# A general approach to the joint asymptotic analysis of statistics from sub-samples

Stanislav Volgushev and Xiaofeng Shao*

May 27, 2013

## Abstract

In time series analysis, statistics based on collections of estimators computed from subsamples play a crucial role in an increasing variety of important applications. Proving results about the joint asymptotic distribution of such statistics is challenging since it typically involves a nontrivial verification of technical conditions and tedious case-by-case asymptotic analysis. In this paper, we provide a novel technique that allows to circumvent those problems in a general setting. Our approach consists of two major steps: a probabilistic part which is mainly concerned with weak convergence of sequential empirical processes, and an analytic part providing general ways to extend this weak convergence to functionals of the sequential empirical process. Our theory provides a unified treatment of asymptotic distributions for a large class of statistics, including recently proposed self-normalized statistics and sub-sampling based p-values. In addition, we comment on the consistency of bootstrap procedures and obtain general results on compact differentiability of certain mappings that seem to be of independent interest.

Keywords and Phrases: Empirical processes, sub-sampling, self-normalization, change point, weak convergence, Time series, compact differentiability

AMS Subject Classification: Primary 62E20, 62M10, 62G09, 62G15. Secondary 62G20.

## 1 Introduction and Motivation

In time series analysis, a large class of statistics can be expressed as smooth functions of estimators computed on consecutive portions (i.e., subsamples) of data. Since time series observations are naturally ordered by time, the use of such statistics has been a common theme in time series inference and examples

are abundant in areas such as sequential monitoring [Chu and White (1995), Aue and Reimherr (2009)], retrospective change point detection [Csörgö and Horváth (1997), Perron (2006)] and subsampling-based inference [Politis and Romano (1994), Politis et al. (1999)], among others. More recent examples include the self-normalized (SN, hereafter) statistics [Shao (2010a)], a new SN-based test statistic for change point detection [Shao and Zhang (2010)] and the p-value of the subsampling-based inference under the fixed-b asymptotics [Shao and Politis (2013)]. To obtain the asymptotic distributions of statistics of such kind, a traditional approach is to express the estimator as a sum of three parts, including the parameter, an average of influence functions, and a remainder term, followed by certain assumptions that ensure asymptotic negligibility of remainder terms and a routine analysis of the leading term which is of linear form. For many statistics of practical interest, theoretical analysis based on this approach can be quite challenging and tedious. In particular verifying the negligibility of remainder terms can be technically involved, since it requires a careful case-by-case study. The situation is further complicated by the fact that in time series settings, the underlying data are dependent. The aim of the present paper is to provide a general approach which allows to easily obtain the asymptotic distribution of statistics based upon infinite collections subsample estimates without long and tedious arguments.

In statistical applications, many important statistics can be expressed as smooth [more precisely: compactly differentiable] functionals of simple quantities such as the empirical distribution function. The analysis of the asymptotic properties of such statistics in the non-sequential setting can be elegantly performed in two distinct steps: an analytic part which consists in establishing the smoothness of the functional and a probabilistic part that is concerned with the analysis of the underlying quantity. One of the many appealing features of such an approach lies in the fact that the analytic properties need to be established only once. Moreover, quantities such as the empirical distribution function are often rather well analyzed for a wide range of data types. This approach has been successfully applied to the analysis of quantiles [Doss and Gill (1992)], survival data [Gill and Johansen (1990)], copulas and scalar measures of dependence [Fermanian et al. (2004); Bücher and Volgushev (2011)] and to the setting of dependent data.

A slightly more formal description of the situation above is as follows. Assume that we have a collection of estimators, say $(\hat{x}_{n,\kappa})_{\kappa \in K}$ of a quantity $x$. A classical example of such a collection is given by estimators computed from various fractions of the sample $X_1, ..., X_n$. For illustration purposes, assume that $x$ is the distribution function, $\hat{x}_{n,\kappa}$ denotes the empirical distribution function computed from $X_1, ..., X_{\lfloor n\kappa \rfloor + 1}$ and $K = [0, 1]$. Also, assume that the parameter of interest, say $\theta$, can be expressed as $\phi(x)$ where $\phi$ denotes some functional. For example, it is possible to express the copula as a functional of the cumulative distribution function. If the map $\phi$ is compactly differentiable, the asymptotic distribution of a suitably normalized version of $\phi(\hat{x}_{n,\kappa})$ for fixed $\kappa$ can be derived from a corresponding result for $\hat{x}_{n,\kappa}$. More precisely, denoting by $\alpha_n$ a sequence diverging to infinity and by $w(\kappa)$ a weight function, weak convergence of $\alpha_n w(\kappa)(\hat{x}_{n,\kappa} - x)$ in a suitable function space implies weak convergence of $\alpha_n w(\kappa)(\phi(\hat{x}_{n,\kappa}) - \phi(x))$ *for a finite collection of fixed values of $\kappa$*. However, in many important applications the *joint* weak convergence of the whole *collection* $\mathbb{V}_n := \left( \alpha_n w(\kappa)(\phi(\hat{x}_{n,\kappa}) - \phi(x)) \right)_{\kappa \in K}$ in a suitable functional sense is required. For the purpose of illustration, consider the following simple example.

**Example 1.1.** For the sake of concreteness, assume that we observe data, say $(Y_i = T_i \wedge C_i, \delta_i = I\{Y_i = T_i\})_{i=1,\dots,n}$ from a censored time series [here, $T_i$ denote survival times, $C_i$ censoring times and $\delta_i$ denote censoring indicators] and want to test if there is a change in the location parameter of the marginal distribution $F_i$ of $T_i$. A general way to quantify the location of censored observations, that is well-defined even under heavy censoring, is provided by the median. Typical test statistics for the null hypothesis of a constant median are based on comparing the medians of the Kaplan-Meier estimators which are computed from portions of the data. For simplicity, assume that the estimator $\hat{m}_\kappa$ with $\kappa \in [0,1]$ is based on the data $(Y_i, \delta_i)_{i=1,\dots,\lfloor n\kappa \rfloor \vee 1}$. A simple test statistic for the null hypothesis of a constant median is given by $\sup_{\kappa \in [0,1]} w(\kappa)|\hat{m}_\kappa - \hat{m}_1|$ with $w$ denoting a suitable weighting function. In order to derive the null distribution of our test statistic, we would typically establish a process convergence result for $\sqrt{n} w(\kappa)(\hat{m}_\kappa - \hat{m}_1)$ viewed as element in the space $D[0,1]$ and apply the continuous mapping theorem. Classical results on compact differentiability [see Example 2.2 and 2.3] imply that the median of the Kaplan-Meier estimator can be represented as a compactly differentiable functional of the two empirical (sub-)distribution functions $\hat{H}_{0,\lfloor n\kappa \rfloor}(y) := \lfloor n\kappa \rfloor^{-1} \sum_{i=1}^{\lfloor n\kappa \rfloor} \delta_i I\{Y_i \le y\}$ and $\hat{F}_{Y,\lfloor n\kappa \rfloor}(y) := \lfloor n\kappa \rfloor^{-1} \sum_{i=1}^{\lfloor n\kappa \rfloor} I\{Y_i \le y\}$. If we want to apply the classical delta-method to derive the process asymptotics of $\sqrt{n} w(\kappa)(\hat{m}_\kappa - \hat{m}_1)$ we are faced with two problems: first, we need process convergence of suitably normalized versions of $\hat{H}_{0,\lfloor n\kappa \rfloor}(y) - \mathbb{E}[H_{0,\lfloor n\kappa \rfloor}(y)]$ and $\hat{F}_{Y,\lfloor n\kappa \rfloor}(y) - \mathbb{E}[F_{Y,\lfloor n\kappa \rfloor}(y)]$. Second, as we shall argue below, the classical delta method does *not* provide results on weak convergence of the quantity $\sqrt{n} w(\kappa)(\hat{m}_\kappa - \hat{m}_1)$ as a *process* indexed in $\kappa$.

Returning to a more general setting, we can say that the classical delta method and a large collection of results on the behavior of general empirical processes allow to establish weak convergence results for a wide class of statistics as long as we consider a fixed, finite collection of values $\kappa$. Informally, we call this the 'non-sequential' case. However, the tools available to date do not allow the same conclusion when we are interested in collections of sub-samples, or, stated informally, in the 'sequential' case. The fundamental aim of the present article is thus to provide general ways of importing the tools mentioned above from the 'non-sequential' into the 'sequential' setting.

For example, let us consider what we would need to apply a delta method in the 'sequential' case if only compact differentiability of the map $\phi$ in the 'non-sequential' case is available. Essentially, such an approach would require us to show compact differentiability of the map

$$\Phi : (h_\kappa)_{\kappa \in K} \mapsto \left( w(\kappa)\phi\left(\frac{h_\kappa}{w(\kappa)}\right) \right)_{\kappa \in K}$$

viewed as a map between suitable metric spaces since we can write

$$\mathbb{V}_n = \alpha_n \left( \Phi\left( (w(\kappa)\hat{x}_{n,\kappa})_{\kappa \in K} \right) - \Phi\left( (w(\kappa)\hat{x}_{n,\kappa})_{\kappa \in K} \right) \right).$$

Given the fact that a large amount of important maps $\phi$ that are known to be compactly differentiable, we would like to make use of this information in the sequential setting. A natural question to ask thus is: given compact differentiability of $\phi$, what can we say about compact differentiability of $\Phi$? As we shall see in Section 2.1, such an implication does not hold in full generality, see in particular Example 2.5 and

the discussion preceding it. At the same time, we obtain a positive result if we additionally assume that the map $\phi$ possesses certain boundedness properties. Additionally, even when compact differentiability of $\Phi$ fails, there still are many relevant settings where additional arguments can be applied to obtain the desired weak convergence of $\mathbb{V}_n$. In fact, in Section 2.1 we show that, given weak convergence of $\mathbb{Y}_n := \left(\alpha_n w(\kappa)(\hat{x}_{n,\kappa} - x)\right)_{\kappa \in K}$, we can derive properties of $\mathbb{V}_n$ in a very general setup. Additionally, some general results on compact differentiability that seem to be of independent interest can be found in Section 2.3.

Another fundamental question that needs to be taken care of before we can apply the functional delta method is the weak convergence of the process $\mathbb{Y}_n$. In fact, results on weak convergence of $\mathbb{Y}_n$ in settings where the data $X_1, ..., X_n$ are allowed to be dependent are limited. A summary of available results as well as new insights providing considerable extensions of those findings are collected in Section 2.2.

Finally, in Section 3, we illustrate how the general results presented in Section 2 can be applied to obtain new insights regarding the properties of recently proposed methods including self-normalization and generalizations thereof [Section 3.1], fixed-b corrections for sub-sampling methods [Section 3.2], and SN-based testing procedures for change-points [Section 3.3]. Some comments on the applicability of our results to bootstrap methods are also provided.

## 2 General results

We begin by introducing some relevant notation. For arbitrary sets $\mathcal{F}_1, ..., \mathcal{F}_J, K$ define the vector space

$$\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K) := \left\{ (H_{1,t}, ..., H_{J,t})_{t \in K} \,\middle|\, H_{j,t} \in \ell^\infty(\mathcal{F}_j) \forall j, t \quad \sup_t \sup_j \sup_{f \in \mathcal{F}_j} |H_{j,t}(f)| < \infty \right\}$$

with norm

$$\|(H_{1,t}, ..., H_{J,t})_{t \in K}\|_\mathcal{L} := \sup_t \sup_j \sup_{f \in \mathcal{F}_j} |H_{j,t}(f)|.$$

Note that $\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K)$ can be identified with $\ell^\infty(K \times \mathcal{F}_1) \times ... \times \ell^\infty(K \times \mathcal{F}_J)$ by considering the relation

$$(H_{1,t}, ..., H_{J,t})_{t \in K} \in \mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K) \quad \leftrightarrow \quad \Big((t,f) \mapsto H_{1,t}(f), ..., (t,f) \mapsto H_{J,t}(f)\Big).$$

By the definition of $\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K)$, we have $\sup_t \sup_f |H_{j,t}(f)| < \infty$ for all $j = 1, ..., J$ so that the maps $(t_j, f_j) \mapsto H_{j,t_j}(f_j)$ are indeed bounded and thus elements of $\ell^\infty(K \times \mathcal{F}_j)$. In particular, if the product space $\ell^\infty(K \times \mathcal{F}_1) \times ... \times \ell^\infty(K \times \mathcal{F}_J)$ is equipped with the maximum norm $\|(x_1, ..., x_J)\|_{\max} := \max_j \|x_j\|_\infty$ induced by the supremum norms on its components, the identification given above is an isometry, that is

$$\|(H_{1,t}, ..., H_{J,t})_{t \in K}\|_\mathcal{L} = \left\| \Big((t,f) \mapsto H_{1,t}(f), ..., (t,f) \mapsto H_{J,t}(f)\Big) \right\|_{\max}.$$

Weak convergence in $\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K)$ is henceforth understood as weak convergence in the Hoffmann-Jørgensen sense in the space $\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K)$ as a subspace of $\ell^\infty(K \times \mathcal{F}_1) \times ... \times \ell^\infty(K \times \mathcal{F}_J)$ [see Van der Vaart and Wellner (1996), Chapters 1.4 and 1.5 for more details].

**Remark 2.1.** In most situations, the sets $\mathcal{F}_1, ..., \mathcal{F}_J$ can be viewed as subsets of $\mathbb{R}^d$. For example, the empirical distribution function $(n^{-1} \sum I\{X_i \leq y\})_{y \in \mathbb{R}^d}$ of a sample of $d$-dimensional random variables $X_1, ..., X_n$ is naturally indexed by the set $\mathbb{R}^d$. Another approach that fits nicely into the empirical process setting and will play a central role in Section 2.2, is to consider classes of functions $\{x \mapsto f(x) | f \in \mathcal{F}_j\}$. In this setting, the empirical process can be elegantly written as $\left(n^{-1} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X_i)\right)_{f \in \mathcal{F}_j}$, see Van der Vaart and Wellner (1996) for examples. For example, the empirical distribution function can also be viewed as element of $\ell^\infty(\mathcal{F})$ with $\mathcal{F}$ denoting the collection of indicators of rectangles, that is $\mathcal{F} = \{x \mapsto I\{x \leq y\} | y \in \mathbb{R}^d\}$. By identifying the function $x \mapsto I\{x \leq y\}$ with the point $y \in \mathbb{R}^d$ we obtain a way to index $\mathcal{F}$ by $\mathbb{R}^d$ and vice versa. In most of the following theoretical developments, the form of $\mathcal{F}_j$ will be arbitrary unless explicitly specified otherwise.

