Functional estimation and change detection for nonstationary time series*

Fabian Mies
RWTH Aachen University

Abstract

Tests for structural breaks in time series should ideally be sensitive to breaks in the parameter of interest, while being robust to nuisance changes. Statistical analysis thus needs to allow for some form of nonstationarity under the null hypothesis of no change. In this paper, estimators for integrated parameters of locally stationary time series are constructed and a corresponding functional central limit theorem is established, enabling change-point inference for a broad class of parameters under mild assumptions. The proposed framework covers all parameters which may be expressed as nonlinear functions of moments, for example kurtosis, autocorrelation, and coefficients in a linear regression model. To perform feasible inference based on the derived limit distribution, a bootstrap variant is proposed and its consistency is established. The methodology is illustrated by means of a simulation study and by an application to high-frequency asset prices.

Keywords: gradual change; locally stationary process; p-variation; bootstrap inference

1   Introduction

While statistical theory has historically been concerned with the study of data which is identically distributed, or at least stationary, this temporal homogeneity is often violated in practice. For example, economic time series are typically heteroscedastic, which is especially prominent for returns of financial assets. Statistical inference which does not account for nonstationarity may thus lead to wrong conclusions. However, the treatment of nonstationary models poses methodological challenges. Specifying the temporal variation nonparametrically, inference is often based on asymptotic methods, and it is in general not clear how to define a suitable limit in a nonstationary setting. As a remedy, Dahlhaus (1997) suggested to regard the time index as a fraction of the sample size, which allows for infill asymptotics. This perspective leads to the concept of locally

stationary time series, which has since been extended and generalized, see e.g. Wu and Zhou (2011); Dahlhaus et al. (2019); Truquet (2019).

In this paper, we study a multivariate, locally stationary time series model given by the causal representation

$$X_{t,n} = G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t) \in \mathbb{R}^d, \quad t = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_t, \epsilon_{t-1}, \ldots) \in \mathbb{R}^\infty$ for iid random variables ϵ_i. For each fixed $u \in [0,1]$ and $n \in \mathbb{N}$, the sequence $(G_n(u, \boldsymbol{\epsilon}_t))_{t \in \mathbb{Z}}$ is stationary. By letting the kernel depend on the fraction $\tfrac{t}{n}$, the model explicitly accounts for nonstationarity. We assume that the kernel G_n converges to a limiting kernel G as $n \to \infty$ in $L_q(P)$, and that $u \mapsto G_n(u, \boldsymbol{\epsilon}_t)$ admits some form of regularity. The nonlinear model (1) has been introduced by Zhou and Wu (2009), and later been refined by Zhou (2013). In contrast to the existing studies, the regularity assumptions imposed on the mapping $u \mapsto G_n(u, \boldsymbol{\epsilon}_t)$ in this paper are much less restrictive, as we only require finiteness of some p-variation instead of Hölder continuity; see the detailed discussion in Section 2.

A statistician might be interested in various properties of the time series $X_{t,n}$. We denote the quantity of interest by a local parameter $\theta_u, u \in [0,1]$, or θ_u^n, which is a functional of the law of $G(u, \boldsymbol{\epsilon}_t)$, resp. $G_n(u, \boldsymbol{\epsilon}_t)$. The estimator proposed in this paper is applicable for any parameter of the form $\theta_u^n = f(\mu_u^n)$, where $\mu_u^n = \mathbb{E}G_n(u, \boldsymbol{\epsilon}_t)$, and f is a sufficiently smooth function. Upon replacing $X_{t,n}$ by $Y_{t,n} = h(X_{t,n})$ for a function h, we may also study parameters $\theta_u^n = f(\mathbb{E}[h(G_n(u, \boldsymbol{\epsilon}_t))])$, i.e. any parameter which may be expressed as a function of moments of $X_{t,n}$. This framework is rather general, and contains for example the variance, kurtosis, and autocorrelations at fixed lag.

Depending on the application, different functionals of the temporal variation $u \mapsto \theta_u$ might be of interest, e.g. its temporal average $\int_0^1 \theta_u \, du$ as studied by Potiron and Mykland (2020), its maximum value $\sup_{u \in [0,1]} \|\theta_u\|$, or its value θ_{u_0} at some $u_0 \in [0,1]$. For instance, estimation of θ_{u_0} is studied by Cui et al. (2020), who estimate the autocovariance function of a locally stationary time series via polynomial smoothing, and by Dahlhaus and Richter (2019). The latter example is a nonparametric problem, and thus in general suffers from slow rates of convergence, depending on the regularity of $u \mapsto \theta_u$. In this paper, we tackle the temporal variation by studying the integrated parameter

$$u \mapsto \Theta(u) = \int_0^u \theta_v \, dv,$$

and we propose a corresponding estimator $M_n(u)$ of $\Theta(u)$. The function $\Theta(u)$ contains all information about the local parameter θ_u and is thus a nonparametric object as well. However, from a statistical perspective, it is attractive to formulate hypotheses in terms of $\Theta(u)$ since the latter may be estimated at a parametric rate $\sqrt{n}$, as demonstrated by our functional central limit theorem presented in Section 3. We note that the idea of recovering a $\sqrt{n}$ rate of convergence via integration has also been employed in other areas of nonparametric estimation, e.g. when performing inference for integrated squared density derivatives in an iid setting (Hall and Marron, 1987; Bickel and Ritov, 1988), for convolutions of nonparametrically specified densities (Schick and Wefelmeyer, 2004),

2

and for quadratic integrals of derivatives of a regression function (Huang and Jianqing, 1999). The previous references consider estimation of a single integrated quantity, but the functional estimation of a local parameter θ_v via its integral function $u \mapsto \Theta(u)$ is also common practice in high-frequency econometrics, when estimating integrated volatility or integrated nonlinear functionals of volatility; see Jacod and Rosenbaum (2013), Aït-Sahalia and Jacod (2014), and the references therein. In the context of nonstationary time series, this is applied, e.g., by Dahlhaus (2009), who considers linear functionals of the time-varying spectral density.

A general method to estimate integrated parameters has been suggested by Potiron and Mykland (2020). They construct block-wise estimators $\hat{\theta}_{i,n}$ and average them to obtain an estimator of $\Theta(1)$. While this approach could be adapted to the functional estimation of $u \mapsto \Theta(u)$, the verification of their assumptions for the estimators $\hat{\theta}_{i,n}$ entails additional analytical effort for each special case. In particular, they require strong conditions on the bias of $\hat{\theta}_{i,n}$, and present explicit debiasing procedures for specific examples. In contrast, our proposed estimator $M_n(u)$ is based on a linearization procedure around a nonparametric pilot estimator $\hat{\mu}_{t,n}$, and may be regarded as a generic approach for removing leading bias terms. We suggest a pilot estimator based on local smoothing, but our results are deliberately formulated under much weaker conditions, requiring only assumptions on the rate of convergence of $\hat{\mu}_{t,n}$. Many modern approaches to filtering and regression are based on statistical learning theory, which typically yields satisfactory rates of convergence, but does not lend itself to statistical inference. Our linearized estimator $M_n(u)$ thus enables rigorous asymptotic inference based on these pilot estimates. Details are presented in Section 3.

Our estimator is particularly useful to test for change-points in the local parameter. Here, the null hypothesis is that the parameter θ_u is the same for all $u \in [0, 1]$, which may be formulated equivalently as

$$H_0 : \theta_u \equiv \theta_0 \quad \Leftrightarrow \quad H_0 : \Theta(u) \equiv u\Theta(1). \tag{2}$$

Analysis of this hypothesis of structural stability has a long history in statistics, see Aue and Horváth (2013) for a recent review. Early studies were concerned with the stability of the mean (Page, 1954, 1955). The methodology has since been extended, and there exist procedures to test for, e.g., the stability of variances (Gao et al., 2019), regression coefficients (Horváth, 1995), or autocovariances (Berkes et al., 2009; Killick et al., 2013; Preuss et al., 2015). Our approach provides a unifying framework to study these problems for parameters θ_u which may be written as a function of nonlinear moments. Besides the mentioned examples, this also includes novel change-point tests which have not been studied previously, e.g. a test for the temporal stability of kurtosis. The proposed change-point tests are robust against various model misspecifications, e.g. nonstationarity and nuisance changes under the null hypothesis. In particular, our test only monitors changes in the parameter $\theta_u = f(\mu_u)$, but not in the parameter μ_u itself. For example, our statistic is sensitive to changes in the variance, but robust to changes in the unknown and time-varying mean value. The change-point tests based on our estimator $M_n(u)$ are discussed in greater detail in Section 4 below.

There are parameters of interest which may not be expressed as a function of finitely many moments, for example quantiles of the marginal distribution, or functionals of the local spectral measure as considered by Dahlhaus (2009). A very general framework is presented by Shao and Zhang (2010), where an arbitrary functional of the time series' distribution is studied. The assumptions therein are formulated in terms of the influence function corresponding to the statistical functional of interest. Their verification is far from simple and basically amounts to proving claims very similar to the steps we take in this article. In contrast, we believe that the conditions imposed in the present paper are conveniently verified for a broad range of practical problems. The framework of Shao and Zhang (2010) is also adopted by Dette and Gösmann (2020) and applied to the monitoring of quantiles, as well as by Gösmann et al. (2021). Note that these authors study the stationary case only, while we allow for nonstationarity. It might be of interest to extend the methodology introduced in the present paper to more general functionals. We leave this question for future work.

The outline of this paper is as follows. After defining the model in Section 2, we describe the functional estimator $M_n(u)$ in Section 3 and present our asymptotic results., we give the rigorous definition of our model and present the technical assumptions. The application of our results to change-point problems is discussed in Section 4, and the finite sample properties of our proposed procedure are assessed via simulations study in Section 5. Our methodology is illustrated by an application to financial data in Section 6. Appendix A in the supplement contains additional remarks, and Appendix B contains further examples and simulation results for change point testing. All technical proofs are deferred to Appendix C.

Notation

For a function $f : \mathbb{R}^d \to \mathbb{R}$, we denote $Df(x) = (\partial_{x_1}, \ldots, \partial_{x_d})f(x) \in \mathbb{R}^{1 \times d}$ the first order differential, and by $D^2 f(x) \in \mathbb{R}^{d \times d}$ the Hessian matrix. For two sequences a_n, b_n, we write $a_n \ll b_n$ if $a_n/b_n \to 0$ as $n \to \infty$. For $x \in \mathbb{R}$, we denote $(x)_+ = \max(x, 0)$. Weak convergence of probability measures and random elements is denoted by $\Rightarrow$. For a vector $x \in \mathbb{R}^d$, the Euclidean norm is denoted by $\|x\|$, and for a matrix $A \in \mathbb{R}^{d \times d}$, we denote by $\|A\| = \|A\|_{op}$ the operator norm, i.e. $\|A\| = \sup_{x \neq 0} \|Ax\|/\|x\|$. The spectral radius of a matrix A is denoted as $\rho(A) = \max(|\lambda_1|, \ldots |\lambda_d|)$, where the λ_i are the complex eigenvalues of A. The transpose of a matrix A is denoted by A^T. For a random vector X, we denote by $\|X\|_{L_q} = (\mathbb{E}\|X\|^q)^{\frac{1}{q}}$ the L_q norm, for $q \geq 1$. The notation $\|G_n\|_{p-var}$ is introduced in Section 2.

2   Model

Let $X_{t,n}, t = 1, \ldots, n$ be a triangular array of random variables which is causal in the sense that $X_{t,n} = G_n(t/n, \boldsymbol{\epsilon}_t)$ for a sequence of functions $G_n : \mathbb{R} \times \mathbb{R}^\infty \to \mathbb{R}^d$. Here, we denote

$$\boldsymbol{\epsilon}_t = (\epsilon_t, \epsilon_{t-1}, \ldots) \in \mathbb{R}^\infty,$$

where the ϵ_i are iid random variables. The functions G_n are assumed to be measurable, where we endow the sequence space $\mathbb{R}^\infty$ with the projection σ-Algebra, see (Billingsley, 1999, Example 1.2).

By using a sequence of functions G_n, we will be able to apply our results to investigate the power of the proposed tests against local alternatives. Furthermore, letting the kernel G_n depend on n allows us to account for potential discretization errors, see Example 1 below. We assume that the kernel G_n tends towards a limiting kernel $G : \mathbb{R} \times \mathbb{R}^\infty \to \mathbb{R}^d$, in the sense that

$$\sup_{u \in [0,1]} \|G_n(u, \boldsymbol{\epsilon}_0) - G(u, \boldsymbol{\epsilon}_0)\|_{L_q} \to 0, \qquad n \to \infty, \tag{A.1}$$

for some $q > 2$. Note that we do not require a rate of convergence in (A.1). Furthermore, we require the function $u \mapsto G_n(u, \boldsymbol{\epsilon}_0) \in L_q(P)$ to satisfy some regularity conditions. A very mild condition can be formulated in terms of p-variation for some $p \geq 1$, defined as

$$\|G_n\|_{p-var} = \sup_{\substack{0=u_0<u_1<\ldots<u_m=1 \\ m \in \mathbb{N}}} \left(\sum_{i=1}^m \|G_n(u_i, \boldsymbol{\epsilon}_t) - G_n(u_{i-1}, \boldsymbol{\epsilon}_t)\|_{L_q}^p \right)^{\frac{1}{p}}.$$

The latter definition is independent of the chosen t, since $\boldsymbol{\epsilon}_t \sim \boldsymbol{\epsilon}_0$. The p-variation of the limiting kernel is denoted analogously as $\|G\|_{p-var}$. We assume that

$$\sup_n \sup_{u \in [0,1]} \|G_n(u, \boldsymbol{\epsilon}_0)\|_{L_q} + \sup_n \|G_n\|_{p-var} \leq C_G < \infty, \tag{A.2}$$

and $p \geq 1$ will be further specified if necessary. Note that $\|G_n\|_{p-var} \leq \|G_n\|_{r-var}$ for $1 \leq r \leq p$, hence assumption (A.2) is stronger for smaller p.

The assumption of finite p-variation is less restrictive than the piecewise-locally-stationary (PLS) framework suggested by Zhou (2013), which amounts to requiring piecewise Lipschitz continuity with finitely many breakpoints. The PLS framework has been applied for change-point analysis by Dette and Wu (2019), Dette et al. (2019), among others. In contrast, finite p-variation still allows for infinitely many discontinuities of the mapping $u \mapsto G_n(u, \boldsymbol{\epsilon}_0) \in L_q(P)$. On the other hand, if the latter mapping is Hölder continuous with exponent β, then $\|G_n\|_{p-var} < \infty$ for $p \geq \frac{1}{\beta}$. Thus, assumption (A.2) is more general than requiring Hölder continuity, and it combines classical smoothness conditions as well as discontinuities in a single framework. The special case of bounded 1-variation has been considered by Dahlhaus and Polonik (2009) for linear processes. Our framework also contains the model of Dahlhaus et al. (2019) as a special case, see Section A.1 in the supplement.

The third and last assumption imposed on the causal kernel is uniform ergodicity. We employ the physical dependence measure introduced by Wu (2005). To define the dependence measure, we introduce an iid copy ϵ_i^* of the ϵ_i Denote for $t \in \mathbb{Z}, j \in \mathbb{Z}_{\geq 0}$,

$$\boldsymbol{\epsilon}_{t,j}^* = \left(\epsilon_t, \ldots, \epsilon_{t-j+1}, \epsilon_{t-j}^*, \epsilon_{t-j-1}^*, \ldots \right) \in \mathbb{R}^\infty,$$
$$\tilde{\boldsymbol{\epsilon}}_{t,j} = \left(\epsilon_t, \ldots, \epsilon_{t-j+1}, \epsilon_{t-j}^*, \epsilon_{t-j-1}, \ldots \right) \in \mathbb{R}^\infty.$$

We assume that there exists a value $\rho \in (0,1)$ such that, for all $j, n \in \mathbb{N}, u \in [0,1]$

$$\|G_n(u, \boldsymbol{\epsilon}_t) - G_n(u, \tilde{\boldsymbol{\epsilon}}_{t,j})\|_{L_q} \leq C_G \rho^j. \tag{A.3}$$

This implies that

$$\|G_n(u, \boldsymbol{\epsilon}_t) - G_n(u, \boldsymbol{\epsilon}_{t,j}^*)\|_{L_q} \leq \frac{C_G}{1-\rho} \rho^j,$$

see Proposition C.1 in the appendix.

This set of assumptions suffices to establish a functional central limit theorem for the partial sums of the $X_{t,n}$. An analogous limit theorem has been proven by Zhou (2013) under the more restrictive PLS assumption.

Theorem 2.1. *Let (A.1), (A.2), and (A.3) hold, for some $q > 2$. Suppose furthermore that $d = 1$, and denote by*

$$\sigma^2(u) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left[G(u, \boldsymbol{\epsilon}_h), G(u, \boldsymbol{\epsilon}_0)\right],$$

the local long-run-variance. Then, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nv \rfloor} [X_{t,n} - \mathbb{E}X_{t,n}] \Rightarrow B\left(\int_0^v \sigma^2(u)\,du\right),$$

where $B(v), v \geq 0$ is a standard Brownian motion. The weak convergence holds in the Skorokhod space $D[0,1]$.

By virtue of the Cramer-Wold device, Theorem 2.1 can be extended to the multivariate case to yield weak convergence in the space $(D[0,1])^d$ endowed with the product topology. Note that this topology is different from the Skorokhod topology on the space $D^d[0,1]$ of cadlag functions $g : [0,1] \to \mathbb{R}^d$, see (Jacod and Shiryaev, 2003, VI.1.23).

We conclude this section by giving an example of a process $X_{t,n}$ which satisfies assumptions (A.1)-(A.3), highlighting that the imposed conditions are rather weak.

Example 1. Consider the time-varying vector autoregressive (tvVAR) process in d dimensions which satisfies

$$X_{t,n} = A(\tfrac{t}{n})\left[X_{t-1,n} - \mu(\tfrac{t-1}{n})\right] + B(\tfrac{t}{n})\epsilon_t + \mu(\tfrac{t}{n}),$$

$$X_{0,n} = \sum_{i=0}^{\infty} A(0)^i B(0)\epsilon_{-i} + \mu(0),$$

where ϵ_t is a sequence of iid, d-dimensional random vectors with finite moments of all orders, $\mu : [0,1] \to \mathbb{R}^d$, and $A, B : [0,1] \to \mathbb{R}^{d \times d}$ are matrix valued functions. Note that higher-order autoregressive processes may also be studied in this framework by stacking the lagged values of the process, effectively increasing the dimension d. For

6

example, an autoregression of order two may be described in terms of the state vector $Y_{t,n} = (X_{t,n}, X_{t-1,n})$, taking values in $\mathbb{R}^{2d}$. Time-varying autoregressive models are classical examples of locally-stationary time series. Early investigations of these models include Subba Rao (1970) and Grenier (1983), and more recent contributions are due to Moulines et al. (2005), Dahlhaus and Polonik (2009), and Giraud et al. (2015), among others.

We extend A, B to the domain $(-\infty, 1]$ by setting $A(u) = A(0)$ and $B(u) = B(0)$ for $u < 0$. The autoregressive process may be cast into our framework $X_{t,n} = G_n(\frac{t}{n}, \boldsymbol{\epsilon}_t)$ as

$$
G_n(u, \boldsymbol{\epsilon}_t) = \sum_{i=0}^{\infty} \left[\prod_{j=1}^{i} A\left(u - \tfrac{j-1}{n}\right) \right] B\left(u - \tfrac{i}{n}\right) \epsilon_{t-i} + \mu(\tfrac{t}{n}).
$$

This infinite sum is well-defined if we suppose that $\sup_u \|B(u)\|, \sup_u \|A(u)\| < \infty$, that the spectral radius $\rho(A(u))$ of $A(u), u \in [0,1]$, is at most $\rho(A(u)) \leq \rho_0 < 1$, and that the function $u \mapsto A(u)$ admits some minimal regularity. In particular, if the function $u \mapsto A(u)$ has bounded p-variation for some $p \geq 1$, then $\| \prod_{j=1}^{i} A\left(u - \tfrac{j-1}{n}\right) \| \leq C\rho^i$ for any $\rho \in (\rho_0, 1)$, see Lemma C.8 in the appendix. Hence, assumption (A.3) is satisfied. Furthermore, if the functions μ, A, and B are left-continuous, we have $\|G_n(u, \boldsymbol{\epsilon}_0) - G(u, \boldsymbol{\epsilon}_0)\|_{L_q} \to 0$ by dominated convergence for any $q > 2$, with limiting kernel

$$
G(u, \boldsymbol{\epsilon}_t) = \sum_{i=0}^{\infty} A(u)^i B(u) \epsilon_{t-i} + \mu(u).
$$

In particular, condition (A.1) holds. Finally, Proposition C.9 in the appendix shows that assumption (A.2) holds if $\|A\|_{p-var} + \|B\|_{p-var} + \|\mu\|_{p-var} < \infty$. Moreover, in the more regular case that μ, A, and B are β-Hölder continuous, the kernel $G_n(u, \boldsymbol{\epsilon}_0)$ is also β-Hölder continuous in $L_q(P)$, uniformly in n, and thus the same holds for $u \mapsto \mu_u^n = \mathbb{E}G_n(u, \boldsymbol{\epsilon}_0)$. The latter property is required to apply Proposition 3.1 discussed in the following section.

3 Estimating integrated parameters

For the locally stationary model introduced in Section 2, we denote the local moments by

$$
\mu_u^n = \mathbb{E}G_n(u, \boldsymbol{\epsilon}_0) \in \mathbb{R}^d, \qquad \mu_u = \lim_{n \to \infty} \mu_u^n = \mathbb{E}G(u, \boldsymbol{\epsilon}_0) \in \mathbb{R}^d, \qquad u \in [0,1].
$$

Inference for the function μ_u^n resp. $\theta_u^n = f(\mu_u^n)$ can be performed in various ways. For example, one might assume a parametric form for the mapping $u \mapsto \mu_u^n$. Here, we are interested in testing nonparametric hypotheses imposed on the function μ_u. To this end, instead of treating the moment function directly, we suggest to consider its integral. In particular, for a nonlinear function $f : \mathbb{R}^d \to \mathbb{R}$, we study the integrated quantity

$$
F_n(u) = \int_0^u f(\mu_v^n)\, dv, \qquad u \in [0,1].
$$

Note that most hypotheses on $\theta_u^n = f(\mu_u^n)$ may be reformulated in terms of $F_n(u)$. In Section 4, we will study change-point detection, where the null hypothesis $H_0 : f(\mu_u^n) \equiv f(\mu_0^n)$ for $u \in [0,1]$ is equivalent to the hypothesis $H_0 : F_n(u) \equiv u F_n(1)$. It is also possible to study the hypothesis $H_0 : f(\mu_u^n) \equiv 0$, which is equivalent to $H_0 : F_n(u) \equiv 0$. Another appealing aspect of studying the integrated quantity $F_n(u)$ instead of $f(\mu_u^n)$ is that the former may be estimated at a parametric rate $\sqrt{n}$, even in a nonparametric framework, see our results below. Though this might be surprising at first sight, note that a similar phenomenon occurs in the classical statistical setting of iid observations, where the empirical distribution function is $\sqrt{n}$ consistent, while nonparametric density estimation only allows for slower, nonparametric rates of convergences.

A straightforward way to estimate $F_n(u)$ is to consider the partial sum process

$$\tilde{M}_n(u) = \frac{1}{n} \sum_{t=\tau}^{\lfloor un \rfloor} f(\hat{\mu}_{t,n}), \tag{3}$$

where $\hat{\mu}_{t,n}$ is a nonparametric estimator of $\mu_{\frac{t}{n}}$, and $\tau = \tau_n \geq 1$ may be introduced to alleviate boundary issues. This approach has two shortcomings. First, its distributional properties strongly depend on the specific estimator $\hat{\mu}_{t,n}$. Thus, additional theoretical effort is required when adapting corresponding inferential procedures to new situations. Second, for nonlinear f, the estimator $\tilde{M}_n(u)$ may be bias-dominated, impeding statistical inference.

