*

$$\lim_{z\to\infty} \limsup_{t\to\infty} \int_z^\infty tP(Y > xb(t))x^{-1}\,\mathrm{d}x = 0.$$

Proof: The function $b(\cdot)$ is defined by

$$P(Y > b(t)) = t^{-1}.$$

We know that $b(\cdot) \in \mathrm{RV}_{1/\alpha}$ and

$$\lim_{t\to\infty} tP(Y > xb(t)) = x^{-\alpha}, \quad x > 0. \tag{23}$$

Set $f_t(x) = tP(Y > xb(t))x^{-1}$. We want to show

$$\lim_{z\to\infty} \limsup_{t\to\infty} \int_z^\infty f_t(x)\,\mathrm{d}x = 0.$$

By (23),

$$\forall\, x > 0 \quad f_t(x) \to x^{-\alpha-1}, \quad \text{as } t \to \infty.$$

To conclude that

$$\int_z^\infty f_t(x)\,\mathrm{d}x \to \int_z^\infty x^{-\alpha-1}\,\mathrm{d}x, \quad \text{as } t \to \infty,$$

we must find a function g such that for $t > t_0$,

$$f_t(x) \le g(x) \quad \text{and} \quad \int_z^\infty g(x)\,\mathrm{d}x < \infty.$$

Set $U(y) = P(Y > y)$. Potter bounds state that $\forall\, \delta > 0, \exists\, u_0, \forall\, u \ge u_0, \forall\, y \ge 1$,

$$(1-\delta)y^{-\alpha-\delta} \le \frac{U(yu)}{U(u)} \le (1+\delta)y^{-\alpha+\delta}.$$

Since $b(t) \to \infty$ as $t \to \infty$, $\exists\, t_0, \forall\, t > t_0$,

$$U(xb(t)) \le (1+\delta)x^{-\alpha+\delta}U(b(t)).$$

Since $U(b(t)) = 1/t$, we obtain, for $t \ge t_0$,

$$f_t(x) = tU(xb(t))x^{-1} \le (1+\delta)x^{-\alpha+\delta-1} =: g(x).$$

The function g is integrable if $\delta < \alpha$. ■

Proof of Theorem 2.1: Since

$$\widehat{H}_{k,n} = \int_1^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x,$$

we must show that

$$\int_1^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x \xrightarrow{P} \int_1^\infty \nu_\alpha(x,\infty]x^{-1}\,\mathrm{d}x = \alpha^{-1}.$$

The verification is based on the commonly used truncation argument, Theorem 3.2 in Billingsley [22], also stated as Theorem 3.5 in Resnick [12]. Set

$$V_n = \int_1^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x, \quad V = \int_1^\infty \nu_\alpha(x,\infty]x^{-1}\,\mathrm{d}x;$$
$$V_n^{(M)} = \int_1^M \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x, \quad V^{(M)} = \int_1^M \nu_\alpha(x,\infty]x^{-1}\,\mathrm{d}x.$$

To establish the desired convergence $V_n \xrightarrow{P} V$, equivalently $V_n \xrightarrow{d} V$, we must verify that

$$\forall\, M > 1, \quad V_n^{(M)} \xrightarrow{d} V_n^{(M)}, \quad \text{as } n \to \infty; \tag{24}$$

$$V^{(M)} \xrightarrow{d} V, \quad \text{as } M \to \infty; \tag{25}$$

$$\forall\, \varepsilon > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} P\left(|V_n^{(M)} - V_n| > \varepsilon\right) = 0. \tag{26}$$

Convergence (24) follows from Proposition 4.3 and Lemma 3.2. Convergence (25) is trivial because $\int_M^\infty \nu_\alpha(x,\infty]x^{-1}\,\mathrm{d}x = \alpha^{-1}M^{-\alpha}$. Since $|V_n^{(M)} - V_n| = \int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x$, (26) is equivalent to

$$\forall\, \varepsilon > 0, \quad \lim_{M\to\infty} \limsup_{n\to\infty} P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x > \varepsilon\right) = 0.$$

The steps of the verification of the above relation, up to (27), are pretty much the same as those developed by Resnick [12, pp. 84–85]. We provide the details because we work with the measure $\nu_n^\dagger$ rather than with the measure ν_n, and the context for the remainder of the proof is helpful. Following (27), we use a different argument.