As discussed previously, the asymptotic analysis of statistics based on the process $\mathbb{V}_n$ can be performed by considering two distinct questions: the stochastic properties of $\mathbb{Y}_n$ and the analytic properties of the map $\phi$. Both questions will be addressed in this section in a general setting. In section 2.1, we present our analytic considerations. An overview of existing results regarding the stochastic part as well as their extension will be considered in section 2.2. Finally, some general results on compact differentiability that seem to be of independent interest are provided in section 2.3.

## 2.1 Analytic considerations

This section is primarily concerned with the following questions: given a collection of estimators $(\hat{y}_{n,s,t})_{(s,t) \in K}$ such that for fixed $(s,t) \in K$ each $\hat{y}_{n,s,t}$ is an element of $\ell^\infty(\mathcal{F}_1) \times ... \times \ell^\infty(\mathcal{F}_J)$ with $K \subset \Delta := \{(s,t) \in [0,1]^2 | s \leq t\}$, a smooth (in a suitable sense) map $\phi : \ell^\infty(\mathcal{F}_1) \times ... \times \ell^\infty(\mathcal{F}_J) \to \ell^\infty(\mathcal{G}_1) \times ... \times \ell^\infty(\mathcal{G}_L)$, and weak convergence of the process [$\alpha_n$ denotes some deterministic sequence diverging to infinity]

$$\mathbb{Y}_n(s,t,f_1, ..., f_J) := (t-s)\alpha_n(\hat{y}_{n,s,t}(f_1, ..., f_J) - x(f_1, ..., f_J))$$

viewed as element of $\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K)$, what can we say about weak convergence of $\mathbb{V}_n$ where

$$\mathbb{V}_n(s,t,g_1, ..., g_L) := (t-s)\alpha_n(\phi(\hat{y}_{n,s,t})(g_1, ..., g_L) - \phi(x)(g_1, ..., g_L))$$

as element of $\mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K)$? And what can we say about bootstrap validity for $\mathbb{V}_n$ given a valid bootstrap procedure for $\mathbb{Y}_n$?

For instance, consider the situation where we have a sample $X_1, ..., X_n$. Assume that the quantity $x$ can be represented as $x = \left((\mathbb{E}[f(X)])_{f \in \mathcal{F}_1}, ...., (\mathbb{E}[f(X)])_{f \in \mathcal{F}_J}\right)$ for some classes of functions $\mathcal{F}_1, ..., \mathcal{F}_J$, see the examples below. A prime example for the quantity $\hat{y}_{n,s,t}$ is given by the estimator computed from the sub-sample $X_{\lfloor ns \rfloor + 1}, ..., X_{\lfloor nt \rfloor}$, that is

$$\hat{y}_{n,s,t} := \left(\left(\frac{1}{\lfloor nt \rfloor - \lfloor ns \rfloor} \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} f(X_i)\right)_{f \in \mathcal{F}_1}, ..., \left(\frac{1}{\lfloor nt \rfloor - \lfloor ns \rfloor} \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} f(X_i)\right)_{f \in \mathcal{F}_J}\right) \tag{1}$$

where the empty sum is defined as zero and we set $'0/0 = 0'$ to take care of the case $\lfloor ns \rfloor = \lfloor nt \rfloor$.

Regarding the smoothness of $\phi$, we impose the following condition

(C) The map

$$\phi : \ell^\infty(\mathcal{F}_1) \times ... \times \ell^\infty(\mathcal{F}_J) \supset D_\phi \to R_\phi \subset \ell^\infty(\mathcal{G}_1) \times ... \times \ell^\infty(\mathcal{G}_L).$$

is compactly differentiable at $x$ tangentially to $V \subset \ell^\infty(\mathcal{F}_1) \times ... \times \ell^\infty(\mathcal{F}_J)$. Additionally, $0 \in V$ as well as $f \in V \Rightarrow cf \in V$ for all $c > 0$.

In the 'classical' setting, compact differentiability is known to provide a good balance between strength of the differentiability concept that is needed for establishing a general functional delta method and the number of statistically relevant functionals that can actually be shown to be compactly differentiable. See Van der Vaart and Wellner (1996), Chapter 3.9 for a more detailed discussion of this topic. Two particular examples are discussed below. Of course, there exists a vast collection of further examples [copulas, dependence measures, M- and L-estimators to name just a few] that are equally important but not discussed here because of space considerations. For a more detailed list we refer the interested reader to Chapter 3.9 in Van der Vaart and Wellner (1996) and the recent paper by Gao and Zhao (2011).

**Example 2.2. Empirical quantiles**
Consider the class of functions $\mathcal{F} : \{y \mapsto I\{y \leq t\} | t \in \mathbb{R}\}$. In this case, $\hat{y}_{n,s,t}$ is simply the empirical distribution function of the sub-sample $X_{\lfloor ns \rfloor + 1}, ..., X_{\lfloor nt \rfloor}$. Consider the quantile map $\phi : F \mapsto (F^{-1}(\tau))_{\tau \in S}$ for some $S \subset (0, 1)$ which now corresponds to $\mathcal{G}_1$. Applying this map to $\hat{y}_{n,s,t}$ yields collections of empirical quantiles of the sub-samples $X_{\lfloor ns \rfloor + 1}, ..., X_{\lfloor nt \rfloor}$. Compact differentiability of the quantile map can be established under appropriate conditions, see Lemma 3.9.23 in Van der Vaart and Wellner (1996).

**Example 2.3. Kaplan-Meier estimator**
Assume that we have right-censored observations of the form $(Y_i, \delta_i)_{i=1,...,n}$. It is a well-known fact that the Kaplan-Meier estimator $\hat{F}_{KM}$ [Kaplan and Meier (1958)], viewed as a map into the set of distribution functions on $[0, V]$ for a suitable $V < \infty$, is a compactly differentiable functional of the two functions

$$\hat{F}_1(t) := \frac{1}{n} \sum_i \delta_i I\{Y_i \leq t\}, \quad \hat{F}_Y(t) := \frac{1}{n} \sum_i I\{Y_i \leq t\},$$

see Chapter 3.9 in Van der Vaart and Wellner (1996). This suggests to consider the classes of functions

$$\mathcal{F}_1 := \Big\{ (y, \delta) \mapsto \delta I\{y \leq t\} \Big| t \in \mathbb{R} \Big\}, \quad \mathcal{F}_2 := \Big\{ (y, \delta) \mapsto I\{y \leq t\} \Big| t \in \mathbb{R} \Big\}.$$

Combining this with the quantile mapping [see Example 2.2] easily allows to consider quantiles of the Kaplan-Meier estimator.

Regarding the process $\mathbb{Y}_n$, we need the following assumption,

(W) Assume that

$$\mathbb{Y}_n \rightsquigarrow \mathbb{Y} \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; \Delta)$$

where

$$\mathbb{Y}(s, t, f_1, ..., f_J) := (\mathbb{Y}_1(s, t, f_1), ..., \mathbb{Y}_J(s, t, f_J))$$

and $\mathbb{Y}_j, j = 1, ..., J$ are centered, Borel measurable processes.

**Remark 2.4.** A detailed discussion of condition (W) for estimators $\hat{y}_{n,s,t}$ of the form (1) is provided in the next section. However, there are interesting examples that go beyond the framework described above. For example, the classical empirical copula process [see Rüschendorf (1976)] is of the form

$$\mathbb{C}_n^\circ(s, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor} \left( I\{X_i \leq F_n^-(u)\} - C(u) \right)$$

where $F_n^-(u) := (F_{n1}^-(u_1), ..., F_{nd}^-(u_d))$ denotes the vector of the generalized inverses of the marginal empirical distribution functions $F_{nj}(y) = n^{-1} \sum_i I\{X_{ij} \leq y\}$ and $C$ is the copula of the distribution of $X$. Note that $F_n^-(u)$ depends on all the data regardless of the value of $s$. The process $\mathbb{C}_n^\circ(s, u)$ can be coerced into the general framework of this section by considering the collection of estimators $\frac{1}{\lfloor ns \rfloor} \sum_{i=1}^{\lfloor ns \rfloor} I\{X_i \leq F_n^-(u)\}$ indexed by $\mathcal{F}_1 := [0, 1]^d$. Weak convergence of the process $\mathbb{C}_n^\circ$ under weak assumptions on the copula with possibly dependent data was recently established by Bücher and Volgushev (2011).

The limit $\mathbb{Y}$ in assumption (W) needs to satisfy certain technical conditions that are not very restrictive as we shall demonstrate later.

(A1) Assume that $\sup_{|s-s'|+|t-t'| \leq \delta} \sup_j \sup_{f_j \in \mathcal{F}_j} |\mathbb{Y}_j(s, t, f_j) - \mathbb{Y}_j(s', t', f_j)| = o_P(1)$ as $\delta \to 0$.

(A2) Define the set

$$U_K := \left\{ (h_{s,t})_{(s,t) \in K} : h_{s,t} \in V \ \forall \ (s, t) \in K, \ \sup_{(s,t) \in K} \|h_{s,t}\| < \infty \right\}.$$

Assume that the sample paths of $\mathbb{Y}$ are in $U_K$ with probability one.

Condition (A2) is non-restrictive in the sense that it is needed to apply the functional delta method to $\mathbb{V}_n(s, t, \cdot)$ for each fixed $(s, t)$. Assumption (A1) is needed for the application of the general compact differentiability result in Section 2.3. As we shall discuss in the next section [see Remark 2.13], assumption (A1) is typically satisfied in a wide variety of practically relevant settings. Assumptions (W), (A1), (A2) are already sufficient to derive weak convergence of $\mathbb{V}_n$ if the set $K$ satisfies $\inf_{(s,t) \in K} |t - s| > 0$. Without this condition, (W), (A1), (A2) are not sufficient as the following example demonstrates.

**Example 2.5.** Consider the map $\phi$ that takes a distribution function to its median and let $K = \{0\} \times [0, 1] \subset \Delta$. Define $\mathcal{F}_1 := \{x \mapsto I\{x \leq y\} | y \in \mathbb{R}\}$ and identify the functions $x \mapsto I\{x \leq y\} \in \mathcal{F}_1$ with $y \in \mathbb{R}$. Assume that $\hat{y}_{n,0,t}(y) = \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} I\{X_i \leq y\}$. Consider a triangular scheme of data that is of the form $X_{jn} = n$ for $1 \leq j < n^{1/3}$ and $X_{jn} \sim U[0, 1]$ i.i.d. for $n^{1/3} \leq j \leq n$. Elementary calculations show that $\mathbb{Y}_n(0, \cdot, \cdot)$ converges weakly to the Kiefer-Müller process $\mathbb{K}$ with covariance $\text{Cov}(\mathbb{K}(t, y), \mathbb{K}(t', y')) = \min(t, t')(\min(y, y') - yy')$. On the other hand, setting $t = n^{-3/4}$ we have almost surely

$$\mathbb{V}_n((0, n^{-3/4})) = n^{1/2}t(\phi(\hat{y}_{n,0,t}) - \phi(x)) = n^{-1/4}(n - 1/2) \to \infty,$$

and thus weak convergence of $\mathbb{V}_n$ can not hold.

The underlying problem in the above example is that due to the weighting with $t - s$, weak convergence of $\alpha_n(t - s)(\hat{y}_{n,t,s} - x)$ is not informative about $\hat{y}_{n,t,s}$ for values of $t - s$ that can be arbitrarily close to zero. Additional assumptions are needed to exclude this kind of behavior presented in the above example. It turns out that for this purpose the following condition is sufficient. As we shall discuss later, there are many situations where it is easily satisfied.

(A3) For any $k_n \to 0$ we have $\sup_{(s,t)\in K, |t-s|\leq k_n}(t - s)\|\phi(\hat{y}_{n,s,t})\| = o_P^*(1)$ where the asterisk denotes outer probability.

**Remark 2.6.** Note that condition (A3) is automatically satisfied if $\sup_{(s,t)\in K}\|\phi(\hat{y}_{n,s,t})\| = O_P^*(1)$. This is trivially true for uniformly bounded maps $\phi$, which includes many interesting examples such as copulas, dependence measures or the Kaplan-Meier estimator (which per definition is a distribution function). Moreover for specific sets $K$, further conditions implying (A3) can be derived. See Remark 2.15 in Section 2.2 for further details.

Now we are ready to state the first main result of this section.

**Theorem 2.7.** *For any compact $K \subset \Delta$, with $\inf_{(s,t)\in K}|t - s| \geq a > 0$ conditions (C), (W), (A1) and (A2) imply $\mathbb{V}_n \rightsquigarrow \mathbb{V}$ in $\mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K)$ where*

$$\mathbb{V}((s,t), g_1, ..., g_L) := \left(\phi_x' \mathbb{Y}(s, t, \cdot)\right)(g_1, ..., g_L).$$

*If additionally (A3) holds, the assumption $\inf_{(s,t)\in K}|t - s| \geq a > 0$ can be dropped.*

**Remark 2.8.** Although assumption (A3) often holds, there are situations where verifying it can be very tedious or requires additional assumptions on the underlying data structure. For example, consider the setting where $K = \Delta$ and $\phi$ denotes the map that takes a distribution function to its median. In that case, assumption (A3) would require that $\frac{1}{n}\max_{i=1,...,n}|X_i| = o_P(1)$ since the median of one observation is the observation itself. Effectively, this places moment assumptions on $X$ that are not needed for the median from large samples to be well-behaved. A closer look at the proofs reveals that for any $\gamma \in (0,1)$ the following modified version of the process $\mathbb{V}_n$

$$\tilde{\mathbb{V}}_n := \left((t - s)I\{t - s \geq \alpha_n^{-\gamma}\}\alpha_n(\phi(\hat{y}_{n,s,t}) - \phi(x))\right)_{(s,t)\in K}$$

converges to the same limit $\mathbb{V}$ without assumption (A3) or the condition $\inf_{(s,t)\in K}|t - s| \geq a > 0$. In the applications discussed in Section 3, the modification above essentially amounts to not using information from extremely small sub-samples. As the discussion above indicates, for certain sets $K$ this can be viewed as a robustification.