In particular, a Taylor expansion yields

$$\frac{1}{n} \sum_{t=\tau}^{\lfloor un \rfloor} \left[f(\hat{\mu}_{t,n}) - f(\mu_{\frac{t}{n}}) \right] = \frac{1}{n} \sum_{t=\tau}^{\lfloor un \rfloor} Df(\mu_{\frac{t}{n}})(\hat{\mu}_{t,n} - \mu_{\frac{t}{n}}) + \mathcal{O}\left(\frac{1}{n} \sum_{t=\tau}^{n} \|\hat{\mu}_{t,n} - \mu_{\frac{t}{n}}\|^2 \right). \tag{4}$$

For most estimators $\hat{\mu}_{t,n}$, the bias of this expression is not smaller than $1/\sqrt{n}$. If one considers, for example, a local average with bandwidth k_n, and assumes $u \mapsto \mu_u$ to be Lipschitz continuous, then the bias of $\tilde{M}_n(u)$ is of order $\mathcal{O}(\frac{k}{n} + \frac{1}{k})$, which is at best $\mathcal{O}(1/\sqrt{n})$. In a related situation, Jacod and Rosenbaum (2013) suggest to solve this problem by undersmoothing, i.e. choosing $k_n \ll \sqrt{n}$, and correcting the quadratic bias term in (4) explicitly; see also (Potiron and Mykland, 2020, Section 4.2). However, the latter quadratic bias term strongly depends on the specific estimator $\hat{\mu}_{t,n}$, so that the approach of Jacod and Rosenbaum (2013) may not be easily transfered to different problems.

As a generic approach for asymptotically unbiased estimation of integrated functionals, we propose the linearized partial sum estimator

$$M_n(u) = \frac{1}{n} \sum_{t=\tau+L}^{\lfloor nu \rfloor} \left[f(\hat{\mu}_{t-L,n}) + Df(\hat{\mu}_{t-L,n})(X_{t,n} - \hat{\mu}_{t-L,n}) \right], \quad u \in [0,1]. \tag{5}$$

for some initial offset $\tau = \tau_n \geq 1$, lag $L = L_n \to \infty$, $L_n \ll \tau_n$, to be specified later, and assuming f to be sufficiently smooth. The pilot estimator $\hat{\mu}_{t,n}$ needs to satisfy minimal

high-level assumptions formulated below. Then, a Taylor expansion of f readily yields, for some $\tilde{\mu}_{t,n}$ between $\hat{\mu}_{t-L,n}$ and $\mu_{t/n}^n$,

$$
M_n(u) = \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} f\left(\mu_{\frac{t}{n}}^n\right) + \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} \left(\mu_{\frac{t}{n}}^n - \hat{\mu}_{t-L,n}\right)^T \frac{D^2 f\left(\tilde{\mu}_{t,n}\right)}{2} \left(\mu_{\frac{t}{n}}^n - \hat{\mu}_{t-L,n}\right)
$$

$$
+ \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} Df(\hat{\mu}_{t-L,n}) \left(X_{t,n} - \mu_{\frac{t}{n}}^n\right)
$$

$$
= I_n^1(u) + I_n^2(u) + I_n^3(u).
$$

It can be shown that $I_n^1(u) - F_n(u)$ converges to zero sufficiently fast. Moreover, if f is sufficiently regular, and the estimator $\hat{\mu}_{t,n}$ is good enough, then $\sqrt{n} I_n^2(u) \to 0$, whereas $\sqrt{n} I_n^3(u)$ is asymptotically unbiased and tends towards a Gaussian process.

We require the function f to have bounded first and second derivatives in a neighborhood of the path of μ_v. Formally, introduce the convex set $\mathcal{M} = \text{conv}\{\mu_u : u \in [0,1]\} \subset \mathbb{R}^d$, and its δ-neighborhood $\mathcal{M}^\delta = \{m \in \mathbb{R}^d : \|m - \tilde{m}\| < \delta \text{ for some } \tilde{m} \in \mathcal{M}\}$. We require that there exists a $\delta \in (0, \infty]$ and a $C_f < \infty$ such that

$$
|f(x)| + \|Df(x)\| + \|D^2 f(x)\| \leq C_f, \quad x \in \mathcal{M}^\delta. \tag{A.4}
$$

By restricting the boundedness assumption to the set $\mathcal{M}^\delta$, we may also consider functions of the form $f(x,y) = x/y$, if $y \geq c > 0$ on the set $\mathcal{M}^\delta$. Without localization, (A.4) would not hold for this choice of f.

Regarding the pilot estimator, we require that $\hat{\mu}_{t,n}$ is measurable w.r.t. the past innovations ϵ_t, i.e. $\hat{\mu}_{t,n}$ is a (potentially nonlinear) filter. By additionally introducing the lag L, the measurability condition on $\hat{\mu}_{t,n}$ serves to de-bias the term $I_n^3(u)$. In particular, as $L_n \to \infty$, the random vectors $Df(\hat{\mu}_{t-L,n})$ and $X_{t,n}$ decouple by virtue of (A.3). Furthermore, we require the estimator $\hat{\mu}_{t,n}$ to be consistent in the sense that

$$
\sum_{t=\tau_n}^{n} \|\hat{\mu}_{t,n} - \mu_{\frac{t}{n}}^n\|^2 = o_P(\sqrt{n}), \tag{A.5}
$$

$$
P\left(\hat{\mu}_{t,n} \in \mathcal{M}_\delta, \ t = \tau_n, \dots, n\right) = 1 + o(1). \tag{A.6}
$$

The initial offset τ_n allows to circumvent boundary issues of the estimator $\hat{\mu}_{t,n}$. The assumptions (A.5) and (A.6) are rather mild. Property (A.6) requires some weak form of uniform consistency of the estimator. If f is globally smooth, i.e. $\delta = \infty$, then (A.6) is vacuous, and $\tau_n = 1$ is a valid choice. Property (A.5) is a requirement on the rate of convergence of $\hat{\mu}_{t,n}$, in a form routinely studied in nonparametric statistics. The latter assumption is discussed in detail in Section A.2 of the supplement. If we are willing to impose some additional smoothness conditions, a suitable nonparametric estimator may be obtained by local averaging. In particular, we define

$$
\hat{\mu}_{t,n}^{NW} = \frac{1}{k_n \wedge t} \sum_{i=(t-k_n)\vee 1}^{t} X_{i,n},
$$

for a sequence $k_n \to \infty, k_n \ll n$. Note that $\hat{\mu}_{t,n}^{NW}$ can be interpreted as a one-sided kernel smoother of Nadaraya-Watson type.

Proposition 3.1. *Let (A.3) hold for some $q > 2$, and choose $k_n \gg n^{\frac{2}{q}}$, then $\hat{\mu}_{t,n}^{NW}$ satisfies (A.6) for any offset sequence $\tau_n \to \infty$. Suppose that for all n, we have $\mu_u^n = \mu_u^{n,1} + \mu_u^{n,2}$, such that $\|\mu_u^{n,2}\|_{p-var} \le a_n$ for $p \in [1,2)$, and such that $u \mapsto \mu_u^{n,1}$ is β-Hölder continuous with Hölder constant C, then*

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \hat{\mu}_{t,n}^{NW} - \mu_{\frac{t}{n}}^n \right\|^2 = \mathcal{O}_P \left(\frac{a_n k_n}{n} + \left(\frac{k_n}{n} \right)^{2\beta} + \frac{\log n}{k_n} \right).$$

In particular, $\hat{\mu}_{t,n}^{NW}$ satisfies (A.5) if $a_n k_n \ll \sqrt{n}$, and $\sqrt{n} \log(n) \ll k_n \ll n^{\frac{4\beta-1}{4\beta}}$. The latter condition is feasible for any $\beta > \frac{1}{2}$.

By allowing the vanishing discontinuous part $\mu^{n,2}$ in Proposition 3.1, we may account for potential discretization errors.

For the local smoother $\hat{\mu}_{t,n}^{NW}$, as well as for any other estimator $\hat{\mu}_{t,n}$ satisfying our assumptions (A.5) and (A.6), the functional estimator $M_n(u)$ admits a central limit theorem with parametric rate $\sqrt{n}$. Our main result may be formulated as follows.

Theorem 3.2. *Suppose that (A.1), (A.2), (A.3), (A.4), (A.5), and (A.6) hold, and let $p \in [1,4)$. Suppose that $\tau_n \ll n$, and that L_n satisfies $L_n \gg \log(n)^{1+a}$ for some $a \in (0,1)$, and that $L_n \ll n^{\frac{1}{p} - \frac{1}{4}}$, $L_n \ll n^{\frac{1}{2p}}$. Then, as $n \to \infty$,*

$$\sqrt{n} \left(M_n(u) - \frac{1}{n} \sum_{t=\tau_n + L_n}^{\lfloor un \rfloor} f(\mu_{\frac{t}{n}}^n) \right) \Rightarrow B \left(\int_0^u Df(\mu_v)\Sigma(v)Df(\mu_v)^T \, dv \right) = M(u),$$

$$\Sigma(v) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(G(v, \boldsymbol{\epsilon}_0), G(v, \boldsymbol{\epsilon}_h) \right) \in \mathbb{R}^{d \times d},$$

where B denotes a standard Brownian motion. The weak convergence holds in the Skorokhod space $D[0,1]$. If $p \in [1,2)$, it also holds that

$$\sqrt{n} \left(M_n(u) - \int_{\frac{\tau_n + L_n}{n} \wedge u}^{u} f(\mu_v^n) \, dv \right) \Rightarrow M(u).$$

Just as in Theorem 2.1, the Cramer-Wold device may be used to extend Theorem 3.2 to the multivariate setting where f takes values in $\mathbb{R}^{d^*}$. The functional weak convergence then holds in the product space $(D[0,1])^{d^*}$. If $p < 2$ and $\tau_n \ll n^{\frac{1}{2}}$, we may alternatively choose $\int_0^u f(\mu_v^n) \, dv$ as centering term, such that $\sqrt{n}[M_n(u) - \int_0^u f(\mu_v^n) \, dv]$ has the asymptotic distribution given in Theorem 3.2.

Remark 1. The suitable choice of the lag parameter L depends on the strength of the dependency of the time series. On the one hand, the bias term I_n^2 grows polynomially with L, i.e. $I_n^2 = \mathcal{O}_P(L_n^{\min(p,2)} n^{-2/\max(p,2)})$, see Lemma C.5 in the supplement. This

suggests to choose the lag as small as possible. On the other hand, L needs to be large enough such that $X_{t,n}$ and $\hat{\mu}_{t-L,n}$ decouple, in order for I_n^3 to be asymptotically unbiased. In particular, in view of assumption A.3, we need to ensure that $\rho^{L_n} \ll 1/\sqrt{n}$. To ensure that both, I_n^2 and the bias of I_n^3, are negligible, a conservative choice is $L_n = c \log(n)^2$ for some factor c. The sensitivity of our methodology with respect to L is assessed by simulations in Section 5.

Remark 2. In contrast to Theorem 2.1, the central limit theorem of the linearized estimator $M_n(u)$ requires $p < 4$, i.e. the kernel G_n needs to be more regular. This restriction is due to the bias incurred by the lag L. In particular, we can only ensure that $I_n^2 = o_P(1/\sqrt{n})$ if $p < 4$, see Lemma C.5 in the supplement. On the other hand, the criticality of $p = 2$ occurs because $|\frac{1}{n}\sum_{t=1}^n f(\mu_{\frac{t}{n}}^n) - \int_0^1 f(\mu_u^n)\,du| = \mathcal{O}(n^{-\frac{1}{p}}\|\mu^n\|_{p-var})$, see Lemma C.4 in the supplement. Note that both issues are not present in the linear case of Theorem 2.1.

The integrand of the asymptotic variance of the limit process $M(u)$ corresponds to the long-run variance under the local, stationary model $G(u, \epsilon_t)$. In particular, a direct application of the delta method shows that

$$\sqrt{n}\left[f(\tfrac{1}{n}\sum_{t=1}^n G(u, \epsilon_t)) - f(\mu_u)\right] \Rightarrow \mathcal{N}\left(0, Df(\mu_u)^T \Sigma(u) Df(\mu_u)\right). \tag{6}$$

If our model is indeed stationary, Theorem 3.2 shows that $M_n(1)$ has the same asymptotic distribution as (6). In this sense, accounting for the nonstationarity does not increase the asymptotic variance.

To perform feasible inference based on Theorem 3.2, we need to handle the unknown asymptotic variance process. A consistent estimator may be constructed via blocked subsampling, similar to the suggestion of Carlstein (1986).

Theorem 3.3. *Let the conditions of Theorem 3.2 hold for some $q > 4$, and $p \in [1, 4)$. Choose some $b_n \to \infty$ such that $b_n \ll n^{\frac{2}{3\max(p,2)}}$, $b_n \ll n^{\frac{q-4}{2q+4}}$. Then, as $n \to \infty$,*

$$Q_n(u) = \frac{1}{n}\sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \frac{1}{b_n}\left[Df(\hat{\mu}_{t-L,n})\sum_{i=1}^{b_n}(X_{t+i,n} - \hat{\mu}_{t-L,n})\right]^2$$

$$\xrightarrow{P} Q(u) = \int_0^u Df(\mu_v)\Sigma(v)Df(\mu_v)^T\,dv.$$

The convergence holds uniformly in $u \in [0, 1]$ since Q_n is monotone.

Theorem 3.3 is a special case of the slightly more general Theorem C.7 in the appendix. Note that the upper bound on b_n reduces to $b_n \ll n^{\frac{1}{3}}$ if $q \geq 10$.

The estimator $Q_n(u)$ may be used to perform inference based on $M_n(u)$ via the following multiplier bootstrap scheme, similar to Zhou (2013).

Theorem 3.4. *Let the conditions of Theorem 3.3 hold for some $q > 4$. Let $Y_t \sim \mathcal{N}(0, 1)$ be iid standard normal random variables, independent of the ϵ_i, and define the process*

$$\widehat{M}_n(u) = \frac{1}{\sqrt{n}}\sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} Y_t\left[\frac{1}{\sqrt{b_n}}\sum_{i=1}^{b_n} Df(\hat{\mu}_{t-L,n})(X_{t+i,n} - \hat{\mu}_{t-L,n})\right], \qquad u \in [0, 1].$$

11

Then the conditional distribution of $\widehat{M}_n$ given $\mathbb{X}_n = (X_{1,n}, \ldots, X_{n,n})$ converges weakly in the Skorokhod space to $M(u)$ in probability, where $M(u)$ is the limit process from Theorem 3.2.

It can also be shown that the bootstrap consistency of Theorem 3.4 holds under weaker rate constraints on $\hat{\mu}_{t,n}$, replacing the rate $o(\sqrt{n})$ by $o(n^{1-\kappa})$ for some $\kappa > 0$. However, a smaller value of κ requires stronger conditions on b_n, see Theorem C.7 in the appendix. The local smoother $\hat{\mu}_{t,n}^{NW}$ is still consistent in the non-smooth case, where only $\|\mu\|_{p-var} < \infty$, although at a slower rate. Hence, the latter estimator may still be utilized for consistent variance estimation. This is of particular interest for applications to change-point tests, as described in the following section, where the bootstrap procedure is still consistent under various alternative hypotheses.

4 Change-point detection

A major motivation to perform inference for the integrated parameter $F_n(u)$ resp. $F(u) = \int_0^u f(\mu_v)\, dv$ is that the estimator $M_n(u)$ may be used to test for change-points. Our framework lends itself to test the hypothesis

$$H_0 : f(\mu_u^n) = f(\mu_0^n) \text{ for all } u \in [0,1] \quad \leftrightarrow \quad H_1 : f(\mu_u^n) \neq f(\mu_0^n) \text{ for some } u \in [0,1]. \quad (7)$$

To perform a test for this problem, a common approach is to formulate the CUSUM statistic, which in our case reads as

$$T_n^* = \sup_{u \in [u_n, 1]} |T_n(u)|, \quad \text{where } T_n(u) = M_n(u) - \frac{u - u_n}{1 - u_n} M_n(1)$$

$$\text{and} \qquad u_n = \frac{\tau_n + L_n - 1}{n}. \tag{8}$$

The main result Theorem 3.2 yields that $T_n(u) \xrightarrow{P} F(u) - uF(1)$ as $n \to \infty$, which is identically zero if the null hypothesis holds. In this case, Theorem 3.2 yields that

$$\sqrt{n} T_n(u) \Rightarrow M(u) - uM(1),$$
$$\sqrt{n} T_n^* \Rightarrow T^* = \sup_{u \in [0,1]} |M(u) - uM(1)|,$$

where $M(u)$ is the Gaussian limit process. The limit distribution T^* may be approximated via the bootstrap procedure outlined in Theorem 3.4, i.e. by sampling the random variable $\widehat{T}_n^* = \sup_{u \in [0,1]} |\widehat{M}_n(u) - u\widehat{M}_n(1)|$, so that $(\widehat{T}_n^* | \mathbb{X}_n) \Rightarrow T^*$ by virtue of Theorem 3.4. In particular, denote by t_α the $1 - \alpha$ quantile of T^*, and by $t_{\alpha,n}$ the $1 - \alpha$ quantile of $(\widehat{T}_n^* | \mathbb{X}_n)$. In practice, the quantile $t_{\alpha,n}$ may be approximated up to arbitrary precision by sampling from the conditional distribution $\hat{T}_n^*$. The corresponding test procedure may then be formulated as follows.

Proposition 4.1. *Let the conditions of Theorem 3.4 hold, and denote by $t_{\alpha,n}$ the $1-\alpha$ quantile of the conditional distribution $\widehat{T}_n^* | \mathbb{X}_n$. If the null hypothesis (7) holds, and if $\mathrm{Var}(M(u_0)) > 0$ for some $u_0 \in (0,1)$, then*

$$\lim_{n \to \infty} P\left(T_n^* > t_{\alpha,n}\right) = \alpha, \quad n \to \infty.$$

Hence, rejecting H_0 if the test statistic T_n^* exceeds the critical value $t_{\alpha,n}$ leads to a test with nominal size $\alpha \in (0,1)$ asymptotically.

Although we focus on the uniform CUSUM test statistic, the functional central limit theorem for the process $M_n(u)$ also enables the consideration of alternative statistics, e.g. the MOSUM statistic introduced by Bauer and Hackl (1978), see also Chu et al. (1995), or the Cramér-von Mises statistic $\int_0^1 T_n(u)^2 \, du$.

A desirable property of change-point tests is robustness against nuisance changes. For the quantity $f(\mu_u^n)$ to be non-constant, it is necessary that the local moment μ_u^n changes. It is thus tempting to instead test the null hypothesis $H_0^* : \mu_u^n \equiv \mu_0^n$, which is methodologically simpler to achieve. For example, the methods of Zhou (2013) and Vogt and Dette (2015) are applicable to test for H_0^*. However, this approach bears the risk to falsely detect a change although $f(\mu_u^n)$ remains constant. For example, it might happen that the variance of a time series is non-constant, while the autocorrelation structure remains constant, as studied by Dette et al. (2019). Furthermore, Schmidt et al. (2020) tests for homoscedasticity with a non-constant mean function. A related approach is presented by Demetrescu and Wied (2018). By design, our test is only sensitive to changes in the quantity $f(\mu_u^n)$.

Another type of nuisance change might occur in the parameters which are not explicitly described by the local moment function μ_u^n. For example, when testing for changes in the mean of a heteroscedastic time series, the variance is a nuisance parameter not contained in the vector μ_u^n, but relevant for statistical inference, see Górecki et al. (2018) and Pešta and Wendler (2020). We account for this type of nonstationarity by working in a locally stationary framework which allows not only for heteroscedasticity, but also for a varying dependency structure. This has also been suggested by Zhou (2013), who designs a corresponding test for changes in the mean. The recent articles Vogt and Dette (2015), Dette et al. (2019), and Cui et al. (2020), also employ a locally stationary model.

Test statistics for the change point problem usually need to be standardized by an estimator of their asymptotic variance. However, variance estimators designed for the stationary case might be inconsistent under the alternative, resulting in a loss of power, see (Juhl and Xiao, 2009; Shao and Zhang, 2010) and the discussion therein. In contrast, our bootstrap procedure is consistent under the alternative where $f(\mu_u^n)$ is not constant and potentially discontinuous, see Theorem 3.4 and the discussion thereafter. Moreover, we may investigate the behavior of our test statistic under local alternatives at rate $1/\sqrt{n}$.

Proposition 4.2. *Suppose that the conditions of Theorem 3.2 hold, and assume furthermore that $\mu_u^n = \mu_u + \frac{1}{\sqrt{n}}\delta_u$ for some function $u \mapsto \delta_u \in \mathbb{R}^d$, such that $f(\mu_u) = f(\mu_0)$*

and $\|\delta\|_{p-var} < \infty$. Then

$$\sqrt{n}T_n^* \Rightarrow \sup_{u \in [0,1]} |T(u) + \Delta(u)|,$$

$$\Delta(u) = \int_0^u Df(\mu_v)\delta_v \, dv - u \int_0^1 Df(\mu_v)\delta_v \, dv,$$

where $T(u) = M(u) - uM(1)$, and $M(u)$ is the limit process from Theorem 3.2.

Note that under the local alternative of Proposition 4.2, the simple estimator $\hat{\mu}_{t,n}^{NW}$ still satisfies (A.5) and (A.6) if $u \mapsto \mu_u$ is smooth, see Proposition 3.1. Hence, Theorem 3.4 is still applicable so that the bootstrap is consistent, and the CUSUM test with bootstrapped critical values has non-trivial power against local alternatives in direction δ_v, given that $\Delta(u) \not\equiv 0$. We also point out that the test has power not only against abrupt changes, but also against changes which occur gradually in time.

The proposed procedure allows for a unified treatment of change-point tests for a wide range of parameters of interest, as demonstrated by the examples below. Previously, suitable test statistics have been constructed individually for these problems, while our results show that they may be treated in a rather generic way.

4.1 Changes in autocorrelation

The dependency structure of time series is commonly described in terms of their autocovariance function. It is thus natural to test the latter for structural stability, as suggested by Berkes et al. (2009). They construct a CUSUM test based on the partial sums which form the empirical autocovariance estimator at fixed lag, and derive limit theorems under the assumption of stationarity. A method to detect changes without fixing the lag is mentioned by (Steland, 2020, Example 3). The case of a nonparametric mean function is investigated by Li and Zhao (2013), and multiple change-points are studied by Preuss et al. (2015). The non-stationary case is investigated by Killick et al. (2013), although without a rigorous analysis of the type I error.

Alternatively, in the same univariate setting, Dette et al. (2019) test whether the autocorrelation $\text{Cor}(X_{t,n}, X_{t-h,n})$ remains constant, for some fixed $h > 0$. This problem is more involved, since it requires standardization by the marginal variances, which are an additional nuisance quantity. They allow the marginal variance to be non-constant and estimate it non-parametrically, in order to standardize the observations. Furthermore, Dette et al. (2019) study a nonlinear, locally stationary specification of the underlying time series. Hence, they account for potential nonstationarity under the null hypothesis.

