Available online at www.sciencedirect.com

SciVerse ScienceDirect

ELSEVIER

Stochastic Processes and their Applications 122 (2012) 910–929

stochastic processes and their applications

www.elsevier.com/locate/spa

Asymptotics for statistical functionals of long-memory sequences

Eric Beutner[a,*], Wei Biao Wu[b,1], Henryk Zähle[c]

[a] *Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, The Netherlands*

[b] *Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA*

[c] *Department of Mathematics, Saarland University, Postfach 151150, D-66041 Saarbrücken, Germany*

Received 14 March 2011; received in revised form 29 July 2011; accepted 13 October 2011
Available online 24 October 2011

Abstract

We present two general results that can be used to obtain asymptotic properties for statistical functionals based on linear long-memory sequences. As examples for the first one we consider L- and V-statistics, in particular tail-dependent L-statistics as well as V-statistics with unbounded kernels. As an example for the second result we consider degenerate V-statistics. To prove these results we also establish a weak convergence result for empirical processes of linear long-memory sequences, which improves earlier ones.

Keywords: Noncentral limit theorem; Linear long-memory sequence; Quasi-Hadamard differentiability; Modified Functional Delta Method; Weighted empirical process; L-statistic; U- and V-statistics; Degenerate V-statistic

1. Introduction

The appearance of *strongly* dependent data, i.e. data with long-memory, has been observed in many areas, such as climate warming, economics and finance; cf. [3–5,27]. A lot of research has focused on statistical inferences for long-memory sequences; for instance [5,19,22–24]. However, for several statistics, including L-statistics as well as U- and V- (von Mises-) statistics, the asymptotic distribution has only been established in some special cases. In [10] degenerate

* Corresponding author.
E-mail addresses: e.beutner@maastrichtuniversity.nl (E. Beutner), wbwu@galton.uchicago.edu (W.B. Wu), zaehle@math.uni-sb.de (H. Zähle).

[1] Wei Biao Wu's research was supported in part from DMS-0906073 and DMS-1106970.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2011.10.006

U-statistics of transformations of Gaussian sequences were considered, and in [20] the limit of U-statistics with bounded kernel was derived. But for general L-, U- and V-statistics there seem to be no respective results in the literature so far. In this article, we will present two general theorems that can be used to derive the asymptotic distribution of statistical functionals based on linear long-memory sequences. We will also demonstrate that the results yield in particular noncentral limit theorems (NCLTs) for tail-dependent L-statistics (cf. Example 3.1) as well as U- and V-statistics with unbounded kernels (cf. Example 3.3).

Before presenting our results, we briefly explain the methods which will be used for deriving them. The first method we explain can be used to establish an NCLT for, among others, L-statistics as well as nondegenerate U- and V-statistics (Theorem 2.3). It is well known that L- and V-statistics can be expressed as $T(\hat{F}_n)$ for some functional $T : \mathbb{F} \to \mathbf{V}$, where $\mathbb{F}$ is a class of distribution functions (DFs) on the real line, $\mathbf{V}$ is a vector space (in fact for L- and V-functionals we have $\mathbf{V} = \mathbb{R}$) and $\hat{F}_n$ is the empirical DF at stage n of the underlying data. Now, roughly speaking, if T is Hadamard differentiable at F, then by the Functional Delta Method (FDM; [15,17,26]) the asymptotic distribution of $T(\hat{F}_n)$ can be expressed by the asymptotic distribution of $\hat{F}_n$. But the FDM was repeatedly criticized for its restricted range of applications since many *tail-dependent* statistical functionals T, including popular L- and V-functionals, are known to be non-Hadamard differentiable at F. However, recently the concept of *quasi-Hadamard differentiability* was introduced in [6]. This is a weaker concept of differentiability than Hadamard differentiability, but it is still strong enough to obtain an FDM; cf. [6, Section 4]. In [6], the latter was called *Modified FDM*. In particular, it can be shown that general L- and V-functionals are quasi-Hadamard differentiable and hence that their asymptotic distributions can be obtained by the Modified FDM; cf. [6,7].

The basic idea of quasi-Hadamard differentiability is to impose a norm only on a suitable subspace $\mathbb{D}_0$ of the space $\mathbb{D}$ of all bounded càdlàg functions on $\overline{\mathbb{R}}$ (and not on all of $\mathbb{D}$), and to differentiate only in directions which lie in (some subset of) $\mathbb{D}_0$. It should be stressed that this is not simply the notion of *tangential* Hadamard-differentiability where the tangential space is equipped with the same norm as the space in which F lies. The crucial point is that norms, which assign to F a finite length, are often not strict enough to obtain "differentiability". On the other hand, "differentiability" w.r.t. such good-natured norms is typically not necessary. For details the reader is referred to the introduction of [6]. Upon having established quasi-Hadamard differentiability of a given statistical functional T, an application of the Modified FDM typically requires to prove a weak convergence result for the underlying empirical process w.r.t. a norm being stricter than the sup-norm $\|\cdot\|_\infty$, for instance w.r.t. a weighted sup-norm $\|\cdot\|_\lambda := \|(\cdot)\phi_\lambda\|_\infty$ with $\phi_\lambda(x) := (1 + |x|)^\lambda$ for some $\lambda > 0$. Here λ depends on the statistical functional whose quasi-Hadamard differentiability one wants to prove. Hence in the context of strongly dependent data, the crucial point is an NCLT for weighted empirical processes, which will be given in Theorem 2.1. Corresponding CLTs can be found in [30] for independent data, in [8] for weakly dependent β-mixing data, in [29] for weakly dependent α- and ρ-mixing data, and in [35] for weakly dependent causal data.

Let us now turn to the second method (Theorem 2.4), which can be used to obtain an NCLT for, among others, degenerate U- and V-statistics. To determine the asymptotic distribution of U-statistics with a degenerate kernel, it was used in [9] that for kernels $g : \mathbb{R}^2 \to \mathbb{R}$ of bounded variation the corresponding U-statistic can be represented as

$$\iint (\hat{F}_n - F)(x_1)(\hat{F}_n - F)(x_2)\, dg(x_1, x_2) \tag{1}$$

with $\hat{F}_n - F$ the empirical error process; an application of the Continuous Mapping Theorem then yields the asymptotic distribution (provided the asymptotics of $\hat{F}_n - F$ is known). Obviously, for g of locally bounded variation only, the map $\Phi : \mathbb{D} \to \mathbb{R}$, $\Phi(f) := \iint f(x_1) f(x_2)\, dg(x_1, x_2)$ is not continuous when using the sup-norm $\|\cdot\|_\infty$. However, if we use the weighted sup-norm $\|\cdot\|_\lambda$ (see above) for some $\lambda > 0$ and if we require $\iint \phi_{-\lambda}(x)\phi_{-\lambda}(y)\, dg(x, y)$ to be well defined, then we can still apply the Continuous Mapping Theorem, although g might only be of locally bounded variation. This concept also applies to other functionals that admit a representation similar to (1).

2. NCLTs based on long-memory sequences

Let

$$X_t := \sum_{s=0}^{\infty} a_s\, \varepsilon_{t-s}, \quad t \in \mathbb{N}, \tag{2}$$

where $(\varepsilon_i)_{i\in\mathbb{Z}}$ are i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with zero mean and finite variance, and the coefficients a_s satisfy $\sum_{s=0}^{\infty} a_s^2 < \infty$ (so that $(X_t)_{t\in\mathbb{N}}$ is an L^2-process). The sequence $(X_t)_{t\in\mathbb{N}}$ is stationary, and we denote by F its marginal DF. Many important time series models, such as the autoregressive moving average (ARMA) and fractional autoregressive integrated moving average (FARIMA), take this form. If $a_0 = 1$ and $a_1 = a_2 = \cdots = 0$, then the X_t are i.i.d. If a_t decays to zero at a sufficiently fast rate, then the covariances $\mathbb{C}\mathrm{ov}(X_0, X_t)$ are absolutely summable over $t \in \mathbb{N}$ and thus the process exhibits short-range dependence (weak dependence). If a_t decays to zero at a sufficiently slow rate, then the covariances $\mathbb{C}\mathrm{ov}(X_0, X_t)$ are *not* absolutely summable over $t \in \mathbb{N}$ and thus the process exhibits long-range dependence (strong dependence).

Our starting point for the derivation of a limit theorem for $T(\hat{F}_n)$ is a limit theorem for the empirical DF $\hat{F}_n := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[X_i,\infty)}$. If the X_t are i.i.d., then it is commonly known that the empirical process $n^{1/2}(\hat{F}_n - F)$ converges in distribution to an F-Brownian bridge, i.e. to a centered Gaussian process with covariance function $\Gamma(s, t) = F(s \wedge t)(1 - F(s \vee t))$. If the X_t are subject to a certain mixing condition (weak dependence), then the limit in distribution of the empirical process $n^{1/2}(\hat{F}_n - F)$ is known to be a centered Gaussian process with covariance function

$$\begin{aligned}\Gamma(s, t) = F(s \wedge t)(1 - F(s \vee t)) + \sum_{k=2}^{\infty} \big[&\mathbb{C}\mathrm{ov}\left(\mathbb{1}_{\{X_1 \le s\}}, \mathbb{1}_{\{X_k \le t\}}\right) \\ &+ \mathbb{C}\mathrm{ov}\left(\mathbb{1}_{\{X_1 \le t\}}, \mathbb{1}_{\{X_k \le s\}}\right)\big];\end{aligned}$$

see [8,12,29,35]. If the X_t exhibit long-range dependence (strong dependence, long-memory), then the situation changes drastically: Assuming a moving average structure (2) with $a_s = s^{-\beta}\,\ell(s)$, $s \ge 1$, for $\beta \in (\frac{1}{2}, 1)$ and a slowly varying function ℓ, and some additional regularity and moment conditions on the distribution of ε_0, one has

$$r_n\big(\hat{F}_n(\cdot) - F(\cdot)\big) \xrightarrow{\ \mathrm{d}\ } c_\beta\, f(\cdot) Z \quad (\text{in } (\mathbb{D}, \mathcal{D}, \|\cdot\|_\infty)) \tag{3}$$

where Z is a standard normally distributed random variable, f is the Lebesgue density of F, (r_n) is a norming sequence depending on the dependence structure of the X_t and increasing slower than $n^{1/2}$ ("noncentral rate"), c_β is some constant, and $\mathcal{D}$ is the σ-algebra on $\mathbb{D}$ generated by the usual coordinate projections; see e.g. [9,18,19,31,32]. Notice the asymptotic degeneracy of the limit process in (3) which shows that the increments of the standardized empirical DF $\hat{F}_n$ over disjoint intervals, or disjoint observation sets, are asymptotically completely correlated.

