A bootstrap functional central limit theorem for time-varying linear processes

Carina Beering\textsuperscript{a} and Anne Leuch\textsuperscript{b}

\textsuperscript{a}Research group of Mathematics and Statistics, Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg, Hamburg, Germany; \textsuperscript{b}Institute of Statistics, Otto-Friedrich-Universität Bamberg, Bamberg, Germany

ABSTRACT
We provide a functional central limit theorem for a broad class of smooth functions for possibly non-causal multivariate linear processes with time-varying coefficients. Since the limiting processes depend on unknown quantities, we propose a local block bootstrap procedure to circumvent this inconvenience in practical applications. In particular, we prove bootstrap validity for a very large class of processes. Our results are illustrated by some numerical examples.

1. Introduction

Unifying asymptotic theory is a powerful tool to develop statistical test procedures or to quantify the uncertainty of parameter estimators. Still, in many applications the limiting random objects of interest depend on unknown quantities that rely on the data generating process, e.g. its variance or the underlying dependence structure. Therefore, they cannot be used directly to construct (asymptotically) valid confidence sets or critical values of hypothesis tests. The bootstrap offers a convenient way to overcome these difficulties and is therefore the key to enable practical use of asymptotic results. From a statistical perspective, (functional) central limit theorems (CLT) and their bootstrap counterparts are particularly appealing as they can be applied to approximate confidence sets for parameters or critical values of $L_2$-test statistics. While there exists a large body of literature on bootstrap validity for empirical processes based stationary processes, see e.g. Künsch (1989), Naik-Nimbalkar and Rajarshi (1994), Bühlmann (1995), Doukhan et al. (2015) and Wieczorek (2016), there are no comparable results for locally stationary processes. The goal of the present paper is to fill this gap as the assumption of a gradually changing probabilistic structure over time is much more realistic than a stationary setting in many applications. Typical examples of use are medical or economical data, see Dette et al. (2011), Anderson and Sandsten (2019) and Jentsch et al. (2020), for instance. The idea of approximating non-stationary time
series on segments by stationary ones can be found in Priestley (1965), whereas the concept of local stationarity, on which this paper is based, goes back to Dahlhaus (1997). An overview of the state of the art is to be found in Dahlhaus (2012). More recently, Dahlhaus et al. (2019) and Phandoidaen and Richter (2022) developed a broad asymptotic theory including laws of large numbers and (functional) CLTs for non-linear, causal locally stationary processes. However, on the bootstrap side methods and validity checks for locally stationary data are mostly tailor-made for specific applications so far: Sergides and Paparoditis (2009) used a semiparametric bootstrap in a testing framework, while wild bootstrap methods have been considered for instance by Vogt (2015), Brunotte (2022) and Karmakar et al. (2022). Sergides and Paparoditis (2008) established a frequency domain bootstrap for statistics of the local periodogram, and Kreiss and Paparoditis (2015) developed a time-frequency domain bootstrap for a broad class of periodogram based statistics. In the present paper, we verify that a local version of the block bootstrap can be applied successfully to mimic the distributional behaviour of a large class of empirical processes indexed by smooth functions. This local block bootstrap was initially proposed by Paparoditis and Politis (2002) for the mean and by Dowla et al. (2013) for heteroscedastic time series with trend. Moreover, it was used by Franzier and Koo (2021) for indirect estimation in locally stationary structured models. In our case, we investigate possibly two-sided time-varying linear processes. The reasons are two-fold: On the one hand, to the best of our knowledge, there is no functional central limit theorem (FCLT) for non-causal locally stationary linear processes. However, non-causal models play an important role in the context of economic data. Examples for the application of non-causal AR models to stock volume data can be found in Breidt et al. (2001), Andrews et al. (2009) and in Lanne and Saikkonen (2014) for non-causal VAR models. On the other hand, a generalisation to non-linear processes would rely on high-level assumptions. Since the technical notation in the proofs would blow up, we restrict ourselves to linear processes here and leave an adaptation of the proof to causal non-linear processes for future research. It is important to note that the method of proof used by Phandoidaen and Richter (2022) to verify a FCLT for locally stationary Bernoulli shifts cannot be adapted in a straightforward manner to prove validity of block bootstrap methods as it relies on martingale difference approaches for Bernoulli shifts. However, the bootstrap process does not inherit this structure.

The rest of the paper is organised as follows: In Section 2, we describe the setting under consideration and provide a FCLT for smooth functions of locally stationary linear processes. Following this, Section 3 is devoted to a local block bootstrap procedure for empirical processes. Afterwards, we illustrate the finite-sample performance of our method in Section 4. The proof of the main asymptotic result is carried out in Section 5, while all remaining proofs as well as several auxiliary results including their proofs are deferred to an Appendix in an online supplement to this paper.

Notation: Let $|\cdot|_1$ denote the max column sum matrix norm, i.e. $|M|_1 = \max_{1 \leq j \leq r} \sum_{i=1}^d |m^{(ij)}|$ for some $(d \times r)$ matrix $M = (m^{(ij)})$. Note that this matrix norm is submultiplicative. For a $d$-dimensional vector $v = (v_1, \ldots, v_d)'$, its $\ell_p$-norm is denoted by $|v|_p$, $p \in [1, \infty]$, whereas the $\mathcal{L}^p$-norm with $p \in [1, \infty)$ for $d$-dimensional random vectors $X$ is signified by $\|X\|_p := (E|X|^p)^{1/p}$. Additionally, said notation is transferred to the bootstrap world by defining $\|X^*\|_{p,*} := (E^*|X^*|^p)^{1/p}$ as the (conditional) bootstrap $\mathcal{L}^p$-norm of $X^*$. 

Besides, the Lipschitz seminorm \(| \cdot |_{\text{Lip}}\) of a function \(f\) as above signifies

\[
|f|_{\text{Lip}} := \sup_{x,y \in T, x \neq y} |f(x) - f(y)|/|x - y|.
\]

2. A functional CLT for locally stationary linear processes

Let \((\varepsilon_t)_{t \in \mathbb{Z}}\) be a sequence of i.i.d. centred \(\mathbb{R}^d\)-valued random vectors and \((X_{t,T})_{t=1}^T\) a \(d\)-variate (possibly) two-sided linear process

\[
X_{t,T} = \mu\left(\frac{t}{T}\right) + \sum_{j \in \mathbb{Z}} A_{t,T}(j)\varepsilon_{t-j},
\]

where \(\mu = (\mu_1, \ldots, \mu_d)'\) is a \(d\)-variate time-varying mean function and \((A_{t,T}(j))_{j \in \mathbb{Z}}\) are coefficient matrices of dimension \((d \times d)\). Note that the sequence \((\varepsilon_t)_{t \in \mathbb{Z}}\) is not necessarily a white noise because we do not assume \(E\varepsilon_t^2 < \infty\). Hence, \((X_{t,T})_{t=1}^T\) is only having a MA\((\infty)\)-representation but is not automatically a MA\((\infty)\)-process. To ensure existence of the afore-defined process and to exclude rapid changes in the coefficients over time leading to a meaningful statistical methodology, we impose the following conditions concerning \((X_{t,T})_{t=1}^T\):

**Assumption 2.1 (Locally stationary linear processes):** The process \((X_{t,T})_{t=1}^T\) is of form (1) with the following specifications:

(i) The innovations \((\varepsilon_t)_{t \in \mathbb{Z}}\) are i.i.d., centred and \(E|\varepsilon_1|_1 < \infty\).

(ii) For some \(\vartheta \in (0, 1)\) and a constant \(B < \infty\)

\[
\sup_{t,T} \left| A_{t,T}(j) \right|_1 \leq B \vartheta^j.
\]

Further, there exists an entry-wise continuously differentiable function \(A(\cdot,j) : [0, 1] \rightarrow \mathbb{R}^{d \times d}\) with \(A(u, j) = (a^{(p,q)}(u,j))_{p,q=1,\ldots,d}\) such that for all \(p, q = 1, \ldots, d\) and \(s \leq k \in \{0, 1\}\), it holds

\[
\sup_u \left| \frac{\partial^s a^{(p,q)}(u,j)}{\partial u^s} \right| \leq B \vartheta^j \quad \text{and} \quad \sup_{t,T} \left| A_{t,T}(j) - A\left(\frac{t}{T}, j\right) \right|_1 \leq B \vartheta^j, \quad j \in \mathbb{Z}.
\]

(iii) Each component of the mean function \(\mu\) is continuously differentiable.

**Remark 2.1:** (i) This kind of assumptions represents a classical framework for statistical inference based on locally stationary processes, see Dahlhaus (2012) and Jentsch et al. (2020). Still, many papers, e.g. Dahlhaus and Subba Rao (2006) and Cardinali and Nason (2010), only require a polynomial decay instead of a geometric one as in (2) and (3). In fact, a polynomial decay is sufficient in the present context as well. However, the degree of decay depends on the presumed order of absolute moments of the function \(f\) introduced later in a complicated manner, see Beering (2021) for details.
in a comparable context. For sake of notational simplicity, we stick to the exponential decay here.

(ii) Other and more general definitions of local stationarity are invoked for example in Vogt (2012) and Dahlhaus et al. (2019). They do not require a linear representation of the process \((X_{t,T})_{t=1}^T\) to the price of cited causality.