We may formulate the problem to test for constant autocorrelations in our general framework. To this end, we set $Y_{t,n} = G_n(\frac{t}{n}, \epsilon_t) = (X_{t,n}, X_{t-h,n}, X_{t,n}^2, X_{t-h,n}^2, X_{t,n}X_{t-h,n})$ for $t = 1, \ldots, n$, assuming for simplicity that $X_{r,n} = \tilde{G}_n(0, \epsilon_r)$ for $r \le 0$. Now set $f : \mathbb{R}^5 \to \mathbb{R}, x \mapsto (x_5 - x_1 x_2)/\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)}$, so that $\text{Cor}(X_{t,n}, X_{t-h,n}) = f(\mathbb{E}Y_{t,n}) = f(\mu_{t/n}^n)$. Again, assumptions (A.1)-(A.3) are a direct consequence of the corresponding properties of $X_{t,n}$. The function f is bounded on any compact set $K \subset \{x \in \mathbb{R}^5 :$

$(x_3 - x_1^2) > 0, (x_4 - x_2^2) > 0\}$. Thus, if $\text{Var}(X_{t,n}) > 2\delta > 0$, then (A.4) holds for $\delta > 0$ as well. We may thus construct $M_n(u)$ based on this time series $Y_{t,n}$ and function f to obtain an estimator for the integrated autocorrelation $F_n(u) = \int_0^u \text{Cor}(G_n(v, \boldsymbol{\epsilon}_0), G_n(v, \boldsymbol{\epsilon}_{-h})) \, dv$. The corresponding CUSUM statistic satisfies (8).

The resulting CUSUM statistic is similar to the statistic suggested by Dette et al. (2019). However, our framwork allows for many potential choices of $\hat{\mu}_{t,n}$, while Dette et al. (2019) only consider a special case. Moreover, our assumptions regarding the regularity of G_n are weaker.

4.2 Further examples

The kurtosis of a random variable X is defined as $\text{Kurt}(X) = \mathbb{E}(X - \mathbb{E}X)^4 / \text{Var}(X)^2$. For a univariate time series $X_{t,n}$, let $Y_{t,n} = (X_t, X_t^2, X_t^3, X_t^4)$. Then $\text{Kurt}(X_{t,n})$ can be written as a function of $\mathbb{E}(Y_{t,n})$. In particular,

$$\text{Kurt}(X_{t,n}) = f(\mathbb{E}Y_{t,n}) = \frac{\mathbb{E}X_{t,n}^4 - 4\mathbb{E}X_{t,n}\mathbb{E}X_{t,n}^3 + 6(\mathbb{E}X_{t,n}^2)^2 - 3(\mathbb{E}X_{t,n})^4}{\mathbb{E}X_{t,n}^2 - (\mathbb{E}X_{t,n})^2}$$

If $\text{Var}(X_{t,n}) > c > 0$, then f satisfies (A.4) so that our results are applicable, and the CUSUM statistic (8) in combination with the bootstrap procedure yields a feasible change-point test. To the best of our knowledge, the proposed method is the first test for structural stability of the marginal kurtosis.

In a similar way, we may consider the skewness $\text{Skew}(X) = \mathbb{E}(X - \mathbb{E}X)^3 / \text{Var}(X)^{\frac{3}{2}}$ of a random variable provided that the variance of X is bounded away from zero, or the coefficient of variation $\text{CV}(X) = \sqrt{\text{Var}(X)}/\mathbb{E}(X)$ if the expectation of X is bounded away from zero. A further example which may be cast in our framework are time-varying autoregressive models, as presented in example 1, where the coefficients may be identified in terms of finitely many autocovariances by means of the Yule-Walker equations. Our methodology could thus be used to test for changes in the second autoregressive component of a univariate tvAR model.

In the supplement, we also discuss change point tests for the marginal variance (Section B.1) and for the coefficients of a linear regression model (Section B.2).

5 Finite sample performance

To assess the finite sample performance of our proposed change-point test, we evaluate its size and power properties via simulations. We consider the locally stationary autoregressive process $X_{t,n}$ given by

$$X_{t,n} = a(\tfrac{t}{n})X_{t-1,n} + \sigma(\tfrac{t}{n})\eta_t(\tfrac{t}{n}). \tag{9}$$

The innovations $\eta_t(\tfrac{t}{n})$ are chosen as independent, zero-mean random variables having a symmetrized Gamma distribution with shape parameter $\alpha(\tfrac{t}{n})$, standardized to unit

	$L_n = \lceil \log(n)^2 \rceil$			$L_n = \lceil \frac{1}{2}\log(n)^2 \rceil$			$L_n = \lceil \frac{1}{5}\log(n)^2 \rceil$			$L_n = \lceil \frac{1}{10}\log(n)^2 \rceil$			DWZ		
	H_0	H_1	H_2	H_0	H_1	H_2	H_0	H_1	H_2	H_0	H_1	H_2	H_0	H_1	H_2
$n = 100$	0.26	0.19	0.28	0.04	0.03	0.05	0.04	0.03	0.04	0.05	0.04	0.04	0.50	0.54	0.49
500	0.02	0.03	0.02	0.02	0.13	0.03	0.04	0.37	0.05	0.05	0.49	0.07	0.18	0.67	0.22
1000	0.01	0.18	0.01	0.02	0.55	0.04	0.04	0.81	0.09	0.06	0.89	0.11	0.13	0.91	0.21
5000	0.02	1.00	0.20	0.04	1.00	0.33	0.07	1.00	0.42	0.08	1.00	0.46	0.07	1.00	0.47
10000	0.03	1.00	0.56	0.06	1.00	0.68	0.08	1.00	0.74	0.09	1.00	0.76	0.07	1.00	0.78

Table 1: Size and power of the bootstrap-based CUSUM test for constant autocorrelation, with nominal level 10%. Reported values are based on 5000 independent samples of the test statistic.

variance. We use

$$\sigma(u) = 0.5 + |\sin(2\pi u)|, \qquad \alpha(u) = \begin{cases} 1, & u \leq 0.7, \\ 2, & u > 0.7, \end{cases}$$

and for $a(u)$, either of the three functions

$$a_0(u) = 0.2, \quad a_1(u) = 0.2 + \frac{u}{2}, \quad a_2(u) = 0.2 + \frac{u}{10}.$$

We want to test for stability of the lag-1 autocorrelation, which is equivalent to the stability of the autoregressive coefficient a. To this end, we apply the change-point test presented in Section 4.1 , in combination with the local estimator $\hat{\mu}_{t,n}^{NW}$.

As described in Remark 2, the lag parameter L should be chosen just big enough such that the functional dependence measure at lag L is negligible. We choose $L_n = c \log(n)^2$ and analyze the effect of the factor c below. Once L_n is specified, the smoothing bandwidth $k = k_n$ may be determined via cross-validation, by minimizing the prediction error

$$\Lambda(k) = \sum_{t=1}^{n-L} \|\hat{\mu}_{t,n}^{NW,k} - X_{t+L,n}\|^2,$$

where $\hat{\mu}_{t,n}^{NW,k}$ denotes the local average with bandwidth k. In our simulations, we consider bandwidths from the interval $k \in [n^{0.35}, n^{0.75}]$. Then, a natural choice for the offset is $\tau_n = k_n$. Finally, we choose the window size for the bootstrap procedure as $b_n = L_n$. Thus, L_n is the only parameter to be chosen manually. While a data-driven choice of the lag parameter L_n is desirable, deriving a corresponding method is out of scope of this article.

To find the critical value of the CUSUM test, for each individual sample, we resample $M = 10^3$ independent realizations based on the bootstrap approximation of Theorem 3.4. Equivalently, we may use the bootstrap samples to compute an approximate p-value for the CUSUM test statistic. When assessing the size of the CUSUM test, we set $a = a_0$, and for the power analysis, we set $a = a_1$ respectively $a = a_2$.

Table 1 presents the size and power of the proposed test in the presented example, for various values of n and different choices L_n. The power values correspond to the
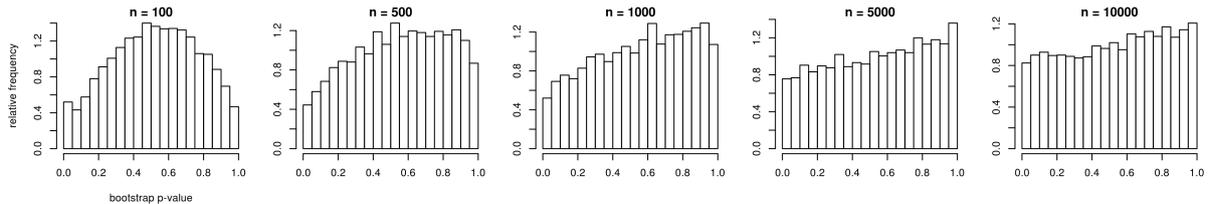
Figure 1: Distribution of bootstrap-based p-values for the CUSUM test for constant autocorrelation, under the null hypothesis, with lag parameter $L_n = \frac{1}{10}\log(n)^2$. The p-values are computed based on $M = 10^3$ bootstrap samples, and the histograms are based on 10^4 independent samples of the test statistic.

alternatives H_1 based on a_1, and H_2 based on a_2. We find that our test is rather conservative, i.e. the type-I error is actually smaller than the nominal level, and this conservativeness vanishes asymptotically as n increases. In particular, our test does not falsely detect a structural break even though the nuisance parameter is non-constant. On the other hand, the method consistently detects deviations from the null hypothesis, as demonstrated by the increasing power against the alternative. Moreover, it is found that the smallest lag value $L_n = \lceil \frac{1}{10}\log(n)^2 \rceil$ yields the best size approximation. Note that the differences in power for various choices of L_n may be partially explained by the different test sizes. For the latter choice of L_n, we also depict the distribution of the simulated p-values in Figure 1. The p-values should ideally be uniformly distributed. Indeed, for large sample size n, the accuracy of the p-values increases, in line with our theoretical results.

For comparison, we also implement the change point test for constant lag-1 auto-correlation proposed by Dette, Wu, and Zhou (Dette et al., 2019). The corresponding size and power are also presented in Table 1, labeled as DWZ. In small samples, the latter test achieves a higher power, which may be explained by a correspondingly higher rate of false positives. For large sample sizes, the power is similar to our proposed test, showing that our broadly applicable method is competitive against the specialized test of Dette et al. (2019).

We also assess the quality of $M_n(1)$ as an estimator of $\int_0^1 a_i(u)\,du$, $i = 0, 1, 2$, in comparison to the plug-in estimator $\tilde{M}_n(1)$. Table 2 presents the mean absolute errors of both estimators, as well as their corresponding bias. Except for the smallest sample size $n = 100$, the proposed linearized estimator performs better than the simple plug-in estimator. In particular, the linearization greatly decreases the bias of the estimator.

We also assess the finite sample performance of a change point test for the coefficients of a linear regression model. The simulation results are presented in Section B.3 of the supplement.

		$n = 100$		$n = 500$		$n = 1000$		$n = 5000$		$n = 10000$	
		M_n	$\tilde{M}_n$	M_n	$\tilde{M}_n$	M_n	$\tilde{M}_n$	M_n	$\tilde{M}_n$	M_n	$\tilde{M}_n$
H_0	error	0.172	0.120	0.053	0.058	0.031	0.039	0.012	0.018	0.008	0.013
	bias	0.071	-0.114	0.022	-0.054	0.011	-0.036	-0.001	-0.016	-0.002	-0.012
H_1	error	0.242	0.197	0.061	0.093	0.035	0.066	0.011	0.031	0.008	0.023
	bias	0.163	-0.197	0.039	-0.093	0.020	-0.066	0.002	-0.031	-0.001	-0.023
H_2	error	0.180	0.134	0.053	0.062	0.032	0.044	0.012	0.020	0.008	0.014
	bias	0.084	-0.131	0.025	-0.060	0.012	-0.043	-0.000	-0.019	-0.002	-0.014

Table 2: Mean absolute error and bias of $M_n(1)$ and $\tilde{M}_n(1)$ as estimators of $\int_0^u a_i(u)\,du$, $i = 0, 1, 2$, with $L_n = \frac{1}{10}\log(n)^2$ for the linearized estimator. All values are based on 5000 simulations.

6 Empirical illustration

To demonstrate the use of our results in practice, we study an application to high-frequency financial data. In particular, we study the price p_t of the german mid-cap stock-index MDAX on April 4, 2016, from 9:00-15:30, at a sampling frequency of 1 second. The data is available as a free sample from the data shop of Deutsche Börse, and part of the supplementary material of this article. We study the log-returns $d_t = \log(p_t) - \log(p_{t-1})$, $t = 1, \ldots, n$, with sample size $n = 23400$.

While many models for asset prices imply uncorrelated returns, empirical research suggests that autocorrelation may be non-zero, especially at high sampling frequencies (Hansen and Lunde, 2006). This dependence structure is typically attributed to the microstructure of the market, e.g. rounding effects or bid-ask rebounds. Recently, Andersen et al. (2019) studied the intraday returns of the NASDAQ100 stocks and found evidence for autocorrelation which is not only non-zero, but also non-constant.

Using the framework laid out in Section 4.1 above, we may rigorously perform asymptotic inference for the local autocorrelation $\text{Cor}(d_t, d_{t-1})$. To this end, we use the local estimator $\hat{\mu}_{t,n}^{NW}$ and choose its bandwidth via cross-validation as in Section 5, with $L_n = \lceil \log(n)^2/10 \rceil$.

The functional estimator $M_n(u)$ of the lag-1 autocorrelation is depicted in Figure 2 (left). First, we observe that $M_n(u)$ is roughly increasing, which indicates that the lag-1 autocorrelation is positive on average. Indeed, the average lag-1 autocorrelation is estimated as $M_n(1) = 0.1314$, with asymptotic standard deviation $\sqrt{Q_n(1)} = 0.0209$. Moreover, visual inspection of $M_n(u)$ suggests that the slope is varying, which corresponds to a non-constant autocorrelation. To test this hypothesis rigorously, we perform the CUSUM test suggested in Section 4. The right panel of Figure 2 shows the CUSUM process $T_n(u)$, and the critical thresholds for a significance level of 10% and 5%, respectively. The critical values are obtained using the bootstrap approximation of Theorem 3.4, with 10^4 bootstrap samples. The bootstrap-based p-value of the CUSUM test statistic is 0.167, based on 10^4 bootstrap samples. Note that the sample size n is rather large, and the simulation results of Section 5 suggest that the bootstrap approx-
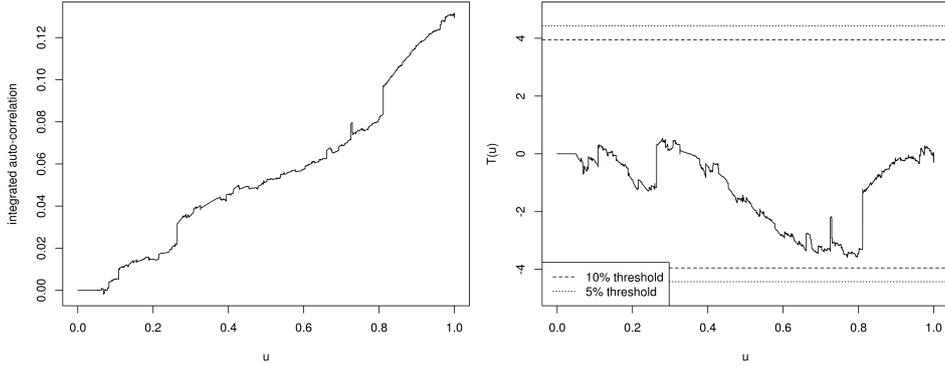
Figure 2: Left: the estimator $M_n(u)$ of the integrated autocorrelation. Right: the corresponding CUSUM process $\bar{T}_n(u)$. The critical thresholds are based on the bootstrap approximation with 10^4 bootstrap samples.

Figure 3: Left: the estimator $M_n(u)$ of the integrated autocorrelation fo the transformed log returns $\tilde{d}_t$. Right: the corresponding CUSUM process $\bar{T}_n(u)$. The critical thresholds are based on the bootstrap approximation with 10^4 bootstrap samples.
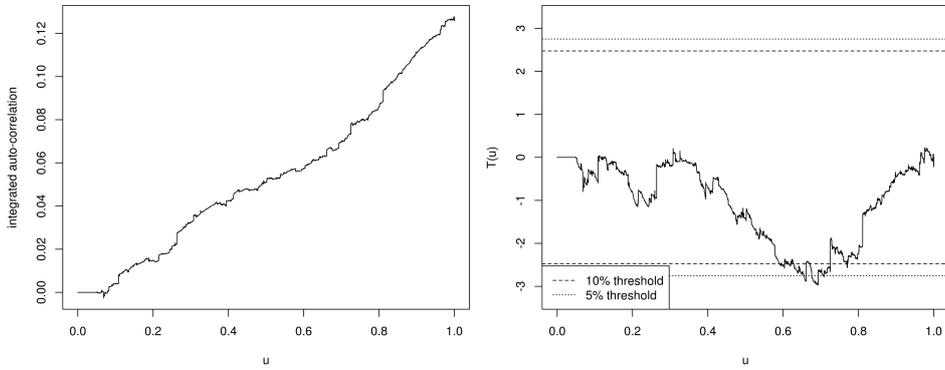
imation is satisfactory for this regime. Hence, we find that the variation of the lag-1 autocorrelation is not significant. That is, based on the bootstrapped CUSUM test, we may not reject the null hypothesis of constant lag-1 autocorrelations for this particular dataset at a significance level of 5%.

The visible discontinuities of the path of $M_n(u)$ in Figure 2, suggest that the estimator is influenced by few very large price changes d_t. While our bootstrap procedure automatically accounts for this, the resulting large variance decreases the power of the change point test. To reduce the effect of the heavy tails of the log returns, we repeat our analysis for the transformed increments $\tilde{d}_t = \arctan(d_t/\gamma)$. We choose $\gamma = 10^{-4}$, which corresponds to the average size of d_t and leads to a unimodal distribution of transformed returns (not depicted). The estimator $M_n(u)$ of the lag-1 autocorrelation of the transformed returns $\tilde{d}_t$ is depicted in Figure 3, as well as the corresponding CUSUM process. The p-value of the CUSUM test is 0.027, hence the hypothesis of constant lag-1 autocorrelation is rejected for the series $\tilde{d}_t$ at a significance level of 5%. Although

19

the autocorrelations of d_t and $\tilde{d}_t$ are not the same parameters, both may be interpreted similarly in the present application. In particular, our findings support the claim of Andersen et al. (2019) that the serial correlation of intraday log returns is non-constant for the present data set. In contrast to Andersen et al. (2019), our change point test does not assess whether the autocorrelation changes its sign.

We also analyze the variance, mean, and kurtosis of d_t and $\tilde{d}_t$ as outlined in Sections B.1 and 4.2. For d_t, the CUSUM tests for constant variance, mean, and kurtosis yield the p-values 0.081, 0.020, and 0.285, respectively. For $\tilde{d}_t$, the respective p-values are 0.000, 0.0484, and 0.000, respectively. In combination with our statistical results on the autocorrelation, this provides strong evidence for nonstationarity of $\tilde{d}_t$ and d_t.

Many models for asset returns at very high frequencies describe the observed price as the sum of two latent components: the fundamental price p_t^*, and the so-called microstructure noise η_t, such that $p_t = p_t^* + \eta_t$. The fundamental price is typically modeled as a semimartingale, and inference for this component needs to account for the microstructure effects, see e.g. Jacod et al. (2009). The microstructure noise is typically assumed to be independent, or dependent but stationary (Hansen and Lunde, 2006; Aït-Sahalia et al., 2011). Nonstationary dependent noise is considered by Jacod et al. (2017), but such that the autocorrelation of the microstructure is constant. The approach we pursue in the present paper does not distinguish between the fundamental price and the microstructure effects. Nevertheless, our empirical findings may motivate the investigation of microstructure models which allow for a nonstationary dependence structure.

A Further remarks

A.1 Alternative definition of local stationarity

The model described in Section 2 may also be compared to the definition of locally-stationary processes introduced by Dahlhaus et al. (2019). They require that for each $u \in [0, 1]$, there exists a stationary process $X_t(u)$ such that

(i) $\|X_{t,n} - X_t(\frac{t}{n})\|_{L_q} \le Cn^{-\alpha}$, and

(ii) $\|X_t(u) - X_t(v)\|_{L_q} \le C|u - v|^\alpha$, for some $C > 0$ and $\alpha \in (0, 1]$.

If the time series $X_{t,n}$ and $X_t(u)$, $u \in [0, 1]$, are $\boldsymbol{\epsilon}_t$-measurable, we may represent them as $X_t(u) = G(u, \boldsymbol{\epsilon}_t)$, and $X_{t,n} = G_n(\frac{t}{n}, \boldsymbol{\epsilon}_t)$, for kernels G and G_n which are measurable w.r.t. $\boldsymbol{\epsilon}_t$. Without loss of generality, we may suppose that $u \mapsto G_n(u, \boldsymbol{\epsilon})$ is a left-continuous, piecewise constant mapping with finitely many break points at t/n, for $t = 1, \dots, n$. Then

$$
\|G_n\|_{p-var} = \left(\sum_{t=2}^n \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}^p \right)^{\frac{1}{p}}
$$

$$
\le 2 \left(\sum_{t=1}^n \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G(\tfrac{t}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}^p \right)^{\frac{1}{p}} + \|G\|_{p-var}.
$$

Conditions (i) and (ii) imply that $\|G_n\|_{p-var}$ is bounded for any $p \ge \frac{1}{\alpha}$. Hence, provided that the time series is $\boldsymbol{\epsilon}_t$ measurable, our framework contains the model of Dahlhaus et al. (2019) as a special case. For instance, we may describe temporally varying auto-regressive processes, as described in Example 1.

A.2 Rates of convergence of the pilot estimator $\hat{\mu}_t^n$

Property (A.5) is a requirement on the rate of convergence of $\hat{\mu}_{t,n}$, in a form routinely studied in nonparametric statistics, see e.g. (van de Geer, 2010, Ch. 9). The achievable rate of convergence in nonparametric regression depends in particular on the smoothness of the function $u \mapsto \mu_u^n$. In our model, the only regularity assumption on μ_u^n is (A.2), which implies that $\|\mu^n\|_{p-var} < \infty$. Thus, empirical process theory suggests an achievable rate of convergence of $\mathcal{O}(n^{\frac{p}{2+p}})$ for $p \in [1, 2)$. This rate can be derived by combining general results for least squares regression with subgaussian errors (van de Geer, 2010, Thm. 9.1) with entropy bounds for the class of functions of bounded p-variation (Giné and Nickl, 2016, Cor. 3.7.50). Note that this rate is in line with the requirement (A.5) if $p \in [1, 2)$. Furthermore, we point out that this rate of convergence matches the rate under the assumption of Hölder continuity with exponent $\beta = \frac{1}{p}$. While the Hölder-continuous case may be treated via local smoothing, the generic approach to achieve this rate of convergence under the assumption of finite p-variation is via empirical risk

minimization, which in particular uses the whole sample. However, we additionally require that $\hat{\mu}_{t,n}$ is ϵ_t-measurable, rendering this approach infeasible. While it might be possible to construct a suitable online estimator which achieves the desired rate of convergence for functions of bounded p-variation, this is out of the scope of this article. See for example Baby and Wang (2019) and Raj et al. (2020) for the case $p = 1$ based on iid observations.

Another alternative is to formulate a parametric estimator of μ_u^n, which will typically satisfy (A.5) with the stronger rate $\sum_{t=1}^n \|\hat{\mu}_{t,n} - \mu_{t/n}^n\|^2 = \mathcal{O}_P(1)$. For example, this approach is feasible if μ_u^n is piecewise constant with a single breakpoint. However, in order to apply a parametric estimator, additional assumptions need to be imposed on the underlying time series. If these parametric assumptions do not hold, any inference based on the corresponding asymptotic results will be flawed. This encourages the use of nonparametric estimators for the local moment function μ_u^n.

