Towards a general theory for nonlinear locally stationary processes

RAINER DAHLHAUS[1,*] STEFAN RICHTER[1,**] and WEI BIAO WU[2]

[1]*Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany. E-mail:* [*]*dahlhaus@statlab.uni-heidelberg.de;* [**]*stefan.richter@iwr.uni-heidelberg.de*
[2]*Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA. E-mail:* *wbwu@galton.uchicago.edu*

In this paper, some general theory is presented for locally stationary processes based on the stationary approximation and the stationary derivative. Laws of large numbers, central limit theorems as well as deterministic and stochastic bias expansions are proved for processes obeying an expansion in terms of the stationary approximation and derivative. In addition it is shown that this applies to some general nonlinear non-stationary Markov-models. In addition the results are applied to derive the asymptotic properties of maximum likelihood estimates of parameter curves in such models.

Keywords: derivative processes; non-stationary processes

1. Introduction

One of the challenges in statistics for stochastic processes is always to develop a general theory that goes beyond the investigation of specific models. An elegant example are stationary Gaussian time series or linear time series which are identifiable from the covariance structure or the spectral density of the process. This leads to such powerful tools as the Whittle-likelihood and quite general asymptotic results on statistical inference. This linear theory has been extended to locally stationary processes (cf. [2,3,6], for an overview see [4], Chapter 5) leading again to a general framework in which problems such as bootstrap methods for locally stationary processes (cf. [12, 21]), testing problems (cf. [17,22]) long memory models (cf. [16,19]) or dynamic non-stationary factor models (cf. [9,15]) can be considered.

In the nonlinear case, the situation is more challenging since there is no natural framework similar to the linear Gaussian case. A general theory, however, has been introduced in [29] for Bernoulli shift processes in combination with the functional dependence measure – an important example being Markov processes (and also linear processes). By using this calculus, it is possible to transfer a large number of results from the i.i.d. case to such processes (for instance, *M*-estimation [30], empirical process theory [31], high-dimensional covariance matrix estimation [1] – see also the overview in [32]).

The functional dependence measure for Bernoulli shift processes can also be extended to locally stationary processes (cf. [33]) leading to a general framework for nonlinear locally stationary processes. Within this framework for example [36] and [28] discuss quantile regression, [34] inference for weighted *V*-statistics and [35] nonparametric regression for locally stationary processes.

Another general concept for locally stationary processes is the use of stationary approximations and derivative processes introduced in the context of time varying ARCH-processes in [8]. The concept has been investigated further in [24] in the context of random coefficient models. [27] uses the stationary approximation for a definition of local stationarity. The concept has also been used for diffusion processes in [11].

In this paper, the general theory for nonlinear locally stationary processes will be developed further. Our contribution is twofold: First, we consider in Sections 2 and 3 processes which admit an expansion in terms of the stationary approximation and the derivative process, and prove several asymptotic results for such processes. We then consider in Section 4 a general Markov-structured locally stationary process and show that it fulfills such an expansion. As a consequence, all results of Sections 2 and 3 immediately can be applied for such processes. In addition, we use in Section 5 the stationary approximation and the derivative process to derive the asymptotic theory for maximum likelihood estimates.

More precisely, we use in Section 2 the stationary approximation to prove global and local laws of large numbers and a central limit theorem which hold under minimal moment assumptions on the process. The proofs make use of the asymptotic theory for sums of stationary sequences. In Section 3, we use the differential calculus connected to derivative processes to derive deterministic and stochastic bias expansions of localized sums. In addition, we show accurate error estimations of the Wigner-Ville spectrum and the distribution function of the process.

In Section 4, we consider a class of Markov processes and prove that they satisfy the expansion from Section 3 in terms of the stationary approximation and the derivative process. A difficult part of the proof is the existence of a continuous modification of the stationary approximation and the proof that the derivative process can be obtained as the solution of a functional equation. These results are the prerequisite to apply the functional calculus used in Sections 2 and 3 and in Section 5 thereafter. We also prove that the functional dependence measure of such processes decays exponentially. Locally stationary Markov processes have recently also been investigated with different type of results in [25].