Having introduced nonstationarity as in Assumption 2.1, the process \((X_{t,T})_{t=1}^T\) can be approximated locally by a (strictly) stationary linear process, its so-called companion process \((\tilde{X}_t(u))_{t \in \mathbb{Z}}\)

\[
\tilde{X}_t(u) = \mu(u) + \sum_{j \in \mathbb{Z}} A(u,j) \varepsilon_{t-j},
\]

as long as \(u\) is close to the rescaled time \(t/T\). Hence, compared to the original process the mean function \(\mu\) stays the same for \(u = t/T\), whereas the function \(A_{t,T}(j)\) is replaced by \(A(u,j)\). From Assumption 2.1(ii), we can conclude

\[
\sup_{u \in [0,1]} |A(u,j)|_1 \leq \tilde{B} \vartheta^j
\]

for some finite constant \(\tilde{B}\). This inequality connotes that (4) possesses a strictly stationary solution for each fixed \(u\) while Assumption 2.1 is satisfied.

**Remark 2.2:**

(i) Closeness of the locally stationary process and its companion process as well as closeness of companion processes for nearby rescaled time points can be specified. We obtain

\[
\sup_{1 \leq t \leq T} \left\| X_{t,T} - \tilde{X}_t \left( \frac{t}{T} \right) \right\|_1 = O(T^{-1})
\]

if Assumption 2.1 holds true for \(k = 0\) and \(\|\tilde{X}_0(u_1) - \tilde{X}_0(u_2)\|_1 \leq C|u_1 - u_2|_1 \forall u_1, u_2 \in [0,1] \) for some \(C < \infty\) if Assumption 2.1 holds true for \(k = 1\); see Jentsch et al. (2020, Lemma 2.1).

(ii) Although the construction with \(A_{t,T}(j)\) and \(A(t/T,j)\) appears to be unnecessarily complicated, it is required to include time-varying ARMA-processes, see Dahlhaus (2012) for details.

Statistical methods for locally stationary processes can either rely on local or global characteristics of the process. Local quantities of interest are, for instance, the local variance \(\text{Var}(\tilde{X}_0(u))\) for any fixed \(u \in (0,1)\) or the local characteristic function, introduced in Jentsch et al. (2020) as \(\varphi(u,\varsigma) := E(e^{i \langle \varsigma, \tilde{X}_0(u) \rangle})\), \(\varsigma \in \mathbb{R}^d\). These quantities can be estimated using kernel estimates based on the observations \(X_{1,T}, \ldots, X_{T,T}\); e.g. \(\varphi(u,\varsigma)\) can be estimated by the local empirical characteristic function (ECF)

\[
\hat{\varphi}(u,\varsigma) = \frac{1}{b_T} \sum_{t=1}^T K \left( \frac{t/T - u}{b_T} \right) e^{i \langle \varsigma, \tilde{X}_t \rangle}, \quad \varsigma \in \mathbb{R}^d,
\]

using a suitable kernel function \(K\) and an appropriate bandwidth \(b_T\). A simple and prominent example of a global quantity of interest is the integrated volatility in high-frequency
finance, see Feng (2015, sec. 2) and references therein. Consider a discrete-time model for the intraday log returns \( Y_{t,T} = \sqrt{T}^{-1} \sigma(t/T) \varepsilon_t \) for a smooth deterministic spot volatility function \( \sigma \) and a centred stationary process \((\varepsilon_t)_t\). Then, a natural estimator for the integrated volatility is given by the realised volatility

\[
RV = \sum_{t=1}^{T} Y_{t,T}^2.
\] (7)

Our goal is to derive a FCLT that is flexible enough to deduce the asymptotic distribution of both quantities (6) and (7). Therefore, we consider function classes changing with \( T \)

\[
\mathcal{F}_T := \left\{ \sum_{t=1}^{T} w_tT f(\xi_tX_{t,T}) : s \in \mathcal{S} \subseteq \mathbb{R}^d \right\}
\]

where the underlying Assumptions 2.2 and 2.4 are specified below. Obviously, the local ECF is included considering its real and imaginary part separately by setting \( w_{t,T} := (b_T T)^{-1/2} K(\frac{T-t}{b_T}) \) and \( f(\xi_tX_{t,T}) = \cos(\xi_tX_{t,T}) \) or \( f(\xi_tX_{t,T}) = \sin(\xi_tX_{t,T}) \), respectively. Secondly, if we choose \( w_{t,T} := T^{-1/2} \), \( X_{t,T} = \sigma(t/T)\varepsilon_t \) and \( f(\xi_tX_{t,T}) = X_{t,T}^2 \), we end up with \( RV \). The first example perfectly illustrates that several weights may be zero. We control for the number of zero weights and the magnitude of the non-zero weights as follows:

**Assumption 2.2 (Weights):** The sequence of non-negative weights \((w_{t,T})_{t=1}^{T}\) fulfills

\[
\sup_{1 \leq t \leq T} w_{t,T} \leq C_w d_T^{-1/2}
\]

for some finite constant \( C_w \), where \( d_T \xrightarrow{t \to \infty} \infty \) denotes the number of non-zero weights in \((w_{t,T})_{t=1}^{T}\).

Before we specify the class of functions for our Donsker-type result, we need to introduce some auxiliary quantities to properly handle the dependence structure of the observed and the companion processes within the proofs relying on truncation arguments. Considering a truncation parameter \( M \in \mathbb{N} \), we set

\[
X_{t,T}^{(M)} := \mu \left( \frac{t}{T} \right) + \sum_{|j| < M} A_{t,T}(j) \varepsilon_{t-j} \quad \text{and} \quad \tilde{X}_{t}^{(M)}(u) := \mu(u) + \sum_{|j| < M} A(u,j) \varepsilon_{t-j}.
\] (8)

**Assumption 2.3 (Function I):** Let \((\mathcal{S}, \rho) \) be a compact semimetric space with \( \mathcal{S} \subseteq \mathbb{R}^d \).

The function \( f : \mathcal{S} \times \mathbb{R}^d \to \mathbb{R} \) satisfies

\[
\sup_{x \in \mathcal{S}} |f(\xi_x \tilde{x}) - f(\xi_x \bar{x})| \leq C_{\text{lip}} |\tilde{x} - \bar{x}|_1, \quad \tilde{x}, \bar{x} \in \mathbb{R}^d,
\] (9)

for some \( C_{\text{lip}} < \infty \). Additionally, for some \( \delta \in (0, 1/2) \) it holds

\[
\sup_{t \leq T, x \in \mathcal{S}, M \in \mathbb{N}} E[|f(\xi_x X_{t,T})|^{2+\delta} + |f(\xi_x \tilde{X}_{t}^{(M)})|^{2+\delta}] < \infty \quad \text{and}
\]
\[ \sup_{u \in [0,1], s \in S, M \in \mathbb{N}} E[|f(s, \tilde{X}_0(u))|^2 + |f(s, \tilde{X}_0^{(M)}(u))|^2] < \infty. \]

**Remark 2.3:**

(i) If \( \|\varepsilon_0\|_{2+\delta} < \infty \), the moment conditions in Assumption 2.3 follow from (9). However, only \( \|\varepsilon_0\|_1 < \infty \) is presumed in Assumption 2.1. In doing so, we can allow for data generating processes with infinite variance as long as the function applied to it is regular enough, see Section 4.2 for an example.

(ii) Obviously, assuming validity of Assumption 2.3 additionally to Assumption 2.1 allows for a generalization of the results stated in Remark 2.2(i) towards

\[ \sup_{1 \leq t \leq T} \left\| f(s, \tilde{X}_{t,T}) - f(s, \tilde{X}_t \left( \frac{t}{T} \right)) \right\|_1 = O(T^{-1}). \]

(iii) Our assumptions are slightly different compared to those in Phandoidaen and Richter (2022), where they allow for H"older continuity of \( f \) with respect to \( (w.r.t.) x \). We expect that it is possible to relax our assumption in a similar way, which, however, would have an effect on the choice of tuning parameters of the bootstrap procedure in Section 3. For sake of notational simplicity, we stick to Lipschitz continuity here. Moreover, note that we work under weaker moment constraints regarding the data generating process in their case \( s = 1 \).

We abbreviate the centred version of \( f \) by

\[ \tilde{f}(s, \cdot) := f(s, \cdot) - Ef(s, \cdot) \] (10)

and state a CLT for the finite-dimensional distributions first. Note that this result is sufficient to deduce asymptotic normality of the realised volatility defined in (7).

**Theorem 2.4 (Central limit theorem):** Suppose Assumption 2.1 holds true for \( k = 1 \) and Assumptions 2.2 as well as 2.3 are valid. Then, for any \( J \in \mathbb{N} \) and \( s_j \in \mathcal{S}, j = 1, \ldots, J \), we have

\[ \left( \sum_{t=1}^{T} w_{t,T} \tilde{f}(s_j, \tilde{X}_{t,T}) , j = 1, \ldots, J \right) \xrightarrow{d} \mathcal{N}(0, V) \]

as \( T \to \infty \), where \( V := (V(s_{j_1}, s_{j_2}))_{j_1,j_2=1,\ldots,J} \) is a \((J \times J)\) covariance matrix with

\[ V(s_{j_1}, s_{j_2}) = \lim_{h \to \infty} \sum_{t=1}^{T} w_{t,T} w_{t+h,T} \text{Cov} \left( f(s_{j_1}, \tilde{X}_0 \left( \frac{t}{T} \right)) , f(s_{j_2}, \tilde{X}_h \left( \frac{t}{T} \right)) \right). \]

We aim at deriving a FCLT. For that purpose, some additional conditions on \( f \) are imposed to assure tightness. To this end, let \( D(u, \mathcal{S}, \rho) = \max(\# \mathcal{S}_0 | \mathcal{S}_0 \subseteq \mathcal{S}, \rho(s_1, s_2) > u \forall s_1 \neq s_2 \in \mathcal{S}_0) \) denote the usual packing number defined e.g. in van der Vaart and Wellner (2000, Def. 2.2.3).

