Journal of Multivariate Analysis 101 (2010) 2452–2463

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A modified functional delta method and its application to the estimation of risk functionals

Eric Beutner [a,*], Henryk Zähle [b]

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

[b] Saarland University, Department of Mathematics, Postfach 151150, 66041 Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 22 February 2010
Available online 26 June 2010

AMS subject classifications:
60F05
62G05
62G20
62G30
62M10
62M99

Keywords:
Central limit theorem
Functional delta method
Hadamard derivative
Weighted empirical process
Mixing
Censoring
Distortion risk measure

ABSTRACT

The classical functional delta method (FDM) provides a convenient tool for deriving the asymptotic distribution of statistical functionals from the weak convergence of the respective empirical processes. However, for many interesting functionals depending on the tails of the underlying distribution this FDM cannot be applied since the method typically relies on Hadamard differentiability w.r.t. the uniform sup-norm. In this article, we present a version of the FDM which is suitable also for nonuniform sup-norms, with the outcome that the range of application of the FDM enlarges essentially. On one hand, our FDM, which we shall call the modified FDM, works for functionals that are "differentiable" in a weaker sense than Hadamard differentiability. On the other hand, it requires weak convergence of the empirical process w.r.t. a nonuniform sup-norm. The latter is not problematic since there exist strong respective results on weighted empirical processes obtained by Shorack and Wellner (1986) [25], Shao and Yu (1996) [23], Wu (2008) [32], and others. We illustrate the gain of the modified FDM by deriving the asymptotic distribution of plug-in estimates of popular risk measures that cannot be treated with the classical FDM.

1. Introduction

In this article, we consider estimators for the functional ρ_g defined by

$$\rho_g(F) := -\int_{-\infty}^{\infty} x\,\mathrm{d}g(F(x)), \tag{1}$$

where the domain of ρ_g is the set of all distribution functions (df) F on the real line for which the integral in (1) exists, and g is some fixed df on the interval $[0, 1]$. Our considerations are motivated by the fact that, in recent years, functionals of the form (1) have commonly received interest in financial and actuarial mathematics as risk measures. More precisely, in this field the functional ρ_g is called distortion risk measure with distortion function g. At the end of the introduction we will explain the meaning of ρ_g in this context and we will provide some examples.

At first, however, we would like to point out that ρ_g is strongly related to functionals used to represent L-statistics. It is indeed well known (see, e.g., [21, p. 265]) that a wide class of L-statistics can be represented as $-\rho_g(F_n)$, where F_n denotes the empirical df of n i.i.d. random variables with df F. The asymptotic distribution of $-\rho_g(F_n)$ was already derived in [26]; see also [21, Section 8.2.4]. Thus, if $\rho_g(F)$ is estimated by $\rho_g(F_n)$ based on i.i.d. data, the results in [26,21] immediately provide the asymptotic distribution of the plug-in estimate $\rho_g(F_n)$. However, there are several situations where the estimate F_n of F differs from the empirical df or where the data are dependent. Then the results in [26,21] do not apply any longer. A

* Corresponding author.
E-mail addresses: e.beutner@maastrichtuniversity.nl (E. Beutner), zaehle@math.uni-sb.de (H. Zähle).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.06.015

unifying approach for overcoming this problem is the functional delta method (FDM) in the sense of [8,10,15]; see also [28, Section 3.9] or [29, Section 20.2]. Once one has shown the Hadamard differentiability of ρ_g w.r.t. the sup-norm $\|\cdot\|_\infty$, the asymptotic distribution can be immediately derived from the weak limit w.r.t. to $\|\cdot\|_\infty$ of the corresponding empirical process $\sqrt{n}(F_n - F)$. For an illustration and examples see, e.g., [8,10,20,28].

However, there is an unsatisfactory fact. The Hadamard differentiability of ρ_g at F w.r.t. the sup-norm requires quite a restrictive assumption on g. If g (regarded as a measure) has compact support strictly within the open interval (0, 1) then the functional ρ_g is easily seen to be Hadamard differentiable (cf. [29, Lemma 22.10]), and there are several related results assuming that the compact support of g is strictly within (0, 1); see, e.g., [11,16,22]. On the other hand, if the compact support supp(g) of g contains at least one of the boundary points 0 or 1 (which is the case for several popular distortion risk measures), then Hadamard differentiability may fail. To see this, we first note that $\rho_g(F)$ has the alternative representation

$$\rho_g(F) = \int_{(-\infty,0)} g(F(t))\mathrm{d}t - \int_{[0,\infty)} \big(1 - g(F(t))\big)\mathrm{d}t \tag{2}$$

provided the integral in (1) exists; recall that g is a df on [0, 1], so $g(F(\cdot))$ is a df on the real line (If we use (2) as the definition of ρ_g, then g may be any nondecreasing function $g : [0, 1] \to [0, 1]$ with $g(0) = 0$ and $g(1) = 1$). In view of (2), the candidate for the Hadamard derivative is easily identified to be $\dot{\rho}_{g,F}(V) := \int_{\mathbb{R}} g'(F(t))V(t)\mathrm{d}t$ for bounded and continuous directions V. To verify that $\dot{\rho}_{g,F}$ does indeed provide the Hadamard derivative of ρ_g at F w.r.t. the sup-norm $\|\cdot\|_\infty$, one has to show that $\|V - V_n\|_\infty \to 0$ implies $|\dot{\rho}_{g,F}(V) - h_n^{-1}(\rho_g(F + h_nV_n) - \rho_g(F))| \to 0$, i.e.

$$\lim_{n\to\infty} \left| \int_{\mathbb{R}} \Big(g'(F(t))V(t) - \frac{g(F(t) + h_nV_n(t)) - g(F(t))}{h_n} \Big)\mathrm{d}t \right| = 0, \tag{3}$$

for any sequence $(h_n) \subset \mathbb{R} \setminus \{0\}$ converging to 0. However, if supp(g) = [0, 1] (or at least $0 \in$ supp(g) or $1 \in$ supp(g)), then in general (3) cannot be deduced from $\|V - V_n\|_\infty \to 0$. To give a simple example, we set $g(x) := x$ (so $\rho_g(F)$ is nothing but the negative mean of F). In this case we have $g' \equiv 1$ and therefore (3) reads as

$$\lim_{n\to\infty} \left| \int_{\mathbb{R}} \big(V(t) - V_n(t)\big)\mathrm{d}t \right| = 0. \tag{4}$$

Now, it is obvious that $\|V - V_n\|_\infty \to 0$ does not imply (4). It was already emphasized in, e.g., [29, Section 22.1] that methods based on Hadamard differentiability w.r.t. the sup-norm do not cover the simplest L-statistic: the sample mean. Similar arguments apply to more interesting distortion functions g with $0 \in$ supp(g); see, e.g., (6) below.

The convergence in (3) would hold if we had assumed that V_n converges to V w.r.t. the nonuniform sup-norm $\|\cdot\|_\lambda := \|(\cdot)\phi_\lambda\|_\infty$ based on the weight function $\phi_\lambda(t) := (1 + |t|)^\lambda$, $t \in \mathbb{R}$, with $\lambda > 1$. So one might tend to replace the uniform sup-norm by the nonuniform sup-norm (λ-norm) in the classical FDM (in the sense of [10,28,29]) applied to the functional ρ_g. However, then the argument F of ρ_g would not have a finite norm: As $\lim_{t\to\infty} F(t) = 1$, we clearly have $\|F\|_\lambda = \infty$ for all $\lambda > 0$. Therefore the classical FDM does not work. Nevertheless the approach of the nonuniform sup-norm is not that bad. Studying the proof of the FDM in detail, one easily observes that the imposed norm is essential only for the tangential space. And in contrast to F itself, the empirical process $\sqrt{n}(F - F_n)$ may perfectly have a finite λ-norm under suitable assumptions on the tails of F, so that the space of all continuous functions with finite λ-norm should be suitable for being the tangential space. In Section 2, we will present a corresponding notion of differentiability which is weaker than Hadamard differentiability but still strong enough for obtaining a (modified) FDM. The modified FDM will be given in Section 4. Before, in Section 3, we will illustrate our method by means of three examples. We will consider the case where F_n is a smooth empirical df, and the cases of censored and dependent data. The examples show in particular how powerful the method presented is.

Before we turn to our theoretical results, we will briefly discuss the motivation of ρ_g in financial economics. For a random variable X, regarded as a financial position, the value $\rho_g(F_X)$ should be seen as the minimal amount of capital (solvency capital requirement; SCR) which has to be added to the position X in order to make it acceptable; see, e.g., [2,6,9]. A naive choice of the SCR would be the negative mean $\rho_{\mathbb{I}}(F_X) := -\int_{-\infty}^{\infty} x\mathrm{d}F_X(x)$, so that "on average" the position X will not erode more capital than the SCR. However the mean does not take into account the riskiness of a position. One can overcome this problem by manipulating the df F_X using a risk-averse (i.e. concave) distortion function g, i.e. by choosing the SCR as the negative mean of the conservatively distorted df $g(F_X(\cdot))$ (cf. [12,30,31]). This leads to the functional ρ_g in (1) which has the alternative representation (2).