However, for our purposes, as explained in the Introduction, the use of the sup-norm $\|\cdot\|_\infty$ in (3) is insufficient. We need a corresponding result for the weighted sup-norm $\|\cdot\|_\lambda := \|(\cdot)\phi_\lambda\|_\infty$ based on the weight function $\phi_\lambda(x) := (1+|x|)^\lambda$ for some $\lambda > 0$. Such a result can be proved using methods of [34]; see Theorem 2.1. For $\lambda \geq 0$, let $\mathbb{D}_\lambda$ be the space of all càdlàg functions ψ on $\overline{\mathbb{R}}$ with $\|\psi\|_\lambda < \infty$, and $\mathbb{C}_\lambda$ be the subspace of all continuous functions in $\mathbb{D}_\lambda$. We equip $\mathbb{D}_\lambda$ with the σ-algebra $\mathcal{D}_\lambda := \mathcal{D} \cap \mathbb{D}_\lambda$ to make it a measurable space, where as before $\mathcal{D}$ is the σ-algebra generated by the usual coordinate projections $\pi_x : \mathbb{D} \to \mathbb{R}$, $\psi \mapsto \psi(x)$. Without loss of generality we assume $a_0 = 1$.

Theorem 2.1. *Let $\lambda \geq 0$, and assume that*

(a) $a_s = s^{-\beta}\,\ell(s)$, $s \in \mathbb{N}$, *where* $\beta \in (\frac{1}{2}, 1)$ *and* ℓ *is slowly varying at infinity,*
(b) $\mathbb{E}[|\varepsilon_0|^{2+2\lambda}] < \infty$,
(c) *the DF* G *of* ε_0 *is twice differentiable and* $\sum_{j=1}^{2}\int |G^{(j)}(x)|^2\phi_{2\lambda}(x)\,dx < \infty$.

Then we have the following analogue of (3):

$$r_n\big(\hat{F}_n(\cdot) - F(\cdot)\big) \xrightarrow{\ \mathrm{d}\ } c_{1,\beta}\, f(\cdot) Z \quad (in\ (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)), \tag{4}$$

where $r_n := n^{\beta-1/2}\,\ell(n)^{-1}$, f *is the Lebesgue density of* F, Z *is a standard normally distributed random variable, and*

$$c_{1,\beta} := \left(\mathbb{E}[\varepsilon_0^2]\,\frac{\left(1-\left(\beta-\frac{1}{2}\right)\right)(1-(2\beta-1))}{\int_0^\infty (x+x^2)^{-\beta}dx}\right)^{1/2}.$$

Notice that assumption (c) implies in particular that the DF F of X_0 is differentiable with derivative $f \in \mathbb{C}_\lambda$. The proof of Theorem 2.1 can be found in the Appendix A. In the case $\lambda = 0$ (sup-norm), and under stronger moment assumptions on ε_0, the convergence in (4) is already known from [19,34,18]. In [11] one can find a proof of (4) for $\lambda = 0$ and a sequence which is given by a sum of a linear long-memory sequence and a weakly dependent nonlinear Bernoulli shift. Even earlier and still in the case $\lambda = 0$, the convergence in (4) was established in [9] where $X_t = G(Y_t)$ for some mean zero, stationary Gaussian sequence (Y_t) with long-memory and some measurable function G satisfying certain conditions. Finally we note that in [22] the convergence in (4) for $\lambda = 0$ is extended to the infinite variance case, where Z is not necessarily Gaussian but only symmetric and α-stable.

Remark 2.2. We note that assumption (c) in Theorem 2.1 can be relaxed in that it suffices to require that there is some $m \in \mathbb{N}$ such that the DF G_m of $\overline{X}_{m,0} := \sum_{s=0}^{m-1} a_s\varepsilon_{m-s}$ is twice differentiable and satisfies $\sum_{j=1}^{2}\int |G_m^{(j)}(x)|^2\phi_{2\lambda}(x)\,dx < \infty$. The proof still works in this setting; see also [34]. □

As a consequence of Theorem 2.1 and the Modified FDM given in [6, Theorem 4.1] we will obtain an NCLT for statistical functionals; cf. Theorem 2.3. Let $\mathbb{F}$ be a class of DF on the real line containing F, $(\mathbf{V}, \|\cdot\|_{\mathbf{V}})$ be a normed vector space, $\mathcal{V}$ be a σ-algebra on $\mathbf{V}$ not larger than the Borel σ-algebra on $\mathbf{V}$, and $T : \mathbb{F} \to \mathbf{V}$ be a mapping. Theorem 2.3 involves the notion of quasi-Hadamard differentiability. For the reader's convenience we recall the definition from [6, Definition 2.1]. If $(\mathbb{D}_0, \|\cdot\|_{\mathbb{D}_0})$ is some normed subspace of $\mathbb{D}$ and $\mathbb{C}_0$ is some subset of $\mathbb{D}_0$, then T is said to be *quasi-Hadamard differentiable at* $F \in \mathbb{F}$ *tangentially to* $\mathbb{C}_0\langle\mathbb{D}_0\rangle$ if there is some continuous mapping $D^{\mathrm{Had}}_{F;\mathbb{C}_0\langle\mathbb{D}_0\rangle}T : \mathbb{C}_0 \to \mathbf{V}$ such that

$$\lim_{n\to\infty} \left\| D^{\mathrm{Had}}_{F;\mathbb{C}_0\langle\mathbb{D}_0\rangle}T\,(v) - \frac{T(F + h_n v_n) - T(F)}{h_n} \right\|_{\mathbf{V}} = 0 \tag{5}$$

holds for each triplet $(v, (v_n), (h_n))$, with $v \in \mathbb{C}_0$, $(v_n) \subset \mathbb{D}_0$ satisfying $\|v_n - v\|_{\mathbb{D}_0} \to 0$ as well as $F + h_n v_n \in \mathbb{F}$ for every $n \in \mathbb{N}$, and $(h_n) \subset (0, \infty)$ satisfying $h_n \to 0$. In this case the mapping $D^{\mathrm{Had}}_{F;\mathbb{C}_0\langle\mathbb{D}_0\rangle}T$ is called quasi-Hadamard derivative of T at F tangentially to $\mathbb{C}_0\langle\mathbb{D}_0\rangle$.

Theorem 2.3. *Let* $\lambda \geq 0$*, and assume that*

(i) $\hat{F}_n$ *takes values only in* $\mathbb{F}$*,*
(ii) *the assumptions of Theorem 2.1 are fulfilled,*
(iii) $\widetilde{\omega} \mapsto T(W(\widetilde{\omega}) + F)$ *is* $(\widetilde{\mathcal{F}}, \mathcal{V})$*-measurable whenever* W *is a measurable mapping from some measurable space* $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ *to* $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$ *such that* $W(\widetilde{\omega}) + \theta \in \mathbb{F}$ *for all* $\widetilde{\omega} \in \widetilde{\Omega}$*,*
(iv) T *is quasi-Hadamard differentiable at* F *tangentially to* $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ *(in the sense of* [6, *Definition 2.1*]*) with quasi-Hadamard derivative* $D^{\mathrm{Had}}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}T$.

Then

$$r_n \left(T(\hat{F}_n(\cdot)) - T(F(\cdot))\right) \xrightarrow{\mathrm{d}} D^{\mathrm{Had}}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}T(c_{1,\beta}\, f(\cdot)Z) \quad (in\ (\mathbf{V}, \mathcal{V}, \|\cdot\|_{\mathbf{V}})), \tag{6}$$

where $r_n, c_{1,\beta}, f$ *and* Z *are as in Theorem 2.1.*

Proof. Assumptions (i)–(iv) exactly match the assumptions of the Modified FDM, i.e. of Theorem 4.1 in [6], in our particular setting. The Modified FDM (which still holds when replacing $\sqrt{n}$ by r_n) thus ensures that (6) holds. □

In some situations the quasi-Hadamard derivative vanishes (cf. Example 3.3 and Section 3.3), so that in these cases the criterion of Theorem 2.3 yields little. However, sometimes one can use the following Theorem 2.4 instead of Theorem 2.3. An application of Theorem 2.4 can be found in Section 3.3.

Theorem 2.4. *Let* $\lambda \geq 0$*, and assume that*

(i) $\hat{F}_n$ *takes values only in* $\mathbb{F}$*,*
(ii) *the assumptions of Theorem 2.1 are fulfilled,*
(iii) *for some* $\gamma > 0$ *and some* $(\|\cdot\|_\lambda, \|\cdot\|_{\mathbf{V}})$*-continuous mapping* $\Psi : \mathbb{D}_\lambda \to \mathbf{V}$*,*

$$r^\gamma\,(T(\hat{F}_n(\cdot)) - T(F(\cdot))) = \Psi(r^\gamma(\hat{F}_n(\cdot) - F(\cdot))) \quad \forall\, n \in \mathbb{N},\ r > 0.$$

Then

$$r_n^\gamma\,(T(\hat{F}_n(\cdot)) - T(F(\cdot))) \xrightarrow{\mathrm{d}} \Psi(c_{1,\beta}\, f(\cdot)Z) \quad (in\ (\mathbf{V}, \mathcal{V}, \|\cdot\|_{\mathbf{V}})), \tag{7}$$

where $r_n, c_{1,\beta}, f$ *and* Z *are as in Theorem 2.1.*

Proof. Assumption (ii) ensures (4). Assumption (iii), the Continuous Mapping Theorem and (4) then yield (7). □

3. Examples

3.1. L-functionals

Let K be the DF of a probability measure on $([0,1], \mathcal{B}([0,1]))$, and $\mathbb{F}_K$ be the class of all DFs F on the real line for which $\int |x|\, dK(F(x)) < \infty$. The functional $\mathcal{L}$, defined by

$$\mathcal{L}(F) := \mathcal{L}_K(F) := \int x\, dK(F(x)), \quad F \in \mathbb{F}_K, \tag{8}$$

is called L-functional associated with K; cf., e.g., [28, p. 265]. The value $\mathcal{L}(F)$ can be seen as the mean of the distorted DF $K \circ F$ on $\mathbb{R}$. It was shown in [6] that if K is continuous and piecewise differentiable, the (piecewise) derivative K' is bounded above and $F \in \mathbb{F}_K$ takes the value $d \in (0,1)$ at most once if K is not differentiable at d, then for every $\lambda > 1$ the functional $\mathcal{L} : \mathbb{F}_K \to \mathbb{R}$ is quasi-Hadamard differentiable at F tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with quasi-Hadamard derivative

$$D^{\mathrm{Had}}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}\mathcal{L}(v) = -\int K'(F(x))\, v(x)\, dx \quad \forall\, v \in \mathbb{C}_\lambda.$$

Thus, if also the assumptions of Theorem 2.1 are fulfilled with $f \in \mathbb{C}_\lambda$, Theorem 2.3 (with $\mathbf{V} = \mathbb{R}$) yields

$$r_n\left(\mathcal{L}(\hat{F}_n) - \mathcal{L}(F)\right) \stackrel{\mathrm{d}}{\longrightarrow} \widetilde{Z} \quad (\text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R}))), \tag{9}$$

where $\widetilde{Z}$ is normally distributed with mean zero and variance $c_{1,\beta}^2(\int K'(F(x))f(x)dx)^2$, and r_n and $c_{1,\beta}$ are as in Theorem 2.1.