**Assumption 2.4 (Function II):**

(i) Additionally to Assumption 2.3, it holds \( \|f(\cdot, \cdot)\|_\infty < \infty \) for any \( x \in \mathbb{R}^d \).
Algorithm 3.1 (Bootstrap algorithm): (a) Consider a blocklength $L_T$ depending on $T$.
(b) Select a window parameter $D_T \in (0, 1)$ such that $TD_T \in \mathbb{N}$.
(c) Generate i.i.d. integers $k_0, \ldots, k_{\lfloor T/L_T \rfloor}$−1 using a discrete uniform distribution on $[-TD_T, TD_T]$.

Theorem 2.5 (Functional central limit theorem): Suppose Assumption 2.1 holds true for $k = 1$ and Assumption 2.2 as well as 2.4 are valid. Then,

$$\left( \sum_{i=1}^{T} w_{i,T} \tilde{f}(\tilde{s}, X_{i,T}) \right)_{\tilde{s}\in\mathcal{S}} \overset{d}{\rightarrow} \left( G(\tilde{s}) \right)_{\tilde{s}\in\mathcal{S}}$$

as $T$ tends to $\infty$, where $(G(\tilde{s}))_{\tilde{s}\in\mathcal{S}}$ is a centred Gaussian process with continuous sample paths w.r.t. $\rho$ and covariance function $V(\tilde{s}, \tilde{s}')$ originating from Theorem 2.4.

3. Locally blockwise bootstrapped empirical processes

As already pointed out in Paparoditis and Politis (2002), the classical block bootstrap algorithm for stationary time series has to be modified in the case of locally stationary time series to capture not only the dependence structure but also the time-changing characteristics of the process. More precisely, given $X_{1,T}, \ldots, X_{T,T}$, a block of a bootstrap analogue starting at time point $t$ should only consist of a stretch of the original time series with time index close to $t$. This is achieved by the introduction of an additional tuning parameter, the so-called window parameter, that controls for the range of observations a certain bootstrap block can be drawn from. An adaption of the local block bootstrap (LBB) proposed by Dowla et al. (2013) to the present setting reads as follows:

(ii) It holds $|f(\tilde{s}, \tilde{x}) - f(\tilde{s}', \tilde{x})| \leq g(\tilde{x})\rho(\tilde{s}, \tilde{s}')$ with some function $g: \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ satisfying one of the following conditions:
(a) $E[g^{2+\delta}(Y)] < K$ for some $K < \infty$ with $Y \in \{\tilde{X}_t(u), \tilde{X}_t^{(M)}(u)\}_{t \in \mathbb{Z}, M \in \mathbb{N}}$,
(b) $\|f\|_{\infty} < \infty$ and $E[g^{1+\delta/2}(Y)] < K$ for some $K < \infty$ with $Y \in \{\tilde{X}_t(u), \tilde{X}_t^{(M)}(u)\}_{t \in \mathbb{Z}, M \in \mathbb{N}}$.
(iii) For any $u > 0$, let $D(u, \mathcal{S}, \rho) \leq C_D(1 + u^{-1})^d$.

Note that in Assumption 2.4(ii), we distinguish between bounded and unbounded functions $f$. At first glance, boundedness seems much more restrictive as unboundedness, but the different assumptions concerning the moments of the function $g$ open up the field of applications. As an example, consider the ECF case in Jentsch et al. (2020), where the function $g$ is equal to the $|\cdot|_1$-norm. Combined with an $\alpha$-stable distribution with $\alpha = 1.5$, which we will use in our simulation study later on, we are not able to fulfil case (a) of the second part of Assumption 2.4 due to the lack of second absolute moments. Especially in finance, the absence of those moments is quite common. Thus, instead of being mostly excluding, the separate handling of bounded functions broadens the scope.

Assumption 2.4(iii) holds, for instance, for $(\mathcal{S}, \rho) = ([-S, S]^d, \rho(\tilde{s}_1, \tilde{s}_2) = |\tilde{s}_1 - \tilde{s}_2|_1)$ for any $S \in (0, \infty)$ as well as for $(\mathcal{S}, \rho) = (\mathbb{R}^d, \rho(\tilde{s}_1, \tilde{s}_2) = \sum_{i=1}^{d} |\arctan(\tilde{s}_1,i) - \arctan(\tilde{s}_2,i)|_1)$.
(d) For \( i = 0, \ldots, \lfloor T/L_T \rfloor - 1 \), define \( X_{i,T}^* \), \( X_{i+1,T}^* \), \( X_{i+2,T}^* \), by \( X_{j+iL_T,T}^* := X_{j+iL_T+k_i,T} \) for \( j = 1, \ldots, L_T \), if the resulting set of indices is in \([1, T]\) and use \(-k_i\) instead of \( k_i\) otherwise.

(e) Construct the bootstrap estimator by replacing \( X_{t,T} \) with \( X_{t,T}^* \), that is
\[
\sum_{t=1}^{T} w_t f \left( s, X_{t,T}^* \right).
\]

Remark 3.1: (i) The distribution used to generate \( k_0, \ldots, k_{\lfloor T/L_T \rfloor - 1} \) does not need to assign uniform weights to every choice of \( k_i\), see Paparoditis and Politis (2002). Here, we choose the discrete uniform distribution as it is easy to handle, analogously to Dowla et al. (2013). Besides, it matches the choice made for the ordinary moving block bootstrap algorithm designed for stationary processes.

(ii) The case differentiation in part (d) of Algorithm 3.1 ensures that if a block is in danger of going over the edge, there is a sound way out. By adjusting the sign of \( k_i\) for the whole block, the interrelated dependence structure is preserved, and moreover, no observation is used twice in the same block.

To establish asymptotic validity, we have to modify our assumptions towards more restrictive moment conditions in the case of unbounded \( f\). Especially when it comes to covariance results, finite \((2 + \delta)\)th absolute moments of \( f\) are not always sufficient but we require the following:

Assumption 3.1 (Function III): Assumption 2.1 for \( k = 1 \) plus Assumptions 2.2 and 2.3 are satisfied and
\[
\sup_{u \in [0,1], s \in S, M \in \mathbb{N}} E|f(s, \tilde{X}_0(u))|^{4+\delta} + |f(s, \tilde{X}_0(M)(u))|^{4+\delta} < \infty.
\]

With the blocklength \( L_T \) and the window parameter \( D_T \), two new parameters are involved, which need to behave good-naturedly in combination with the number of non-zero weights \( d_T \):

Assumption 3.2 (Bootstrap rates): For the blocklength \( L_T \in \mathbb{N} \) and the window parameter \( D_T \) with \( TD_T \in \mathbb{N} \), it holds
\[
L_T \to \infty, \quad L_T = o \left( d_T^{\frac{\delta}{2(2+\delta)}} \right), \quad d_T^{\frac{25}{2+\delta}} = O(TD_T) \quad \text{and} \quad TD_T = O \left( d_T^{\frac{1}{2+\delta}} \right).
\]

Remark 3.2: (i) In particular, these assumptions imply \( L_T = o(TD_T) \).

(ii) To the best of our knowledge, optimal choices of the tuning parameters \( L_T \) and \( D_T \) of the local block bootstrap have not been derived in the literature so far. As already mentioned in Dowla et al. (2013), an adaptation of the subsampling cross-validation approach proposed in Hall et al. (1995) might be a feasible option. A deep consideration of this issue would go far beyond the scope of the paper and is left for future research.

In the style of (10), we define
\[
\tilde{f}^\ast (s, X_{t,T}^*) := f(s, X_{t,T}^*) - E^\ast f(s, X_{t,T}^*)
\]
for \( t = 1, \ldots, T \) and \( s \in \mathbb{R}^d \) as the bootstrap version of the centred function \( \tilde{f} \), which enables us to state the bootstrap counterpart to Theorem 2.4 in a comprehensive way:
Theorem 3.3 (Bootstrap central limit theorem): Under Assumptions 3.1 and 3.2, it holds for any \( J \in \mathbb{N} \) and \( j \in \mathcal{J}, j = 1, \ldots, J, \)

\[
\left( \sum_{t=1}^{T} w_t T f^* \left( \tilde{s}_j, X^*_t, T \right), j = 1, \ldots, J, \right) \overset{d}{\longrightarrow} Z \sim \mathcal{N}(0, V)
\]

in probability, where \( V \) is defined in Theorem 2.4. If additionally \( V(s_1, s_1) > 0 \), then

\[
\sup_{v \in \mathbb{R}} \left| \mathbb{P} \left( \sum_{t=1}^{T} w_t T f^* \left( s_1, X^*_t, T \right) \leq v \right) - \Phi \left( \frac{v}{V(s_1, s_1)} \right) \right| \overset{P}{\longrightarrow} 0.
\]

Note that this result is in line with Paparoditis and Politis (2002, Th. 3.1), where the mean of locally stationary time series \( X_t, T = \mu + V(t/T) \varepsilon_t \) with \( \mu \in \mathbb{R} \), a smooth function \( V \), and \( \alpha \)-mixing stationary innovations \( (\varepsilon_t)_t \) satisfying the stronger moment assumption \( E|\varepsilon_t|^6 < \infty \) were considered.