In Section 5, we use these results to investigate nonparametric maximum likelihood estimation of parameter curves in locally stationary processes. Concluding remarks are given in Section 6. Some proofs are postponed to the Supplementary Material [7], Supplement A.

2. General asymptotic results for locally stationary processes

In this and the next section, we restrict ourselves to the basic idea of local stationarity, namely that the nonstationary process $X_{t,n}$ can be approximated locally by a stationary process $\tilde{X}_t(u)$ in some neighborhood of u, that is for those t where $|t/n - u|$ is small. An even better approximation can usually be achieved by using the derivative process $\frac{\partial \tilde{X}_t(u)}{\partial u}$ leading heuristically to

$$X_{t,n} \approx \tilde{X}_t\left(\frac{t}{n}\right) \approx \tilde{X}_t(u_0) + \left(\frac{t}{n} - u_0\right) \frac{\partial \tilde{X}_t(u)}{\partial u}\bigg|_{u=u_0} + \text{remainder} \qquad (1)$$

(or higher order approximations by using higher order derivative processes). An example are the processes defined below in (3) and (4) which are investigated in detail in Section 4. In this and the next section, we explore what kind of results can be obtained just based on this framework where

in this section we just use the stochastic approximation $\tilde{X}_t(u)$ (to derive law of large numbers and central limit theorems) while in the next section we use in addition the derivative process $\frac{\partial \tilde{X}_t(u)}{\partial u}$ (to derive deterministic and stochastic bias approximations among other results). For a better understanding, we make some comments about the situation in advance:

(1) In Assumption 2.1(S2) and (S3) we assume that $\tilde{X}_t(u)$ is almost surely continuous or continuously differentiable in u respectively. This is a strong assumption since $\tilde{X}_t(u)$ is initially defined in most cases pointwise in u as in (4). The existence of continuous or continuously differentiable versions of $\tilde{X}_t(u)$ will be proved for (4) in Section 4.

We mention that almost sure differentiability could be replaced by the weaker assumption of differentiability in L^q (see Remark 4.3 and Proposition 3.3(b)). The stronger assumption of almost sure differentiability leads to some weaker assumptions in applications (see Remark 3.5) and has the advantage that several results can be obtained in a more straightforward way and that the presentation is simpler.

(2) It is obvious that for asymptotic results some form of mixing is needed in addition which we require in Assumption 2.3. We distinguish between mixing conditions on $X_{t,n}$ and the stationary approximation $\tilde{X}_t(u)$ since our aim is to show that most of the results can be obtained by only posing mixing assumptions on $\tilde{X}_t(u)$ which on the other side leads to an additional approximation error in the results and subsequently in come cases to stronger assumptions.

(3) One of the nice features of Assumptions 2.1 and 2.3 is that, provided they hold for the process $X_{t,n}$, they then automatically also hold for a large class of functionals $g(X_{t,n}, \ldots, X_{t-r+1,n})$ (under modified moment assumptions). Thus, all results for the process $X_{t,n}$ immediately transfer to $g(X_{t,n}, \ldots, X_{t-r+1,n})$. This is stated in Proposition 2.5 below.

As usual we are working in the infill asymptotic framework with rescaled time $t/n \in [0, 1]$ where n denotes the number of observations. We now assume:

Assumption 2.1 (Stationary approximation). Let $q > 0$ and $\|W\|_q := (\mathbb{E}|W|^q)^{1/q}$. Let $X_{t,n}$, $t = 1, \ldots, n$ be a triangular array of stochastic processes. For each $u \in [0, 1]$, let $\tilde{X}_t(u)$ be a stationary and ergodic process such that the following holds.

(S1) $\sup_{u \in [0,1]} \|\tilde{X}_t(u)\|_q < \infty$. There exists $1 \geq \alpha > 0$, $C_B > 0$ such that uniformly in $t = 1, \ldots, n$ and $u, v \in [0, 1]$,

$$\left\| \tilde{X}_t(u) - \tilde{X}_t(v) \right\|_q \leq C_B |u - v|^\alpha, \qquad \left\| X_{t,n} - \tilde{X}_t\left(\frac{t}{n}\right) \right\|_q \leq C_B n^{-\alpha}. \qquad (2)$$

(S2) $u \mapsto \tilde{X}_t(u)$ is a.s. continuous for all $t \in \mathbb{Z}$ and $\| \sup_{u \in [0,1]} |\tilde{X}_t(u)| \|_q < \infty$.

