Electronic Journal of Statistics
Vol. 8 (2014) 390–431
ISSN: 1935-7524
DOI: 10.1214/14-EJS888

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

Stanislav Volgushev

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany
e-mail: stanislav.volgushev@rub.de

and

Xiaofeng Shao

Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, 61820
e-mail: xshao@illinois.edu

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 are of independent interest.

AMS 2000 subject classifications: Primary 62E20, 62M10, 62G09, 62G15; secondary 62G20.
Keywords and phrases: Change point, compact differentiability, empirical processes, self-normalization, sub-sampling, time series, weak convergence.

Received October 2013.

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) statistic [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 of 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 (2013)] 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 of a quantity $\mathbf{F}$. A classical example of such a collection is given by estimators computed from various fractions of the sample $X_1, \dots, X_n$. For illustration purposes, assume that $\mathbf{F}$ is the distribution function and $\hat{\mathbf{F}}_{1,k}$ denotes the empirical distribution function computed from $X_1, \dots, X_{k\vee 1}$. Also, assume that the parameter of interest, say θ, can be expressed as $\phi(\mathbf{F})$ where ϕ denotes some functional. For example, it is possible to express the copula as a functional of the cumulative distribution function. If the map ϕ is compactly differentiable, the asymptotic distribution of a suitably normalized version of $\phi(\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor})$ for fixed κ can be derived from a corresponding result for $\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor}$. More precisely, denoting by α_n a sequence diverging to infinity and by $w(\kappa)$ a weight function, weak convergence of $\mathbb{Y}_n := \alpha_n w(\kappa)(\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor} - \mathbf{F})$ in a suitable function space implies weak convergence of $\alpha_n w(\kappa)(\phi(\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor}) - \phi(\mathbf{F}))$ *for a finite collection of fixed values of* κ. However, in many important applications the *joint* weak convergence of the whole *collection* $\mathbb{W}_n := (\alpha_n w(\kappa)(\phi(\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor}) - \phi(\mathbf{F}))_{\kappa\in[0,1]}$ in a suitable functional sense is required.

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 κ. 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 to provide general ways of importing the tools mentioned above from the 'non-sequential' into the 'sequential' setting.

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

$$\Phi : (h(\cdot;\kappa))_{\kappa\in[0,1]} \mapsto \Big(w(\kappa)\phi\Big(\frac{h(\cdot;\kappa)}{w(\kappa)}\Big)\Big)_{\kappa\in[0,1]}$$

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

$$\mathbb{W}_n = \alpha_n\Big(\Phi\Big((w(\kappa)\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor})_{\kappa\in[0,1]}\Big) - \Phi\Big((w(\kappa)\mathbf{F})_{\kappa\in[0,1]}\Big)\Big).$$

Given the fact that many important maps ϕ 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 ϕ, what can we say about compact differentiability of Φ? As we shall see in Section 2, such an implication does not hold in full generality, see in particular Example 2.2 and the discussion preceding it. At the same time, we obtain a positive result if we additionally assume that the map ϕ possesses certain boundedness properties. Additionally, even when compact differentiability of Φ fails, there still are many relevant settings where additional arguments can be applied to obtain the desired weak convergence of $\mathbb{W}_n$. In fact, in Section 4.1 we show that, given weak convergence of $\mathbb{Y}_n = (\alpha_n w(\kappa)(\hat{\mathbf{F}}_{1,\lfloor n\kappa \rfloor} - \mathbf{F}))_{\kappa\in[0,1]}$, 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 5.

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 $\mathbb{Y}_n$. Unfortunately, results on weak convergence of $\mathbb{Y}_n$ in settings where the data $X_1, \dots, X_n$ are allowed to be dependent are limited. A summary of available results as well as new insights providing extensions of those findings are collected in Section 4.2.

Throughout the paper, we will use standard notation from empirical process theory. In particular, we denote by $\ell^\infty(D)$ the space of bounded, real-valued functions on D and by $\|\cdot\|_\infty$ the supremum distance. Weak convergence, denoted by $\rightsquigarrow$, will always be understood in the sense of Hoffmann-Jørgensen [see Van der Vaart and Wellner (1996), Chapter 1.3].

We conclude with an overview of the main theoretical results in this paper and their connections.

- In Section 2, we provide an illustration of our approach based on empirical distribution functions. The main result in this section is Theorem 2.5. which states conditions that allow to obtain weak convergence of processes of the type $\mathbb{W}_n$ defined above from weak convergence of $\mathbb{Y}_n$.
- In Section 4.1, we extend the results from Section 2. Essentially, we consider a general collection of (potentially function-valued) estimators, say $(\hat{\mathbf{G}}_n(\cdot;\kappa))_{\kappa\in K}$, which is indexed by a general compact metric space K. The main result is Theorem 4.5, which describes settings where weak convergence of $(\alpha_n w(\kappa)(\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}(\cdot)))_{\kappa\in K}$ in a suitable functional sense implies weak convergence of $(\alpha_n w(\kappa)(\phi(\hat{\mathbf{G}}_n(\cdot;\kappa)) - \phi(\mathbf{G}(\cdot))))_{\kappa\in K}$. This contains Theorem 2.5 as a special case.
- In Section 4.2, we collect results on sequential empirical processes indexed by general classes of functions. The main result here is Theorem 4.10, which provides new criteria for the weak convergence of processes of the form
$$\Big(\frac{1}{\lfloor ns\rfloor}\sum_{i=1}^{\lfloor ns\rfloor} f(X_i)\Big)_{f\in\mathcal{F},s\in[0,1]}$$
where $\mathcal{F}$ denotes a suitable of functions. In particular, this result can be applied to obtain weak convergence of the 'simple' sequential empirical process $\mathbb{Y}_n$.
- Section 4.3 contains a discussion of bootstrap procedures. In Theorem 4.11, we provide a version of Theorem 4.5 in the bootstrap setting.
- In Section 5, we state a general result on compact differentiability of certain maps that play a role in the analysis discussed above [see Theorem 5.2 for details]. The specific example of the map Φ considered above is further elaborated in Example 5.3.

We also provide applications of the general theoretical results to various practical examples. Sections 3.1 and 3.2 contain a detailed discussion of self-normalized statistics and sequential empirical copula processes, respectively. In Section 4.4, we additionally discuss sub-sampling and fixed-b corrections. Finally, Section C in the appendix provides details on a test for change-points.

2. An illustration based on empirical distribution functions

In order to illustrate the kind of results that can be obtained with our methodology without overwhelming the reader with too much notation, we begin by considering a particularly interesting special case. Throughout this section, assume that we have a triangular array consisting of $\mathbb{R}^d$-valued random variables $X_{1,n},\ldots,X_{n,n}$ that are defined on the same probability space $(\Omega,\mathcal{A},\mathbb{P})$.

Throughout, we assume that $X_{i,n}, i = 1, \dots, n$ have the same distribution for all i, n but are not necessarily independent. For the sake of a simple notation, we will denote the observations by $X_1, \dots, X_n$. Many quantities of interest such as L-statistics, quantiles, the copula or scalar measures of multivariate dependence can be represented as a compactly differentiable map of the empirical distribution function $\hat{\mathbf{F}}_{1,n}(v) := n^{-1} \sum_i I\{X_i \le v\}$.

Given smoothness of a map $\phi : \ell^\infty(D) \supset \boldsymbol{D}_\phi \to \boldsymbol{R}_\phi \subset \ell^\infty(R)$ with $D \subset \mathbb{R}^d, R \subset \mathbb{R}^k$ and weak convergence of $\sqrt{n}(\hat{\mathbf{F}}_{1,n} - \mathbf{F})$ as an element of $\ell^\infty(D)$, the functional delta method [see e.g. Van der Vaart and Wellner (1996), Chapter 3.9] asserts that $\sqrt{n}(\phi(\hat{\mathbf{F}}_{1,n}) - \phi(\mathbf{F}))$ also converges in distribution. However, in many settings we are interested in a more general process. More precisely, assume that for each $k \le l = 1, \dots, n$ we compute the estimator of interest from the observations $X_{k\vee 1}, \dots, X_{l\vee k\vee 1}$ and denote the corresponding empirical distribution function by $\hat{\mathbf{F}}_{k,l}$. In many cases, the asymptotic properties of $n^{-1/2}(l-k)(\phi(\hat{\mathbf{F}}_{k,l}) - \phi(\mathbf{F}))$ indexed by $k \le l = 0, \dots, n$ are of interest. In other words, we often are interested in the 'sequential' process

$$\Big(\mathbb{W}_n(u; s, t)\Big)_{u \in R, (s,t) \in K} \in \ell^\infty(R \times K)$$

where $K \subseteq \Delta := \{(s,t) \in [0,1]^2 | s \le t\}$ and

$$\mathbb{W}_n(u; s, t) := n^{-1/2}(\lfloor nt \rfloor - \lfloor ns \rfloor)(\phi(\hat{\mathbf{F}}_{\lfloor ns \rfloor, \lfloor nt \rfloor})(u) - \phi(\mathbf{F})(u)).$$

For the sake of a shorter notation, we will often write $\mathbb{W}_n(\cdot; s, t)$ for the process $(\mathbb{W}_n(u; s, t))_{u \in R}$ viewed as element of the space $\ell^\infty(R)$ and $\mathbb{W}_n$ for the process $(\mathbb{W}_n(u; s, t))_{u \in R, (s,t) \in K}$ viewed as element of $\ell^\infty(R \times K)$.

Processes of the kind discussed above are of particular interest in settings where the data stem from a time series. To illustrate ideas, consider the following simple example. More elaborate examples, including general versions of self-normalized statistics and sequential empirical copula processes can be found in Section 3.

Example 2.1. Testing change points in a time series is an important topic in econometrics and statistics; see Perron (2006) for a recent review. A large class of tests in the literature are based on comparing estimators from various fractions of the data. For instance, denoting by $\|\cdot\|$ some norm on $\ell^\infty(R)$, one could consider Kolmogorov-Smirnov type statistics of the form

$$\max_{k=1,\dots,n-1} n^{1/2} \frac{k(n-k)}{n^2} \|\phi(\hat{\mathbf{F}}_{1,k}) - \phi(\hat{\mathbf{F}}_{k+1,n})\|.$$

The limiting distribution of the corresponding test statistics under the null can typically be derived from the limit of the process $\mathbb{W}_n$, which can be obtained by the methods described in this section. Additional details for an alternative test for change points are provided in Section C.

For fixed values of s, t, weak convergence of the quantity $\mathbb{W}_n(\cdot; s, t)$ can be derived by an application of the classical functional delta method. However, this is not the case if we are interested in the process indexed by $(s,t) \in K$ with infinite sets K. Is compact differentiability of ϕ together with weak convergence of $(\sqrt{n}(t-s)(\hat{\mathbf{F}}_{\lfloor ns \rfloor, \lfloor nt \rfloor} - \mathbf{F}))_{(s,t)\in K}$ enough to yield process convergence in this setting? A simple example given below shows that in full generality this cannot be true.

Example 2.2. Consider the map ϕ that takes a distribution function to its median. Consider a triangular array of data that is of the form $X_{jn} = n$ for $1 \le j < n^{1/3}$ and $X_{jn} \sim U[0,1]$ i.i.d. for $n^{1/3} \le j \le n$. Let $K = \{0\} \times [0,1]$. Let $\mathbf{F}$ denote the uniform distribution function on $[0,1]$. Elementary calculations show that $(n^{-1/2}\lfloor nt \rfloor(\hat{\mathbf{F}}_{1,\lfloor nt \rfloor} - \mathbf{F}))_{t\in[0,1]}$ converges weakly to the Kiefer-Müller process $\mathbb{K}$ with covariance $\mathrm{Cov}(\mathbb{K}(t,y), \mathbb{K}(t',y')) = \min(t,t')(\min(y,y') - yy')$. On the other hand, setting $t = n^{-3/4}$ we have almost surely

$$\mathbb{W}_n(\cdot; 0, n^{-3/4}) = n^{1/2} n^{-3/4}(\phi(\hat{\mathbf{F}}_{1,\lfloor n^{1/4} \rfloor}) - \phi(\mathbf{F})) = n^{-1/4}(n - 1/2) \to \infty,$$

and thus weak convergence of $\mathbb{W}_n$ cannot hold.

This example demonstrates that employing results from the 'non-sequential' in the 'sequential' setting requires careful consideration. Taking a closer look at what goes wrong in Example 2.2 we see that, due to the scaling, weak convergence of the process $(n^{-1/2}\lfloor nt \rfloor(\hat{\mathbf{F}}_{1,\lfloor nt \rfloor} - \mathbf{F}))_{t\in[0,1]}$ is not very informative about the estimator $\hat{\mathbf{F}}_{1,\lfloor nt \rfloor}$ when $nt \to 0$, so that a pathological behavior of $\hat{\mathbf{F}}_{1,\lfloor nt \rfloor}$ in those instances cannot be excluded. There are two ways of handling this problem. First, we could still hope for sensible results if we restrict our attention to values of $s - t$ that are bounded away from zero. Second, if the map ϕ is bounded, the process $\mathbb{W}_n(\cdot; s, t)$ will still be well-behaved near zero due to the scaling with $\lfloor ns \rfloor - \lfloor nt \rfloor$. Below, we provide a detailed discussion of both settings.

We now give the assumptions that are needed to ensure weak convergence of the process $\mathbb{W}_n$.

The first assumption states that ϕ is smooth in a suitable sense. This assumption is needed to apply the classical functional delta method in the non-sequential setting.