For the more general case of a functional CLT, stronger assumptions concerning the function \( g \) introduced in Assumption 2.4 are required, too:

Assumption 3.3 (Function IV): The function \( g \) originating from Assumption 2.4 satisfies one of the following conditions:

(a) \( E g^{4+\delta}(Y) < K \) for some \( K < \infty \) with \( Y \in \{ \tilde{X}_t(u), \tilde{X}^{(M)}_t(u) \}_{t \in \mathbb{Z}, M \in \mathbb{N}} \) and for some constant \( C_{g,a} < \infty \)

\[
\sup_{t \in \{1, \ldots, T\}} \left\| g \left( X_t, T \right) - g \left( \tilde{X}_t \left( \frac{t}{T} \right) \right) \right\|_{4+\delta} \leq \frac{C_{g,a}}{T},
\]

(b) \( \|f\|_\infty < \infty, E g^{1+\delta/2}(Y) < K \) for some \( K < \infty \) with \( Y \in \{ \tilde{X}_t(u), \tilde{X}^{(M)}_t(u) \}_{t \in \mathbb{Z}, M \in \mathbb{N}} \) and for some constant \( C_{g,b} < \infty \)

\[
\sup_{t \in \{1, \ldots, T\}} \left\| g \left( X_t, T \right) - g \left( \tilde{X}_t \left( \frac{t}{T} \right) \right) \right\|_{2+\frac{\delta}{2}} \leq \frac{C_{g,b}}{T}.
\]

Finally, we use the previously established bootstrap CLT in combination with a tightness result to prove the desired bootstrap FCLT:

Theorem 3.4 (Bootstrap functional central limit theorem): Let Assumptions 2.4 to 3.3 be true. Then, for \( \tilde{f}^* \) as in (11) it holds

\[
\left( \sum_{t=1}^{T} w_t T \tilde{f}^* \left( s, X^*_t, T \right) \right) \overset{d}{\longrightarrow} \left( G(s) \right) s \in \mathcal{S}
\]

in \( P \)-probability as \( T \) tends to \( \infty \), where \( (G(s)) s \in \mathcal{S} \) is defined in Theorem 2.5.

We expect similar results to hold for a suitable version of the dependent wild bootstrap. Results on its validity in related settings can be found in Brunotte (2022). For our example in Subsection 4.1, the local block bootstrap outperformed the dependent wild bootstrap in a simulation study in Feng (2015). A detailed comparison of both methods is beyond the scope of the paper.
4. Numerical results

We illustrate the finite sample performance of the proposed local bootstrap procedure by two small numerical examples. In both cases, we replicate the simulations \(N = 500\) times each with \(B = 500\) bootstrap resamplings. The implementations are carried out in R; see R Core Team (2022).

4.1. An application to realised volatility

We investigate the coverage of bootstrap-based confidence sets for the realised volatility introduced in (7) as a global characteristic of the data generating process. As a straightforward consequence of Theorems 2.4 and 3.3, we obtain the following result:

Corollary 4.1: Assume that \((\varepsilon_t)_t\) is a linear process with exponentially decaying coefficients and innovations that have finite absolute \(8 + \delta\) moments. Further, suppose that the long-run variance of \((\varepsilon_t)_t\) is positive and that \(\sigma\) is continuously differentiable on \([0,1]\). Then, under Assumption 3.2,

\[
\sup_{v \in \mathbb{R}} \left| P^* \left( \sqrt{T} \sum_{t=1}^{T} (X_{t,T}^* - E^* X_{t,T}^*) \leq v \right) - P \left( \sqrt{T} \left( RV - \sigma^2 \int_0^1 \sigma^2(u) du \right) \leq v \right) \right| \to 0.
\]

In order to illustrate the finite-sample performance of the local bootstrap for RV, we revisit a scenario similar to Feng (2015). We consider the coverage of symmetric 90% confidence intervals derived by local bootstrap for samples of size \(T = 1000\) and \(T = 2000\) with \(\sigma(u) = 0.32(u - 0.5)^2 + 0.04, u \in [0, 1]\), reflecting a volatility smile, and \(\varepsilon_t = 0.5 \eta_{t-1} + \eta_t\). Here, \((\eta_t)_t\) is a sequence of i.i.d. innovations satisfying \(\eta_t \sim \mathcal{N}(0, 0.8)\). As it can be seen from Figure 1, the coverage of the confidence intervals for RV obtained by the local bootstrap is close to the desired level. In particular, the results are robust w.r.t. appropriate choices of blocklength \(L_T\) and half window size \(DT = D_T T\). Note that the optimal block-lengths of the ordinary moving block bootstrap in the sense of Lahiri (2003, Cor. 7.1, \(k = 2\)) are \(L_{1000} = 3\) and \(L_{2000} = 4\) in this example.

![Figure 1](image-url)
4.2. An application to local empirical characteristic functions

We consider local ECFs as introduced in (6). Therefore, we need to specify our requirements for both kernel and bandwidth.

Assumption 4.1 (Kernel and bandwidth):

(i) The function \( K : \mathbb{R} \rightarrow [0, \infty) \) is non-negative, symmetric and Lipschitz continuous. Additionally, it integrates up to 1 and has compact support \([-1, 1]\).

(ii) The sequence of bandwidths \((b_T)_{T \in \mathbb{N}}\) is non-negative and fulfills \( b_T \to 0 \), \( b_T^2 T \to \infty \) and \( b_T^3 T = o(1) \) as \( T \to \infty \).

Now we are able to state the following corollary:

Corollary 4.2: Suppose Assumptions 2.1 and 2.4 are satisfied for \( k = 1 \) plus Assumptions 3.2 and 4.1. Additionally, let

\[
\sum_{h_1, h_2 \in \mathbb{Z}} \left( \text{Cov} \left( \cos \left( \langle s, \tilde{X}_0(u) \rangle \right), \cos \left( \langle s, \tilde{X}_{h_2}(u) \rangle \right) \right) \text{Cov} \left( \sin \left( \langle s, \tilde{X}_0(u) \rangle \right), \sin \left( \langle s, \tilde{X}_{h_1}(u) \rangle \right) \right) \\
- \left( \text{Cov} \left( \cos \left( \langle s, \tilde{X}_0(u) \rangle \right), \sin \left( \langle s, \tilde{X}_{h_1}(u) \rangle \right) \right) \text{Cov} \left( \cos \left( \langle s, \tilde{X}_0(u) \rangle \right), \sin \left( \langle s, \tilde{X}_{h_2}(u) \rangle \right) \right) \right) > 0
\]

be fulfilled. Then, it holds for every \( s \in \mathbb{R}^d \)

\[
\sup_{v \in \mathbb{R}} \left| \mathbb{P}^* \left( \left| (b_T T)^{1/2} \left( \tilde{\varphi}^*(u; s) - E^* \tilde{\varphi}^*(u; s) \right) \right| \leq v \right) \\
- \mathbb{P} \left( \left| (b_T T)^{1/2} \left( \tilde{\varphi}(u; s) - \varphi(u; s) \right) \right| \leq v \right) \right| \xrightarrow{T \to \infty} 0
\]

as \( T \) tends to \( \infty \).

We aim to examine the finite-sample impact of different sample and window size choices with regard to coverage results in the ECF setup addressed in Jentsch et al. (2020, sec. 5). That is, we generate a locally stationary process \((X_{t,T})_{t=1}^T\) by

\[
X_{t,T} = \begin{cases} 
0.9 \sin \left( \frac{2\pi}{T} t \right) \varepsilon_t + \varepsilon_1, & t = 1, \\
0.9 \sin \left( \frac{2\pi}{T} t \right) X_{t-1,T} + \varepsilon_t, & t = 2, \ldots, T,
\end{cases}
\]

with \((\varepsilon_t)_{t \in \mathbb{Z}}\) forming an i.i.d. sequence and following an \( \alpha \)-stable marginal distribution with parameters \( \mu = 0, \alpha = 1.5, \beta = 0 \) and \( \gamma = 0.5 \). The innovations bequeaths the \( \alpha \)-stable distribution to the companion process with slightly different parameters:

\[
\tilde{\mu}(u) = 0, \quad \tilde{\alpha}(u) = 1.5, \quad \tilde{\beta}(u) = 0 \quad \text{and} \quad \tilde{\gamma}(u) = \frac{0.5}{1 - |0.9 \sin(2\pi u)|^{1.5}}.
\]

This leads to

\[
\varphi(u; s) = \exp \left( -\frac{0.5 |s|^{1.5}}{1 - |0.9 \sin(2\pi u)|^{1.5}} \right)
\]

as belonging characteristic function for the companion process. Remembering (6), we need to specify some other parameters, which are \( s = 6, u = 0.4 \) as well as \( b_T = T^{-0.35} \) and
Figure 2. Coverage for different half window sizes $TD_T$ with $L_T = 15$ in the first and $L_T = 25$ in the second row, both with $b_T = T^{-0.35}$ on the left and $b_T = T^{-0.4}$ on the right and for $T = 2000$ (•) and $T = 5000$ (°).