Fix $\varepsilon > 0$ and $\eta > 0$. Observe that

$$P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x > \varepsilon\right) \le Q_1(n) + Q_2(n),$$

where

$$Q_1(n) = P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x > \varepsilon,\ \left|\frac{\widehat{Y}_{(k)}}{b(n/k)} - 1\right| < \eta\right),$$
$$Q_2(n) = P\left(\left|\frac{\widehat{Y}_{(k)}}{b(n/k)} - 1\right| \ge \eta\right).$$

By Proposition 4.2, $\limsup_{n\to\infty} Q_2(n) = 0$, so we focus on $Q_1(n)$. We start with the bound

$$\begin{aligned} Q_1(n) &\le P\left(\int_M^\infty \hat{\nu}_n(x,\infty]x^{-1}\,\mathrm{d}x > \varepsilon,\ \frac{\widehat{Y}_{(k)}}{b(n/k)} > 1-\eta\right) \\ &= P\left(\int_M^\infty \frac{1}{k}\sum_{i=1}^n I_{\widehat{Y}_i/\widehat{Y}_{(k)}}(x,\infty]x^{-1}\,\mathrm{d}x > \varepsilon,\ \frac{\widehat{Y}_{(k)}}{b(n/k)} > 1-\eta\right). \end{aligned}$$

Conditions $\widehat{Y}_i/\widehat{Y}_{(k)} > x$ and $\widehat{Y}_{(k)}/b(n/k) > 1-\eta$ imply $\widehat{Y}_i/b(n/k) > x(1-\eta)$, so

$$\begin{aligned} Q_1(n) &\le P\left(\int_M^\infty \frac{1}{k}\sum_{i=1}^n I_{\widehat{Y}_i/b(n/k)}(x(1-\eta),\infty]x^{-1}\,\mathrm{d}x > \varepsilon\right) \\ &= P\left(\int_M^\infty \nu_n^\dagger(x(1-\eta),\infty]x^{-1}\,\mathrm{d}x > \varepsilon\right) \\ &= P\left(\int_{M(1-\eta)}^\infty \nu_n^\dagger(x,\infty]x^{-1}\,\mathrm{d}x > \varepsilon\right). \end{aligned}$$

Consequently, because $\widehat{Y}_i$, $1 \le i \le n$, have the same distribution,

$$\begin{aligned} Q_1(n) &\le \frac{1}{\varepsilon}\int_{M(1-\eta)}^\infty E\left[\nu_n^\dagger(x,\infty]\right]x^{-1}\,\mathrm{d}x \\ &= \frac{1}{\varepsilon}\int_{M(1-\eta)}^\infty \frac{n}{k}P\left(\frac{\widehat{Y}}{b(n/k)} > x\right)x^{-1}\,\mathrm{d}x. \end{aligned}$$

It thus remains to show that

$$\lim_{z\to\infty}\limsup_{n\to\infty}\int_z^\infty \frac{n}{k}P\left(\widehat{Y} > xb(n/k)\right)x^{-1}\,\mathrm{d}x = 0. \tag{27}$$

We use the decomposition

$$\begin{aligned} P\left(\widehat{Y} > xb(n/k)\right) &= P\left(\widehat{Y} > xb(n/k),\ Y > \tfrac{1}{2}xb(n/k)\right) \\ &\quad + P\left(\widehat{Y} > xb(n/k),\ Y \le \tfrac{1}{2}xb(n/k)\right) \\ &\le P\left(Y > \tfrac{1}{2}xb(n/k)\right) + P\left(|\widehat{Y}-Y| > \tfrac{1}{2}xb(n/k)\right). \end{aligned}$$