B   More examples of change point problems

B.1   Changes in variance

For a univariate time series $X_{t,n}$, various researches have designed tests for constancy of the variances $\text{Var}(X_{t,n})$, starting with the investigation of asset returns by Hsu et al. (1974) and the corresponding methodology of Wichern et al. (1976). A test based on cumulative sums of squared observations has been suggested by Inclán and Tiao (1994), and Chen and Gupta (1997) suggest an alternative procedure based on the Schwarz information criterion. The approach of Inclán and Tiao (1994) has been generalized by Lee and Park (2001) to dependent processes, and by Aue et al. (2009) to the multivariate case. A common shortfall of many procedures is that they require constancy of the mean $\mathbb{E}(X_{t,n})$, which is a nuisance parameter in the present situation. More recent work studies the case of non-constant mean by using a suitable nonparametric estimator thereof, see Gao et al. (2019) and Schmidt et al. (2020). Note that the latter references still require stationarity of all remaining nuisance quantities, such as autocorrelations and higher order moments.

An alternative test for homoscedasticity can be formulated using our results. Based on the univariate time series $X_{t,n} = \tilde{G}_n(\frac{t}{n}, \epsilon_t)$, we define the vector valued time series $Y_{t,n} = G_n(\frac{t}{n}, \epsilon_t) = (X_{t,n}, X_{t,n}^2)$ with mean $\mu_u^n = \mathbb{E}G_n(u, \epsilon_0)$. Then $\text{Var}(X_{t,n}) = \mathbb{E}X_{t,n}^2 - (\mathbb{E}X_{t,n})^2 = f(\mu_{t/n}^n)$, for $f : \mathbb{R}^2 \to \mathbb{R}$, $f(a, b) = b - a^2$. It is straight-forward to check that if $\tilde{G}_n$ satisfies assumptions (A.1)-(A.3) for $q > 8$, then G_n satisfies the assumptions for $\frac{q}{2} > 4$. Moreover, f and all its derivatives are bounded on compacts, such that (A.4) holds if $\mu_u = \lim_{n\to\infty} \mu_u^n$ is bounded. Thus, the asymptotic result (8) holds and yields a test for the hypothesis $H_0 : \text{Var}(X_{t,n}) \equiv \text{const}$, allowing for non-stationary mean under very mild regularity conditions on dependency structure of $X_{t,n}$.

B.2 Changes in regression coefficients

Suppose we observe a univariate time series $Z_{t,n}$ and a d-dimensional vector time series $X_{t,n}$, such that $Z_{t,n} = \beta_{t,n}^T X_{t,n} + \eta_{t,n}$. The noise values $\eta_{t,n}$ satisfy $\mathbb{E}(\eta_{t,n}) = 0$ and $\mathrm{Cov}(X_{t,n}, \eta_{t,n}) = 0$. We want to test whether $\beta_{t,n}$ is constant, or varies with t. This problem has been studied, among others, by Horváth (1995), Horváth et al. (2004), and Aue et al. (2008) for uncorrelated innovations $\eta_{t,n}$. Correlated regression errors are treated, for example, by Robbins et al. (2016). In this setting, the hypothesis $H_0 : \beta_{t,n} \equiv \beta_{1,n}$ may be formulated in the form (7), since

$$\beta_{t,n} = \mathrm{Cov}(X_{t,n})^{-1} \mathrm{Cov}(X_{t,n}, Z_{t,n}),$$

where $\mathrm{Cov}(X_{t,n}) \in \mathbb{R}^{d \times d}$ and $\mathrm{Cov}(X_{t,n}, Z_{t,n}) \in \mathbb{R}^{d \times 1}$. Thus, $\beta_{t,n} = f(\mathbb{E} Y_{t,n})$, for

$$Y_{t,n} = (X_{t,n}, Z_{t,n}, X_{t,n} X_{t,n}^T, X_{t,n} Z_{t,n}) \in \mathbb{R}^{2d+1+d^2},$$

where the matrix $X_{t,n} X_{t,n}^T$ is interpreted as a d^2-dimensional vector. Assumptions (A.1)-(A.3) are a direct consequence of the corresponding properties of $X_{t,n}$ and $\eta_{t,n}$. Moreover, f satisfies (A.4) if we ensure that the smallest eigenvalue of $\mathrm{Cov}(X_{t,n})$ admits a uniform lower bound.

Note that $\beta_{t,n}$ is a d-dimensional parameter. It is straight-forward to extend the CUSUM statistic T_n^* to the multivariate setting, e.g. by replacing the absolute value by an arbitrary vector norm. Alternatively, one might test for changes in a single coordinate of $\beta_{t,n}$.

B.3 Further simulation results

To further assess the finite sample performance of our proposed procedure, we study the regression model

$$Z_{t,n} = \beta(\tfrac{t}{n})^T W_{t,n} + X_{t,n},$$

for regression coefficients $\beta(u) \in \mathbb{R}^2$, and noise process $X_{t,n}$ as in equation (9) of the article, using $\sigma(u)$ and $\alpha(u)$ as specified, and autoregression coefficient $a(u) = a_1(u)$. The cofactors $W_{t,n} \sim \mathcal{N}(0, \Sigma(\tfrac{t}{n}))$ are independent, bivariate normal random vectors, independent of $X_{t,n}$, with covariance matrix

$$\Sigma(u) = A(u)^T A(u), \quad \text{for} \quad A(u) = \begin{pmatrix} 1 & 2 + |\sin(2\pi u)| \\ 0 & 1 \end{pmatrix}^2.$$

Here, we want to test for structural stability of the first regression coefficient $\beta(u)_1$, and we apply the CUSUM test as outlined in Section B.2. For our simulations, we employ the model with regression coefficients $\beta(u) = \beta^0(u) \equiv (1, 2)^T$ as null hypothesis, and $\beta(u) = \beta^1(u) = (1 + u, 2 + u^2)^T$, $\beta^2(u) = (1 + \frac{u}{3}, 2 + \frac{u^2}{3})$ as alternatives, i.e. a gradual change.
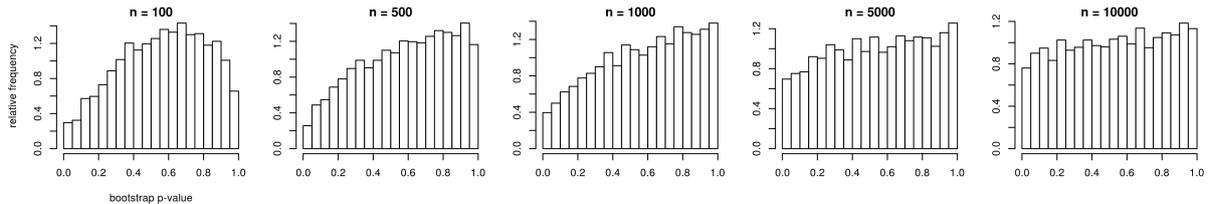
Figure 4: Distribution of bootstrap-based p-values for the CUSUM test for a constant regression coefficient, under the null hypothesis, with lag parameter $L_n = \frac{1}{10}\log(n)^2$. The p-values are computed based on $M = 10^3$ bootstrap samples, and the histograms are based on 5000 independent samples of the test statistic.

| | $L_n = \lceil \log(n)^2 \rceil$ | | | $L_n = \lceil \frac{1}{2}\log(n)^2 \rceil$ | | | $L_n = \lceil \frac{1}{5}\log(n)^2 \rceil$ | | | $L_n = \lceil \frac{1}{10}\log(n)^2 \rceil$ | | |
	H_0	H_1	H_2	H_0	H_1	H_2	H_0	H_1	H_2	H_0	H_1	H_2
$n = 100$	0.16	0.14	0.15	0.08	0.06	0.08	0.03	0.03	0.03	0.03	0.03	0.03
500	0.01	0.02	0.01	0.01	0.06	0.01	0.02	0.16	0.03	0.04	0.21	0.04
1000	0.01	0.10	0.01	0.02	0.28	0.04	0.04	0.41	0.07	0.04	0.45	0.09
5000	0.03	0.96	0.20	0.05	0.99	0.28	0.06	0.99	0.33	0.07	0.99	0.35
10000	0.05	1.00	0.46	0.06	1.00	0.53	0.07	1.00	0.58	0.08	1.00	0.58

Table 3: Size and power of the bootstrap-based CUSUM test for a constant regression coefficient, with nominal level 10%. Reported values are based on 5000 independent samples of the test statistic.

The distribution of the bootstrap-based p-values under null hypothesis is depicted in Figure 4. Values of the test's size and power are presented in Table 3, for different choices of the lag parameter L_n. Just as for the autocorrelation, we find that our proposed test for the regression coefficient is conservative, and that the size approximation is better for smaller L_n. As the sample size increases, the distribution of the p-values approaches the desired uniform distribution, and the size of the test tends towards the nominal level. This demonstrates the robustness of our test against non-stationary nuisance parameters. Furthermore, the CUSUM test consistently detects the nonstationarity of the regression coefficient as the sample size increases.

C   Technical proofs

C.1   Equivalence of the physical dependence measure

Proposition C.1. *Let $\epsilon_t, \epsilon_{t,j}^*, \tilde{\epsilon}_{t,j}$, be as in Section 2. Let $G : \mathbb{R}^\infty \to \mathbb{R}^d$ be a measurable function, where we endow $\mathbb{R}^\infty$ with the σ-Algebra generated by all finite projections. If $\|G(\epsilon_0)\|_{L_q} < \infty$ for some $q \geq 1$, then*

$$\left\| G(\epsilon_0) - G(\epsilon_{0,j}^*) \right\|_{L_q} \leq \sum_{s=0}^{\infty} \|G(\epsilon_0) - G(\tilde{\epsilon}_{0,j+s})\|_{L_q}.$$

24

Proof of Proposition C.1. Consider the martingale $G_t = \mathbb{E}(G(\boldsymbol{\epsilon}_0)|\epsilon_0, \ldots, \epsilon_{-t+1})$. Since $\|G(\boldsymbol{\epsilon}_0)\|_{L_q} < \infty$, the martingale convergence theorem guarantees that that $\|G_t - G(\boldsymbol{\epsilon}_0)\|_{L_q} \to 0$ as $t \to \infty$. Now note that $G_t - G(\boldsymbol{\epsilon}_0) \stackrel{d}{=} G_t - G(\boldsymbol{\epsilon}_{0,t}^*)$, such that we also obtain $\|G_t - G(\boldsymbol{\epsilon}_{0,t}^*)\|_{L_q} \to 0$ as $t \to \infty$ and thus $\|G(\boldsymbol{\epsilon}_{0,t}^*) - G(\boldsymbol{\epsilon}_0)\|_{L_q} \to 0$ as $t \to \infty$. Hence, for any $t \in \mathbb{N}$, and any $j \in \mathbb{N}$,

$$\|G(\boldsymbol{\epsilon}_0) - G(\boldsymbol{\epsilon}_{0,j}^*)\|_{L_q} \le \|G(\boldsymbol{\epsilon}_0) - G(\boldsymbol{\epsilon}_{0,j+t}^*)\|_{L_q} + \sum_{s=1}^{t} \|G(\boldsymbol{\epsilon}_{0,j+s}^*) - G(\boldsymbol{\epsilon}_{0,j+s-1}^*)\|_{L_q}.$$

But $\|G(\boldsymbol{\epsilon}_{0,j+s}^*) - G(\boldsymbol{\epsilon}_{0,j+s-1}^*)\|_{L_q} = \|G(\boldsymbol{\epsilon}_0) - G(\tilde{\boldsymbol{\epsilon}}_{0,j+s-1})\|_{L_q}$. Thus, letting $t \to \infty$, we find that

$$\|G(\boldsymbol{\epsilon}_0) - G(\boldsymbol{\epsilon}_{0,j}^*)\|_{L_q} \le \sum_{s=0}^{\infty} \|G(\boldsymbol{\epsilon}_0) - G(\tilde{\boldsymbol{\epsilon}}_{0,j+s})\|_{L_q}.$$

$\square$

Applying Proposition (C.1) under assumption (A.3), we find that

$$\|G(\boldsymbol{\epsilon}_t) - G(\boldsymbol{\epsilon}_{t,j}^*)\|_{L_q} \le \sum_{s=j}^{\infty} C_G \rho^j \le \frac{C_G}{1-\rho} \rho^j.$$

C.2 Asymptotics of partial sums

Proof of Theorem 2.1. We exploit the geometric decay of the dependence measure (A.3) to employ a coupling construction. To this end, we split the n observations into $m = m_n \ll n$ blocks of length $\tilde{L} = \tilde{L}_n = \lfloor n/m \rfloor$ and let $m \to \infty$. Furthermore, let $r = r_n \ll \tilde{L}_n$ with $r_n \to \infty$ be a smaller block length. It turns out that the rates $\tilde{L}_n = \lceil n^\epsilon \rceil, r_n = \lceil n^{\epsilon/2} \rceil$ are appropriate, for some $0 < \epsilon < \min(\frac{1}{2} - \frac{1}{q}, \frac{1}{p})$. We decompose

$$S_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nu \rfloor} X_{t,n} - \mathbb{E}X_{t,n}$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} [X_{t,n} - \mathbb{E}X_{t,n}] + \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r-1} [X_{t,n} - \mathbb{E}X_{t,n}]$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=\tilde{L}\lfloor um \rfloor+1}^{\lfloor nu \rfloor} [X_{t,n} - \mathbb{E}X_{t,n}]$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} Y_{j,n} + \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} \tilde{Y}_{j,n} + R_n(u). \tag{10}$$

Consider the remainder term $R_n(u)$. By assumptions (A.1) and (A.2), we have $\|X_{t,n}\|_{L_q} \le 2C_G$ for all n, t. The union bound and Markov's inequality yield, for any $a > 0$,

$$P\left(\sup_{u \in [0,1]} |R_n(u)| > a\right) \le \sum_{k=\tilde{L}+1}^{n} P\left(\frac{1}{\sqrt{n}} \sum_{t=k-\tilde{L}}^{k} |X_{t,n} - \mathbb{E}X_{t,n}| > a\right)$$

$$\le \frac{n}{\sqrt{n}^q} \frac{(2C_G\tilde{L})^q}{a^q}.$$

Since $q > 2$, this term tends to zero for our choice of $\tilde{L}$, since $\tilde{L} = \tilde{L}_n = n^\epsilon$ for $\epsilon < \frac{1}{2} - \frac{1}{q}$.

In (10), the random variables $Y_{j,n}$, $j = 1, \ldots, m$, may be replaced by independent copies $Y_{j,n}^*$ as follows. For each $j = 1, \ldots, m$, let $\tilde{\epsilon}_t^j$ be an independent copy of the $\epsilon_t, t \in \mathbb{Z}$. Then

$$Y_{j,n} = \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} \left[G_n(\tfrac{t}{n}, \epsilon_t, \epsilon_{t-1}, \ldots) - \mathbb{E}X_{t,n}\right]$$

$$= \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} \left[G_n(\tfrac{t}{n}, \epsilon_t, \ldots, \epsilon_{(j-1)\tilde{L}+1}, \tilde{\epsilon}_{(j-1)\tilde{L}}^j, \tilde{\epsilon}_{(j-1)\tilde{L}-1}^j \ldots) - \mathbb{E}X_{t,n}\right] + \delta_{j,n}$$

$$= Y_{j,n}^* + \delta_{j,n},$$

for a random variable $\delta_{j,n}$ with $\|\delta_{j,n}\|_{L_q} \le C_G\tilde{L}\sum_{k=r}^{\infty} \rho^k \le C\tilde{L}\rho^r$, by virtue of (A.3). Note that the $Y_{j,n}^*$ are independent and satisfy $Y_{j,n}^* \sim Y_{j,n}$, for $j = 1, \ldots, m$. Analogously, we may construct independent copies $\tilde{Y}_{j,n}^*$ of $\tilde{Y}_{j,n}$, $j = 1, \ldots, m$, such that $\|\tilde{Y}_{j,n} - \tilde{Y}_{j,n}^*\|_{L_q} = \|\tilde{\delta}_{j,n}\|_{L_q} \le C\tilde{L}\rho^r$. Hence,

$$S_n(u) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} Y_{j,n}^* + \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} \tilde{Y}_{j,n}^* + \tilde{R}_n(u) + R_n(u).$$

The remainder term $\tilde{R}_n(u)$ satisfies

$$\|\sup_{u \in [0,1]} \tilde{R}_n(u)\|_{L_q} \le \frac{1}{\sqrt{n}} \sum_{j=1}^{m} \left(\|\delta_{j,n}\|_{L_q} + \|\tilde{\delta}_{j,n}\|_{L_q}\right) \le C\frac{\tilde{L}m}{\sqrt{n}}\rho^r,$$

which is asymptotically negligible since $r = r_n = n^{\epsilon/2}$ and $\epsilon > 0$, and $\tilde{L}_n m_n \le n$.

We now establish a central limit theorem for the term $S_n^*(u) = \frac{1}{\sqrt{n}}\sum_{j=1}^{\lfloor um \rfloor} Y_{j,n}^*$. This result will also apply to the term $\frac{1}{\sqrt{n}}\sum_{j=1}^{m} \tilde{Y}_{j,n}^*$ with a different rate of convergence, and we will be able to conclude that the latter is negligible. To this end, consider

$$\mathrm{Var}(Y_{j,n}^*) = \sum_{t,s=(j-1)\tilde{L}+r}^{j\tilde{L}} \mathrm{Cov}\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t), G_n(\tfrac{s}{n}, \boldsymbol{\epsilon}_s)\right]$$

$$= \sum_{t,s=(j-1)\tilde{L}+r}^{j\tilde{L}} \text{Cov}\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_s)\right] + g_n^{s,t},$$

for some $|g_n^{s,t}| \leq C\|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_s) - G_n(\tfrac{s}{n}, \boldsymbol{\epsilon}_s)\|_{L_2}$, using the $L_2(P)$ boundedness of G. Furthermore, condition (A.3) implies that $|\text{Cov}\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_s)\right]| \leq C\rho^{|t-s|}$. Hence, it can be checked that

$$\sum_{s,t=(j-1)\tilde{L}+r}^{j\tilde{L}} \text{Cov}\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_s)\right]$$

$$= \bar{\delta}_{j,n} + \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} \sum_{s=-\infty}^{\infty} \text{Cov}\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_t), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_s)\right]$$

$$= \mathcal{O}(1) + \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} \sigma_n^2(\tfrac{t}{n})$$

for some $|\bar{\delta}_{j,n}| \leq C$, and $\sigma_n(u) = \sum_{h=-\infty}^{\infty} \text{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0))$. Thus, for any $u \in [0,1]$,

$$\sum_{j=1}^{\lfloor um \rfloor} \text{Var}\left(\frac{Y_{j,n}^*}{\sqrt{n}}\right) = \frac{1}{n} \sum_{j=1}^{\lfloor um \rfloor} \bar{\delta}_{j,n} + \frac{1}{n} \sum_{j=1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} \sigma_n^2(\tfrac{t}{n}) + \frac{1}{n} \sum_{j=1}^{\lfloor um \rfloor} \sum_{t,s=(j-1)\tilde{L}+1}^{j\tilde{L}} g_{s,t}$$

$$= V_n^1(u) + V_n^2(u) + V_n^3(u). \tag{11}$$

The term $V_n^1(u)$ tends to zero uniformly in u because the $\bar{\delta}_{j,n}$ are bounded, and $m_n \ll n$. The third term may be bounded as

$$\sup_{u \in [0,1]} |V_n^3(u)| \leq \frac{1}{n} \sum_{j=1}^{m} \sum_{t,s=(j-1)\tilde{L}+1}^{j\tilde{L}} C\|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{s}{n}, \boldsymbol{\epsilon}_0)\|_{L_2}$$

$$\leq \frac{C}{n} \sum_{t=1}^{n} \sup_{s=1,\dots,n, |s-t| \leq \tilde{L}} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{s}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}$$

$$\leq \frac{\tilde{L}C}{n} \sum_{t=1}^{n} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}$$

$$\leq \frac{\tilde{L}C}{n} n^{1-\frac{1}{p}} \left(\sum_{t=1}^{n} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}^p\right)^{\frac{1}{p}}$$

$$\leq \tilde{L}_n C n^{-\frac{1}{p}} \|G_n\|_{p-var},$$

which tends to zero since $\|G_n\|_{p-var} \leq C_G$ by assumption, and $\tilde{L}_n n^{-\frac{1}{p}} \to 0$.

To treat the term $V_n^2(u)$, we show that $\|\sigma^2\|_{p-var} < \infty$. First, we find that for $u, v \in [0,1]$ and any $h \in \mathbb{Z}$,

$$|\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0)) - \mathrm{Cov}(G_n(v, \boldsymbol{\epsilon}_h), G_n(v, \boldsymbol{\epsilon}_0))|$$
$$\leq |\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0) - G_n(v, \boldsymbol{\epsilon}_0))| + |\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h) - G_n(v, \boldsymbol{\epsilon}_h), G_n(v, \boldsymbol{\epsilon}_0))|$$
$$= |\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0) - G_n(v, \boldsymbol{\epsilon}_0))| + |\mathrm{Cov}(G_n(v, \boldsymbol{\epsilon}_{-h}), G_n(u, \boldsymbol{\epsilon}_0) - G_n(v, \boldsymbol{\epsilon}_0))|.$$

Both summands may be treated identically using (A.3). In particular, assuming w.l.o.g. $h \geq 0$,

$$|\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0) - G_n(v, \boldsymbol{\epsilon}_0))|$$
$$= \left|\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_{h,h}^*), G_n(u, \boldsymbol{\epsilon}_0) - G_n(v, \boldsymbol{\epsilon}_0))\right|$$
$$+ \left|\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h) - G_n(u, \boldsymbol{\epsilon}_{h,h}^*), G_n(u, \boldsymbol{\epsilon}_0) - G_n(v, \boldsymbol{\epsilon}_0))\right|$$
$$\leq 0 + C\rho^{|h|} \|G_n(u) - G_n(v)\|_{L_2}.$$

Hence,

$$|\sigma_n^2(u) - \sigma_n^2(v)| \leq \sum_{h=-\infty}^{\infty} |\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0)) - \mathrm{Cov}(G_n(v, \boldsymbol{\epsilon}_h), G_n(v, \boldsymbol{\epsilon}_0))|$$
$$\leq \sum_{h=-\infty}^{\infty} C\rho^{|h|} \|G_n(u) - G_n(v)\|_{L_2}$$
$$\leq C\|G_n(u) - G_n(v)\|_{L_2}, \tag{12}$$

which implies that $\|\sigma_n^2\|_{p-var} \leq C\|G_n\|_{p-var} \leq C \cdot C_G$. Moreover, we have that $\sigma_n^2(u) \to \sigma^2(u)$, because the same argument as above yields

$$|\sigma_n^2(u) - \sigma^2(u)| \leq \sum_{h=-\infty}^{\infty} C\rho^{|h|} \|G_n(u, \boldsymbol{\epsilon}_0) - G(u, \boldsymbol{\epsilon}_0)\|_{L_2} \quad \to 0. \tag{13}$$

Hence, we find that $u \mapsto \sigma_n^2(u)$ is bounded and of bounded p-variation, uniformly in n. The same holds for $\sigma^2(u)$. This suffices to establish convergence of the Riemann sum $V_n^2(u)$ in (11), since

$$\sup_{u \in [0,1]} \left| V_n^2(u) - \int_0^u \sigma_n^2(v)\,dv \right|$$
$$= \sup_{u \in [0,1]} \left| \frac{1}{n} \sum_{j=1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+r}^{j\tilde{L}} \sigma_n^2(\tfrac{t}{n}) - \int_0^u \sigma_n^2(v)\,dv \right|$$
$$= \sup_{u \in [0,1]} \left| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} \sigma_n^2(\tfrac{t}{n}) - \int_0^{\frac{\lfloor un \rfloor}{n}} \sigma_n^2(v)\,dv \right| + \mathcal{O}\left(\frac{mr}{n}\right) + \mathcal{O}\left(\frac{\tilde{L}}{n}\right)$$

$$
= \sup_{u \in [0,1]} \left| \sum_{t=1}^{\lfloor un \rfloor} \int_{\frac{t-1}{n}}^{\frac{t}{n}} [\sigma_n^2(\tfrac{t}{n}) - \sigma_n^2(v)] \, dv \right| + \mathcal{O}\left(\frac{mr}{n}\right) + \mathcal{O}\left(\frac{\tilde{L}}{n}\right) + \mathcal{O}\left(\frac{1}{n}\right)
$$

$$
\leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\frac{t-1}{n} \leq v < w \leq \frac{t}{n}} |\sigma_n^2(v) - \sigma_n^2(w)| + \mathcal{O}\left(\frac{mr}{n}\right) + \mathcal{O}\left(\frac{\tilde{L}}{n}\right)
$$

$$
\leq n^{-\frac{1}{p}} \|\sigma_n^2\|_{p-var} + \mathcal{O}\left(\frac{mr}{n}\right) + \mathcal{O}\left(\frac{\tilde{L}}{n}\right).
$$

By our choice of $\tilde{L}$ and r, in particular $r/\tilde{L} \asymp \frac{mr}{n} \to 0$, the latter term tends to zero. Moreover, $\int_0^u \sigma_n^2(v) \, dv \to \int_0^u \sigma^2(v) \, dv$ uniformly in u by the dominated convergence theorem. We have thus shown that

$$
\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} Y_{j,n}^*\right) = \int_0^u \sigma^2(v) \, dv + o(1),
$$

uniformly in u. On the other hand, following the same steps, we can show that

$$
\mathrm{Var}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} \tilde{Y}_{j,n}\right) = o(1) + \frac{1}{n} \sum_{j=1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r-1} \sigma_n^2(\tfrac{t}{n}) = o(1) + \mathcal{O}\left(\frac{mr}{n}\right),
$$

which tends to zero by our choice of m, r. Doob's inequality thus yields that

$$
\sup_{u \in [0,1]} \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} \tilde{Y}_{j,n} \to 0,
$$

in probability, as $n \to \infty$. Hence, $S_n(u) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} Y_{j,n}^* + o(1)$, uniformly in u.