$b_T = T^{-0.4}$, respectively. Moreover, we choose the blocklength $L_T$ equal to 15 as well as to 25. Regarding the sample size, we look at both $T = 2000$ and $T = 5000$. Because of these choices, there is no need to consider endpoints as they are filtered out by the kernel function. Furthermore, our simulations are based on a significance level of 0.05. Figure 2 shows the increase of the coverage results for higher choices of the bootstrap window size towards the aimed 0.95. Overall, the coverage results for $T = 2000$ grow faster than those for $T = 5000$. In most cases, they also exceed the target value of 0.95 around $TD_T = 200$ and decline again for larger values of $TD_T$. For the larger sample size, in turn, there is only an upward trend discernible, but with more variation within for $b_T = T^{-0.35}$. As to $b_T = T^{-0.4}$, the variation within the upward trend is higher for $T = 2000$. Concerning the blocklengths, the coverage results for $L_T = 15$ start closer to the target value than those for $L_T = 25$. On the other hand, the differences between the values for $T = 2000$ and $T = 5000$ are smaller for the larger blocklength. A wider simulation study to examine parameter choice impact can be found in Beering (2021, ch. 5).

For a more involved example in the field of hypothesis testing based on $L_2$ statistics and relying on the FCLTs stated in Theorems 2.5 and 3.4 instead of the CLTs, see Beering (2021, ch. 7), where a characteristic function based test for local independence is established.

5. Proof of Theorem 2.5

Throughout this section, $C$ denotes a generic constant that may change its value from line to line.
Proof of Theorem 2.5: Following van der Vaart and Wellner (2000, Th. 1.5.4), we need to show convergence of the fidis and asymptotic tightness in order to prove process convergence. Finally, the continuity of the sample path of the limiting process can be concluded with the help of van der Vaart and Wellner (2000, Add. 1.5.8). Theorem 2.4 gives the required convergence of the fidis. Using van der Vaart and Wellner (2000, Th. 1.5.7), we show uniform equicontinuity. In view of

$$\lim_{T \to \infty} P \left( \sup_{s \in \mathcal{F}} \left| \sum_{t=1}^{T} w_{t,T} \tilde{f} \left( s, X_{t,T} \right) \right| > \lambda \right) = \lim_{T \to \infty} P \left( \sup_{s \in \mathcal{F}} \left| \sum_{t=1}^{T} w_{t,T} \tilde{f} \left( s, \tilde{X}_{t} \left( \frac{t}{T} \right) \right) \right| > \lambda \right),$$

which follows straightforwardly from

$$E \left( \sup_{s \in \mathcal{F}} \left| \sum_{t=1}^{T} w_{t,T} \left( f \left( s, X_{t,T} \right) - f \left( s, \tilde{X}_{t} \left( \frac{t}{T} \right) \right) \right) \right| \right) \leq C \sum_{t=1}^{T} w_{t,T} E \left| X_{t,T} - \tilde{X}_{t} \left( \frac{t}{T} \right) \right|_1 \leq C d_{T}^{-1/2},$$

it remains to show

$$\lim_{r \to 0} \lim_{T \to \infty} P \left( \sup_{\rho(\hat{s}_1, \hat{s}_2) < r} \left| \sum_{i \in R} w_{i,T} \left( \tilde{f} \left( \hat{s}_1, \tilde{X}_{i} \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \hat{s}_2, \tilde{X}_{i} \left( \frac{i}{T} \right) \right) \right) \right| > \lambda \right) = 0$$

for any $\lambda > 0$. For this purpose, we define

$$\kappa_T := \left\lfloor \frac{d_{T}^{\frac{1}{m_T}}}{} \right\rfloor \quad \text{and} \quad \mu_T := \left\lfloor \frac{d_{T}^{\frac{1}{2\kappa_T}}}{} \right\rfloor$$

for some case-specific $m > 1$ which will be particularised for the cases (a) and (b) in Assumption 2.4 later on. In the style of Arcones and Yu (1994), we divide our set of indices into blocks $H_t$, $T_t$ and $R$ in such a way that the indices of the first $\kappa_T$ non-negative weights are in $H_1$, the indices of the second $\kappa_T$ non-negative weights in $T_1$, the indices of the second $\kappa_T$ non-negative weights in $H_2$ and so on until we have eventually $\mu_T H$-blocks and $\mu_T T$-blocks each. The remaining indices are arranged in block $R$. We establish an upper bound for the RHS of (13) considering the $H$-blocks, $T$-blocks and the $R$-block separately. Regarding the last one, we obtain from Assumptions 2.2 and 2.4

$$\lim_{r \to 0} \lim_{T \to \infty} P \left( \sup_{\rho(\hat{s}_1, \hat{s}_2) < r} \left| \sum_{i \in R} w_{i,T} \left( \tilde{f} \left( \hat{s}_1, \tilde{X}_{i} \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \hat{s}_2, \tilde{X}_{i} \left( \frac{i}{T} \right) \right) \right) \right| > \frac{\lambda}{3} \right) \leq \lim_{r \to 0} \lim_{T \to \infty} C \sup_{\rho(\hat{s}_1, \hat{s}_2) < r} \sum_{i \in R} w_{i,T} E \left[ g \left( \tilde{X}_{i} \left( \frac{i}{T} \right) \right) \right] \rho(\hat{s}_1, \hat{s}_2) \leq \lim_{T \to \infty} C d_{T}^{-1/2} \kappa_T = 0.$$
variables whose indices are situated in different blocks $H_1, \ldots, H_{\mu_T}$ are independent, we rely on \( \tilde{\xi}_i(M) (\frac{1}{T}) \) with $M = \lfloor \kappa T/2 \rfloor$:

\[
\sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left( \tilde{f} \left( \xi_1, \tilde{\xi}_i \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) \right)
\]

\[
= \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left( \tilde{f} \left( \xi_1, \tilde{\xi}_i \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) \right)
\]

\[
+ \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left( \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) \right)
\]

\[
= : I_a + I_b + I_c.
\]

First, similar arguments as used in the proof of Lemma A.2 yield for $I_a$ (and similarly $I_b$)

\[
\lim_{r \to 0} \lim_{T \to \infty} \mathbb{E} \left( \sup_{\rho(\xi_1, \xi_2) < \kappa T} \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left| \tilde{f} \left( \xi_1, \tilde{\xi}_i \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) \right| \right)
\]

\[
\leq C \lim_{T \to \infty} \sum_{t=1}^{T} w_{i,T} \vartheta \left[ \kappa T/2 \right],
\]

which is zero. Hence, it remains to show asymptotic negligibility of

\[
\lim_{r \to 0} \lim_{T \to \infty} \mathbb{P} \left( \sup_{\rho(\xi_1, \xi_2) < \kappa T} \left| \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left( \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) \right) \right| > \frac{\lambda}{9} \right).
\]

(14)

Since the summands with different indices $t$ are independent we can make use of standard empirical process theory. To this end, consider $\xi, \xi_1, \xi_2 \in \mathcal{I}$ and define

\[
v_T(\xi) := \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \tilde{f} \left( \xi_1, \tilde{\xi}_i^{(M)} \left( \frac{i}{T} \right) \right) \quad \text{and} \quad v_T(\xi_1, \xi_2) := v_T(\xi_1) - v_T(\xi_2).
\]

Inspired by Arcones and Yu (1994), we use a classical chaining argument. For this purpose, let

\[
r_k := r 2^{-k}, \quad k = 0, \ldots, k_T,
\]

for some $r, k_T$ which will be specified thereinafter. Moreover, let $\mathcal{F}_k \subseteq \mathcal{I}$ be an index set satisfying

\[
\# \mathcal{F}_k = D(k) = D(r_k, \mathcal{I}, \rho) \quad \text{and} \quad \sup_{\xi_1 \in \mathcal{I}} \min_{\xi_2 \in \mathcal{F}_k} \rho(\xi_1, \xi_2) < r_k, \quad k \in \{0, \ldots, k_T\}.
\]

By Assumption 2.4(iii), it holds $D(k) \leq r_k^{-d}$ for $r > 0$ chosen sufficiently small. This gives us the existence of maps $\pi_k: \mathcal{I} \to \mathcal{F}_k$ for $k = 0, \ldots, k_T$ such that $|\xi - \pi_k \xi|_1 \leq r_k \forall \xi \in \mathcal{I}$.
Subsequently, we get the following two inequalities for \( \mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S} \) with \( \rho(\mathcal{S}_1, \mathcal{S}_2) < r \):

\[
\rho\left(\pi_0 \mathcal{S}_1, \pi_0 \mathcal{S}_2\right) \leq 3r \quad \text{and} \quad \rho\left(\pi_k \mathcal{S}_1, \pi_k^{-1} \mathcal{S}_2\right) \leq 3r_k, \quad k \in \{1, \ldots, k_T\}.
\]

Thus, we get

\[
\sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S}, \rho(\mathcal{S}_1, \mathcal{S}_2) < r} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big| \leq 2 \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S}, \rho(\mathcal{S}_1, \mathcal{S}_2) \leq r_k} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big| + \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S}, \rho(\mathcal{S}_1, \mathcal{S}_2) \leq 3r} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big| + 2 \sum_{k=1}^{k_T} \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{F}_k, \rho(\mathcal{S}_1, \mathcal{S}_2) \leq 3r_k} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big|.
\]

Again, we introduce some auxiliary quantities. Let

\[
\lambda_k := \frac{1}{r_k^{1+\frac{1}{4}}} \vee \left( \frac{4}{C} r_k^{-4} \log D(k) \right)^{1/2}, \quad k \in \{1, \ldots, k_T\},
\]

for a constant \( \tilde{C} \in (0, \infty) \) specified later, which may take different values in cases (a) and (b). We get

\[
\log D(k) \leq \lambda_k^2 \tilde{C} r_k^{-1}. \tag{16}
\]

Additionally, let \( r \) be small enough to allow for

\[
4 \sum_{k=1}^{\infty} \lambda_k \leq \frac{\lambda}{27}. \tag{17}
\]