Surprisingly, most of the popular risk measures in practice can be represented as in (2). Let us give two examples. First, the Value-at-Risk at level $\alpha \in (0, 1)$ of a position X with df F is defined by

$$\mathrm{VaR}_\alpha(F) := -F^{\rightarrow}(\alpha), \tag{5}$$

where $F^{\rightarrow}(\alpha) := \inf\{t \in \mathbb{R} : F(t) > \alpha\}$ is the right-continuous inverse of F at α. It is a distortion risk measure w.r.t. $g(x) = \mathbb{1}_{(\alpha,1]}(x)$. If the position X is secured by the capital $\mathrm{VaR}_\alpha(F_X)$ (SCR), then the position will not erode more capital than the SCR with probability $1-\alpha$ (with α preferably close to 0). This comprehensible interpretation is impaired by the fact that the Value-at-Risk does not take into account the magnitude of loss "on the rare occasions". Second, the Average Value-at-Risk at level $\alpha \in (0, 1)$ of a random variable X with df F is defined by

$$\mathrm{AVaR}_\alpha(F) := \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_a(F)\mathrm{d}a, \tag{6}$$

provided the integral exists. It is a distortion risk measure w.r.t. $g(x) = (x/\alpha) \wedge 1$. If $F_X(F_X^{\rightarrow}(\alpha)) = \alpha$, then we have $\mathrm{AVaR}_\alpha(F_X) = \mathbb{E}[-X| - X \geq \mathrm{VaR}_\alpha(F_X)]$ indicating that the Average Value-at-Risk overcomes the drawback of the Value-at-Risk. This is why the Average Value-at-Risk is that popular in practice. We emphasize that the distortion function g corresponding to AVaR_α has the interval $[0, \alpha]$ as compact support (in particular $0 \in \mathrm{supp}(g)$) and has a similar behavior at the left boundary of $[0, 1]$ as the identity $g(x) = x$. Thus, according to our preceding discussion, in general the asymptotic distribution of empirical estimates $\rho_g(F_n)$ of $\rho_g(F)$ cannot be derived by using the classical FDM. However, it can be derived by using the modified FDM to be presented in this article.

2. Quasi-Hadamard differentiability, and the asymptotic distribution of nonparametric estimates of $\rho_g(F)$

Recall that tangential Hadamard differentiability is the right type of differentiability in connection with the delta method; see, for example, [10,28,29]. As indicated in the introduction, we start our considerations with introducing a slightly modified notion of tangential Hadamard differentiability which we call *quasi tangential Hadamard differentiability* (or just quasi-Hadamard differentiability). It is shown in Section 4 that this type of "differentiability" is sufficient for proving a FDM.

Definition 2.1 (*Quasi-Hadamard Differentiability*). Let $\mathbf{V}$ be a vector space, $(\mathbf{V}', \|\cdot\|_{\mathbf{V}'})$ be a normed vector space, and f be a mapping $f : \mathbf{V}_f \to \mathbf{V}'$ defined on a subset $\mathbf{V}_f$ of $\mathbf{V}$. Let $\mathbf{V}_0$ be a subspace of $\mathbf{V}$ equipped with a norm $\|\cdot\|_{\mathbf{V}_0}$, and $\mathbb{C}_0$ be a subset of $\mathbf{V}_0$. Then f is said to be quasi-Hadamard differentiable at $\theta \in \mathbf{V}_f$ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$ if there is some continuous map $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f : \mathbb{C}_0 \to \mathbf{V}'$ such that

$$\lim_{n\to\infty}\left\| D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f\,(v) - \frac{f(\theta + h_n v_n) - f(\theta)}{h_n}\right\|_{\mathbf{V}'} = 0 \tag{7}$$

holds for each triplet $(v, (v_n), (h_n))$ with $v \in \mathbb{C}_0$, $(v_n) \subset \mathbf{V}_0$ satisfying $\|v_n - v\|_{\mathbf{V}_0} \to 0$ as well as $\theta + h_n v_n \in \mathbf{V}_f$ for every $n \in \mathbb{N}$, and $(h_n) \subset \mathbb{R}_0 := \mathbb{R} \setminus \{0\}$ satisfying $h_n \to 0$. In this case the map $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f$ is called the quasi-Hadamard derivative of f at θ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$.

Of course, the norm $\|\cdot\|_{\mathbf{V}_0}$ induces a topological structure on $\mathbb{C}_0$: If $\mathbb{C}_0$ is not a subspace but only a subset of $\mathbf{V}_0$, then we may regard $\|\cdot\|_{\mathbf{V}_0}$ at least as a metric. Actually the definition of quasi-Hadamard differentiability still makes sense if the norms $\|\cdot\|_{\mathbf{V}_0}$ and $\|\cdot\|_{\mathbf{V}'}$ are replaced by metrics; see Remark 4.2 below. The notion of quasi-Hadamard differentiability differs from the "classical" tangential Hadamard differentiability (in the sense of [1,10,28]) primarily in that $\mathbf{V}$ is *not* required to be a *normed* vector space. Also notice that in Definition 2.1 it is not required that $\theta \in \mathbf{V}_0$. It is further worth mentioning that if $\|\cdot\|_{\mathbf{V}_0}$ provides a norm on all of $\mathbf{V}$ and if $\mathbb{C}_0 = \mathbf{V}_0$, then the two notions of "differentiability" coincide. Finally we emphasize that the notion of quasi-Hadamard differentiability differs from the concept of quasi-differentials introduced in [21, p. 221]. The latter concept was introduced to simplify the method of checking "Fréchet differentiability" (in the sense of [21, p. 217]) of statistical functionals; the benefit of this concept is illustrated in [21, p. 255] in the context of M-estimates.

We are now going to show that the functional ρ_g given by (1) (resp. (2)) is quasi-Hadamard differentiable tangentially to some suitable tangential space. We denote by $\mathbb{D}$ the vector space of all càdlàg functions on $\overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ denotes the usual two-point compactification of $\mathbb{R}$. We write $\mathbb{F}_g$ for the set of all df F on $\overline{\mathbb{R}}$ for which the integral in (1) exists. Recall that $\phi_\lambda : \overline{\mathbb{R}} \to [1, \infty]$ was defined by $\phi_\lambda(t) = (1 + |t|)^\lambda$, $t \in \overline{\mathbb{R}}$, for some $\lambda > 1$. We write $\mathbb{D}_\lambda$ for the space of all càdlàg functions $V : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ satisfying $\|V\|_\lambda := \|V\phi_\lambda\|_\infty < \infty$. Notice that $\|\cdot\|_\lambda$ provides a norm on $\mathbb{D}_\lambda$, that $\mathbb{D}_\lambda$ is a subspace of $\mathbb{D}$, and that $\|\cdot\|_\lambda$ does *not* provide a norm on all of $\mathbb{D}$. We write $\mathbb{C}_\lambda$ for the subspace of all continuous functions of $\mathbb{D}_\lambda$, and note that $\mathbb{D}$, $\mathbb{F}_g$, $\mathbb{R}$, ρ_g, $\mathbb{D}_\lambda$, $\mathbb{C}_\lambda$ and $\|\cdot\|_\lambda$ will play the roles of $\mathbf{V}$, $\mathbf{V}_f$, $\mathbf{V}'$, f, $\mathbf{V}_0$, $\mathbb{C}_0$ and $\|\cdot\|_{\mathbf{V}_0}$ (respectively) in the setting of Definition 2.1.

Now, g as a nondecreasing function is differentiable dt-almost everywhere, the set $\mathcal{DP}_g$ of all differentiability points is a Borel set, and the derivative $g'(x)$ is nonnegative for every $x \in \mathcal{DP}_g$. We set $g' :\equiv 0$ outside $\mathcal{DP}_g$, so that the derivative g' is a Borel measurable function. Thus we may set for every $F \in \mathbb{F}_g \subset \mathbb{D}$

$$\dot{\rho}_{g,F}(V) := \int_{\mathbb{R}} g'(F(t))\,V(t)\,\mathrm{d}t, \quad V \in \mathbb{C}_\lambda, \tag{8}$$

provided the integral exists (for instance if g' is bounded). The following result shows that if g and F are sufficiently regular, then the map $\dot{\rho}_{g,F}$ defined in (8) is the quasi-Hadamard derivative of ρ_g at F tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$.

Theorem 2.2 (*Quasi-Hadamard Differentiability of ρ_g*). *Suppose that*

(a) *g is continuous and piecewise differentiable, and g' is bounded above by some constant $M > 0$,*
(b) *$F \in \mathbb{F}_g$ takes the value $d \in (0, 1)$ at most once if g is not differentiable at d.*

Then the mapping $\rho_g : \mathbb{F}_g \to \mathbb{R}$ *is quasi-Hadamard differentiable at* F *tangentially to* $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ *with quasi-Hadamard derivative*

$$D^{\mathrm{Had}}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}\rho_g(V) = \dot\rho_{g,F}(V), \quad V \in \mathbb{C}_\lambda.$$

Example 2.3. Condition (a) is fulfilled for, e.g., $g(x) = (x/\alpha) \wedge 1$ which corresponds to the Average Value-at-Risk at level $\alpha \in (0, 1)$ defined in (6). $\diamond$

Proof of Theorem 2.2. The operator $\dot\rho_{g,F}$ is clearly linear. It is also $\|\cdot\|_\lambda$-bounded (and therefore $\|\cdot\|_\lambda$-continuous) since g' is ($\mathrm{d}t$-almost everywhere) bounded above by some constant $M > 0$:

$$|\dot\rho_{g,F}(V)| \le \int_{\mathbb{R}} M\|V\|_\lambda\,\phi_\lambda(t)^{-1}\,\mathrm{d}t \le \Big(M\int_{\mathbb{R}}\phi_\lambda(t)^{-1}\mathrm{d}t\Big)\|V\|_\lambda, \quad V \in \mathbb{C}_\lambda.$$

Thus, according to Definition 2.1, it only remains to show that

$$\left|\frac{\rho_g(F+h_nV_n)-\rho_g(F)}{h_n} - \Big(\int_{\mathbb{R}} g'(F(t))\,V(t)\,\mathrm{d}t\Big)\right| \to 0, \quad n \to \infty \tag{9}$$

holds true for each triplet $(V, (V_n), (h_n))$, with $V \in \mathbb{C}_\lambda$, $(V_n) \subset \mathbb{D}_\lambda$ satisfying $\|V_n - V\|_\lambda \to 0$ as well as $F + h_nV_n \in \mathbb{F}_g$, $n \in \mathbb{N}$, and $(h_n) \subset \mathbb{R}\setminus\{0\}$ satisfying $h_n \to 0$.