(S3) $\alpha = 1$ and $u \mapsto \tilde{X}_t(u)$ is a.s. continuously differentiable for all $t \in \mathbb{Z}$ and $\| \sup_{u \in [0,1]} |\partial_u \tilde{X}_t(u)| \|_q < \infty$.

(S1) allows us to replace $X_{t,n}$ by the stationary approximation $\tilde{X}_t(u)$ with rate $|t/n - u|^\alpha + n^{-\alpha}$. In many models and statistical applications, $\alpha = 1$. In Section 4 (cf. Corollary 4.9), we will show that Assumption 2.1 is fulfilled for example for processes $X_{t,n}$ defined by the recursion

$$X_{t,n} = G_{\varepsilon_t}\left(X_{t-1,n}, \ldots, X_{t-p,n}, \frac{t}{n} \vee 0 \right), \qquad t \leq n, \qquad (3)$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i.i.d. random variables, $G : \mathbb{R} \times \mathbb{R}^p \times [0, 1] \to \mathbb{R}$ and $a \vee b := \max\{a, b\}$. Here, ε_t take the role of i.i.d. innovations which enter the process in the tth step. For $u \in [0, 1]$, the *stationary approximation* $\tilde{X}_t(u)$, $t \in \mathbb{Z}$, in this case is given by the recursion

$$\tilde{X}_t(u) = G_{\varepsilon_t}\big(\tilde{X}_{t-1}(u), \ldots, \tilde{X}_{t-p}(u), u\big), \qquad t \in \mathbb{Z}, \tag{4}$$

(or an a.s. continuous modification). We first give some examples which are covered by our results. These include in particular several classical parametric time series models where the constant parameters have been replaced by time-dependent parameter curves.

Example 2.2.

(i) The tvAR(p) process: Given parameter curves $a_i, \sigma : [0, 1] \to \mathbb{R}$ $(i = 1, \ldots, p)$,

$$X_{t,n} = a_1\left(\frac{t}{n}\right)X_{t-1,n} + \cdots + a_p\left(\frac{t}{n}\right)X_{t-p,n} + \sigma\left(\frac{t}{n}\right)\varepsilon_t.$$

(ii) The tvARCH(p) process (cf. [8]): Given parameter curves $a_i : [0, 1] \to \mathbb{R}$ $(i = 0, \ldots, p)$,

$$X_{t,n} = \left(a_0\left(\frac{t}{n}\right) + a_1\left(\frac{t}{n}\right)X_{t-1,n}^2 + \cdots + a_p\left(\frac{t}{n}\right)X_{t-p,n}^2\right)^{1/2}\varepsilon_t.$$

(iii) The tvTAR(1) process (cf. [36]): Given parameter curves $a_1, a_2 : [0, 1] \to \mathbb{R}$, define

$$X_{t,n} = a_1\left(\frac{t}{n}\right)X_{t-1,n}^+ + a_2\left(\frac{t}{n}\right)X_{t-1,n}^- + \varepsilon_t,$$

where $x^+ := \max\{x, 0\}$ and $x^- := \max\{-x, 0\}$.

(iv) The time-varying random coefficient model (cf. [24]): With some parameter functions $a_i(\cdot)$, $i = 0, \ldots, p$,

$$X_{t,n} = a_0\left(\varepsilon_t, \frac{t}{n}\right) + a_1\left(\varepsilon_t, \frac{t}{n}\right)X_{t-1,n} + \cdots + a_p\left(\varepsilon_t, \frac{t}{n}\right)X_{t-p,n}.$$