(C) The map

$$\phi : \ell^\infty(D) \supset \boldsymbol{D}_\phi \to \boldsymbol{R}_\phi \subset \ell^\infty(R)$$

is compactly differentiable at $\mathbf{F}$ tangentially to the vector space $V \subset \ell^\infty(D)$. Denote the derivative by $\phi'_{\mathbf{F}}$.

Note that assumption (C) is a completely 'non-sequential' assumption. In the 'non-sequential' 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. Examples include quantiles, copulas, dependence measures, M- and L-estimators to name just a few. 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). The next assumption is an analogue of the weak convergence of the sequential empirical process.

(W) Assume that for

$$\mathbb{Y}_n(v;t) := n^{1/2}t(\hat{\mathbf{F}}_{1,\lfloor nt\rfloor}(v) - \mathbf{F}(v))$$

we have

$$\mathbb{Y}_n \rightsquigarrow \mathbb{Y} \quad \text{in} \quad \ell^\infty(D \times [0,1])$$

where $\mathbb{Y}$ denotes a centered process such that $\omega \mapsto \mathbb{Y}$ is $(\mathcal{A}, \mathcal{B})$-measurable [here, $\mathcal{B}$ denotes the Borel sigma-Algebra on $(\ell^\infty(D \times [0,1]), \|\cdot\|_\infty)$].

Assumption (W) was established in a variety of settings. Results for $d \geq 1$ are derived by Sen (1974) and Rüschendorf (1974) under $\phi-$mixing and by Yoshihara (1975), Inoue (2001) and Bücher (2013) under strong mixing. Additionally, in Section 4.2 we will show how a similar result can be derived under the so-called GMC property which provides an alternative measure of dependence that has been recently introduced in Wu and Shao (2004). The case $d = 1$ has been considered in Berkes et al. (2009) under S-mixing, where the authors derived a stronger result than weak convergence of the process. The paper by Dehling and Taqqu (1989) establishes a condition that is similar to (W) for long-range dependent data with a scaling different from $n^{1/2}$.

Finally, we need some additional assumptions on the sample paths of the limiting processes $\mathbb{Y}$ in condition (W). As discussed below, these assumptions are not very restrictive.

(A1) Assume that $\sup_{|t-t'|\leq\delta} \sup_{v\in D} |\mathbb{Y}(v;t) - \mathbb{Y}(v;t')| = o_P(1)$ as $\delta \to 0$.

(A2) Define the set

$$U_K := \Big\{(h_t)_{t\in[0,1]} : h_t \in V \ \forall\ t \in [0,1], \sup_{t\in[0,1]} \|h_t\| < \infty\Big\}$$

where V is the vector space from condition (C). 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 classical functional delta method for each fixed (s,t). Assumption (A1) is needed for the application of the general compact differentiability result in Section 5. As we shall discuss in Section 4.2, both conditions are satisfied for a wide variety of dependent data.

Remark 2.3. At first glance, it might be surprising that we state conditions (W), (A1), (A2) for the process $\sqrt{n}t(\hat{\mathbf{F}}_{1,\lfloor nt\rfloor} - \mathbf{F})$ indexed by $t \in [0,1]$ instead of $\sqrt{n}(t-s)(\hat{\mathbf{F}}_{\lfloor ns\rfloor,\lfloor nt\rfloor} - \mathbf{F})$ indexed by $(s,t) \in \Delta$. However, some simple arguments [which essentially boil down to observing that $(t-s)\hat{\mathbf{F}}_{\lfloor ns\rfloor,\lfloor nt\rfloor} \approx t\hat{\mathbf{F}}_{1,\lfloor nt\rfloor} - s\hat{\mathbf{F}}_{1,\lfloor ns\rfloor}$] show that this is already sufficient to obtain results for the more general process. More details in a general setting are provided in Section 4.

Assumptions (C), (W), (A1), (A2) are sufficient to establish weak convergence of the process $\mathbb{W}_n(\cdot;s,t)$ if we index it by $(s,t) \in K \subset \Delta$ such that $\inf_{(s,t)\in K}(t-s) > 0$. As we have seen in Example 2.2, those conditions are not sufficient if $\inf_{(s,t)\in K}(t-s) = 0$. In the latter case, we need the following additional condition

(A3) For any $k_n \to 0$ we have $\sup_{(s,t)\in K, |t-s|\le k_n}(t-s)\|\phi(\hat{\mathbf{F}}_{\lfloor ns\rfloor,\lfloor nt\rfloor})\| \to 0$ in outer probability.

Remark 2.4. Note that condition (A3) is automatically satisfied if the quantity $\sup_{(s,t)\in K}\|\phi(\hat{\mathbf{F}}_{\lfloor ns\rfloor,\lfloor nt\rfloor})\|$ is bounded in outer probability. This is trivially true for uniformly bounded maps ϕ, 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. For instance, Remark 4.9 in Section 4.2 shows that (A3) holds if we take $K = \{0\}\times[0,1]$ under the additional assumption that the data $X_1, X_2, \ldots$ form a strictly stationary sequence and if $\hat{\mathbf{F}}_{1,n}$ converges to $\mathbf{F}$ outer almost surely.

We are now ready to state our main result.

Theorem 2.5. *For any compact $K \subset \Delta$ with $\inf_{(s,t)\in K}(t-s) > 0$ conditions (C), (W), (A1) and (A2) imply $\mathbb{W}_n \rightsquigarrow \mathbb{W}$ in $\ell^\infty(R\times K)$ where*

$$\mathbb{W}(u;s,t) := \Big(\phi'_{\mathbf{F}}(\mathbb{Y}(\cdot;t) - \mathbb{Y}(\cdot;s))\Big)(u).$$

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

Informally, one could say that if the map ϕ is bounded, applicability of the delta method in the 'non-sequential' setting implies a stronger, 'sequential' version of this result without additional restrictions. Without boundedness of ϕ, the 'sequential' version follows for certain types of sets K. The usefulness of Theorem 2.5 in various settings is demonstrated in the next section.

Moreover, the findings in this section can be generalized in several directions. First, the collection of empirical distribution functions can be replaced by a general collection of estimators indexed by a compact subset of $\mathbb{R}^d$, or, even more generally, by a compact metric space K. Second, the map ϕ can be defined to live on general metric spaces. Those generalizations are discussed in Section 4.1. Moreover, Section 4.2 contains results that imply condition (W) while in Section 4.3 we briefly discuss bootstrap procedures. Additionally, compact differentiability can be replaced by quasi Hadamard differentiability [see Beutner and Zähle (2010)]. General results in this direction can be found in Section 5.

3. Some examples and applications

In order to demonstrate the power of the results from the previous section, we consider two applications. The first application discusses a recently proposed approach to inference for time series data, also called self-normalization. The second application concerns a generalization of the sequential empirical copula process. Here, we show that results derived in Bücher and Kojadinovic (2013) by long and tedious proofs follow easily with our general theory. Details for an additional application to change-point detection are provided in the Appendix [see Section C].

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 this traditional approach lies in the bandwidth parameter(s) involved in the consistent estimation, which also occurs for other existing approaches, such as subsampling [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 class of 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 the remainder terms. To be precise, assume that we observe data $X_1, \dots, X_n$ that stem from a strictly stationary time series. Let $\theta \in \mathbb{R}^q$ be the quantity of interest which depends on the distribution of X_t, and assume that it can be represented as $\theta = \phi(\mathbf{F})$ with $\mathbf{F}$ denoting the distribution function of X_1. Also, denote by $\hat{\theta}_{1,k}$ an estimator for θ that is computed from the sub-sample $X_1, \dots, X_k$. In what follows, we assume that it is of the form $\hat{\theta}_{1,k} = \phi(\hat{\mathbf{F}}_{1,k})$. Given the above notation, Shao (2010a) assumed that

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

where $\{R_n(k/n)\}_{k=1}^n$ denotes a sequence of remainder terms that are negligible uniformly in k. 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 ϕ, $R_n(1) = o_P(n^{-1/2})$ and

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

where $\Sigma = \sum_{k\in\mathbb{Z}} \mathrm{cov}(L(X_0), L(X_k)) > 0$ 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 θ is then based on estimating the covariance matrix Σ 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 quantity

$$G_n = n(\phi(\hat{\mathbf{F}}_{1,n}) - \theta)^T V_n^{-1}(\phi(\hat{\mathbf{F}}_{1,n}) - \theta),$$

where $V_n = n^{-2}\sum_{j=1}^{n} j^2(\phi(\hat{\mathbf{F}}_{1,j}) - \phi(\hat{\mathbf{F}}_{1,n}))(\phi(\hat{\mathbf{F}}_{1,j}) - \phi(\hat{\mathbf{F}}_{1,n}))^T$ is the self-normalization matrix.

In Shao (2010a,b), the asymptotic distribution of G_n was derived under the following assumptions:

$$\Big(n^{-1/2}\sum_{j=1}^{\lfloor nt\rfloor} L(X_j)\Big)_{t\in[0,1]} \rightsquigarrow \Sigma^{1/2}(\mathbb{B}_t)_{t\in[0,1]}, \tag{3}$$

$$R_n(1) = o_P(n^{-1/2}), \quad n^{-2}\sum_{j=1}^{n} |jR_n(j/n)|^2 = o_p(1) \tag{4}$$

with $(\mathbb{B}_t)_{t\in[0,1]}$ denoting a q-dimensional vector of independent Brownian motions on $[0,1]$. To verify (4), a common approach is to derive a uniform Bahadur representation for $\hat{\theta}_{1,\lfloor nt\rfloor}$ 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

$$G_n \rightsquigarrow U_q := \mathbb{B}_1^T\Big(\int_0^1 \Big(\mathbb{B}_t - t\mathbb{B}_1\Big)\Big(\mathbb{B}_t - t\mathbb{B}_1\Big)^T dt\Big)^{-1}\mathbb{B}_1, \tag{5}$$

where the limiting distribution is pivotal and does not depend on the unknown covariance matrix Σ. Then the SN-based $100(1-\alpha)\%$ confidence region for θ is of the form

$$\{\theta : n(\phi(\hat{\mathbf{F}}_{1,n}) - \theta)^T V_n^{-1}(\phi(\hat{\mathbf{F}}_{1,n}) - \theta) \le U_{q,\alpha}\}, \tag{6}$$

where $U_{q,\alpha}$ is the $100(1-\alpha)$th percentile of the distribution for U_q. See Lobato (2001) for simulated critical values for U_q.

Using the results in Section 4, 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 (5) is weak convergence of the process $(\sqrt{n}t(\phi(\hat{\mathbf{F}}_{1,\lfloor nt\rfloor}) - \theta))_{t\in[0,1]}$ where we make use of the notation from Section 2. In the language of Section 2, this amounts to setting $K = \{0\}\times[0,1]$. Assuming that $\phi(\mathbf{F})$ is an element of $\mathbb{R}^q$, the quantity $\mathbb{W}_n(\cdot; s,t)$ can be viewed as a $\mathbb{R}^q-$valued vector. Some straightforward calculations show that under assumptions (A1)–(A3), (C), (W) the statistic G_n

can be represented as

$$G_n = \mathbb{W}_n(\cdot;0,1)^T \Big[\int_0^1 \Big(\mathbb{W}_n(\cdot;0,t) - t\mathbb{W}_n(\cdot;0,1) \Big) \times \Big(\mathbb{W}_n(\cdot;0,t) - t\mathbb{W}_n(\cdot;0,1) \Big)^T dt \Big]^{-1} \mathbb{W}_n(\cdot;0,1) + o_P(1).$$

An application of Theorem 4.5 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

$$G_n \rightsquigarrow \mathbb{W}(\cdot;0,1)^T \Big[\int_0^1 \Big(\mathbb{W}(\cdot;0,t) - t\mathbb{W}(\cdot;0,1) \Big) \times \Big(\mathbb{W}(\cdot;0,t) - t\mathbb{W}(\cdot;0,1) \Big)^T dt \Big]^{-1} \mathbb{W}(\cdot;0,1).$$

Under the assumption that $\mathbb{W}(\cdot;0,t) = \Sigma^{1/2}\mathbb{B}_t$, 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 4.8.

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

$$V_n(H) := \int_\Delta (\phi(\hat{\mathbf{F}}_{\lfloor ns \rfloor, \lfloor nt \rfloor}) - (t-s)\phi(\hat{\mathbf{F}}_{1,n})) \times (\phi(\hat{\mathbf{F}}_{\lfloor ns \rfloor, \lfloor nt \rfloor}) - (t-s)\phi(\hat{\mathbf{F}}_{1,n}))^T dH(s,t)$$

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

$$W(H) := \int_\Delta \Big(\mathbb{W}(\cdot;s,t) - (t-s)\mathbb{W}(\cdot;0,1) \Big) \Big(\mathbb{W}(\cdot;s,t) - (t-s)\mathbb{W}(\cdot;0,1) \Big)^T dH(s,t).$$

Assuming that $W(H)$ is non-singular almost surely [which is the case as soon as H places mass on sufficiently many different points], the asymptotic distribution of the generalized self-normalized quantity $G_n(H)$ 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 $\{(t,t) | t \in [0,1]\} \subset \Delta$ or that (A3) holds. Additionally, assume that $W(H)$ is non-singular almost surely. Then the generalized SN quantity $G_n(H)$ satisfies*

$$\begin{aligned} G_n(H) &:= \mathbb{W}_n(\cdot;0,1)^T V_n(H)^{-1} \mathbb{W}_n(\cdot;0,1) \\ &\rightsquigarrow \mathbb{W}(\cdot;0,1)^T W(H)^{-1} \mathbb{W}(\cdot;0,1) =: U_q(H). \end{aligned}$$

The confidence region for θ can be constructed similarly as in (6) and the critical values for $U_q(H)$ can be approximated numerically as done in Lobato (2001).