Using the representation (2) one plainly verifies that the expression on the left-hand side of (9) is bounded above by

$$\int_{\mathbb{R}}\left|\frac{g(F(t)+h_nV_n(t))-g(F(t))}{h_n} - g'(F(t))\,V(t)\right|\,\mathrm{d}t. \tag{10}$$

We denote the integrand of the integral in (10) by $I_n(t)$. The distortion function g is differentiable $\mathrm{d}t$-almost everywhere and F takes the value d at most once if g is not differentiable at d. Thus $I_n(t)$ converges to 0 as $n \to \infty$ for $\mathrm{d}t$-almost all t. By the dominated convergence theorem it thus suffices to find some Lebesgue integrable majorant of (I_n).

With the help of the triangle inequality and the mean value theorem (along with the continuity and piecewise differentiability of g) we obtain for $\mathrm{d}t$-almost every $t \in \mathbb{R}$

$$\begin{aligned} I_n(t) &\le \left|\frac{g(F(t)+h_nV_n(t))-g(F(t))}{h_n}\right| + g'(F(t))\,|V(t)| \\ &= \left|\frac{g'(\xi_n(t))h_nV_n(t)}{h_n}\right| + g'(F(t))\,|V(t)| \\ &\le M\,|V_n(t)| + M\,|V(t)| \end{aligned}$$

for some suitable function ξ_n in between F and $F+h_nV_n$. Since V_n converges to V in $(\mathbb{D}_\lambda, \|\cdot\|_\lambda)$, we have $M_1 := \sup_{n\in\mathbb{N}}\|V_n\|_\lambda < \infty$. Therefore we obtain

$$\begin{aligned} I_n(t) &= I_n(t)\phi_\lambda(t)\phi_\lambda(t)^{-1} \\ &\le \Big(M\sup_{n\in\mathbb{N}}|V_n(t)\phi_\lambda(t)| + M\,|V(t)\phi_\lambda(t)|\Big)\phi_\lambda(t)^{-1} \\ &\le \Big(M\,M_1 + M\,\|V\|_\lambda\Big)\phi_\lambda(t)^{-1}. \end{aligned} \tag{11}$$

The latter expression obviously provides a Lebesgue integrable majorant of (I_n). $\square$

For the bound (11), more precisely for finding some $\mathrm{d}t$-integrable majorant of (I_n), it is essential that we work with the nonuniform sup-norm $\|\cdot\|_\lambda$ and not, as usual, with the uniform sup-norm $\|\cdot\|_\infty$. Of course, other weight functions ϕ instead of ϕ_λ for which (11) holds and which satisfy $\int_{\mathbb{R}}\phi(t)^{-1}\mathrm{d}t < \infty$ are also possible. If F has compact support or if g has compact support in the *open* interval $(0, 1)$ then we could also work with $\|\cdot\|_\infty$ (and $\mathbb{D}$ instead of $\mathbb{D}_\lambda$). However, we do not want to rule out standard (claim) distributions (as, e.g., the Pareto distribution) or the Average Value-at-Risk at level α (which corresponds to $g(x) = (x/\alpha) \wedge 1$ whose compact support is $[0, \alpha]$). For this obvious reason we work with the norm $\|\cdot\|_\lambda$. The next remark illustrates how one can proceed if g does not fulfill the continuity requirement of Theorem 2.2. The remark shows in particular that our proposed method works also for the Value-at-Risk at level α defined in (5) (although in this case one would certainly apply the classical FDM).

Remark 2.4. Suppose that the distortion function g can be written as $g(x) = w_0\,g_0(x) + \sum_{i=1}^m w_i\,\mathbb{1}_{[\alpha_i,1]}(x)$, $x \in [0, 1]$, with $w_0, \dots, w_m \in \mathbb{R}$, $\alpha_1, \dots, \alpha_m \in (0, 1)$ and g_0 as in Theorem 2.2. Notice that $\rho_{g_i}(F)$, with $g_i = \mathbb{1}_{[\alpha_i,1]}(x)$, is nothing but $-F^{\rightarrow}(\alpha_i)$, i.e. the upper α_i-quantile of F multiplied by -1. As is generally known, it is easy to show that, if F is differentiable at $F^{\rightarrow}(\alpha_i)$ with derivative $F'(F^{\rightarrow}(\alpha_i)) > 0$, the mapping ρ_{g_i} is Hadamard differentiable at F tangentially to $\mathbb{C} := \{v \in \mathbb{D} : v \text{ is continuous}\}$ with Hadamard derivative $\dot\rho_{g_i,F}(V) := V(F^{\rightarrow}(\alpha_i))/F'(F^{\rightarrow}(\alpha_i))$. Using exactly the same arguments, one easily obtains that ρ_{g_i} is also quasi-Hadamard differentiable at F, tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with the same "derivative" $\dot\rho_{g_i,F}$. Thus, if g_0 and F are as in Theorem 2.2, and F is in addition differentiable at $F^{\rightarrow}(\alpha_i)$ with derivative $F'(F^{\rightarrow}(\alpha_i)) > 0$ for every $i = 1, \ldots, m$, we immediately obtain that ρ_g is quasi-Hadamard differentiable at F, tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with quasi-Hadamard derivative $\sum_{i=0}^m w_i\dot\rho_{g_i,F}$, where $\dot\rho_{g_0,F}$ is defined as in (8). $\diamond$

Typically, one is interested in the asymptotic distribution of the estimate $\rho_g(F_n)$; for example, to obtain asymptotic confidence intervals for $\rho_g(F)$ or for hypotheses tests. Theorem 2.2 along with the modified FDM of Theorem 4.1 now yields a corresponding result (Theorem 2.5). Illustrating examples for F_n can be found in Section 3. We equip $\mathbb{D}_\lambda$ with the σ-algebra $\mathcal{D}_\lambda := \mathcal{D} \cap \mathbb{D}_\lambda$ to make it also a measurable space, where $\mathcal{D}$ is the σ-algebra generated by the usual coordinate projections $\pi_t : \mathbb{D} \to \overline{\mathbb{R}}$, $t \in \overline{\mathbb{R}}$. Recall that $\mathbb{D}$ is the space of all càdlàg functions on $\overline{\mathbb{R}}$. Also notice that $\mathcal{D}_\lambda$ coincides with the σ-algebra on $\mathbb{D}_\lambda$ generated by the $\|\cdot\|_\lambda$-closed balls (this can easily be shown by following the instructions in [14, Problem IV.2.4] adapted to the norm $\|\cdot\|_\lambda$). For every $n \in \mathbb{N}$, we let F_n be a mapping from a probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ to $\mathbb{D}$.

Theorem 2.5 (*Asymptotic Distribution of $\rho_g(F_n)$*). *Suppose that g and F satisfy conditions* (a) *and* (b) *of* Theorem 2.2, *respectively. Moreover let $\lambda > 1$ and suppose that*

(i) *F_n is $(\mathcal{F}_n, \mathcal{D})$ measurable, and $F_n \in \mathbb{F}_g$ for every $n \in \mathbb{N}$,*
(ii) *$F_n - F$ takes values only in $\mathbb{D}_\lambda$ for every $n \in \mathbb{N}$,*
(iii) *there is some random element B of $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$ with continuous samples such that*

$$\sqrt{n}(F_n - F) \xrightarrow{d} B \quad (\text{in } (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)).$$

Then, if $\dot\rho_{g,F}$ is defined as in (8),

$$\sqrt{n}(\rho_g(F_n) - \rho_g(F)) \xrightarrow{d} \dot\rho_{g,F}(B) \quad (\text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R}))). \tag{12}$$

Proof. First of all we note that $F_n - F$ is $(\mathcal{F}_n, \mathcal{D}_\lambda)$-measurable because we assumed that F_n is $(\mathcal{F}_n, \mathcal{D})$-measurable, $F_n - F \in \mathbb{D}_\lambda$ and $\mathcal{D}_\lambda = \mathcal{D} \cap \mathbb{D}_\lambda$. Moreover, $\dot\rho_{g,F}(B)$ is well defined because B takes values in $\mathbb{C}_\lambda$. Thus, since the subspace $\mathbb{C}_\lambda$ of all continuous functions of $\mathbb{D}_\lambda$ is $\|\cdot\|_\lambda$-separable, assertion (12) is a consequence of the Theorems 2.2 and 4.1. For an application of Theorem 4.1 it only remains to ensure that the mapping $\tilde\omega \mapsto \rho_g(W(\tilde\omega) + F)$ is $(\tilde{\mathcal{F}}, \mathcal{B}(\mathbb{R}))$-measurable whenever W is a measurable mapping from some measurable space $(\tilde\Omega, \tilde{\mathcal{F}})$ to $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$ such that $W(\tilde\omega) + F \in \mathbb{F}_g$ for all $\tilde\omega \in \tilde\Omega$. Since W is $(\tilde{\mathcal{F}}, \mathcal{D}_\lambda)$-measurable and $\mathcal{D}_\lambda$ is the projection σ-field, we obtain in particular $(\tilde{\mathcal{F}}, \mathcal{B}(\mathbb{R}))$-measurability of $\tilde\omega \mapsto W(t, \tilde\omega)$ for every $t \in \overline{\mathbb{R}}$. Along with the representation (2), this yields $(\tilde{\mathcal{F}}, \mathcal{B}(\mathbb{R}))$-measurability of $\tilde\omega \mapsto \rho_g(W(\tilde\omega) + F)$. $\square$