By Lemma 4.1,

$$\lim_{z\to\infty}\limsup_{n\to\infty}\int_z^\infty \frac{n}{k}P\left(Y > \frac{1}{2}xb(n/k)\right)x^{-1}\,\mathrm{d}x = 0.$$

If $\alpha > 4$, by (8) and (9)

$$\begin{aligned} \int_z^\infty \frac{n}{k}P\left(|\widehat{Y}-Y| > \frac{1}{2}xb(n/k)\right)x^{-1}\,\mathrm{d}x &\le \frac{n}{k}\int_z^\infty \frac{2}{xb(n/k)}E|\widehat{Y}-Y|x^{-1}\,\mathrm{d}x \\ &= O\left(\frac{n^{1/2}}{kb(n/k)}\right)\frac{1}{z}. \end{aligned}$$

By Assumption 2.3, for a slowly varying function L and $\gamma \in (\gamma(\alpha), 1)$,

$$\frac{n^{1/2}}{kb(n/k)} = \left\{\frac{n^{\gamma}}{k}\right\}^{1-1/\alpha} \left\{n^{\gamma(\alpha)-\gamma} L\left(\frac{n}{k}\right)\right\}^{1-1/\alpha} \to 0, \quad \text{as } n \to \infty.$$

If $\alpha \in (2, 4)$, we use the bound ($r \in (0, 1]$):

$$\begin{aligned}\int_z^{\infty} \frac{n}{k} P\left(|\widehat{Y} - Y| > \frac{1}{2} x b(n/k)\right) x^{-1}\, \mathrm{d}x &\le \frac{n}{k} \int_z^{\infty} \left(\frac{2}{xb(n/k)}\right)^r E|\widehat{Y} - Y|^r x^{-1}\, \mathrm{d}x \\ &= O\left(\frac{n}{k} \frac{E|\widehat{Y} - Y|^r}{b^r(n/k)}\right) \frac{1}{z^r}.\end{aligned}$$

The value of r will depend on α. Choosing it, and checking that it is available, requires some work. As in the proof of Proposition 4.1,

$$E|\widehat{Y} - Y|^r = O\left(n^{-\kappa r/\beta}\right),$$

provided (21) holds. Set

$$\gamma^* = \frac{1 - \frac{r}{\alpha} - \frac{\kappa r}{\beta}}{1 - \frac{r}{\alpha}}.$$

Then, for some slowly varying L,

$$\frac{n}{k} \frac{E|\widehat{Y} - Y|^r}{b^r(n/k)} = O\left(\frac{n^{\gamma^*}}{k} L\left(\frac{n}{k}\right)\right)^{1-r/\alpha}.$$

Clearly $\gamma^* < 1$. We must verify that there are β, κ and r, in permitted ranges, such that γ^* can be arbitrarily close to $\gamma(\alpha)$ given by (15). With α and r fixed, γ^* will approach its smallest possible value as κ/β approaches its largest possible value, i.e. $1 - 2/\alpha$. In this case, γ^* is greater than and approaches

$$\gamma_L(\alpha, r) := \frac{1 - \frac{r}{\alpha} - \left(1 - \frac{2}{\alpha}\right) r}{1 - \frac{r}{\alpha}} = \frac{\alpha - \alpha r + r}{\alpha - r}.$$

Condition (21) restricts the available values of r. A direct calculation shows that it is equivalent to

$$r < \frac{\beta\alpha}{\beta + \alpha}.$$

For a fixed α, the right-hand side is an increasing function of β and attains its upper limit if $\beta = \alpha/2$. This means that r must be less than, but can be arbitrarily close to, $\alpha/3$. Thus γ^* can be arbitrarily close to

$$\gamma_L\left(\alpha, \frac{\alpha}{3}\right) = 2 - \frac{\alpha}{2}.$$

Combining it with Remark 4.1 concludes the proof. ■

Acknowledgments

We thank the Associate Editor and two referees for reading the paper carefully and pointing out ways of improving the exposition.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research has been partially supported by National Science Foundation (NSF) grants DMS-1462067 and DMS-1737795.