**Remark 2.9.** A closer look at the proof of the above result shows that the special structure of $\Delta$ does not play a crucial role. In fact, the same approach yields a more general result. Let $(K, d_K)$ denote a general compact metric space. Assume that

$$\mathbb{Y}_n(\kappa, f_1, ..., f_J) = w(\kappa)\alpha_n\Big(\hat{y}_{n,\kappa}(f_1, ..., f_J) - x(f_1, ..., f_J)\Big)$$

and that $\mathbb{Y}_n \rightsquigarrow \mathbb{Y} \in \mathcal{L}(\mathcal{F}_1, ..., \mathcal{F}_J; K)$ with a centered process $\mathbb{Y}$ with $w(\cdot)$ denoting a bounded weight function. If additionally $\sup_{d_K(\kappa,\kappa')\leq\delta}\sup_j\sup_{f_j\in\mathcal{F}_j}|\mathbb{Y}_j(\kappa, f_j) - \mathbb{Y}_j(\kappa', f_j)| = o_P(1)$ and if the sample paths of $\mathbb{Y}$, are, with probability one in the set

$$\tilde{U}_K := \Big\{(h_\kappa)_{\kappa\in K} : h_\kappa \in V \ \forall \ \kappa \in K, \sup_{\kappa\in K}\|h_\kappa\| < \infty\Big\}$$

it follows that with $\gamma \in (0, 1)$ arbitrary

$$\Big(w(\kappa)I\{w(\kappa) \geq \alpha_n^{-\gamma}\}\alpha_n(\phi(\hat{y}_{n,\kappa}) - \phi(x))\Big)_{\kappa\in K} \rightsquigarrow \Big(\phi'_x\mathbb{Y}(\kappa, \cdot)\Big)_{\kappa\in K} \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K)$$

as long as $\inf_{\kappa\in K}|w(\kappa)| > 0$. If additionally a modified version of condition (A3) holds, i.e. if for any $k_n \to 0$ we have $\sup_{\kappa\in K,|w(\kappa)|\leq k_n}w(\kappa)\|\phi(\hat{y}_{n,\kappa})\| = o_P^*(1)$, the weak convergence above holds without the assumption $\inf_{\kappa\in K}|w(\kappa)| > 0$.

Next, we discuss bootstrap procedures. In particular, consider the following bootstrap version of the quantity $\hat{y}_{n,s,t}$ defined in (1)

$$\hat{y}_{n,s,t}^b(s, t, f_1, ..., f_J) := \Big(\frac{1}{\lfloor nt\rfloor - \lfloor ns\rfloor}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor}M_if(X_i), ..., \frac{1}{\lfloor nt\rfloor - \lfloor ns\rfloor}\sum_{i=\lfloor ns\rfloor+1}^{\lfloor nt\rfloor}M_if(X_i)\Big). \quad (2)$$

with $M_1, ..., M_n$ denoting random variables independent of the original sample $X_1, ..., X_n$. The corresponding bootstrap version of the process $\mathbb{G}_n$ is given by

$$\mathbb{Y}_n^b := (t - s)\alpha_n(\hat{y}_{n,s,t}^b - \hat{y}_{n,s,t}) =: (\mathbb{Y}_{n,1}^b, ..., \mathbb{Y}_{n,J}^b) \quad (3)$$

Under suitable assumptions on the data and random variables $M_1, ..., M_n$ a conditional version of assumption (W) holds. Specifically, assume that

(WB) $\mathbb{Y}_n^b$ weakly converges to $\mathbb{Y}$ conditionally on the data in probability, or

$$\Big(\mathbb{Y}_{n,1}^b, ..., \mathbb{Y}_{n,J}^b\Big) \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} \Big(\mathbb{Y}_1, ..., \mathbb{Y}_J\Big) \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K).$$

Here, weak convergence conditional on the data in probability ($\overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}}$-convergence) is understood in the Hoffmann-Jørgensen sense as defined in Kosorok (2008), that is $\mathbb{Y}_n^b \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} \mathbb{Y}$ if and only if

(i) $\sup_{f\in\mathrm{BL}_1}\big|\mathbb{E}_Mf(\mathbb{Y}_n^b) - \mathbb{E}f(\mathbb{Y})\big| \to 0$ in outer probability,

(ii) $\mathbb{E}_Mf(\mathbb{Y}_n^b)^* - \mathbb{E}_Mf(\mathbb{Y}_n^b)_* \overset{\mathbb{P}}{\to} 0$ for all $f \in \mathrm{BL}_1$,

where $\mathrm{BL}_1$ denotes the set of all Lipschitz-continuous functions $f : \mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; K) \to \mathbb{R}$ that are uniformly bounded by 1 and have Lipschitz constants bounded by 1, and where the asterisks in (ii) denote measurable majorants (and minorants, respectively) with respect to the joint data $(X_1, ..., X_n, M_1, ..., M_n)$. Also, note that the map $(M_1, ..., M_n) \mapsto \mathbb{Y}_n^b$ is measurable conditionally on the original data $X_1, ..., X_n$ outer almost surely [for fixed $X_1, ..., X_n$, this mapping is Lipschitz-continuous] and thus we do not need to consider measurable majorants. Settings where results of this kind hold are discussed in the next section.

The classical delta method for the bootstrap [see e.g. Theorem 12.1 in Kosorok (2008)] asserts that for a map $\phi$ that is compactly differentiable at $x$ with derivative $\phi'_x$ and additionally satisfies suitable measurability conditions, we have

$$\alpha_n(t - s)(\phi(\hat{y}_{n,s,t}^b) - \phi(\hat{y}_{n,s,t})) \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_x \mathbb{G} \quad in \quad \ell^\infty(\mathcal{G}_1) \times ... \times \ell^\infty(\mathcal{G}_L)$$

for every fixed $(s, t)$. The next Theorem provides a generalization of this finding. More precisely, it states conditions that allow for a generalization of Theorem 2.7 to conditional weak convergence in $\mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K)$.

**Theorem 2.10.** *With the notation above, assume that (WB), (A1), (A2) and (C) hold. Then for any compact $K \subset \Delta$ with $\inf_{(s,t)\in K} |t - s| > 0$ we have*

$$\mathbb{V}_n^b := \alpha_n(t - s)(\phi(\hat{y}_{n,s,t}^b) - \phi(\hat{y}_{n,s,t})) \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_x \mathbb{Y} = \mathbb{V} \quad in \quad \mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K).$$

*If additionally (A3) holds and $\sup_{(s,t)\in K, |t-s|\leq k_n}(t - s)\|\phi(\hat{y}_{n,s,t}^b)\| = o_P^*(1)$, the convergence holds for arbitrary compact $K \subset \Delta$.*

**Remark 2.11.** Suitable modifications of the extensions discussed in Remark 2.8 and Remark 2.9 continue to hold in the bootstrap setting. More precisely, we can replace sets $K \subset \Delta$ by arbitrary compact sets and the weighting $t - s$ with arbitrary bounded weighting functions $w$, in which case the assumption $\inf_{(s,t)\in K} |t - s| > 0$ needs to be replaced by $\inf_{\kappa\in K} |w(\kappa)| > 0$. Also, conditional weak convergence of $\tilde{\mathbb{V}}_n^b = (w(\kappa)I\{|w(\kappa)\}| > \alpha_n^{-\gamma}\}\alpha_n(\phi(\hat{y}_{n,s,t}^b) - \phi(\hat{y}_{n,s,t}))_{\kappa\in K}$ holds without assumption (A3) and the condition $\sup_{(s,t)\in K, |t-s|\leq k_n}(t - s)\|\phi(\hat{y}_{n,s,t}^b)\| = o_P^*(1)$ used in the above theorem.

## 2.2 Probabilistic considerations

In this section, we focus our attention on the setting where $\mathbb{Y}_n$ has a specific structure that typically arises in applications. More precisely, consider the multi-parameter sequential empirical process

$$\mathbb{Y}_n := \left( (t - s)\alpha_n(\hat{y}_{n,s,t} - x) \right)_{(s,t)\in\Delta}$$

where $\Delta := \{(s, t) \in [0, 1]^2 | s \leq t\}$ and the quantity

$$\hat{y}_{n,s,t} := \left( \left( \frac{1}{\lfloor nt \rfloor - \lfloor ns \rfloor} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} f(X_i) \right)_{f\in\mathcal{F}_1}, ..., \left( \frac{1}{\lfloor nt \rfloor - \lfloor ns \rfloor} \sum_{i=\lfloor ns \rfloor+1}^{\lfloor nt \rfloor} f(X_i) \right)_{f\in\mathcal{F}_J} \right)$$

10

denotes an estimator for $x$ that is computed based on the sub-sample $X_{\lfloor ns \rfloor + 1}, ..., X_{\lfloor nt \rfloor}$. It turns out that conditions (W), (A1), (A2) in the previous section can be derived from simpler conditions that involve only a collection of 'classical' one-parameter sequential processes

$$\mathbb{G}_n(t, f_1, ..., f_J) := (\mathbb{G}_{n,1}(t, f_1), ..., \mathbb{G}_{n,J}(t, f_J))$$

where $\mathbb{G}_{n,j}(t, f) := t\alpha_n(\hat{x}_{n,t}^{(j)}(f) - x)$ and

$$\hat{x}_{n,t} := \left( \left( \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} f(X_i) \right)_{f \in \mathcal{F}_1}, ..., \left( \frac{1}{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} f(X_i) \right)_{f \in \mathcal{F}_J} \right) =: (\hat{x}_{n,t}^{(1)}, ..., \hat{x}_{n,t}^{(J)}).$$

Consider the assumptions

(W') Assume that
$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; [0, 1])$$

where
$$\mathbb{G}(t, f_1, ..., f_J) := (\mathbb{G}_1(t, f_1), ..., \mathbb{G}_J(t, f_J))$$

and $\mathbb{G}_j, j = 1, ..., J$ are centered, Borel measurable processes.

(A1') Assume that $\sup_{|s-t| \le \delta} \sup_j \sup_{f_j \in \mathcal{F}_j} |\mathbb{G}_j(t, f_j) - \mathbb{G}_j(s, f_j)| = o(1)$ as $\delta \to 0$.

The conditions above turn out to be sufficient for (W) and (A1).

**Proposition 2.12.** *Under conditions (W') and (A1'), we have*

$$\mathbb{Y}_n \rightsquigarrow \mathbb{Y} \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; \Delta)$$

*where $\Delta = \{s, t \in [0, 1] : s \le t\}$ and*

$$\mathbb{Y}(s, t, f_1, ..., f_J) := \mathbb{G}(t, f_1, ..., f_J) - \mathbb{G}(s, f_1, ..., f_J).$$

*Moreover, $\mathbb{Y}$ satisfies assumption (A1).*

**Remark 2.13.** For many kinds of weakly dependent data [including, of course, the independent case], the process $\mathbb{G}$ is a vector of centered Gaussian processes with covariance of the form

$$\text{Cov}(\mathbb{G}(s, f_1, ..., f_J), \mathbb{G}(t, g_1, ..., g_J)) = (s \wedge t) K(f_1, ..., f_J, g_1, ..., g_J)$$

for some uniformly bounded covariance kernel $K$. In this case, assumption (A1') holds. To see this, note that under (A1') the process $\mathbb{G}$ has paths that are uniformly continuous with respect to the metric $\rho_2((t, f_1, ..., f_J), (t', f_1', ..., f_J')) := \mathbb{E}[(\mathbb{G}(t, f_1, ..., f_J) - \mathbb{G}(t', f_1', ..., f_J'))^2]$, see Example 1.5.10 in Van der Vaart and Wellner (1996). The discussion at the beginning of Example 1.5.10 in Van der Vaart and Wellner (1996) thus yields the desired result. The special structure of $\mathbb{Y}_n$ implies that its sample paths have the same property.

**Remark 2.14.** There are interesting cases where condition (A1') holds for limiting processes that are non-Gaussian. More precisely, defining $\mathcal{F}_1 = [-\infty, \infty]$, the results in Dehling and Taqqu (1989) imply weak convergence of the process $\mathbb{G}_n$ if the data $X_i$ exhibit long-range dependence. The limiting process, which can be non-Gaussian, is of the form $\mathbb{G}(t, y) = f(y)Z_m(t)$ with $f$ denoting a deterministic, uniformly bounded function and $Z_m$ a so-called m'th order Hermite-process [see Dehling and Taqqu (1989) for a definition]. In particular, the sample paths of this process are Hölder-continuous [see Maejima and Tudor (2007)] and thus assumptions (A1') and (W') hold.

**Remark 2.15.** Consider the special case $K = \{0\} \times [0, 1]$. In this case, assumption (A3) is satisfied as soon as $\hat{x}_{n,t}$ is of the form given in (1) with the data $X_1, X_2, \ldots$ stemming from a strictly stationary sequence and $\hat{x}_{n,1} \to x$ outer almost surely. To see this, note that under the assumptions discussed above we have $\sup_t \|\phi(\hat{x}_{n,t})\| = \max_{j=1,\ldots,n} \|\phi(\hat{x}_{j,1})\|$ and that by Lemma B.1 together with the continuous mapping theorem $\|\phi(\hat{x}_{n,1})\| \to \|\phi(x)\|$ outer almost surely. This in turn implies that $(\sup_{n\geq 1} \|\phi(\hat{x}_{n,1})\|)^*$ [the asterisk denoting a measurable majorant] is bounded in probability. For results implying almost sure convergence in a very general setting, see Adams and Nobel (2010) and the references cited therein.

For independent data, assumption (W') is known to hold as soon as the classes of functions $\mathcal{F}_1, \ldots, \mathcal{F}_J$ are Donsker [see Van der Vaart and Wellner (1996), Chapter 2.12.1]. For dependent data, much less is known. Available results are, to the best of our knowledge, limited to classes of functions of of the form $\mathcal{F}_1 = \{u \mapsto I\{u \leq y\}|y \in \mathbb{R}^d\}$ [the inequality is understood component-wise]. Here, results for $d > 1$ are derived by Sen (1974) and Rüschendorf (1974) under $\phi-$mixing and by Yoshihara (1975) and Inoue (2001) under strong mixing. Berkes et al. (2009) considered the case $d = 1$ under S-mixing, and derived a stronger result than weak convergence of the process. Finally, the paper by Dehling and Taqqu (1989) contains a similar result for the class of functions $\mathcal{F}_1 = \{u \mapsto I\{u \leq y\}|y \in \mathbb{R}\}$ and long-range dependent data. To the best of our knowledge, nothing is known for general classes of functions.

Note that by Lemma 1.4.3 in Van der Vaart and Wellner (1996), asymptotic tightness of $\mathbb{G}_n$ is equivalent to asymptotic tightness of $\mathbb{G}_{n,j}$ for all $j = 1, \ldots, J$. Thus, Problem 1.5.3 in the same reference implies that in order to obtain weak convergence of $\mathbb{G}_n$ to $\mathbb{G}$, we need to show that first $\mathbb{G}_{n,j}$ is asymptotically tight for all $j = 1, \ldots, J$ and second that the following condition holds

(F) For all finite collections $s_{i,j} \in [0, 1], i = 1, \ldots, N, j = 1, \ldots, J, f_{ij} \in \mathcal{F}_j, i = 1, \ldots, N, j = 1, \ldots, J$ the collection $(\mathbb{G}_{j,n}(s_{ij}, f_{ij}))_{j=1,..,J,i=1,...,N}$ converges weakly to $(\mathbb{G}_j(s_{ij}, f_{ij}))_{j=1,...,J,i=1,...,N}$ in the usual $\mathbb{R}^{NJ}$-dimensional sense.

