Functional Estimation and Change Detection for Nonstationary Time Series

Fabian Mies
Institute of Statistics, RWTH Aachen University, Aachen, Germany

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 article, 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.

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. Forexample,economictimeseriesaretypicallyheteroscedastic,thistemporalhomogeneityisoftenviolatedinpractice.

In this article, we study a multivariate, locally stationary time series model given by the causal representation
\[ X_{t,n} = G_n(u, \varepsilon_i), \quad t = 1, \ldots, n, \]
where \( \varepsilon_i = (\varepsilon_i, \varepsilon_{i-1}, \ldots) \in \mathbb{R}^\infty \) for iid random variables \( \varepsilon_i \). For each fixed \( u \in [0,1] \) and \( n \in \mathbb{N} \), the sequence \( (G_n(u, \varepsilon_i))_{i \in \mathbb{Z}} \) is stationary. By letting the kernel depend on the fraction \( \frac{t}{n} \), 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 \( u \mapsto G_n(u, \varepsilon_i) \) 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 \( u \mapsto G_n(u, \varepsilon_i) \) in this article 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 \( \theta^n_u \), which is a functional of the law of \( G(u, \varepsilon_i) \), resp. \( G_n(u, \varepsilon_i) \). The estimator proposed in this article is applicable for any parameter of the form \( \theta^n_u = f(\mu^n_u) \), where \( \mu^n_u = \mathbb{E} G_n(u, \varepsilon_i) \), 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 \( \theta^n_u = f(\mathbb{E}[h(G_n(u, \varepsilon_i))]) \), that is, 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, for example, its temporal average \( \frac{1}{t_0} \int_0^t \theta_u \, du \) as studied by Potiron and Mykland (2020), its maximum value \( \sup_{u \in [0,1]} ||\theta_u|| \), or its value \( \theta_{u_0} \) at some \( u_0 \in [0,1] \). For instance, estimation of \( \theta_{u_0} \) is studied by Cui, Levine, and Zhou (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 article, 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 \( \hat{M}_n(u) \) of \( \Theta(u) \). The function \( \Theta(u) \) contains all information about the local
parameter $\theta_0$, 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, for example, 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), 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 $\theta_i$ 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, for example, by Dahlhaus (2009), who considers linear functionals of the time-varying spectral density.

A general method to estimate integrated parameters was suggested by Potirron and Mykland (2020). They constructed blockwise estimators $\hat{\theta}_n$, and averaged 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 $\hat{\theta}_n$ entails additional analytical effort for each special case. In particular, they require strong conditions on the bias of $\hat{\theta}_n$, 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 $\hat{\mu}_n$, 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 $\hat{\mu}_n$. 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 $\theta_0$ is the same for all $u \in [0,1]$, which may be formulated equivalently as

$$H_0 : \theta_u \equiv \theta_0 \iff H_0 : \Theta(u) \equiv u\Theta(1).$$

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, for example, the stability of variances (Gao et al. 2019), regression coefficients (Horváth 1995), or autocovariances (Berkes, Gombay, and Horváth 2009; Killick, Eckley, and Jonathan 2013; Preuss, Puchstein, and Dette 2015). Our approach provides a unifying framework to study these problems for parameters $\theta_u$ 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, for example, a test for the temporal stability of kurtosis. The proposed change-point tests are robust against various model misspecifications, for example, 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 $\mu_u$ 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.

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 was 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 article are conveniently verified for a broad range of practical problems. The framework of Shao and Zhang (2010) was also adopted by Dette and Gösmann (2020) and applied to the monitoring of quantiles, as well as by Gösmann, Kley, and Dette (2021). Note that these authors studied the stationary case only, while we allow for nonstationarity. It might be of interest to extend the methodology introduced in the present article to more general functionals. We leave this question for future work.

The outline of this article 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. 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 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 by $Df(x) = (\partial_{x_1}, \ldots, \partial_{x_d})f(x) \in \mathbb{R}^{1 \times d}$ the first-order differential, and by $D^2f(x) \in \mathbb{R}^{d \times d}$ 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 a vector $x \in \mathbb{R}^d$, 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, that is, $||A|| = \sup_{x \neq 0} |\frac{Ax}{||x||}|$. The spectral radius of a matrix $A$ is denoted as $\rho(A) = \max(\lambda_1, \ldots, \lambda_d)$, where the $\lambda_i$ 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 $||X||_{L_q} = (E||X||^q)^{1/q}$ 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 $X_{t,n} = G_n(t/n, \epsilon_t)$ for a sequence of functions $G_n : \mathbb{R} \times \mathbb{R}_n \rightarrow \mathbb{R}^d$. Here, we denote $\epsilon_t = (\epsilon_1, \epsilon_{t-1}, \ldots) \in \mathbb{R}^\infty$, where the $\epsilon_t$ are iid random variables. The functions $G_n$ are assumed to be measurable, where we endow the sequence space $\mathbb{R}_n^\infty$ with the projection $\sigma$-Algebra, see Billingsley (1999, exampl. 1.2).