Since we have \( D(k) = \Theta\left(r_k^{-d}\right) \), summability of \( (\lambda_k r_k)^k \) for \( T \to \infty \) is assured. At this point, we come back to (14). With the preassigned notation and (17), we can split up as follows:

\[
P \left( \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S}, \rho(\mathcal{S}_1, \mathcal{S}_2) < r} \sum_{i=1}^{\mu_T} \sum_{H_i} w_{i, T} \left| f(\mathcal{S}_1, \bar{X}_j(M) \left( \frac{i}{T} \right)) - \tilde{f}(\mathcal{S}_2, \bar{X}_j(M) \left( \frac{i}{T} \right)) \right| > \frac{\lambda}{9} \right)
\]

\[
\leq P \left( \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{S}, \rho(\mathcal{S}_1, \mathcal{S}_2) \leq r_k} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big| > \frac{\lambda}{27} \right)
\]

\[
+ P \left( \sum_{k=1}^{k_T} \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{F}_k, \rho(\mathcal{S}_1, \mathcal{S}_2) \leq 3r_k} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big| > 4 \sum_{k=1}^{k_T} \lambda_k \right)
\]

\[
+ P \left( \sup_{\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{F}_0, \rho(\mathcal{S}_1, \mathcal{S}_2) \leq 3r} \big| \nu_T(\mathcal{S}_1, \mathcal{S}_2) \big| > \frac{\lambda}{27} \right)
\]

\[
=: I + II + III. \tag{18}
\]

In order to show asymptotic negligibility of terms II and III, we want to make use of Bernstein’s inequality for sums of independent random variables exerted on the outer sum of
\( \nu_T \). Term I, however, will be discussed by using a symmetrisation lemma at the end of the proof.

The remaining part of the proof presumes Assumption 2.4(a) to hold. For (b) see Lemma A.6 in the Appendix. Before starting with the examination of term II in (18), we specify the lower bound of \( m \) as \( m > \frac{\delta^{5} - 4 - 3\delta^2}{2(1 - \delta^2)} \). Moreover, we need \( r_k \) in (15) to meet the following bounding conditions:

\[
T \leq \frac{\text{Cov}_\nu \left( s_1, X_i^1 \right)}{\text{var}_\nu \left( s_1, s_2 \right)} \leq T \quad \text{for} \quad k \geq 1, \quad T \leq \frac{\text{Cov}_\nu \left( s_1, X_i^1 \right)}{\text{var}_\nu \left( s_1, s_2 \right)} \leq T \quad \text{for} \quad k \geq 1.
\]

Note that our choice of \( \nu \) guarantees that the left-hand side (LHS) is strictly smaller than the RHS.

Now we turn our attention to the second summand in (18). To be able to apply Bernstein's inequality, we need to establish an upper bound for the variance of the inner sum of \( \nu_T \). Consider \( l := |i_1 - i_2| \). Then,

\[
\text{Var} \left( \nu_T \left( s_1, s_2 \right) \right) \leq \sum_{t=1}^{\mu_T} \sum_{i_1, i_2 \in H_t} w_{i_1, T} w_{i_2, T} \left| \text{Cov} \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_2}{T} \right) \right) \right) \right|
\]

\[
- \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_1}{T} \right) \right) \right),
\]

\[
+ \sum_{t=1}^{\mu_T} \sum_{i_1, i_2 \in H_t} w_{i_1, T} w_{i_2, T} \left| \text{Cov} \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_2}{T} \right) \right) \right) \right|
\]

\[
- \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_2}{T} \right) \right) \right),
\]

for \( M(l) := \lfloor \min \{M, l/2 \} \rfloor \). For the first covariance of (20) (and similarly for the second), we have

\[
\left| \text{Cov} \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_1}{T} \right) \right) \right) \right|
\]

\[
- \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_1}{T} \right) \right) \right),
\]

\[
f \left( s_1, X_i^1 \left( \frac{i_2}{T} \right) \right) - f \left( s_2 \tilde{X}^M_i \left( \frac{i_2}{T} \right) \right),
\]

\[
\leq \text{Cov} \left( f \left( s_1, X_i^1 \left( \frac{i_1}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_1}{T} \right) \right), f \left( s_1, X_i^1 \left( \frac{i_2}{T} \right) \right), f \left( s_2 \tilde{X}^M_i \left( \frac{i_2}{T} \right) \right) \right)
\]

\[
- f \left( s_2 \tilde{X}^M_i \left( \frac{i_2}{T} \right) \right).
\]
\[
\begin{align*}
&+ \left| \text{Cov} \left( f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) - f \left( \tilde{x}_{11}^{(M(I))} \left( \frac{i_1}{T} \right) \right) \right) \\
&\quad \quad - f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M(I))} \left( \frac{i_2}{T} \right) \right) \right| \\
&f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) \right| 
\end{align*}
\]  

(21)

and, again, we only examine the first covariance of (21). Similarly to Lemma A.2 we have

\[
\begin{align*}
&\left| \text{Cov} \left( f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) - f \left( \tilde{x}_{11}^{(M(I))} \left( \frac{i_1}{T} \right) \right) \right) \\
&\quad \quad - f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M(I))} \left( \frac{i_2}{T} \right) \right) \right| \\
&\quad \quad \leq E \left| f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) - f \left( \tilde{x}_{11}^{(M(I))} \left( \frac{i_1}{T} \right) \right) \right| \\
&\quad \quad \times \left| f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M(I))} \left( \frac{i_2}{T} \right) \right) \right| \\
&\quad \quad + C \rho (\xi_1, \xi_2) \sum_{M > |j| \geq M(I)} \tilde{g}^j,
\end{align*}
\]

(22)

In the later following calculations to bound the variance of \( \nu_T \), we will need two suitable but different bounds. Therefore, we establish two alternative bounds for the first summand of (22):

(i) Using Hölder’s inequality, we get

\[
E \left| f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) - f \left( \tilde{x}_{11}^{(M(I))} \left( \frac{i_1}{T} \right) \right) \right| \\
\times \left| f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M(I))} \left( \frac{i_2}{T} \right) \right) \right| \\
\leq C \| \xi_0 \|_{2+\delta} \sum_{M > |j| \geq M(I)} \tilde{g}^j.
\]

(ii) On the other hand, using similar arguments as in the proof of Lemma A.3, we get

\[
\sup_{M \in \mathcal{N}} \left\| f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) \right\|_{2+\delta} \left\| f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M(I))} \left( \frac{i_2}{T} \right) \right) \right\|_{2+\delta} \\
\leq C \rho (\xi_1, \xi_2).
\]

These two bounds for (22) and similarly for (21) allow us to bound the covariance in (20) via

\[
\left| \text{Cov} \left( f \left( \tilde{x}_{11}^{(M)} \left( \frac{i_1}{T} \right) \right) - f \left( \tilde{x}_{11}^{(M(I))} \left( \frac{i_1}{T} \right) \right) \right) \\
- \left( f \left( \tilde{x}_{22}^{(M)} \left( \frac{i_2}{T} \right) \right) - f \left( \tilde{x}_{22}^{(M(I))} \left( \frac{i_2}{T} \right) \right) \right) \right|,
\]
Thus, we obtain for any $R_0 \geq 2$

\[
\text{Var} \left( v_T (\xi_1, \xi_2) \right) \leq C \mu_T \kappa_T d_T^{-1} \sum_{t=0}^{d_T} \min \left\{ \rho (\xi_1, \xi_2), t^{-\frac{1+\delta}{2}} \right\} \leq C \left( R_0 \rho (\xi_1, \xi_2) + R_0^{-1/\delta} \right).
\]

With $R_0 := [\rho (\xi_1, \xi_2)^{\frac{\delta}{1+\delta}}]$ and for any $r$ chosen sufficiently small, we get

\[
\text{Var} \left( v_T (\xi_1, \xi_2) \right) \leq C \rho (\xi_1, \xi_2)^{\frac{1}{1+\delta}}. \tag{23}
\]

Since we aim at the application of Bernstein’s inequality to bound II in (18), we first provide a suitable approximation of II by a sum of bounded random variables. To this end, we define

\[
\hat{f} \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) := f \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) \mathcal{K}_{\Omega_{\text{sup},t}},
\]

\[
\tilde{f} \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) = \hat{f} \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) - \mathbb{E} \hat{f} \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right)
\]

with

\[
\Omega_{\text{sup},t} := \left\{ \omega \in \Omega \mid \sup_{\xi \in \mathcal{X}, i \in H_T} w_{i,T} \left| f \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) \right| \leq d_T^{-\frac{\delta}{4+3\delta}} \right\}.
\]

Note that in view of compactness of $(\mathcal{X}, \rho)$, we have

\[
P \left( \Omega \setminus \Omega_{\text{sup},t} \right) \leq P \left( \sup_{\xi \in \mathcal{X}, i \in H_T} w_{i,T} \left| f \left( \xi, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) - f \left( 0, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) \right| > \frac{1}{2} d_T^{-\frac{\delta}{4+3\delta}} \right) + P \left( \sup_{\xi \in \mathcal{X}, i \in H_T} w_{i,T} \left| f \left( 0, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) \right| > \frac{1}{2} d_T^{-\frac{\delta}{4+3\delta}} \right)
\]

\[
\leq C \sum_{i \in H_T} w_{i,T}^{2+\delta} d_T^{2+\delta} \left[ E \left[ g \left( \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) \right] \right]^{2+\delta} + E \left| f \left( 0, \tilde{X}_i (M) \left( \frac{t}{T} \right) \right) \right|^{2+\delta}
\]

\[
\leq C d_T^{-\frac{2}{2(4+3\delta)}-1} + \frac{1}{2m} \rightarrow o(1).
\]