There is a vast literature containing results that imply the finite-dimensional convergence (F), see Dehling et al. (2002) and the references cited therein for an overview. Criteria establishing asymptotic tightness of the processes $\mathbb{G}_{n,j}$ for dependent data on the other hand are not as widely available, and one general result along those lines is provided below. This result is of independent interest. In particular, it can be used to verify condition (W') in a number of settings that have not been considered before.

**Theorem 2.16.** *Assume that the process $\mathbb{G}_n$ is of the form $\mathbb{G}_n = t\alpha_n(\hat{x}_{n,t} - x)$ where $\hat{x}_{n,t}$ is defined in (1) and the data $X_1, X_2, \ldots$ come from a strictly stationary sequence. Assume that for each $j = 1, \ldots, J$*

*there exists a semi-metric $\rho_j$ on $\mathcal{F}_j$ which makes $\mathcal{F}_j$ totally bounded, and for each $j = 1, ..., J$ we have $\sup_{f \in \mathcal{F}_j} \mathbb{E}|f|^q < \infty$. Define $\mathcal{F}_{j,\delta} := \{f - g | f, g \in \mathcal{F}_j, \rho_j(f, g) \leq \delta\}$. Assume that the process $\mathbb{G}_{n,j}(1, \cdot)$ satisfies for some $q > 2$ and $j = 1, ..., J$*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \mathbb{E}^* \|\mathbb{G}_{n,j}(1, \cdot)\|_{\mathcal{F}_{j,\delta}}^q = 0 \tag{4}$$

*[remember that the asterisk denotes outer expectation], that*

$$\max_{j=1,...,J} \sup_{n \in \mathbb{N}} \mathbb{E}^* \|\mathbb{G}_{n,j}(1, f)\|^q < \infty \quad \forall f \in \mathcal{F}. \tag{5}$$

*and that for every $j$ the class of functions $\mathcal{F}_j$ has envelope $F_j$ which has finite $q$'th moment. Let condition (F) hold. Then $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\mathcal{L}^\infty(\mathcal{F}_1, ..., \mathcal{F}_J; [0, 1])$.*

Condition (4) has been established by Andrews and Pollard (1994) for strongly mixing data, and inequality (3.1) in Andrews and Pollard (1994) reveals that (5) holds under the same assumption. Moreover Hagemann (2012) established (4) for stationary sequences with geometric moment contraction properties [see Wu and Shao (2004)], and the results in his appendix show that again (5) holds under the same assumptions.

Next, consider bootstrap procedures. In the case of independent data, a mild assumption on the multipliers $M_i$ suffices. More precisely, assuming that $M_i$ are i.i.d., independent of the data $X_i$, and that $\int \sqrt{P(|M_1| > u)} du$ is finite [which follows if $M_1$ has finite moment of order $2 + \varepsilon$], the classes of functions $\mathcal{F}_1, ..., \mathcal{F}_J$ being Donsker [see Van der Vaart and Wellner (1996), page 81 for a definition of this property] implies (WB). To see this, note that by arguments similar to the ones given in the proof of Proposition 2.12 it suffices to derive (WB) for the set $K = \{0\} \times [0, 1]$. To do so, apply Lemma B.3 in the appendix where the approximating mappings $A_i$ and $A_{i,n}^b$ are defined through projections on piecewise constant functions, see the arguments in the proof of Theorem 1.5.6 in Van der Vaart and Wellner (1996). Then assumption (i) of Lemma B.3 corresponds to conditional finite-dimensional convergence which can be established by arguments similar to those given in Lemma 2.9.5 in Van der Vaart and Wellner (1996). Condition (ii) corresponds to tightness of the limit process $\mathbb{Y}$. Condition (iii) follows from the unconditional asymptotic tightness of $\mathbb{Y}_n^b$, which can be established by combining Theorem 2.12.1 and 2.9.2 in Van der Vaart and Wellner (1996).

Under dependence, much less is known about bootstrap validity for empirical processes, even in the non-sequential setting. For an overview of available results, see Radulović (2009). In the sequential setting, some results along those lines were recently considered by Bücher and Ruppert (2013) based on arguments from . More precisely, those authors proposed to consider variables $M_{1,n}, ..., M_{n,n}$ from a triangular scheme that satisfy certain conditions [see assumptions A1-A3 in their paper]. In particular, the results in Bücher and Ruppert (2013) imply (WB) for $K = \{0\} \times [0, 1]$ under strong mixing conditions for the class of functions $\mathcal{F} = \{u \mapsto I\{u \leq w\} | w \in \mathbb{R}^d\}$. Moreover, using the techniques in that paper, in particular the Ottaviani type inequality [Lemma 1 in Appendix B of the corresponding

paper], it should be possible to derive (WB) for $K = \{0\} \times [0,1]$ by combining arguments similar to those in the proof of Theorem 2.12.1 in Van der Vaart and Wellner (1996) with the Ottaviani-type inequality of Bücher and Ruppert (2013) and results on the validity of bootstrap procedures in the non-sequential setting. For an overview of such results, see Radulović (2009) and the references cited therein.

## 2.3 A general result on (quasi) Hadamard differentiability

This section contains an abstract result on compact differentiability that seems to be of independent interest. It plays a crucial role in the proofs of Theorems 2.7 and 2.10. The result in this section applies to both classical Hadamard differentiability [also known as compact differentiability], and the more general concept of quasi-Hadamard differentiability which was recently introduced by Beutner and Zähle (2010). The main advantage of this more general approach is that it allows to apply a modified delta method in settings where the classical delta method fails, the simplest example being the mean. In particular, the distribution of U- and V-statistics and value-at-risk functionals can be derived in settings where the classical delta method fails. See Beutner and Zähle (2010, 2012); Beutner et al. (2012) for further details. For the reader's convenience, we state the definition from Beutner and Zähle (2010).

**Definition 2.17. (Beutner and Zähle (2010))** Consider a metrized topological vector space $(R, d_R)$, a vector space $D$ with subsets $D_\phi, D_0 \subset D, C_0 \subset D_0$ and assume that $(D_0, d_D)$ is a metrized topological vector space. A map $\phi : D_\phi \to R$ is said to be *quasi-Hadamard differentiable at $x \in D_\phi$ tangentially to $C_0\langle D_0 \rangle$ with derivative $\phi'_x$* if for every $t_n \searrow 0$ and sequence $h_n \to h$ with $h_n \in D_0 \forall n$, $h \in C_0$ such that $x + t_n h_n \in D_\phi \forall n$ we have

$$d_R(t_n^{-1}(\phi(x + t_n h_n) - \phi(x)), \phi'_x h) \to 0$$

with $\phi'_x$ denoting a continuous, linear map $C_0 \to R$.

Consider the following general setting.

(S) Denote by $(R, d_R)$ a metrized topological vector space. Consider a second vector space $D$ with subsets $D_\phi, D_0 \subset D, C_0 \subset D_0$ and assume that $(D_0, d_D)$ is a metrized topological vector space. Let $\phi : D_\phi \to R$ be quasi Hadamard differentiable at $x$ tangentially to $C_0\langle D_0 \rangle$ and denote the derivative by $\phi'_x$. Let $(K, d_K)$ be a compact metric space. Define the sets

$$\mathbf{D} := \left\{ (h_t)_{t \in K} \,\Big|\, h_t \in D \;\forall t \right\}$$

$$\mathbf{R} := \left\{ (h_t)_{t \in K} \,\Big|\, h_t \in R \;\forall t, \sup_{s,t} d_R(h_t, h_s) < \infty \right\}$$

$$\mathbf{D}_\Phi := \left\{ (h_t)_{t \in K} \,\Big|\, h_t \in D_\phi \;\forall t, \; \sup_{s,t} d_R(\phi(h_t), \phi(h_s)) < \infty \right\}$$

$$\mathbf{R}_\Phi := \left\{ (h_t)_{t \in K} \,\Big|\, h_t \in D \;\forall t, \sup_{s,t} d_R(h_t, h_s) < \infty \right\}$$

$$\mathbf{D}_0 := \left\{ (h_t)_{t \in K} \,\Big|\, h_t \in D_0 \;\forall t, \; \sup_{s,t} d_D(h_t, h_s) < \infty \right\}$$

$$\mathbf{C}_0 := \left\{ (h_t)_{t \in K} \,\Big|\, h_t \in C_0 \;\forall t, \; \sup_{s,t} d_D(h_t, h_s) < \infty \right\}$$

14

On the sets $\mathbf{R}_\Phi$ and $\mathbf{D}_0$, define the metrics

$$d_{R,\Phi}((h_t)_{t \in K}, (g_t)_{t \in K}) := \sup_t d_R(h_t, g_t) \text{ and } d_{D,\Phi}((h_t)_{t \in K}, (g_t)_{t \in K}) := \sup_t d_D(h_t, g_t),$$

respectively. For elements $(h_t)_{t \in K}, (g_t)_{t \in K}$ set $(h_t)_{t \in K} + a(g_t)_{t \in K} := (h_t + a g_t)_{t \in K}$ and assume that with this definition, $(\mathbf{D}_0, d_{D,\Phi})$ and $(\mathbf{R}, d_{R,\Phi})$ are metrized topological vector spaces. Define the map

$$\Phi : \begin{cases} \mathbf{D}_\Phi & \to & \mathbf{R}_\Phi \\ (h_t)_{t \in K} & \mapsto & (\phi(h_t))_{t \in K}. \end{cases}$$

**Theorem 2.18.** *Under setup (S) the map $\Phi$ is pseudo-Hadamard differentiable at $X := (x)_{t \in K}$ tangentially to $\mathbf{U}\langle \mathbf{D}_0 \rangle$ where*

$$\mathbf{U} := \left\{ (h_t)_{t \in K} : h_t \in \mathbf{C}_0 \; \forall t \in K, \; \sup_{d_K(s,t) \le \delta} d_D(h_s, h_t) = o(1) \; as \; \delta \to 0 \right\}$$

*and the derivative is given by*

$$\Phi'_X : \begin{cases} \mathbf{C}_0 & \to & \mathbf{R} \\ (h_t)_{t \in K} & \mapsto & (\phi'_x h_t)_{t \in K}. \end{cases}$$

Since quasi-Hadamard differentiability also implies classical Hadamard differentiability, the above result continues to hold in the classical setting.

As an illustration of the above result, let us consider the specific setting of Remark 2.9 in Section 2.1 where $(R, d_R) = (\ell^\infty(\mathcal{G}_1) \times ... \times \ell^\infty(\mathcal{G}_L), \|\cdot\|_{\max})$, $D = \ell^\infty(\mathcal{F}_1) \times ... \times \ell^\infty(\mathcal{F}_J)$, $d_D((f_1, ..., f_J), (f'_1, ..., f'_J)) := \max_j \|f_j - f'_j\|_\infty$. Consider the map

$$\Psi_K : \begin{cases} \mathbf{D}_\Psi & \to & \mathbf{R}_\Psi \\ (h_\kappa)_{\kappa \in K} & \mapsto & \left( w(\kappa)\phi\left(\frac{h_\kappa}{w(\kappa)}\right) \right)_{\kappa \in K} \end{cases}$$

where $K$ is a compact set, $\mathbf{R}_\Psi \subset \mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K)$ and

$$\mathbf{D}_\Psi := \left\{ (H_\kappa)_{\kappa \in K} \Big| \frac{H_\kappa}{w(\kappa)} \in D_\phi \forall \kappa \in K, \sup_{\kappa \in K} \left\| w(\kappa)\phi\left(\frac{H_\kappa}{w(\kappa)}\right) \right\| < \infty \right\}.$$

Note that with this definition, $\mathbb{V}_n = \alpha_n \left( \Psi_K(w(\kappa)\hat{y}_{n,\kappa}) - \Psi_K(X_K) \right)$ where $X_K := ((\kappa, f_1, ..., f_J) \mapsto w(\kappa)x(f_1, ..., f_J))$ and that $\mathbb{Y}_n = \alpha_n w(\kappa)(\hat{y}_{n,\kappa} - X_K)$. As long as $\inf_{\kappa \in K} |w(\kappa)| > 0$, compact differentiability of $\Psi_K$ with derivative

$$(\Psi_K)'_X : \begin{cases} \mathbf{V}_K & \to & \mathbf{R} \\ (h_\kappa)_{\kappa \in K} & \mapsto & (\phi'_x h_\kappa)_{\kappa \in K}. \end{cases}$$

is a direct consequence of Theorem 2.18 [here, $\mathbf{V}_K$ is defined similarly to $\mathbf{U}$ with $U$ replaced by $V$]. To see this, consider a sequence of real numbers $r_n \searrow 0$ and $h_n \in \mathbf{D}$ such that $X_K + r_n h_n \in \mathbf{D}_\Psi$ for all $n \in \mathbb{N}$ with $h_n \to h \in \mathbf{V}_K$. Then, by compact differentiability of $\Phi$,

$$r_n^{-1}(\Psi_K(X_K + r_n h_n) - \Psi_K(X_K)) = r_n^{-1}w(\cdot)(\Phi_K(\tilde{X}_K + r_n \tilde{h}_n) - \Phi_K(\tilde{X}_K)) \to w(\cdot)\Phi'_K \tilde{h}.$$

15

where $\tilde{X}_K := ((\kappa, f_1, ..., f_J) \mapsto x(f_1, ..., f_J))$, $\tilde{h}_n := ((\kappa, f_1, ..., f_J) \mapsto h_n(\kappa, f_1, ..., f_J)/w(\kappa))$ and $\tilde{h} := ((\kappa, f_1, ..., f_J) \mapsto h(\kappa, f_1, ..., f_J)/w(\kappa))$. Finally, observing that $\Phi'_K$ is linear, compact differentiability of $\Psi_K$ and the definition of its derivative follow.

This result is of independent interest. For example, Gao and Zhao (2011) recently demonstrated that compact differentiability can be used to establish large and moderate deviation principles. The findings above allow to carry their results into the setting of statistics from subsamples and could for example be used to analyze rejection probabilities of various breakpoint tests.