By using a sequence of functions $G_n$, we will be able to apply our results to investigate 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. We assume that the kernel $G_n$ tends to a limiting kernel $G : \mathbb{R} \times \mathbb{R}_\infty \rightarrow \mathbb{R}^d$, in the sense that

$$\sup_{u \in [0,1]} ||G_n(u, \epsilon_0) - G(u, \epsilon_0)||_{L_q} \rightarrow 0, \quad n \rightarrow \infty, \quad (A.1)$$

for some $q > 2$. Note that we do not require a rate of convergence in Equation (A.1). Furthermore, we require the function $u \mapsto G_n(u, \epsilon_0) \in L_q(P)$ to satisfy some regularity conditions. A very mild condition can be formulated in terms of $p$-variation for some $p \geq 1$, defined as

$$||G_n||_{p-var} = \sup_{0 = u_0 < u_1 < \cdots < u_m = 1} \left( \sum_{i=1}^{m} ||G_n(u_i, \epsilon_t) - G_n(u_{i-1}, \epsilon_t)||_{L_q}^p \right)^{1/p}. \quad (A.2)$$

The latter definition is independent of the chosen $t$, since $\epsilon_t \sim \epsilon_0$. The $p$-variation of the limiting kernel is denoted analogously as $||G||_{p-var}$. We assume that

$$\sup_{n u \in [0,1]} ||G_n(u, \epsilon_0)||_{L_q} + \sup_{n} ||G_n||_{p-var} \leq C_G < \infty, \quad (A.2)$$

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, Wu, and Zhou (2019), among others. In contrast, finite $p$-variation still allows for infinitely many discontinuities of the mapping $u \mapsto G_n(u, \epsilon_0) \in L_q(P)$. On the other hand, if the latter mapping is Hölder continuous with exponent $\beta$, then $||G_n||_{p-var} < \infty$ for $p \geq \frac{\beta}{\beta - 1}$. 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, Richter, and Wu (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 $\epsilon_t^*$ of the $\epsilon_t$. Denote for $t \in \mathbb{Z}_+, j \in \mathbb{Z}_\geq 0$,

$$\epsilon_{t,j} = (\epsilon_{t+1}, \ldots, \epsilon_{t+j}, \epsilon_{t-j}, \ldots) \in \mathbb{R}_\infty, \quad \epsilon_{t,j}^* = (\epsilon_{t+1}^*, \ldots, \epsilon_{t+j}^*, \epsilon_{t-j}^*, \ldots) \in \mathbb{R}_\infty.$$

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

$$||G_n(u, \epsilon_t) - G_n(u, \epsilon_{t,j}^*)||_{L_q} \leq C_G \rho^j. \quad (A.3)$$

This implies that

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

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 Equations (A.1)–(A.3) hold, for some $q \geq 2$. Suppose furthermore that $d = 1$, and denote by

$$\sigma^2(u) = \sum_{h=-\infty}^{\infty} \text{Cov} [G(u, \epsilon_h), G(u, \epsilon_0)],$$

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

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

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 càdlàg functions $g : [0,1] \rightarrow \mathbb{R}^d$, 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

$$X_{t,n} = A(t) X_{t-1,n} + \mu(t) + B(t) \epsilon_t + \mu(t),$$

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

where $\epsilon_t$ is a sequence of iid, $d$-dimensional random vectors with finite moments of all orders, $\mu : [0,1] \rightarrow \mathbb{R}^d$, and $A, B : [0,1] \rightarrow \mathbb{R}^{d \times d}$ 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 example, an autoregression of order two may be described in terms of the state vector \(Y_{t,u} = (X_{t,n}, X_{t-1,n})\), 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, Priouret, and Roueff (2005), Dahlhaus and Polonik (2009), and Giraud, Roueff, and Sanchez-Perez (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 \(X_{t,n} = Gn(u^n, \epsilon^n)\) as

\[
G(u^n, \epsilon^n) = \sum_{i=0}^{\infty} \prod_{j=1}^{i} A \left( u - \frac{u^n}{n} \right) B \left( u - \frac{u^n}{n} \right) \epsilon_{n-i} + \mu(u^n).
\]

This infinite sum is well-defined if we suppose that \(\sup_{u^n} ||B(u^n)||, \sup_{u^n} ||A(u^n)|| < \infty\), 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 \(\sup_{u^n} \left| A \left( u - \frac{u^n}{n} \right) \right| \leq C_\rho^p\) for any \(\rho \in (\rho_0, 1)\), see Lemma C.8 in the appendix. Hence, assumption (A.3) is satisfied. Furthermore, if the functions \(A, B\) are left-continuous, we have \(|G_{n}(u^n, \epsilon_0) - G(u^n, \epsilon_0)|_{L^q} \rightarrow 0\) by dominated convergence for any \(q > 2\), with limiting kernel

\[
G(u^n, \epsilon^n) = \sum_{i=0}^{\infty} A(u^n)B(u^n)\epsilon_{n-i} + \mu(u^n).
\]

In particular, condition (A.1) holds. Finally, Proposition C.9 in the appendix shows that assumption (A.2) holds if \(||A||_{p-\text{var}} + ||B||_{p-\text{var}} + ||\mu||_{p-\text{var}} < \infty\). Moreover, in the more regular case that \(\mu, A, B\) are \(\beta\)-Hölder continuous, the kernel \(G_{n}(u^n, \epsilon_0)\) is also \(\beta\)-Hölder continuous in \(L_{q}(P)\), uniformly in \(n\), and thus the same holds for \(u \mapsto \mu_n^n = \mathbb{E}G_{n}(u^n, \epsilon_0)\). 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

\[
\mu_n^n = \mathbb{E}G_{n}(u^n, \epsilon_0) \in \mathbb{R}^d,
\]

\[
\mu_n = \lim_{n \to \infty} \mu_n^n = \mathbb{E}G(u, \epsilon_0) \in \mathbb{R}^d, \quad u \in [0,1].
\]