Example 3.1 (*Average Value at Risk*)**.** In mathematical finance, L-functionals are also known as distortion risk measures, and K is often referred to as distortion function. The risk measure $\mathcal{L}_K$ is coherent in the sense of [1] if and only if K is convex; cf. [33]. Since for every convex K the right boundary of the compact support of the probability measure dK is 1, every coherent distortion risk measure depends on the right tail of the argument. Such risk measures cannot be treated by the classical FDM, because, roughly speaking, tail-dependent functionals are *not* Hadamard differentiable w.r.t. the sup-norm $\|\cdot\|_\infty$; for details see the introduction of [6]. On the other hand, the FDM was modified in [6] in order to obtain also the asymptotic distribution of plug-in estimators for general distortion risk measures. A very popular example for a coherent distortion risk measure with K satisfying the assumptions stated subsequent to (8) is the Average Value at Risk (also called Expected Shortfall) at level $\alpha \in (0,1)$. The latter corresponds to the distortion function $K(x) = \frac{1}{1-\alpha}\max\{0, x-\alpha\}$. In this case the variance of $\widetilde{Z}$ in (9) is given by

$$\frac{c_{1,\beta}^2}{(1-\alpha)^2}\left(\int_{F^{\rightarrow}(\alpha)}^{\infty} f(x)\, dx\right)^2,$$

where $F^{\rightarrow}$ denotes the right-continuous inverse of F. □

3.2. V-functionals

Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function, and $\mathbb{F}_g$ be the class of all DFs F on the real line for which $\iint |g(x_1, x_2)| dF(x_1)dF(x_2) < \infty$. The functional $\mathcal{V}$, defined by

$$\mathcal{V}(F) := \mathcal{V}_g(F) := \iint g(x_1, x_2)\, dF(x_1)dF(x_2), \quad F \in \mathbb{F}_g, \tag{10}$$

is called V-functional associated with g. For background see, e.g., [25]. Let $\mathbb{BV}_{\mathrm{loc,rc}}$ be the space of all functions $\psi : \mathbb{R} \to \mathbb{R}$ that are right-continuous and locally of bounded variation. For $\psi \in \mathbb{BV}_{\mathrm{loc,rc}}$, we denote by $d\psi^+$ and $d\psi^-$ the unique positive Radon measures induced by the Jordan decomposition of ψ, and we set $|d\psi| := d\psi^+ + d\psi^-$. We impose the following assumptions.

Assumption 3.2. For some $\lambda > \lambda' \geq 0$ the following two assertions hold

(a) For every $x_2 \in \mathbb{R}$ fixed, the function $g_{x_2}(\cdot) := g(\cdot\,, x_2)$ lies in $\mathbb{BV}_{\mathrm{loc}} \cap \mathbb{D}_{-\lambda'}$. Moreover, the function $x_2 \mapsto \int \phi_{-\lambda}(x_1)|dg_{x_2}|(x_1)$ is measurable and finite w.r.t. $\|\cdot\|_{-\lambda'}$.
(b) The functions $g_{1,F}(\cdot) := \int g(\cdot\,, x_2)dF(x_2)$ and $g_{2,F}(\cdot) := \int g(x_1,\,\cdot)dF(x_1)$ lie in $\mathbb{BV}_{\mathrm{loc,rc}}$, and we have $\int \phi_{-\lambda}(x)\,|dg_{i,F}|(x) < \infty$ for $i = 1, 2$. Moreover, the functions $\overline{g_{1,F}}(\cdot) := \int |g(\cdot\,, x_2)|dF(x_2)$ and $\overline{g_{2,F}}(\cdot) := \int |g(x_1,\,\cdot)|dF(x_1)$ lie in $\mathbb{D}_{-\lambda'}$.

It is shown in [7] that under Assumption 3.2 the functional $\mathcal{V}$ is quasi-Hadamard differentiable at F tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with quasi-Hadamard derivative

$$D^{\mathrm{Had}}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}\mathcal{V}(v) = -\int v(x)dg_{1,F}(x) - \int v(x)dg_{2,F}(x) \quad \forall\, v \in \mathbb{C}_\lambda. \tag{11}$$

Thus, if also the assumptions of Theorem 2.1 are fulfilled with $f \in \mathbb{C}_\lambda$, Theorem 2.3 (with $\mathbf{V} = \mathbb{R}$) yields

$$r_n\left(\mathcal{V}(\hat{F}_n) - \mathcal{V}(F)\right) \xrightarrow{\mathrm{d}} \widetilde{Z} \quad (\text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R}))), \tag{12}$$

where $\widetilde{Z}$ is normally distributed with mean zero and variance $c^2_{1,\beta}(\int f(x)dg_{1,F}(x) + \int f(x)dg_{2,F}(x))^2$, and r_n and $c_{1,\beta}$ are as in Theorem 2.1.

Example 3.3. It is easy to show that the variance kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ and Gini's mean difference kernel $g(x_1, x_2) = |x_1 - x_2|$ satisfy conditions (a)–(b) in Assumption 3.2 for $\lambda' = 2$ and $\lambda' = 1$, respectively; cf. [7]. In the former case, however, it is straightforwardly seen that the asymptotic variance in (12) vanishes, so that the right-hand side in (12) degenerates to zero. This is consistent with Example 1 in [10]. □

Remark 3.4. Notice that the V-statistic $\mathcal{V}(\hat{F}_n) = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n g(X_i, X_j)$ slightly differs from the U-statistic $\mathcal{U}_n := \frac{1}{n(n-1)}\sum_{i=1}^n \sum_{i \neq j=1}^n g(X_i, X_j)$. However, our method is suitable also for U-statistics; see Remark 2.5 in [7] and note that r_n grows slower than $\sqrt{n}$. □

Remark 3.5. An NCLT for U- and V-statistics has already been established in [20, Section 5(b)] using different techniques. However, the assumptions there are more technical and more restrictive. In particular, the kernel g has to be bounded. □

3.3. Degenerate V-functionals

Among V-functionals (introduced in Section 3.2) the functionals with a so-called degenerate kernel have attracted special interest; see, e.g., [9,10,13]. A kernel g is degenerate w.r.t. $F \in \mathbb{F}_g$ if the functions $g_{1,F}$ and $g_{2,F}$ defined in part (b) of Assumption 3.2 are identically zero. In this case, we refer to $\mathcal{V}$ (defined in (10)) as a *degenerate* V-functional w.r.t. F. Moreover, in this case the right-hand side in (11) vanishes and thus the right-hand side in (12) degenerates to zero. Nevertheless one can establish a nondegenerate NCLT for $\mathcal{V}(\hat{F}_n)$. In contrast to the considerations in Section 3.2, where the derivation of the asymptotic distribution of $\mathcal{V}(\hat{F}_n)$ relies on quasi-Hadamard differentiability and Theorem 2.3, we will now exploit the degeneracy of the kernel g and Theorem 2.4. The crucial points will be that the degeneracy of the kernel g leads to the representation

$$\mathcal{V}(\hat{F}_n) = \iint g(x_1, x_2)\, d(\hat{F}_n - F)(x_1) d(\hat{F}_n - F)(x_2) \tag{13}$$

and that under certain conditions on g and F this equals

$$\mathcal{V}(\hat{F}_n) = \iint (\hat{F}_n - F)(x_1)(\hat{F}_n - F)(x_2)\, dg(x_1, x_2). \tag{14}$$

The representation (13) was pointed out in [9, Section 2].

Moreover, in [9] it was also pointed out that, using integration-by-parts, relation (14) holds true. To apply integration-by-parts, it was assumed in [9] that the kernel g is right-continuous and has bounded total variation. However, as the assumption that g be of bounded total variation is too restrictive, the result of [9, Section 2] was extended in [10] to more general kernels. In [10] it was shown, that the result of [9, Section 2] can be extended to kernels g such that $[g]$ has finite $\|\cdot\|_F$-norm (for the definition and properties of $[g]$ and $\|\cdot\|_F$, respectively, see [10]). The extension was based on the fact that kernels g, for which $[g]$ has finite $\|\cdot\|_F$-norm, can be approximated by kernels g that have bounded total bivariation.

Here, we proceed differently, and it seems that by tendency the method presented here covers more examples; see Remark 3.8. Instead of approximating functions with unbounded total bivariation we extend the integration-by-parts formula, which was used in [10] to establish equality of (13) and (14), to right-continuous kernels with unbounded total bivariation. To this end, recall that a function $g : \mathbb{R}^2 \to \mathbb{R}$ is said to be of *locally* bounded bivariation if for every half-open rectangle $(a_1, b_1] \times (a_2, b_2]$, with $a = (a_1, a_2),\ b = (b_1, b_2) \in \mathbb{R}^2$,

$$\sup_{\Pi} \sum_{(x_1,x_2]\times(y_1,y_2]\in\Pi} \left|g(x_2, y_2) - g(x_1, y_2) - g(x_2, y_1) + g(x_1, y_1)\right| < \infty,$$

where the supremum is taken over all partitions Π of $(a_1, b_1] \times (a_2, b_2]$ consisting of finitely many half-open rectangles. We denote by $\mathbb{BV}^2_{\mathrm{loc,rc}}$ the space of all upper right-continuous functions $g : \mathbb{R}^2 \to \mathbb{R}$ of locally bounded bivariation. For $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$, we set $|dg| := dg^+ + dg^-$ with dg^+ and dg^- the unique positive Radon measures induced by the Jordan decomposition of g into the difference of two bimonotonically increasing functions; cf. [16, Proposition 1.17]. The following lemma, which allows us to prove an NCLT for $\mathcal{V}(\hat{F}_n)$ for a degenerate kernel g, gives sufficient conditions for the validity of the equation

$$\iint g(x_1, x_2)\, dh(x_1)dh(x_2) = \iint h(x_1)h(x_2)\, dg(x_1, x_2). \tag{15}$$

The lemma is based on a general integration-by-parts formula given in the Appendix B.