To establish a central limit theorem, we verify Lyapunov's condition. By our assumptions on G_n, in particular convergence and bounded p-variation in $L_q(P)$, we know that $\|G_n(u)\|_{L_q} \leq C$ for some constant C, uniformly in n and u. Hence,

$$
\sum_{j=1}^{m} \left\| Y_{j,n}^* / \sqrt{n} \right\|_{L_q}^q \leq \sum_{j=1}^{m} \frac{(C\tilde{L})^q}{n^{\frac{q}{2}}} = \mathcal{O}(\tilde{L}^{q-1}/n^{q/2-1}) = \mathcal{O}(\tilde{L}n^{\frac{2-q}{2q-2}})^{\frac{1}{q-1}}, \tag{14}
$$

since $m\tilde{L}/n \to 1$. This term tends to zero since $\tilde{L}_n n^{\frac{2-q}{2q-2}} \leq \tilde{L}_n n^{\frac{1}{2}-\frac{1}{q}} \to 0$ by our choice of $\tilde{L}_n$. Thus, the functional central limit theorem is applicable, see e.g. (Jacod and Shiryaev, 2003, Thm. VIII.3.33). We obtain

$$
\frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor um \rfloor} Y_{j,n}^* \Rightarrow B_{\int_0^u \sigma^2(v) \, dv}.
$$

$\square$

C.3 Properties of the local smoother $\hat{\mu}_{t,n}^{NW}$

Proposition 3.1 is a consequence of the following two Lemmas.

Lemma C.2. *Let (A.3) hold for some $q > 2$. Let $\tau_n \to \infty$ and choose $k = k_n$ such that $k_n \gg n^{\frac{2}{q}}$. Then, as $n \to \infty$,*

$$\sup_{t=\tau_n,\dots,n} \left\| \hat{\mu}_{t,n}^{NW} - \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} \mu_{\frac{i}{n}}^n \right\| \xrightarrow{P} 0.$$

Proof of Lemma C.2. The union bound and Markov's inequality yield, for any $a > 0$,

$$P\left(\sup_{t=\tau_n,\dots,n} \left\| \hat{\mu}_{t,n}^{NW} - \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} \mu_{\frac{i}{n}}^n \right\| > a \right)$$

$$= P\left(\sup_{t=\tau_n,\dots,n} \left\| \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} [X_{i,n} - \mathbb{E}X_{i,n}] \right\| > a \right)$$

$$\leq a^{-q} \sum_{t=\tau_n}^{n} \left\| \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} [X_{i,n} - \mathbb{E}X_{i,n}] \right\|_{L_q}^q. \tag{15}$$

To bound the latter L_q norm, we apply Theorem 1 of Liu et al. (2013). The latter result is formulated for stationary time series of the type $X_t = G(\boldsymbol{\epsilon}_t)$. However, the stationarity is only strictly required for the last equation of the proof therein, where we may replace $\|X_1\|_{L_q}$ by $s_q = \sup_n \sup_{u \in [0,1]} \|G_n(u, \boldsymbol{\epsilon}_0)\|_{L_q}$, which is bounded by virtue of (A.3) for $j = 0$. At all other occasions in the proof of Liu et al. (2013), only the quantity $\theta_{j,q} = \|G(\boldsymbol{\epsilon}_0) - G(\tilde{\boldsymbol{\epsilon}}_{0,j})\|_{L_q}$ is relevant, which may be replaced in our nonstationary setting by

$$\theta_{j,q}^* = \sup_n \sup_{u \in [0,1]} \|G_n(u, \boldsymbol{\epsilon}_0) - G_n(u, \tilde{\boldsymbol{\epsilon}}_{0,j})\|_{L_q}, \qquad q \geq 2.$$

By assumption (A.3), we have $\theta_{j,q}^* \leq C\rho^j$. Hence, for our case, the bound of Liu et al. (2013) reads as follows: For all k and all $1 \leq s < t \leq n$ such that $t - s = k$,

$$\left\| \sum_{i=s}^{t} [X_{i,n} - \mathbb{E}X_{i,n}] \right\|_{L_q}$$

$$\leq k^{1/2} \left[\frac{87\,q}{\log q} \sum_{j=1}^{k} \theta_{j,2}^* + 3(q-1)^{1/2} \sum_{j=k+1}^{\infty} \theta_{j,q}^* + \frac{29\,q}{\log q} s_2 \right]$$

$$+ k^{1/q} \left[\frac{87\,q(q-1)^{1/2}}{\log q} \sum_{j=1}^{k} j^{1/2-1/q} \theta_{j,q}^* + \frac{29\,q}{\log q} s_q \right]$$

$$\leq k^{1/2} \left(s_q + \sum_{j=1}^{\infty} \theta_{j,q}^* \right) \left(\frac{87\,q}{\log q} + 3(q-1)^{1/2} + \frac{58\,q}{\log q} + \frac{87\,q(q-1)^{1/2}}{\log q} \right)$$

$$\leq C k^{1/2},$$

for a factor C which does not depend on k nor n. For the latter two inequalities, we used that $\theta_{j,2}^* \leq \theta_{j,q}^*$ and $s_2 \leq s_q$ for $q \geq 2$, and the geometric decay of $\theta_{j,q}^*$. Thus,

$$\sum_{t=\tau_n}^{n} \left\| \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} [X_{i,n} - \mathbb{E}X_{i,n}] \right\|_{L_q}^{q} \leq C \sum_{t=\tau_n}^{n} (k \wedge t)^{-q/2}$$

$$\leq C \left(nk^{-q/2} + \tau_n^{1-q/2} \right).$$

This term tends to zero if $q > 2$. We hence obtain the uniform convergence via the union bound (15). $\square$

Lemma C.3. *Let (A.3) hold for some $q \geq 2$. Suppose that $u \mapsto \mu_u^n$ is β-Hölder continuous with Hölder constant C independent of n. Then*

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \hat{\mu}_{t,n}^{NW} - \mu_{\frac{t}{n}}^n \right\|^2 = \mathcal{O}_P \left(\left(\frac{k}{n} \right)^{2\beta} + \frac{\log n}{k} \right).$$

If, instead of Hölder continuity, we assume that $\|\mu^n\|_{p-var} < C < \infty$ for some $p \in [1,2)$, then

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \hat{\mu}_{t,n}^{NW} - \mu_{\frac{t}{n}}^n \right\|^2 = \mathcal{O}_P \left(\frac{k}{n} + \frac{\log n}{k} \right).$$

Furthermore, if $\mu_u^n = \mu_u^{n,1} + \mu_u^{n,2}$ such that $u \mapsto \mu_u^{n,1}$ is β-Hölder continuous with Hölder constant C independent of n, and $\|\mu^{n,2}\|_{p-var} \leq a_n$, then

$$\frac{1}{n} \sum_{t=1}^{n} \left\| \hat{\mu}_{t,n}^{NW} - \mu_{\frac{t}{n}}^n \right\|^2 = \mathcal{O}_P \left(\frac{a_n k}{n} + \left(\frac{k}{n} \right)^{2\beta} + \frac{\log n}{k} \right).$$

Proof of Lemma C.3. Assume without loss of generality that the $X_{i,n}$ are scalar, i.e. $d = 1$. Otherwise, we may treat each component individually. We have

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \|\hat{\mu}_{t,n}^{NW} - \mu_{\frac{t}{n}}^n\|^2 \leq \frac{2}{n} \sum_{t=1}^{n} \mathbb{E} \left\| \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} [X_{i,n} - \mathbb{E}X_{i,n}] \right\|^2$$

$$+ \frac{2}{n} \sum_{t=1}^{n} \left\| \frac{1}{k \wedge t} \sum_{i=(t-k)\vee 1}^{t} \left[\mu_{\frac{i}{n}}^n - \mu_{\frac{t}{n}}^n \right] \right\|^2. \tag{16}$$

The Hölder continuity of μ_u^n yields

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{1}{k\wedge t}\sum_{i=(t-k)\vee 1}^{t}\left[\mu_{\frac{i}{n}}^n-\mu_{\frac{t}{n}}^n\right]\right\|^2 \leq C\left(\frac{k}{n}\right)^{2\beta}.$$

Alternatively, if we only assume $\|\mu^n\|_{p-var} < \infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\frac{1}{k\wedge t}\sum_{i=(t-k)\vee 1}^{t}\left[\mu_{\frac{i}{n}}^n-\mu_{\frac{t}{n}}^n\right]\right\|^2 \leq \frac{1}{n}\sum_{t=1}^{n}\sup_{s\in\left[\frac{t-k}{n}\vee 0,t\right]}\|\mu_{t/n}^n-\mu_s^n\|^2$$

$$\leq \frac{1}{n}(2\|\mu\|_\infty)^{2-p}\sum_{t=1}^{n}\sup_{s\in\left[\frac{t-k}{n}\vee 0,t\right]}\|\mu_{t/n}^n-\mu_s^n\|^p$$

$$\leq \frac{k}{n}(2\|\mu^n\|_\infty)^{2-p}\|\mu^n\|_{p-var}^p$$

Note that $\|\mu^n\|_\infty \leq |\mu_0^n| + \|\mu^n\|_{p-var}$, which is bounded. The third case is obtained by a combination of the two previous bounds.

Regarding the first term in (16), note that (A.3) implies $|\mathrm{Cov}(G_n(u,\boldsymbol{\epsilon}_i), G_n(v,\boldsymbol{\epsilon}_j))| \leq C\rho^{|i-j|}$, for any $u,v\in[0,1]$ and any n. Hence,

$$\mathbb{E}\left\|\frac{1}{k\wedge t}\sum_{i=(t-k)\vee 1}^{t}[X_{i,n}-\mathbb{E}X_{i,n}]\right\|^2 = \mathrm{Var}\left(\frac{1}{k\wedge t}\sum_{i=(t-k)\vee 1}^{t}G_n(\tfrac{i}{n},\boldsymbol{\epsilon}_i)\right)$$

$$= \frac{1}{(k\wedge t)^2}\sum_{i,j=(t-k)\vee 1}^{t}\mathrm{Cov}\left(G_n(\tfrac{i}{n},\boldsymbol{\epsilon}_i), G_n(\tfrac{j}{n},\boldsymbol{\epsilon}_j)\right)$$

$$\leq \frac{1}{k\wedge t}\sum_{j=1}^{\infty}C\rho^j.$$

Therefore,

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\|\frac{1}{k\wedge t}\sum_{i=(t-k)\vee 1}^{t}[X_{i,n}-\mathbb{E}X_{i,n}]\right\|^2$$

$$\leq \frac{C}{n}\sum_{t=1}^{n}\frac{1}{k\wedge t}$$

$$\leq \frac{C}{k}+\frac{C}{n}\sum_{t=1}^{k}\frac{1}{t}$$

$$\leq \frac{C}{k}+\frac{C\log k}{n} \quad \leq \quad \frac{2C\log n}{k},$$

completing the proof. $\qquad\qquad\square$

Of course, Lemma C.2 and Lemma C.3 also hold with μ_u^n replaced by its limit μ_u if we use the samples $X_t = G(\frac{t}{n}, \boldsymbol{\epsilon}_t)$ instead of $X_{t,n} = G_n(\frac{t}{n}, \boldsymbol{\epsilon}_t)$.

Proof of Proposition 3.1. Clearly, Lemma C.3 implies that (A.5) holds. Regarding (A.6), note that $\sup_{u \in [0,1]} |\mu_u^n - \mu_u| \to 0$ by virtue of (A.1). Thus, for any $\delta > 0$, we have that $\frac{1}{k \wedge t} \sum_{i=(t-k) \vee 1}^{t} \mu_{\frac{i}{n}}^n \in \mathcal{M}_{\delta/2}$ for n sufficiently large. Together with Lemma C.2, this implies (A.6). $\qquad\square$

C.4 Functional central limit theorem for $M_n(u)$

Recall the decomposition $M_n(u) = I_n^1(u) + I_n^2(u) + I_n^3(u)$, where

$$I_n^1(u) = \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} f\left(\mu_{\frac{t}{n}}^n\right),$$

$$I_n^2(u) = \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} \left(\mu_{\frac{t}{n}}^n - \hat{\mu}_{t-L,n}\right)^T \frac{D^2 f\left(\tilde{\mu}_{t,n}\right)}{2} \left(\mu_{\frac{t}{n}}^n - \hat{\mu}_{t-L,n}\right),$$

$$I_n^3(u) = \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} Df(\hat{\mu}_{t-L,n}) \left(X_{t,n} - \mu_{\frac{t}{n}}^n\right).$$

The first term $I_n^1(u)$ is a discretized version of the integral $\int_0^u f(\mu_v^n) \, dv$, and its convergence depends on the regularity of the function $u \mapsto \mu_u^n$. Assumption (A.2) implies $\|\mu^n\|_{p-var} < C_G$, which may be used to treat the term $I_n^1(u)$ as follows.

Lemma C.4. *For any function $u \mapsto g_u \in \mathbb{R}^d$, it holds that*

$$\sup_{u \in [0,1]} \left\| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} g_{\frac{t}{n}} - \int_0^{\frac{\lfloor un \rfloor}{n}} g_v \, dv \right\| \leq n^{-\frac{1}{p}} \|g\|_{p-var}.$$

Furthermore, if (A.1) and (A.2) hold for $p \geq 1$, then there exists $N > 0$ such that for all $n \geq N$,

$$\sup_{u \in [0,1]} \left| I_n^1(u) - \int_{\frac{\tau_n+L_n}{n}}^{u} f(\mu_v^n) \, dv \right| \leq C_f \|\mu^n\|_{p-var} n^{-\frac{1}{p}}, \tag{17}$$

$$\sup_{u \in [0,1]} \left| I_n^1(u) - \int_0^{u} f(\mu_v^n) \, dv \right| \leq C_f \|\mu^n\|_{p-var} n^{-\frac{1}{p}} + C_f \frac{\tau_n + L_n}{n}. \tag{18}$$

Proof of Lemma C.4. It holds that

$$\left\| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} g_{\frac{t}{n}} - \int_0^{\frac{\lfloor un \rfloor}{n}} g_v \, dv \right\| \leq \int_0^1 \left\| g_v - g_{\frac{\lceil vn \rceil}{n}} \right\| dv \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{v \in [\frac{t-1}{n}, \frac{t}{n}]} \left\| g_v - g_{\frac{t}{n}} \right\|$$

$$\leq \frac{n^{1-\frac{1}{p}}}{n} \left(\sum_{t=1}^{n} \sup_{v \in [\frac{t-1}{n}, \frac{t}{n}]} \left\| g_v - g_{\frac{t}{n}} \right\|^p \right)^{\frac{1}{p}}$$

$$\leq n^{-\frac{1}{p}} \|g\|_{p-var}.$$

Regarding the term $I_n^1(u)$, note that $\mu_u^n \in \mathcal{M}_\delta$ for all $u \in [0,1]$, and n large enough. Then the mean value theorem and the boundedness of Df on $\mathcal{M}_\delta$ yield $|f(\mu_u^n) - f(\mu_v^n)| \leq C_f \|\mu_u^n - \mu_v^n\|$ for any $u, v \in [0,1]$, such that $\|f(\mu^n)\|_{p-var} \leq C_f \|\mu^n\|_{p-var}$. Thus, for n large enough and for any u such that $\lfloor nu \rfloor > \tau_n + L_n$,

$$\left| I_n^1(u) - \int_{\frac{\tau_n + L_n}{n}}^{\frac{\lfloor nu \rfloor}{n}} f(\mu_v^n)\, dv \right| \leq C_f n^{-\frac{1}{p}} \|\mu^n\|_{p-var}.$$

Moreover, since $f(\mu_v^n)$ is bounded,

$$\left| \int_0^{\frac{\tau_n + L_n}{n}} f(\mu_v^n)\, dv \right| \leq C_f \frac{\tau_n + L_n}{n}, \qquad \left| \int_{\frac{\lfloor nu \rfloor}{n}}^{u} f(\mu_v^n)\, dv \right| \leq \frac{C_f}{n} \to 0,$$

completing the proof. $\qquad\qquad\square$

In particular, for $(\tau_n + L_n) \ll n^{\frac{1}{2}}$ and $p < 2$, we find that

$$\sup_{u \in [0,1]} \left| I_n^1(u) - \int_0^u f(\mu_v^n)\, dv \right| = o(1/\sqrt{n}).$$

The term $I_n^2(u)$ may be treated directly by our assumptions on the nonparametric estimator $\hat{\mu}_{t,n}$, and using additionally the regularity of $u \mapsto \mu_u^n$.

Lemma C.5. *Let* (A.1), (A.2), (A.4), (A.5), *and* (A.6) *hold, for some $p \geq 1$. Then*

$$\sup_{u \in [0,1]} |I_n^2(u)| = o_P(1/\sqrt{n}) + \mathcal{O}_P(L_n^{\min(p,2)} n^{-\frac{2}{\max(p,2)}}).$$

In particular, $\sup_{u \in [0,1]} |I_n^2(u)| = o_P(1/\sqrt{n})$ *if* $L_n \ll n^{\frac{1}{p} - \frac{1}{4}}$ *for $p \geq 2$, and $L_n \ll n^{\frac{1}{2p}}$ for $p \in [1,2)$.*

Proof of Lemma C.5. Assumption (A.1) implies that for n sufficiently large, $\mu_u^n \in \mathcal{M}_\delta$ for all $u \in [0,1]$. Moreover, $\hat{\mu}_{t-L,n} \in \mathcal{M}_\delta$ for all $t \geq \tau_n + L_n$, with probability tending to one. Since $\mathcal{M}_\delta$ is convex, we find that $\tilde{\mu}_{t,n} \in \mathcal{M}_\delta$ for all $t \geq \tau_n + L_n$, with probability tending to one. On this event, it holds that

$$|I_n^2(u)| \leq \frac{2C_f}{n} \sum_{t=\tau_n + L_n}^{\lfloor nu \rfloor} \|\hat{\mu}_{t-L,n} - \mu_{\frac{t-L}{n}}^n\|^2 + \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^2$$

$$\leq o_P(1/\sqrt{n}) + \frac{2C_f C_G}{n} \sum_{t=\tau_n + L_n}^{\lfloor nu \rfloor} \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^2.$$

34

Moreover, if $p < 2$, boundedness of $\|\mu_t^n\|$ yields

$$
\begin{aligned}
\frac{1}{n} \sum_{t=\tau_n+L_n}^{n} \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^2 &\leq \frac{C}{n} \sum_{t=\tau_n+L_n}^{n} \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^p \\
&\leq \frac{C}{n} \sum_{t=\tau_n+L_n}^{n} \left(\sum_{i=1}^{L_n} \|\mu_{\frac{t-i+1}{n}}^n - \mu_{\frac{t-i}{n}}^n\| \right)^p \\
&\leq \frac{C}{n} \sum_{t=\tau_n+L_n}^{n} L_n^{p-1} \sum_{i=1}^{L_n} \|\mu_{\frac{t-i+1}{n}}^n - \mu_{\frac{t-i}{n}}^n\|^p \\
&\leq \frac{L_n^p}{n} \sum_{t=1}^{n} \|\mu_{\frac{t}{n}}^n - \mu_{\frac{t-1}{n}}^n\|^p \qquad = L_n^p n^{-1} \|\mu^n\|_{p-var}^p.
\end{aligned}
$$

If $p > 2$, we may apply Hölder's inequality to obtain

$$
\begin{aligned}
\frac{1}{n} \sum_{t=\tau_n+L_n}^{n} \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^2 &\leq \frac{1}{n} n^{1-\frac{2}{p}} \left(\sum_{t=\tau_n+L_n}^{n} \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^p \right)^{\frac{2}{p}} \\
&= \left(\frac{1}{n} \sum_{t=\tau_n+L_n}^{n} \|\mu_{\frac{t-L}{n}}^n - \mu_{\frac{t}{n}}^n\|^p \right)^{\frac{2}{p}} \qquad \leq L_n^2 n^{-\frac{2}{p}} \|\mu^n\|_{p-var}^2.
\end{aligned}
$$

This yields the claimed upper bound. $\qquad\square$

The last term $I_n^3(u)$ admits a functional central limit theorem with scaling factor $\sqrt{n}$, as shown in the following Lemma. Its proof follows the same ideas as the proof of Theorem 2.1, but additional care is needed when treating the random but consistent factor $Df(\hat{\mu}_{t,n}) \approx Df(\mu_{t/n}^n)$.

Lemma C.6. *Suppose that (A.1), (A.2), (A.3), (A.4), (A.5), and (A.6) hold for some $p \geq 1$. Suppose that L_n satisfies $L_n \gg \log(n)^{1+a}$ for some $a \in (0,1)$, and that $L_n \ll n^{\frac{1}{p}}$. Then, as $n \to \infty$,*

$$
\sqrt{n} I_n^3(u) \Rightarrow B\left(\int_0^u Df(\mu_v)\Sigma(v)Df(\mu_v)^T \, dv \right),
$$

$$
\Sigma(v) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(G(v, \boldsymbol{\epsilon}_0), G(v, \boldsymbol{\epsilon}_h) \right) \in \mathbb{R}^{d \times d},
$$

where B denotes a standard Brownian motion. The weak convergence holds in the Skorokhod space $D[0,1]$.

Proof of Lemma C.6. As in the proof of Theorem 2.1, we split the sum into blocks of size $\tilde{L} = \tilde{L}_n = \lfloor n^\epsilon \wedge \frac{L_n}{2} \rfloor$, and $r = r_n = \lfloor \tilde{L}_n^{1-\epsilon} \rfloor$, for some arbitrarily small $\epsilon \in (0,1)$ to be chosen in the sequel. We assume n to be sufficiently large such that $1 \leq r_n < \tilde{L}_n \leq \frac{L_n}{2}$. Denote by $m = m_n = \lfloor n/\tilde{L}_n \rfloor$ the number of blocks.