3.2. A general version of the sequential empirical copula process

Denote by X a d-dimensional random vector with continuous distribution function $\mathbf{F}$ and marginal distribution functions $\mathbf{F}_1, \dots, \mathbf{F}_d$. As shown by Sklar (1959), the joint distribution function of the vector X has a unique representation of the form $\mathbf{F}(x) = C(\mathbf{F}_1(x_1), \dots, \mathbf{F}_d(x_d))$ where the function $C : [0,1]^d \to [0,1]$ is called the copula of $\mathbf{F}$. One interpretation of this decomposition is that the copula captures all the dependence between the components of X while being invariant under strictly increasing transformations of the marginals. For this reason, copulas have recently received a lot of attention in various areas such as econometrics [see e.g. Patton (2009)] and environmental modelling [see e.g. Salvadori (2007)]. One question that is of particular interest in the analysis of time series data is whether the dependency structure within a time series is constant over time. To answer this question in a completely non-parametric fashion, empirical copulas that are computed from various fractions of realizations, say $X_1, \dots, X_n$, from the time series must be compared. Here, the empirical copula of the sample $X_1, \dots, X_n$ is defined through

$$C_{1,n}(u_1, \dots, u_d) := \frac{1}{n} \sum_{i=1}^{n} I\{\hat{\mathbf{F}}_{1,n}^{(1)}(X_{i1}) \le u_1, \dots, \hat{\mathbf{F}}_{1,n}^{(d)}(X_{id}) \le u_d\}$$

where $\hat{\mathbf{F}}_{1,n}^{(j)}(y) := n^{-1} \sum_i I\{X_{ij} \le y\}$ is the j'th marginal empirical distribution function evaluated at y. One then needs to consider the following sequential empirical copula process

$$\mathbb{C}_n(u; s, t) := n^{-1/2}(\lfloor nt \rfloor - \lfloor ns \rfloor)(C_{\lfloor ns \rfloor + 1, \lfloor nt \rfloor}(u) - C(u))$$

which was recently introduced by Bücher and Kojadinovic (2013). Here, for $k \ge l$, $C_{k,l}$ denotes the empirical copula of the observations $X_k, \dots, X_l$. Additionally we define $C_{k,k-1} \equiv 0$ for all $k = 1, \dots, n$.

An observation that is very useful in the analysis of empirical copulas is that, given continuity of the marginal distribution functions $\mathbf{F}_1, \dots, \mathbf{F}_d$ of $\mathbf{F}$, the empirical copula of the original sample $X_1, \dots, X_n$ and the transformed sample $Y_1, \dots, Y_n$ with $Y_i := (\mathbf{F}_1(X_{i1}), \dots, \mathbf{F}_d(X_{id}))$ coincide almost surely. In condition (c2) given below it thus is sufficient to consider the empirical distribution function of the Y_i.

In order to establish weak convergence of $\mathbb{C}_n$, Bücher and Kojadinovic (2013) made the following assumptions.

(c1) For any $j \in \{1, \dots, d\}$, the partial derivatives $\dot{C}_j := \partial C / \partial u_j$ of the copula C exist and are continuous on the set $\{u \in [0,1]^d : u_j \in (0,1)\}$.

(c2) The data $X_1, \dots, X_n$ are drawn from a strictly stationary sequence $(X_i)_{i \in \mathbb{Z}}$ with continuous marginal distributions $\mathbf{F}_1, \dots, \mathbf{F}_d$. Define the transformed random variables $Y_i := (\mathbf{F}_1(X_{i1}), \dots, \mathbf{F}_d(X_{id}))$ and denote by $\hat{\mathbf{F}}_{1,\lfloor nt \rfloor}^Y$ the corresponding empirical distribution functions. Assume that with those definitions,

$$\tilde{\mathbb{G}}_n^Y(u; t) := n^{-1/2} \lfloor nt \rfloor (\hat{\mathbf{F}}_{1,\lfloor nt \rfloor}^Y(u) - C(u))$$

converges weakly in $\ell^\infty([0,1]^{d+1})$ to a tight Gaussian process $\mathbb{B}_C$ concentrated on

$$\Big\{f \in \mathcal{C}([0,1]^{d+1}) : f(v) = 0 \text{ if one of the components of } v \text{ is } 0, \\ f(1,\ldots,1;t) = 0\ \forall t \in [0,1]\Big\}.$$

Here $\mathcal{C}([0,1]^{d+1})$ denotes the space of continuous functions on $[0,1]^{d+1}$.

Theorem 2.3 in Bücher and Kojadinovic (2013) states that under (c1) and (c2) the process $\mathbb{C}_n$ converges weakly to the centered Gaussian process [recall that $u^{(j)} = (1,\ldots,1,u_j,1,\ldots,1)$ with u_j at position j]

$$\mathbb{C}_C(u;s,t) := \mathbb{B}_C(u;t) - \mathbb{B}_C(u;s) - \sum_{j=1}^d \dot{C}_j(u)(\mathbb{B}_C(u^{(j)};t) - \mathbb{B}_C(u^{(j)};s)).$$

The proof given in Bücher and Kojadinovic (2013) is long and it involves many technicalities. The reason is that they derive certain smoothness properties of the map

$$\big(\hat{\mathbf{F}}^Y_{\lfloor ns\rfloor+1,\lfloor nt\rfloor}\big)_{s\le t\in[0,1]} \mapsto \big(C_{\lfloor ns\rfloor+1,\lfloor nt\rfloor}\big)_{s\le t\in[0,1]}$$

in a direct way. Here, we demonstrate how a similar result can be obtained by an application of Theorem 2.5. This approach is much simpler since we combine known results about the 'non-sequential' copula map with our general results. In the notation of Section 2, let $D = R = [0,1]^d$ and denote by $\boldsymbol{D}_\phi$ the set of all distribution functions on $[0,1]^d$ that have no mass in zero. Denote by ϕ the 'copula map' taking a distribution function $\mathbf{F}$ on $[0,1]^d$ with no mass at zero to the corresponding copula, i.e. let

$$\phi(\mathbf{F}) := \mathbf{F}(\mathbf{F}_1^{-1},\ldots,\mathbf{F}_d^{-1}), \quad \mathbf{F}_j^{-1}(x) := \inf\{y \in [0,1] : \mathbf{F}_j(y) \ge x\}.$$

Compact differentiability of the map ϕ has been studied, among others, by Fermanian et al. (2004) and Bücher and Volgushev (2013). The last named authors established that ϕ is compactly differentiable at any copula C that satisfies condition (c1) tangentially to the space

$$V := \{f \in \mathcal{C}([0,1]^d) : f(u) = 0 \text{ if one of the components of } u \text{ is } 0, \\ f(1,\ldots,1) = 0\}.$$

The derivative of ϕ at C is given by $(\phi'_C(h))(u) = h(u) - \sum_{j=1}^d \dot{C}_j(u)h(u^{(j)})$. Thus condition (C) from the previous section holds. Now it is easy to see that assumption (c2) implies conditions (W), (A1) and (A2) from the preceding section. Finally, condition (A3) holds since the copula map is bounded by construction. Thus all conditions of Theorem 2.5 are fulfilled and we obtain the weak convergence $\tilde{\mathbb{C}}_n \rightsquigarrow \mathbb{C}_C$ in $\ell^\infty([0,1]^d \times \Delta)$ where we defined

$$\tilde{\mathbb{C}}_n(u;s,t) := n^{-1/2}(\lfloor nt\rfloor - \lfloor ns\rfloor)(\phi(\hat{\mathbf{F}}_{\lfloor ns\rfloor,\lfloor nt\rfloor})(u) - C(u)).$$

Additionally, by arguments similar to the ones given in the proof of Lemma 1 in Kojadinovic and Rohmer (2012), we obtain the bound

$$\sup_{u\in[0,1]^d} |C_{k,l}(u) - \phi(\hat{\mathbf{F}}_{k,l})| \leq \frac{d}{l-k} \quad \forall k < l$$

which shows that

$$\tilde{\mathbb{C}}_n(u; s, t) = n^{-1/2}(\lfloor nt\rfloor - \lfloor ns\rfloor)(C_{\lfloor ns\rfloor,\lfloor nt\rfloor}(u) - C(u)) + o_P(1).$$

The only difference between the right-hand side of the above equation and the process $\mathbb{C}_n(u; s, t)$ of Bücher and Kojadinovic (2013) is the fact that $C_{\lfloor ns\rfloor,\lfloor nt\rfloor}(u)$ is replaced by $C_{\lfloor ns\rfloor+1,\lfloor nt\rfloor}(u)$. However, standard calculations employing the continuity of the sample paths of the limiting process $\mathbb{C}$ imply that

$$\sup_{u,s,t} |C_{\lfloor ns\rfloor,\lfloor nt\rfloor}(u) - C_{\lfloor ns\rfloor+1,\lfloor nt\rfloor}(u)| = o_P(n^{-1/2}).$$

The details are omitted for the sake of brevity.

4. Extension to a general setting

In this section, we provide general versions of the results presented in Section 2. This is motivated by the fact, that the setting of Section 2 excludes some interesting examples. First, it does not allow to handle the Kaplan-Meier estimator [see Example 4.3 for additional details]. Second, some processes such as the sequential empirical copula process considered by Rüschendorf (1976) do not fit in the simple setting of Section 2, see Example 4.4.

To be able to handle those situations, we replace the empirical distribution functions computed from fractions of the data by general collections of estimators that need not be based on sub-samples and need not be indexed by subsets of Δ. We begin by introducing some relevant notation. For arbitrary sets $\mathcal{F}_1, \dots, \mathcal{F}_J, K$ define the vector space

$$\mathcal{L}^\infty(\mathcal{F}_1, \dots, \mathcal{F}_J; K) := \Big\{ (H^{(1)}(\cdot;\kappa), \dots, H^{(J)}(\cdot;\kappa))_{\kappa\in K} \ : \ H^{(j)}(\cdot;\kappa) \in \ell^\infty(\mathcal{F}_j) \forall j, \kappa \sup_\kappa \sup_j \sup_{f\in\mathcal{F}_j} |H^{(j)}(f;\kappa)| < \infty \Big\}$$

with norm

$$\|(H^{(1)}, \dots, H^{(J)})\|_{\mathcal{L}} := \sup_\kappa \sup_j \sup_{f\in\mathcal{F}_j} |H^{(j)}(f;\kappa)|.$$

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

$$\begin{aligned} & (H^{(1)}(\cdot;\kappa), \dots, H^{(J)}(\cdot;\kappa))_{\kappa\in K} \in \mathcal{L}^\infty(\mathcal{F}_1, \dots, \mathcal{F}_J; K) \\ \leftrightarrow \quad & \Big((f,\kappa) \mapsto H^{(1)}(f;\kappa), \dots, (f,\kappa) \mapsto H^{(J)}(f;\kappa) \Big). \end{aligned}$$

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

$$\begin{aligned} &\|(H^{(1)}(\cdot;\kappa),\ldots,H^{(J)}(\cdot;\kappa))_{\kappa\in K}\|_{\mathcal{L}} \\ &= \Big\|\Big((f,\kappa)\mapsto H^{(1)}(f;\kappa),\ldots,(f,\kappa)\mapsto H^{(J)}(f;\kappa)\Big)\Big\|_{\max}. \end{aligned}$$

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

Remark 4.1. In many situations, the sets $\mathcal{F}_1,\ldots,\mathcal{F}_J$ can be viewed as subsets of $\mathbb{R}^d$. For example, the empirical distribution function $(n^{-1}\sum I\{X_i \le y\})_{y\in\mathbb{R}^d}$ of a sample of d-dimensional random variables $X_1,\ldots,X_n$ 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 4.2, is to consider classes of functions $\mathcal{F}_j$. In this setting, the empirical process can be written as $(n^{-1}\sum_{i=1}^n f(X_i) - \mathbb{E}f(X_i))_{f\in\mathcal{F}_j}$, 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 $\mathcal{F} = \{x \mapsto I\{x \le y\} | y \in \mathbb{R}^d\}$. By identifying the function $x \mapsto I\{x \le y\}$ 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.

The rest of this Section is organized as follows. In Section 4.1, we present analytic considerations that can be viewed as a generalization of the findings in Section 2. An overview of existing results regarding sequential empirical processes under dependence as well as extensions of those findings will be considered in Section 4.2. Section 4.3 contains some results on bootstrap approximation. Finally, in Section 4.4 we demonstrate how those findings can be utilized for fixed-b corrections of sub-sampling estimators.

4.1. Analytic considerations

Denote by (K,d_K) a compact metric space. Assume that we are given a collection of estimators $(\hat{\mathbf{G}}_n(\cdot;\kappa))_{\kappa\in K}$ of an $\ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_J)$-valued quantity $\mathbf{G}$. Assume that $\hat{\mathbf{G}}_n(\cdot;\kappa)$ are defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$ [in principle, this space is allowed to depend on n. However, following Van der Vaart and Wellner (1996) we will not stress this in the notation]. Denote by $\phi : \ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_J) \to \ell^\infty(\mathcal{G}_1) \times \ldots \times \ell^\infty(\mathcal{G}_L)$ some smooth map [this will be made precise below]. For a deterministic sequence, say α_n, of real numbers diverging to infinity and a bounded function $w : K \to \mathbb{R}$ consider the following two processes

$$\mathbb{Y}_n(f_1, \dots, f_J; \kappa) := \alpha_n w(\kappa)(\hat{\mathbf{G}}_n(f_1, \dots, f_J; \kappa) - \mathbf{G}(f_1, \dots, f_J))$$

and

$$\mathbb{V}_n(g_1, \dots, g_L; \kappa) := \alpha_n w(\kappa)(\phi(\hat{\mathbf{G}}_n(\cdot; \kappa))(g_1, \dots, g_L) - \phi(\mathbf{G})(g_1, \dots, g_L)).$$

This section is primarily concerned with the following question: given compact differentiability of the map ϕ and weak convergence of the process $\mathbb{Y}_n$ viewed as element of $\mathcal{L}^\infty(\mathcal{F}_1, \dots, \mathcal{F}_J; K)$, what can we say about weak convergence of $\mathbb{V}_n$?