# 3 Applications

In this section, we demonstrate how the results in Section 2 can be applied to various subsample based methodologies studied in the recent literature. Throughout this section, we will assume that we have a sample of data $X_1, ..., X_n$ from a strictly stationary time series. The process $\hat{y}_{n,s,t}$ is assumed to be based on the sub-sample $X_{\lfloor ns \rfloor + 1}, ..., X_{\lfloor nt \rfloor}$, i.e. of the form given in equation (1) . In what follows, write $\theta = \phi(x)$ for the parameter of interest and define $\hat{\vartheta}_{n,s,t} := \phi(\hat{y}_{n,s,t})$. For notational convenience, we also consider the quantity $\hat{\theta}_{n,k,j}$ which is computed from the data $X_k, X_{k+1}, ..., X_j$. Note that $\hat{\theta}_{n,k,j} = \hat{\vartheta}_{n,k/n-1,j/n}$. For the sake of a shorter notation, introduce the abbreviation $\mathbb{V}_{s,t} := \mathbb{V}(s, t, \cdot)$.

## 3.1 Self-normalization

For a weakly dependent stationary time series, inference on a finite-dimensional quantity (say, mean or median) typically involves a consistent estimation of the asymptotic variance matrix of the sample estimator. The difficulty with the traditional approach lies in the bandwidth parameter(s) involved in the consistent estimation, which also occurs for other existing approaches, such as sub-sampling [Politis and Romano (1994)], moving block bootstrap [Künsch (1989)] and block-wise empirical likelihood [Kitamura (1997)]. To avoid the bandwidth selection, a general self-normalized approach to confidence interval construction and hypothesis testing for a stationary time series has been developed in Shao (2010a). The basic idea is to use recursive estimates to form an inconsistent estimator of asymptotic variance (matrix) of a statistic and use a non-standard but pivotal limiting distribution to perform the inference. The SN approach is convenient to implement as recursive estimates can be easily calculated with no need to develop new algorithms. Moreover, it does not involve any bandwidth parameters and its finite sample performance is comparable or could be superior to some other existing bandwidth-dependent inference methods, as shown in Shao (2010a). Owing to these nice features, it has been recently extended to a few important inference problems in time series; see Shao and Zhang (2010); Shao (2011, 2012); Zhou and Shao (2013), among others.

The theory for the SN approach was first developed in Shao (2010a,b) by adopting a traditional approach, which is based on a linearization of the statistic and assumptions on uniform negligibility of remainder

terms. More precisely, Shao (2010a) assumed that

$$\hat{\theta}_{n,1,k} = \theta + k^{-1} \sum_{i=1}^{k} L(X_i) + R_n(k/n) \tag{6}$$

where $\{R_n(k/n)\}_{k=1}^{n}$ denote negligible remainder terms. To describe the basic idea of Shao's approach, note that we generally expect that for a weakly dependent stationary time series and smooth functional $\phi$, $R_n(1) = o_P(n^{-1/2})$ and

$$n^{-1/2} \sum_{j=1}^{n} L(X_j) \rightsquigarrow N(0, \Sigma), \tag{7}$$

where $\Sigma = \sum_{k \in \mathbb{Z}} \operatorname{cov}(L(X_0), L(X_k)) > 0$ is the so-called long run variance matrix. Further note that we implicitly assume $\mathbb{E}[L(X_j)] = 0$, which is trivially satisfied in many cases. Inference on $\theta$ is then based on estimating the covariance matrix $\Sigma$ consistently, which can be difficult as it involves a choice of bandwidth parameters. To avoid those complications, Shao (2010a) proposed to consider the self-normalized statistic

$$G_n = n(\hat{\theta}_{n,1,n} - \theta)' V_n^{-1} (\hat{\theta}_{n,1,n} - \theta)$$

where $V_n = n^{-2} \sum_{j=1}^{n} j^2 (\hat{\theta}_{n,1,j} - \hat{\theta}_{n,1,n})(\hat{\theta}_{n,1,j} - \hat{\theta}_{n,1,n})'$ is the self-normalization matrix. In Shao (2010a,b), the asymptotic distribution of $G_n$ was derived under the following assumptions:

$$n^{-1/2} \sum_{j=1}^{\lfloor nt \rfloor} L(X_j) \rightsquigarrow \Sigma^{1/2} \mathbb{B}(t), \tag{8}$$

$$R_n(1) = o_P(n^{-1/2}), \quad n^{-2} \sum_{j=1}^{n} |j R_n(j/n)|^2 = o_p(1) \tag{9}$$

with $\mathbb{B}$ denoting a vector of independent Brownian motions on $[0,1]$. To verify (9), a common approach is to derive a uniform Bahadur representation for $\hat{\theta}_{n,1,\lfloor nt \rfloor}$ and control the order of $R_n(t)$ uniformly over $t \in [0,1]$. Such a task is in general not easy and it requires a tedious case-by-case study. Under the assumptions above, Shao (2010a) proved that

$$G_n \rightsquigarrow \mathbb{B}^T(1) \Big( \int_0^1 \Big( \mathbb{B}(t) - t\mathbb{B}(1) \Big) \Big( \mathbb{B}(t) - t\mathbb{B}(1) \Big)^T dt \Big)^{-1} \mathbb{B}(1), \tag{10}$$

where the limiting distribution is pivotal and does not depend on the unknown covariance matrix $\Sigma$. Using the results in Section 2, we can both considerably generalize the findings in Shao (2010a) and at the same time avoid tedious calculations required to bound remainder terms. The key observation is that the only result required to derive (10) is weak convergence of the process $\Big( \sqrt{n} t(\hat{\vartheta}_{n,0,t} - \theta) \Big)_{t \in [0,1]}$. In the language of Section 2.1, this amounts to setting $K = \{0\} \times [0,1]$. Assuming that $\phi(x)$ is an element of $\mathbb{R}^p$, the quantity $\mathbb{V}_n(s,t,\cdot)$ can be viewed as a $\mathbb{R}^p-$valued vector. Abusing notation, denote this vector by $\mathbb{V}_n(s,t)$. Similarly, denote by $\mathbb{V}_{s,t}$ the vector $\mathbb{V}(s,t,\cdot)$. Some straightforward calculations show that under assumptions (A1)-(A3), (C), (W) the statistic $G_n$ can be represented as

$$G_n = \mathbb{V}_n^T(0,1) \Big( \int_0^1 \Big( \mathbb{V}_n(0,t) - t\mathbb{V}_n(0,1) \Big) \Big( \mathbb{V}_n(0,t) - t\mathbb{V}_n(0,1) \Big)^T dt \Big)^{-1} \mathbb{V}_n(0,1) + o_P(1).$$

An application of Theorem 2.7 with the set $K = \{0\} \times [0,1]$ in combination with the discussion at the beginning of this section and the continuous mapping theorem yields

$$G_n \rightsquigarrow \mathbb{V}_{0,1}^T \left( \int_0^1 \left( \mathbb{V}_{0,t} - t\mathbb{V}_{0,1} \right) \left( \mathbb{V}_{0,t} - t\mathbb{V}_{0,1} \right)^T dt \right)^{-1} \mathbb{V}_{0,1}.$$

Under the assumption that $\mathbb{V}(0, t, \cdot) = \Sigma^{1/2}\mathbb{B}(0, t, \cdot)$, the limit of the statistic $G_n$ is pivotal. Note that the limiting process will typically have this form in most settings with weakly dependent data, see Remark 2.13.

With the general machinery of Section 2 at hand, there are several extensions and remarks that can be made to the self-normalization approach. First, observe that we can replace the self-normalization matrix $V_n$ with a more general statistic of the form

$$V_n(H) := \int_\Delta (\hat{\vartheta}_{n,s,t} - (t-s)\hat{\vartheta}_{n,0,1})(\hat{\vartheta}_{n,s,t} - (t-s)\hat{\vartheta}_{n,0,1})^T dH(s,t)$$

with $H$ denoting an arbitrary probability measure on $\Delta$. By the continuous mapping theorem, we have joint convergence of $(V_n(H), \hat{\vartheta}_{n,0,1})$ to $(W(H), \mathbb{V}_{0,1})$ where

$$W(H) := \int_\Delta \left( \mathbb{V}_{s,t} - (t-s)\mathbb{V}_{0,1} \right) \left( \mathbb{V}_{s,t} - (t-s)\mathbb{V}_{0,1} \right)^T dH(s,t).$$

Assuming that $W(H)$ is non-singular almost surely [which happens as soon as $H$ places mass on sufficiently many different points], the asymptotic distribution of the generalized self-normalized statistic $G_n(H)$ follows. We thus have derived the following result.

**Proposition 3.1.** *Let assumptions (A1), (A2), (W), (C) hold and assume that either the support of $H$ is bounded away from the set $\{(t,t)|t \in [0,1]\}$ or that (A3) holds. Additionally, assume that $W(H)$ is non-singular almost surely. Then the generalized SN-statistic $G_n$ satisfies*

$$G_n(H) := \mathbb{V}_n^T(0,1)V_n(H)^{-1}\mathbb{V}_n(0,1) \rightsquigarrow \mathbb{V}_{0,1}^T W(H)^{-1}\mathbb{V}_{0,1}.$$

Finally, note that by the discussion in Remark 2.8 it might be advantageous to exclude estimators $\hat{\theta}_{n,k,l}$ that are based on too small proportions of data. By considering a modified version of the statistic $G_n$ of the form $\bar{G}_n(H) := \mathbb{V}_n^T(0,1)\bar{V}_n(H)^{-1}\mathbb{V}_n(0,1)$ with

$$\bar{V}_n(H) := \frac{\int_\Delta (\hat{\vartheta}_{n,s,t} - (t-s)\hat{\vartheta}_{n,0,1})(\hat{\vartheta}_{n,s,t} - (t-s)\hat{\vartheta}_{n,0,1})^T I\{t-s > n^{-\gamma}\}dH(s,t)}{\int_\Delta I\{t-s > n^{-\gamma}\}dH(s,t)},$$

and arbitrary $\gamma \in (0, 1/2)$, we would obtain the convergence $\bar{G}_n(H) \rightsquigarrow \mathbb{V}_{0,1}^T W(H)^{-1}\mathbb{V}_{0,1}^T$ without the need for assumption (A3).


## 3.2 Subsampling and fixed-b corrections

Sub-sampling [Politis and Romano (1994)] has been used in a wide range of inference problems for time series. The basic idea is that the distribution of an estimator computed from a sufficiently large sub-sample of the data should be close to that of the estimator from the whole data set. Confidence intervals

and tests can then be constructed by approximating the unknown distribution of the estimator with sub-sampling counterparts. To accommodate the time series dependence non-parametrically, it involves the sub-sampling window width $l$, which needs to go to infinity as sample size goes to infinity but at a slower rate to achieve consistent approximation. In practice, the choice of $l$ affects the sub-sampling distribution estimator and related operating characteristics, although its role does not show up in the conventional first order asymptotics. In Shao and Politis (2013), the traditional sub-sampling method was calibrated using a p-value based argument under the so-called fixed-$b$ asymptotics [Kiefer and Vogelsang (2005)], where $b = l/n$. For simplicity, assume that $\theta$ is $\mathbb{R}^d$-valued. Defining $N = n - l + 1$, the sub-sampling based estimator of the distribution function of $\|\sqrt{n}\{\hat{\theta}_{n,1,n} - \theta\}\|$ evaluated at $x$ is

$$L_{n,l}(x) = N^{-1} \sum_{j=1}^{N} I\{\|\sqrt{l}(\hat{\theta}_{n,j,j+l-1} - \hat{\theta}_{n,1,n})\| \le x\}.$$

The corresponding p-value of the test statistic $\|\sqrt{n}(\hat{\theta}_{n,1,n} - \theta_0)\|$ for the null hypothesis $\theta = \theta_0$ is

$$\hat{p}_n(b) = N^{-1} \sum_{j=1}^{N} I\{\|\sqrt{n}(\hat{\theta}_{n,1,n} - \theta_0)\| \le \|\sqrt{l}(\hat{\theta}_{n,j,j+l-1} - \hat{\theta}_{n,1,n})\|\}.$$

Note that under the conditions $l/n + 1/l = o(1)$ and additional regularity assumptions, $\hat{p}_n(b)$ has a uniform asymptotic distribution, see Politis et al. (1999). Under the fixed-$b$ asymptotic framework, $l/n = b \in (0,1]$ is held fixed. Following an elementary approach, the limiting null distribution of $\hat{p}_n(b)$, which equals

$$G(b) = (1-b)^{-1} \int_0^{1-b} I\{\|\Sigma^{1/2}\mathbb{B}(1)\| \le \|\Sigma^{1/2}(\mathbb{B}(b+t) - \mathbb{B}(t) - b\mathbb{B}(1))\|/\sqrt{b}\}dt$$

was derived in Shao and Politis (2013) by assuming that

$$\hat{\theta}_{n,j,j+l-1} = \theta + l^{-1} \sum_{i=j}^{j+l-1} L(X_i) + R_n(j, j+l-1),$$

that a similar representation holds for $\hat{\theta}_{n,1,n}$, that (8) holds for $\{L(X_t)\}$ with remainder $R_n(1,n)$, and that the remainder terms satisfy $\sqrt{n}|R_n(1,n)| = o_p(1)$ and $\sqrt{l}\sup_{j=1,\cdots,N}|R_n(j,j+l-1)| = o_p(1)$. Verifying the latter assumption for general functionals can be quite tedious and challenging.

Now, consider the general setup of Section 2 and let conditions (C), (W), (A1) and (A2) hold. We apply Theorem 2.7 with $K := \{(t, t+b)|t \in [0, 1-b]\} \cup \{(0,1)\}$ and assume that the map

$$h \mapsto \frac{1}{1-b} \int_0^{1-b} I\{\|h(0,1)\| \le \|h(t, t+b) - bh(0,1)\|/\sqrt{b}\}dt$$

is continuous on a set of functions that contains the sample paths of $\|\mathbb{V}\|$ with probability one. In particular, this is the case if $\mathbb{V}(s,t) = (t-s)\Sigma^{1/2}(\mathbb{B}(t) - \mathbb{B}(s))$ with $\Sigma$ denoting a non-singular matrix and $\mathbb{B}$ a vector of independent Brownian motions [see the arguments in Shao and Politis (2013)], which is typically the case for weakly dependent stationary time series. From now on, assume that this is the case. Observe that for $\theta = \theta_0$ we have in the setting discussed above

$$\hat{p}_n(b) = \frac{1}{1-b} \int_0^{1-b} I\{\mathbb{V}_n(0,1) \le \|\mathbb{V}_n(t, t+b) - b\mathbb{V}_n(0,1)\|/\sqrt{b}\}dt + o_P(1),$$

where the negligibility of remainder follows from an application of the continuous mapping theorem. The results in Theorem 2.7 in combination with the continuous mapping theorem thus yield

$$\hat{p}_n(b) \rightsquigarrow \frac{1}{1-b} \int_0^{1-b} I\{\|\mathbb{V}(0,1,\cdot)\| \leq \|\mathbb{V}(t,t+b,\cdot) - b\mathbb{V}(0,1,\cdot)\|/\sqrt{b}\}dt := P$$

as soon as assumptions (C), (W), (A1) and (A2) hold.