Step (i): As a first step, we match the upper and lower limits of the sum $I_n^3(u)$ with the block size $L = L_n$. Define

$$\bar{I}_n^3(u) = \frac{1}{n} \sum_{t=\lceil \frac{\tau_n + L_n}{\tilde{L}} \rceil \tilde{L}+1}^{\lfloor um \rfloor \tilde{L}} Df(\hat{\mu}_{t-L,n}) \left(X_{t,n} - \mu_{\frac{t}{n}}^n \right).$$

By virtue of Lemma C.2 and (A.4), $\|Df(\hat{\mu}_{t,n})\| \leq C_f$ for all t, with probability tending to one, such that

$$\left| I_n^3(u) - \bar{I}_n^3(u) \right| = \mathcal{O}_P \left(\frac{2\tilde{L}_n C_f}{n} \sup_{t=1,\ldots,n} |X_{t,n} - \mathbb{E}X_{t,n}| \right).$$

Since $\|X_{t,n}\|_{L_q} \leq C < \infty$, the union bound and Markov's inequality yield the upper bound $\sup_{t=1,\ldots,n} |X_{t,n} - \mathbb{E}X_{t,n}| = \mathcal{O}_P(n^{\frac{1}{q}})$. Using $\tilde{L}_n \leq n^\epsilon$, we obtain

$$\sup_{u \in [0,1]} \left| I_n^3(u) - \bar{I}_n^3(u) \right| = \mathcal{O}_P \left(n^{\epsilon + \frac{1}{q} - 1} \right).$$

Since $q > 2$, we may choose $\epsilon > 0$ sufficiently small such that the latter term is of order $o_P(1/\sqrt{n})$.

Step (ii): Next, we introduce $\chi_{t,n} = \mathbb{1}(\hat{\mu}_{t,n} \in \mathcal{M}_\delta)$, and

$$Y_{j,n} = \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n})(X_{t,n} - \mathbb{E}X_{t,n}),$$

$$\tilde{Y}_{j,n} = \sum_{t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n})(X_{t,n} - \mathbb{E}X_{t,n}),$$

$$\tilde{I}_n^3(u) = \frac{1}{n} \sum_{j=\lceil \frac{\tau_n + L_n}{\tilde{L}} \rceil +1}^{\lfloor um \rfloor} (Y_{j,n} + \tilde{Y}_{j,n}).$$

By virtue of assumption (A.6), we have $\sup_{u \in [0,1]} |\bar{I}_n^3(u) - \tilde{I}_n^3(u)| \to 0$ in probability. Now, for each j, let $\epsilon_i^{(j)}, i \in \mathbb{Z}$ be an independent copy of the ϵ_i, and define

$$Y_{j,n}^* = \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n}) \left[G_n \left(\frac{t}{n}, \epsilon_t, \ldots, \epsilon_{(j-1)\tilde{L}+1}, \epsilon_{(j-1)\tilde{L}}^{(j)}, \ldots \right) - \mathbb{E}X_{t,n} \right]$$

$$= \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n}) Z_{t,n}^j,$$

$$\tilde{Y}_{j,n}^* = \sum_{t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n}) \left[G_n \left(\frac{t}{n}, \epsilon_t, \ldots, \epsilon_{(j-1)\tilde{L}+1-r}, \epsilon_{(j-1)\tilde{L}-r}^{(j)}, \ldots \right) - \mathbb{E}X_{t,n} \right]$$

$$= \sum_{t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n}) \tilde{Z}_{t,n}^j.$$

This construction yields that the $Y_{j,n}^*$ and the $\tilde{Y}_{j,n}^*$ are martingale differences, since $\hat{\mu}_{t-L,n}$ is measurable w.r.t. $\boldsymbol{\epsilon}_{t-L}$, and since $\tilde{L}_n + r_n < L_n$. It also holds that $\tilde{Z}_{t,n}^j \sim Z_{t,n}^j \sim (X_{t,n} - \mathbb{E}X_{t,n})$. Moreover, by virtue of (A.3), it holds that $\|Y_{j,n}^* - Y_{j,n}\|_{L_q} \le C\,\tilde{L}\rho^r$ and $\|\tilde{Y}_{j,n}^* - \tilde{Y}_{j,n}\|_{L_q} \le C\,r\rho^r \le C\,\tilde{L}_n\rho^r$. Now define the martingales

$$J_n^1(u) = \frac{1}{n} \sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um \rfloor} Y_{j,n}^*,$$

$$J_n^2(u) = \frac{1}{n} \sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um \rfloor} \tilde{Y}_{j,n}^*,$$

and note that

$$\left\| \sup_{u\in[0,1]} \left| \tilde{I}_n^3(u) - J_n^1(u) - J_n^2(u) \right| \right\|_{L_q} \le \frac{C}{n} \sum_{j=1}^{m_n} \tilde{L}_n \rho^{r_n} \le C\rho^{r_n} = o(\sqrt{n}),$$

since $\tilde{L}_n m_n \le n$, and $r_n \gg \log(n)^{(1+a)(1-\epsilon)} \gg \log(n)^{1+\frac{a}{2}}$, upon choosing ϵ sufficiently small.

Step (iii): Define the two sequences of filtrations $\mathcal{F}_{k,n}^1 = \sigma(\epsilon_i, \epsilon_i^{(j)} : j \le k, i \le k\tilde{L}_n)$, and $\mathcal{F}_{k,n}^2 = \sigma(\epsilon_i, \epsilon_i^{(j)} : j \le k, i \le k\tilde{L}_n - r_n)$. Then $J_n^1(k/m)$ is an $\mathcal{F}_{k,n}^1$ martingale, and $J_n^2(k/m)$ is a $\mathcal{F}_{k,n}^2$ martingale. The predictable quadratic variation of $\sqrt{n}J_n^1$ is given by

$$\langle \sqrt{n}J_n^1(u) \rangle = \frac{1}{n} \sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um \rfloor} \mathbb{E}\left[\left(Y_{j,n}^*\right)^2 | \mathcal{F}_{j-1,n}^1 \right]$$

$$= \frac{1}{n} \sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um \rfloor} \sum_{s,t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \chi_{t-L,n}\chi_{s-L,n} Df(\hat{\mu}_{t-L,n}) \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) Df(\hat{\mu}_{s-L,n})^T.$$

We observe that

$$\left| Df(\hat{\mu}_{t-L,n}) \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) Df(\hat{\mu}_{s-L,n})^T - Df(\mu_{\frac{t}{n}}^n) \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) Df(\mu_{\frac{s}{n}}^n)^T \right| \chi_{t-L,n}\chi_{s-L,n}$$

$$\le \left| Df(\hat{\mu}_{t-L,n}) \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) \left[Df(\hat{\mu}_{s-L,n}) - Df(\mu_{\frac{s}{n}}^n) \right]^T \right| \chi_{t-L,n}\chi_{s-L,n}$$

$$+ \left| \left[Df(\hat{\mu}_{t-L,n}) - Df(\mu_{\frac{t}{n}}^n) \right] \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) Df(\hat{\mu}_{s-L,n})^T \right| \chi_{t-L,n}\chi_{s-L,n}$$

$$\le C\rho^{|t-s|} \left[\left\| Df(\hat{\mu}_{t-L,n}) - Df(\mu_{\frac{t}{n}}^n) \right\| + \left\| Df(\hat{\mu}_{s-L,n}) - Df(\mu_{\frac{s}{n}}^n) \right\| \right] \chi_{t-L,n}\chi_{s-L,n}$$

$$\leq C\rho^{|t-s|}\left[\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|+\left\|\hat{\mu}_{s-L,n}-\mu^n_{\frac{s}{n}}\right\|\right],$$

where the last inequalities are a consequence of (A.3), and the boundedness of Df and D^2f on $\mathcal{M}_\delta$. Moreover,

$$\frac{1}{n}\sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um\rfloor}\sum_{s,t=(j-1)\tilde{L}+r+1}^{j\tilde{L}}\rho^{|t-s|}\left[\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|+\left\|\hat{\mu}_{s-L,n}-\mu^n_{\frac{s}{n}}\right\|\right]$$

$$=\frac{2}{n}\sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um\rfloor}\sum_{s,t=(j-1)\tilde{L}+r+1}^{j\tilde{L}}\rho^{|t-s|}\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|$$

$$\leq\frac{C}{n}\sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um\rfloor}\sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}}\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|$$

$$\leq\frac{C}{n}\sum_{t=\tau_n}^{n}\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|$$

$$\leq\frac{C}{\sqrt{n}}\sqrt{\sum_{t=\tau_n}^{n}\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|^2}. \tag{19}$$

In the proof of Lemma C.5, we have shown that $\frac{1}{n}\sum_{t=\tau_n}^{n}\left\|\hat{\mu}_{t-L,n}-\mu^n_{\frac{t}{n}}\right\|^2=o_p(1/\sqrt{n})+\mathcal{O}_P(L_n^{\min(p,2)}n^{-\frac{2}{\max(p,2)}})$, which tends to zero because $L_n\ll n^{\frac{1}{p}}$ by assumption. Hence, (19) tends to zero as $n\to\infty$, and we thus obtain

$$\langle\sqrt{n}J^1_n(u)\rangle=o_P(1)+\sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um\rfloor}\sum_{s,t=(j-1)\tilde{L}+r+1}^{j\tilde{L}}Df(\mu^n_{\frac{t}{n}})\operatorname{Cov}(Z^j_{t,n},Z^j_{s,n})Df(\mu^n_{\frac{s}{n}})^T \tag{20}$$

where the $o_P(1)$ term vanishes uniformly in $u\in[0,1]$. Note that we may omit the indicators $\chi_{t-L,n}$ asymptotically by virtue of assumption (A.6). Analogously, we obtain

$$\langle\sqrt{n}J^2_n(u)\rangle=o_P(1)+\frac{1}{n}\sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um\rfloor}\sum_{s,t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r}Df(\mu^n_{\frac{t}{n}})\operatorname{Cov}(Z^j_{t,n},Z^j_{s,n})Df(\mu^n_{\frac{s}{n}})^T.$$

Step (iv): We further simplify the asymptotic expression for $\langle\sqrt{n}J^1_n(u)\rangle$. In (20), we want to replace $\mu^n_{\frac{s}{n}}$ by $\mu^n_{\frac{t}{n}}$. Observe that

$$\left|\frac{1}{n}\sum_{j=\left\lceil\frac{\tau_n+L_n}{\tilde{L}}\right\rceil+1}^{\lfloor um\rfloor}\sum_{s,t=(j-1)\tilde{L}+r+1}^{j\tilde{L}}Df(\mu^n_{\frac{t}{n}})\operatorname{Cov}(Z^j_{t,n},Z^j_{s,n})\left[Df(\mu^n_{\frac{s}{n}})-Df(\mu^n_{\frac{t}{n}})\right]^T\right|$$

$$\leq \frac{C}{n} \sum_{t=1}^{n} \sum_{s=t-\tilde{L}}^{t+\tilde{L}} \rho^{|s-t|} \left\| Df(\mu_{\frac{s}{n}}^{n}) - Df(\mu_{\frac{t}{n}}^{n}) \right\|$$

$$\leq \frac{C}{n} \sum_{t=1}^{n} \max_{|s-t| \leq \tilde{L}} \left\| \mu_{\frac{s}{n}}^{n} - \mu_{\frac{t}{n}}^{n} \right\|$$

$$\leq \frac{C\tilde{L}_n}{n} \sum_{t=2}^{n} \left\| \mu_{\frac{t}{n}}^{n} - \mu_{\frac{t-1}{n}}^{n} \right\|$$

$$\leq C\tilde{L}_n n^{-\frac{1}{p}} \left(\sum_{t=2}^{n} \left\| \mu_{\frac{t}{n}}^{n} - \mu_{\frac{t-1}{n}}^{n} \right\|^{p} \right)^{\frac{1}{p}} \quad \leq C\tilde{L}_n n^{-\frac{1}{p}} \| \mu^{n} \|_{p-var}.$$

By (A.2), and since $\tilde{L}_n \ll n^{\epsilon}$ the latter term is of order $\mathcal{O}(n^{\epsilon - \frac{1}{p}}) = o(1)$ if ϵ is small enough. Hence,

$$\langle \sqrt{n} J_n^1(u) \rangle$$

$$= o_P(1) + \frac{1}{n} \sum_{j=\left\lceil \frac{\tau_n + L_n}{\tilde{L}} \right\rceil + 1}^{\lfloor um \rfloor} \sum_{s,t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} Df(\mu_{\frac{t}{n}}^{n}) \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) Df(\mu_{\frac{t}{n}}^{n})^T$$

$$= o_P(1) + \frac{1}{n} \sum_{j=\left\lceil \frac{\tau_n + L_n}{\tilde{L}} \right\rceil + 1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} Df(\mu_{\frac{t}{n}}^{n}) \left[\sum_{s=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) \right] Df(\mu_{\frac{t}{n}}^{n})^T.$$

$$(21)$$

The covariance term may be simplified as follows. Define

$$\Sigma_n(u) = \sum_{h=-\infty}^{\infty} \operatorname{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0)) \quad \in \mathbb{R}^{d \times d}, \qquad u \in [0,1].$$

Note that (A.3) implies that $\| \operatorname{Cov}(G_n(u, \boldsymbol{\epsilon}_h), G_n(u, \boldsymbol{\epsilon}_0)) \| \leq C \rho^{|h|}$ for some constant C, where $\| \cdot \|$ denotes an arbitrary matrix norm. In particular, $\Sigma_n(u)$ is well defined since the series is finite.

Now, for any $t \geq (\tau_n + L_n)$, let $\delta_{t,n} = \sup_{|s-t| \leq \tilde{L}} \| G_n(\frac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\frac{s}{n}, \boldsymbol{\epsilon}_0) \|_{L_q}$. Since $q > 2$, we find that for $(j-1)\tilde{L} + 1 \leq t \leq j\tilde{L}$,

$$\left\| \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \left[\Sigma_n(\tfrac{t}{n}) - \sum_{s=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \operatorname{Cov}(Z_{t,n}^j, Z_{s,n}^j) \right] \right\|$$

$$= \left\| \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \left[\Sigma_n(\tfrac{t}{n}) - \sum_{s=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \operatorname{Cov}(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t-s}), G_n(\tfrac{s}{n}, \boldsymbol{\epsilon}_0)) \right] \right\|$$

$$\leq \left\| \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \left[\Sigma_n(\tfrac{t}{n}) - \sum_{s=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \operatorname{Cov}(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t-s}), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0)) \right] \right\| + \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} C\tilde{L}\delta_{t,n}$$

$$
= \left\| \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \sum_{\substack{[s\le (j-1)\tilde{L}+r] \\ \vee\, [s>j\tilde{L}]}} \mathrm{Cov}(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t-s}), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0)) \right\| + C \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \tilde{L}\delta_{t,n}
$$

$$
\le \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \sum_{\substack{[s\le (j-1)\tilde{L}+r] \\ \vee\, [s>j\tilde{L}]}} C\rho^{|t-s|} + C \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \tilde{L}\delta_{t,n}
$$

$$
\le C \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} [\rho^{|t-(j-1)\tilde{L}|} + \rho^{|t-j\tilde{L}|}] + C \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \tilde{L}\delta_{t,n}
$$

$$
\le C + C \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \tilde{L}\delta_{t,n}.
$$

Therefore,

$$
\left\| \langle \sqrt{n} J_n^1(u) \rangle - \frac{1}{n} \sum_{j=\lceil \frac{\tau_n+L_n}{\tilde{L}} \rceil + 1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} Df(\mu_{\frac{t}{n}}^n)\Sigma_n(\tfrac{t}{n})Df(\mu_{\frac{t}{n}}^n)^T \right\|
$$

$$
\le o_P(1) + C\frac{\tilde{L}_n}{n} \sum_{t=\tau_n+L_n}^{n} \delta_{t,n} + C\frac{m_n}{n}
$$

$$
\le o_P(1) + C\frac{\tilde{L}_n^2}{n} \sum_{t=\tau_n+L_n}^{n} \|\mu_{\frac{t}{n}}^n - \mu_{\frac{t-1}{n}}^n\| + C\frac{m_n}{n}
$$

$$
\le o_P(1) + C\tilde{L}_n^2 n^{-\frac{1}{p}} \|\mu^n\|_{p-var} + C\frac{m_n}{n}.
$$

The latter term tends to zero upon choosing ϵ small enough, since $\tilde{L}_n = \mathcal{O}(n^\epsilon)$. Note also that $m_n/n \to 0$. Finally, since $r_n \ll \tilde{L}_n$, we have shown that

$$
\langle \sqrt{n} J_n^1(u) \rangle = o_P(1) + \frac{1}{n} \sum_{j=\lceil \frac{\tau_n+L_n}{\tilde{L}} \rceil + 1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} Df(\mu_{\frac{t}{n}}^n)\Sigma_n(\tfrac{t}{n})Df(\mu_{\frac{t}{n}}^n)^T
$$

$$
= o_P(1) + \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} Df(\mu_{\frac{t}{n}}^n)\Sigma_n(\tfrac{t}{n})Df(\mu_{\frac{t}{n}}^n)^T, \tag{22}
$$

where the last step holds because $\tau_n + L_n \ll n$ and due to the boundedness of $Df(\mu_u^n)$ and $\Sigma_n(u)$, which is in turn implied by (A.3) and (A.2) for $q > 2$. Note also that the $o_P(1)$ term in (22) vanishes uniformly in $u \in [0,1]$.

Analogously, we may show that

$$
\langle \sqrt{n} J_n^2(u) \rangle = o_P(1) + \frac{1}{n} \sum_{j=\lceil \frac{\tau_n+L_n}{\tilde{L}} \rceil + 1}^{\lfloor um \rfloor} \sum_{t=(j-1)\tilde{L}+1}^{(j-1)\tilde{L}+r} Df(\mu_{\frac{t}{n}}^n)\Sigma_n(\tfrac{t}{n})Df(\mu_{\frac{t}{n}}^n)^T \longrightarrow 0,
$$

40

using again the boundedness, and the fact that $r_n \ll \tilde{L}_n$ such that the number of summands is asymptotically negligible. This implies that $\sup_{u \in [0,1]} |J_n^2(u)| \xrightarrow{P} 0$ by the Burkholder-Davis-Gundy inequality for martingales.

Observe that the function $u \mapsto \Sigma_n(u)$ has bounded p-variation, see (12). Moreover, $u \mapsto Df(\mu_u^n)$ has bounded p-variation by assumptions (A.2) and (A.4). Since both $\|\Sigma_n(u)\|$ and $\|Df(\mu_u^n)\|$ are bounded, we find that $u \mapsto Df(\mu_u^n)^T \Sigma_n(u) Df(\mu_u^n)$ has bounded p-variation. Thus, as in Lemma C.4,

$$\left| \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} Df(\mu_{\frac{t}{n}}^n) \Sigma_n(\tfrac{t}{n}) Df(\mu_{\frac{t}{n}}^n)^T - \int_0^u Df(\mu_v^n) \Sigma_n(v) Df(\mu_v^n)^T \, dv \right| = \mathcal{O}(n^{-\frac{1}{p}}).$$

Assumption (A.1) implies that $\mu_u^n \to \mu_u$ and $\Sigma_n(u) \to \Sigma(u)$ uniformly in $u \in [0,1]$. The latter uniform convergence holds because $\mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_0), G_n(u, \boldsymbol{\epsilon}_h)) \to \mathrm{Cov}(G(u, \boldsymbol{\epsilon}_0), G(u, \boldsymbol{\epsilon}_h))$ uniformly in u, and $\sup_{u \in [0,1]} \| \mathrm{Cov}(G_n(u, \boldsymbol{\epsilon}_0), G_n(u, \boldsymbol{\epsilon}_h)) \| \leq C\rho^{|h|}$ by virtue of (A.3). Hence,

$$\sup_{u \in [0,1]} \left| \langle \sqrt{n} J_n^1(u) \rangle - \int_0^u Df(\mu_v) \Sigma(v) Df(\mu_v)^T \, dv \right| = o_P(1).$$

Step (v): Having at hand the limit of $\langle \sqrt{n} J_n^1(u) \rangle$, a functional central limit theorem for $\sqrt{n} J_n^1(u)$ holds if we can verify Lyapunov's condition. We have, for any $2 < \delta < q$,

$$\sum_{j=\lceil \frac{\tau_n + L_n}{\tilde{L}} \rceil + 1}^{m} \mathbb{E}\left[\|Y_{j,n}^*/\sqrt{n}\|^\delta \, | \mathcal{F}_{j-1,n}^1 \right]$$

$$= n^{-\frac{\delta}{2}} \sum_{j=\lceil \frac{\tau_n + L_n}{\tilde{L}} \rceil + 1}^{m} \mathbb{E}\left[\left\| \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \chi_{t-L,n} Df(\hat{\mu}_{t-L,n}) Z_{t,n}^j \right\|^\delta \, \Bigg| \mathcal{F}_{j-1,n}^1 \right]$$

$$\leq C_f^\delta n^{-\frac{\delta}{2}} \sum_{j=\lceil \frac{\tau_n + L_n}{\tilde{L}} \rceil + 1}^{m} \mathbb{E}\left[\left(\sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \|Z_{t,n}^j\| \right)^\delta \, \Bigg| \mathcal{F}_{j-1,n}^1 \right]$$

$$\leq C_f^\delta n^{-\frac{\delta}{2}} \tilde{L}_n^{\delta-1} \sum_{j=\lceil \frac{\tau_n + L_n}{\tilde{L}} \rceil + 1}^{m} \sum_{t=(j-1)\tilde{L}+r+1}^{j\tilde{L}} \mathbb{E}\|Z_{t,n}^j\|^\delta$$

$$\leq C_f^\delta n^{-\frac{\delta}{2}} \tilde{L}_n^{\delta-1} \sum_{t=1}^{n} \mathbb{E}\|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0)\|^\delta$$

$$= \mathcal{O}\left(n^{1-\frac{\delta}{2}} \tilde{L}_n^{\delta-1} \right),$$

where we used that $\|G_n(u, \boldsymbol{\epsilon}_0)\|_{L_q}$ is uniformly bounded. By choosing ϵ sufficiently small, and using $\tilde{L}_n \leq n^\epsilon$, the latter term tends to zero. Hence, the functional central limit

theorem for martingales, e.g. (Jacod and Shiryaev, 2003, Thm. VIII.3.33), yields

$$\sqrt{n} J_n^1(u) \Rightarrow B\left(\int_0^u Df(\mu_s)\Sigma(s)Df(\mu_s)^T \, ds\right).$$

In the previous steps, we have established that $\sup_{u\in[0,1]} |I_n^3(u) - J_n^1(u)| = o_P(\sqrt{n})$, which yields the desired result. $\qquad\square$

Theorem 3.2 is now a direct consequence of Lemmas C.4, C.5, and C.6.

C.5 Bootstrap inference

Theorem 3.3 is a special case of the following more general theorem about consistent estimation of the limiting variance. In contrast to the central limit theorem for $M_n(u)$, here, we only require the weaker assumption

$$\sum_{t=\tau_n}^n \|\hat{\mu}_{t,n} - \mu_{\frac{t}{n}}^n\|^2 = o(n^{1-\kappa}), \tag{C.5*}$$

for some $\kappa \in (0, \frac{1}{2}]$.