References

[1] Bosq D. Linear processes in function spaces. New York: Springer; 2000.
[2] Ramsay JO, Silverman BW. Functional data analysis. New York: Springer; 2005.
[3] Horváth L, Kokoszka P. Inference for functional data with applications. New York: Springer; 2012.
[4] Hsing T, Eubank R. Theoretical foundations of functional data analysis, with an introduction to linear operators. New York: Wiley; 2015.
[5] Yao F, Müller H-G, Wang J-L. Functional linear regression analysis for longitudinal data. Ann Stat. 2005;33:2873–2903.
[6] Gabrys R, Kokoszka P. Portmanteau test of independence for functional observations. J Amer Stat Assoc. 2007;102:1338–1348.
[7] Berkes I, Gabrys R, Horváth L, et al. Detecting changes in the mean of functional observations. J Roy Stat Soc B. 2009;71:927–946.
[8] Horváth L, Kokoszka P, Reeder R. Estimation of the mean of functional time series and a two sample problem. J Roy Stat Soc B. 2013;75:103–122.
[9] Horváth L, Kokoszka P, Rice G. Testing stationarity of functional time series. J Econom. 2014;179:66–82.
[10] Embrechts P, Klüppelberg C, Mikosch T. Modelling extremal events for insurance and finance. Berlin: Springer; 1997.
[11] Beirlant J, Goegebeur Y, Segers J, et al. Statistics of extremes: theory and applications. John Wiley & Sons; 2004.
[12] Resnick SI. Heavy-tail phenomena. New York: Springer; 2007.
[13] Mason DM. Laws of large numbers for sums of extreme values. Ann Probab. 1982;10:754–764.
[14] Hsing T. On tail index estimation using dependent data. Ann Stat. 1991;19:1547–1569.
[15] Resnick SI. Extreme values, regular variation, and point processes. New York: Springer; 1987.
[16] Resnick SI, Stărică C. Consistency of Hill's estimator for dependent data. J Appl Probab. 1995;32:139–167.
[17] Resnick SI, Stărică C. Tail index estimation for dependent data. Ann Appl Probab. 1998;8:1156–1183.
[18] Wang T, Resnick SI, Consistency of hill estimators in a linear preferential attachment model. Extremes. 2018. doi:10.1007/s10687-018-0335-7
[19] Hult H, Lindskog F. Regular variation for measures on metric spaces. Publ Inst Math (NS). 2006;80(94):121–140.
[20] Meiguet T. Heavy tailed functional time series. Universite catholigue de Louvain; 2010.
[21] Kokoszka P, Stoev S, Xiong Q, Principal components analysis of regularly varying functions. Bernoulli. 2019. Revision invited.
[22] Billingsley P. Convergence of probability measures. 2nd ed. New York: Wiley; 1999.
[23] Lucca DO, Moench E. The pre-FOMC announcement drift. J Finance. 2015;70:329–371.
[24] Hall P. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. J Multivar Anal. 1990;32:177–203.

Appendices

Appendix 1. Background in regular variation

We utilize the concept of regular variation of measures in L^2. We start by recalling some terminology and fundamental facts about regularly varying functions.

A measurable function $L : (0, \infty) \to \mathbb{R}$ is said to be slowly varying (at infinity) if, for all $\lambda > 0$,

$$\frac{L(\lambda u)}{L(u)} \to 1, \quad \text{as } u \to \infty.$$

Functions of the form $R(u) = u^{\rho} L(u)$ are said to be regularly varying with exponent $\rho \in \mathbb{R}$.

The notion of regular variation has been extended to Banach and even metric spaces using the notion of M_0 convergence [19]. Even though we will work only with Hilbert spaces, we review the theory in a more general context.