Hence, II in (18) can be bounded from above by

\[
P \left( 2 \sum_{k=1}^{k_T} \sup_{\xi_1, \xi_2 \in \mathcal{X}_{k+1} \setminus \mathcal{X}_k \mid \rho (\xi_1, \xi_2) \leq \lambda_k} \left| v_T (\xi_1, \xi_2) \right| > 4 \sum_{k=1}^{k_T} \lambda_k \right)
\]
Having in mind that

\[ H_t \leq \rho(\xi_1, \xi_2) \leq 3r_k \]

apply Bernstein’s inequality and get

\[ \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left[ f^c \left( s_1, \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) - f^c \left( s_2, \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right] > \sum_{k=1}^{k_T} \lambda_k + \sum_{k=1}^{\mu_T} \left| C_{r_k} \sum_{i \in H_t} w_{i,T} E \left( g \left( \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right) \right| \geq \frac{\lambda_k}{\lambda_T} + o(1). \]  

Note that asymptotic negligibility of the middle term on the RHS of (24) follows from

\[ \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} E \left[ g \left( \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right] \leq Cd_T^{1/2} \left( d_T^{-\frac{1+3}{2m}} + 1 \right) \]

since \((r_k/\lambda_k)_k\) is uniformly bounded. To bound the first summand on the RHS of (24), we apply Bernstein’s inequality and get

\[ \sum_{k=1}^{k_T} \sum_{\xi_1, \xi_2 \in \mathcal{F}_T : \rho(\xi_1, \xi_2) \leq 3r_k} P \left( \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left[ f^c \left( s_1, \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) - f^c \left( s_2, \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right] > \lambda_k \right) \]

\[ \leq 2 \sum_{k=1}^{k_T} D(k)D(k-1) \exp \left( -\frac{1}{2} \cdot \frac{\lambda_k^2}{V_{H,k} + M_{k,k}} \right), \]  

(25)

where, by definition of \(H_t\) and \(\tilde{f}\),

\[ \tilde{M} := 4d_T^{\frac{4+3(2m-2m)}{2m(4+3m)}} \]

and, in view of (23), for some appropriately chosen \(C < \infty\)

\[ V_{H,k} := C r_k^{\frac{1}{1+\delta}} \]

\[ \geq \sup_{\xi_1, \xi_2 \in \mathcal{F}_T} \text{Var} \left( v_T \left( \xi_1, \xi_2 \right) \right) \]

\[ \geq \sup_{\xi_1, \xi_2 \in \mathcal{F}_T} \text{Var} \left( \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,T} \left( f^c \left( s_1, \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) - f^c \left( s_2, \bar{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right) \right). \]

Having in mind that \(r_{k,T} \leq r_k\) is fulfilled by construction, we take up on (25) to get

\[ \Pi \leq 2 \sum_{k=1}^{k_T} \exp \left( 2 \log \left( D(k) \right) - \frac{1}{2} \cdot \frac{\lambda_k^2}{C_1 r_k^\frac{1}{1+\delta} + C_2 r_k^\frac{1}{1+\delta} \lambda} \right) + o(1) \]
\[ \leq 2 \sum_{k=1}^{k_T} \exp \left( -\frac{\bar{C}}{2} \lambda^2 r_k \frac{1}{r_T^3} \right) + o(1) \] 

which vanishes with \( r \to 0 \) in view of (16) for suitably chosen constants \( C_1, C_2, \bar{C} < \infty \).

Concerning term III of (18), we can follow the same steps with \( V_{III} := Cr^{1/2} \) for some expeditiously chosen \( C < \infty \) and obtain

\[ III \leq \sum_{\xi_1 \in \mathcal{F}_k, \xi_2 \in \mathcal{F}_0} P \left( \sum_{t=1}^{\mu_T} \sum_{i \in H_t} w_{i,t} \left( \tilde{f} \left( \xi_1, \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) - \tilde{f} \left( \xi_2, \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right) > \frac{\lambda}{27} \right) \]

\[ + o(1) \]

\[ \leq 2D^2(0) \exp \left( -\frac{1}{2} \cdot \frac{C_1 \lambda^2}{V_{III} + \frac{M C \lambda}{\delta}} \right) + o(1) \] 

which vanishes with \( r \to 0 \) with the same arguments as before for suitably chosen constants \( C_1, C_2 < \infty \).

Now we move on with the remaining first summand in (18) and aim at verifying

\[ \lim_{r \to 0} \lim_{T \to \infty} E \left( \sup_{\xi_1, \xi_2 \in \mathcal{S}, \rho(\xi_1, \xi_2) \leq r_T} \left| v_T (\xi_1, \xi_2) \right| \right) = 0. \] 

(28)

Once again, we need some further notation. For \( \xi \in \mathcal{S} \) let

\[ L_{i,T} (\xi) := \sum_{i \in H_t} w_{i,t} f \left( \xi, \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \quad \text{and} \quad L_{i,T}^0 (\xi) := \xi_i L_{i,T} (\xi), \]

where \( (\xi_t)_{t=1}^{\mu_T} \) are i.i.d. Rademacher variables independent of \( (\xi_t)_{t \in \mathbb{Z}} \). As \( (L_{i,T}(\xi))_{t=1}^{\mu_T} \) consists of independent random variables by construction, we can apply a standard symmetrisation lemma (see e.g. van der Vaart and Wellner (2000, Lem. 2.3.1)) to get

\[ E \left( \sup_{\xi_1, \xi_2 \in \mathcal{S}, \rho(\xi_1, \xi_2) \leq r_T} \left| v_T (\xi_1, \xi_2) \right| \right) \leq 2E \left( \sup_{\xi_1, \xi_2 \in \mathcal{S}, \rho(\xi_1, \xi_2) \leq r_T} \left| \sum_{t=1}^{\mu_T} \left( L_{i,T}^0 (\xi_1) - L_{i,T}^0 (\xi_2) \right) \right| \right). \] 

(29)

Note that \( \sum_{t=1}^{\mu_T} L_{i,T}^0 \) has sub-Gaussian increments conditionally on \( L_{1,T}, \ldots, L_{\mu_T,T} \). This is the case since for \( \xi_1, \xi_2 \in \mathcal{S} \) and \( \eta > 0 \), we get by applying Hoeffding’s inequality

\[ P \left( \left| \sum_{t=1}^{\mu_T} L_{i,T}^0 (\xi_1) - L_{i,T}^0 (\xi_2) \right| > \hat{\rho}_{T,2} (\xi_1, \xi_2) \eta \right| L_{1,T}, \ldots, L_{\mu_T,T} \) \]

\[ \leq 2 \exp \left( -\frac{\hat{\rho}_{T,2} (\xi_1, \xi_2)^2 \eta^2}{2 \sum_{t=1}^{\mu_T} \left( L_{i,T} (\xi_1) - L_{i,T} (\xi_2) \right)^2} \right) \]
C. BEERING AND A. LEUCHT

Regarding the expectations on the RHS, we obtain

\[ = 2 \exp \left( -\frac{\eta^2}{2} \right) \]  

(30)

with the random semimetric

\[ \hat{\rho}_{T,2}(\hat{x}_1, \hat{x}_2) := \left( \sum_{t=1}^{\mu_T} (L_{t,T}(\hat{x}_1) - L_{t,T}(\hat{x}_2))^2 \right)^{1/2} \]  

(31)

on \( \mathcal{S} \). We aim at verifying (28) with the help of a maximal inequality for sub-Gaussian processes, which will be more convenient with a different semimetric. To obtain this new semimetric, we note that

\[ (L_{t,T}(\hat{x}_1) - L_{t,T}(\hat{x}_2))^2 \leq 2^{2+\delta} |L_{t,T}|^2 \frac{2+\delta}{\infty} |L_{t,T}|_{\text{Lip}}^\frac{4-\delta}{3} \rho(\hat{x}_1, \hat{x}_2)^\frac{4-\delta}{3} \]  

(32)

holds on \( (\mathcal{S}, \rho) \) with \( |L_{t,T}|_{\text{Lip}} \) denoting the Lipschitz constant of \( L_{t,T} \). By defining

\[ Q_T := 2^{2+\delta} \left( \sum_{t=1}^{\mu_T} |L_{t,T}|_{\infty}^\frac{2+\delta}{3} |L_{t,T}|_{\text{Lip}}^\frac{4-\delta}{3} \right)^{1/2}, \]

we get

\[ \mathbb{E}Q_T \leq C \left( \sum_{t=1}^{\mu_T} \left( \mathbb{E} \left| L_{t,T} \right|_{\infty}^{2+\delta} \right)^{1/3} \left( \mathbb{E} \left| L_{t,T} \right|_{\text{Lip}}^{4-\delta} \right)^{2/3} \right)^{1/2}. \]  

(33)