Before proceeding, we illustrate the abstract setting described above with several examples.

Example 4.2. Sequential empirical processes
Assume that we have a sample of identically distributed random variables, say $X_1, \dots, X_n$. Assume that the quantity $\mathbf{G}$ can be represented as

$$\mathbf{G} = \Big((\mathbb{E}[f(X)])_{f \in \mathcal{F}_1}, \dots, (\mathbb{E}[f(X)])_{f \in \mathcal{F}_J}\Big)$$

for some classes of functions $\mathcal{F}_1, \dots, \mathcal{F}_J$ [see, for instance, Example 4.3]. A prime example for the quantity $\hat{\mathbf{G}}_n(\cdot; \kappa)$ with $\kappa = (s,t) \in \Delta$ is given by the estimator computed from the sub-sample $X_{\lfloor ns \rfloor \vee 1}, \dots, X_{\lfloor nt \rfloor \vee 1}$, that is

$$\begin{aligned} \hat{\mathbf{G}}_n(f_1, \dots, f_J; s, t) &:= (\hat{\mathbf{G}}_n^{(1)}(f_1; s, t), \dots, \hat{\mathbf{G}}_n^{(J)}(f_J; s, t)) \\ \hat{\mathbf{G}}_n^{(j)}(\cdot; s, t) &:= \frac{1}{1 + \lfloor nt \rfloor \vee 1 - \lfloor ns \rfloor \vee 1} \Big(\sum_{i = \lfloor ns \rfloor \vee 1}^{\lfloor nt \rfloor \vee 1} f(X_i) \Big)_{f \in \mathcal{F}_j}. \end{aligned} \tag{7}$$

The reason for considering general classes of functions instead of just indicators as we have done in Section 2 is that some interesting estimators are not simply functionals of empirical distribution functions.

Example 4.3. Kaplan-Meier estimator
Assume that we have right-censored observations of the form $(Y_i, \delta_i)_{i=1,\dots,n}$. 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

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

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

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

In particular, this example fits in the present setting while it cannot be handled with the methods from Section 2.

Finally, we would like to point out that, while the setting of Example 4.2 occurs in practise most frequently, there are interesting processes that are of a more general structure.

Example 4.4. Sequential empirical copula process [Rüschendorf (1976)] The 'classical' sequential empirical copula process is of the form

$$\mathbb{C}_n^{\circ}(u;s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor sn \rfloor \vee 1} \Big(I\big\{\hat{\mathbf{F}}_{1,n}^{(1)}(X_{i1}) \le u_1, \dots, \hat{\mathbf{F}}_{1,n}^{(d)}(X_{id}) \le u_d)\big\} - C(u) \Big)$$

where $\hat{\mathbf{F}}_{1,n}^{(j)}(y) := n^{-1}\sum_i I\{X_{ij} \le y\}$ is the j'th marginal empirical distribution function evaluated at y [see also Section 3.2] and C is the copula of the distribution of X. Note that $\hat{\mathbf{F}}_{1,n}^{(j)}$ depends on all the data regardless of the value of s, so that the estimator $\hat{\mathbf{G}}_n(u;\kappa) := \frac{1}{\lfloor \kappa n \rfloor \vee 1}\sum_{i=1}^{\lfloor \kappa n \rfloor \vee 1} I\{\hat{\mathbf{F}}_{1,n}^{(1)}(X_{i1}) \le u_1, \dots, \hat{\mathbf{F}}_{1,n}^{(d)}(X_{id}) \le u_d)\}$ is not of the form (7). Still, the process $\mathbb{C}_n^{\circ}(u;s)$ can be coerced into the general framework of this section by considering the collection of estimators $\hat{\mathbf{G}}_n(\cdot;\kappa)$ by setting $K = [0,1]$ and $\mathcal{F}_1 := [0,1]^d$. 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 (2013).

We now proceed to state the assumptions needed for the general result in this section. Regarding the smoothness of ϕ, we impose the following condition which corresponds to assumption (C)

(Cg) Denote by $\mathcal{G}_1, \dots, \mathcal{G}_L$ arbitrary sets. The map

$$\phi : \ell^{\infty}(\mathcal{F}_1) \times \ldots \times \ell^{\infty}(\mathcal{F}_J) \supset \boldsymbol{D}_{\phi} \to \boldsymbol{R}_{\phi} \subset \ell^{\infty}(\mathcal{G}_1) \times \ldots \times \ell^{\infty}(\mathcal{G}_L).$$

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

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

(Wg) Assume that

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

where

$$\mathbb{Y}(f_1, \dots, f_J; \kappa) := (\mathbb{Y}^{(1)}(f_1;\kappa), \dots, \mathbb{Y}^{(J)}(f_J;\kappa))$$

and $\mathbb{Y}^{(j)}, j = 1, \dots, J$ are centered, Borel measurable random elements of $\ell^{\infty}(\mathcal{F}_j \times K)$.

A detailed discussion of condition (Wg) for estimators $\hat{\mathbf{G}}_n(\cdot;\kappa)$ of the form (7) is provided in the next section.

The limit $\mathbb{Y}$ in assumption (Wg) needs to satisfy certain technical conditions that correspond to assumptions (A1)–(A3) from Section 2.

(Ag1) Assume that $\sup_{d_K(\kappa,\kappa')\le\delta}\sup_j\sup_{f_j\in\mathcal{F}_j}|\mathbb{Y}^{(j)}(f_j;\kappa)-\mathbb{Y}^{(j)}(f_j;\kappa')| = o_P(1)$ as $\delta\to 0$.

(Ag2) Define the set

$$U_K := \Big\{(h_\kappa)_{\kappa\in K} : h(\cdot;\kappa)\in V\ \forall\ \kappa\in K, \sup_{\kappa\in K}\|h(\cdot;\kappa)\|<\infty\Big\}.$$

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

(Ag3) For any $k_n \to 0$ we have $\sup_{\kappa\in K,|w(\kappa)|\le k_n} w(\kappa)\|\phi(\hat{\mathbf{G}}_n(\cdot;\kappa))\| = o_P^*(1)$ where the asterisk denotes outer probability. Additionally, it holds that $\sup_{|w(\kappa)|\le\delta}\|\mathbb{Y}(\cdot;\kappa)\| = o_P(1)$ as $\delta\to 0$.

Condition (Ag2) is nonrestrictive in the sense that it is needed to apply the functional delta method to $\mathbb{V}_n(\cdot;\kappa)$ for each fixed κ. Assumption (Ag1) is needed for the application of the general compact differentiability result in Section 5. As we shall discuss in the next section [see Remark 4.8], assumption (Ag1) is typically satisfied in a wide variety of practically relevant settings. As in Section 2, assumptions (Wg), (Ag1), (Ag2) are already sufficient to derive weak convergence of $\mathbb{V}_n$ if the set K satisfies $\inf_{\kappa\in K}|w(\kappa)|>0$. If we want to drop this condition, we additionally need (Ag3).

Our main result is a generalization of Theorem 2.5 from Section 2

Theorem 4.5. *Assume that the function w is uniformly bounded. For any compact metric space (K,d_K), with $\inf_{\kappa\in K}|w(\kappa)|\ge a>0$ conditions (Cg), (Wg), (Ag1) and (Ag2) imply $\mathbb{V}_n \rightsquigarrow \mathbb{V}$ in $\mathcal{L}^\infty(\mathcal{G}_1,\dots,\mathcal{G}_L;K)$ where*

$$\mathbb{V}(g_1,\dots,g_L;\kappa) := \Big(\phi'_{\mathbf{G}}\mathbb{Y}(\cdot;\kappa)\Big)(g_1,\dots,g_L).$$

If additionally (Ag3) holds, the assumption $\inf_{\kappa\in K}|w(\kappa)|\ge a>0$ can be dropped.

Remark 4.6. Although assumption (Ag3) 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 ϕ denotes the map that takes a distribution function to its median. In that case, assumption (Ag3) would require that $\frac{1}{n}\max_{i=1,\dots,n}|X_i| = o_P(1)$ 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 $\mathbb{V}_n$

$$\tilde{\mathbb{V}}_n(g_1,\dots,g_L;\kappa) := w(\kappa)I\{|w(\kappa)|\ge\alpha_n^{-\gamma}\}\alpha_n(\phi(\hat{\mathbf{G}}_n(\cdot;\kappa))-\phi(\mathbf{G}))(g_1,\dots,g_L)$$

converges to the same limit $\mathbb{V}$ without assumption (Ag3) or the additional condition $\inf_{\kappa\in K}|w(\kappa)|\ge a>0$. 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.

Finally, we remark that applying Theorem 4.5 instead of Theorem 2.5 directly leads to an extension of the findings in Section 3 to the more general setting considered in the present section.

4.2. 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

$$\mathbb{Y}_n(f_1,\dots,f_J;s,t) := \Big(n^{1/2}(t-s)(\hat{\mathbf{G}}_n(f_1,\dots,f_J;s,t) - \mathbf{G}(f_1,\dots,f_J))\Big)_{(s,t)\in\Delta}$$

where $\Delta := \{(s,t)\in[0,1]^2 | s\le t\}$ and the quantity $\hat{\mathbf{G}}_n(f_1,\dots,f_J;s,t) := (\hat{\mathbf{G}}_n^{(1)}(f_1;s,t),\dots,\hat{\mathbf{G}}_n^{(J)}(f_J;s,t))$ with

$$\hat{\mathbf{G}}_n^{(j)}(\cdot;s,t) := \Big(\frac{1}{1+\lfloor nt\rfloor\vee 1 - \lfloor ns\rfloor \vee 1}\sum_{i=\lfloor ns\rfloor\vee 1}^{\lfloor nt\rfloor\vee 1} f(X_i)\Big)_{f\in\mathcal{F}_j}, \quad j=1,\dots,J$$

denotes an estimator for $\mathbf{G}$ that is based on the sub-sample $X_{\lfloor ns\rfloor\vee 1},\dots,X_{\lfloor nt\rfloor\vee 1}$. It turns out that conditions (Wg), (Ag1), (Ag2) in the previous section can be derived from simpler conditions that involve only a collection of 'classical' one-parameter sequential processes

$$\mathbb{G}_n(f_1,\dots,f_J;t) := (\mathbb{G}_n^{(1)}(f_1;t),\dots,\mathbb{G}_n^{(J)}(f_J;t))$$

where $\mathbb{G}_n^{(j)}(f;t) := n^{1/2}t(\hat{\mathbf{H}}_n^{(j)}(f;t) - \mathbf{G}(f))$ and

$$\hat{\mathbf{H}}_n^{(j)}(f;t) := \frac{1}{\lfloor nt\rfloor\vee 1}\sum_{i=1}^{\lfloor nt\rfloor\vee 1} f(X_i), \quad j=1,\dots,J.$$

Additionally, let

$$\hat{\mathbf{H}}_n(f_1,\dots,f_J;t) := (\hat{\mathbf{H}}_n^{(1)}(f_1;t),\dots,\hat{\mathbf{H}}_n^{(J)}(f_J;t)). \tag{8}$$

Consider the following assumptions

(W') Assume that
$$\mathbb{G}_n \rightsquigarrow \mathbb{G} \quad \text{in} \quad \mathcal{L}^\infty(\mathcal{F}_1,\dots,\mathcal{F}_J;[0,1])$$
where
$$\mathbb{G}(f_1,\dots,f_J;t) := (\mathbb{G}^{(1)}(f_1;t),\dots,\mathbb{G}^{(J)}(f_J;t))$$
and $\mathbb{G}^{(j)}, j=1,\dots,J$ are centered, Borel measurable random elements in $\ell^\infty(\mathcal{F}_j)$.

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

The conditions above turn out to be sufficient for (Wg) and (Ag1).

Proposition 4.7. *Under conditions (W') and (A1'), we have*

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

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

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

Moreover, $\mathbb{Y}$ *satisfies assumption (Ag1).*

Remark 4.8. For many types 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

$$\mathrm{Cov}(\mathbb{G}(f_1, \dots, f_J; s), \mathbb{G}(g_1, \dots, g_J; t)) = (s \wedge t)\eta(f_1, \dots, f_J, g_1, \dots, g_J)$$

for some uniformly bounded covariance kernel η. In this case, assumption (A1') holds. To see this, note that under (W') and Gaussianity of the limit, the process $\mathbb{G}$ has paths that are uniformly continuous with respect to the metric $\rho_2((s, f_1, \dots, f_J), (t, g_1, \dots, g_J)) := \mathbb{E}[(\mathbb{G}(f_1, \dots, f_J; s) - \mathbb{G}(g_1, \dots, g_J; t))^2]$, 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 $\mathbb{Y}$ implies that its sample paths have the same property.