Unless $\theta$ is real-valued, the asymptotic distribution of the statistic $\hat{p}_n(b)$ is in general not pivotal. Shao and Politis (2013) proposed to estimate its distribution based on further sub-sampling. An alternative is to consider block bootstrap approximations such as those discussed in Section 2.2. More precisely, consider a bootstrap version for $y_{n,s,t}$ which is of the form given in (2) and denote it by $\hat{y}_{n,s,t}^B$. Define a bootstrap version for $\hat{\vartheta}_{n,s,t}$ through $\hat{\vartheta}_{n,s,t}^B := \phi(\hat{y}_{n,s,t}^B)$. Assume that the map $\phi$ is continuous. Now Theorem 2.10 combined with the continuous mapping theorem for the bootstrap in probability [see Theorem 10.8 in Kosorok (2008)] directly yields that under condition (WB)

$$\hat{p}_n^B(b) := \frac{1}{1-b} \int_0^{1-b} I\{\sqrt{n}\|\hat{\vartheta}_{n,0,1}^B - \theta\| \leq \sqrt{nb}\|\hat{\vartheta}_{n,t,t+b}^B - \hat{\vartheta}_{n,0,1}^B\|\}dt \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} P.$$

Finally, note that the reasoning above does not rely on $\theta$ being $\mathbb{R}^p$-valued and that it is thus also possible to handle infinite dimensional parameters.

## 3.3 Testing for change points

Testing change points in a time series is a well-studied topic in econometrics and statistics; see Perron (2006) for a recent review. A large class of tests in the literature is based on the so-called CUSUM (cumulative sum) process and the test statistic is a smooth functional of the CUSUM process with Kolmogorov-Smirnov ($L^\infty$) test and Cramer-von-Mises ($L^2$) test being two prominent examples. To accommodate the time series dependence and make the limiting null distribution pivotal, one needs to obtain a consistent estimator of the long run variance as a studentizer. As mentioned previously, consistent estimation involves a bandwidth parameter, the choice of which is even more difficult in the change point testing problem. In particular, the fixed bandwidth (e.g., $n^{1/3}$) is not adaptive to the magnitude of dependence and the data-dependent bandwidth could lead to the so-called non-monotonic power problem [Vogelsang (1999)], i.e., the power of the test can decrease when the alternative gets farther away from the null. To overcome the non-monotonic power problem, Shao and Zhang (2010) proposed SN-based tests in a general framework. Let $\theta_t = T(\mathcal{D}(X_t)) \in \mathbb{R}^q$ be the quantity of interest which depends on the distribution of $X_t$ denoted by $\mathcal{D}(X_t)$. The goal is to test if there is a change point in $\{\theta_t\}_{t=1}^n$, i.e.

$$H_0 : \theta_1 = \cdots = \theta_n$$

and the alternative hypothesis is

$$H_1 : \theta_1 = \cdots = \theta_{k^*} \neq \theta_{k^*+1} = \cdots = \theta_n \text{ for some unknown } k^*, \ 1 \leq k^* < N.$$

This framework is general enough to include mean, median, autocorrelation at certain lags of a univariate time series. Let $T_n(k) = k/\sqrt{n}(\hat{\theta}_{n,1,k} - \hat{\theta}_{n,1,n})$ and

$$V_n(k) = n^{-2}\left\{\sum_{t=1}^k t^2(\hat{\theta}_{n,1,t} - \hat{\theta}_{n,1,k})(\hat{\theta}_{n,1,t} - \hat{\theta}_{n,1,k})'\right.$$

20

$$+ \sum_{t=k+1}^{n} (n - t + 1)^2 (\hat{\theta}_{n,t,n} - \hat{\theta}_{n,k+1,n})(\hat{\theta}_{n,t,n} - \hat{\theta}_{n,k+1,n})' \Bigg\}.$$

Then the test statistic is defined as $G_n = \sup_{k=1,\cdots,n-1} T_n(k)' V_n(k)^{-1} T_n(k)$. The asymptotic null distribution was derived in Shao and Zhang (2010) using an elementary approach. Specifically, they rely on the expansion of $\hat{\theta}_{n,t_1,t_2}$, i.e.,

$$\hat{\theta}_{n,t_1,t_2} = \theta + (t_2 - t_1 + 1)^{-1} \sum_{t=t_1}^{t_2} L(X_t) + R_n(t_1, t_2).$$

Again the functional central limit theorem is assumed for $\{L(X_t)\}$ (i.e., (8) holds) and the remainder terms are assumed to be asymptotically negligible. In particular, Shao and Zhang (2010) assume that

$$\sup_{k=1,\cdots,n} |kR_n(1,k)| = o_p(n^{1/2}) \quad \text{and} \quad \sup_{k=1,\cdots,n} |kR_n(n-k+1,n)| = o_p(n^{1/2}). \tag{11}$$

The above condition (11) is not easy to verify and a detailed case-by-case study is needed. Alternatively, consider the setting of Section 2. Under conditions (W), (A1), (A2) and (A3) with $K = (\{0\} \times [0,1]) \cup ([0,1] \times \{0\})$ it is possible to show that $G_n = o_P(1) + \sup_{r \in [0,1]} H_n(r)$ where

$$H_n(r) := (\hat{\vartheta}_{n,0,r} - \hat{\vartheta}_{n,0,1})^T \hat{W}_{r,n}^{-1} (\hat{\vartheta}_{n,0,r} - \hat{\vartheta}_{n,0,1})$$

with

$$\hat{W}_{r,n} := \int_0^r (\hat{\vartheta}_{n,0,s} - \hat{\vartheta}_{n,0,r})^T (\hat{\vartheta}_{n,0,s} - \hat{\vartheta}_{n,0,r}) ds + \int_r^1 (\hat{\vartheta}_{n,s,1} - \hat{\vartheta}_{n,r,1})^T (\hat{\vartheta}_{n,s,1} - \hat{\vartheta}_{n,r,1}) ds.$$

Applying Theorem 2.7 in combination with the continuous mapping theorem yields weak convergence of $G_n$ to

$$\sup_{r \in [0,1]} (\mathbb{V}_{0,r} - r\mathbb{V}_{0,1})^T W_r^{-1} (\mathbb{V}_{0,r} - r\mathbb{V}_{0,1})$$

where

$$W_r := \int_0^r \left(\mathbb{V}_{0,s} - \frac{s}{r}\mathbb{V}_{0,r}\right)^T \left(\mathbb{V}_{0,s} - \frac{s}{r}\mathbb{V}_{0,r}\right) ds + \int_r^1 \left(\mathbb{V}_{s,1} - \frac{1-s}{1-r}\mathbb{V}_{r,1}\right)^T \left(\mathbb{V}_{s,1} - \frac{1-s}{1-r}\mathbb{V}_{r,1}\right) ds.$$

Finally, note that by considering the modification $\tilde{G}_n := \sup_{r \in [n^{-\gamma}, 1-n^{-\gamma}]} H_n(r)$ with $\gamma \in (0, 1/2)$ arbitrary, assumption (A3) can be dropped. See Remark 2.8 for further details.

# A    Proofs of main results

**Proof of Theorem 2.7** The proof consists of two steps. First, we show that the convergence holds for $K$ with $\inf_{(s,t) \in K} t - s > 0$, and second, we extend the result to general sets $K \subset \Delta$ under assumption (A3). The first step follows by an application of the functional delta method [see Theorem 3.9.4 in Van der Vaart and Wellner (1996)] in combination with Theorem 2.18. In particular, the space $\mathbf{D}_0$ can be identified with $\mathcal{L}(\mathcal{F}_1, ..., \mathcal{F}_J; K)$ since a finite norm $\|H\|_{\mathcal{L}}$ is equivalent to the distance $\|H - (x)_{(s,t) \in K}\|_{\mathcal{L}}$

being finite. Similarly, the space $R_\Psi$ is identified with $\ell^\infty(\mathcal{G} \times K)$ and the metric $d_{R,\Phi}$ corresponds to the supremum norm on $\mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K)$. Observe that

$$\alpha_n(\hat{y}_{n,s,t}(f_1, ..., f_J) - x(f_1, ..., f_J)) = \frac{1}{t-s} \, \mathbb{Y}_n(s, t, f_1, ..., f_J).$$

The functional delta method in combination with elementary considerations thus implies

$$\alpha_n(\phi(\hat{y}_{n,s,t}) - \phi(x)) \rightsquigarrow \frac{1}{t-s} \phi'_x \mathbb{Y}(s, t, \cdot) \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{G}_1, ..., \mathcal{G}_L; K),$$

the factor $\frac{1}{t-s}$ can be moved in front since $\phi'_x$ is a linear map. Multiplying both sides by $t - s$, the Continuous Mapping Theorem [see Theorem 1.3.6 in Van der Vaart and Wellner (1996)] completes the first step of the proof.

For the second step, define the set $K_S := \{(s, t) \in K | t - s \in S\}$ and consider the approximating processes

$$A_{i,n} := (t - s)\alpha_n I_{(s,t)\in K_{[1/i,1]}}(\phi(\hat{y}_{n,s,t}) - \phi(x)), \quad A_i := \phi'_x \mathbb{Y}(s, t, \cdot) I_{(s,t)\in K_{[1/i,1]}}.$$

It then suffices to verify the following three statements [see Bücher et al. (2011)]

    (i)    For every $i \in \mathbb{N}$:  $A_{i,n} \rightsquigarrow A_i$  for $n \to \infty$,

    (ii)   $A_i \rightsquigarrow \phi'_x \mathbb{Y}(s, t, \cdot)$  for $i \to \infty$,

    (iii)  For every $\varepsilon > 0$:  $\lim\limits_{i\to\infty} \limsup\limits_{n\to\infty} \mathbb{P}^*(\|A_{i,n} - \mathbb{V}_n\| > \varepsilon) = 0$.

The first statement is the weak convergence established in the first step. For (ii), note that

$$\left\| A_i - \phi'_x \mathbb{Y}(s, t, \cdot) \right\| \leq \|\phi'_x\|_{op} \sup_{(s,t)\in K_{[0,1/i]}} \sup_j \sup_{f\in\mathcal{F}_j} |\mathbb{Y}_j(s, t, f)|,$$

[here, $\|\cdot\|_{op}$ denotes the operator norm] and the right-hand side converges to zero in probability, this is a direct consequence of assumption (A1).

Finally, for a proof of (iii) note that for $\beta_n := \gamma_n \vee \alpha_n^{-1/2} = o(1)$ from Lemma B.2

$$\begin{aligned}
\|A_{i,n} - \mathbb{V}_n\| &\leq \alpha_n \sup_{(s,t)\in K_{[\beta_n n^{-1/2}, 1/i]}} (t-s)\|\phi(\hat{y}_{n,s,t}) - \phi(x)\| + \sup_{(s,t)\in K_{[0,\beta_n n^{-1/2}]}} (t-s)\|\phi(\hat{y}_{n,s,t}) - \phi(x)\| \\
&\leq C \sup_{(s,t)\in K_{[\beta_n n^{-1/2}, 1/i]}} \sup_j \sup_{f\in\mathcal{F}_j} |\mathbb{Y}_{n,j}(s, t, f)| + \sup_{(s,t)\in K_{[0,\beta_n n^{-1/2}]}} (t-s)\Big( \|\phi(\hat{x}_{n,t})\| + \|\phi(x)\| \Big) \\
&\quad + I\Big\{ \sup_{(s,t)\in K_{[\beta_n n^{-1/2}, 1/i]}} (t-s)\alpha_n\|\hat{y}_{n,s,t} - x\| > \varepsilon \Big\} \alpha_n \sup_{(s,t)\in K_{[\beta_n n^{-1/2}, 1/i]}} \|\phi(\hat{y}_{n,s,t}) - \phi(x)\| \\
&=: R_{n,1} + R_{n,2} + R_{n,3}.
\end{aligned}$$

Here the second inequality follows by an application of Lemma B.1 on the set $\Big\{ \sup_{(s,t)\in K_{[\beta_n n^{-1/2}, 1/i]}} \alpha_n\|\hat{y}_{n,s,t} - x\| \leq \varepsilon \Big\}$ after observing that by definition

$$\alpha_n\|\hat{y}_{n,s,t} - x\| = \frac{1}{t-s} \sup_j \sup_{f\in\mathcal{F}_j} |\mathbb{Y}_{n,j}(s, t, f)|.$$

22

Condition (A3) implies that $R_{n,2} = o_P^*(1)$. To see that $R_{n,1} + R_{n,3}$ converge to zero in outer probability, define the set

$$S_j(i,\varepsilon) := \Big\{ y \in \ell^\infty(K \times \mathcal{F}_j) \Big| \sup_{(s,t) \in K_{[0,1/i]}} \sup_{f \in \mathcal{F}_j} |y((s,t),f)| \geq \varepsilon \Big\}.$$

This set is closed, and by the Portmanteau theorem [Theorem 1.3.4 in Van der Vaart and Wellner (1996)] combined with the weak convergence of $\mathbb{Y}_{nj}$ and assumption (A1) on $\mathbb{Y}$ we obtain

$$\limsup_{n\to\infty} P^*(\mathbb{Y}_{nj} \in S_j(i,\varepsilon)) \leq P(\mathbb{Y}_j \in S_j(i,\varepsilon))$$

for $j = 1, ..., J$. By condition (A1), $\lim_{i\to\infty} P(\mathbb{Y}_j \in S_j(i,\varepsilon)) = 0$ for every $\varepsilon > 0$. This shows that $R_{n,1} = o_P^*(1)$ and $R_{n,3} = o_P^*(1)$. Thus the proof is complete. □

**Proof of Theorem 2.10** The first assertion follows by an application of the bootstrap functional delta method [see e.g. Theorem 12.1 in Kosorok (2008)]. For more details on the appropriate identification of spaces, see the proof of Theorem 2.7 in the present note. In order to prove the second part, define the set $K_S := \{(s,t) \in K | t - s \in S\}$ and consider

$$A_{i,n}^b := (t-s)\alpha_n I_{(s,t) \in K_{[1/i,1]}}(\phi(\hat{y}_{n,s,t}^b) - \phi(\hat{y}_{n,s,t})), \quad A_i := \phi_x' \mathbb{Y}(t,\cdot) I_{(s,t) \in K_{[1/i,1]}}.$$

By Lemma B.3 it then suffices to verify the following three statements which can be regarded as adaptation of Theorem 4.2 in Billingsley (1968) to the present setting

(i)    For every $i \in \mathbb{N}:$ $A_{i,n}^b \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} A_i$  for $n \to \infty$,

(ii)   $A_i \rightsquigarrow \phi_x' \mathbb{Y}(s,t,\cdot)$  for $i \to \infty$,

(iii)  For every $\varepsilon > 0:$ $\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\|A_{i,n}^b - \mathbb{V}_n^b\| > \varepsilon) = 0$.