Inference for the function \(\mu_n\) resp. \(\theta_n = f(\mu_n^n)\) can be performed in various ways. For example, one might assume a parametric form for the mapping \(u \mapsto \mu_n^n\). Here, we are interested in testing nonparametric hypotheses imposed on the function \(\mu_n\). 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

\[
F_n(u) = \int_0^u f(\mu^n_n) \, dv, \quad u \in [0,1].
\]

Note that most hypotheses on \(\theta_n = f(\mu^n_n)\) may be reformulated in terms of \(F_n(u)\). In Section 4, we will study change-point detection, where the null hypothesis \(H_0 : f(\mu^n_n) \equiv f(\mu^n_{0n})\) for \(u \in [0,1]\) is equivalent to the hypothesis \(H_0 : F_n(u) \equiv uF(1)\). It is also possible to study the hypothesis \(H_0 : f(\mu^n_n) \equiv 0\), which is equivalent to \(H_0 : F_n(u) \equiv 0\). Another appealing aspect of studying the integrated quantity \(F_n(u)\) instead of \(f(\mu^n_n)\) 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 convergence.

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

\[
\hat{M}_n(u) = \frac{1}{n} \sum_{i=\tau}^{\lfloor un \rfloor} f(\hat{\mu}_{i,n}),
\]

where \(\hat{\mu}_{i,n}\) is a nonparametric estimator of \(\mu_{i,n}\), 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 \(\hat{\mu}_{i,n}\). Thus, additional theoretical effort is required when adapting corresponding inferential procedures to new situations. Second, for nonlinear \(f\), the estimator \(\hat{M}_n(u)\) may be bias-dominated, impeding statistical inference.

In particular, a Taylor expansion yields

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

For most estimators \(\hat{\mu}_{i,n}\), 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 \(u \mapsto \mu_{n}\) to be Lipschitz continuous, then the bias of \(\hat{M}_n(u)\) is of order \(O(\frac{k}{n} + \frac{1}{\sqrt{n}})\), which is at best \(O(1/\sqrt{n})\). In a related situation, Jacod and Rosenbaum (2013) suggested to solve this problem by undersmoothing, that is, choosing \(k_n \ll \sqrt{n}\), and correcting the quadratic bias term in Equation (4) explicitly; see also (Potiron and Mykland 2020, sec. 4.2). However, the latter quadratic bias term strongly depends on the specific estimator \(\hat{\mu}_{i,n}\), so that the approach of Jacod and Rosenbaum (2013) may not be easily transferred to different problems.

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

\[
M_n(u) = \frac{1}{n} \sum_{i=\tau + L}^{\lfloor un \rfloor} \left[ f(\hat{\mu}_{i-L,n}) + Df(\hat{\mu}_{i-L,n})(X_{i,n} - \hat{\mu}_{i-L,n}) \right],
\]

\[u \in [0,1].\]

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 $\tilde{\mu}_{t-L,n}$ and $\mu^n_{t/n}$,

$$M_n(u) = \frac{1}{n} \sum_{i=t-a+n}^{\lfloor nu \rfloor} f(\mu^n_{i/n}) + \frac{1}{n} \sum_{i=t-a+n}^{\lfloor nu \rfloor} (\mu^n_{i/n} - \tilde{\mu}_{t-L,n})^T Df(\mu^n_{i/n}) + \frac{1}{n} \sum_{i=t-a+n}^{\lfloor nu \rfloor} Df(\mu^n_{i/n}) (X_{i,n} - \mu^n_{i/n})$$

$$= I^n_n(u) + I^n_{a,n}(u) + I^n_{b,n}(u).$$

It can be shown that $I^n_{a,n}(u) - F_n(u)$ converges to zero sufficiently fast. Moreover, if $f$ is sufficiently regular, and the estimator $\hat{\mu}_{t,n}$ is good enough, then $\sqrt{n} I^n_{a,n}(u) \to 0$, whereas $\sqrt{n} I^n_{b,n}(u)$ 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 $\mu_t$. Formally, introduce the convex set $M = \{\mu_t : u \in [0, 1]\} \subset \mathbb{R}^d$, and its $\delta$-neighborhood $M^\delta = \{m \in \mathbb{R}^d : ||m - \tilde{m}|| < \delta \}$ for some $\tilde{m} \in M$. We require that there exists a $\delta > 0$ such that

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

By restricting the boundedness assumption to the set $M^\delta$, we may also consider functions of the form $f(x, y) = x/y$, if $y \geq \varepsilon > 0$ on the set $M^\delta$. Without localization, Equation (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 $\epsilon_i$, that is, $\hat{\mu}_{t,n}$ 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 $I^n_{a,n}(u)$. 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

$$\sum_{t=s}^{n} ||\hat{\mu}_{t,n} - \mu^n_t||^2 = o_P(\sqrt{n}), \quad (A.5)$$

$$P(\hat{\mu}_{t,n} \in M_{\delta}, t = n_0, \ldots, n) = 1 + o(1). \quad (A.6)$$

The initial offset $t_0$ 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, that is, $\delta = \infty$, then (A.6) is vacuous, and $t_0 = 1$ 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 supplementary material. If we are willing to impose some additional smoothness conditions, then a suitable nonparametric estimator may be obtained by local averaging. In particular, we define