Lemma 3.6. *Assume that, for some* $0 \le \lambda' < \lambda$,

(a) $g \in \mathbb{BV}^2_{\mathrm{loc,rc}}$, $K := \sup_{b_1,b_2\in\mathbb{R}} |\phi_{-\lambda'}(b_1)\phi_{-\lambda'}(b_2)g(b_1, b_2)|$ *is finite, and that the integral* $\iint \phi_{-\lambda}(x_1)\phi_{-\lambda}(x_2)\, |dg|(x_1, x_2)$ *is finite,*

(b) *the function* $g_{x_2}(\,\cdot\,) := g(\,\cdot\,, x_2)$ *satisfies* Assumption 3.2(a), *and the same holds for* $g_{x_1}(\,\cdot\,) := g(x_1,\,\cdot\,)$,

(c) $h \in \mathbb{D}_\lambda \cap \mathbb{BV}_{\mathrm{loc}}$, *and* $\iint |g(x_1, x_2)|\,|d\widetilde{h}|(x_1, x_2) < \infty$ *is finite, where* $\widetilde{h} : \mathbb{R}^2 \to \mathbb{R}$ *is defined by* $\widetilde{h}(x_1, x_2) := h(x_1)h(x_2)$,
(d) *the functions* $\widetilde{h}$ *and* g *have no joint discontinuities.*

Then (15) *holds.*

Proof. To prove (15) we show that the conditions of Lemma B.1 given in the Appendix B hold true for $v := g$ and $u := \widetilde{h}$ with $k_1 = k_2 = k_3 = 0$. Conditions (ii) and (iv) of Lemma B.1 hold true by assumptions (b) and (d). By condition (c) to verify condition (i) of Lemma B.1, it only remains to show that the integral $\iint |h(x_1)h(x_2)|\,|dg|(x_1, x_2)$ is finite. Since $h \in \mathbb{D}_\lambda$, we have

$$\iint |h(x_1)h(x_2)|\,|dg|(x_1, x_2) \le C^2 \iint \phi_{-\lambda}(x_1)\phi_{-\lambda}(x_2)\,|dg|(x_1, x_2),$$

where $C := \sup_{x_1} |\phi_\lambda(x_1)h(x_1)|$. Let us turn to condition (iii) of Lemma B.1. We first show that $k_1 = k_2 = 0$. We have

$$\begin{aligned}\left|\int_{a_1}^{b_1} h(x_1)h(b_2)\,dg_{b_2}(x_1)\right| &\le \left(\int |h(x_1)|\,|dg_{b_2}|(x_1)\right)|h(b_2)|\,\phi_{-\lambda'}(b_2)\,\phi_{\lambda'}(b_2)\\ &\le C\left(\int \phi_{-\lambda}(x_1)\,|dg_{b_2}|(x_1)\phi_{-\lambda'}(b_2)\right)|h(b_2)|\,\phi_{\lambda'}(b_2),\end{aligned}$$

where C is as above. Hence $|\int_{a_1}^{b_1} h(x_1)h(b_2)\,dg_{b_2}(x_1)| \to 0$ as $a_i \to -\infty, b_i \to \infty, i = 1, 2$, because the mapping $b_2 \mapsto \int \phi_{-\lambda}(x_1)\,|dg_{b_2}|(x_1)$ is finite w.r.t. $\|\cdot\|_{-\lambda'}$, $\|h\|_\lambda$ is finite, and $\lambda' \in [0, \lambda)$. Applying the same arguments to the other terms appearing in the definition of k_1 and k_2 in condition (iii) of Lemma B.1, we obtain that $k_1 = k_2 = 0$. Finally we show that $k_3 = 0$. Since $|h(b_1)h(b_2)\,g(b_1, b_2)| \le K|h(b_1)\,h(b_2)\,\phi_{\lambda'}(b_1)\phi_{\lambda'}(b_2)|$ by assumption (a), we obtain that $|h(b_1)h(b_2)\,g(b_1, b_2)|$ converges to zero as $b_1, b_2 \to \infty$ since $h \in \mathbb{D}_\lambda$. Similar arguments show that the three other terms in the definition of k_3 converge to zero, too. This finishes the proof. □

Now, we can provide an NCLT for $\mathcal{V}(\hat{F}_n)$ for a degenerate kernel g. Recall that $\mathcal{V}(F) = 0$ whenever g is degenerate, and notice that the factor on the left-hand side in (16) is r_n^2, which differs from the left-hand side in (12) where the factor is r_n. Recall also that $\mathbb{F}_g$ is the class of all DFs F on the real line for which $\iint |g(x_1, x_2)|dF(x_1)dF(x_2) < \infty$.

Theorem 3.7. *Assume that, for some* $\lambda > \lambda' \ge 0$,

(a) *The functions* $g_{1,F}$ *and* $g_{2,F}$ *defined in* Assumption 3.2(b) *are identically zero, i.e. the kernel* g *is degenerate,*
(b) *conditions* (a)–(b) *of* Lemma 3.6 *hold for* g,
(c) $F \in \mathbb{F}_g$, *and the assumptions of* Theorem 2.1 *are fulfilled,*
(d) *The sets* $D_1 := \{x \in \mathbb{R} : \exists\, y \in \mathbb{R}$ *such that* g *is discontinuous in* $(x, y)\}$ *and* $D_2 := \{y \in \mathbb{R} : \exists\, x \in \mathbb{R}$ *such that* g *is discontinuous in* $(x, y)\}$ *are* dF *null sets.*

Then

$$r_n^2\,\mathcal{V}(\hat{F}_n) \xrightarrow{\mathrm{d}} \left(c_{1,\beta}^2 \iint f(x_1)f(x_2)dg(x_1, x_2)\right)Z^2 \quad (in\ (\mathbb{R}, \mathcal{B}(\mathbb{R}))), \tag{16}$$

where Z^2 *is* χ_1^2*-distributed, and* r_n *and* $c_{1,\beta}$ *are as in* Theorem 2.1.

Proof. We adapt the arguments of [9, Section 2]. Under condition (a) we have the representation (13) for $\mathcal{V}(\hat{F}_n)$. By an application of Lemma 3.6 to g and $h := \hat{F}_n - F$, we can conclude from (13) that the alternative representation (14) holds $\mathbb{P}$-almost surely. We note that Lemma 3.6 can be applied because assumptions (a)–(b), (c), and (d) of Lemma 3.6 hold (to be exact, condition (d) holds only $\mathbb{P}$-almost surely) by conditions (b), (c), and (d). From (14) we immediately obtain (recall that $\mathcal{V}(F) = 0$ by the degeneracy of g which was imposed by condition (a))

$$r^2\,\mathcal{V}(\hat{F}_n(\cdot)) = r^2\,(\mathcal{V}(\hat{F}_n(\cdot)) - \mathcal{V}(F(\cdot))) = \Psi_g(r(\hat{F}_n(\cdot) - F(\cdot))) \qquad \mathbb{P}\text{-a.s.}$$

for every $n \in \mathbb{N}$ and $r > 0$, where $\Psi_g(v) := \iint v(x_1)\,v(x_2)\,dg(x_1, x_2)$, $v \in \mathbb{D}_\lambda$, v measurable. Since the mapping $\Psi_g : \mathbb{D}_\lambda \to \mathbb{R}$ is $(\|\cdot\|_\lambda, |\cdot|)$-continuous, Theorem 2.4 yields (16). □

Remark 3.8. Consider the kernel $g(x_1, x_2) = x_1^3 x_2^3$ and a distribution with DF F that is symmetric about 0. Then the corresponding V-functional is degenerate. Additionally, let us assume that the tails of F are of order $x^{-(3+\varepsilon)}$ for some $\varepsilon > 0$. Then, for every $\varepsilon > 0$ we can apply Theorem 3.7 provided the assumptions of Theorem 2.1 are fulfilled with $f \in \mathbb{C}_\lambda$. On the other hand, since g is differentiable, the $\|\cdot\|_F$-norm of $[g]$ equals (cf. [10, Lemma 2.1])

$$\iint \sqrt{F(x_1)(1 - F(x_1))}\sqrt{F(x_2)(1 - F(x_2))}\,x_1^2\,x_2^2\,dx_1\,dx_2.$$

Obviously, this quantity is not finite for every $\varepsilon > 0$ implying that the results of [10] cannot be applied to all $\varepsilon > 0$. □

Example 3.9. *(Goodness-of-fit test)* For a given DF F_0 and any measurable (weight) function $w : \mathbb{R} \to \mathbb{R}_+$, the weighted Cramér–von Mises test statistic

$$T_n^0 := \int w(x)\big(\hat{F}_n(x) - F_0(x)\big)^2\,dF_0(x)$$

was introduced for testing the null hypothesis $F = F_0$; see, e.g., [10, Example 3]. The test statistic T_n^0 can be expressed as V-statistic $\mathcal{V}(\hat{F}_n)$ with kernel

$$g(x_1, x_2) := \int w(x)\big(\mathbb{1}_{[x_1,\infty)}(x) - F_0(x)\big)\big(\mathbb{1}_{[x_2,\infty)}(x) - F_0(x)\big)\,dF_0(x)$$

and we have

$$dg(x_1, x_2) = dg^+(x_1, x_2) = |dg|(x_1, x_2) = \int w(x)\,\delta_{(x,x)}(dx_1, dx_2)\,dF(x);$$

see also [10, Example 3]. Moreover, under the null hypothesis $F = F_0$ the kernel g is obviously degenerate w.r.t. F. In this case, choosing $w = \phi_{2\lambda'}$ (implying that the assumptions on g in Theorem 3.7 are fulfilled), the double integral on the right-hand side in (16) reads as $\int f(x)^2 w(x)\,dF(x)$. □

Acknowledgments

The authors wish to thank the reviewers for their useful hints and comments.

Appendix A. Proof of Theorem 2.1

First of all we introduce some notation which will be needed for the proof of Theorem 2.1. Let $\mathcal{F}_t := \sigma(\varepsilon_s : s \in \mathbb{Z} \text{ with } s \le t)$ for every $t \in \mathbb{Z}$. For every $u \in \mathbb{Z}$ fixed with $u \le t - 1$, we define truncated processes $\overline{X}_{\cdot,u}$ and $\underline{X}_{\cdot,u}$ by $\overline{X}_{t,u} := \sum_{s=0}^{t-u-1} a_s \varepsilon_{t-s}$, $t \in \mathbb{N}$, and $\underline{X}_{t,u} := \sum_{s=t-u}^{\infty} a_s \varepsilon_{t-s} = \mathbb{E}[X_t|\mathcal{F}_u]$, $t \in \mathbb{N}_0$, respectively.

In view of the decomposition $X_t = \overline{X}_{t,u} + \underline{X}_{t,u}$ and the $\mathcal{F}_u$-measurability of $\underline{X}_{t,u}$, Theorem 5.4 in [21] yields

$$\mathbb{E}[\mathbb{1}_{[X_t,\infty)}(x)|\mathcal{F}_u] = F_{\overline{X}_{t,u}}(x - \underline{X}_{t,u}) \quad \mathbb{P}\text{-a.s.} \tag{17}$$

for all $u \leq t-1$, $t \in \mathbb{N}$, and $x \in \mathbb{R}$, where $F_{\overline{X}_{t,u}}$ denotes the DF of $\overline{X}_{t,u}$. For notational simplicity, we set $G_m := G_m^{(0)} := F_{\overline{X}_{m,0}}$ and denote by $G_m^{(1)}$ the derivative of G_m.

Now, let us turn to the actual proof of Theorem 2.1. For every $n \in \mathbb{N}$, we set $\sigma_n := n^{1-(\beta-1/2)}\ell(n)$ and

$$S_{n,0}(x) := n(\hat{F}_n(x) - F(x)), \quad x \in \mathbb{R}$$
$$S_{n,1}(x) := n(\hat{F}_n(x) - F(x)) + f(x)\sum_{i=1}^n X_i, \quad x \in \mathbb{R}.$$

Hence,

$$\frac{S_{n,0}(\cdot)}{\sigma_n} = \frac{S_{n,1}(\cdot)}{\sigma_n} - f(\cdot)\frac{\sum_{i=1}^n X_i}{\sigma_n}.$$

As $f \in \mathbb{D}_\lambda$, $\sigma_n^{-1} = n^{-1} r_n$ and $\frac{1}{\sigma_n}\sum_{i=1}^n X_i \xrightarrow{d} c_\beta Z$ (cf. [2, Theorem 2]; for the shape of c_β see [19, Lemma 6.1]), for the statement of Theorem 2.1 to be true it suffices to show that $\|\frac{S_{n,1}(\cdot)}{\sigma_n}\|_\lambda$ converges in probability to zero. In the remainder of this section we will show that this convergence holds.