Consider a separable Banach space $\mathbb{B}$ and let $B_\epsilon := \{z \in \mathbb{B} : \|z\| < \epsilon\}$ be the open ball of radius $\epsilon > 0$, centred at the origin. A Borel measure μ defined on $\mathbb{B}_0 := \mathbb{B} \backslash \{\mathbf{0}\}$ is said to be *boundedly finite* if $\mu(A) < \infty$, for all Borel sets that are bounded away from $\mathbf{0}$, that is, such that $A \cap B_\epsilon = \emptyset$, for some $\epsilon > 0$. Let $\mathbb{M}_0$ be the collection of all such measures. For $\mu_n, \mu \in \mathbb{M}_0$, we say that μ_n converge to μ in the M_0 topology, if $\mu_n(A) \to \mu(A)$, for all bounded away from $\mathbf{0}$, μ-continuity Borel sets A, i.e., such that $\mu(\partial A) = 0$, where $\partial A := \overline{A} \setminus A^{\circ}$ denotes the boundary of A. M_0 convergence can be metrized such that $\mathbb{M}_0$ becomes a complete separable metric space (Theorem 2.3 in [19] and also Section 2.2. of [20]). The following result is known, see e.g. Chapter 2 of Meiguet [20] and references therein.

Proposition A.1: *Let X be a random element in a separable Banach space $\mathbb{B}$ and $\alpha > 0$. The following three statements are equivalent*:

(i) *For some slowly varying function L,*

$$P(\|X\| > u) = u^{-\alpha} L(u)$$

and

$$\frac{P(u^{-1}X \in \cdot)}{P(\|X\| > u)} \xrightarrow{M_0} \mu(\cdot), \quad u \to \infty,$$

where μ is a non-null measure on the Borel σ-field $\mathcal{B}(\mathbb{B}_0)$ of $\mathbb{B}_0 = \mathbb{B} \backslash \{\mathbf{0}\}$.

(ii) *There exists a probability measure Γ on the unit sphere $\mathbb{S}$ in $\mathbb{B}$ such that, for every $t > 0$,*

$$\frac{P(\|X\| > tu, X/\|X\| \in \cdot)}{P(\|X\| > u)} \xrightarrow{w} t^{-\alpha} \Gamma(\cdot), \quad u \to \infty.$$

(iii) *Relation (4) holds, and for the same spectral measure Γ in* (ii),

$$P\left(X/\|X\| \in \cdot \,|\, \|X\| > u\right) \xrightarrow{w} \Gamma(\cdot), \quad u \to \infty.$$

Definition A.1: If any one of the equivalent conditions in Proposition A.1 hold, we shall say that X is regularly varying with index α. The measures μ and Γ will be referred to as exponent and angular measures of X, respectively.

Appendix 2. A motivating data example

We present a motivating example based on financial data. Similar questions arise in the analysis of annual precipitation or other climate related curves.

Denote by $P_i(t)$ the price of an asset at time t of trading day i. For the assets we consider in our example, t is time in minutes between 9:30 and 16:00 EST (NYSE opening times) rescaled to the unit interval $(0, 1)$. The intraday return curve on day i is defined by $X_i(t) = \log P_i(t) - \log P_i(0)$. In practice, $P_i(0)$ is the price after the first minute of trading. The curves X_i show how the return

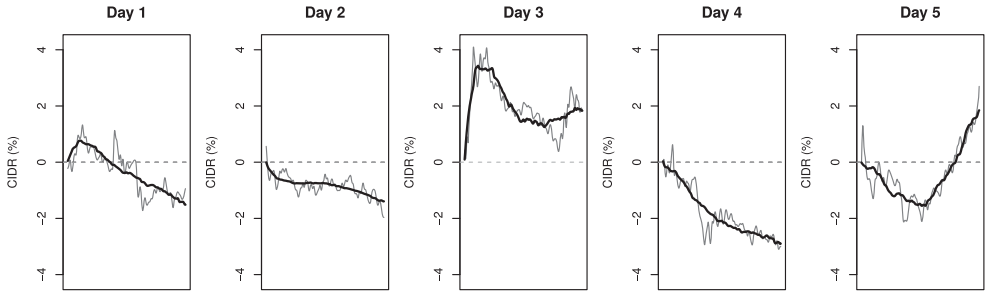

Figure A1. Five consecutive intraday return curves, Walmart stock. The raw returns are noisy grey lines. The smoother black lines are approximations $\widehat{X}_i(t) = \sum_{j=1}^{3} \hat{\xi}_{ij}\hat{v}_j$.