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

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 Equation (A.3) hold for some $q > 2$, and choose $k_n \gg n^{2/q}$, then $\hat{\mu}_{t,n}^{NW}$ satisfies Equation (A.6) for any offset sequence $t_n \to \infty$. Suppose that for all $n$, we have $\mu^n_{t,n} = \mu^n_{t,n}^{1} + \mu^n_{t,n}^{2}$, such that $||\mu^n_{t,n}^{1}||_{p-\text{var}} \leq a_n$ for $p \in [1, 2]$, and such that $u \mapsto \mu^n_{t,n}^{1}$ is $\beta$-Hölder continuous with Hölder constant $C$, then

$$\sum_{t=1}^{n} \left|\hat{\mu}_{t,n}^{NW} - \mu^n_{t} \right| \leq C \left( a_n k_n \right).$$

In particular, $\hat{\mu}_{t,n}^{NW}$ satisfies Equation (A.5) if $a_n k_n \ll \sqrt{n}$, and

$$\sqrt{n} \log(n) \ll k_n \ll n^{2/q-1}. \quad (A.7)$$

The latter condition is feasible for any $\beta > 2$.

By allowing the vanishing discontinuous part $\mu^n_{t,n}^{2}$ 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 3.1.** Suppose that Equations (A.1)–(A.6) hold, and let $p \in [1, 4]$. Suppose that $t_n \ll n$, and that $L_n$ satisfies $L_n \gg \log(n)^{1+a}$ for some $a \in (0, 1)$, and that $L_n \ll n^{2/q-1}, L_n \ll n^{1/2}$. Then, as $n \to \infty$,

$$\sqrt{n} \left( M_n(u) - \frac{1}{n} \sum_{t=a+n}^{\lfloor nu \rfloor} f(\mu^n_t) \right) \Rightarrow B \left( \int_0^u Df(\mu_t) \Sigma(v) Df(\mu_t)^T dv \right) = M(u),$$

$$\Sigma(v) = \sum_{h=-\infty}^{\infty} \text{cov}(G(v, \epsilon_h), G(v, \epsilon_{h,b})) \in \mathbb{R}^{d \times d},$$

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

$$\sqrt{n} \left( M_n(u) - \int_0^u f(\mu^n_t) dv \right) \Rightarrow M(u).$$

Just as in Theorem 2.1, the Cramer-Wold device may be used to extend Theorem 3.1 to the multivariate setting where $f$ takes values in $\mathbb{R}^d$. The functional weak convergence then holds in the product space $D[0, 1]^{d^2}$. If $p < 2$ and $n \ll n^1$, then we may alternatively choose $\int_0^u f(\mu^n_t) dv$ as centering term, such that $\sqrt{n} [M_n(u) - \int_0^u f(\mu^n_t) dv]$ has the asymptotic distribution given in Theorem 3.1.

**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 $I^n_{a,n}$ grows polynomially with $L$, that is, $I^n_{a,n} = \ldots$
In contrast to Theorem 2.1, the central limit theorem 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 \( p < 4 \), that is, 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 \( I_n^2 \approx \alpha_p(1/\sqrt{n}) \) if \( p < 4 \), see Lemma C.5 in the supplementary material. On the other hand, the criticality of \( p = 2 \) occurs because \( I_n^2 \leq \frac{1}{n} \sum_{i=1}^{n} f(\mu_\nu^n) - f(\mu_\nu) \) $\Rightarrow$ $\mathcal{N}(0, Df(\mu_\nu) T \Sigma(\mu_\nu) Df(\mu_\nu))$. (6)

If our model is indeed stationary, then Theorem 3.1 shows that \( M_n(1) \) has the same asymptotic distribution as Equation (6). In this sense, accounting for the nonstationarity does not increase the asymptotic variance.

To perform feasible inference based on Theorem 3.1, we need to handle the unknown asymptotic variance process. A consistent estimator may be constructed via blocked subsampling, similar to the suggestion of Carlin 

Theorem 3.2. Let the conditions of Theorem 3.1 hold for some \( q > 4 \), and \( p \in [1,4] \). Choose some \( b_n \to \infty \) such that \( b_n \leq n^{-2/\max(p,2)} \), then, as \( n \to \infty \),

\[
Q_n(u) = \frac{1}{n} \sum_{t=L_n}^{n} \frac{1}{b_n} \left[ Df(\mu_{t-L_n}) \left( \sum_{i=1}^{b_n} (X_{t+i-L_n} - \mu_{t-L_n}) \right) \right]^2 \to Q(u) = \int_0^u Df(\mu_\nu) \Sigma(\mu_\nu) Df(\mu_\nu) \, dv.
\]

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

Theorem 3.2 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 \( b_n \leq n^{-1/2} \) 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 3.3. Let the conditions of Theorem 3.2 hold for some \( q > 4 \). Let \( Y_i \sim \mathcal{N}(0,1) \) be iid standard normal random variables, independent of the \( \epsilon_i \), and define the process

\[
\hat{M}_n(u) = \frac{1}{\sqrt{n}} \sum_{t=L_n}^{n} Y_t \left[ \frac{1}{b_n} \sum_{i=1}^{b_n} Df(\mu_{t-L_n})(X_{t+i-L_n} - \mu_{t-L_n}) \right],
\]

Then the conditional distribution of \( \hat{M}_n \) given \( X_n = (X_1, \ldots, X_n) \) converges weakly in the Skorokhod space to \( M(u) \) in probability, where \( M(u) \) is the limit process from Theorem 3.1.