We clearly have $S_{n,1} = M_n + T_n$, where

$$M_n(x) := \sum_{i=1}^n \Big(\mathbb{1}_{[X_i,\infty)}(x) - \mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_{i-1}]\Big), \quad x \in \mathbb{R}$$
$$\begin{aligned} T_n(x) &:= \sum_{i=1}^n \Big(\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_{i-1}] - F(x) + f(x)X_i\Big) \\ &= \sum_{i=1}^n \Big(F_{\overline{X}_{i,i-1}}(x - \underline{X}_{i,i-1}) - F(x) + f(x)X_i\Big) \\ &= \sum_{i=1}^n \Big(G_1(x - \underline{X}_{i,i-1}) - F(x) + f(x)X_i\Big), \quad x \in \mathbb{R} \end{aligned}$$

(recall (17) and note that $F_{\overline{X}_{i,i-1}} = F_{\overline{X}_{1,0}} = G_1$). By Slutzky's lemma, it thus suffices to show that both $\|\frac{M_n}{\sigma_n}\|_\lambda$ and $\|\frac{T_n}{\sigma_n}\|_\lambda$ converge in probability to zero.

As for $\|\frac{M_n}{\sigma_n}\|_\lambda$, we observe that for every $\varepsilon > 0$

$$\mathbb{P}\Big[\sigma_n^{-1}\|M_n\|_\lambda > \varepsilon\Big] \leq \frac{1}{\varepsilon^2}\frac{\mathbb{E}\big[\|M_n^2\|_{2\lambda}\big]}{n^{2-(2\beta-1)}\ell(n)^2}.$$

Since $\mathbb{E}[\|M_n^2\|_{2\lambda}] = \mathcal{O}(n\log^2 n)$ by Lemma 13 in [34], and $\beta \in (\frac{1}{2}, 1)$, we immediately obtain that $\|\frac{M_n}{\sigma_n}\|_\lambda$ converges in probability to zero.

As for $\|\frac{T_n}{\sigma_n}\|_\lambda$, we note that by Lemma 4 in [34]

$$\|T_n^2\|_{2\lambda} \leq 2^{1+4\lambda}\sum_{j=0}^1 \int \{T_n^{(j)}(x)\}^2 \phi_{2\lambda}(x)\,dx,$$

where $T_n^{(j)}$ denotes the jth derivative of T_n, i.e.

$$T_n^{(j)}(x) = \sum_{i=1}^{n}\Big(G_1^{(j)}(x - \underline{X}_{i,i-1}) - F^{(j)}(x) + f^{(j)}(x)X_i\Big) \quad \forall x \in \mathbb{R}.$$

Thus we have for every $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\Big[\sigma_n^{-1}\,\|T_n\|_\lambda > \varepsilon\Big] &\le \frac{\mathbb{E}\big[\,\|T_n^2\|_{2\lambda}\big]}{\varepsilon^2\sigma_n^2}\\ &\le \frac{2^{1+4\lambda}}{\varepsilon^2\sigma_n^2}\sum_{j=0}^{1}\int \mathbb{E}\Big[\{T_n^{(j)}(x)\}^2\Big]\phi_{2\lambda}(x)\,dx\\ &=: \frac{2^{1+4\lambda}}{\varepsilon^2\sigma_n^2}\sum_{j=0}^{1}\mathrm{I}_n(j).\end{aligned}$$

It follows from Lemma A.5 that $\mathrm{I}_n(j) = o(\sigma_n^2)$ for $j \in \{0, 1\}$. That is $\|\frac{T_n}{\sigma_n}\|_\lambda$ converges in probability to zero, which completes the proof of Theorem 2.1.

The proof of Lemma A.5 is based on Lemmas A.1–A.4, for which we need some notation. For every $k \in \mathbb{Z}$, we define the projection operator $\mathcal{P}_k : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{F}_k, \mathbb{P})$ by

$$\mathcal{P}_k(Y) := \mathbb{E}[Y|\mathcal{F}_k] - \mathbb{E}[Y|\mathcal{F}_{k-1}], \quad Y \in L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Notice that $\mathcal{P}_k(Y) = 0$ for every $Y \in L^1(\Omega, \mathcal{F}_{k-1}, \mathbb{P})$. For every $i \in \mathbb{N}$ and $x \in \mathbb{R}$, we define $L_i(x) := G_1(x - \underline{X}_{i,i-1}) - F(x) + f(x)X_i$. Hence, $T_n^{(j)}(x) = \sum_{i=1}^n L_i^{(j)}(x)$ and

$$\begin{aligned}L_i^{(j)}(x) &= G_1^{(j)}(x - \underline{X}_{i,i-1}) - F^{(j)}(x) + f^{(j)}(x)X_i\\ &= \frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_{i-1}] - F^{(j)}(x) + f^{(j)}(x)X_i\end{aligned}$$

for every $i \in \mathbb{N}$, $x \in \mathbb{R}$, and $j \in \{0, 1\}$.

Lemma A.1. *For every $n \in \mathbb{N}$, $x \in \mathbb{R}$, and $j \in \{0, 1\}$,*

$$T_n^{(j)}(x) = \sum_{k=-\infty}^{n}\mathcal{P}_k(T_n^{(j)}(x)) \quad \mathbb{P}\text{-a.s.}$$

Proof. Of course, it suffices to show $L_i^{(j)}(x) = \sum_{k=-\infty}^{n}\mathcal{P}_k(L_i^{(j)}(x))$ for every $i = 1, \ldots, n$. Since $L_i^{(j)}(x)$ is $\mathcal{F}_i$-measurable, we obtain

$$\begin{aligned}&\sum_{k=-\infty}^{n}\mathcal{P}_k(L_i^{(j)}(x))\\ &\quad= \lim_{l\to-\infty}\sum_{k=l+1}^{n}\Big(\mathbb{E}[L_i^{(j)}|\mathcal{F}_k] - \mathbb{E}[L_i^{(j)}|\mathcal{F}_{k-1}]\Big)\\ &\quad= L_i^{(j)}(x) - \lim_{l\to-\infty}\left\{\frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_l] - F^{(j)}(x) + f^{(j)}(x)\mathbb{E}[X_i|\mathcal{F}_l]\right\}.\end{aligned}$$

Since X_i is in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and therefore $\mathbb{P}$-almost surely finite, we obtain that $\mathbb{E}[X_i|\mathcal{F}_l] = \sum_{s=i-l}^{\infty} a_s\varepsilon_{i-s}$ converges $\mathbb{P}$-almost surely to zero as $l \to -\infty$. For the statement of Lemma A.1

to be true, it thus suffices to show that $\frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_l]$ converges $\mathbb{P}$-almost surely to $F^{(j)}(x)$ as $l \to -\infty$. By (17),

$$\frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_l] = F^{(j)}_{\overline{X}_{i,l}}(x - \underline{X}_{i,l}) = F^{(j)}_{\overline{X}_{i,l}}(x - \mathbb{E}[X_i|\mathcal{F}_l]). \tag{18}$$

Since $\lim_{l\to-\infty} \overline{X}_{i,l} = X_i$ and $f = F^{(1)}$ is continuous, we have $\lim_{l\to-\infty} F^{(j)}_{\overline{X}_{i,l}}(y) = F^{(j)}(y)$ uniformly in y on compact sets. As seen above, we further have $\lim_{l\to-\infty}\mathbb{E}[X_i|\mathcal{F}_l] = 0$ $\mathbb{P}$-almost surely. Thus, the expression in (18) converges indeed $\mathbb{P}$-almost surely to $F^{(j)}(x)$ as $l \to -\infty$. □

Lemma A.2. *For every $i \in \mathbb{N}$, $j \in \{0, 1\}$, $x \in \mathbb{R}$, and every $k \in \mathbb{Z}$ with $k < i$,*

$$\mathcal{P}_i(L_i^{(j)}(x)) = f^{(j)}(x)\varepsilon_i,$$
$$\mathcal{P}_k(L_i^{(j)}(x)) = G^{(j)}_{i-k}(x - \underline{X}_{i-k,0}) - G^{(j)}_{i-k+1}(x - \underline{X}_{i-k,-1}) + f^{(j)}(x)a_{i-k}\varepsilon_k,$$

where $G^{(j)}_m$ was defined to be the jth derivative of the DF of $\overline{X}_{m,0} := \sum_{s=0}^{m-1} a_s\varepsilon_{m-s}$.

Proof. Using (17) along with $i > k$, we obtain

$$\begin{aligned}
\mathcal{P}_k(L_i^{(j)}(x)) &= \mathcal{P}_k\left(\frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_{i-1}] - F^{(j)}(x) + f^{(j)}(x)X_i\right)\\
&= \frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_k] - F^{(j)}(x) + f^{(j)}(x)\mathbb{E}[X_i|\mathcal{F}_k]\\
&\quad - \frac{\partial^j}{\partial x^j}\mathbb{E}[\mathbb{1}_{[X_i,\infty)}(x)|\mathcal{F}_{k-1}] + F^{(j)}(x) - f^{(j)}(x)\mathbb{E}[X_i|\mathcal{F}_{k-1}]\\
&= \frac{\partial^j}{\partial x^j}F_{\overline{X}_{i,k}}(x - \underline{X}_{i,k}) + f^{(j)}(x)\sum_{s=i-k}^{\infty} a_s\varepsilon_{i-s}\\
&\quad - \frac{\partial^j}{\partial x^j}F_{\overline{X}_{i,k-1}}(x - \underline{X}_{i,k-1}) - f^{(j)}(x)\sum_{s=i-k+1}^{\infty} a_s\varepsilon_{i-s}\\
&= G^{(j)}_{i-k}(x - \underline{X}_{i,k}) - G^{(j)}_{i-k+1}(x - \underline{X}_{i,k-1}) + f^{(j)}(x)a_{i-k}\varepsilon_k\\
&= G^{(j)}_{i-k}(x - \underline{X}_{i-k,0}) - G^{(j)}_{i-k+1}(x - \underline{X}_{i-k,-1}) + f^{(j)}(x)a_{i-k}\varepsilon_k.
\end{aligned}$$

This proves the second identity. The first identity can be proved analogously, noting that we assumed $a_0 = 1$. □

Lemma A.3. *For every $n \in \mathbb{N}$ and $j \in \{0, 1\}$, we have $\mathrm{I}_n(j) \le \sum_{k=-\infty}^{n}\left(\sum_{i=1\vee k}^{n}\lambda_{i-k,j}\right)^2$, where, for $m \in \mathbb{N}$,*

$$\lambda_{0,j} := \left(\int \{f^{(j)}(x)\}^2\phi_{2\lambda}(x)\,dx\right)^{1/2}\mathbb{E}[\varepsilon_0^2]^{1/2},$$

$$\lambda_{m,j} := \left(\int \mathbb{E}\left[\left\{G^{(j)}_m(x - \underline{X}_{m,0}) - G^{(j)}_{m+1}(x - \underline{X}_{m,-1}) + f^{(j)}(x)a_m\varepsilon_0\right\}^2\right]\phi_{2\lambda}(x)\,dx\right)^{1/2}$$

with $G^{(j)}_m$ the jth derivative of the DF of $\overline{X}_{m,0} := \sum_{s=0}^{m-1} a_s\varepsilon_{m-s}$.