Remark 2.6 (*Asymptotic Normality*). If B in Theorem 2.5 is a Gaussian process with zero mean and measurable covariance function Γ satisfying $\sigma^2 := \int_{\mathbb{R}}\int_{\mathbb{R}} g'(F(s))\Gamma(s,t)g'(F(t))\,\mathrm{d}s\,\mathrm{d}t < \infty$, then the random variable $\dot\rho_{g,F}(B)$ is normally distributed with mean 0 and variance σ^2. This can easily be shown via the characteristic functions by approximating the integral $\int_{\mathbb{R}} g'(F(t))B(t)\,\mathrm{d}t$ by Riemann sums (of Gaussian variables). For the sake of brevity we omit the details. If one is interested in confidence intervals or hypotheses tests then, of course, the asymptotic variance σ^2 has to be estimated consistently. In the next section we also take this problem into account (cf. Remarks 3.3 and 3.8). $\diamond$

3. Examples

3.1. Uncensored i.i.d. data; smoothed empirical df

Suppose $X_1, X_2, \ldots$ are i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with df F. We denote by $\hat F_n := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[X_i,\infty)}$ the corresponding empirical df at stage n. For some purposes it might be beneficial to consider a smoothed version F_n of $\hat F_n$ (smoothed versions of $\hat F_n$ have already been studied in [33]). For instance, for the estimation of the Value-at-Risk $\mathrm{VaR}_\alpha(F)$ by $\rho_g(F_n)$ (with $g(x) = \mathbb{1}_{[\alpha,1]}(x)$) a smoothing leads to a significantly smaller mean square error; cf. [24]. Analogous statements concerning the estimation of the Average Value-at-Risk $\mathrm{AVaR}_\alpha(F)$ by $\rho_g(F_n)$ (with $g(x) = (x/\alpha) \wedge 1$) can be found in [5,19].

Here, we consider a smoothing by the heat kernel (Gaussian kernel). We set $p_\varepsilon(y) := (2\pi\varepsilon)^{-1/2}\exp(-y^2/(2\varepsilon))$ ($\varepsilon > 0$, $y \in \mathbb{R}$), and we denote by $(P_\varepsilon)_{\varepsilon\ge 0}$ the corresponding (heat) semigroup, i.e., $P_\varepsilon\psi(.) := \int_{\mathbb{R}} \psi(y)p_\varepsilon(. - y)\mathrm{d}y$ for $\varepsilon > 0$, and $P_0 := \mathbb{I}$. We focus on the following estimate of F:

$$F_n(t) := P_{\varepsilon_n}\hat F_n(t) = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} \mathbb{1}_{[X_i,\infty)}(y)p_{\varepsilon_n}(t-y)\mathrm{d}y, \quad t \in \mathbb{R}.$$

The next result specifies the asymptotic distribution of the estimate $\rho_g(F_n)$ of $\rho_g(F)$. For a Lebesgue absolutely continuous g the case without smoothing ($\varepsilon_n = 0$ for all n) was already treated in [26] in the context of L-statistics. A related problem was also treated in [18] by use of the "classical" FDM but the conditions on the distortion function g were essentially stronger: g

was assumed to have compact support in the *open* interval (0, 1). Here, and at several places later on, we will assume that F satisfies

$$\limsup_{t\to-\infty} F(t)|t|^{\gamma} < \infty \quad \text{and} \quad \limsup_{t\to\infty} \overline{F}(t)t^{\gamma} < \infty \tag{13}$$

for some suitable exponent γ, where $\overline{F} := 1 - F$ is the tail function of F.

Theorem 3.1 (*Asymptotic Normality*)**.** *Suppose that*

(a) *g satisfies assumption* (a) *of Theorem 2.2,*
(b) *$F \in \mathbb{F}_g$ is Lipschitz continuous, and satisfies assumption* (b) *of Theorem 2.2 and condition (13) for some $\gamma > 2$,*
(c) *$\sqrt{n}\,\varepsilon_n^{\theta} \to 0$ for some $\theta \in (1/4, (\gamma-1)/(2\gamma))$.*

Then

$$law\big(\sqrt{n}(\rho_g(F_n) - \rho_g(F))\big) \xrightarrow{w} N(0, \sigma^2)$$

with

$$\sigma^2 := \int_{\mathbb{R}}\int_{\mathbb{R}} g'(F(s))\Gamma(s,t)g'(F(t))\,\mathrm{d}s\,\mathrm{d}t, \tag{14}$$

where $\Gamma(s,t) = F(s\wedge t)\overline{F}(s\vee t)$. In the case $\varepsilon_n = 0$, $n \in \mathbb{N}$, the df F only needs to be continuous rather than Lipschitz continuous.

Examples for a distortion function g, which satisfies condition (a) of Theorem 2.2, were already given in Example 2.3. Theorem 3.1 yields in particular consistency of $\rho_g(F_n)$ for $\rho_g(F)$. For results on strong consistency see [35].

Proof of Theorem 3.1. First of all notice that (13) with $\gamma > 2$ implies that σ^2 is finite. The limit process of $\sqrt{n}(F_n - F)$ is known to be an F-Brownian bridge B_F°. Corollary A.2 shows that the weak convergence holds in $(\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)$ with $\lambda := \gamma - 2\gamma\theta (\in (1, \gamma/2))$. Thus, in view of Theorem 2.5 and Remark 2.6, the assertion of Theorem 3.1 holds if we can verify conditions (i)–(ii) of Theorem 2.5 and that B_F° has continuous samples. Condition (i) was already verified in [35, Lemma 3], and condition (ii) follows from Corollary A.2. Moreover B_F° has continuous samples since F is continuous and B_F° has the same law as $B^{\circ}(F)$ for some classical Brownian bridge (cf. [25, p.103]) with continuous samples. Thus the result follows from Theorem 2.5 and Remark 2.6. If $\varepsilon_n = 0$, $n \in \mathbb{N}$, we may use Theorem A.1 instead of Corollary A.2. In this case F does not need to be Lipschitz continuous. However, continuity of F is still needed in order to apply Theorem 2.5. □

Remark 3.2. It is worth mentioning that for $\varepsilon_n = 0$ for all $n \in \mathbb{N}$, Theorem 3.1 specifies the asymptotic distribution of L-statistics. The conditions imposed here are very similar to those imposed in [26] for the proof of the asymptotic normality of L-statistics. However, the above result is not restricted to the case where the estimate F_n of F (to be plugged in ρ_g) is given by the empirical df $\hat{F}_n$. Furthermore, the result shows that even in the case where g does not have compact support strictly within the open interval (0, 1), L-statistics can be analyzed by a (modified) FDM. We can, for example, choose $g(x) = x$ leading to the sample mean (cf. the introduction). ◇

Remark 3.3. In order to derive from Theorem 3.1 asymptotic confidence intervals for $\rho_g(F_n)$ one needs a consistent estimate of the asymptotic variance σ^2. A natural estimator is given by the right-hand side of (14) with F replaced by the empirical df at stage n, which we denote by σ_n^2. In this case, strong consistency of σ_n^2 for σ^2 is known from [26, Theorem 1]. More generally, one can replace F by the smoothed empirical df F_n (defined in (13)) with $\varepsilon_n \downarrow 0$. Then, in the setting of Theorem 3.1, one can still prove strong consistency of σ_n^2 for σ^2. For the sake of brevity we omit the details. ◇

3.2. Censored data

In insurance practice one often encounters the problem that the data set is censored. For instance, one might have the relation $X = \tilde{X}/C$ between the actual (but unobservable) claim X and the observable fraction $\tilde{X}$, where C is some random variable taking values in the interval (0, 1]. Here X and $\tilde{X}$ are nonpositive random variables (a negative value of X corresponds to a payout to the client), and $\tilde{X}$ and C are assumed to be independent. A similar assumption appears in the factor model for claim reserving which is the basis of the widely used chain ladder method. We denote by F, $\tilde{F}$ and H the df of X, that of $\tilde{X}$ and that of C (respectively), and we assume that H is known. The goal is the estimation of

$$\rho_g(F) = -\int_{\mathbb{R}} x\,\mathrm{d}g(F(x)) = \int_{-\infty}^{0} g(F(t))\mathrm{d}t$$

based on i.i.d. copies $\tilde{X}_1, \ldots, \tilde{X}_n$ of $\tilde{X}$. We clearly have the representation $F(t) = \int_0^1 \tilde{F}(tz)\mathrm{d}H(z)$, $t \le 0$, for the df F. Thus a natural estimator F_n for F based on the censored observations $\tilde{X}_1, \ldots, \tilde{X}_n$ is

$$F_n(t) = \int_0^1 \tilde{F}_n(tz)\mathrm{d}H(z) = \frac{1}{n}\sum_{i=1}^{n} H((\tilde{X}_i/t)\wedge 1), \quad t < 0, \tag{15}$$

with $\tilde{F}_n := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[\tilde{X}_i,0]}$ the empirical df of $\tilde{X}_1, \ldots, \tilde{X}_n$.

We now turn to the asymptotic distribution of the estimate $\rho_g(F_n)$ of $\rho_g(F)$. It will be given in Theorem 3.7 below. For every bounded $V \in \mathbb{D}$ we set

$$\tau_H(V)(t) := \int_0^1 V(zt)\mathrm{d}H(z), \quad t \in \overline{\mathbb{R}}. \tag{16}$$

The following lemma shows that τ_H provides a continuous linear operator on $\mathbb{D}_\lambda$ if H's mass near the origin is only moderate; Lemma 3.5 below gives a more transparent sufficient condition.