Remark 4.9. Consider the special case $K = \{0\} \times [0,1]$. In this case, assumption (Ag3) is satisfied as soon as $\hat{\mathbf{H}}_n$ is of the form given in (8) with the data $X_1, X_2, \dots$ stemming from a strictly stationary sequence and $\hat{\mathbf{H}}_n(\cdot; 1) \to \mathbf{G}$ outer almost surely. To see this, note that under the assumptions discussed above we have $\sup_t \|\phi(\hat{\mathbf{H}}_n(\cdot; t))\| = \max_{j=1,\dots,n} \|\phi(\hat{\mathbf{H}}_j(\cdot; 1))\|$ and that by the continuous mapping theorem $\|\phi(\hat{\mathbf{H}}_n(\cdot; 1))\| \to \|\phi(\mathbf{G})\|$ outer almost surely. This in turn implies that $(\sup_{n \geq 1} \|\phi(\hat{\mathbf{H}}_n(\cdot; 1))\|)^*$ [the asterisk denoting a measurable majorant] is bounded in probability. For results implying almost sure convergence of processes 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 $\mathcal{F}_1, \dots, \mathcal{F}_J$ 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 the form $\mathcal{F}_1 = \{u \mapsto I\{u \leq y\} | y \in \mathbb{R}^d\}$ [the inequality is understood componentwise], and a list of corresponding results can be found in Section 2. To the best of our knowledge, nothing is known for general classes of functions. Such results can be established based on Theorem 4.10 provided below.

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 $j = 1, \dots, J$. 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 $(\mathbb{G}_n^{(j)}(f_{ij}; s_{ij}))_{j=1,\ldots,J,i=1,\ldots,N}$ converges weakly to $(\mathbb{G}^{(j)}(f_{ij}; s_{ij}))_{j=1,\ldots,J,i=1,\ldots,N}$ 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 4.10. *Assume that the process $\mathbb{G}_n$ is of the form*

$$\mathbb{G}_n(v;t) = n^{1/2} t(\hat{\mathbf{H}}_n(v;t) - \mathbf{G}(v))$$

where $\hat{\mathbf{H}}_n(\cdot;t)$ is defined in (8) 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 ρ_j on $\mathcal{F}_j$ which makes $\mathcal{F}_j$ totally bounded. Define $\mathcal{F}_{j,\delta} := \{f - g | f, g \in \mathcal{F}_j, \rho_j(f,g) \le \delta\}$. Assume that the process $\mathbb{G}_n^{(j)}(\cdot;1)$ satisfies for some $q > 2$ and $j = 1, \ldots, J$

$$\lim_{\delta \downarrow 0} \limsup_{n\to\infty} \mathbb{E}^* \sup_{f \in \mathcal{F}_{j,\delta}} \|\mathbb{G}_n^{(j)}(f;1)\|^q = 0 \tag{9}$$

[recall that the asterisk denotes outer expectation], that

$$\max_{j=1,\ldots,J} \sup_{n \in \mathbb{N}} \sup_{f \in \mathcal{F}_j} \mathbb{E}^* \|\mathbb{G}_n^{(j)}(f;1)\|^q < \infty, \tag{10}$$

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 $\mathcal{L}^\infty(\mathcal{F}_1, \ldots, \mathcal{F}_J; [0,1])$.

Condition (9) has been established by Andrews and Pollard (1994) for strongly mixing data, and inequality (3.1) in Andrews and Pollard (1994) reveals that (10) holds under the same assumption. Moreover Hagemann (2012) established (9) for stationary sequences with geometric moment contraction properties [see Wu and Shao (2004)], and the results in his appendix show that again (10) holds under the same assumptions.

4.3. Some comments on bootstrap procedures

In this section, we briefly discuss bootstrap procedures. In particular, we consider the following bootstrap version of the quantity $\hat{\mathbf{G}}_n$ defined in (7)

$$\hat{\mathbf{G}}_n^B(f_1,\dots,f_J;s,t) := (\hat{\mathbf{G}}_n^{(1),B}(f_1;s,t),\dots,\hat{\mathbf{G}}_n^{(J),B}(f_J;s,t)), \tag{11}$$
$$\hat{\mathbf{G}}_n^{(j),B}(\cdot;s,t) := \frac{1}{1+\lfloor nt\rfloor\vee 1-\lfloor ns\rfloor\vee 1}\Big(\sum_{i=\lfloor ns\rfloor\vee 1}^{\lfloor nt\rfloor\vee 1} M_{i,n}f(X_i)\Big)_{f\in\mathcal{F}_j},$$

with $M_{1,n},\dots,M_{n,n}$ denoting a triangular array of random variables independent of the original sample $X_1,\dots,X_n$. The corresponding bootstrap version of the process $\mathbb{Y}_n$ is given by

$$\mathbb{Y}_n^B := \alpha_n(t-s)(\hat{\mathbf{G}}_n^B(\cdot;s,t)-\hat{\mathbf{G}}_n(\cdot;s,t)) =: (\mathbb{Y}_n^{(1),B},\dots,\mathbb{Y}_n^{(J),B}). \tag{12}$$

In the literature, this approach is known under the name 'multiplier bootstrap', see for example Section 2.9 in Van der Vaart and Wellner (1996) for the case of independent multipliers and Bühlmann (1993) for an extension to the setting of dependent observations. Under suitable assumptions on the data and random variables $M_{1,n},\dots,M_{n,n}$, a conditional version of assumption (W) holds [see Remark 4.13 for a more detailed discussion]. Specifically, assume that

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

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

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

(i) $\sup_{f\in \mathrm{BL}_1}\left|\mathbb{E}_M f(\mathbb{Y}_n^B)-\mathbb{E}f(\mathbb{Y})\right|\to 0$ in outer probability,

(ii) $\mathbb{E}_M f(\mathbb{Y}_n^B)^* - \mathbb{E}_M f(\mathbb{Y}_n^B)_* \xrightarrow{\mathbb{P}} 0$ for all $f\in \mathrm{BL}_1$,

where BL_1 denotes the set of all functions $f:\mathcal{L}^\infty(\mathcal{F}_1,\dots,\mathcal{F}_J;K)\to\mathbb{R}$ that are bounded by 1 and are Lipschitz-continuous with Lipschitz constants bounded by 1, and where the asterisks in (ii) denote measurable majorants (and minorants, respectively) with respect to the joint data $(X_1,\dots,X_n,M_{1,n},\dots,M_{n,n})$. Also, note that the map $(M_{1,n},\dots,M_{n,n})\mapsto \mathbb{Y}_n^B$ is measurable conditionally on the original data $X_1,\dots,X_n$ outer almost surely [for fixed $X_1,\dots,X_n$, 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 Remark 4.13.

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

$$\alpha_n(t-s)(\phi(\hat{\mathbf{G}}_n^B(\cdot;s,t))-\phi(\hat{\mathbf{G}}_n(\cdot;s,t))) \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_{\mathbf{G}}\mathbb{Y}(\cdot;s,t) \text{ in } \ell^\infty(\mathcal{G}_1)\times\ldots\times\ell^\infty(\mathcal{G}_L)$$

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 4.5 to conditional weak convergence in $\mathcal{L}^\infty(\mathcal{G}_1,\ldots,\mathcal{G}_L;K)$.

Theorem 4.11. *With the notation above, assume that (WB), (Ag1), (Ag2) and (Cg) hold. Then for any compact $K \subset \Delta$ with $\inf_{(s,t)\in K}|t-s|>0$ we have for*

$$\mathbb{V}_n^B(\cdot;s,t) := \alpha_n(t-s)(\phi(\hat{\mathbf{G}}_n^B(\cdot;s,t)) - \phi(\hat{\mathbf{G}}_n(\cdot;s,t)))$$

that

$$\mathbb{V}_n^B \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \phi'_{\mathbf{G}}\mathbb{Y} = \mathbb{V} \quad \textit{in} \quad \mathcal{L}^\infty(\mathcal{G}_1,\ldots,\mathcal{G}_L;K).$$

If additionally (Ag3) holds and $\sup_{(s,t)\in K,|t-s|\le k_n}(t-s)\|\phi(\hat{\mathbf{G}}_n^B(\cdot;s,t))\| = o_P^*(1)$, the convergence holds for arbitrary compact $K \subset \Delta$.

Remark 4.12. Suitable modifications of the extension discussed in Remark 4.6 continue to hold in the bootstrap setting. More precisely, conditional weak convergence of the process

$$\tilde{\mathbb{V}}_n^B = \Big(\alpha_n(t-s)I\{|t-s|>\alpha_n^{-\gamma}\}(\phi(\hat{\mathbf{G}}_n^B(\cdot;s,t)) - \phi(\hat{\mathbf{G}}_n(\cdot;s,t))\Big)_{(s,t)\in\Delta}$$

holds without additional assumptions.

We conclude this section by providing some discussion of settings where condition (WB) holds.

Remark 4.13. In the case of independent data, a mild assumption on the multipliers $M_{i,n}$ suffices. More precisely, assuming that $M_{1,n},\ldots,M_{n,n} = M_1,\ldots,M_n$ where 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 $\mathcal{F}_1,\ldots,\mathcal{F}_J$ being Donsker implies (WB). To see this, note that by arguments similar to the ones given in the proof of Proposition 4.7 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 Bühlmann (1993). More precisely, those authors proposed to consider variables $M_{1,n}, \ldots, M_{n,n}$ from a triangular scheme that satisfy certain conditions [see assumptions A1–A3 in their paper]. In particular, the results in Bücher and Kojadinovic (2013) imply (WB) for $K = \{0\} \times [0,1]$ under strong mixing conditions for the class of functions $\mathcal{F} = \{u \mapsto I\{u \le w\} | w \in \mathbb{R}^d\}$. Moreover, using the techniques in that paper, it should be possible to derive (WB) for more general classes of functions $\mathcal{F}$ and $K = \{0\} \times [0,1]$ 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 (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.

4.4. An application to sub-sampling 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 the 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 $b = l/n$. From now on, we will assume that we have a sample of data $X_1, \ldots, X_n$ from a strictly stationary time series. The estimator $\hat{\mathbf{G}}_n(\cdot; s, t)$ is assumed to be based on the sub-sample $X_{\lfloor ns \rfloor \vee 1}, \ldots, X_{\lfloor nt \rfloor \vee 1}$, i.e. of the form given in equation (7). In what follows, write $\theta = \phi(\mathbf{G})$ for the parameter of interest. For notational convenience, we also consider the quantity $\hat{\theta}_{k,j}$ which is computed from the data $X_k, X_{k+1}, \ldots, X_j$. Note that $\hat{\theta}_{k,j} = \phi(\hat{\mathbf{G}}(\cdot; k/n, j/n))$. For simplicity, assume that θ is $\mathbb{R}^d$-valued. Defining $N := n - l + 1$, the sub-sampling based estimator of the distribution function of $\|\sqrt{n}(\hat{\theta}_{1,n} - \theta)\|$ evaluated at x is given by

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

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

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

Note that under the conditions $l/n + 1/l = o(1)$ and additional regularity assumptions, the asymptotic distribution of $\hat{p}_n(b)$ is $U[0,1]$, 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

$$G(b) = (1-b)^{-1} \int_0^{1-b} I\{\|\Sigma^{1/2}\mathbb{B}_1\| \le \|\Sigma^{1/2}(\mathbb{B}_{b+t} - \mathbb{B}_t - b\mathbb{B}_1)\|/\sqrt{b}\}dt$$

[here, $\mathbb{B}$ denotes a vector of independent Brownian motions on [0,1] and Σ is a positive definite matrix] was derived in Shao and Politis (2013) by assuming that

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

that a similar representation holds for $\hat{\theta}_{1,n}$, that (3) 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 4 and let conditions (Cg), (Wg), (Ag1) and (Ag2) hold. We apply Theorem 4.5 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 $\mathbb{V}(\cdot;s,t) = (t-s)\Sigma^{1/2}(\mathbb{B}_t - \mathbb{B}_s)$ with Σ denoting a non-singular matrix and $\mathbb{B}$ a vector of independent Brownian motions [see the arguments in Shao and Politis (2013)], which typically holds 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

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

where the negligibility of remainder follows from an application of the continuous mapping theorem. The results in Theorem 4.5 in combination with the continuous mapping theorem thus yield

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

as soon as assumptions (Cg), (Wg), (Ag1) and (Ag2) hold.

Unless θ 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 4.2. More precisely, consider a bootstrap version for $\hat{\mathbf{G}}_n(\cdot; s, t)$ which is of the form given in (11) and denote it by $\hat{\mathbf{G}}_n^B(\cdot; s, t)$. Define a bootstrap version for $\phi(\hat{\mathbf{G}}_n(\cdot; s, t))$ through $\phi(\hat{\mathbf{G}}_n^B(\cdot; s, t))$. Assume that the map ϕ is continuous. Now Theorem 4.11 combined with the continuous mapping theorem for the bootstrap in probability [see Theorem 10.8 in Kosorok (2008)] directly yields that if we additionally assume condition (WB), it follows that

$$\hat{p}_n^B(b) := \frac{1}{1-b}\int_0^{1-b} I\Big\{\sqrt{n}\|\phi(\hat{\mathbf{G}}_n^B(\cdot;0,1)) - \theta_0\| \\ \leq \sqrt{nb}\|\phi(\hat{\mathbf{G}}_n^B(\cdot;t,t+b)) - \phi(\hat{\mathbf{G}}_n^B(\cdot;0,1))\|\Big\}dt \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} P.$$

Finally, note that the reasoning above does not rely on θ_0 being $\mathbb{R}^p$-valued and that it is thus also possible to handle infinite dimensional parameters.