Proof. By Lemma A.1, we have $T_n^{(j)}(x) = \sum_{k=-\infty}^{n} \mathcal{P}_k(T_n^{(j)}(x))$. Further, it is easily seen that $\mathbb{E}[\mathcal{P}_k(T_n^{(j)}(x))\mathcal{P}_l(T_n^{(j)}(x))] = 0$ for every $k, l \leq n$ with $k \neq l$. Finally, notice that $\mathcal{P}_k(L_i^{(j)}(x)) = 0$ for all $i < k$ since $L_i^{(j)}(x)$ is $\mathcal{F}_i$-measurable. Then,

$$\begin{aligned}
\mathrm{I}_n(j) &= \int \mathbb{E}\Big[\{T_n^{(j)}(x)\}^2\Big]\phi_{2\lambda}(x)\,dx \\
&= \int \mathbb{E}\left[\left\{\sum_{k=-\infty}^{n} \mathcal{P}_k(T_n^{(j)}(x))\right\}^2\right]\phi_{2\lambda}(x)\,dx \\
&= \sum_{k=-\infty}^{n} \mathbb{E}\left[\int \{\mathcal{P}_k(T_n^{(j)}(x))\}^2 \phi_{2\lambda}(x)\,dx\right] \\
&= \sum_{k=-\infty}^{n} \mathbb{E}\left[\int \left\{\sum_{i=1\vee k}^{n} \mathcal{P}_k(L_i^{(j)}(x))\right\}^2 \phi_{2\lambda}(x)\,dx\right] \\
&= \sum_{k=-\infty}^{n} \mathbb{E}\left[\int \left\{\sum_{i=1\vee k}^{n} \frac{\mathcal{P}_k(L_i^{(j)}(x))}{\sqrt{\lambda_{i-k,j}}}\sqrt{\lambda_{i-k,j}}\right\}^2 \phi_{2\lambda}(x)\,dx\right] \\
&\leq \sum_{k=-\infty}^{n} \mathbb{E}\left[\int \left(\sum_{i=1\vee k}^{n} \frac{\{\mathcal{P}_k(L_i^{(j)}(x))\}^2}{\lambda_{i-k,j}}\right)\left(\sum_{i=1\vee k}^{n} \lambda_{i-k,j}\right)\phi_{2\lambda}(x)\,dx\right] \\
&= \sum_{k=-\infty}^{n} \left(\sum_{i=1\vee k}^{n} \frac{\int \mathbb{E}[\{\mathcal{P}_k(L_i^{(j)}(x))\}^2]\phi_{2\lambda}(x)\,dx}{\lambda_{i-k,j}}\right)\left(\sum_{i=1\vee k}^{n} \lambda_{i-k,j}\right) \\
&= \sum_{k=-\infty}^{n} \left(\sum_{i=1\vee k}^{n} \lambda_{i-k,j}\right)^2,
\end{aligned}$$

where the "$\leq$" follows from Hölder's inequality and $f^{(j)}(\cdot)a_m\varepsilon_{i-m} \overset{d}{=} f^{(j)}(\cdot)a_m\varepsilon_0$. The last equality is immediate from Lemma A.2. □

Lemma A.4. *For $j \in \{0, 1\}$, we have $\lambda_{m,j} = o(a_m)$, where $\lambda_{m,j}$ is as in Lemma A.3.*

Proof. Let $\varepsilon_0' \sim G$ be independent of $(\varepsilon_s)_{s\in\mathbb{Z}}$, probably defined on an extension of the original probability space. Since $G_{m+1}^{(j)}(x - \underline{X}_{m,-1}) = \mathbb{E}[G_m^{(j)}(x - \underline{X}_{m,-1} - a_m\varepsilon_0')|\mathcal{F}_0], \mathbb{E}[\varepsilon_0'|\mathcal{F}_0] = \mathbb{E}[\varepsilon_0'] = 0$ and both $\underline{X}_{m,0}$ and ε_0 are $\mathcal{F}_0$-measurable, we have

$$\begin{aligned}
&G_m^{(j)}(x - \underline{X}_{m,0}) - G_{m+1}^{(j)}(x - \underline{X}_{m,-1}) + f^{(j)}(x)a_m\varepsilon_0 \\
&\quad = \mathbb{E}\Big[G_m^{(j)}(x - \underline{X}_{m,0}) - G_m^{(j)}(x - \underline{X}_{m,-1} - a_m\varepsilon_0') + f^{(j)}(x)a_m(\varepsilon_0 - \varepsilon_0')|\mathcal{F}_0\Big].
\end{aligned} \tag{19}$$

By the Mean Value Theorem there is some ξ_m between $x - \underline{X}_{m,-1} - a_m\varepsilon_0'$ and $x - \underline{X}_{m,-1}$ such that $G_m^{(j)}(x - \underline{X}_{m,-1} - a_m\varepsilon_0') = G_m^{(j)}(x - \underline{X}_{m,-1}) - G_m^{(j+1)}(\xi_m)a_m\varepsilon_0'$. So, introducing the telescoping sum $G_m^{(j+1)}(x - \underline{X}_{m,-1})a_m\varepsilon_0 - G_m^{(j+1)}(x - \underline{X}_{m,-1})a_m\varepsilon_0$ on the right-hand side in (19) and noting that both $\underline{X}_{m,0}$ and $\underline{X}_{m,-1}$ are $\mathcal{F}_0$-measurable and that ε_0' is independent of $\mathcal{F}_0$, we obtain from (19)

$$
\begin{aligned}
&G_m^{(j)}(x-\underline{X}_{m,0}) - G_{m+1}^{(j)}(x-\underline{X}_{m,-1}) + f^{(j)}(x)a_m\varepsilon_0 \\
&\quad = G_m^{(j)}(x-\underline{X}_{m,0}) - G_m^{(j)}(x-\underline{X}_{m,-1}) + G_m^{(j+1)}(x-\underline{X}_{m,-1})a_m\varepsilon_0 \\
&\qquad + f^{(j)}(x)a_m\varepsilon_0 - G_m^{(j+1)}(x-\underline{X}_{m,-1})a_m\varepsilon_0 \\
&\qquad + \mathbb{E}\Big[G_m^{(j+1)}(\xi_m)a_m\varepsilon_0' - f^{(j)}(x)a_m\varepsilon_0'|\mathcal{F}_0\Big].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\int\Big\{G_m^{(j)}(x-\underline{X}_{m,0}) - G_{m+1}^{(j)}(x-\underline{X}_{m,-1}) + f^{(j)}(x)a_m\varepsilon_0\Big\}^2\phi_{2\lambda}(x)\,dx \\
&\quad \le 4\Big(\int\Big\{G_m^{(j)}(x-\underline{X}_{m,0}) - G_m^{(j)}(x-\underline{X}_{m,-1}) + G_m^{(j+1)}(x-\underline{X}_{m,-1})a_m\varepsilon_0\Big\}^2\phi_{2\lambda}(x)\,dx \\
&\qquad + \int\Big\{f^{(j)}(x)a_m\varepsilon_0 - G_m^{(j+1)}(x-\underline{X}_{m,-1})a_m\varepsilon_0\Big\}^2\phi_{2\lambda}(x)\,dx \\
&\qquad + \int\mathbb{E}\Big[G_m^{(j+1)}(\xi_m)a_m\varepsilon_0' - f^{(j)}(x)a_m\varepsilon_0'|\mathcal{F}_0\Big]^2\phi_{2\lambda}(x)\,dx\Big) \\
&\quad =: 4\,(\mathrm{I}_{m,j} + \mathrm{II}_{m,j} + \mathrm{III}_{m,j}).
\end{aligned}
$$

Noting $\lambda_{m,j}^2 \le 4(\mathbb{E}[\mathrm{I}_{m,j}] + \mathbb{E}[\mathrm{II}_{m,j}] + \mathbb{E}[\mathrm{III}_{m,j}])$, it remains to show that $\mathbb{E}[\mathrm{I}_{m,j}] = o(a_m^2)$, $\mathbb{E}[\mathrm{II}_{m,j}] = o(a_m^2)$ and $\mathbb{E}[\mathrm{III}_{m,j}] = o(a_m^2)$. We will proceed in three steps.

Step I. On the one hand, we have by Lemma 7 (29) in [34]

$$
\begin{aligned}
\mathrm{I}_{m,j} &= \int\Big\{G_m^{(j)}(x-\underline{X}_{m,-1}+a_m\varepsilon_0) \\
&\qquad - G_m^{(j)}(x-\underline{X}_{m,-1}) + G_m^{(j+1)}(x-\underline{X}_{m,-1})a_m\varepsilon_0\Big\}^2\phi_{2\lambda}(x)dx \\
&\le C_1\,(a_m\varepsilon_0)^4\,\phi_{2\lambda}(a_m\varepsilon_0)\phi_{2\lambda}(\underline{X}_{m,-1})\int G_m^{(j+1)}(x)^2\phi_{2\lambda}(x)\,dx \\
&= C_2\,a_m^4\,\Big(\varepsilon_0^4\,\phi_{2\lambda}(a_m\varepsilon_0)\phi_{2\lambda}(\underline{X}_{m,-1})\Big)
\end{aligned} \tag{20}
$$

for some constants $C_1, C_2 > 0$. On the other hand, using Lemma 7 (28) in [34] twice, we obtain

$$
\begin{aligned}
\mathrm{I}_{m,j} &\le 4\int\Big\{G_m^{(j)}(x-\underline{X}_{m,-1}+a_m\varepsilon_0) - G_m^{(j)}(x-\underline{X}_{m,-1})\Big\}^2\phi_{2\lambda}(x)\,dx \\
&\quad + 4\int\Big\{G_m^{(j+1)}(x-\underline{X}_{m,-1}) - G_m^{(j+1)}(x)\Big\}^2(a_m\varepsilon_0)^2\phi_{2\lambda}(x)\,dx \\
&\quad + 4\int G_m^{(j+1)}(x)^2(a_m\varepsilon_0)^2\phi_{2\lambda}(x)\,dx \\
&\le 4C_3(a_m\varepsilon_0)^2\,\phi_{2\lambda}(a_m\varepsilon_0)\,\phi_{2\lambda}(\underline{X}_{m,-1})\int G_m^{(j+1)}(x)^2\phi_{2\lambda}(x)\,dx \\
&\quad + 4C_3\,(a_m\varepsilon_0)^2\,(\underline{X}_{m,-1})^2\,\phi_{2\lambda}(\underline{X}_{m,-1})\,\phi_{2\lambda}(0)\int G_m^{(j+2)}(x)^2\phi_{2\lambda}(x)\,dx \\
&\quad + 4\,(a_m\varepsilon_0)^2\int G_m^{(j+1)}(x)^2\phi_{2\lambda}(x)\,dx \\
&\le C_4\,a_m^2\,\Big(\phi_{2+2\lambda}(\underline{X}_{m,-1})\big(\varepsilon_0^2\phi_{2\lambda}(a_m\varepsilon_0) + \varepsilon_0^2\big)\Big)
\end{aligned} \tag{21}
$$