Lemma 3.4. *If H satisfies*

$$\sup_{t\in\mathbb{R}} \phi_\lambda(t) \int_0^1 \phi_\lambda(zt)^{-1}\mathrm{d}H(z) < \infty \tag{17}$$

then (16) provides a continuous linear operator $\tau_H : \mathbb{D}_\lambda \to \mathbb{D}_\lambda$, $V \mapsto \tau_H(V)$.

Proof. For $V \in \mathbb{D}_\lambda$ we obtain

$$\begin{aligned}\|\tau_H(V)\|_\lambda &= \sup_{t\in\mathbb{R}} |\tau_H(V)(t)\phi_\lambda(t)| \\ &= \sup_{t\in\mathbb{R}} \phi_\lambda(t) \int_0^1 \phi_\lambda(zt)^{-1} \{\phi_\lambda(zt)V(zt)\}\, \mathrm{d}H(z) \\ &\le K\|V\|_\lambda\end{aligned}$$

with $K := \sup_{t\in\mathbb{R}} \phi_\lambda(t) \int_0^1 \phi_\lambda(zt)^{-1}\mathrm{d}H(z)$. Since K is finite by (17), τ_H restricted to $\mathbb{D}_\lambda$ is indeed a continuous linear operator from $\mathbb{D}_\lambda$ to $\mathbb{D}_\lambda$. □

Lemma 3.5. *Let $K > 0$, $\delta > 1$, and suppose that H possesses a Lebesgue density h which satisfies $h(x) \le Kx^{\delta-1}$ for all $x \in (0, 1]$. Then assumption (17) of Lemma 3.4 is satisfied for every $\lambda \in (1, \delta)$.*

Proof. We have

$$\begin{aligned}\phi_\lambda(t) \int_0^1 \phi_\lambda(tz)^{-1}\mathrm{d}H(z) &\le (1+|t|)^\lambda \int_0^1 \frac{1}{(1+|tz|)^\lambda} Kz^{\delta-1}\mathrm{d}z \\ &= K\frac{(1+|t|)^\lambda}{|t|^\delta} \int_0^{|t|} \frac{u^{\delta-1}}{(1+u)^\lambda}\mathrm{d}u.\end{aligned}$$

The latter term is bounded above uniformly in $t \in \mathbb{R}$, which proves (17). □

Example 3.6. Let H be the df of the beta distribution on $(0, 1]$ with parameters $a, b > 0$. If $a > 1$, then H satisfies the conditions of Lemma 3.5 for $\delta := a$. ◇

Theorem 3.7 (*Asymptotic Normality*)**.** *Suppose that*

(a) *g satisfies assumption* (a) *of Theorem 2.2,*
(b) *$F \in \mathbb{F}_g$ satisfies assumption* (b) *of Theorem 2.2 and condition (13) for some $\gamma > 2$,*
(c) *$\tilde{F}$ is continuous,*
(d) *H satisfies the assumptions of Lemma 3.5.*

Then

$$law\big(\sqrt{n}(\rho_g(F_n) - \rho_g(F))\big) \xrightarrow{w} N(0, \sigma^2) \tag{18}$$

with

$$\sigma^2 := \int_{\mathbb{R}}\int_{\mathbb{R}} g'(F(s))\Gamma(s,t)g'(F(t))\,\mathrm{d}s\,\mathrm{d}t, \tag{19}$$

where $\Gamma(s,t) := \int_0^1\int_0^1 [\tilde{F}(xt \wedge ys)(1 - \tilde{F}(xt \vee ys))]\mathrm{d}H(x)\mathrm{d}H(y)$.

Theorem 3.7 yields in particular consistency of $\rho_g(F_n)$ for $\rho_g(F)$. For results on strong consistency see again [35]. Examples for a distortion function g, which satisfies condition (a) of Theorem 2.2, were already given in Example 2.3.

Proof of Theorem 3.7. By Lemmas 3.4 and 3.5, τ_H defined in (16) provides a continuous linear operator from $\mathbb{D}_\lambda$ to $\mathbb{D}_\lambda$ for any $\lambda \in (1, \min\{\gamma/2, \delta\})$. Further, condition (13) for F implies that condition (13) also holds for $\tilde{F}$, so that Theorem A.1 yields

$$\sqrt{n}(\tilde{F}_n - \tilde{F}) \xrightarrow{d} B^\circ_{\tilde{F}} \quad (\text{in } (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)).$$

Therefore the Continuous Mapping Theorem gives

$$\begin{aligned}\sqrt{n}(F_n - F) &= \sqrt{n}(\tau_H(\tilde{F}_n) - \tau_H(\tilde{F}))\\ &= \tau_H(\sqrt{n}(\tilde{F}_n - \tilde{F}))\\ &\xrightarrow{d} \tau_H(B^\circ_{\tilde{F}}) \quad (\text{in } (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)),\end{aligned}$$

where the Continuous Mapping Theorem is applicable since the process $B^\circ_{\tilde{F}} = B^\circ(\tilde{F})$ may be assumed to be continuous ($\tilde{F}$ was assumed to be continuous) and the subspace of continuous functions in $\mathbb{D}_\lambda$ is $\|\cdot\|_\lambda$-separable. The process $\tau_H(B^\circ_{\tilde{F}})(t) = \int_0^1 B^\circ_{\tilde{F}}(zt)\mathrm{d}H(z)$, $t \le 0$, is easily seen to be Gaussian. It is centered since $\mathbb{E}[\tau_H(B^\circ_{\tilde{F}})(t)] = \mathbb{E}[\int_0^1 B^\circ_{\tilde{F}}(zt)\,\mathrm{d}H(z)] = 0$ for all $t \le 0$ by Fubini's theorem. Moreover its covariance function is given by

$$\begin{aligned}\Gamma(s,t) &= \mathbb{E}\big[\tau_H(B^\circ_{\tilde{F}})(t)\tau_H(B^\circ_{\tilde{F}})(s)\big]\\ &= \mathbb{E}\Big[\Big(\int_0^1 B^\circ_{\tilde{F}}(zt)\mathrm{d}H(z)\Big)\Big(\int_0^1 B^\circ_{\tilde{F}}(zs)\mathrm{d}H(z)\Big)\Big]\\ &= \int_0^1\int_0^1 \mathbb{E}\big[B^\circ_{\tilde{F}}(xt)B^\circ_{\tilde{F}}(ys)\big]\mathrm{d}H(x)\mathrm{d}H(y)\\ &= \int_0^1\int_0^1 \Big(\tilde{F}(xt\wedge ys)(1-\tilde{F}(xt\vee ys))\Big)\mathrm{d}H(x)\mathrm{d}H(y), \quad t,s \le 0,\end{aligned}$$

where we again used Fubini's theorem. Now $\dot{\rho}_{g,F}(\tau_H(B^\circ_{\tilde{F}})) \sim N(0,\sigma^2)$ follows from Remark 2.6.

Thus, in view of Theorem 2.5 and Remark 2.6, for (18) it remains to verify conditions (i), (ii) of Theorem 2.5. Condition (i) was already verified in [35, Lemma 5] under the assumptions of Lemma 3.5, and condition (ii) with $\lambda \in (1, \min\{\gamma/2, \delta\})$ is included in the statement of Theorem A.1. □

Remark 3.8. In order to derive from Theorem 3.7 asymptotic confidence intervals for $\rho_g(F_n)$ one needs a consistent estimate of the asymptotic variance σ^2. A natural estimator is given by the right-hand side of (19) with F replaced by F_n, which we denote by σ_n^2. One can indeed show that σ_n^2 converges $\mathbb{P}$-almost surely to σ^2, provided g' is Lebesgue almost everywhere continuous and $\tilde{F}$ takes the value d at most once if d is a discontinuity of g'. So as not to break the flow of presentation we do not go into details. ◇

3.3. Dependent data

Finally, we briefly discuss how the method can be used to estimate $\rho_g(F)$ based on dependent data. Let (X_i) be a strictly stationary sequence of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $F_n = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[X_i,\infty)}$ be the corresponding empirical df at stage n. We will consider two popular dependency structures: α- and ϱ-mixing. For the definition of these (and other) mixing conditions and for examples of strictly stationary α- or ϱ-mixing sequences see, e.g., [3,7,13]. As usual, the corresponding mixing coefficients will be referred to as $\alpha(n)$ and $\rho(n)$, respectively.

If (X_i) is strictly stationary and α-mixing, then we can combine Theorem 2.5 and results of [23] on weighted empirical processes of α-mixing sequences to derive the asymptotic distribution of $\rho_g(F_n)$.