Regarding the expectations on the RHS, we obtain

\[
\mathbb{E} |L_{t,T}|_{\infty}^{2+\delta} \leq \left\| \sum_{i \in H_t} w_{i,T} \left( C_g \left( \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) + \left| f \left( 0, \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right| \right) \right\|_{2+\delta}^{2+\delta}
\]

\[
\leq C \left( \sum_{i \in H_t} d_{T}^{1/2} \left\| g \left( \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right\|_{2+\delta}^{2+\delta} + \left\| f \left( 0, \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right\|_{2+\delta}^{2+\delta} \right)^{2+\delta} 
\]

\[
\leq C \delta \frac{2+\delta}{2} d_{T}^{2+\delta} ,
\]

\[
\mathbb{E} |L_{t,T}|_{\text{Lip}}^{4-\delta} \leq \left( \sum_{i \in H_t} w_{i,T} \left\| g \left( \tilde{X}_i^{(M)} \left( \frac{i}{T} \right) \right) \right\|_{4-\delta}^{4-\delta} \right)^{\frac{4-\delta}{2}} \leq C \frac{\delta}{2} d_{T}^{4} .
\]

Together, we have for (33)

\[
\mathbb{E}Q_T \leq C \left( \sum_{t=1}^{\mu_T} \kappa_T^{(2+\delta) \frac{1}{3} + \frac{4-\delta}{2} \frac{3}{3}} \left( \frac{2+\delta}{3} + \frac{4-\delta}{2} \frac{3}{3} \right) \delta d_{T}^{\frac{4-\delta}{2}} \right)^{1/2} = C (\mu_T \kappa_T^2 d_{T}^{-1})^{1/2} \leq C d_{T}^{1/2}. \]  

(34)

From \( \hat{\rho}_{T,2}(\hat{x}_1, \hat{x}_2) \leq Q_T \rho(\hat{x}_1, \hat{x}_2) \) \( \frac{4-\delta}{2} =: \tilde{\rho}_T(\hat{x}_1, \hat{x}_2) \), we see that \( \tilde{\rho}_T \) is again a random semimetric as \( \frac{4-\delta}{2} \in (0, 1) \). Now we make use of van der Vaart and Wellner (2000, Cor. 2.2.8)
to get

$$E\left( \sup_{\xi_1, \xi_2 \in \mathcal{S}, \rho(\xi_1, \xi_2) \leq r_T} \left| \sum_{t=1}^{\mu_T} (L_{t,T}(\xi_1) - L_{t,T}(\xi_2)) \right| \right) \leq C_3 \int_0^{QTr_k_T} \left( \log (D(u, \mathcal{S}, \rho_T)) \right)^{1/2} du$$

$$\leq C_3 \int_0^{QTr_k_T} \left( \log \left( D \left( \left( \frac{u}{Q_T} \right)^{\frac{6}{4-\delta}}, \mathcal{S}, \rho \right) \right) \right)^{1/2} du$$

$$\leq CQT \int_0^{r_k_T} u^{-\frac{3}{4-\delta}} du$$

$$\leq CQT^{\frac{1-\delta}{6}}.$$

Returning to (29), we get from (19) and (??)

$$E\left( \sup_{\xi_1, \xi_2 \in \mathcal{S}, \rho(\xi_1, \xi_2) \leq r_T} \left| \sum_{t=1}^{\mu_T} (L_{t,T}(\xi_1) - L_{t,T}(\xi_2)) \right| \right) \leq CQT^{\frac{1-\delta}{6}} \leq Cd_T^{-\frac{1}{12m}},$$

which tends to 0 as $T \to \infty$, and the proof is completed.

Acknowledgments

We thank the editors and an anonymous referee for their helpful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is partly funded by the Volkswagen Foundation, Germany (Professorinnen für Niedersachsen des Niedersächsischen Vorab).

ORCID

Carina Beering http://orcid.org/0000-0003-1339-9439

Anne Leucht http://orcid.org/0000-0003-3295-723X

References

Anderson, R., and Sandsten, M. (2019), 'Inference for Time-Varying Signals Using Locally Stationary Processes', *Journal of Computational and Applied Mathematics*, 347, 24–35.

Andrews, B., Calder, M., and Davis, R.A. (2009), 'Maximum Likelihood Estimation for $\alpha$-Stable Autoregressive Processes', *The Annals of Statistics*, 37(4), 1946–1982.

Arcones, M.A., and Yu, B. (1994), 'Central Limit Theorems for Empirical and $u$-Processes of Stationary Mixing Sequences', *Journal of Theoretical Probability*, 7, 47–71.
Beering, C. (2021), ‘A Functional Central Limit Theorem and Its Bootstrap Analogue for Locally Stationary Processes with Application to Independence Testing’, Ph.D. thesis, Technische Universität Braunschweig.
Breedt, F.J., Davis, R.A., and Trindade, A.A. (2001), ‘Least Absolute Deviation Estimation for All-Pass Time Series Models’, The Annals of Statistics, 29, 919–946.
Brunotte, G.-N. (2022), ‘A Test of Independence Under Local Stationarity Based on the Local Characteristic Function’. http://doi.org/10.13140/RG.2.2.36779.31523
Bühlmann, P. (1995), ‘The Blockwise Bootstrap for General Empirical Processes of Stationary Sequences’, Stochastic Processes and Their Applications, 58, 247–265.
Cardinali, A., and Nason, G.P. (2010), ‘Costationarity of Locally Stationary Time Series’, Journal of Time Series Econometrics, 2(2), 1–35.
Dahlhaus, R. (1997), ‘Fitting Time Series Models to Nonstationary Processes’, The Annals of Statistics, 25, 1–37.
Dahlhaus, R. (2012), ‘Locally Stationary Processes’, Handbook of Statistics, 30, 351–413.
Dahlhaus, R., Richter, S., and Wu, W.B. (2019), ‘Towards a General Theory for Non-Linear Locally Stationary Processes’, Bernoulli, 25, 1013–1044.
Dahlhaus, R., and Subba Rao, S. (2006), ‘Statistical Inference for Time-Varying ARCH Processes’, The Annals of Statistics, 34, 1075–1114.
Dette, H., Preuß, P., and Vetter, M. (2011), ‘Measure of Stationarity in Locally Stationary Processes with Applications to Testing’, Journal of the American Statistical Association, 106, 1113–1124.
Doukhan, P., Lang, G., Leucht, A., and Neumann, M.H. (2015), ‘Dependent Wild Bootstrap for the Empirical Process’, Journal of Time Series Analysis, 36, 290–314.
Dowla, A., Paparoditis, E., and Politis, D.N. (2013), ‘Local Block Bootstrap Inference for Trending Time Series’, Metrika, 76, 733–764.
Feng, G. (2015), ‘Bootstrap Methods for Univariate and Multivariate Volatility’, Ph.D. thesis, Technische Universität Braunschweig.
Franzier, D.T., and Koo, B. (2021), ‘Indirect Inference for Locally Stationary Models’, Journal of Econometrics, 223, 1–27.
Hall, P., Horowitz, J.L., and Jing, B.-Y. (1995), ‘On Blocking Rules for the Bootstrap with Dependent Data’, Biometrika, 50, 561–574.
Jentsch, C., Leucht, A., Meyer, M., and Beering, C. (2020), ‘Empirical Characteristic Functions-Based Estimation and Distance Correlation for Locally Stationary Processes’, Journal of Time Series Analysis, 41, 110–133.
Karmakar, S., Richter, S., and Wu, W.B. (2022), ‘Simultaneous Inference for Time-Varying Models’, Journal of Econometrics, 227, 408–428.
Kreiss, J.-P., and Paparoditis, E. (2015), ‘Bootstrapping Locally Stationary Processes’, Journal of the Royal Statistical Society Series B: Statistical Methodology, 77, 267–290.
Künsch, H.R. (1989), ‘The Jackknife and the Bootstrap for General Stationary Observations’, The Annals of Statistics, 17, 1217–1241.
Lahiri, S.N. (2003), Resampling Methods for Dependent Data, New York: Springer.
Lanne, M., and Saikkonen, P. (2014), ‘Noncausal Vector Autoregression’, Econometric Theory, 29, 447–481.
Naik-Nimbalkar, U.V., and Rajarshi, M.B. (1994), ‘Validity of Blockwise Bootstrap for Empirical Processes with Stationary Observations’, The Annals of Statistics, 22, 980–994.
Paparoditis, E., and Politis, D.N. (2002), ‘Local Block Bootstrap’, Comptes rendus de l’Académie des sciences Paris, Série I, 335, 952–962.
Phandoidaen, N., and Richter, S. (2022), ‘Empirical Process Theory for Locally Stationary Processes’, Bernoulli, 28, 453–480.
Priestley, M.B. (1965), ‘Evolutionary Spectra and Non-Stationary Processes’, Journal of the Royal Statistical Society: Series B, 27, 204–237.
R Core Team (2022), R: A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing.
Sergides, M., and Paparoditis, E. (2008), ‘Bootstrapping the Local Periodogram of Locally Stationary Processes’, Journal of Time Series Analysis, 29, 264–299.
Sergides, M., and Paparoditis, E. (2009), ‘Frequency Domain Tests of Semiparametric Hypotheses for Locally Stationary Process’, Scandinavian Journal of Statistics, 36, 800–821.

van der Vaart, A.W., and Wellner, J.A (2000), Weak Convergence and Empirical Processes. With Applications to Statistics, New York: Springer.

Vogt, M. (2012), ‘Nonparametric Regression for Locally Stationary Time Series’, The Annals of Statistics, 40, 2601–2633.

Vogt, M. (2015), ‘Testing for Structural Change in Time-Varying Nonparametric Regression Models’, Econometric Theory, 31, 811–859.

Wieczorek, B. (2016), ‘Blockwise Bootstrap of the Estimated Empirical Process Based on $\psi$-Weakly Dependent Observations’, Statistical Inference for Stochastic Processes, 19, 111–129.