for some constants $C_3, C_4 > 0$. From (20)–(21) we deduce

$$\begin{aligned}
&\mathbb{E}[\mathrm{I}_{m,j}] \\
&\quad\le C_5 a_m^2 \mathbb{E}\Big[\min\Big\{a_m^2\,\varepsilon_0^4\,\phi_{2\lambda}(a_m\varepsilon_0)\,\phi_{2\lambda}(\underline{X}_{m,-1}),\,\phi_{2+2\lambda}(\underline{X}_{m,-1})\big(\varepsilon_0^2\phi_{2\lambda}(a_m\varepsilon_0)+\varepsilon_0^2\big)\Big\}\Big] \\
&\quad\le C_5\,a_m^2\,\mathbb{E}[\phi_{2+2\lambda}(\underline{X}_{m,-1})]\,\mathbb{E}\Big[\min\Big\{a_m^2\varepsilon_0^4\phi_{2\lambda}(a_m\varepsilon_0),\,\varepsilon_0^2\phi_{2\lambda}(a_m\varepsilon_0)+\varepsilon_0^2\Big\}\Big]
\end{aligned}$$

for some constant $C_5 > 0$, where we used the independence of $\underline{X}_{m,-1}$ and ε_0. Now, the latter expectation in the last line converges to zero by the dominated convergence theorem (a majorant is given by $\varepsilon_0^2\phi_{2\lambda}(a_m\varepsilon_0) + \varepsilon_0^2 \le 2^{2\lambda}(\varepsilon_0^2 + \max_{s\in\mathbb{N}_0}|a_s|\varepsilon_0^{2+2\lambda}) + \varepsilon_0^2$ due to assumption (b), and $a_m^2\varepsilon_0^4\phi_{2\lambda}(a_m\varepsilon_0)$ converges $\mathbb{P}$-almost surely to zero as $m \to \infty$). To show that the first expectation in the last line is bounded above uniformly in $m \in \mathbb{N}$, we first note that

$$|\underline{X}_{m,-1}| = \left|\sum_{s=m+1}^{\infty} s^{-\beta}\ell(s)\varepsilon_{m-s}\right| \le L\sum_{s=m+1}^{\infty}|s^{-\beta}\varepsilon_{m-s}| \le L\sum_{s=1}^{\infty}|s^{-\beta}\varepsilon_{-s}|$$

with $L := \max_{s\in\mathbb{N}_0}|\ell(s)|$, and so

$$\mathbb{E}[\phi_{2+2\lambda}(\underline{X}_{m,-1})] \le 2^{1+2\lambda}\left\{1 + L^{2+2\lambda}\,\mathbb{E}\left[\left(\sum_{s=1}^{\infty}|s^{-\beta}\varepsilon_{-s}|\right)^{2+2\lambda}\right]\right\}.$$

Now the latter is bounded above uniformly in $m \in \mathbb{N}$, because by the Rosenthal inequality and assumptions (a)–(b)

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{s=1}^{\infty}|s^{-\beta}\varepsilon_{-s}|\right)^{2+2\lambda}\right] &\le \max\left\{\sum_{s=1}^{\infty}\mathbb{E}[|s^{-\beta}\,\varepsilon_{-s}|^{2+2\lambda}];\,\left(\sum_{s=1}^{\infty}\mathbb{E}[|s^{-\beta}\,\varepsilon_{-s}|^2]\right)^{(2+2\lambda)/2}\right\} \\
&\le \max\left\{\mathbb{E}[|\varepsilon_0|^{2+2\lambda}]\sum_{s=1}^{\infty}s^{-\beta(2+2\lambda)};\,\left(\mathbb{E}[\varepsilon_0^2]\sum_{s=1}^{\infty}s^{-2\beta}\right)^{1+\lambda}\right\} \\
&< \infty.
\end{aligned}$$

Thus, $\mathbb{E}[\mathrm{I}_{m,j}] = o(a_m^2)$.

Step II. Notice that $f^{(j)}(x) = \mathbb{E}[G_m^{(j+1)}(x - \underline{X}_{m,0})]$ due to Lemma 7 (27) in [34]. Setting $\sigma^2 := \mathbb{E}[\varepsilon_0^2]$, using the independence of ε_0 and $\underline{X}_{m,-1}$ as well as Jensen's inequality, we obtain

$$\begin{aligned}
\mathbb{E}[\mathrm{II}_{m,j}] \le\;& 2\,\mathbb{E}\left[\int\left\{f^{(j)}(x)a_m\varepsilon_0 - G_m^{(j+1)}(x)a_m\varepsilon_0\right\}^2\phi_{2\lambda}(x)\,dx\right] \\
&+ 2\mathbb{E}\left[\int\left\{G_m^{(j+1)}(x)a_m\varepsilon_0 - G_m^{(j+1)}(x-\underline{X}_{m,-1})a_m\varepsilon_0\right\}^2\phi_{2\lambda}(x)\,dx\right] \\
=\;& a_m^2 2\mathbb{E}[\varepsilon_0^2]\int\left\{\mathbb{E}[G_m^{(j+1)}(x-\underline{X}_{m,0})] - G_m^{(j+1)}(x)\right\}^2\phi_{2\lambda}(x)\,dx \\
&+ a_m^2 2\mathbb{E}[\varepsilon_0^2]\,\mathbb{E}\left[\int\left\{G_m^{(j+1)}(x) - G_m^{(j+1)}(x-\underline{X}_{m,-1})\right\}^2\phi_{2\lambda}(x)\,dx\right] \\
\le\;& a_m^2\sigma^2 2\Bigg(\mathbb{E}\left[\int\left\{G_m^{(j+1)}(x-\underline{X}_{m,0}) - G_m^{(j+1)}(x)\right\}^2\phi_{2\lambda}(x)\,dx\right]
\end{aligned}$$

$$+\mathbb{E}\left[\int\left\{G_m^{(j+1)}(x)-G_m^{(j+1)}(x-\underline{X}_{m,-1})\right\}^2\phi_{2\lambda}(x)\,dx\right]\Bigg)$$
$$=: a_m^2\sigma^2 2(R_{m,j}(1)+R_{m,j}(2)).$$

Using Lemma 7 (28) in [34], the Rosenthal inequality and the dominated convergence theorem, one can now easily show that $R_{m,j}(1)=o(1)$ and $R_{m,j}(2)=o(1)$. That is, $\mathbb{E}[\mathrm{II}_{m,j}]=o(a_m^2)$.

Step III. Using the conditional Hölder inequality as well as the independence of ε_0' and $\mathcal{F}_0$,

$$\begin{aligned}
&\mathbb{E}[\mathrm{III}_{m,j}]\\
&=a_m^2\mathbb{E}\left[\int\mathbb{E}\Big[\big(G_m^{(j+1)}(\xi_m)-f^{(j)}(x)\big)\varepsilon_0'|\mathcal{F}_0\Big]^2\phi_{2\lambda}(x)\,dx\right]\\
&\le a_m^2\mathbb{E}\left[\int\Big(\mathbb{E}\big[(\varepsilon_0')^2|\mathcal{F}_0\big]^{1/2}\mathbb{E}\big[\big(G_m^{(j+1)}(\xi_m)-f^{(j)}(x)\big)^2|\mathcal{F}_0\big]^{1/2}\Big)^2\phi_{2\lambda}(x)\,dx\right]\\
&=a_m^2\sigma^2\mathbb{E}\left[\int\mathbb{E}\big[\big(G_m^{(j+1)}(\xi_m)-f^{(j)}(x)\big)^2|\mathcal{F}_0\big]\phi_{2\lambda}(x)\,dx\right]\\
&\le a_m^2\sigma^2 2\Bigg(\mathbb{E}\left[\int\big(G_m^{(j+1)}(\xi_m)-G_m^{(j+1)}(x)\big)^2\phi_{2\lambda}(x)\,dx\right]\\
&\quad+\mathbb{E}\left[\int\big(G_m^{(j+1)}(x)-f^{(j)}(x)\big)^2\phi_{2\lambda}(x)\,dx\right]\Bigg)\\
&=: a_m^2\sigma^2 2(R_{m,j}(3)+R_{m,j}(4)).
\end{aligned}$$

Proceeding as for the first summand in Step II, we obtain $R_{m,j}(4)=o(1)$. Moreover, we have $\xi_m=x-\underline{X}_{m,-1}+r_m$ with $|r_m|\le|a_m\varepsilon_0'|$, and ε_0' and $\underline{X}_{m,-1}$ are independent. Thus, using Lemma 7 (28) in [34],

$$\begin{aligned}
&R_{m,j}(3)\\
&=\mathbb{E}\left[\int\left\{G_m^{(j+1)}(x-\underline{X}_{m,-1}+r_m)-G_m^{(j+1)}(x)\right\}^2\phi_{2\lambda}(x)\,dx\right]\\
&\le\mathbb{E}\left[C_4\,(\underline{X}_{m,-1}-r_m)^2\,\phi_{2\lambda}(\underline{X}_{m,-1}-r_m)\int G_m^{(j+1)}(x)^2\phi_{2\lambda}(x)\,dx\right]\\
&\le C_4\,2^{2\lambda+1}\,\mathbb{E}\left[\big(\underline{X}_{m,-1}^2+r_m^2\big)\big(\phi_{2\lambda}(\underline{X}_{m,-1})+\phi_{2\lambda}(r_m)\big)\int G_m^{(j+1)}(x)^2\phi_{2\lambda}(x)\,dx\right]\\
&\le C_6\,\mathbb{E}\left[\big(\underline{X}_{m,-1}^2+(a_m\varepsilon_0)^2\big)\big(\phi_{2\lambda}(\underline{X}_{m,-1})+\phi_{2\lambda}(a_m\varepsilon_0)\big)\int G_m^{(j+1)}(x)^2\phi_{2\lambda}(x)\,dx\right]
\end{aligned}$$

for some constants $C_4, C_6>0$. Using again the Rosenthal inequality and the dominated convergence theorem, we obtain $R_{m,j}(3)=o(1)$. That is, we also have $\mathbb{E}[\mathrm{III}_{m,j}]=o(a_m^2)$. □

Lemma A.5. *For* $j\in\{0,1\}$, *we have* $\mathrm{I}_n(j)=o(\sigma_n^2)$.

Proof. By Lemma A.4, we have $\lambda_{i,j}=o(a_i)$, and by Karamata's theorem (cf. [14, p. 281]) we have $\sum_{i=1}^n a_i\sim n^{1-\beta}\ell(n)/(1-\beta)\sim na_n/(1-\beta)$. Also note that $\sigma_n\sim n^{3/2-\beta}\ell(n)/c_{1,\beta}$.