It can also be shown that the bootstrap consistency of Theorem 3.3 holds under weaker rate constraints on \( \mu_{t-L_n} \), replacing the rate \( o(\sqrt{n}) \) by \( o(n^{1-k}) \) for some \( k > 0 \). However, a smaller value of \( k \) requires stronger conditions on \( b_n \), see Theorem C.7 in the appendix. The local smoother \( \hat{M}_{n,W} \) 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 used 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 \( F_n(u) \) is that the estimator \( M_n(u) \) may be used to test for change-points. Our framework lends itself to test the hypothesis

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

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

\[
T_n^a = \sup_{u \in [u_0,1]} |T_n(u)|, \quad \text{where} \quad T_n(u) = M_n(u) - \frac{u - u_n}{1 - u_n} M_n(1), \quad \text{and} \quad u_n = \frac{\tau_n + L_n - 1}{n}.
\]

The main result Theorem 3.1 yields that \( T_n(u) \to F(u) - u F(1) \) as \( n \to \infty \), which is identically zero if the null hypothesis holds. In this case, Theorem 3.1 yields that

\[
\sqrt{n} T_n^a \Rightarrow M(u) - u M(1), \quad \text{and} \quad \sqrt{n} T_n^a \Rightarrow \mathcal{T}^a = \sup_{u \in [0,1]} |M(u) - u M(1)|,
\]

where \( M(u) \) is the Gaussian limit process. The limit distribution \( T^n \) may be approximated via the bootstrap procedure outlined in Theorem 3.3, that is, by sampling the random variable \( T_n^B = \sup_{u \in [0,1]} |\hat{M}_n(u) - u \hat{M}_n(1)| \), so that \( T_n^B \to \mathcal{T}^a \) by virtue of Theorem 3.3. In particular, denote by \( t_{a} \) the \( 1-\alpha \) quantile of \( T^n \), and by \( t_{a,n} \), the \( 1-\alpha \) quantile of \( T_n^B \). In practice, the quantile \( t_{a,n} \) may be approximated up to arbitrary precision by sampling from the conditional distribution \( \hat{T}_n^B \). The corresponding test procedure may then be formulated as follows.
Proposition 4.1. Let the conditions of Theorem 3.3 hold, and denote by \( t_{a,n} \) the \( 1 - \alpha \) quantile of the conditional distribution \( T_n^a | X_{n} \). If the null hypothesis (7) holds, and if \( \text{var}(M(u_0)) > 0 \) for some \( u_0 \in (0, 1) \), then

\[
\lim_{n \to \infty} P \{ T_{n}^a > t_{a,n} \} = \alpha, \quad n \to \infty.
\]

Hence, rejecting \( H_0 \) if the test statistic \( T_n^a \) exceeds the critical value \( t_{a,n} \) 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, for example, the MOSUM statistic introduced by Bauer and Hackl (1978), see also Chu, Hornik, and Kuan (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 \( f(\mu_n^u) \) to be non-constant, it is necessary that the local moment \( \mu_n^u \) changes. It is thus tempting to instead test the null hypothesis \( H^*_0 : \mu_n^u \equiv \mu_0^u \), 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 \( f(\mu_n^u) \) 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, Wu, and Zhou (2019). Furthermore, Schmidt et al. (2020) tested for homoscedasticity with a nonconstant mean function. A related approach was presented by Demetrescu and Wied (2018). By design, our test is only sensitive to changes in the quantity \( f(\mu_n^u) \).

Another type of nuisance change might occur in the parameters which are not explicitly described by the local moment function \( \mu_n^u \). 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 \( \mu_n^u \), but relevant for statistical inference, see Görecki, Horváth, and Kokoszka (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 designed a corresponding test for changes in the mean. The recent articles Vogt and Dette (2015), Dette, Wu, and Zhou (2019), and Cui, Levine, and Zhou (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) and Shao and Zhang (2010) and the discussion therein. In contrast, our bootstrap procedure is consistent under the alternative where \( f(\mu_n^u) \) is not constant and potentially discontinuous, see Theorem 3.3 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 3.1 hold, and assume furthermore that \( \mu_n^u = \mu_u + \frac{1}{\sqrt{n}} \delta_u \) 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

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

\[
\Delta(u) = \int_0^u Df(\mu_\beta)\delta_\beta \, dv - u \int_0^1 Df(\mu_\beta)\delta_\beta \, dv,
\]

where \( T(u) = M(u) - uM(1) \), and \( M(u) \) is the limit process from Theorem 3.1.

Note that under the local alternative of Proposition 4.2, the simple estimator \( \hat{\mu}_n^c \) still satisfies Equations (A.5) and (A.6) if \( u \mapsto \mu_u \) is smooth, see Proposition 3.1. Hence, Theorem 3.3 is still applicable so that the bootstrap is consistent, and the CUSUM test with bootstrapped critical values has nontrivial power against local alternatives in direction \( \delta_u \), given that \( \Delta(u) \neq 0 \). 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, Gombay, and Horváth (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, exampl. 3). The case of a nonparametric mean function is investigated by Li and Zhao (2013), and multiple change-points are studied by Preuss, Puchstein, and Dette (2015). The nonstationary case is investigated by Killick, Eckley, and Jonathan (2013), although without a rigorous analysis of the Type I error.