5. A general result on (quasi) Hadamard differentiability

This section contains an abstract result on compact differentiability is of independent interest. It plays a crucial role in the proofs of Theorems 4.5 and 4.11. 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 5.1 (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 ϕ'_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 $\phi'_x : C_0 \to R$ denoting a continuous map.

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 ϕ'_x. Let (K, d_K) be a compact metric space. Define the sets

$$\mathscr{D} := \Big\{(h(\cdot;\kappa))_{\kappa\in K}\Big|h(\cdot;\kappa) \in D\ \forall\kappa\Big\}$$

$$\mathscr{R} := \Big\{(h(\cdot;\kappa))_{\kappa\in K}\Big|h(\cdot;\kappa) \in R\ \forall\kappa,\ \sup_{\kappa,\kappa'} d_R(h(\cdot;\kappa), h(\cdot;\kappa')) < \infty\Big\}$$

$$\mathscr{D}_\Phi := \Big\{(h(\cdot;\kappa))_{\kappa\in K}\Big|h(\cdot;\kappa) \in D_\phi\ \forall\kappa,\ \sup_{\kappa,\kappa'} d_R(\phi(h(\cdot;\kappa)), \phi(h(\cdot;\kappa'))) < \infty\Big\}$$

$$\mathscr{D}_0 := \Big\{(h(\cdot;\kappa))_{\kappa\in K}\Big|h(\cdot;\kappa) \in D_0\ \forall\kappa,\ \sup_{\kappa,\kappa'} d_D(h(\cdot;\kappa), h(\cdot;\kappa')) < \infty\Big\}$$

$$\mathscr{C}_0 := \Big\{(h(\cdot;\kappa))_{\kappa\in K}\Big|h(\cdot;\kappa) \in C_0\ \forall\kappa,\ \sup_{\kappa,\kappa'} d_D(h(\cdot;\kappa), h(\cdot;\kappa')) < \infty\Big\}$$

On the sets $\mathscr{R}$ and $\mathscr{D}_0$, define the metrics

$$d_{\mathscr{R}}((h(\cdot;\kappa))_{\kappa\in K}, (g(\cdot,\kappa))_{\kappa\in K}) := \sup_t d_R(h(\cdot;\kappa), g(\cdot,\kappa))$$

and

$$d_{\mathscr{D}_\Phi}((h(\cdot;\kappa))_{\kappa\in K}, (g(\cdot,\kappa))_{\kappa\in K}) := \sup_t d_D(h(\cdot;\kappa), g(\cdot,\kappa)),$$

respectively. For elements $(h(\cdot;\kappa))_{\kappa\in K}, (g(\cdot,\kappa))_{\kappa\in K}$ set

$$(h(\cdot;\kappa))_{\kappa\in K} + a(g(\cdot,\kappa))_{\kappa\in K} := (h(\cdot;\kappa) + ag(\cdot,\kappa))_{\kappa\in K}$$

and assume that with this definition, $(\mathscr{D}_0, d_{D,\Phi})$ and $(\mathscr{R}, d_{R,\Phi})$ are metrized topological vector spaces. Define the map

$$\Phi : \begin{cases} \mathscr{D}_\Phi & \to & \mathscr{R}_\Phi \\ (h(\cdot;\kappa))_{\kappa\in K} & \mapsto & (\phi(h(\cdot;\kappa)))_{\kappa\in K}. \end{cases}$$

Theorem 5.2. *Under setup (S) the map Φ is quasi-Hadamard differentiable at $X := (x)_{\kappa\in K}$ tangentially to $\mathscr{U}\langle\mathscr{D}_0\rangle$ where*

$$\mathscr{U} := \Big\{(h(\cdot;\kappa))_{\kappa\in K} : h(\cdot;\kappa) \in C_0\ \forall\kappa \in K, \sup_{d_K(\kappa,\kappa')\le\delta} d_D(h(\cdot,\kappa'), h(\cdot;\kappa)) = o(1) \text{ as } \delta \to 0\Big\}$$

and the derivative is given by

$$\Phi'_X : \begin{cases} \mathscr{C}_0 & \to & \mathscr{R} \\ (h(\cdot;\kappa))_{\kappa\in K} & \mapsto & (\phi'_x h(\cdot;\kappa))_{\kappa\in K}. \end{cases}$$

If additionally the map ϕ'_x is linear, then so is the map Φ'_X.

Note that Beutner and Zähle (2010) do not assume the derivative map ϕ'_x to be linear. Additionally, as pointed out by a referee, a closer look at the proof of the functional delta method [see e.g. Theorem 3.9.4 in Van der Vaart and Wellner (1996)] shows that linearity of the derivative map is in fact not required. On the other hand, linearity of the derivative enters our proofs at several places and also seems necessary for the bootstrap version of the functional delta method. While we believe that some of our results can be generalized to settings where the derivative is not linear, we do not pursue this question further in order to not make the technical proofs even more involved.

Example 5.3. As an illustration of the above result, let us consider the specific setting of Section 4.1 where $(R, d_R) = (\ell^\infty(\mathcal{G}_1) \times \ldots \times \ell^\infty(\mathcal{G}_L), \|\cdot\|_{\max})$, $D = \ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_J)$, $d_D((f_1, \ldots, f_J), (f'_1, \ldots, f'_J)) := \max_j \|f_j - f'_j\|_\infty$. Assume that ϕ satisfies condition (Cg). Consider the map

$$\Psi_K : \begin{cases} \mathscr{D}_\Psi & \to & \mathscr{R}_\Psi \\ (h(\cdot;\kappa))_{\kappa\in K} & \mapsto & \Big(w(\kappa)\phi\Big(\frac{h(\cdot;\kappa)}{w(\kappa)}\Big)\Big)_{\kappa\in K} \end{cases}$$

where (K, d_K) is a compact metric space, $\mathscr{R}_\Psi \subset \mathcal{L}^\infty(\mathcal{G}_1, \ldots, \mathcal{G}_L; K)$ and

$$\mathscr{D}_\Psi := \Big\{(h(\cdot;\kappa))_{\kappa\in K}\Big|\frac{h(\cdot;\kappa)}{w(\kappa)} \in D_\phi \forall\kappa \in K, \sup_{\kappa\in K}\Big\|w(\kappa)\phi\Big(\frac{h(\cdot;\kappa)}{w(\kappa)}\Big)\Big\| < \infty\Big\}.$$

Note that with this definition, $\mathbb{V}_n(\cdot;\kappa) = \alpha_n(\Psi_K(w(\kappa)\hat{\mathbf{G}}_n(\cdot;\kappa)) - \Psi_K(X_K(\cdot;\kappa)))$ where we defined the map $X_K := ((f_1, \ldots, f_J, \kappa) \mapsto w(\kappa)\mathbf{G}(f_1, \ldots, f_J))$. Additionally we have $\mathbb{Y}_n(\cdot;\kappa) = \alpha_n(w(\kappa)\hat{\mathbf{G}}_n(\cdot;\kappa) - X_K(\cdot;\kappa))$. As long as it holds that $\inf_{\kappa\in K} |w(\kappa)| > 0$, compact differentiability of Ψ_K with derivative

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

is a direct consequence of Theorem 5.2 [here, $\mathscr{V}_K$ is defined similarly to $\mathscr{U}$ with C_0 replaced by V]. To see this, consider a sequence of real numbers $r_n \searrow 0$ and $h_n \in \mathscr{D}$ such that $X_K + r_n h_n \in \mathscr{D}_\Psi$ for all $n \in \mathbb{N}$ with $h_n \to h \in \mathscr{V}_K$. Then, by compact differentiability of Φ,

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

where

$$\begin{aligned} \tilde{X}_K &:= ((f_1, \ldots, f_J, \kappa) \mapsto \mathbf{G}(f_1, \ldots, f_J)), \\ \tilde{h}_n &:= ((f_1, \ldots, f_J, \kappa) \mapsto h_n(f_1, \ldots, f_J; \kappa)/w(\kappa)), \\ \tilde{h} &:= ((f_1, \ldots, f_J, \kappa) \mapsto h(f_1, \ldots, f_J; \kappa)/w(\kappa)). \end{aligned}$$

Finally, observing that Φ'_K is linear, compact differentiability of Ψ_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.

Acknowledgements

Stanislav Volgushev was supported by the DFG grant Vo1799/1-1 and by the Sonderforschungsbereich 823 of the DFG (projects A1, C1). Xiaofeng Shao was supported in part by NSF grant DMS-11-04545. This research was conducted while Volgushev was visiting the University of Illinois at Urbana-Champaign. He would like to thank the people at the Statistics and Economics departments for their hospitality. The authors would also like to thank Axel Bücher and Holger Dette for comments on a preliminary version of this manuscript. Furthermore, we thank the editor and two anonymous referees for their constructive comments that helped to considerably improve the presentation of this manuscript.

Appendix A: Proofs of main results

Proof of Theorem 2.5. Theorem 2.5 is a consequence of the more general result Theorem 4.5. This can be seen by considering the weight function $w((s,t)) := t-s$, setting $J=1$, defining $\mathcal{F}_1 := D$, and considering the collection of estimators $\hat{\mathbf{G}}_n(u;s,t) := \hat{\mathbf{F}}_{\lfloor ns \rfloor, \lfloor nt \rfloor}(u)$. In particular, in this case the space of functions $\mathcal{L}(\mathcal{F}_1;K)$ can be identified with $\ell^\infty(\mathcal{F}_1 \times K)$. Assumption (Cg) is a direct consequence of condition (C). Condition (Wg) follows from (W) by the results in Proposition 4.7 while (Ag1) and (Ag2) are direct consequences of (A1) and (A2), respectively. Similarly, the first part of (Ag3) follows directly from (A3) while the statement $\sup_{\kappa:|w(\kappa)|\le\delta} \|\mathbb{Y}(\cdot;\kappa)\| = o_P(1)$ as $\delta \to 0$ follows since $|w((s,t))| \le \delta$ is equivalent to $|t-s| \le \delta$ and since by construction $\mathbb{Y}(\cdot;t,t) \equiv 0$ almost surely. □

Proof of Theorem 4.5. The proof consists of two steps. First, we show that the convergence holds for K, w with $\inf_{\kappa\in K} |w(\kappa)| > 0$, and second, we extend the result to general K, w under assumption (Ag3). 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 the discussion in Example 5.3.

For the second step, consider the approximating processes

$$\begin{aligned} A_{i,n} &:= \alpha_n w(\kappa) I\{|w(\kappa)| \ge 1/i\}(\phi(\hat{\mathbf{G}}_n(\cdot;\kappa)) - \phi(\mathbf{G})) \\ A_i &:= \phi'_{\mathbf{G}}\big(\mathbb{Y}(\cdot;\kappa) I\{|w(\kappa)| \ge 1/i\}\big). \end{aligned}$$

It then suffices to verify the following three statements [see Lemma B.1 in Bücher et al. (2011)]

(i) For every $i \in \mathbb{N}: \; A_{i,n} \; \rightsquigarrow \; A_i \;$ for $n \to \infty$,
(ii) $A_i \; \rightsquigarrow \; \phi'_{\mathbf{G}}\mathbb{Y}(\cdot;\kappa)$ for $i \to \infty$,
(iii) For every $\varepsilon > 0: \; \lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\|A_{i,n} - \mathbb{V}_n\| > \varepsilon) = 0.$

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

$$\Big\| A_i - \phi'_{\mathbf{G}}\mathbb{Y}(\cdot;\kappa)\Big\| \le \|\phi'_{\mathbf{G}}\|_{op} \sup_{\kappa:|w(\kappa)|\in[0,1/i]} \sup_j \sup_{f\in\mathcal{F}_j} |\mathbb{Y}^{(j)}(f;\kappa)|,$$

[here, $\|\cdot\|_{op}$ denotes the operator norm] and the right-hand side converges to zero (as $i \to \infty$) in probability, this is a direct consequence of assumption (Ag3).