Theorem C.7. *Let the conditions of Theorem 3.2 hold for some $q > 4$, replacing (A.5) by (C.5*). Choose some $b_n \to \infty$ such that*

$$b_n \ll n^{\frac{2}{3}\frac{1}{\max(p,2)}}, \quad b_n \ll n^{\frac{\kappa q-2}{q+2}}.$$

Then, as $n \to \infty$,

$$Q_n(u) = \frac{1}{n}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor - b_n} \frac{1}{b_n}\left[Df(\hat{\mu}_{t-L,n})\sum_{i=1}^{b_n}(X_{t+i,n} - \hat{\mu}_{t-L,n})\right]^2$$

$$\xrightarrow{P} Q(u) = \int_0^u Df(\mu_v)\Sigma(v)Df(\mu_v)^T \, dv.$$

The convergence holds uniformly in $u \in [0,1]$ since Q_n is monotone.

Proof of Theorem C.7. Step (i): As a first step, we are concerned with the estimators $\hat{\mu}_{t-L,n}$. We decompose

$$\sum_{i=1}^{b_n} Df(\hat{\mu}_{t-L,n})(X_{t+i,n} - \hat{\mu}_{t-L,n})$$

$$= \sum_{i=1}^{b_n}\left\{Df(\mu_{\frac{t}{n}}^n)\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n\right] + \left[Df(\hat{\mu}_{t-L,n}) - Df(\mu_{\frac{t}{n}}^n)\right]\left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{\frac{t}{n}}^n\right]\right.$$

$$\left. + Df(\hat{\mu}_{t-L,n})\left[G_n(\tfrac{t+i}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i})\right] + Df(\hat{\mu}_{t-L,n})\left[\mu_{\frac{t}{n}}^n - \hat{\mu}_{t-L,n}\right]\right\}.$$

With probability tending to one, it holds that $\hat{\mu}_{t-L,n} \in \mathcal{M}_\delta$ for all $t = \tau_n + L_n, \ldots, n$ by virtue of (A.6). On this event, we may exploit the regularity and boundedness of f such that

$$\left| \sum_{i=1}^{b_n} Df(\hat{\mu}_{t-L,n})(X_{t+i,n} - \hat{\mu}_{t-L,n}) - \sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right|$$

$$\leq C \sum_{i=1}^{b_n} Z_{t,i,n} \| \hat{\mu}_{t-L,n} - \mu_{\frac{t}{n}}^n \| + \| G_n(\tfrac{t+i}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) \| \quad =: \Delta_{t,n},$$

for $Z_{t,i,n} = \| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{\frac{t}{n}}^n \| + 1$. Now, via the inequality $|a^2 - b^2| \leq |a - b||a + b|$, we obtain

$$\left| \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \frac{1}{b_n} \left[\sum_{i=1}^{b_n} Df(\hat{\mu}_{t-L,n})(X_{t+i,n} - \hat{\mu}_{t-L,n}) \right]^2 \right.$$

$$\left. - \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \frac{1}{b_n} \left[\sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right]^2 \right|$$

$$\leq \frac{1}{n} \frac{1}{b_n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \Delta_{t,n} \left[\Delta_{t,n} + \left| \sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right| \right]$$

$$\leq \frac{1}{n} \frac{1}{b_n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \Delta_{t,n}^2 + \sqrt{\frac{1}{n} \frac{1}{b_n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \Delta_{t,n}^2} \sqrt{\frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \frac{1}{b_n} \left[\sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right]^2}.$$

It thus suffices to show that

$$\frac{1}{n} \frac{1}{b_n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \Delta_{t,n}^2 \xrightarrow{P} 0, \tag{23}$$

and

$$\frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \frac{1}{b_n} \left[\sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right]^2 \xrightarrow{P} \int_0^u Df(\mu_v)^T \Sigma(v) Df(\mu_v) \, dv. \tag{24}$$

Step (ii): Regarding the term (23), we find that

$$\frac{1}{n} \frac{1}{b_n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \Delta_{t,n}^2$$

$$\leq \frac{C}{n} \frac{1}{b_n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \left\{ \left[\sum_{i=1}^{b_n} Z_{t,i,n} \| \hat{\mu}_{t-L,n} - \mu_{\frac{t}{n}}^n \| \right]^2 + \left[\sum_{i=1}^{b_n} \| G_n(\tfrac{t+i}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) \| \right]^2 \right\}$$

$$\leq \frac{C}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \left\{ \|\hat{\mu}_{t-L,n} - \mu_{\frac{t}{n}}^n\|^2 \sum_{i=1}^{b_n} |Z_{t,i,n}|^2 + \sum_{i=1}^{b_n} \|G_n(\tfrac{t+i}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i})\|^2 \right\}$$

$$\leq \frac{C}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \left\{ \|\hat{\mu}_{t-L,n} - \mu_{\frac{t}{n}}^n\|^2 b_n \max_{i=1,\dots,b_n} |Z_{t,i,n}|^2 + \sum_{i=1}^{b_n} \left[\sum_{j=1}^{i} \|G_n(\tfrac{t+j}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t+j-1}{n}, \boldsymbol{\epsilon}_{t+i})\| \right]^2 \right\}$$

$$\leq \frac{C b_n}{n} \max_{\substack{t=\tau_n+L_n,\dots,n \\ i=1,\dots,b_n}} |Z_{t,i,n}|^2 \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \|\hat{\mu}_{t-L,n} - \mu_{\frac{t}{n}}^n\|^2$$

$$+ \frac{C b_n}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \sum_{i,j=1}^{b_n} \|G_n(\tfrac{t+j}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t+j-1}{n}, \boldsymbol{\epsilon}_{t+i})\|^2. \tag{25}$$

The second term may be bounded by noting that, for $p \leq 2$,

$$\frac{b_n}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \sum_{i,j=1}^{b_n} \mathbb{E}\|G_n(\tfrac{t+j}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t+j-1}{n}, \boldsymbol{\epsilon}_{t+i})\|^2$$

$$\leq \frac{b_n^3}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_2}^2$$

$$\leq \frac{b_n^3}{n} \sum_{t=1}^{n} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}^p \qquad \leq \frac{b_n^3}{n} \|G_n\|_{p-var}^p.$$

In the second inequality, we used that $q > 2$, $p < 2$, and that $\|G_n(u, \boldsymbol{\epsilon}_0)\|_{L_q}$ is uniformly bounded due to (A.2). If $p > 2$, we find that

$$\sum_{t=1}^{n} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}^2 \leq n^{1-\frac{2}{p}} \left(\sum_{t=1}^{n} \|G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0) - G_n(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_0)\|_{L_q}^p \right)^{\frac{2}{p}}$$

$$\leq n^{1-\frac{2}{p}} \|G_n\|_{p-var}^2,$$

such that the second term in (25) is of order $\mathcal{O}(b_n^3 n^{-\frac{2}{\max(p,2)}})$.

Concerning the first term in (25), we note that

$$\max_{t=\tau_n+L_n,\dots,n} |Z_{t,i,n}| \leq |Z_{1,i,n}| + \sum_{t=2}^{n} |Z_{t,i,n} - Z_{t-1,i,n}|,$$

$$\left\| \max_{t=\tau_n+L_n,\dots,n} |Z_{t,i,n}| \right\|_{L_q} \leq 1 + 2\|G_n(\tfrac{1}{n}, \boldsymbol{\epsilon}_{t+i})\|_{L_q} + 2\sum_{t=2}^{n} \|G(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - G(\tfrac{t-1}{n}, \boldsymbol{\epsilon}_{t+i})\|_{L_q}$$

$$\leq 1 + 2C_G + n^{1-\frac{1}{p}} \|G_n\|_{p-var}$$

$$\leq 2n^{1-\frac{1}{p}} C_G + 1.$$

44

Since $q > 2$, this yields

$$\left\|\max_{t=\tau_n+L_n,\ldots,n}|Z_{t,i,n}|^2\right\|_{L_{\frac{q}{2}}} = \left\|\max_{t=\tau_n+L_n,\ldots,n}|Z_{t,i,n}|\right\|_{L_q}^2 \leq C\,n^{2-\frac{2}{p}},$$

for some finite C. An alternative upper bound, which is sharper for $p > 2$, is given by

$$\left\|\max_{t=\tau_n+L_n,\ldots,n}|Z_{t,i,n}|^2\right\|_{L_{\frac{q}{2}}} \leq \sum_{t=1}^{n}\|Z_{t,i,n}\|_{L_q}^2 \leq Cn,$$

such that

$$\left\|\max_{t=\tau_n+L_n,\ldots,n}|Z_{t,i,n}|^2\right\|_{L_{\frac{q}{2}}} \leq Cn^{2-\frac{2}{\min(p,2)}}.$$

Thus, the union bound and Markov's inequality yield that

$$\max_{i=1,\ldots,b_n}\max_{t=1,\ldots,n}|Z_{t,i,n}|^2 = \mathcal{O}_P(|n^{2-\frac{2}{\min(p,2)}}b_n|^{\frac{2}{q}}).$$

Furthermore, as in the proof of Lemma C.5,

$$\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor-b_n}\|\hat\mu_{t-L,n}-\mu_{\frac{t}{n}}^n\|^2 \leq o_P(n^{1-\kappa}) + \mathcal{O}_P(L_n^{\min(p,2)}n^{1-\frac{2}{\max(2,p)}}).$$

Plugging this into (25), we have thus shown that

$$\frac{1}{n}\frac{1}{b_n}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor-b_n}\Delta_{t,n}^2 = \frac{b_n}{n}\mathcal{O}_P(|n^{2-\frac{2}{\min(p,2)}}b_n|^{\frac{2}{q}})\left[o_P(n^{1-\kappa}) + \mathcal{O}_P(L_n^{\min(p,2)}n^{1-\frac{2}{\max(2,p)}})\right] + \mathcal{O}_P\left(b_n^3 n^{-\frac{2}{\max(p,2)}}\right)$$

For $p \leq 2$, this bound reduces to

$$\frac{1}{n}\frac{1}{b_n}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor-b_n}\Delta_{t,n}^2 = \mathcal{O}_P\left(n^{\frac{4}{q}(1-\frac{1}{p})-\kappa}b_n^{1+\frac{2}{q}}\right) + \mathcal{O}_P\left(n^{\frac{4}{q}(1-\frac{1}{p})-\kappa}b_n^{1+\frac{2}{q}}n^{\kappa-1}L_n^p\right) + \mathcal{O}_P\left(b_n^3 n^{-1}\right)$$

$$\leq \mathcal{O}_P\left(n^{\frac{2}{q}-\kappa}b_n^{1+\frac{2}{q}}\right) + \mathcal{O}_P\left(n^{\frac{2}{q}-\kappa}b_n^{1+\frac{2}{q}}n^{\kappa-1}L_n^p\right) + \mathcal{O}_P\left(b_n^3 n^{-1}\right),$$

which tends to zero by the assumptions $b_n \ll n^{-\frac{1}{3}}$, $b_n \ll n^{\frac{\kappa q-2}{q+2}}$, and $L_n \ll n^{\frac{1}{2p}}$. For $p > 2$, we find that

$$\frac{1}{n}\frac{1}{b_n}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor-b_n}\Delta_{t,n}^2 = \mathcal{O}_P\left(n^{\frac{2}{q}-\kappa}b_n^{1+\frac{2}{q}}\right) + \mathcal{O}_P\left(n^{\frac{2}{q}-\kappa}b_n^{1+\frac{2}{q}}n^{\kappa-1}L_n^2\right) + \mathcal{O}_P\left(b_n^3 n^{-\frac{2}{p}}\right),$$

which also tends to zero as $n \to \infty$.

Step (iii): It remains to study the term (24). Introduce the random variables

$$\zeta_{t,n} = \frac{1}{b_n} \left[\sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right]^2,$$

such that we are led to study the term $\frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \zeta_{t,n}$. To bound the covariances of the $\zeta_{t,n}$, we write $\zeta_{t,n} = \zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n})$, such that (A.3) may be exploited. Let $b \geq 0$ be an integer, and recall that

$$\boldsymbol{\epsilon}_{t+b_n,b+b_n}^* = (\epsilon_{t+b_n}, \ldots, \epsilon_{t-b+1}, \epsilon_{t-b}^*, \ldots) \in \mathbb{R}^\infty,$$
$$\boldsymbol{\epsilon}_{t+i,b+i}^* = (\epsilon_{t+i}, \ldots, \epsilon_{t-b+1}, \epsilon_{t-b}^*, \ldots) \in \mathbb{R}^\infty,$$

where the ϵ_j^* are an independent copy of the ϵ_j. Then

$$\left\| \zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n}) - \zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n,b+b_n}^*) \right\|_{L_2}$$

$$= \frac{1}{b_n} \left\| \left[\sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right]^2 - \left[\sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i,b+i}^*) - \mu_{t,n}^n \right] \right]^2 \right\|_{L_2}$$

$$\leq \frac{C}{b_n} \left\| \sum_{i=1}^{b_n} \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i,b+i}^*) \right\| \sum_{j=1}^{b_n} \left[\left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+j}) \right\| + \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+j,b+j}^*) \right\| \right] \right\|_{L_2}$$

via the inequality $|x^2 - y^2| \leq |x - y||x + y|$,

$$\leq \frac{C}{b_n} \sum_{i=1}^{b_n} \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i,b+i}^*) \right\|_{L_4} \left\| \sum_{j=1}^{b_n} \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+j}) \right\| + \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i,b+i}^*) \right\| \right\|_{L_4}$$

by Hölder's inequality,

$$\leq \frac{C}{b_n} \sum_{i=1}^{b_n} \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i,b+i}^*) \right\|_{L_q} \sum_{j=1}^{b_n} \left[\left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+j}) \right\|_{L_q} + \left\| G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+j,b+j}^*) \right\|_{L_q} \right]$$

$$\leq C b_n \rho^b.$$

In the last step, we used $q > 4$, the boundedness of $\| G_n(u, \boldsymbol{\epsilon}_0) \|_{L_q}$ due to (A.2), and the ergodicity (A.3).

Moreover, it holds that

$$\| \zeta_{t,n} \|_{L_2} = \left\| \frac{1}{\sqrt{b_n}} \sum_{i=1}^{b_n} Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right\|_{L_4}^2$$

$$\leq \frac{1}{b_n} \left[\sum_{i=1}^{b_n} \left\| Df(\mu_{\frac{t}{n}}^n) \left[G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_{t+i}) - \mu_{t,n}^n \right] \right\|_{L_4} \right]^2$$

$$\leq Cb_n,$$

because $\|G_n(\frac{t}{n}, \boldsymbol{\epsilon}_0)\|_{L_4}$ is bounded by virtue of (A.2), and $Df(\mu_{t/n}^n)$ is bounded by virtue of (A.4).

Now consider s, t such that $s < t$. If $|s - t| > b_n$, then $\zeta_{s,n}$ is independent of $\boldsymbol{\epsilon}_{t+b_n,|s-t|+b_n}^*$, such that

$$
\begin{aligned}
|\mathrm{Cov}(\zeta_{t,n}, \zeta_{s,n})| &= \left|\mathrm{Cov}\left(\zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n}) - \zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n,|s-t|+b_n}^*), \zeta_{s,n}\right)\right| \\
&\leq \|\zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n}) - \zeta_{t,n}(\boldsymbol{\epsilon}_{t+b_n,|s-t|+b_n}^*)\|_{L_2} \|\zeta_{s,n}\|_{L_2} \\
&\leq Cb_n^2 \rho^{|s-t|}.
\end{aligned}
$$

These findings may be summarized as

$$|\mathrm{Cov}(\zeta_{t,n}, \zeta_{s,n})| \leq Cb_n^2 \rho^{(|s-t|-b_n)\vee 0}. \tag{26}$$

In particular,

$$
\begin{aligned}
\mathrm{Var}\left(\frac{1}{n}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor - b_n} \zeta_{t,n}\right) &= \frac{1}{n^2}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor - b_n} \mathrm{Cov}(\zeta_{t,n}, \zeta_{s,n}) \\
&\leq \frac{Cb_n^2}{n^2}\sum_{s,t=1}^{n} \rho^{(|s-t|-b_n)\vee 0} \\
&\leq \frac{Cb_n^2}{n^2}\sum_{t=1}^{n}\left[b_n + \frac{2}{1-\rho}\right] \quad = \mathcal{O}\left(\frac{b_n^3}{n}\right),
\end{aligned}
$$

which tends to zero by assumption. Thus, $\frac{1}{n}\sum_{t=\tau_n+L_n}^{\lfloor nu\rfloor - b_n}[\zeta_{t,n} - \mathbb{E}\zeta_{t,n}] \xrightarrow{P} 0$ as $n \to \infty$.

<u>Step (iv)</u>: The expected value of $\zeta_{t,n}$ may be computed as

$$
\begin{aligned}
&\mathbb{E}\zeta_{t,n} \\
&= Df(\mu_{\frac{t}{n}}^n)\left[\frac{1}{b_n}\sum_{i,j=1}^{b_n} \mathrm{Cov}\left(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_i), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_j)\right)\right] Df(\mu_{\frac{t}{n}}^n)^T \\
&= Df(\mu_{\frac{t}{n}}^n)\left[\frac{1}{b_n}\sum_{i=1}^{b_n}\sum_{j=-\infty}^{\infty} \mathrm{Cov}\left(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_i), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_j)\right)\right] Df(\mu_{\frac{t}{n}}^n)^T + \mathcal{O}\left(\frac{1}{b_n}\sum_{i=1}^{b_n}[\rho^i + \rho^{b_n-i}]\right) \\
&= Df(\mu_{\frac{t}{n}}^n)\left[\sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_h), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0)\right)\right] Df(\mu_{\frac{t}{n}}^n)^T + \mathcal{O}(\tfrac{1}{b_n}) \\
&= Df(\mu_{\frac{t}{n}}^n)\Sigma_n(\tfrac{t}{n})Df(\mu_{\frac{t}{n}}^n)^T + \mathcal{O}(\tfrac{1}{b_n}),
\end{aligned}
$$

using the geometric ergodicity (A.3) in the second step. Note that the remainder term is of order $1/b_N$, uniformly in t and n, because $\|Df(\mu_{\frac{t}{n}}^n)\| \leq C_f$ for all t, and n large enough.

Just as in the proof of Lemma C.6, step (iv) therein, we find that

$$\left| \frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \mathbb{E}\zeta_{t,n} - \int_0^u Df(\mu_v^n)\Sigma_n(v)Df(\mu_v^n)^T \, dv \right| \to 0,$$

as $n \to \infty$. Moreover, since $\|G_n(\frac{t}{n}, \boldsymbol{\epsilon}_0) - G(\frac{t}{n}, \boldsymbol{\epsilon}_0)\|_{L_2} \to 0$ as $n \to \infty$, uniformly in t, the dominated convergence theorem yields that

$$\left\| \Sigma(\tfrac{t}{n}) - \sum_{h=-\infty}^{\infty} \text{Cov}\left(G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_h), G_n(\tfrac{t}{n}, \boldsymbol{\epsilon}_0)\right) \right\| \to 0,$$

uniformly in t. Furthermore, $\|Df(\mu_{t/n}^n) - Df(\mu_{t/n})\| \to 0$ uniformly in t, such that

$$\frac{1}{n} \sum_{t=\tau_n+L_n}^{\lfloor nu \rfloor - b_n} \mathbb{E}\zeta_{t,n} = \int_0^u Df(\mu_v^n)\Sigma_n(v)Df(\mu_v^n)^T \, dv + o(1)$$

$$= \int_0^u Df(\mu_v)\Sigma(v)Df(\mu_v)^T \, dv + o(1).$$

Together with the variance bound derived in step (iii), this establishes (24) and thus completes the proof. $\qquad\square$

Theorem 3.3 is a consequence of Theorem C.7 for $\kappa = \frac{1}{2}$. Regarding the bootstrap consistency, Theorem 3.4, note that $Q_n(u)$ is precisely the variance process of $\widehat{M}_n(u)$.

Proof of Theorem 3.4. Conditionally on $\mathbb{X}_n$, $\widehat{M}_n(u)$ is a zero mean Gaussian process with independent increments, and variance function $\text{Var}(\widehat{M}_n(u)|\mathbb{X}_n) = Q_n(u)$ for Q_n as in Theorem C.7. It may hence be represented as $\widehat{M}_n(u) = B(Q_n(u))$ for a standard Brownian motion B which is independent of $\mathbb{X}$. The uniform convergence of $Q_n(u) \to Q(u) = \int_0^u Df(\mu_v)\Sigma(v)Df(\mu_v)^T \, dv$ thus yields $\sup_{u\in[0,1]} |B(Q_n(u)) - B(Q(u))| \xrightarrow{P} 0$. This yields weak convergence of $\widehat{M}_n(u)$ in probability. That is, for any probability metric d which metricizes weak convergence on the Skorokhod space, we have $d(\mathcal{L}(M), \mathcal{L}(\widehat{M}_n|\mathbb{X}_n)) \xrightarrow{P} 0$, where $\mathcal{L}(M)$ denotes the measure induced by the random element M. $\qquad\square$

C.6 Properties of the tvVAR example

To show existence of the tvVAR process of Example 1, we want to show that $\|\prod_{j=1}^i A(u - \frac{j-1}{n})\| \le C\rho^i$, where $\|\cdot\|$ denotes the Euclidean operator norm. If $\|A(u)\| \le \rho$ for all $u \in [0,1]$, this conclusion is obvious. However, the latter condition is too restrictive for many applications. For example, consider the univariate autoregressive process of order m given by $X_{t,n} = \sum_{j=1}^m a_j(\frac{t}{n})X_{t-j,n} + \epsilon_t$. This process may be represented as a tvVAR

process with state vector $Y_{t,n} = (X_{t,n}, \ldots, X_{t-m+1,n})$, and autoregressive matrix

$$
A(u) = \begin{pmatrix} a_1(u) & a_2(u) & a_3(u) & \ldots & a_m(u) \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{m \times m}. \tag{27}
$$

A typical condition for the stability of $X_{t,n}$ is that the autoregressive polynomial $1 - \sum_{j=1}^{m} a_j(u) z^j$ has no (complex) roots of absolute value smaller than $\frac{1}{\rho_0} > 1$, see e.g. (Giraud et al., 2015, Sec. 3.1.3). Equivalently, the spectral radius of $A(u)$ is bounded by $\rho_0 < 1$. On the other hand, we have $\|A(u)\| \geq 1$ for all $u \in [0,1]$, irrespective of the coefficients $a_j(u)$.

In view of the previous example, we do not restrict the Euclidean operator norm to be smaller than one. Instead, we impose an upper bound on the spectral radius of the matrices $A(u)$, i.e. $\rho(A(u)) \leq \rho_0 < 1$. This comes at the price of requiring additional regularity of the mapping $u \mapsto A(u)$.

Lemma C.8. *Let $A_i \in \mathbb{R}^{d \times d}$, $i \in \mathbb{N}$, be a sequence of matrices such that $\|A_i\| \leq A^* < \infty$, and $\sup_{i \in \mathbb{N}} \rho(A_i) = \rho_0 < 1$. Assume furthermore that for some $p \geq 1$,*

$$
\|A\|_{p-var} = \left(\sup_{1 \leq i_1 < i_2 < \ldots} \sum_{k=1}^{\infty} \|A_{i_{k+1}} - A_{i_k}\|^p \right)^{\frac{1}{p}} \leq A^* < \infty.
$$

Then for any $\rho > \rho_0$, there exists a $K = K(\rho, \rho_0, A^*)$ such that

$$
\left\| \prod_{i=s+1}^{t} A_i \right\| \leq K \rho^{t-s}, \qquad 0 \leq s < t.
$$

Proof of Lemma C.8. We proceed as in the proof of (Giraud et al., 2015, Lemma 8). For any $s, t \in \mathbb{N}_0$, $s \leq t$, we have

$$
\prod_{i=s+1}^{t} A_i = A_{s+1}^{t-s} + \sum_{k=s+1}^{t} A_{s+1}^{(k-1)-s} (A_k - A_{s+1}) \prod_{i=k+1}^{t} A_i,
$$

$$
\left\| \prod_{i=s+1}^{t} A_i \right\| \leq \|A_{s+1}^{t-s}\| + (t-s)(A^*)^{t-s-1} \max_{s+1 \leq k \leq t} \|A_k - A_{s+1}\|. \tag{28}
$$

Here, we employ the convention $\prod_{i=u}^{v} A_i = I$ if $u > v$, i.e. the empty product is the identity matrix.