Alternatively, in the same univariate setting, Dette, Wu, and Zhou’s (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 nonconstant and estimate it nonparametrically, in order to standardize the observations. Furthermore, Dette, Wu, and Zhou (2019) studied 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 \( Y_{t,n} = G_n(\varepsilon_t, \varepsilon_t) = (X_{t,n}, X_{t-h,n}, X_{t-2h,n}, Y_{t,n}, X_{t-3h,n}) \) for \( t = 1, \ldots, n \), assuming for simplicity that \( X_{t,n} = G_n(0, \varepsilon_t) \) for \( r \leq 0 \). Now set \( f : \mathbb{R}^3 \to \mathbb{R}, x \mapsto (x_3 - x_1x_2)/\sqrt{(x_3 - x_1^2)(x_4 - x_2^2)} \).
so that \( \text{cor}(X_{t,n}, X_{t-h,n}) = f(\mathbb{E}Y_{t,n}) = f(\mu_{Y_{t,n}}^*) \). 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 \( K \subset \{ x \in \mathbb{R}^2 : (x_1 - x_2^2) > 0, (x_4 - x_2^2) > 0 \} \). Thus, if \( \text{var}(X_{t,n}) > 2\theta > 0 \), then Equation (A.4) holds for \( \delta > 0 \) as well. We may thus construct \( M_n(u) \) based on this time series \( Y_t \) and function \( f \) to obtain an estimator for the integrated autocorrelation \( F_n(u) = \int_0^u \text{cor} (G_n(v, \epsilon_0), G_n(v, \epsilon_{-h})) \, dv \). The corresponding CUSUM statistic satisfies Equation (8).

The resulting CUSUM statistic is similar to the statistic suggested by Dette, Wu, and Zhou (2019). However, our framework allows for many potential choices of \( \mu_{Y_{t,n}} \), while Dette, Wu, and Zhou (2019) only considered 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 Equation (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 \( \text{Skew}(X) = \mathbb{E}(X - \mathbb{E}X)^3/\text{var}(X)^{3/2} \) 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 supplementary material, 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

\[
X_{t,n} = a(\frac{1}{n})X_{t-1,n} + \sigma(\frac{1}{n})\eta_{t}(\frac{1}{n}).
\]

The innovations \( \eta_{t}(\frac{1}{n}) \) are chosen as independent, zero-mean random variables having a symmetrized Gamma distribution with shape parameter \( \alpha(\frac{1}{n}) \), standardized to unit variance. We use

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

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

\[
a_0(u) = 0.2, \quad a_1(u) = 0.2 + \frac{u}{2}, \quad a_2(u) = 0.2 + \frac{u}{4}.\]

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 \( \mu(\hat{\mu}_{Y_{t,n}}^*) \).

As described in Remark 2, the lag parameter \( L \) should be chosen just big enough 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

\[
\Lambda(k) = \sum_{i=1}^{n-L} ||\mu_{i+L_n} - X_{i+L_n,n}||^2,
\]

where \( \mu_{i+L_n} \) denotes the local average with bandwidth \( k \). In our simulations, we consider bandwidths from the interval \( k \in [0.75 \cdot n^{0.35}, n^{0.75}] \). Then, a natural choice for the offset is \( t_n = k_n \). 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 3.3. 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 \( a = a_1 \), 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 of \( L_n \). The power values correspond to the alternatives \( H_1 \) based on \( a_1 \), and \( H_2 \) based on \( a_2 \). We find that our test is rather conservative, that is, 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 \( L_n = \left\lceil \frac{1}{10} \log(n)^2 \right\rceil \) 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 autocorrelation proposed by Dette, Wu, and Zhou.
Table 1. Size and power of the bootstrap-based CUSUM test for constant autocorrelation, with the nominal level 10%.

|        | $L_n = [\log(n)^{1/2}]$ | $L_n = [\log(n)^{1/2}]$ | $L_n = [\log(n)^{1/2}]$ | $L_n = [\log(n)^{1/2}]$ | $L_n = [\log(n)^{1/2}]$ | DWZ |
|--------|------------------------|------------------------|------------------------|------------------------|------------------------|-----|
|        | $H_0$ | $H_1$ | $H_2$ | $H_0$ | $H_1$ | $H_2$ | $H_0$ | $H_1$ | $H_2$ | $H_0$ | $H_1$ | $H_2$ |
| $n = 100$ | 0.26 | 0.19 | 0.28 | 0.04 | 0.03 | 0.05 | 0.04 | 0.03 | 0.04 | 0.05 | 0.04 | 0.04 |
| 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 |
| 1000    | 0.001 | 0.18 | 0.01 | 0.02 | 0.55 | 0.04 | 0.04 | 0.81 | 0.09 | 0.06 | 0.89 | 0.11 |
| 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 |
| 10,000  | 0.03 | 1.00 | 0.56 | 0.06 | 1.00 | 0.68 | 0.08 | 1.00 | 0.74 | 0.09 | 1.00 | 0.76 |

NOTE: Reported values are based on 5000 independent samples of the test statistic.

Figure 1. Distribution of bootstrap-based $p$-values for the CUSUM test for constant autocorrelation, under the null hypothesis, with lag parameter $L_n = [\log(n)/10]$. 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.