Assertion (i) follows from the first part. Assertion (ii) can be established by exactly the same arguments as the corresponding statement in the proof of Theorem 2.7. For a proof of the third assertion, note that $\mathbb{Y}_n^b \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} \mathbb{Y}$ implies $\mathbb{Y}_n^b \rightsquigarrow \mathbb{Y}$, see e.g. the proof of Theorem 10.4, assertion $(ii) \Rightarrow (i)$ in Kosorok (2008). Thus assertion (iii) follows by exactly the same arguments as (iii) in the proof of Theorem 2.7. Hence the proof is complete. □

**Proof of Proposition 2.12** Observe the representation

$$\hat{y}_{n,s,t}(\cdot) - x(\cdot) = \frac{n}{(\lfloor nt \rfloor - \lfloor ns \rfloor) \vee 1}\Big( \frac{\lfloor nt \rfloor}{n}\,(\hat{x}_{n,t} - x) - \frac{\lfloor ns \rfloor}{n}\,(\hat{x}_{n,s} - x) \Big)$$

and thus setting $'0/0 = 0'$

$$\mathbb{Y}_n(s,t,\cdot) = \frac{n(t-s)}{(\lfloor nt \rfloor - \lfloor ns \rfloor) \vee 1}\Big( \frac{\lfloor nt \rfloor}{nt}\,\mathbb{G}_n(t,\cdot) - \frac{\lfloor ns \rfloor}{ns}\,\mathbb{G}_n(s,\cdot) \Big)$$

Observe that $\left|\frac{\lfloor nt \rfloor}{nt} - 1\right| \leq \frac{1}{\lfloor nt \rfloor \vee 1}$ and $\left|\frac{n(t-s)}{(\lfloor nt \rfloor - \lfloor ns \rfloor) \vee 1} - 1\right| \leq \frac{3}{(\lfloor nt \rfloor - \lfloor ns \rfloor) \vee 1}$. Defining $\tilde{\mathbb{Y}}_n(s,t,\cdot) := \mathbb{G}_n(t,\cdot) - \mathbb{G}_n(s,\cdot)$, observe that $\tilde{\mathbb{Y}}_n \rightsquigarrow \mathbb{Y}$ by the continuous mapping theorem. Moreover, $\sup_t \left\|\frac{\lfloor nt \rfloor}{nt}\,\mathbb{G}_n(t,\cdot) - \right.$

$\mathbb{G}_n(t, \cdot)\big\| = o_P^*(1)$ since for $t \geq n^{-1/4}$ the factor $\frac{\lfloor nt \rfloor}{nt}$ tends to one uniformly and since $\sup_{t \leq n^{-1/4}} \|\mathbb{G}_n(t, \cdot)\| = o_P^*(1)$ by arguments similar to those used to establish the negligibility of $R_{n,1}$ at the end of the proof of Theorem 2.7. Thus it remains to show that $\left(\frac{n(t-s)}{(\lfloor nt \rfloor - \lfloor ns \rfloor) \vee 1} - 1\right) \tilde{\mathbb{Y}}_n(s, t, \cdot)$ is uniformly small. This can be done by similar arguments [distinguish the cases $t - s \leq n^{-1/4}$ and $t - s > n^{-1/4}$]. This completes the proof. □

**Proof of Theorem 2.16** Since it suffices to show asymptotic tightness of each process $\mathbb{G}_{n,j}$ individually, we will focus on $\mathbb{G}_{n,1}$. To simplify notation, define $\mathbb{Z}_n := \mathbb{G}_{n,1}$, $\mathcal{F} := \mathcal{F}_1$, $\mathcal{F}_\delta := \mathcal{F}_{1,\delta}$. Start by noting that under the assumptions of the theorem together with (5) we have for some finite constant $C_1$

$$\sup_{k \in \mathbb{N}} \mathbb{E}^* \|\mathbb{Z}_k(1, \cdot)\|_{\mathcal{F}}^q \leq C_1 < \infty \tag{12}$$

To see this, fix $\delta > 0$ and cover the set $\mathcal{F}$ with $N$ balls of radius $\delta$ and centers $f_1, ..., f_N$. Then make use of the bound

$$\sup_{k \in \mathbb{N}} \mathbb{E}^* \|\mathbb{Z}_k(1, \cdot)\|_{\mathcal{F}}^q \leq \max_{1 \leq k \leq n_0} \mathbb{E}^* \|\mathbb{Z}_k(1, \cdot)\|_{\mathcal{F}}^q + \max_{j=1,...,N} \sup_{k \in \mathbb{N}} \mathbb{E}^* \|\mathbb{Z}_k(1, f_j)\|^q + \sup_{n \geq n_0} \mathbb{E}^* \|\mathbb{Z}_n(1, \cdot)\|_{\mathcal{F}_\delta}^q$$

and condition (5).

In order to establish asymptotic tightness of $\mathbb{G}$, apply Theorem 1.5.7 in Van der Vaart and Wellner (1996) with the metric $d((s, f), (t, g)) := \rho(f, g) + |s - t|$. By the triangle inequality, we have

$$\sup_{|s-t|+\rho(f,g)<\delta} |\mathbb{Z}_n(s, f) - \mathbb{Z}_n(t, g)| \leq \sup_{0 \leq t \leq 1} \|\mathbb{Z}_n(t, f)\|_{\mathcal{F}_\delta} + \sup_{|s-t|<\delta} \|\mathbb{Z}_n(s, f) - \mathbb{Z}_n(t, f)\|_{\mathcal{F}}$$

Start by considering the first term. Define $S_k(g) = \sum_{j=1}^k \{g(X_i) - \mathbb{E}g(X_i)\}$ and note that

$$\sup_{0 \leq t \leq 1} \|\mathbb{Z}_n(t, f)\|_{\mathcal{F}_\delta} \leq \max_{1 \leq k \leq n} \sqrt{\frac{k}{n}} \|\mathbb{Z}_k(1, \cdot)\|_{\mathcal{F}_\delta} = \frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \|S_k\|_{\mathcal{F}_\delta}.$$

Fix $\epsilon \in (0, \{1 - 2^{-1/2+1/q}\}^{q/(q-1)}/2^{q/(2q-2)})$. Under (4), there exists a $\delta_0 > 0$ and $n_0 \in \mathbb{N}$, such that when $\delta \in (0, \delta_0)$ and $n \geq n_0(\epsilon)$, $\mathbb{E}^* \|\mathbb{Z}_n(1, \cdot)\|_{\mathcal{F}_\delta}^q \leq \epsilon^2$. Moreover, under (12) we have $\max_{1 \leq k \leq n_0} \|S_k\|_{\mathcal{F}_\delta} \leq C_1 \sqrt{n_0}$ for all $n_0 \in \mathbb{N}$. By the Markov inequality and Proposition 1(ii) in Wu (2007), for $q > 2$, $d = d_n = \lfloor \log n/(\log 2) \rfloor + 1$,

$$\begin{aligned}
P^*(\max_{1 \leq k \leq n} \|S_k\|_{\mathcal{F}_\delta} > \sqrt{n}\epsilon) &\leq (\sqrt{n}\epsilon)^{-q} \mathbb{E}^* [\max_{1 \leq k \leq n} \|S_k\|_{\mathcal{F}_\delta}^q] \\
&\leq (\sqrt{n}\epsilon)^{-q} \left( \sum_{j=0}^d 2^{(d-j)/q} \left\{ \mathbb{E}^* \sup_{g \in \mathcal{F}_\delta} |S_{2^j}|^q \right\}^{1/q} \right)^q \\
&\leq \epsilon^{-q} n^{-q/2} \left( O(n) + \left\{ \sum_{j=\lfloor \log n_0/(\log 2) \rfloor+1}^d 2^{(d-j)/q} (\epsilon^{2q} 2^{jq/2})^{1/q} \right\}^q \right) \\
&\leq \epsilon^{-q} O(n^{-q/2+1}) + n^{-q/2} \epsilon^q \left( \sum_{j=\lfloor \log n_0/(\log 2) \rfloor+1}^d 2^{(d-j)/q} (2^{jq/2})^{1/q} \right)^q
\end{aligned}$$

24

$$\leq \epsilon^{-q}O(n^{-q/2+1}) + \frac{2^{q/2}}{\{1-2^{-1/2+1/q}\}^q}\epsilon^q$$
$$< \epsilon$$

for $n \geq n_1(\epsilon) \vee n_0(\epsilon)$, where $n_1(\epsilon)$ is chosen such that the last inequality holds. Since $\epsilon$ was arbitrary, we have shown that $\limsup_{n\to\infty} P^*(\sup_{0\leq t\leq 1}\|\mathbb{Z}_n(t,f)\|_{\mathcal{F}_\delta} > \epsilon) < \epsilon$ for all $\delta < \delta_0$. It thus remains to consider the second term. Since the increments of $\mathbb{Z}_n(s,f)$ in $s$ are stationary, the above probability can be bounded by

$$\lceil\frac{1}{\delta}\rceil P^*\left(\sup_{0\leq s\leq\delta}\|\mathbb{Z}_n(s,f)\|_{\mathcal{F}} > \epsilon\right) = \lceil\frac{1}{\delta}\rceil P^*\left(\max_{1\leq k\leq n\delta}\|\sqrt{k}\mathbb{G}_k\|_{\mathcal{F}} > \sqrt{n}\epsilon\right). \tag{13}$$

Let $d(\delta) = \lfloor\log(n\delta)/(\log 2)\rfloor + 1$. Again by the Markov inequality and Proposition 1(ii) in Wu (2007),

$$P^*\left(\max_{1\leq k\leq n\delta}\|S_k\|_{\mathcal{F}} > \sqrt{n}\epsilon\right) \leq (\sqrt{n}\epsilon)^{-q}\mathbb{E}^*\max_{1\leq k\leq n\delta}\|S_k\|_{\mathcal{F}}^q$$
$$\leq (\sqrt{n}\epsilon)^{-q}\left\{\sum_{j=0}^{d(\delta)} 2^{(d(\delta)-j)/q}\left(\mathbb{E}^*\|S_{2^j}\|_{\mathcal{F}}^q\right)^{1/q}\right\}^q$$
$$\leq (\sqrt{n}\epsilon)^{-q}\left\{\sum_{j=0}^{d(\delta)} 2^{(d(\delta)-j)/q}C_1^{1/q}2^{j/2}\right\}^q$$
$$\leq (\sqrt{n}\epsilon)^{-q}C_1 2^{d(\delta)q/2}\frac{1}{\{1-2^{-(1/2-1/q)}\}^q}$$
$$\leq C_2\epsilon^{-q}\delta^{q/2}$$

for $n$ sufficiently large. Combined with (13), we get $\limsup_{n\to\infty} P^*(\max_{0\leq j\delta\leq 1}\sup_{j\delta\leq s\leq(j+1)\delta}\|\mathbb{Z}_n(s,f) - \mathbb{Z}_n(j\delta,f)\|_{\mathcal{F}} > \epsilon) < \epsilon$ when $\delta < (\epsilon^{q+1}/C_2)^{1/(q/2-1)}$. The proof is thus complete. □

**Proof of Theorem 2.18** Let $a_n = o(1)$ and $H^{(n)}$ denote a sequence in $\mathbf{D}_0$ with $H^{(n)} \to H \in \mathbf{U}$ such that $X + a_n H^{(n)} \in \mathbf{D}_\Phi \ \forall n \in \mathbb{N}$. We need to show that

$$a_n^{-1}\left(\Phi(X + a_n H^{(n)}) - \Phi(X)\right) \to \Phi'_X H.$$

Assume that this does not hold. Then there exists a sequence $t_n$ and a positive number $b$ such that

$$d_R\left(a_n^{-1}(\phi(x + a_n H_{t_n}^{(n)}) - \phi(x)), (\phi'_x \cdot H_{t_n})\right) \geq b \tag{14}$$

for all $n \geq N_0$. On the other hand, the sequence $H_{t_n}^{(n)}$ has a subsequence $H_{t_{n_k}}^{(n_k)}$ which converges to $H_{t_\infty}$ for some $t_\infty \in K$. To see that this is the case, start by noting that $t_n$ is a sequence in a compact metric space, i.e. it has a convergent subsequence $t_{n_k} \to t_\infty$ with $t_\infty \in K$. The definition of the set $\mathbf{U}$ then implies that $H_{t_{n_k}} \to H_{t_\infty}$. Together with the uniform convergence $\sup_t d_D(H_t^{(n)}, H_t) = o(1)$ this yields $H_{t_{n_k}}^{(n_k)} \to H_{t_\infty}$. Now quasi compact differentiability of $\phi$ tangentially to $C_0\langle D_0\rangle$ implies

$$a_n^{-1}\left(\phi(x + a_n H_{t_{n_k}}^{(n_k)}) - \phi(x)\right) \to \phi'_x H_{t_\infty},$$

and together with continuity of $\phi'_x$ this contradicts (14). Thus the proof is complete. □

# B   Auxiliary technical results

**Lemma B.1.**   *Denote by $(R, \| \cdot \|_R)$ a normed vector space. Consider a second vector space $D$ with subsets $D_\phi, D_0 \subset D, C_0 \subset D_0$ and assume that $(D_0, \| \cdot \|_D)$ is a normed vector space. Let $\phi : D_\phi \to R$ be quasi compactly differentiable at $x$ tangentially to $C_0 \langle D_0 \rangle$ and assume $0 \in C_0$. Then there exist constants $\varepsilon > 0, K < \infty$ such that*

$$\|\phi(x) - \phi(x + y)\|_R \leq K \|y\|_D \quad \forall y \in D_0 : \ \|y\|_D \leq \varepsilon, x + y \in D_\phi. \tag{15}$$

**Proof** Assume that (15) does not hold. Then for any pair $\varepsilon > 0, K < \infty$ there exists a $y_{K,\varepsilon} \in D_0$ such that $x + y_{K,\varepsilon} \in D_\phi$, $\|y_{K,\varepsilon}\|_D \leq \varepsilon$ and $\|\phi(x) - \phi(x + y_{K,\varepsilon})\|_R > K \|y_{K,\varepsilon}\|_D$. Consider the sequence $z_n := y_{n^2, n^{-2}}$ and define $\alpha_n := \|z_n\|_D \neq 0$. Then

$$\left\| \frac{\phi(x + n\alpha_n (n\alpha_n)^{-1}(z_n)) - \phi(x)}{n\alpha_n} \right\|_R > \frac{n^2 \alpha_n}{n\alpha_n} = n \longrightarrow \infty.$$