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

$$\begin{aligned}
\|A_{i,n} - \mathbb{V}_n\| \;\; &\le \;\; \alpha_n \sup_{\kappa:|w(\kappa)|\in[\beta_n n^{-1/2},1/i]} |w(\kappa)|\|\phi(\hat{\mathbf{G}}_n(\cdot;\kappa)) - \phi(\mathbf{G})\| \\
&\quad + \sup_{\kappa:|w(\kappa)|\in[0,\beta_n n^{-1/2}]} |w(\kappa)|\|\phi(\hat{\mathbf{G}}_n(\cdot;\kappa)) - \phi(\mathbf{G})\| \\
&\le \;\; c_\phi \sup_{\kappa:|w(\kappa)|\in[\beta_n n^{-1/2},1/i]} \sup_j \sup_{f\in\mathcal{F}_j} |\mathbb{Y}_n^{(j)}(f;\kappa)| \\
&\quad + \sup_{\kappa:|w(\kappa)|\in[0,\beta_n n^{-1/2}]} |w(\kappa)|\Big(\|\phi(\hat{\mathbf{G}}_n(\cdot;\kappa))\| + \|\phi(\mathbf{G})\|\Big) \\
&\quad + I\Big\{\sup_{\kappa:|w(\kappa)|\in[\beta_n n^{-1/2},1/i]} \alpha_n|w(\kappa)|\|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| > \varepsilon_\phi\Big\} \\
&\qquad \times \alpha_n \sup_{\kappa:|w(\kappa)|\in[\beta_n n^{-1/2},1/i]} \|\phi(\hat{\mathbf{G}}_n(\cdot;\kappa)) - \phi(\mathbf{G})\| \\
&=: \;\; R_{n,1} + R_{n,2} + R_{n,3},
\end{aligned}$$

where c_ϕ, ε_ϕ are from Lemma B.1. Here the second inequality follows by an application of Lemma B.1 on the set

$$\Big\{\sup_{\kappa:|w(\kappa)|\in[\beta_n n^{-1/2},1/i]} \alpha_n\|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| \le \varepsilon_\phi\Big\}$$

after observing that by definition

$$\alpha_n\|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| = \frac{1}{w(\kappa)} \sup_j \sup_{f\in\mathcal{F}_j} |\mathbb{Y}_n^{(j)}(f;\kappa)|.$$

Condition (Ag3) 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

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

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 $\mathbb{Y}_n^{(j)}$ we obtain

$$\limsup_{n\to\infty} P^*(\mathbb{Y}_n^{(j)} \in S_j(i,\varepsilon_\phi)) \le P(\mathbb{Y}^{(j)} \in S_j(i,\varepsilon_\phi))$$

for $j = 1,\dots,J$. By condition (Ag3), $\lim_{i\to\infty} P(\mathbb{Y}^{(j)} \in S_j(i,\varepsilon)) = 0$ 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 4.11. 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 4.5. In order to prove the second part, define

$$\begin{aligned} A_{i,n}^B &:= \alpha_n w(\kappa) I\{|w(\kappa)| \ge 1/i\}(\phi(\hat{\mathbf{G}}_n^B(\cdot;\kappa)) - \phi(\hat{\mathbf{G}}_n(\cdot;\kappa))) \\ A_i &:= \phi'_{\mathbf{G}}\big(\mathbb{Y}(\cdot;\kappa) I\{|w(\kappa)| \ge 1/i\}\big). \end{aligned}$$

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

$$\begin{aligned} &\text{(i)} && \text{For every } i \in \mathbb{N}: \ A_{i,n}^B \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} A_i \ \text{ for } n \to \infty, \\ &\text{(ii)} && A_i \rightsquigarrow \phi'_{\mathbf{G}}\mathbb{Y}(\cdot;\kappa) \ \text{ for } i \to \infty, \\ &\text{(iii)} && \text{For every } \varepsilon > 0: \ \lim_{i\to\infty} \limsup_{n\to\infty} \mathbb{P}^*(\|A_{i,n}^B - \mathbb{V}_n^B\| > \varepsilon) = 0. \end{aligned}$$

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 4.5. For a proof of the third assertion, note that $\mathbb{Y}_n^B \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{Y}$ 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 4.5. Hence the proof is complete. □

Proof of Proposition 4.7. First, observe that for $s \le n^{-1}, t \ge n^{-1}$ we have

$$\hat{\mathbf{G}}_n(\cdot;s,t) - \mathbf{G} = \hat{\mathbf{H}}_n(\cdot;t) - \mathbf{G}.$$

Moreover, for $t \ge s \ge n^{-1}$ we have

$$\begin{aligned} &\hat{\mathbf{G}}_n(\cdot;s,t) - \mathbf{G} \\ = \ &\frac{n}{1 + \lfloor nt \rfloor \vee 1 - \lfloor ns \rfloor \vee 1}\Big(\frac{\lfloor nt \rfloor \vee 1}{n}\ (\hat{\mathbf{H}}_n(\cdot;t) - \mathbf{G}) \\ &\qquad - \frac{-1 + \lfloor ns \rfloor \vee 1}{n}\ (\hat{\mathbf{H}}_n(\cdot;s - n^{-1}) - \mathbf{G})\Big) \end{aligned}$$

and thus setting '0/0 = 0' we obtain that for $t \ge s \ge n^{-1}$

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

Observe that for $t > 0$ we have $|\frac{\lfloor nt\rfloor\vee 1}{nt}-1| \le \frac{1}{\lfloor nt\rfloor \vee 1}$ and $|\frac{n(t-s)}{1+\lfloor nt\rfloor\vee 1-\lfloor ns\rfloor\vee 1}-1| \le \frac{3}{(\lfloor nt\rfloor - \lfloor ns\rfloor)\vee 1}$. Defining $\tilde{\mathbb{Y}}_n(\cdot;s,t) := \mathbb{G}_n(\cdot;t) - \tilde{\mathbb{G}}_n(\cdot;s)$, where $\tilde{\mathbb{G}}_n(\cdot;s) := \mathbb{G}_n(\cdot;s)I\{s\ge n^{-1}\}$ observe that $\tilde{\mathbb{Y}}_n \rightsquigarrow \mathbb{Y}$ by the continuous mapping theorem and some elementary calculations. Moreover, $\sup_t \|\frac{\lfloor nt\rfloor\vee 1}{nt}\,\mathbb{G}_n(\cdot;t) - \mathbb{G}_n(\cdot;s)\| = o_P^*(1)$ since for $t \ge n^{-1/4}$ the factor $\frac{\lfloor nt\rfloor \vee 1}{nt}$ tends to one uniformly and since $\sup_{t\le n^{-1/4}}\|\mathbb{G}_n(\cdot;t)\| = o_P^*(1)$ by arguments similar to those used to establish the negligibility of $R_{n,1}$ at the end of the proof of Theorem 4.5. Thus it remains to show that $(\frac{n(t-s)}{1+\lfloor nt\rfloor\vee 1-\lfloor ns\rfloor\vee 1}-1)\tilde{\mathbb{Y}}_n(\cdot;s,t)$ is uniformly small. This can be done by similar arguments (distinguish the cases $t-s\le n^{-1/4}$ and $t-s > n^{-1/4}$). This completes the proof. □

Proof of Theorem 4.10. 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 $\mathbb{Z}_n(t,f) := \mathbb{G}_n^{(1)}(f;t)$, $\mathcal{F} := \mathcal{F}_1$, $\mathcal{F}_\delta := \mathcal{F}_{1,\delta}$. For functions $\mathbf{G} \in \ell^\infty(\mathcal{F})$ and subsets $\mathcal{G}\subset\mathcal{F}$ introduce the notation $\|\mathbf{G}\|_{\mathcal{G}} := \sup_{f\in\mathcal{G}}|\mathbf{G}(f)|$. Start by noting that under the assumptions of the theorem together with (10) we have for some finite constant C_1

$$\sup_{k\in\mathbb{N}} \mathbb{E}^*\|\mathbb{Z}_k(1,\cdot)\|_{\mathcal{F}}^q \le C_1 < \infty. \tag{13}$$

To see this, fix $\delta > 0$ and cover the set $\mathcal{F}$ with countably many balls of radius δ and centres $\{f_j : j\in\mathbb{N}\}$. Then make use of the bound

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

and conditions (9), (10).

In order to establish asymptotic tightness of $\mathbb{Z}_n$, 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

$$\sup_{|s-t|+\rho(f,g)<\delta}|\mathbb{Z}_n(s,f)-\mathbb{Z}_n(t,g)| \le 2\sup_{0\le t\le 1}\|\mathbb{Z}_n(t,\cdot)\|_{\mathcal{F}_\delta} + \sup_{|s-t|<\delta}\|\mathbb{Z}_n(s,\cdot)-\mathbb{Z}_n(t,\cdot)\|_{\mathcal{F}}$$

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

$$\sup_{0\le t\le 1}\|\mathbb{Z}_n(t,\cdot)\|_{\mathcal{F}_\delta} \le \max_{1\le k\le n} 2\sqrt{\frac{k}{n}}\|\mathbb{Z}_k(1,\cdot)\|_{\mathcal{F}_\delta} = \frac{2}{\sqrt{n}}\max_{1\le k\le n}\|S_k\|_{\mathcal{F}_\delta}.$$

Fix $\epsilon \in (0, \{1-2^{-1/2+1/q}\}^{q/(q-1)}/2^{q/(2q-2)})$. Under (9), there exists a $\delta_0 > 0$ and $n_0 \in \mathbb{N}$, such that when $\delta \in (0, \delta_0)$ and $n \geq n_0(\epsilon)$, $\mathbb{E}^*\|\mathbb{Z}_n(1,\cdot)\|_{\mathcal{F}_\delta}^q \leq \epsilon^{2q}$. Moreover, under (13) 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 $q > 2$, $d = d_n = \lfloor \log n/(\log 2)\rfloor + 1$, we have

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

for $n \geq n_1(\epsilon) \vee n_0(\epsilon)$, where $n_1(\epsilon)$ is chosen such that the last inequality holds. Since ϵ was arbitrary, we have shown that $\limsup_{n\to\infty} P^*(\sup_{0\leq t\leq 1} \|\mathbb{Z}_n(t,\cdot)\|_{\mathcal{F}_\delta} > \epsilon) < \epsilon$ 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, we obtain by arguments similar to those given in the proof of Theorem 2.12.1 in Van der Vaart and Wellner (1996) that it is sufficient to consider the term

$$
\lceil\frac{1}{\delta}\rceil P^*\left(\sup_{0\leq s\leq\delta} \|\mathbb{Z}_n(s,\cdot)\|_{\mathcal{F}} > \epsilon\right) \quad\leq\quad \lceil\frac{1}{\delta}\rceil P^*\left(\max_{1\leq k\leq n\delta} 2\|S_k\|_{\mathcal{F}} > \sqrt{n}\epsilon\right). \tag{14}
$$

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

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

$$\leq \quad (\sqrt{n}\epsilon)^{-q} C_1^q 2^{d(\delta)q} \frac{1}{\{1-2^{-(1/2-1/q)}\}^q}$$
$$\leq \quad C_2 \epsilon^{-q} \delta^{q/2}$$

for n sufficiently large. Combined with (14), we get

$$\limsup_{n\to\infty} P^*(\max_{0\leq j\delta\leq 1} \sup_{j\delta\leq s\leq(j+1)\delta} \|\mathbb{Z}_n(s,\cdot)-\mathbb{Z}_n(j\delta,\cdot)\|_{\mathcal{F}} > \epsilon) < \epsilon$$

when $\delta < (\epsilon^{q+1}/C_2)^{1/(q/2-1)}$. The proof is thus complete. □

Proof of Theorem 5.2. Let $a_n = o(1)$ and $H^{(n)}$ denote a sequence in $\mathscr{D}_0$ with $H_n \to H \in \mathscr{U}$ such that $X + a_n H_n \in \mathscr{D}_\Phi$ $\forall n \in \mathbb{N}$. We need to show that, with respect to the metric $d_{\mathscr{R}}$,

$$a_n^{-1}\Big(\Phi(X+a_nH_n)-\Phi(X)\Big) \to \Phi'_X H.$$

Assume that this does not hold. Then by definition there exists a $b > 0$ and a sub-sequence n_k such that

$$d_{\mathscr{R}}\Big(a_{n_k}^{-1}\Big(\Phi(X+a_{n_k}H_{n_k})-\Phi(X)\Big), \Phi'_X H\Big) \geq 2b \quad \forall k \in \mathbb{N}$$

which by definition of $d_{\mathscr{R}}$ implies that there exists a sequence κ_{n_k} in K such that that

$$d_R\Big(a_{n_k}^{-1}(\phi(x+a_{n_k}H_{n_k}(\cdot;\kappa_{n_k}))-\phi(x)), (\phi'_x \cdot H(\cdot;\kappa_n))\Big) \geq b \qquad (15)$$

for all $k \in \mathbb{N}$. On the other hand, the sequence $H_{n_k}(\cdot;\kappa_{n_k})$ has a subsequence $H_{n_{k_j}}(\cdot;\kappa_{n_{k_j}})$ which converges to $H(\cdot;\kappa_\infty)$ for some $\kappa_\infty \in K$. To see that this is the case, start by noting that κ_{n_k} is a sequence in a compact metric space, and thus it has a convergent subsequence $\kappa_{n_{k_j}} \to \kappa_\infty$ with $\kappa_\infty \in K$. The definition of the set $\mathscr{U}$ then implies that $H(\cdot;\kappa_{n_k}) \to H(\cdot;\kappa_\infty)$. Together with the uniform convergence $\sup_\kappa d_D(H_n(\cdot;\kappa), H(\cdot;\kappa)) = o(1)$ this yields $H_{n_k}(\cdot;\kappa_{n_{k_j}}) \to H(\cdot;\kappa_\infty)$. Now quasi compact differentiability of ϕ tangentially to $C_0\langle D_0\rangle$ implies

$$a_n^{-1}\Big(\phi(x+a_nH_{n_{k_j}}(\cdot;\kappa_{n_{k_j}}))-\phi(x)\Big) \to \phi'_x H(\cdot;\kappa_\infty),$$

and together with continuity of ϕ'_x this contradicts (15). Thus the proof is complete. □

Appendix 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 $\varepsilon_\phi > 0, c_\phi < \infty$ such that*

$$\|\phi(x)-\phi(x+y)\|_R \leq c_\phi\|y\|_D \quad \forall y \in D_0 : \ \|y\|_D \leq \varepsilon_\phi, x+y \in D_\phi. \qquad (16)$$

Proof of Lemma B.1. Assume that (16) does not hold. Then for any pair $\varepsilon > 0, c < \infty$ there exists a $y_{c,\varepsilon} \in D_0$ such that $x + y_{c,\varepsilon} \in D_\phi$, $\|y_{c,\varepsilon}\|_D \leq \varepsilon$ and $\|\phi(x) - \phi(x + y_{c,\varepsilon})\|_R > c\|y_{c,\varepsilon}\|_D$. Consider the sequence $z_n := y_{n^2,n^{-2}}$ and define $\alpha_n := \|z_n\|_D \neq 0$. Then

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

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 ϕ 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. □