Table 2. Mean absolute error and bias of $M_n(1)$ and $\Hat{M}_n(1)$ as estimators of $\int_0^1 a_i(u) \, du, i = 0, 1, 2$, with $L_n = [\log(n)^{2/10}]$ for the linearized estimator.

|        | $n = 100$ | $n = 500$ | $n = 1000$ | $n = 5000$ | $n = 10,000$ |
|--------|-----------|-----------|-----------|-----------|-----------|
|        | $M_n$     | $\Hat{M}_n$ | $M_n$     | $\Hat{M}_n$ | $M_n$     | $\Hat{M}_n$ | $M_n$     | $\Hat{M}_n$ | $M_n$     | $\Hat{M}_n$ | $M_n$     | $\Hat{M}_n$ |
| $H_0$  | bias 0.172 | 0.120 | 0.053 | 0.058 | 0.031 | 0.039 | 0.012 | 0.018 | 0.008 | 0.013 |
|        | error 0.242 | 0.197 | 0.061 | 0.093 | 0.035 | 0.066 | 0.011 | 0.031 | 0.008 | 0.023 |
| $H_1$  | bias 0.163 | 0.197 | 0.039 | 0.093 | 0.020 | 0.066 | 0.002 | 0.031 | 0.001 | 0.023 |
|        | error 0.189 | 0.134 | 0.053 | 0.062 | 0.032 | 0.044 | 0.012 | 0.020 | 0.008 | 0.014 |
| $H_2$  | bias 0.084 | 0.131 | 0.025 | 0.060 | 0.012 | 0.043 | 0.000 | 0.019 | 0.002 | 0.014 |

NOTE: All values are based on 5000 simulations.

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, Wu, and Zhou (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 $\Hat{M}_n(1)$. 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.

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 to 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 nonzero, especially at high sampling frequencies (Hansen and Lunde 2006). This dependence structure is typically attributed to the microstructure of the market, for example, rounding effects or bid–ask rebounds. Recently, Andersen et al. (2021) studied the intraday returns of the NASDAQ100 stocks and found evidence for autocorrelation which is not only nonzero, but also nonconstant. Using the framework laid out in Section 4.1, 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 $\Hat{\mu}_n^{NW}$ and choose its bandwidth via cross-validation as in Section 5, with $L_n = [\log(n)^{2/10}]$.

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 nonconstant autocorrelation. To test this hypothesis rigorously, we perform the CUSUM test suggested in Section 4. The
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 for 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.

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 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. (2021) that the serial correlation of intraday log returns is non-constant for the present dataset. In contrast to Andersen et al. (2021), 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 $d_t$ and $\tilde{d}_t$.

Many models for asset returns at very high frequencies describe the observed price as the sum of two latent components: the fundamental price $p_t^f$, and the so-called microstructure noise $\eta_t$, such that $p_t = p_t^f + \eta_t$. The fundamental price is typically modeled as a semimartingale, and inference
for this component needs to account for the microstructure effects, see, for example, Jacob et al. (2009). The microstructure noise is typically assumed to be independent, or dependent but stationary (Hansen and Lund 2006; Aït-Sahalia, Mykland, and Zhang 2011). Nonstationary and dependent noise is considered by Jacob, Li, and Zheng (2017), but such that the autocorrelation of the microstructure is constant. The approach we pursue in the present article 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.

Supplementary Material

The appendices A and B containing additional examples and simulation results of change point problems, and Appendix C containing all technical proofs, are available as a digital supplement. The R code used for the simulations, and the data used for the empirical example are also available as a supplement.

Acknowledgments

The author gratefully acknowledges the comments of the anonymous reviewers, which led to substantial improvements of this article.