Moreover $\|(n\alpha_n)^{-1}(z_n)\|_D = n^{-1} = o(1)$, i.e. $(n\alpha_n)^{-1}(z_n) \to 0$. This yields a contradiction since quasi compact differentiability of $\phi$ implies that [note that $n\alpha_n \leq n^{-1} = o(1)$]

$$\frac{\phi(x + n\alpha_n (n\alpha_n)^{-1}(z_n)) - \phi(x)}{n\alpha_n} \longrightarrow \phi'_x 0.$$

Thus the proof is complete. $\qquad \square$

**Lemma B.2.**   *Under assumptions (W) and (A1) there exists a sequence of real numbers $\gamma_n = o(1)$ such that $\sup_{(s,t) \in K_{\alpha_n^{-1} \gamma_n}} \|\hat{y}_{n,s,t} - x\| = o_P^*(1)$ where we defined $K_a := \{(s,t) \in K | t - s \geq a\}$.*

**Proof** Define $K_a^C$ as the complement of $K_a$ in $K$ and set

$$B_n := \sup_{(s,t) \in K_{\alpha_n^{-1/2}}^C} \sup_{f_1, \ldots, f_J} \|\mathbb{Y}_n(s, t, f_1, \ldots, f_J)\|.$$

By asymptotic equicontinuity of $\mathbb{Y}_n$ [see the discussion in the proof of Theorem 2.7 for more details and note that $\sup_{s=t} \sup_{f_1, \ldots, f_J} |\mathbb{Y}_n(s, t, f_1, \ldots, f_J)| \equiv 0$ a.s.] we have $B_n = o_P^*(1)$. This implies

$$\forall \varepsilon > 0 \ \exists n_0(\varepsilon) \in \mathbb{N} : \quad (*) \ \forall n \geq n_0(\varepsilon) \ P^*(B_n > \varepsilon) < \varepsilon.$$

Note that $a \mapsto n_0(a)$ is decreasing since for any $a < b$ we have $P^*(B_n > a) < a \Rightarrow P^*(B_n > b) < b$. Set $N_0(\varepsilon) := 2 \inf\{n_0(\varepsilon) | (*) \text{ holds}\}$ and define

$$\delta_n := 2 \inf\{\varepsilon > 0 | n > N_0(\varepsilon)\}.$$

By construction $N_0(\delta_n) < n$, and thus $P^*(B_n > \delta_n) < \delta_n$. Moreover, $\delta_n \to 0$ since by construction $\delta_n \leq \varepsilon \ \forall n \geq N_0(\varepsilon/3)$. Defining $\gamma_n = \delta_n^{1/2}$ yields $B_n = o_P^*(\gamma_n)$. Note that

$$\sup_{(s,t) \in K_{\alpha_n^{-1} \gamma_n}} \|\hat{y}_{n,s,t} - x\| \leq \sup_{(s,t) \in K_{\alpha_n^{-1/2}}} \|\hat{y}_{n,s,t} - x\| + \sup_{(s,t) \in K_{\alpha_n^{-1/2}}^C \cap K_{\gamma_n \alpha_n^{-1}}} \|\hat{y}_{n,s,t} - x\|.$$

Now observe that

$$\sup_{(s,t)\in K_{\alpha_n^{-1/2}}} \|\hat{y}_{n,s,t} - x\| \le 2\alpha_n^{-1/2} \sup_{(s,t)\in K_{\alpha_n^{-1/2}}} \|\mathbb{Y}_n(s,t,\cdot)\| = o_P^*(1),$$

by arguments similar to those used to establish the negligibility of $R_{n,1}$ at the end of the proof of Theorem 2.7. Similarly

$$\sup_{(s,t)\in K_{\alpha_n^{-1/2}}^C \cap K_{\gamma_n\alpha_n^{-1}}} \|\hat{y}_{n,s,t} - x\| \le \gamma_n^{-1} \sup_{(s,t)\in K_{\alpha_n^{-1/2}}^C} \|\mathbb{Y}_n(s,t,\cdot)\| = \gamma_n^{-1} B_n = o_P^*(1)$$

This completes the proof. □

**Lemma B.3.** *Given a sequence of random variables $M_1, M_2, ...,$ and a sequence of random elements $\mathbb{V}_n^b(M_1, ..., M_n)$ in a normed space $(D, \|\cdot\|_D)$, assume that the map $(M_1, ..., M_n) \to \mathbb{V}_n^b(M_1, ..., M_n)$ is measurable for every $n \in \mathbb{N}$ outer almost surely [the randomness in $\mathbb{V}_n^b$ is allowed to come from sources apart from the $M_i$]. Assume that for $i \in \mathbb{N}$ there exist approximations $A_{i,n}^b, A_i$ such that $(M_1, ..., M_n) \to A_{i,n}^b(M_1, ..., M_n)$ is measurable for every $i, n \in \mathbb{N}$ outer almost surely.*

(i)   *For every $i \in \mathbb{N}$ : $A_{i,n}^b \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} A_i$  for $n \to \infty$,*

(ii)  *$A_i \rightsquigarrow \mathbb{V}$  for $i \to \infty$,*

(iii) *For every $\varepsilon > 0$ : $\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\|A_{i,n}^b - \mathbb{V}_n^b\| > \varepsilon) = 0$.*

*where $A_i, \mathbb{V}$ denote a tight processes. Then $\mathbb{V}_n^b \overset{\mathbb{P}}{\underset{M}{\rightsquigarrow}} \mathbb{V}$.*

**Proof of Lemma B.3** We need to show that

(a)  $\sup_{f\in BL_1} \left|\mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}f(\mathbb{V})\right| \to 0$ in outer probability,

(b)  $\mathbb{E}_M f(\mathbb{V}_n^b)^* - \mathbb{E}_M f(\mathbb{V}_n^b)_* \overset{\mathbb{P}}{\to} 0$ for all $f \in BL_1$.

Begin by observing that for every $i \in \mathbb{N}$, every $\omega$, and every $f \in BL_1$

$$\left|\mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}f(\mathbb{V})\right| \le \left|\mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}_M f(A_{i,n}^b)\right| + \left|\mathbb{E}_M f(A_{i,n}^b) - \mathbb{E}f(A_i)\right| + |\mathbb{E}f(A_i) - \mathbb{E}f(\mathbb{V})|.$$

Moreover, for every $\omega$

$$\sup_{f\in BL_1} \left|\mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}_M f(A_{i,n}^b)\right| \le \sup_{f\in BL_1} \mathbb{E}_M \left|f(\mathbb{V}_n^b) - f(A_{i,n}^b)\right| \le \mathbb{E}_M \left[\|\mathbb{V}_n^b - A_{i,n}^b\|^* \wedge 2\right].$$

In particular, this implies that for any $\gamma > 0$

$$\mathbb{E}^*\left[\sup_{f\in BL_1} \left|\mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}_M f(A_{i,n}^b)\right|\right] \le \mathbb{E}\left[\|\mathbb{V}_n^b - A_{i,n}^b\|^* \wedge 2\right] \le 2\mathbb{P}(\|\mathbb{V}_n^b - A_{i,n}^b\|^* > \gamma) + \gamma.$$

Thus (iii) yields

$$\lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{E}^*\left[\sup_{f\in BL_1} \left|\mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}_M f(A_{i,n}^b)\right|\right] = 0.$$

27

Fix arbitrary $\varepsilon, \eta > 0$. The computations above yield the existence of an $i_1 \in \mathbb{N}$ such that for all $i \geq i_1$

$$\limsup_{n \to \infty} \mathbb{P}^* \Big( \sup_{f \in \mathrm{BL}_1} \Big| \mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E}_M f(A_{i,n}^b) \Big| > \varepsilon/3 \Big) < \eta/3.$$

Moreover, by (ii) and the definition of weak convergence, there exists an $i_2 \in \mathbb{N}$ such that for all $i \geq i_2$

$$\mathbb{P}^* \Big( \sup_{f \in \mathrm{BL}_1} |\mathbb{E} f(A_i) - \mathbb{E} f(\mathbb{V})| > \varepsilon/3 \Big) < \eta/3.$$

Set $k = i_1 \vee i_2$. Then (i) implies as $n \to \infty$

$$\sup_{f \in \mathrm{BL}_1} \Big| \mathbb{E}_M f(A_{k,n}^b) - \mathbb{E} f(A_k) \Big| = o_{\mathbb{P}^*}(1),$$

and combining all the results above we see that

$$\limsup_{n \to \infty} \mathbb{P}^* \Big( \sup_{f \in \mathrm{BL}_1} \Big| \mathbb{E}_M f(\mathbb{V}_n^b) - \mathbb{E} f(\mathbb{V}) \Big| > \varepsilon \Big) < \eta.$$

Since $\eta, \varepsilon$ were arbitrary, this establishes (a). For a proof of (b), note that (i) implies $A_{i,n}^b \rightsquigarrow A_i$ since conditional weak convergence implies unconditional weak convergence [see the proof of Theorem 10.4, assertion $(ii) \Rightarrow (i)$ in Kosorok (2008)]. Thus, by an the results in Bücher et al. (2011), (i)-(iii) imply that $\mathbb{V}_n^b \rightsquigarrow \mathbb{V}$. In particular, this implies asymptotic measurability of $\mathbb{V}_n^b$ [see section 1.3 in Van der Vaart and Wellner (1996)], and together with the continuity of $f \in \mathrm{BL}_1$ this shows that $f(\mathbb{V}_n^b) \rightsquigarrow f(\mathbb{V})$ by an application of the continuous mapping theorem. Thus $\mathbb{E}_M f(\mathbb{V}_n^b)^* - \mathbb{E}_M f(\mathbb{V}_n^b)_*$ converges to zero in $L^1$, hence also in probability. Now the proof is complete. $\qquad \square$

# References

Adams, T. and Nobel, A. (2010). Uniform convergence of vapnik–chervonenkis classes under ergodic sampling. *The Annals of Probability*, 38(4):1345–1367.

Andrews, D. W. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review*, 62(1):119–132.

Aue, A. Horváth, L. and Reimherr, M. (2009). Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics*, 149:174–190.

Berkes, I., Hörmann, S., and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and Their Applications*, 119(4):1298–1324.

Beutner, E., Wu, W. B., and Zähle, H. (2012). Asymptotics for statistical functionals of long-memory sequences. *Stochastic Processes and their Applications*, 122(3):910–929.

Beutner, E. and Zähle, H. (2010). A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis*, 101(10):2452–2463.

Beutner, E. and Zähle, H. (2012). Deriving the asymptotic distribution of u-and v-statistics of dependent data using weighted empirical processes. *Bernoulli*, 18(3):803–822.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons Inc., New York.

Bücher, A., Dette, H., and Volgushev, S. (2011). New estimators of the pickands dependence function and a test for extreme-value dependence. *Annals of Statistics*, 39(4):1963–2006.

Bücher, A. and Ruppert, M. (2013). Consistent testing for a constant copula under strong mixing based on the tapered block multiplier technique. *Journal of Multivariate Analysis*, 116(0):208 – 229.

Bücher, A. and Volgushev, S. (2011). Empirical and sequential empirical copula processes under serial dependence. *arXiv preprint arXiv:1111.2778*.

Chu, C-S. Stinchcombe, M. and White, H. (1995). Monitoring structural change. *Econometrica*, 64(5):1045–1065.

Csörgö, M. and Horváth, L. (1997). *Limiting Theorems in Change-point Analysis*. New York, Wiley.

Dehling, H., Mikosch, T., and Sørensen, M. (2002). *Empirical process techniques for dependent data*. Birkhäuser.

Dehling, H. and Taqqu, M. S. (1989). The empirical process of some long-range dependent sequences with an application to u-statistics. *The Annals of Statistics*, pages 1767–1783.

Doss, H. and Gill, R. D. (1992). An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *Journal of the American Statistical Association*, 87(419):869–877.

Fermanian, J. D., Radulović, D., and Wegkamp, M. H. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860.

Gao, F. and Zhao, X. (2011). Delta method in large deviations and moderate deviations for estimators. *The Annals of Statistics*, 39(2):1211–1240.

Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *The Annals of Statistics*, pages 1501–1555.

Hagemann, A. (2012). Stochastic equicontinuity in nonlinear time series models. *Arxiv preprint arXiv:1206.2385*.

Inoue, A. (2001). Testing for distributional change in time series. *Econometric Theory*, 17(1):156–187.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21:1130–1164.

Kitamura, Y. (1997). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Annals of Statistics*, 25:2084–2102.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics, New York.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241.

Maejima, M. and Tudor, C. A. (2007). Wiener integrals with respect to the hermite process and a non-central limit theorem. *Stochastic Analysis and Applications*, 25(5):1043–1056.

Perron, P. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics, vol I: Econometric Theory, eds. K. Patterson and T.C. Mills*, pages 278–352.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22:2031–2050.

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

Radulović, D. (2009). Another look at the disjoint blocks bootstrap. *Test*, 18(1):195–212.

Rüschendorf, L. (1974). On the empirical process of multivariate, dependent random variables. *Journal of Multivariate Analysis*, 4(4):469–478.

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Annals of Statistics*, 4:912–923.

Sen, P. K. (1974). Weak convergence of multidimensional empirical processes for stationary $\phi$-mixing processes. *The Annals of Probability*, 2(1):147–154.

Shao, X. (2010a). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):343–366.

Shao, X. (2010b). corrigendum: A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(5):695–696.

Shao, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, 32(6):598–606.

Shao, X. (2012). Parametric inference in stationary time series models with dependent errors. *Scandinavian Journal of Statistics*, 39(4):772–783.

Shao, X. and Politis, D. (2013). Fixed b subsampling and the block bootstrap: improved confidence sets based on p-value calibration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75:161–184.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag, New York.

Vogelsang, T. J. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *Journal of Econometrics*, 88:283–299.

Wu, W. (2007). Strong invariance principles for dependent random variables. *The Annals of Probability*, 35(6):2294–2320.

Wu, W. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436.

Yoshihara, K.-i. (1975). Weak convergence of multidimensional empirical processes for strong mixing sequences of stochastic vectors. *Probability Theory and Related Fields*, 33(2):133–137.

Zhou, Z. and Shao, X. (2013). Inference for linear models with dependent errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75:323–343.