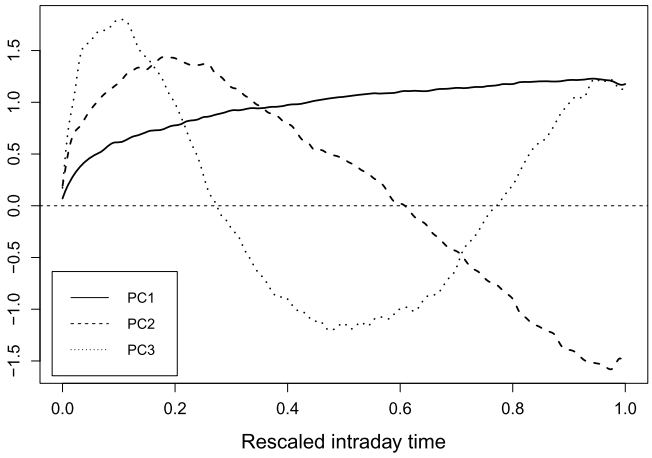

Figure A2. The first three sample FPCs of intraday returns on Walmart stock.

accumulates over the trading day, see e.g. Figure 1 in Lucca and Moench [23]; examples of such curves are shown in Figure A1.

The first three sample FPCs, $\hat{v}_1, \hat{v}_2, \hat{v}_3$, are shown in Figure A2. They are computed, using (3), from minute-by-minute Walmart returns from 5 July 2006 to 30 December 2011, $n = 1378$ trading days. (This period is used for the other assets we consider.) The curves $\widehat{X}_i(t) = \sum_{j=1}^{3} \hat{\xi}_{ij}\hat{v}_j$, with the scores $\hat{\xi}_{ij} = \int X_i(t)\hat{v}_j(t)\,\mathrm{d}t$, approximate the curves X_i well, as shown in Figure A1. Figure A3 shows the Hill plots of the sample scores $\hat{\xi}_{ij}$ for two stocks and for $j = 1,2,3$. These plots show estimates, $\hat{\alpha} = \hat{\alpha}(k)$, of the tail index α as a function of the minimal order statistic k used to compute $\hat{\alpha}$. The formula for these estimates is given in Section 2. In an asymptotic setting, $\hat{\alpha}$ is obtained as the number of upper order statistics, k, tends to infinity with the sample size n, in such a way that $k/n \to 0$. A general practical approach to choosing k is to examine these plots and use the values of k which lead to relatively stable estimates, keeping in mind that k cannot be too large to avoid bias, nor too small to avoid variability. There are also several ways of selecting 'optimal k'. Figure A3 shows the Hill plots centred around the value of k selected by the method of Hall [24], implemented by the function `hall` of the `R` package `tea`. These plots show that it is reasonable to assume that the scores have Pareto tails, with the tail index between 2 and 4.

We emphasize that the Hill plots Figure A3 are computed using the samples score $\hat{\xi}_{ij} = \langle X_i, \hat{v}_j \rangle$, whereas the population parameter is the tail index α of the unobservable scores $\xi_{ij} = \langle X_i, v_j \rangle$. We also note that the tails studied in this example are those of the scores of cumulative return functions, not