References

Aït-Sahalia, Y., and Jacod, J. (2014), *High-Frequency Financial Econometrics*, Princeton, NJ: Princeton University Press. [1012]
Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011), “Ultra High Frequency Volatility Estimation With Dependent Microstructure Noise,” *Journal of Econometrics*, 160, 160–175. [1021]
Andersen, T. G., Archakov, I., Cebiroglu, G., and Hautsch, N. (2021), “Local Mispricing and Microstructural Noise: A Parametric Perspective,” *Journal of Econometrics*, forthcoming. [https://doi.org/10.1016/j.jeconom.2021.06.006] [1019,1020]
Aue, A., and Horváth, L. (2013), “Structural Breaks in Time Series,” *Journal of Time Series Analysis*, 34, 1–16. [1012]
Bauer, P., and Hackl, P. (1978), “The Use of MOSUMS for Quality Control,” *Technometrics*, 20, 431. [1017]
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, 2044–2063. [1012,1017]
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, 381–393. [1012]
Billingsley, P. (1999), *Convergence of Probability Measures*, New York: Wiley. [1013]
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, 1171–1179. [1016]
Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995), “MOSUM Tests for Parameter Constancy,” *Biometrika*, 82, 603–617. [1017]
Cui, Y., Levine, M., and Zhou, Z. (2020), “Estimation and Inference of Time-Varying Auto-Covariance under Complex Trend: A Difference-Based Approach,” arXiv: 2003.05006. [1011,1017]
Dahlhaus, R. (1997), “Fitting Time Series Models to Nonstationary Processes,” *Annals of Statistics*, 25, 1–37. [1011]
— (2009), “Local Inference for Locally Stationary Time Series Based on the Empirical Spectral Measure,” *Journal of Econometrics*, 151, 101–112. [1012]
Dahlhaus, R., and Polonik, W. (2009), “Empirical Spectral Processes for Locally Stationary Time Series,” *Bernoulli*, 15, 1–39. [1013,1014]
Dahlhaus, R., and Richter, S. (2019), “Adaptation for Nonparametric Estimators of Locally Stationary Processes,” arXiv: 1902.10381. [1011]
Dahlhaus, R., Richter, S., and Wu, W. B. (2019), “Towards a General Theory for Nonlinear Locally Stationary Processes,” *Bernoulli*, 25, 1013–1044. [1011,1013]
Demetrescu, M., and Wied, D. (2018), “Testing for Constant Correlation of Filtered Series Under Structural Change,” *Econometrica Journal*, 22, 10–33. [1017]
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, 1361–1377. [1012]
Dette, H., and Wu, W. (2019), “Detecting Relevant Changes in the Mean of Nonstationary Processes—A Mass Excess Approach,” *The Annals of Statistics*, 47, 3578–3608. [1013]
Dette, H., Wu, W., and Zhou, Z. (2019), “Change Point Analysis of Correlation in Non-Stationary Time Series,” *Statistica Sinica*, 29, 611–643. [1013,1017,1018,1019]
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, 773–781. [1012]
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, 2412–2450. [1014]
Görecki, T., Horváth, L., and Kokoszka, P. (2018), “Change Point Detection in Heteroscedastic Time Series,” *Econometrics and Statistics*, 7, 63–88. [1017]
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. [1012]
Grenier, Y. (1983), “Time-Dependent ARMA Modeling of Nonstationary Signals,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31, 899–911. [1014]
Hall, P., and Marron, J. S. (1987), “Estimation of Integrated Squared Derivatives,” *Statistics and Probability Letters*, 6, 109–115. [1012]
Hansen, P. R., and Lunde, A. (2006), “Realized Variance and Market Microstructure Noise,” *Journal of Business & Economic Statistics*, 24, 127–161. [1019,1021]
Horváth, L. (1995), “Detecting Changes in Linear Regressions,” *Statistics*, 26, 189–208. [1012]
Huang, L. S. and Jianqing, F. A. (1999), “Nonparametric Estimation of Quadratic Regression Functions,” *Bernoulli*, 5, 927–949. [1012]
Jacob, 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, 2249 – 2276. [1021]
Jacob, J., Li, Y., and Zheng, X. (2017), “Statistical Properties of Microstructure Noise,” *Econometrica*, 85, 1133–1174. [1021]
Jacob, J., and Rosenbaum, M. (2013), “Quarticity and Other Functionals of Volatility: Efficient Estimation,” *The Annals of Statistics*, 41, 1462–1484. [1012,1014]
Jacob, J., and Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes*, Volume 288 of *Grundlehren der mathematischen Wissenschaften*. Berlin: Springer. [1013]
Juhl, T., and Xiao, Z. (2009), “Tests for Changing Mean With Monotonic Power,” *Journal of Econometrics*, 148, 14–24. [1017]
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, 1167–1183. [1012,1017]
Li, X., and Zhao, Z. (2013), “Testing for Changes in Autocovariances of Nonparametric Time Series Models,” *Journal of Statistical Planning and Inference*, 143, 237–250. [1017]
Moulines, E., Priouret, P., and Roueff, F. (2005), “On Recursive Estimation for Time Varying Autoregressive Processes,” *The Annals of Statistics*, 33, 2610 – 2654. [1014]
Page, E. (1954), “Continuous Inspection Schemes,” *Biometrika*, 41, 100–115. [1012]
— (1955), “A Test for a Change in a Parameter Occurring at an Unknown Point,” *Biometrika*, 42, 523–527. [1012]
Pešta, M., and Wendler, M. (2020), “Nuisance-Parameter-Free Changepoint Detection in Non-Stationary Series,” TEST, 29, 379–408. [1017]

Potiron, Y., and Mykland, P. (2020), “Local Parametric Estimation in High Frequency Data,” Journal of Business & Economic Statistics, 38, 679–692. [1011,1012,1014]

Preuss, P., Puchstein, R., and Dette, H. (2015), “Detection of Multiple Structural Breaks in Multivariate Time Series,” Journal of the American Statistical Association, 110, 654–668. [1012,1017]

Schick, A., and Wefelmeyer, W. (2004), “Root n Consistent Density Estimators for Sums of Independent Random Variables,” Journal of Nonparametric Statistics, 16, 925–935. [1012]

Schmid, S., Wornowizki, M., Fried, R., and Dehling, H. (2020), “An Asymptotic Test for Constancy of the Variance under Short-Range Dependence,” arXiv:2002.10178. [1017]

Shao, X., and Zhang, X. (2010), “Testing for Change Points in Time Series,” Journal of the American Statistical Association, 105, 1228–1240. [1012,1017]

Steland, A. (2020), “Testing and Estimating Change-Points in the Covariance Matrix of a High-Dimensional Time Series,” Journal of Multivariate Analysis, 177, 104582. [1017]

Subba Rao, T. (1970), “The Fitting of Non-Stationary Time-Series Models With Time-Dependent Parameters,” Journal of the Royal Statistical Society, Series B, 32, 312–322. [1014]

Truquet, L. (2019), “Local Stationarity and Time-Inhomogeneous Markov Chains,” The Annals of Statistics, 47, 2023–2050. [1011]

Vogt, M., and Dette, H. (2015), “Detecting Gradual Changes in Locally Stationary Processes,” Annals of Statistics, 43, 713–740. [1017]

Wu, W. B. (2005), “Nonlinear System Theory: Another Look at Dependence,” Proceedings of the National Academy of Sciences, 102, 14150–14154. [1013]

Wu, W. B., and Zhou, Z. (2011), “Gaussian Approximations for Non-Stationary Multiple Time Series,” Statistica Sinica, 21, 1397–1413. [1011]

Zhou, Z. (2013), “Heteroscedasticity and Autocorrelation Robust Structural Change Detection,” Journal of the American Statistical Association, 108, 726–740. [1011,1013,1016,1017]

Zhou, Z., and Wu, W. B. (2009), “Local Linear Quantile Estimation for Nonstationary Time Series,” Annals of Statistics, 37, 2696–2729. [1011]