We now prove laws of large numbers and a central limit theorem. To specify the necessary mixing conditions, we use the uniform functional dependence measure (cf. [13]). Let ε_t, $t \in \mathbb{Z}$ be a sequence of i.i.d. random variables. For $t \geq 0$, let $\mathcal{F}_t := (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and $\mathcal{F}_t^{*(t-k)} := (\varepsilon_t, \ldots, \varepsilon_{t-k+1}, \varepsilon_{t-k}^*, \varepsilon_{t-k-1}, \varepsilon_{t-k-2}, \ldots)$, where ε_{t-k}^* is a random variable which has the same distribution as ε_1 and is independent of all ε_t, $t \in \mathbb{Z}$. For a process $W_t = H_t(\mathcal{F}_t) \in L^q$ with deterministic $H_t : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ define $W_t^{*(t-k)} := H_t(\mathcal{F}_t^{*(t-k)})$ and the uniform functional dependence measure

$$\delta_q^W(k) := \sup_{t\in\mathbb{Z}}\big\|W_t - W_t^{*(t-k)}\big\|_q. \tag{5}$$

If Y_t is stationary, (5) reduces to the form $\delta_q^Y(k) = \|Y_k - Y_k^{*0}\|_q$.

Assumption 2.3 (Dependence measure). For some $q > 0$, assume that

- (M1) (dependence measure of the stat. approximation) for each $u \in [0, 1]$, there exists a measurable function $H(u, \cdot)$ such that $\tilde{X}_t(u) = H(u, \mathcal{F}_t)$ and $\delta_q^{\tilde{X}}(k) := \sup_{u \in [0,1]} \delta_q^{\tilde{X}(u)}(k)$ fulfills $\Delta_{0,q}^{\tilde{X}} := \sum_{k=0}^{\infty} \delta_q^{\tilde{X}}(k) < \infty$.
- (M2) (dependence measure of the process) for each $t, n \in \mathbb{N}$, there exists a measurable function $H_{t,n}$ such that $X_{t,n} = H_{t,n}(\mathcal{F}_t)$ with $\Delta_{0,q}^{X} := \sum_{k=0}^{\infty} \sup_{n \in \mathbb{N}} \delta_q^{X_{\cdot,n}}(k) < \infty$.
- (M3) $\partial_u \tilde{X}_t(u) = \partial_u H(u, \mathcal{F}_t)$ and $\delta_q^{\partial \tilde{X}}(k) := \sup_{u \in [0,1]} \delta_q^{\partial_u \tilde{X}(u)}(k)$ is absolutely summable in the sense that $\Delta_{0,q}^{\partial \tilde{X}} := \sum_{k=0}^{\infty} \delta_q^{\partial \tilde{X}}(k) < \infty$.

In Section 4, we will show that these assumptions are also fulfilled for models which obey (3) and (4). Note that (M1) is a mixing condition on $\tilde{X}_t(u)$ (which in most cases is sufficient for asymptotic results) while (M2) is a mixing condition on $X_{t,n}$. In our results, (M1) or (M2) are assumed alternatively. In general our goal is to use mainly assumptions on the stationary approximations, that is, to use (M1) instead of (M2). This also allows $X_{t,n}$ to have a structure which is different from $H_{t,n}(\mathcal{F}_t)$, for instance, contamination with some random noise which is decreasing in n and is not produced by the innovations ε_i. Note that posing only (M1) forces us to replace $X_{t,n}$ in the proofs by its stationary approximation which naturally leads to an approximation error $n^{-\alpha}$. In many practical cases, we have $\alpha = 1$. It can be seen in our results that this implies negligibility of this error.

For the model $\tilde{X}_t(u) = H(u, \cdot)$, $X_{t,n} = H_{t,n}(\mathcal{F}_t)$ with (S1), then (M2) implies (M1) while the reverse implication is false. This can be seen as follows: For arbitrary $u \in [0, 1]$, choose $t_n \in \{1, \ldots, n\}$ with $|u - t_n/n| \leq n^{-\alpha}$. Then

$$
\begin{aligned}
\delta_q^{\tilde{X}(u)}(k) &= \left\| \tilde{X}_t(u) - \tilde{X}_t(u)^{*(t-k)} \right\|_q \\
&= \left\| \tilde{X}_t(t_n/n) - \tilde{X}_t(t_n/n)^{*(t_n-k)} \right\|_q + O(n^{-\alpha}) \\
&= \left\| X_{t_n,n} - X_{t_n,n}^{*(t_n-k)} \right\|_q + O(n^{-\alpha}) \leq \sup_{n \in \mathbb{N}} \delta_q^{X_{\cdot,n}}(k) + O(n^{-\alpha}),
\end{aligned}
$$