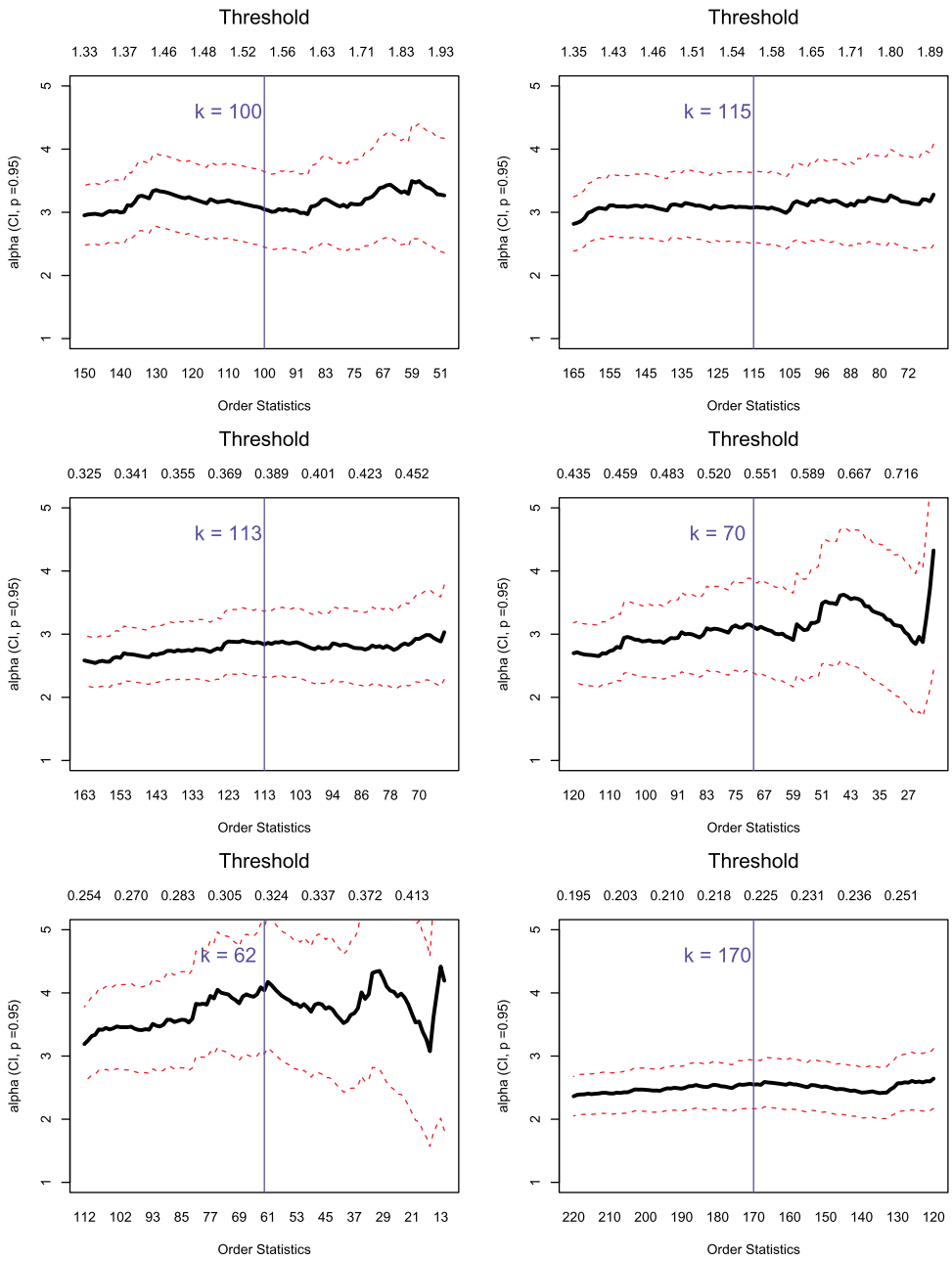

Figure A3. Hill plots for absolute values of the sample FPC scores for Walmart (left) and IBM (right). From top to bottom: levels $j = 1,2,3$. The vertical line shows the optimal k selected by the method of Hall [24].

of point-to-point returns, like daily or weekly returns. For the latter, the tail index can be in different ranges, and in times of financial crises may be even smaller than 1. It is possible that, for different assets or different time periods, even the tail index of the scores studied here is smaller than 2. Our theory would not apply to such data.