Theorem 3.9. *Let (X_i) be a strictly stationary α-mixing sequence of random variables with common df F. Suppose that*

(a) *g satisfies assumption* (a) *of* Theorem 2.2,
(b) *$\alpha(n) = \mathcal{O}(n^{-\theta})$ for some $\theta > 1+\sqrt{2}$,*
(c) *$F \in \mathbb{F}_g$ is continuous, and satisfies assumption* (b) *of* Theorem 2.2 *and condition* (13) *for some $\gamma > \frac{2\theta}{\theta-1}$.*

Then

$$law\big(\sqrt{n}(\rho_g(F_n) - \rho_g(F))\big) \xrightarrow{w} N(0,\sigma^2) \qquad (20)$$

with

$$\sigma^2 := \int_{\mathbb{R}}\int_{\mathbb{R}} g'(F(s))\Gamma(s,t)g'(F(t))\,\mathrm{d}s\,\mathrm{d}t,$$

where

$$\Gamma(s,t) = F(s\wedge t)\bar{F}(s\vee t) + \sum_{k=2}^{\infty}\Big[\mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1\le s\}}, \mathbb{1}_{\{X_k\le t\}}) + \mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1\le t\}}, \mathbb{1}_{\{X_k\le s\}})\Big].$$

Proof. Theorem 2.2 of [23] (along with our assumptions (b) and (c)) and a transformation of the process $(F_n(\cdot)-F(\cdot))\phi_\lambda(\cdot)$ to a uniform empirical process yield that $\sqrt{n}(F_n - F)$ weakly converges to a centered Gaussian process B_F°, with covariance function Γ, in $(\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)$ for any $\lambda \in (1, \frac{\gamma(\theta-1)}{2\theta})$. Thus, in view of Theorem 2.5 and Remark 2.6, it remains to verify conditions (i) and (ii) of Theorem 2.5 and that B_F° has continuous samples. Condition (i) obviously holds by (2) and the fact that $F_n(t) = 0$, $t < t_1(\omega)$, and $F_n(t) = 1$, $t > t_2(\omega)$. Condition (ii) is ensured by Theorem 2.2 of [23] (along with the aforementioned transformation). Finally, the continuity of B_F° follows from the fact that the limit process of the above uniform empirical process has continuous samples (cf. the proof of Theorem 2.2 of [23] and apply the Kolmogorov–Čentsov Theorem), the continuity of F and the Continuous Mapping Theorem. Notice also that σ^2 is finite due to the facts that g' is bounded and that B_F° lies in $\mathbb{D}_\lambda$. □

For the particular case $g(x) = (x/\alpha)\wedge 1$ (which corresponds to the Average Value-at-Risk at level α) and a kernel estimator see also [19]. Notice that for an i.i.d. sequence (X_i) condition (b) of Theorem 3.9 is fulfilled for θ arbitrarily large. Thus, in that case, conditions (b) and (c) of Theorem 3.9 are equivalent to the conditions imposed on F in Theorem 3.1 for the case without smoothing (i.e. $\varepsilon_n = 0$ for all $n \in \mathbb{N}$).

For strictly stationary GARCH(p,q) processes, which are often used in mathematical finance, one has under some mild regularity conditions that $\alpha(n) \le c\varrho^n$, $n\in\mathbb{N}$, for some constants $c>0$ and $\varrho\in(0,1)$; cf. [13]. Thus, these GARCH processes always satisfy assumption (b) of Theorem 3.9.

If (X_i) strictly stationary and ϱ-mixing, then we can combine Theorem 2.5 and a result of [23] on weighted empirical processes of ϱ-mixing sequences to derive the asymptotic distribution of $\rho_g(F_n)$.

Theorem 3.10. *Let (X_i) be a strictly stationary ϱ-mixing sequences of random variables with common df F. Suppose that*

(a) *g satisfies assumption* (a) *of Theorem* 2.2*,*
(b) *$F\in\mathbb{F}_g$ is continuous, and satisfies assumption* (b) *of Theorem* 2.2 *and condition* (13) *for some $\gamma>2$,*
(c) $\sum_{k=2}^{\infty}\left|\mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1\le s\}}, \mathbb{1}_{\{X_k\le t\}}) + \mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1\le t\}}, \mathbb{1}_{\{X_k\le s\}})\right| < \infty$,
(d) $\sum_{n=1}^{\infty}\varrho(2^n) < \infty$.

Then

$$law\big(\sqrt{n}(\rho_g(F_n)-\rho_g(F))\big) \stackrel{w}{\to} N(0,\sigma^2) \tag{21}$$

with σ^2 (and Γ) as in Theorem 3.9*.*

Proof. The proof follows along the lines of the proof of Theorem 3.9 (with any $\lambda\in(1,\gamma/2)$) by using that Theorem 2.3 of [23] is applicable due to our assumptions (b)–(d), and is therefore omitted. □

Finally, it is worth mentioning that results similar to those of Theorems 3.9 and 3.10 can be derived for other dependency concepts by using the quasi-Hadamard differentiability of the map $\rho_g(F)$ and the modified FDM. For instance, analogous results can be proven for strictly stationary associated sequences of random variables by again using results of [23], or for strictly stationary sequences fulfilling the dependence concept of Wu [32]. Strictly stationary associated sequences of random variables arise naturally in mathematical finance. For instance, ARCH(∞) processes are often used to model financial log-returns, and it is well known that strictly stationary squared ARCH(∞) processes are associated.

4. Modified functional delta method

In this section we present our modification of the FDM. For a discussion of the classical FDM see, e.g., [1,10,16,17,20,25,28,29].

Let $\mathbf{V}$ and $\mathbf{V}'$ be vector spaces, and $\mathbf{V}_0$ be a subspace of $\mathbf{V}$. Let $\|\cdot\|_{\mathbf{V}_0}$ and $\|\cdot\|_{\mathbf{V}'}$ be norms on $\mathbf{V}_0$ and $\mathbf{V}'$, respectively. Moreover let $\mathcal{V}_0$ and $\mathcal{V}'$ be σ-algebras on $\mathbf{V}_0$ and $\mathbf{V}'$, respectively. Suppose that $\mathcal{V}_0$ is nested between the open ball and the Borel σ-algebra on $\mathbf{V}_0$, and that $\mathcal{V}'$ is not larger than the Borel σ-algebra on $\mathbf{V}'$. We emphasize again that $\|\cdot\|_{\mathbf{V}_0}$ is not required to be a norm on all of $\mathbf{V}$. For every $n\in\mathbb{N}$, let $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ be a probability space, and T_n be a map from Ω_n to $\mathbf{V}$. We then have the following result which involves the notion of quasi-Hadamard differentiability introduced in Definition 2.1. For the notion of weak convergence in metric spaces we refer to [14, Chapter IV].

Theorem 4.1 (*Modified Functional Delta Method*)**.** *Let $f:\mathbf{V}_f\to\mathbf{V}'$ be a map defined on some subset $\mathbf{V}_f$ of $\mathbf{V}$, let $\theta\in\mathbf{V}_f$, let $\mathbb{C}_0$ be some subset of $\mathbf{V}_0$ being separable w.r.t. $\|\cdot\|_{\mathbf{V}_0}$ (we regard $\|\cdot\|_{\mathbf{V}_0}$ as a metric if $\mathbb{C}_0$ is not a vector space), and suppose that*

(i) *T_n takes values only in $\mathbf{V}_f$,*
(ii) *$T_n - \theta$ takes values only in $\mathbf{V}_0$, is $(\mathcal{F}_n, \mathcal{V}_0)$-measurable and satisfies*

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} V \quad (\textit{in } (\mathbf{V}_0, \mathcal{V}_0, \|\cdot\|_{\mathbf{V}_0})) \tag{22}$$

for some random element V of $(\mathbf{V}_0, \mathcal{V}_0)$, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values only in $\mathbb{C}_0$,
(iii) *$\tilde{\omega} \mapsto f(W(\tilde{\omega}) + \theta)$ is $(\tilde{\mathcal{F}}, \mathcal{V}')$-measurable whenever W is a measurable mapping from some measurable space $(\tilde{\Omega}, \tilde{\mathcal{F}})$ to $(\mathbf{V}_0, \mathcal{V}_0)$ such that $W(\tilde{\omega}) + \theta \in \mathbf{V}_f$ for all $\tilde{\omega} \in \tilde{\Omega}$,*
(iv) *f is quasi-Hadamard differentiable at θ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$ with quasi-Hadamard derivative $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f$.*

Then

$$\sqrt{n}(f(T_n) - f(\theta)) \xrightarrow{d} D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f\,(V) \quad (\textit{in } (\mathbf{V}', \mathcal{V}', \|\cdot\|_{\mathbf{V}'})). \tag{23}$$

Proof. We may and do follow the arguments of the proof sketch for Theorem 3 in [10]. First of all we have to clarify some measurability issue: In order to ensure that assertion (23) makes sense, we have to show that both $f(T_n)$ and $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f\,(V)$ are $\mathcal{V}'$-measurable. Setting $V_n := \sqrt{n}(T_n-\theta)$ and noting that $n^{-1/2}V_n$ is $(\mathcal{F}_n, \mathcal{V}_0)$-measurable by (ii) and that $n^{-1/2}V_n+\theta = T_n$ takes values only in $\mathbf{V}_f$ by (i), the $(\mathcal{F}_n, \mathcal{V}')$-measurability of $f(T_n) = f(n^{-1/2}V_n + \theta)$ is an immediate consequence of (iii). Further, V (seen as a mapping from Ω to $\mathbb{C}_0$) is clearly $(\mathcal{F}, \mathcal{V}_0 \cap \mathbb{C}_0)$-measurable. Since $\mathbb{C}_0$ is $\|\cdot\|_{\mathbf{V}_0}$-separable, the σ-algebra $\mathcal{V}_0 \cap \mathbb{C}_0$ coincides with the Borel σ-algebra $\mathcal{B}(\mathbb{C}_0)$ on $\mathbb{C}_0$ w.r.t. $\|\cdot\|_{\mathbf{V}_0}$. Because $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f$ as the quasi-Hadamard derivative is $(\|\cdot\|_{\mathbf{V}_0}, \|\cdot\|_{\mathbf{V}'})$-continuous and $\mathcal{V}'$ was assumed to be not larger than the Borel σ-algebra, we also have $(\mathcal{B}(\mathbb{C}_0), \mathcal{V}')$-measurability of $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f$. That is, $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f\,(V)$ is indeed $(\mathcal{F}, \mathcal{V}')$-measurable.