which by $n \to \infty$ implies (M1). If for instance for each fixed $u \in [0, 1]$, $\tilde{X}_t(u)$ is a simple AR(1) process $\tilde{X}_t(u) = \sum_{k=0}^{\infty} \alpha(u)^k \varepsilon_{t-k}$ with Lipschitz continuous $\alpha : [0, 1] \to (-1, 1)$ and $\|\varepsilon_i\|_q < \infty$, then it is easy to see that $X_{t,n} = \tilde{X}_t(t/n) + (\varepsilon_t + \cdots + \varepsilon_{t-n})n^{-2}$ satisfies (S1) with $q = 1$ and $\alpha = 1$, but $\delta_q^{X_{\cdot,n}}(k) = \sup_t |\theta(t/n)^k + n^{-2}|$ is not absolutely summable. There are more counterexamples, see the Supplementary Material [7], Supplement B.

Invariance property of the assumptions with respect to transformations

For some fixed $r \in \mathbb{N}$ define $Z_{t,n} := (X_{t,n}, \ldots, X_{t-r+1,n})'$ and $\tilde{Z}_t(u) := (\tilde{X}_t(u), \ldots, \tilde{X}_{t-r+1}(u))'$. We prove that $g(Z_{t,n})$ also fulfills Assumptions 2.1 and 2.3 for the following class of functions g.

Definition 2.4 (The class $\mathcal{L}_r(M, C)$). We say that a function $g : \mathbb{R}^r \to \mathbb{R}$ is in the class $\mathcal{L}_r(M, C)$ if $M \geq 0$ and

$$\sup_{y \neq y'} \frac{|g(y) - g(y')|}{|y - y'|_1 \cdot (1 + |y|_1^M + |y'|_1^M)} \leq C, \tag{6}$$

where $|y|_1 := \sum_{i=1}^r |y_i|$.

Proposition 2.5 (Invariance property of the stationary approximation). *Assume that $g \in \mathcal{L}_r(M, C)$ and $q > 0$.*

  (i) *If Assumption 2.1(S1), (S2) and Assumption 2.3(M1), (M2) are fulfilled for the process $X_{t,n}$ with $\tilde{q} = q \cdot (M + 1)$ and $1 \geq \alpha > 0$, then the same assumptions are fulfilled for the process $g(Z_{t,n})$ with q and α.*

Now let g in addition be continuously differentiable where the partial derivatives $\partial_j g$, $j = 1, \ldots, r$ fulfill $\partial_j g \in \mathcal{L}_r(M - 1, C')$ with some $M \geq 1$.

  (ii) *If Assumptions 2.1(S3) and 2.3(M1), (M3) are fulfilled for the process $X_{t,n}$ with $\tilde{q} = q \cdot (M + 1)$, then the same assumptions are fulfilled for the process $g(Z_{t,n})$ with q.*

The proof is immediate from Hoelder's inequality and therefore omitted. Let us mention that the condition $\partial_j g \in \mathcal{L}_r(M - 1, C')$ is only needed to prove the mixing condition 2.3(M3) since $\| \sup_{u \in [0,1]} |\partial_u g(\tilde{Z}_0(u))| \|_q < \infty$ can be shown by using $|\partial_j g(z)| \leq C(1 + |z|_1^M)$.

With slight changes, the statements of Proposition 2.5 can be extended to Hoelder continuous functions which fulfill

$$\sup_{y \neq y'} \frac{|g(y) - g(y')|}{|y - y'|_1^\beta \cdot (1 + |y|_1^M + |y'|_1^M)} \leq C, \tag{7}$$

with some $1 \geq \beta > 0$.