Lemma B.2. *Under assumptions (Wg) and (Ag3) there exists a sequence of real numbers* $\gamma_n = o(1)$ *such that* $\sup_{\kappa:|w(\kappa)|\geq\alpha_n^{-1}\gamma_n} \|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| = o_P^*(1)$ *[recall that the sequence* α_n *appears in assumption (Wg)].*

Proof of Lemma B.2. Define

$$B_n := \sup_{\kappa:|w(\kappa)|\leq\alpha_n^{-1/2}} \sup_{f_1,\dots,f_J} \|\mathbb{Y}_n(f_1,\dots,f_J;\kappa)\|.$$

By the Portmonteau Theorem, we have for any $\varepsilon > 0, \delta > 0$

$$\limsup_{n\to\infty} P^*(B_n \geq \varepsilon) \leq P\Big(\sup_{\kappa:|w(\kappa)|\leq\delta} \sup_{f_1,\dots,f_J} \|\mathbb{Y}(f_1,\dots,f_J;\kappa)\| \geq \varepsilon\Big).$$

By assumption (Ag3), the right-hand side of the above display tends to zero for $\delta \to 0$, and thus $B_n = o_P^*(1)$. 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

$$\begin{aligned}\sup_{\kappa:|w(\kappa)|\geq\alpha_n^{-1}\gamma_n} \|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| \quad &\leq \quad \sup_{\kappa:|w(\kappa)|\geq\alpha_n^{-1/2}} \|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| \\ &+ \sup_{\kappa:|w(\kappa)|\in[\gamma_n\alpha_n^{-1},\alpha_n^{-1/2}]} \|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\|.\end{aligned}$$

Now observe that

$$\sup_{\kappa:|w(\kappa)|\geq\alpha_n^{-1/2}} \|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| \leq 2\alpha_n^{-1/2} \sup_{\kappa:|w(\kappa)|\geq\alpha_n^{-1/2}} \|\mathbb{Y}_n(\cdot;\kappa)\| = o_P^*(1),$$

by arguments similar to those used to establish the negligibility of $R_{n,1}$ at the end of the proof of Theorem 4.5. Similarly

$$\sup_{\kappa:|w(\kappa)|\in[\gamma_n\alpha_n^{-1},\alpha_n^{-1/2}]} \|\hat{\mathbf{G}}_n(\cdot;\kappa) - \mathbf{G}\| \leq \gamma_n^{-1} \sup_{\kappa:|w(\kappa)|\leq\alpha_n^{-1/2}} \|\mathbb{Y}_n(\cdot;\kappa)\| = \gamma_n^{-1} B_n = o_P^*(1)$$

This completes the proof. □

Lemma B.3. *Given a triangular array random variables $M_{1,n},\dots,M_{n,n}$, and a sequence of random elements $\mathbb{V}_n^B(M_{1,n},\dots,M_{n,n})$ in a normed space $(D,\|\cdot\|_D)$, assume that the map $(M_{1,n},\dots,M_{n,n}) \mapsto \mathbb{V}_n^B(M_{1,n},\dots,M_{n,n})$ 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 $M_{1,n},\dots,M_{n,n}$]. Assume that for $i\in\mathbb{N}$ there exist approximations $A_{i,n}^B, A_i$ such that $(M_{1,n},\dots,M_{n,n}) \mapsto A_{i,n}^B(M_{1,n},\dots,M_{n,n})$ is measurable for every $i,n\in\mathbb{N}$ outer almost surely and that additionally*

(i) *For every $i\in\mathbb{N}$: $A_{i,n}^B \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} A_i$ 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 $\mathbb{V}_n^B \underset{M}{\overset{\mathbb{P}}{\rightsquigarrow}} \mathbb{V}$.

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

(a) $\sup_{f\in\mathrm{BL}_1}\left|\mathbb{E}_M f(\mathbb{V}_n^B) - \mathbb{E}f(\mathbb{V})\right| \to 0$ in outer probability,

(b) $\mathbb{E}_M f(\mathbb{V}_n^B)^* - \mathbb{E}_M f(\mathbb{V}_n^B)_* \xrightarrow{\mathbb{P}} 0$ for all $f\in\mathrm{BL}_1$.

Begin by observing that for every $i\in\mathbb{N}$, every ω, and every $f\in\mathrm{BL}_1$

$$\begin{aligned} &\left|\mathbb{E}_M f(\mathbb{V}_n^B) - \mathbb{E}f(\mathbb{V})\right| \\ \leq\ &\left|\mathbb{E}_M f(\mathbb{V}_n^B) - \mathbb{E}_M f(A_{i,n}^B)\right| + \left|\mathbb{E}_M f(A_{i,n}^B) - \mathbb{E}f(A_i)\right| + \left|\mathbb{E}f(A_i) - \mathbb{E}f(\mathbb{V})\right|. \end{aligned}$$

Moreover, for every ω

$$\begin{aligned} \sup_{f\in\mathrm{BL}_1}\left|\mathbb{E}_M f(\mathbb{V}_n^B) - \mathbb{E}_M f(A_{i,n}^B)\right| &\leq \sup_{f\in\mathrm{BL}_1}\mathbb{E}_M\left|f(\mathbb{V}_n^B) - f(A_{i,n}^B)\right| \\ &\leq \mathbb{E}_M\left[\|\mathbb{V}_n^B - A_{i,n}^B\|^* \wedge 2\right]. \end{aligned}$$

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

$$\begin{aligned} &\mathbb{E}^*\Big[\sup_{f\in\mathrm{BL}_1}\left|\mathbb{E}_M f(\mathbb{V}_n^B) - \mathbb{E}_M f(A_{i,n}^B)\right|\Big] \\ \leq\ &\mathbb{E}\left[\|\mathbb{V}_n^B - A_{i,n}^B\|^* \wedge 2\right] \leq 2\mathbb{P}(\|\mathbb{V}_n^B - A_{i,n}^B\|^* > \gamma) + \gamma. \end{aligned}$$

Thus (iii) yields

$$\lim_{i\to\infty}\limsup_{n\to\infty}\mathbb{E}^*\Big[\sup_{f\in \mathrm{BL}_1}\big|\mathbb{E}_M f(\mathbb{V}_n^B)-\mathbb{E}_M f(A_{i,n}^B)\big|\Big]=0.$$

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 η,ε 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 Lemma B.1 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. □

Appendix C: Details an a test for change points

As is clear from the recent review of Perron (2006), testing change points in a time series has many important applications in econometrics and statistics. 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^∞) 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 a SN-based tests in a general framework. More precisely, consider the following setup. Assume that we observe data $X_1, \ldots, X_n$ that stem from a not necessarily stationary time series. Let $\theta_k = \phi(\mathbf{F}^{(k)}) \in \mathbb{R}^q$ be the quantity of interest which depends on the distribution function of X_k denoted by $\mathbf{F}^{(k)}$. The goal is to test if there is a change point in $\{\theta_k\}_{k=1}^n$, i.e.

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

and a commonly considered alternative 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. Define

$$\begin{aligned} V_n(k) &= n^{-2}\Big\{ \sum_{j=1}^{k} j^2(\phi(\hat{\mathbf{F}}_{1,j}) - \phi(\hat{\mathbf{F}}_{1,k}))(\phi(\hat{\mathbf{F}}_{1,j}) - \phi(\hat{\mathbf{F}}_{1,k}))^T \\ &\quad + \sum_{j=k+1}^{n} (n-j+1)^2(\phi(\hat{\mathbf{F}}_{j,n}) - \phi(\hat{\mathbf{F}}_{k+1,n}))(\phi(\hat{\mathbf{F}}_{j,n}) - \phi(\hat{\mathbf{F}}_{k+1,n}))^T \Big\}. \end{aligned}$$

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

$$\phi(\hat{\mathbf{F}}_{k,j}) = \theta + (j-k+1)^{-1} \sum_{l=k}^{j} L(X_l) + R_n(k,j).$$

Again the functional central limit theorem is assumed for $\{L(X_k)\}$ (i.e., (3) holds) and the remainder terms are assumed to be asymptotically negligible. In particular, Shao and Zhang (2010) assume that

$$\sup_{k=1,\cdots,n} |kR_n(1,k)| = o_p(n^{1/2}), \qquad \sup_{k=1,\cdots,n} |kR_n(n-k+1,n)| = o_p(n^{1/2}). \quad (17)$$

The above condition (17) 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 $G_n = \sup_{r \in [0,1]} H_n(r) + o_P(1)$ where

$$H_n(r) := (\phi(\hat{\mathbf{F}}_{1,\lfloor nr \rfloor}) - \phi(\hat{\mathbf{F}}_{1,n}))^T \hat{W}_{r,n}^{-1} (\phi(\hat{\mathbf{F}}_{1,\lfloor nr \rfloor}) - \phi(\hat{\mathbf{F}}_{1,n}))$$

with

$$\hat{W}_{r,n} := \int_0^r (\phi(\hat{\mathbf{F}}_{1,\lfloor ns \rfloor}) - \frac{s}{r}\phi(\hat{\mathbf{F}}_{1,\lfloor nr \rfloor}))(\phi(\hat{\mathbf{F}}_{1,\lfloor ns \rfloor}) - \frac{s}{r}\phi(\hat{\mathbf{F}}_{1,\lfloor nr \rfloor}))^T ds$$
$$+ \int_r^1 (\phi(\hat{\mathbf{F}}_{\lfloor ns \rfloor,n}) - \frac{1-s}{1-r}\phi(\hat{\mathbf{F}}_{\lfloor nr \rfloor,n}))(\phi(\hat{\mathbf{F}}_{\lfloor ns \rfloor,n}) - \frac{1-s}{1-r}\phi(\hat{\mathbf{F}}_{\lfloor nr \rfloor,n}))^T ds.$$

Applying Theorem 2.5 in combination with the continuous mapping theorem yields weak convergence of G_n to

$$\sup_{r\in[0,1]} (\mathbb{W}(\cdot;0,r) - r\mathbb{W}(\cdot;0,1))^T W_r^{-1} (\mathbb{W}(\cdot;0,r) - r\mathbb{W}(\cdot;0,1))$$

where

$$W_r := \int_0^r \Big(\mathbb{W}(\cdot;0,s) - \frac{s}{r}\mathbb{W}(\cdot;0,r)\Big)\Big(\mathbb{W}(\cdot;0,s) - \frac{s}{r}\mathbb{W}(\cdot;0,r)\Big)^T ds$$
$$+ \int_r^1 \Big(\mathbb{W}(\cdot;s,1) - \frac{1-s}{1-r}\mathbb{W}(\cdot;r,1)\Big)\Big(\mathbb{W}(\cdot;s,1) - \frac{1-s}{1-r}\mathbb{W}(\cdot;r,1)\Big)^T ds.$$

References

Adams, T. and Nobel, A. (2010). Uniform convergence of vapnik–chervonenkis classes under ergodic sampling. *The Annals of Probability*, 38(4):1345–1367. MR2663629

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. MR2518505

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. MR2508575

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. MR2891441

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. MR2719874

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. MR2948902

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons Inc., New York. MR0233396

Bücher, A. (2013). A note on weak convergence of the sequential multivariate empirical process under strong mixing. *arXiv preprint arXiv:1304.5113*.

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. MR2893858

Bücher, A. and Kojadinovic, I. (2013). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *arXiv preprint arXiv:1306.3930*.

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. MR3049901

Bücher, A. and Volgushev, S. (2013). Empirical and sequential empirical copula processes under serial dependence. *Journal of Multivariate Analysis*.

Bühlmann, P. L. (1993). *The blockwise bootstrap in time series and empirical processes*. PhD thesis, Swiss Federal Institute of Technology Zürich.

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. MR2743035

Dehling, H., Mikosch, T., and Sørensen, M. (2002). *Empirical process techniques for dependent data*. Birkhäuser. MR1958776

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. MR1026312

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. MR1185204

Fermanian, J. D., Radulović, D., and Wegkamp, M. H. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860. MR2093613

Gao, F. and Zhao, X. (2011). Delta method in large deviations and moderate deviations for estimators. *The Annals of Statistics*, 39(2):1211–1240. MR2816352

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. MR1074422

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. MR1863569

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481. MR0093867

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21:1130–1164. MR2200988

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

Kojadinovic, I. and Rohmer, T. (2012). Asymptotics and multiplier bootstrap of the sequential empirical copula process with applications to change-point detection. *arXiv preprint arXiv:1206.2557*.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics, New York. MR2724368

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217–1241. MR1015147

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association*, 96:1066–1076. MR1947254

Patton, A. J. (2009). Copula–based models for financial time series. In *Handbook of financial time series*, pages 767–785. Springer.

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. MR1329181

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. New York, Springer. MR1707286

Radulović, D. (2009). Another look at the disjoint blocks bootstrap. *Test*, 18(1):195–212. MR2495971

Rüschendorf, L. (1974). On the empirical process of multivariate, dependent random variables. *Journal of Multivariate Analysis*, 4(4):469–478. MR0353404

Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Annals of Statistics*, 4:912–923. MR0420794

Salvadori, G. (2007). *Extremes in Nature: An Approach to Using Copulas*, volume 56. Springer.

Sen, P. K. (1974). Weak convergence of multidimensional empirical processes for stationary ϕ-mixing processes. *The Annals of Probability*, 2(1):147–154. MR0402845

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. MR2758116

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. MR2758242

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. MR2846559

Shao, X. (2012). Parametric inference in stationary time series models with dependent errors. *Scandinavian Journal of Statistics*, 39(4):772–783. MR3000848

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. MR3008276

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240. MR2752617

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231. MR0125600

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Verlag, New York. MR1385671

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. MR2353389

Wu, W. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436. MR2052582

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. MR0385962

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. MR3021390