Now choose some $\rho_1 \in (\rho_0, \rho)$. By virtue of (Moulines et al., 2005, Lemma 12), there exists a universal $C = C(\rho_0, \rho_1, A^*)$ such that $\|A_{s+1}^{t-s}\| \leq C \rho_1^{t-s}$. We choose $l =$

$l(\rho_0, \rho_1, A^*) \in \mathbb{N}$ such that $C\rho_1^l \leq \frac{\rho^l}{2}$, and let $\epsilon = \frac{\rho^l}{2l(A^*)^{l-1}}$. Split the product into blocks of length l, i.e. we write

$$\prod_{i=s+1}^{t} A_i = \left[\prod_{j=1}^{\lfloor \frac{t-s}{l} \rfloor} \left(\prod_{i=s+(j-1)l+1}^{s+jl} A_i \right) \right] \left[\prod_{i=s+\lfloor \frac{t-s}{l} \rfloor l+1}^{t} A_i \right],$$

$$\left\| \prod_{i=s}^{t} A_i \right\| \leq \left[\prod_{j=1}^{\lfloor \frac{t-s}{l} \rfloor} \left\| \prod_{i=s+(j-1)l+1}^{s+jl} A_i \right\| \right] (A^*)^l.$$

Using (28), we may further bound

$$\delta_j = \left\| \prod_{i=s+(j-1)l+1}^{s+jl} A_i \right\| \leq \begin{cases} \rho^l, & \max_{s+(j-1)l+1 \leq k \leq s+jl} \|A_k - A_{s+(j-1)l+1}\| < \epsilon, \\ (A^*)^l, & \text{otherwise.} \end{cases}$$

There are at most $\kappa = \lceil \frac{(A^*)^p}{\epsilon^p} \rceil$ indices j such that $\delta_j > \rho^l$, because

$$\sum_{j=1}^{\infty} \mathbb{1} \left(\max_{s+(j-1)l+1 \leq k \leq s+jl} \|A_k - A_{s+(j-1)l+1}\| \geq \epsilon \right)$$

$$\leq \epsilon^{-p} \sum_{j=1}^{\infty} \left[\max_{s+(j-1)l+1 \leq k \leq s+jl} \|A_k - A_{s+(j-1)l+1}\|^p \right] \leq \epsilon^{-p} \|A\|_{p-var}^p \leq \kappa.$$

In particular,

$$\prod_{j=1}^{\lfloor \frac{t-s}{l} \rfloor} \left\| \prod_{i=s+(j-1)l+1}^{s+jl} A_i \right\| \leq \rho^{l \lfloor \frac{t-s}{l} \rfloor} \rho^{-l\kappa} (A^*)^{l\kappa},$$

$$\left\| \prod_{i=s+1}^{t} A_i \right\| \leq \rho^{t-s} \rho^{-l(\kappa+1)} (A^*)^{l(\kappa+1)} = \rho^{t-s} K,$$

for some $K = K(\rho, \rho_0, A^*)$. $\qquad \square$

In the literature, results similar to Lemma C.8 have been established to show existence of the tvVAR model. The existing results are formulated for matrices $A_i = A(t - \frac{i}{p})$ for some function $u \mapsto A(u)$. In particular, (Moulines et al., 2005, Proposition 13) requires $u \mapsto A(u)$ to be Hölder continuous, and Lemma 8 of Giraud et al. (2015) replaces Hölder continuity by some arbitrarily weak uniform continuity. Proposition 2.4 of Dahlhaus and Polonik (2009) requires $u \mapsto A(u)$ to be of bounded variation. The bounded p-variation imposed in Lemma C.8 is weaker than the assumptions of Moulines et al. (2005) and Dahlhaus and Polonik (2009). Compared to Giraud et al. (2015), we allow for discontinuities. On the other hand, Giraud et al. only require uniform continuity, without further restrictions on the modulus of continuity, which also includes some irregular continuous cases where $\|A\|_{p-var} = \infty$ for all $p \geq 1$.

Proposition C.9. *Let $G_n(u, \epsilon_t)$ as in Example 1, with $\sup_u \|A(u)\| < \infty$ and $\sup_u \|B(u)\| < \infty$, and suppose that the matrices $A(u)$ have spectral radius $\rho(A(u)) \le \rho_0 < 1$ for all $u \in [0, 1]$.*

1. *If $\|A\|_{p-var} < \infty$, $\|B\|_{p-var} < \infty$, and $\|\mu\|_{p-var} < \infty$ for some $p \ge 1$, then there exists some $C = C(A, B, \mu)$ such that $\|G_n\|_{p-var} + \sup_{u \in [0,1]} \|G_n(u, \epsilon_0)\|_{L_q} \le C$ for all n.*

2. *If $u \mapsto A(u)$, $u \mapsto B(u)$, and $u \mapsto \mu(u)$ are β-Hölder continuous in the Euclidean operator norm $\| \cdot \|$, then there exists $C_H = C_H(A, B, \mu) < \infty$ such that for all $n \in \mathbb{N}$, the mapping $u \mapsto G_n(u, \epsilon_t)$ is β-Hölder continuous in $L_q(P)$ with Hölder constant C_H.*

Proof of Proposition C.9. Note that $\mathbb{E}\|\epsilon_0\|^q < \infty$ by assumption. In the sequel, the factor $C = C(A, B, \mu)$ may vary from line to line. For any $u \in [0, 1]$, we have

$$\|G_n(u, \epsilon_0)\|_{L_q} \le C \sum_{i=0}^{\infty} \left\| \prod_{j=1}^{i} A(u - \tfrac{j-1}{n}) \right\| \left\| B(u - \tfrac{i}{n}) \right\| + \|\mu(u)\|$$

$$\le C \sum_{i=0}^{\infty} \rho^i + \|\mu(0)\| + \|\mu\|_{p-var} \quad \le C(A, B, \mu).$$

In the last step, we used Lemma C.8 for some $\rho \in (\rho_0, 1)$. In the Hölder continuous case, Lemma C.8 is applicable because β-Hölder continuity implies finite $\frac{1}{\beta}$-variation. In the case of finite p-variation, Lemma C.8 is directly applicable.

We now consider the regularity of $u \mapsto G_n(u, \epsilon_0)$. For any $u, v \in [0, 1]$, we have

$$\|G_n(u, \epsilon_0) - G_n(v, \epsilon_0)\|_{L_q}$$

$$\le C \sum_{i=0}^{\infty} \left\| \left[\prod_{j=1}^{i} A\left(u - \tfrac{j-1}{n}\right) \right] B\left(u - \tfrac{i}{n}\right) - \left[\prod_{j=1}^{i} A\left(v - \tfrac{j-1}{n}\right) \right] B\left(v - \tfrac{i}{n}\right) \right\| + \|\mu(u) - \mu(v)\|$$

$$\le C \sum_{i=0}^{\infty} \left\| \prod_{j=1}^{i} A\left(u - \tfrac{j-1}{n}\right) \right\| \left\| B\left(u - \tfrac{i}{n}\right) - B\left(v - \tfrac{i}{n}\right) \right\| + \|\mu(u) - \mu(v)\|$$

$$+ C \sum_{i=0}^{\infty} \sum_{k=1}^{i} \left[\left\| \prod_{j=1}^{k-1} A\left(u - \tfrac{j-1}{n}\right) \right\| \left\| \prod_{j=k+1}^{i} A\left(v - \tfrac{j-1}{n}\right) \right\| \right]$$

$$\cdot \left\| A(u - \tfrac{k-1}{n}) - A(v - \tfrac{k-1}{n}) \right\| \left\| B\left(v - \tfrac{i}{n}\right) \right\|$$

$$\le C \sum_{i=0}^{\infty} \rho^i \left\| B\left(u - \tfrac{i}{n}\right) - B\left(v - \tfrac{i}{n}\right) \right\| + C \sum_{i=0}^{\infty} \sum_{k=1}^{i} \rho^{i-1} \left\| A(u - \tfrac{k-1}{n}) - A(v - \tfrac{k-1}{n}) \right\|$$

$$+ \|\mu(u) - \mu(v)\|. \tag{29}$$

Now if μ, A, and B are β-Hölder continuous with Hölder constants C_μ, C_A, and C_B, we find that

$$\|G_n(u, \epsilon_0) - G_n(v, \epsilon_0)\|_{L_q}$$

$$\leq C \sum_{i=0}^{\infty} \rho^i |u-v|^\beta C_B + C \sum_{i=0}^{\infty} \sum_{k=1}^{i} \rho^{i-1} |u-v|^\beta C_A + C_\mu |u-v|^\beta$$

$$\leq |u-v|^\beta (C_\mu + C_A + C_B) C \left[1 + \sum_{i=0}^{\infty} \rho^i + \sum_{i=0}^{\infty} i\, \rho^{i-1} \right],$$

and the latter series are finite since $\rho < 1$.

Alternatively, we may treat the case of finite p-variation. Let $u_0 < u_1 < \ldots < u_m$. Applying the Minkowski inequality to (29), we find obtain

$$\left[\sum_{l=1}^{m} \| G_n(u_l, \boldsymbol{\epsilon}_0) - G_n(u_{l-1}, \boldsymbol{\epsilon}_0) \|_{L_q}^p \right]^{\frac{1}{p}}$$

$$\leq C \sum_{i=0}^{\infty} \rho^i \left[\sum_{l=1}^{m} \| B\left(u_l - \tfrac{i}{n}\right) - B\left(u_{l-1} - \tfrac{i}{n}\right) \|^p \right]^{\frac{1}{p}}$$

$$+ C \sum_{i=0}^{\infty} \sum_{k=1}^{i} \rho^{i-1} \left[\sum_{l=1}^{m} \| A(u_l - \tfrac{k-1}{n}) - A(u_{l-1} - \tfrac{k-1}{n}) \|^p \right]^{\frac{1}{p}} + \| \mu \|_{p-var}$$

$$\leq C \sum_{i=0}^{\infty} \rho^i \| B \|_{p-var} + C \sum_{i=0}^{\infty} \sum_{k=1}^{i} \rho^{i-1} \| A \|_{p-var} + \| \mu \|_{p-var}$$

$$\leq C(\| B \|_{p-var} + \| A \|_{p-var} + \| \mu \|_{p-var}).$$

Since $C = C(A, B, \mu)$, this completes the proof. $\qquad\square$

C.7 Properties of the CUSUM test

Proof of Proposition 4.1. By virtue of a result of Lifshits (1983), see the Corollary on p. 606 therein, we find that the distribution of T^* is absolutely continuous, apart from a potential atom at 0. Since $\mathrm{Var}(M(u_0)) > 0$ for some $u_0 \in (0,1)$, the variance structure of the process $M(u)$ yields that $\mathrm{Var}(M(u_0) - u_0 M(1)) > 0$. Due to the Gaussianity of $M(u)$, it holds that $P(|M(u_0) - u_0 M(1)| = 0) = 0$, such that $P(T^* = 0) = 0$. Hence, the cumulative distribution function $H(x) = P(T^* \leq x)$ is continuous. Denote furthermore by $H_n(x) = P(\widehat{T}_n | \mathbb{X}_n)$ the (random) distribution function of the bootstrap approximation. We have to show that $\mathbb{E}\left[H(H_n^{-1}(1-\alpha)) \right] \to 1 - \alpha$ as $n \to \infty$, where H^{-1} denotes the quantile function corresponding to a cumulative distribution function $H : \mathbb{R} \to [0,1]$.

Suppose to the contrary that $\mathbb{E}\left[H(H_n^{-1}(1-\alpha)) \right] \not\to 1 - \alpha$, then for some $\epsilon > 0$, there exists a subsequence $n_k \to \infty$ as $k \to \infty$ such that $|\mathbb{E}\left[H(H_{n_k}^{-1}(1-\alpha)) \right] - (1-\alpha)| > \epsilon$. Since $H_{n_k} \Rightarrow H$ in probability, there exists a further subsequence, also denoted by n_k, such that $H_{n_k} \Rightarrow H$ almost surely as $k \to \infty$. This implies that $H_{n_k}^{-1}(p) \to H^{-1}(p)$, for all $p \in [0,1]$ where $H^{-1}(p)$ is continuous (Van der Vaart, 1998, Lemma 21.2). Since H^{-1} is left-continuous and has at most countably many discontinuities, we find that

$$H^{-1}(1-\alpha) \leq \liminf_{k \to \infty} H_{n_k}^{-1}(1-\alpha) \leq \limsup_{k \to \infty} H_{n_k}^{-1}(1-\alpha) \leq H^{-1}((1-\alpha)+),$$

where $H^{-1}(p+) = \lim_{q \downarrow p} H^{-1}(q)$ denotes the right-hand limit. Since F is continuous, we have $H(H^{-1}(1-\alpha)) = H(H^{-1}((1-\alpha)+)) = 1-\alpha$, such that $H(H_{n_k}^{-1}(1-\alpha)) \to 1-\alpha$ almost surely. Via the dominated convergence theorem, we obtain $\mathbb{E}\left[H(H_{n_k}^{-1}(1-\alpha))\right] \to (1-\alpha)$, completing the proof by contradiction. $\qquad\square$

Proof of Proposition 4.2. It holds that

$$
T_n(u) = \left[M_n(u) - \frac{(u-u_n)_+}{1-u_n}M_n(1)\right] + \frac{1}{n}\left[\sum_{t=\tau_n+L_n}^{\lfloor un \rfloor} f(\mu_{\frac{t}{n}}^n) - \frac{(u-u_n)_+}{1-u_n}\sum_{t=\tau_n+L_n}^{n} f(\mu_{\frac{t}{n}}^n)\right]
$$

$$
= \left[M_n(u) - \frac{(u-u_n)_+}{1-u_n}M_n(1)\right] + \Delta_n(u).
$$

By virtue of Theorem 3.2, $\sqrt{n}[M_n(u) - \frac{(u-u_n)_+}{1-u_n}M_n(1)] \Rightarrow M(u)$. Moreover,

$$
\Delta_n(u) = \frac{1}{n}\left[\sum_{t=\tau_n+L_n}^{\lfloor un \rfloor} f(\mu_{\frac{t}{n}}^n) - \frac{(\frac{\lfloor un \rfloor}{n}-u_n)_+}{1-u_n}\sum_{t=\tau_n+L_n}^{n} f(\mu_{\frac{t}{n}}^n)\right] + \mathcal{O}(\tfrac{1}{n})
$$

$$
= \frac{1}{n}\sum_{t=\tau_n+L_n}^{\lfloor un \rfloor}\left[f(\mu_{\frac{t}{n}}^n) - \frac{1}{n+1-\tau_n-L_n}\sum_{s=\tau_n+L_n}^{n} f(\mu_{\frac{s}{n}}^n)\right] + \mathcal{O}(\tfrac{1}{n}).
$$

Now use that $f(\mu_u^n) = f(\mu_u) + \frac{1}{\sqrt{n}}Df(\mu_u)\delta_u + \mathcal{O}(1/n)$, uniformly in $u \in [0,1]$ by virtue of the boundedness of f. Since $f(\mu_u) = f(\mu_0)$ for all $u \in [0,1]$, we obtain

$$
\Delta_n(u) = \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{t=\tau_n+L_n}^{\lfloor un \rfloor}\left[Df(\mu_{\frac{t}{n}})\delta_{\frac{t}{n}} - \frac{1}{n+1-\tau_n-L_n}\sum_{s=\tau_n+L_n}^{n} Df(\mu_{\frac{s}{n}})\delta_{\frac{s}{n}}\right] + \mathcal{O}(\tfrac{1}{n}),
$$

$$
\sqrt{n}\Delta_n(u) = o(1) + \frac{1}{n}\sum_{t=1}^{\lfloor un \rfloor} Df(\mu_{\frac{t}{n}})\delta_{\frac{t}{n}} - \frac{u}{n}\sum_{t=1}^{n} Df(\mu_{\frac{t}{n}})\delta_{\frac{t}{n}}
$$

$$
= o(1) + \Delta(u).
$$

In the last step, we use Lemma C.4 and the fact that $\|u \mapsto Df(\mu_u)\delta_u\|_{p-var} < \infty$. Moreover, the convergence $\sqrt{n}\Delta_n(u) \to \Delta(u)$ holds uniformly in $u \in [0,1]$. This establishes the convergence of T_n^*.

$\qquad\square$

References

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press, Princeton.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175.

Andersen, T. G., Archakov, I., Cebiroglu, G., and Hautsch, N. (2019). Local Mispricing and Microstructural Noise: A Parametric Perspective. Technical report.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2008). Testing for changes in polynomial regression. *Bernoulli*, 14(3):637–660.

Baby, D. and Wang, Y.-X. (2019). Online forecasting of total-variation-bounded sequences. *Advances in Neural Information Processing Systems*, 32:11071–11081.

Bauer, P. and Hackl, P. (1978). The Use of MOSUMS for Quality Control. *Technometrics*, 20(4):431.

Berkes, I., Gombay, E., and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference*, 139(6):2044–2063.

Bickel, P. J. and Ritov, Y. (1988). Estimating Integrated Squared Density Derivatives: Sharp Best Order of Convergence Estimates. *Sankhya: The Indian Journal of Statistics, Series A*, 50(3):381–393.

Billingsley, P. (1999). *Convergence of probability measures*. John Wiley & Sons.

Carlstein, E. (1986). The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence. *The Annals of Statistics*, 14(3):1171–1179.

Chen, J. and Gupta, A. K. (1997). Testing and Locating Variance Changepoints with Application to Stock Prices. *Journal of the American Statistical Association*, 92(438):739–747.

Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995). MOSUM Tests for Parameter Constancy. *Biometrika*, 82(3):603.

Cui, Y., Levine, M., and Zhou, Z. (2020). Estimation and Inference of Time-Varying Auto-Covariance under Complex Trend: A Difference-based Approach. *arXiv preprint 2003.05006*.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.

Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *Journal of Econometrics*, 151(2):101–112.

Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39.

Dahlhaus, R. and Richter, S. (2019). Adaptation for nonparametric estimators of locally stationary processes. *arXiv preprint 1902.10381*.

Dahlhaus, R., Richter, S., and Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2):1013–1044.

Demetrescu, M. and Wied, D. (2018). Testing for constant correlation of filtered series under structural change. *Econometrics Journal*, 22:10–33.

Dette, H. and Gösmann, J. (2020). A Likelihood Ratio Approach to Sequential Change Point Detection for a General Class of Parameters. *Journal of the American Statistical Association*, 115(531):1361–1377.

Dette, H. and Wu, W. (2019). Detecting relevant changes in the mean of nonstationary processes—A mass excess approach. *The Annals of Statistics*, 47(6):3578–3608.

Dette, H., Wu, W., and Zhou, Z. (2019). Change Point Analysis of Correlation in Non-stationary Time Series. *Statistica Sinica*, 29:611–643.

Gao, Z., Shang, Z., Du, P., and Robertson, J. L. (2019). Variance Change Point Detection Under a Smoothly-Changing Mean Trend with Application to Liver Procurement. *Journal of the American Statistical Association*, 114(526):773–781.

Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press.

Giraud, C., Roueff, F., and Sanchez-Perez, A. (2015). Aggregation of predictors for nonstationary sub-linear processes and online adaptive forecasting of time varying autoregressive processes. *The Annals of Statistics*, 43(6):2412–2450.

Górecki, T., Horváth, L., and Kokoszka, P. (2018). Change point detection in heteroscedastic time series. *Econometrics and Statistics*, 7:63–88.

Gösmann, J., Kley, T., and Dette, H. (2021). A new approach for open-end sequential change point monitoring. *Journal of Time Series Analysis*, 42:63–84.

Grenier, Y. (1983). Time-dependent arma modeling of nonstationary signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(4):899–911.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters*, 6(2):109–115.

Hansen, P. R. and Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Horváth, L. (1995). Detecting changes in linear regressions. *Statistics*, 26(3):189–208.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126(1):225–251.

Hsu, D.-A., Miller, R. B., and Wichern, D. W. (1974). On the stable paretian behavior of stock-market prices. *Journal of the American Statistical Association*, 69(345):108–113.

Huang, L. S. and Jianqing, F. A. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli*, 5(5):927–949.

Inclán, C. and Tiao, G. C. (1994). Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance. *Journal of the American Statistical Association*, 89(427):913–923.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, 119(7):2249 – 2276.

Jacod, J., Li, Y., and Zheng, X. (2017). Statistical properties of microstructure noise. *Econometrica*, 85(4):1133–1174.

Jacod, J. and Rosenbaum, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics*, 41(3):1462–1484.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg.

Juhl, T. and Xiao, Z. (2009). Tests for changing mean with monotonic power. *Journal of Econometrics*, 148(1):14–24.

Killick, R., Eckley, I. A., and Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, 7(1):1167–1183.

Lee, S. and Park, S. (2001). The Cusum of Squares Test for Scale Changes in Infinite Order Moving Average Processes. *Scandinavian Journal of Statistics*, 28(4):625–644.

Li, X. and Zhao, Z. (2013). Testing for changes in autocovariances of nonparametric time series models. *Journal of Statistical Planning and Inference*, 143(2):237–250.

Lifshits, M. A. (1983). On the absolute continuity of distributions of functionals of random processes. *Theory of Probability & Its Applications*, 27(3):600–607.

Liu, W., Xiao, H., and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, 23:1257–1272.

Moulines, E., Priouret, P., and Roueff, F. (2005). On recursive estimation for time varying autoregressive processes. *The Annals of Statistics*, 33(6):2610 – 2654.

Page, E. (1954). Continuous Inspection Schemes. *Biometrika*, 41(1/2):100–115.

Page, E. (1955). A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika*, 42(3/4):523–527.

Pešta, M. and Wendler, M. (2020). Nuisance-parameter-free changepoint detection in non-stationary series. *TEST*, 29:379–408.

Potiron, Y. and Mykland, P. (2020). Local Parametric Estimation in High Frequency Data. *Journal of Business & Economic Statistics*, 38(3):679–692.

Preuss, P., Puchstein, R., and Dette, H. (2015). Detection of Multiple Structural Breaks in Multivariate Time Series. *Journal of the American Statistical Association*, 110(510):654–668.

Raj, A., Gaillard, P., and Saad, C. (2020). Non-stationary online regression. *arXiv preprint 2011.06957*.

Robbins, M. W., Gallagher, C. M., and Lund, R. B. (2016). A General Regression Changepoint Test for Time Series Data. *Journal of the American Statistical Association*, 111(514):670–683.

Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *Journal of Nonparametric Statistics*, 16(6):925–935.

Schmidt, S., Wornowizki, M., Fried, R., and Dehling, H. (2020). An Asymptotic Test for Constancy of the Variance under Short-Range Dependence. *arXiv preprint 2002.10178*.

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

Steland, A. (2020). Testing and estimating change-points in the covariance matrix of a high-dimensional time series. *Journal of Multivariate Analysis*, 177:104582.

Subba Rao, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(2):312–322.

Truquet, L. (2019). Local stationarity and time-inhomogeneous Markov chains. *The Annals of Statistics*, 47(4):2023–2050.

van de Geer, S. (2010). *Empirical Processes in M-Estimation*. Cambridge University Press.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes. *Annals of Statistics*, 43(2):713–740.

Wichern, D. W., Miller, R. B., and Hsu, D.-A. (1976). Changes of variance in first-order autoregressive time series models -with an application. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 25(3):248.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, 21(3):1397–1413.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108(502):726–740.

Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Annals of Statistics*, 37(5 B):2696–2729.