So, since $\beta \in (\frac{1}{2}, 1)$, elementary calculations yield

$$\sum_{k=-\infty}^{n} \left(\sum_{i=1}^{n} \lambda_{i-k,j} \right)^2 \leq 2n \left(\sum_{i=1}^{2n} \lambda_{i,j} \right)^2 + \sum_{k=-\infty}^{-n} \{o(na_{-k})\}^2 = o(\sigma_n^2).$$

This proves the claim. □

Appendix B. Integration theoretical auxiliaries

Let $\mathbb{BV}^2_{\text{loc,rc}}$ be defined as in Section 3.3. Recall that for $u \in \mathbb{BV}^2_{\text{loc,rc}}$, we set $|du| := du^+ + du^-$ with du^+ and du^- the unique positive Radon measures induced by the Jordan decomposition of u into the difference of two bimonotonically increasing functions; cf. [16, Proposition 1.17].

Lemma B.1. *Let $u, v \in \mathbb{BV}^2_{\text{loc,rc}}$ and assume that:*

(i) *The integrals $\iint |v(x_1, x_2)|\, |du|(x_1, x_2)$ and $\iint |u(x_1, x_2)|\, |dv|(x_1, x_2)$ are finite.*
(ii) *The functions $v_{x_1}(\,\cdot\,) := v(x_1, \,\cdot\,)$ and $v_{x_2}(\,\cdot\,) := v(\,\cdot\,, x_2)$ are of locally bounded variation for every fixed $x_1, x_2 \in \mathbb{R}$.*
(iii) *The following limits exist*

$$k_1 := \lim_{a_1,a_2\to-\infty,b_1,b_2\to\infty} \int_{a_1}^{b_1} u(x_1, b_2)\, dv_{b_2}(x_1) - \int_{a_1}^{b_1} u(x_1, a_2)\, dv_{a_2}(x_1),$$

$$k_2 := \lim_{a_1,a_2\to-\infty,\, b_1,b_2\to\infty} \int_{a_2}^{b_2} u(b_1, x_2)\, dv_{b_1}(x_2) - \int_{a_2}^{b_2} u(a_1, x_2)\, dv_{a_1}(x_2),$$

$$k_3 := \lim_{a_1,a_2\to-\infty,\, b_1,b_2\to\infty} \Big(u(b_1, b_2)v(b_1, b_2) - u(a_1, b_2)v(a_1, b_2) - u(b_1, a_2)v(b_1, a_2) + u(a_1, a_2)v(a_1, a_2) \Big).$$

(iv) *The functions u and v have no joint discontinuity.*

Then

$$\iint v(x_1, x_2) du(x_1, x_2) = \iint u(x_1, x_2) dv(x_1, x_2) - k_1 - k_2 + k_3. \tag{22}$$

Remark B.2. It should be mentioned that $v_{x_1} \in \mathbb{BV}_{\text{loc,rc}}$ (which is imposed on v through condition (ii)) is not implied by $v \in \mathbb{BV}^2_{\text{loc,rc}}$. To see this, let $h : \mathbb{R} \to \mathbb{R}$ be any (right-continuous) function being of unbounded variation on every finite interval $I \subset \mathbb{R}$. Then define $v : \mathbb{R}^2 \to \mathbb{R}$ by $v(x_1, x_2) := h(x_2)$, and observe that for every half-open rectangle $(a_1, b_1] \times (a_2, b_2]$,

$$\sup_{\Pi} \sum_{(x_1,y_1]\times(x_2,y_2]\in\Pi} \big| v(y_1, y_2) - v(x_2, y_1) - v(x_1, y_2) + v(x_1, x_2) \big| = 0,$$

where the supremum is taken over all partitions Π of $(a_1, b_1] \times (a_2, b_2]$ consisting of finitely many half-open rectangles. Therefore $v \in \mathbb{BV}^2_{\text{loc,rc}}$. On the other hand, $v_{x_1} \notin \mathbb{BV}_{\text{loc,rc}}$. □

Proof of Lemma B.1. Under condition (ii) and (iv), and since $u, v \in \mathbb{BV}^2_{\mathrm{loc,rc}}$, we have (see [10]) for every half-open rectangle $(a_1, b_1] \times (a_2, b_2]$

$$\begin{aligned}\int_{a_1}^{b_1}\int_{a_2}^{b_2} v(x_1,x_2)\,du(x_1,x_2) &= \int_{a_1}^{b_1}\int_{a_2}^{b_2} u(x_1,x_2)\,dv(x_1,x_2)\\ &\quad - \int_{a_1}^{b_1} u(x_1,b_2)\,dv_{b_2}(x_1) + \int_{a_1}^{b_1} u(x_1,a_2)\,dv_{a_2}(x_1)\\ &\quad - \int_{a_2}^{b_2} u(b_1,x_2)\,dv_{b_1}(x_2) + \int_{a_2}^{b_2} u(a_1,x_2)\,dv_{a_1}(x_2)\\ &\quad + u(b_1,b_2)v(b_1,b_2) - u(a_1,b_2)v(a_1,b_2)\\ &\quad - u(b_1,a_2)v(b_1,a_2) + u(a_1,a_2)v(a_1,a_2). \qquad (23)\end{aligned}$$

By assumption the integral $\iint v(x_1, x_2)\, du(x_1, x_2)$ exists, and we have that

$$\begin{aligned}\int_{a_1}^{b_1}\int_{a_2}^{b_2} v(x_1,x_2)\,du(x_1,x_2) &= \iint v(x_1,x_2)\mathbb{1}_{\{a_1<x_1\le b_1, a_2<x_2\le b_2\}}\,du^+(x_1,x_2)\\ &\quad - \iint v(x_1,x_2)\mathbb{1}_{\{a_1<x_1\le b_1, a_2<x_2\le b_2\}}\,du^-(x_1,x_2) \qquad (24)\end{aligned}$$

converges to $\iint v(x_1, x_2)\, du(x_1, x_2)$ as $a_1, a_2 \to -\infty, b_1, b_2 \to \infty$, since by Lebesgue's dominated convergence theorem the two integrals on the right-hand side of (24) converge to $\iint v(x_1, x_2)\, du^+(x_1, x_2)$ and $\iint v(x_1, x_2)\, du^-(x_1, x_2)$, respectively. The integral $\int_{a_1}^{b_1}\int_{a_2}^{b_2} u(x_1, x_2)\, dv(x_1, x_2)$ can be treated in the same way. The result follows now by using condition (iii) for the remaining terms on the right-hand side of (23). □

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, Mathematical Finance 9 (1999) 203–228.

[2] F. Avram, M.S. Taqqu, Noncentral limit theorems and Appell polynomials, Annals of Probability 15 (1987) 767–775.

[3] R.T. Baillie, Long memory processes and fractional integration in econometics, Journal of Econometrics 73 (1996) 5–59.

[4] J. Beran, Statistical methods for data with long range dependence, Statistical Science 7 (1992) 404–427.

[5] J. Beran, Statistics for Long-Memory Processes, Chapman & Hall, New York, 1994.

[6] E. Beutner, H. Zähle, A modified functional delta method and its application to the estimation of risk functionals, Journal of Multivariate Analysis 101 (2010) 2452–2463.

[7] E. Beutner, H. Zähle, Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes, Bernoulli (2011) (in press).

[8] X. Chen, Y. Fan, Estimation of copula-based semiparametric time series models, Journal of Econometrics 130 (2006) 307–335.

[9] H. Dehling, M.S. Taqqu, The empirical process of some long-range dependent sequences with an application to U-statistics, Annals of Statistics 17 (1989) 1767–1783.

[10] H. Dehling, M.S. Taqqu, Bivariate symmetric statistics of long-range dependent observations, Journal of Statistical Planning and Inference 28 (1991) 153–165.

[11] P. Doukhan, G. Lang, D. Surgailis, M.-C. Viano, Functional limit theorem for the empirical process of a class of Bernoulli shifts with long memory, Journal of Theoretical Probability 18 (2005) 161–186.

[12] P. Doukhan, P. Massart, E. Rio, The invariance principle for the empirical measure of a weakly dependent process, Annales de l'Institut Henri Poincaré B 31 (1995) 393–427.

[13] Y. Fan, Q. Li, Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing, Journal of Nonparametric Statistics 10 (1999) 245–271.

[14] W. Feller, An Introduction to Probability Theory and its Applications, vol. 2, Wiley, New York, 1971.
[15] L.T. Fernholz, Von Mises Calculus for Statistical Functionals, in: Lecture Notes in Statistics, vol. 19, Springer-Verlag, New York, 1983.
[16] S.R. Ghorpade, B.V. Limaye, A Course in Multivariate Calculus and Analysis, Springer-Verlag, New York, 2009.
[17] R.D. Gill, Non- and semi-parametric maximum likelihood estimators and the von Mises method (part I), Scandinavian Journal of Statistics 16 (1989) 97–128.
[18] L. Giraitis, D. Surgailis, Central limit theorem for the empirical process of a linear sequence with long memory, Journal of Statistical Planning and Inference 80 (1999) 81–93.
[19] H.-C. Ho, T. Hsing, On the asymptotic expansion of the empirical process of long-memory moving averages, Annals of Statistics 24 (1996) 992–1024.
[20] T. Hsing, W.B. Wu, On weighted U-statistics for stationary processes, Annals of Probability 32 (2004) 1600–1631.
[21] O. Kallenberg, Foundations of Modern Probability, Springer-Verlag, New York, 1997.
[22] H.L. Koul, D. Surgailis, Asymptotics of empirical processes of long memory moving averages with infnite variance, Stochastic Processes and their Applications 91 (2001) 309–336.
[23] R. Kulik, Empirical process of long-range dependent sequences when parameters are estimated, Journal of Statistical Planning and Inference 139 (2009) 287–294.
[24] F. Lavancier, A. Philippe, D. Surgailis, A two-sample test for comparison of long memory parameters, Journal of Multivariate Analysis 101 (2010) 2118–2136.
[25] A.J. Lee, U-Statistics, Marcel Dekker Inc., New York, 1990.
[26] J.A. Reeds, On the definition of von Mises functionals, Research Report S-44, Department of Statistics, Harvard University, 1976.
[27] J. Robinson, Time series with strong dependence, in: Sims, E. (Ed.), Advances in Econometric, Econometric Society Monographs, vol. 23, Cambridge, 1994, pp. 47–94.
[28] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[29] Q.-M. Shao, H. Yu, Weak convergence for weighted empirical processes of dependent sequences, Annals of Probability 24 (1996) 2098–2127.
[30] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.
[31] D. Surgailis, Zones of attraction of self-similar multiple integrals, Lithuanian Mathematical Journal 22 (1982) 327–340.
[32] M.S. Taqqu, Weak convergence to fractional Brownian motion and to the Rosenblatt process, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31 (1975) 287–302.
[33] S. Wang, J. Dhaene, Comontonicity, correlation order and premium principles, Insurance: Mathematics and Economics 22 (1998) 235–242.
[34] W.B. Wu, Empirical processes of long-memory sequences, Bernoulli 9 (2003) 809–831.
[35] W.B. Wu, Empirical processes of stationary sequences, Statistica Sinica 18 (2008) 313–333.