Now, the convergence in (22) and the Skorohod–Dudley–Wichura almost sure representation theorem (see, e.g., [10, Theorem 2]) imply the existence of some random elements $\tilde{V}, \tilde{V}_1, \tilde{V}_2, \ldots$ of $(\mathbf{V}_0, \mathcal{V}_0)$ on a common probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ with $V \stackrel{d}{=} \tilde{V}$, $V_n \stackrel{d}{=} \tilde{V}_n$ as well as $\|\tilde{V} - \tilde{V}_n\|_{\mathbf{V}_0} \to 0$ and $\tilde{V} \in \mathbb{C}_0$ $\tilde{\mathbb{P}}$-almost surely (for the application of Theorem 2 in [10] we need the assumptions that V takes values only in $\mathbb{C}_0$ and that $\mathcal{V}_0$ is nested between the open ball and the Borel σ-algebra). We set $\tilde{T}_n := n^{-1/2}\tilde{V}_n + \theta$ and note that $T_n \stackrel{d}{=} \tilde{T}_n$. We may and do modify $\tilde{V}$ and $\tilde{V}_n$ in such a way that $\tilde{V}$ takes values only in $\mathbb{C}_0$, that $n^{-1/2}\tilde{V}_n + \theta$ takes values only in $\mathbf{V}_f$, and that $\|\tilde{V} - \tilde{V}_n\|_{\mathbf{V}_0} \to 0$ everywhere. As above we then obtain that both $f(\tilde{T}_n) = f(n^{-1/2}\tilde{V}_n + \theta)$ and $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f\,(\tilde{V})$ are $(\tilde{\mathcal{F}}, \mathcal{V}')$-measurable. Further, we plainly have

$$\sqrt{n}\,(f(\tilde{T}_n) - f(\theta)) = \frac{f(\theta + n^{-1/2}\sqrt{n}(\tilde{T}_n - \theta)) - f(\theta)}{n^{-1/2}}. \tag{24}$$

By the quasi-Hadamard differentiability of f at θ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$ (along with $\|\tilde{V} - \tilde{V}_n\|_{\mathbf{V}_0} \to 0$ and the fact that $\tilde{V}$ takes values only in $\mathbb{C}_0$), we have that the right-hand side of (24) converges $\tilde{\mathbb{P}}$-almost surely to $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}f\,(\tilde{V})$ in $\|\cdot\|_{\mathbf{V}'}$ (the role of h_n in (7) is played by $n^{-1/2}$), and so does the left-hand side in (24). Thus, since the left-hand side in (23) and the left-hand side in (24) clearly induce the same law on $(\mathbf{V}', \mathcal{V}')$ and since almost sure convergence induces convergence in distribution, this implies the claim of Theorem 4.1. □

The measurability assumption (iii) in the preceding theorem is similar to the one in Corollary 1 in [10]. It ensures that the left-hand sides in (23) and (24) are indeed random elements of $\mathbf{V}'$. This assumption could be dropped by employing the concept of outer integrals; see, for example, [28,29] for this concept.

Remark 4.2. The result of Theorem 4.1 still holds if in Definition 2.1 and in Theorem 4.1 the norms on $\mathbf{V}_0$ and $\mathbf{V}'$ are replaced by metrics. The above proof obviously still works in this setting. ◇

Acknowledgments

The authors wish to thank an associate editor and a referee for their careful reading, and for useful hints and comments.

Appendix. Weak convergence of empirical processes in $\mathbb{D}_\lambda$

Let $\mathbb{D}, \phi_\lambda, \mathbb{D}_\lambda$ and $\|\cdot\|_\lambda$ be defined as in Section 2, and equip $\mathbb{D}$ and $\mathbb{D}_\lambda$ with the sup-norm $\|\cdot\|_\infty$ and the λ-norm $\|\cdot\|_\lambda$, respectively. Further, let $\mathcal{D}$ and $\mathcal{D}_\lambda$ be defined as in Section 2 before Theorem 2.5. Suppose F is a df on the real line. Further suppose $X_1, X_2, \ldots$ are i.i.d. random variables with df F on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\hat{F}_n := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[X_i,\infty)}$ denote the corresponding empirical df at stage n. It is well known that the empirical process $\hat{V}_n := \sqrt{n}(\hat{F}_n - F)$ converges in distribution in $(\mathbb{D}, \mathcal{D}, \|\cdot\|_\infty)$ to an F-Brownian bridge B_F°, i.e. to a centered Gaussian process with covariance function $\Gamma(s,t) = F(s \wedge t)\overline{F}(s \vee t)$.

For our purposes we need a stronger convergence result. In fact we need the weak convergence w.r.t. the more stringent norm $\|\cdot\|_\lambda$, i.e. $\hat{V}_n := \sqrt{n}(\hat{F}_n - F) \xrightarrow{d} B_F^\circ$ in $(\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)$. In Section 2 we chose $\lambda > 1$, but the results to be given in this section hold for arbitrary $\lambda \geq 0$. In order to obtain the required convergence we have to impose an additional assumption on F. We assume

$$\exists\, \gamma > 2\lambda : \qquad \limsup_{t\to-\infty} F(t)|t|^\gamma < \infty \quad \text{and} \quad \limsup_{t\to\infty} \overline{F}(t)t^\gamma < \infty, \tag{A.1}$$

where $\overline{F} := 1-F$ denotes the tail function of F. The claim of the following theorem is an immediate consequence of Theorem 6.2.1 in [25] since under condition (A.1) we have $\int_{\mathbb{R}} \phi_{2\lambda}(t)\mathrm{d}F(t) < \infty$.

Theorem A.1. *Set $\hat{V}_n := \sqrt{n}(\hat{F}_n - F)$. If F satisfies (A.1) then $\|\hat{V}_n - B_F^\circ\|_\lambda \xrightarrow{p} 0$. In particular, $\hat{V}_n \xrightarrow{d} B_F^\circ$ (in $(\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)$).*

The convergence result of Theorem A.1 remains true when replacing the empirical df $\hat{F}_n$ by its smoothed version $F_n := P_{\varepsilon_n}\hat{F}_n$ (cf. (12)), provided ε_n converges to 0 sufficiently fast and F is sufficiently regular. For corresponding results on the weak convergence in the sup-norm see, e.g., [27,34].

Corollary A.2. *Set $V_n := \sqrt{n}(F_n - F)$, and suppose that F is Lipschitz continuous and satisfies (A.1). Further suppose that $\sqrt{n}\,\varepsilon_n^{(\gamma-\lambda)/(2\gamma)} \to 0$. Then we have $\|V_n - B_F^\circ\|_\lambda \xrightarrow{p} 0$. In particular, $V_n \xrightarrow{d} B_F^\circ$ (in $(\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)$).*

For the proof of Corollary A.2 we need the following lemma. Let $\mathbb{B}_\lambda$ be the space of all measurable functions f on $\mathbb{R}$ with $\|f\|_\lambda < \infty$. Moreover, let $\mathbb{B}_{\lambda,0}$ be the space of all bounded measurable functions f on $\mathbb{R}$ satisfying $f(t)\phi_\lambda(t) \to 0$ as $|t| \to \infty$.

Lemma A.3. *Let $\lambda \geq 0$. For every $f \in \mathbb{B}_\lambda$ there is some constant $C_\lambda > 0$ such that*

$$\|P_\varepsilon f\|_\lambda \leq C_\lambda \|f\|_\lambda \tag{A.2}$$

for all $\varepsilon \in (0, 1]$. Moreover, if $f \in \mathbb{B}_{\lambda,0}$ then

$$\lim_{\varepsilon\downarrow 0} \|P_\varepsilon f - f\|_\lambda = 0. \tag{A.3}$$

Proof. It can easily be shown that $\int_{\mathbb{R}} \phi_\lambda(y)^{-1} p_\varepsilon(t-y)\mathrm{d}y \leq C_\lambda \phi_\lambda(t)^{-1}$ for some suitable constant $C_\lambda > 0$ independent of t (cf. [35, Lemma 1]). In view of this inequality the bound (A.2) is more or less obvious.

In order to verify (A.3) pick $f \in \mathbb{B}_{\lambda,0}$. First we notice that

$$\phi_\lambda(y)^{-1}|\phi_\lambda(t) - \phi_\lambda(y)| \leq C_\lambda\big(|t-y| + |t-y|^\lambda\big) \tag{A.4}$$

for some suitable constant $C_\lambda > 0$. For $\lambda \in [0, 1]$ the inequality (A.4) is obvious, and for $\lambda > 1$ it can be obtained as follows:

$$\begin{aligned}
\phi_\lambda(y)^{-1}|\phi_\lambda(t) - \phi_\lambda(y)| &\leq (1+|y|)^{-\lambda}\, |t-y|\lambda\big(1 + (|t| \vee |y|)\big)^{\lambda-1} \\
&\leq \lambda|t-y|\, \frac{\big(1 + (|t| \wedge |y|) + |t-y|\big)^{\lambda-1}}{(1+|y|)^\lambda} \\
&\leq \lambda 2^{(\lambda-2)\vee 0}\big(|t-y| + |t-y|^\lambda\big).
\end{aligned}$$

With the help of (A.4) we obtain

$$\begin{aligned}
|P_\varepsilon f(t) - f(t)|\phi_\lambda(t) &= \left|\int_{\mathbb{R}} p_\varepsilon(t-y)\big(f(y) - f(t)\big)\phi_\lambda(t)\mathrm{d}y\right| \\
&\leq \left|\int_{\mathbb{R}} p_\varepsilon(t-y)\big(f(y)\phi_\lambda(y) - f(t)\phi_\lambda(t)\big)\mathrm{d}y\right| + \left|\int_{\mathbb{R}} p_\varepsilon(t-y)f(y)\big(\phi_\lambda(t) - \phi_\lambda(y)\big)\mathrm{d}y\right| \\
&\leq \|P_\varepsilon(f\phi_\lambda) - f\phi_\lambda\|_\infty + \|f\|_\lambda \int_{\mathbb{R}} p_\varepsilon(t-y)\phi_\lambda(y)^{-1}\big(\phi_\lambda(t) - \phi_\lambda(y)\big)\mathrm{d}y \\
&\leq \|P_\varepsilon(f\phi_\lambda) - f\phi_\lambda\|_\infty + \|f\|_\lambda \int_{\mathbb{R}} p_\varepsilon(t-y)C_\lambda\big(|t-y| + |t-y|^\lambda\big)\mathrm{d}y \\
&=: S_1(t,\varepsilon) + S_2(t,\varepsilon).
\end{aligned}$$