In view of the above proposition, *all* theorems formulated for $X_{t,n}$ in this and the next section are, under appropriate moment conditions, also valid for transformations $g(Z_{t,n})$ of $X_{t,n}$. An important example is the covariance operator $g : \mathbb{R}^r \to \mathbb{R}$, $g(x_1, \ldots, x_r) = x_1 x_r$ which leads to $g(Z_{t,n}) = X_{t,n} X_{t-r+1,n}$ and fulfills $g \in \mathcal{L}_r(1, 1)$.

Local and global laws of large numbers

The smoothness of $X_{t,n}$ in time direction can be used to obtain laws of large numbers by only assuming the existence of the first moment of $X_{t,n}$. The key step of the proof is to split the sum over $X_{t,n}$ into sums over smaller ranges of t where $X_{t,n}$ can be approximated by stationary processes. We will also provide results for localized sums. Usually, we will need the following.

Assumption 2.6 (Localizing kernel). $K : \mathbb{R} \to \mathbb{R}$ is a bounded function, i.e. with some $|K|_\infty > 0$ it holds that $\sup_{x \in \mathbb{R}} |K(x)| \leq |K|_\infty$, and of bounded variation B_K with compact support $[-\frac{1}{2}, \frac{1}{2}]$ satisfying $\int K \, dx = 1$. Let $K_b(x) := \frac{1}{b} K(\frac{x}{b})$.

The first part of the following theorem can be seen as a generalization of the ergodic theorem to non-stationary processes, while the second part provides uniform convergence rates if more than the first moment is available.

Theorem 2.7 (Law of large numbers). *Let $q = 1$ in (i), (ii) and $q > 1$ in (iii). Suppose that Assumption 2.1(S1) holds with some $1 \geq \alpha > 0$ and that Assumption 2.6 holds. Then we obtain for the process $X_{t,n}$ (or alternatively for the process $g(Z_{t,n})$ if Assumption 2.1(S1) is fulfilled for the process $X_{t,n}$ with $\tilde{q} = q \cdot (M + 1)$ instead of q) the following results:*

(i)
$$\frac{1}{n} \sum_{t=1}^{n} X_{t,n} \to \int_0^1 \mathbb{E}\tilde{X}_0(u)\,\mathrm{d}u \qquad in \ L^1.$$

(ii) *For each $u \in (0, 1)$*
$$\frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{t/n - u}{b}\right) \cdot X_{t,n} \to \mathbb{E}\tilde{X}_0(u) \qquad in \ L^1$$

as $n \to \infty$, $nb \to \infty$ and $b = b_n \to 0$.

(iii) *Additionally, suppose that Assumption 2.3(M1) holds with q, α. Then*
$$\left\| \sup_{u \in [0,1]} \left| \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{t/n - u}{b}\right) \cdot (X_{t,n} - \mathbb{E}X_{t,n}) \right| \right\|_q \tag{8}$$
$$\leq \frac{B_K p}{(q-1)^2} \Delta_{q,0}^{\tilde{X}} \cdot n^{1/q - 1} b^{-1} + 2 C_B B_K \cdot n^{-\alpha} b^{-1}.$$

If $q > 2$, then there exist constants C_1, C_2 not depending on n, b such that for all $x > 0$:
$$\mathbb{P}\left(\sup_{u \in [0,1]} \left| \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{t/n - u}{b}\right) \cdot (X_{t,n} - \mathbb{E}X_{t,n}) \right| > x \right)$$
$$\leq \frac{2 C_1 (B_K \Delta_{0,q}^{\tilde{X}})^q n^{1-q} b^{-q}}{(x/2)^q} + 8 G_{1-2/q}\left(\frac{C_2 n^{1/2} bx}{2 B_K \Delta_{0,q}^{\tilde{X}}} \right) \tag{9}$$
$$+ \frac{(2 B_K C_B)^q}{(x/2)^q} \cdot \left(n^{-\alpha} b^{-1} \right)^q,$$

with positive constants C_1, C_2 not depending on n, b and $G_\gamma(y) := \sum_{j=1}^{\infty} e^{-j^\gamma y^2}$ a Gaussian-like tail function. $G_\gamma(y)$ can be replaced by $\exp(-cy^2)$ for some $c > 0$.

Remark 2.8.