Since $f \in \mathbb{B}_{\lambda,0}$, the function $f\phi_\lambda$ lies in the space of all bounded measurable functions on $\mathbb{R}$ vanishing at infinity. Since on the latter space the operator semigroup (P_ε) is $\|\cdot\|_\infty$-continuous on $[0, \infty)$, the summand $S_1(t,\varepsilon)$ converges to 0 as $\varepsilon \downarrow 0$ uniformly in $t \in \mathbb{R}$. Further, for every $\kappa > 0$ we have $\int_{\mathbb{R}} |t-y|^\kappa p_\varepsilon(t-y)\mathrm{d}y \leq C_\kappa \varepsilon^\kappa/2$ for all $t \in \mathbb{R}$ and $\varepsilon \in (0, 1]$, where $C_\kappa > 0$ is some suitable constant independent of t and ε (see [4]). Thus the summand $S_2(t,\varepsilon)$ converges to 0 as $\varepsilon \downarrow 0$ uniformly in $t \in \mathbb{R}$. Hence, $S_1(t,\varepsilon) + S_2(t,\varepsilon)$ converges to 0 as $\varepsilon \downarrow 0$ uniformly in $t \in \mathbb{R}$. This gives (A.3). □

Proof of Corollary A.2. We have

$$\begin{aligned}\|V_n - B_F^\circ\|_\lambda &= \|\sqrt{n}(F_n - F) - B_F^\circ\|_\lambda \\ &\le \|P_{\varepsilon_n}(\sqrt{n}(\hat{F}_n - F) - B_F^\circ)\|_\lambda + \|P_{\varepsilon_n} B_F^\circ - B_F^\circ\|_\lambda + \sqrt{n}\,\|(P_{\varepsilon_n}F - F)\|_\lambda \\ &=: S_1(n) + S_2(n) + S_3(n).\end{aligned}$$

In [35, proof of Lemma 2] it is shown that the assumptions on F in Corollary A.2 ensure $\|P_{\varepsilon_n}F - F\|_\lambda \le C_{\gamma,\lambda}\varepsilon_n^{(\gamma-\lambda)/(2\gamma)}$ for some suitable constant $C_{\gamma,\lambda} > 0$ (this is the only point where we need the Lipschitz continuity of F). Thus, by the assumption on ε_n, we obtain $S_3(n) \to 0$.

Further, notice that B_F° lies in $\mathbb{B}_{\lambda,0}$ (which was introduced before Lemma A.3) $\mathbb{P}$-almost surely. To see this, it suffices to show $\limsup_{|t|\to\infty} B_F^\circ(t)\phi_\lambda(t) = 0$. Let us suppose the contrary, i.e. suppose there is some $\delta > 0$ and a sequence (t_n) with $|t_n| \to \infty$ and $|B_F^\circ(t_n)|\phi_\lambda(t_n) \ge \delta$ for all n. On the other hand we have

$$\mathbb{E}[(B_F^\circ(t)\phi_\lambda(t))^2] = \phi_\lambda(t)^2 F(t)\overline{F}(t). \tag{A.5}$$

Assumption (A.1) implies in particular that $F(t)\phi_\lambda(t)^2$ and $\overline{F}(t)\phi_\lambda(t)^2$ converge to 0 as $t \to -\infty$ and $t \to \infty$, respectively. Along with (A.5) this implies $\lim_{n\to\infty} B_F^\circ(t_n)\phi_\lambda(t_n) \to 0$ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. In particular we can find some subsequence $(t_{n_k}) \subset (t_n)$ such that $\lim_{k\to\infty} B_F^\circ(t_{n_k})\phi_\lambda(t_{n_k}) \to 0$ $\mathbb{P}$-almost surely. This gives a contradiction. So we do indeed have $B_F^\circ \in \mathbb{B}_{\lambda,0}$ $\mathbb{P}$-almost surely. The second part of Lemma A.3 then implies that $S_2(n)$ converges to 0 $\mathbb{P}$-almost surely.

Finally, we have $\sqrt{n}(\hat{F}_n - F) - B_F^\circ \in \mathbb{B}_\lambda$ since both the minuend and the subtrahend lie in $\mathbb{B}_\lambda$. The first part of Lemma A.3 then implies that $S_1(n)$ is bounded above by $C_\lambda\|\sqrt{n}(\hat{F}_n - F) - B_F^\circ\|_\lambda$ for some constant $C_\lambda > 0$. By Theorem A.1 we deduce $S_1(n) \xrightarrow{p} 0$.

Hence, we have $S_1(n) + S_2(n) + S_3(n) \xrightarrow{p} 0$, from which we can deduce $\|V_n - B_F^\circ\|_\lambda \xrightarrow{p} 0$ (cf. Corollary 2.3.1 of [25]). □

References

[1] P.K. Anderson, Ø. Borgan, R.D. Gill, N. Keiding, Statistical Models Based on Counting Processes, Springer-Verlag, New York, 1993.
[2] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, Mathematical Finance 9 (1999) 203–228.
[3] R.C. Bradley, Basic properties of strong mixing conditions. A survey and some open problems, Probability Surveys 2 (2005) 107–144.
[4] I.N. Bronstein, K.A. Semendyayev, G. Musiol, H. Mühlig, Handbook of Mathematics, Springer-Verlag, Berlin, 2002.
[5] S.X. Chen, Nonparametric estimation of expected shortfall, Journal of Financial Econometrics 6 (2008) 87–107.
[6] F. Delbaen, Coherent Risk Measures, Lecture Notes, Pisa, 2000.
[7] P. Doukhan, Mixing: Properties and Examples, in: Lecture Notes in Statistics, Springer-Verlag, New York, 1994.
[8] L.T. Fernholz, Von Mises calculus for statistical functionals, in: Lecture Notes in Statistics, vol. 19, Springer-Verlag, New York, 1983.
[9] H. Föllmer, A. Schied, Stochastic Finance. An Introduction in Discrete Time, de Gruyter, Berlin, 2004.
[10] 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.
[11] O. Hössjer, Incomplete generalized L-statistics, Annals of Statistics 24 (1996) 2631–2654.
[12] B.L. Jones, R. Zitikis, Empirical estimation of risk measures and related quantities, The North American Actuarial Journal 8 (2004) 114–118.
[13] A.M. Lindner, Stationarity, mixing, distributional properties and moments of GARCH(p, q) processes, in: T.G. Andersen, R.A. Davis, J.-P. Kreiß, T. Mikosch (Eds.), in: Handbook of Financial Time Series, Springer-Verlag, Berlin, 2008, pp. 41–69.
[14] D. Pollard, Convergence of Stochastic Processes, Springer-Verlag, New York, 1984.
[15] J.A. Reeds, On the definition of von Mises functionals, Research Report S-44, Department of Statistics, Harvard University, 1976.
[16] J. Ren, P.K. Sen, On Hadamard differentiability of extended statistical functional, Journal of Multivariate Analysis 39 (1991) 30–43.
[17] J. Ren, P.K. Sen, Hadamard differentiability on $D[0, 1]^p$, Journal of Multivariate Analysis 55 (1995) 14–28.
[18] J. Ren, M. Zhou, L-estimators for doubly censored data, Journal of Nonparametric Statistics 8 (1997) 1–20.
[19] O. Scaillet, Nonparametric estimation and sensitivity analysis of expected shortfall, Mathematical Finance 14 (2004) 115–129.
[20] P.K. Sen, J. Singer, Large Sample Methods in Statistics, Chapman and Hall, New York, 1993.
[21] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[22] R.J. Serfling, Generalized L-, M- and R-statistics, Annals of Statistics 12 (1984) 76–86.
[23] Q.-J. Shao, H. Yu, Weak convergence for weighted empirical processes of dependent sequences, Annals of Probability 24 (1996) 2098–2127.
[24] S.J. Sheather, J.S. Marron, Kernel quantile estimators, Journal of American Statistical Association 85 (1990) 410–416.
[25] G.R. Shorack, J.A. Wellner, Empirical Processes with Applications to Statistics, Wiley, New York, 1986.
[26] S. Stigler, Linear functions of order statistics with smooth weight functions, Annals of Statistics 2 (1974) 676–693.
[27] A.W. van der Vaart, Weak convergence of smoothed empirical processes, Scandinavian Journal of Statistics 21 (1994) 501–504.
[28] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, New York, 1996.
[29] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 1998.
[30] S.S. Wang, Premium calculation by transforming the layer premium density, ASTIN Bulletin 26 (1996) 71–92.
[31] S.S. Wang, J. Dhaene, Comontonicity, correlation order and premium principles, Insurance: Mathematics and Economics 22 (1998) 235–242.
[32] W.B. Wu, Empirical processes of stationary sequences, Statistica Sinica 18 (2008) 313–333.
[33] H. Yamato, Uniform convergence of an estimator of a distribution function, Bulletin of Mathematical Statistics 15 (1972) 69–78.
[34] J.E. Yukich, Weak convergence of smoothed empirical processes, Scandinavian Journal of Statistics 19 (1992) 271–279.
[35] H. Zähle, Rates of almost sure convergence of plug-in estimates for distortion risk measures, Metrika, 2010, in press (doi:10.1007/s00184-010-0302-z).