(i) The additional $O(n^{-\alpha} b^{-1})$ or $O((n^{-\alpha} b^{-1})^{\tilde{q}})$ terms in (8) and (9), respectively, can be omitted under Assumption 2.3(M2). In this case, one has to replace $\Delta_{0,\tilde{q}}^{\tilde{X}}$ by $\Delta_{0,\tilde{q}}^{X}$.

(ii) For $q > 1$, $b = o(n^{1-\frac{1}{q}})$ and $n = o(b^{1/\alpha})$, the results of Theorem 2.7(ii) and Proposition 3.3 (which is about bias expansion) can be used to obtain uniform convergence of the mean estimator $\hat{\mu}_b(u) := \frac{1}{nb} \sum_{t=1}^{n} K(\frac{t/n-u}{b}) X_{t,n}$ towards $\mu(u) := \mathbb{E} \tilde{X}_0(u)$ in the sense that

$$\sup_{u \in [\frac{b}{2}, 1-\frac{b}{2}]} \left| \hat{\mu}_b(u) - \mu(u) \right| \xrightarrow{p} 0.$$

A central limit theorem

We provide local and global central limit theorems under minimal moment conditions which are useful in particular to find asymptotic distributions of (nonparametric) estimators of locally stationary processes, see Section 5. For the proofs, we need the dependence condition from Assumption 2.3(M1). Assumption 2.1(S2), namely $\| \sup_{u \in [0,1]} |\tilde{X}_t(u)| \|_q < \infty$, is crucial to show a Lindeberg-type condition under minimal moment conditions.

Theorem 2.9 (Central limit theorem – global version). *Let Assumption 2.6 hold. Suppose that Assumptions 2.1(S1), (S2) and 2.3(M1) hold with some $q \geq 2$ and $\alpha > \frac{1}{2}$. Define $S_n := \sum_{t=1}^{n} (X_{t,n} - \mathbb{E} X_{t,n})$ (if $X_{t,n}$ is replaced by $g(Z_{t,n})$ in the assertions below the same assumptions must be fulfilled with $\tilde{q} = q \cdot (M+1)$ instead of q). Then we have the following invariance principle*:

$$\{ S_{\lfloor nu \rfloor} / \sqrt{n}, 0 \leq u \leq 1 \} \xrightarrow{d} \left\{ \int_0^u \sigma(v) \, \mathrm{d}B(v), 0 \leq u \leq 1 \right\},$$

where $B(v)$ is a standard-Brownian motion and the long-run variance $\sigma^2(v)$ is given by

$$\sigma^2(v) = \sum_{k \in \mathbb{Z}} \mathrm{Cov}(\tilde{X}_0(v), \tilde{X}_k(v)). \tag{10}$$

The condition $\alpha > \frac{1}{2}$ can be omitted under Assumption 2.3(M2).

Finally, we present a simple localized version of the central limit theorem for general locally stationary processes.

Theorem 2.10 (Central limit theorem – local version). *Let Assumption 2.6 hold. Suppose that Assumptions 2.1(S1), (S2) and 2.3(M1) hold with some $q \geq 2$ and $1 \geq \alpha > 0$ (if $X_{t,n}$ is replaced by $g(Z_{t,n})$ in the assertion below the same assumptions must be fulfilled with $\tilde{q} = q \cdot (M+1)$ instead of q). Then for all $u \in (0,1)$, provided that $\sqrt{nb} \cdot n^{-\alpha} \to 0$, $b \to 0$ and $nb \to \infty$:*

$$\frac{1}{\sqrt{nb}} \sum_{t=1}^{n} K\left(\frac{t/n-u}{b} \right) \cdot \{ X_{t,n} - \mathbb{E} X_{t,n} \} \xrightarrow{d} N\left(0, \int K(x)^2 \, \mathrm{d}x \cdot \sigma^2(u) \right)$$

with $\sigma^2(u)$ defined in (10).

3. Differential calculus for nonstationary processes

In this section, we prove almost sure and uniform L^q Taylor expansions of $X_{t,n}$ and invent a kind of differential calculus for locally stationary processes. Below we will use this stronger kind of approximation for proving deterministic and stochastic bias expansions. While deterministic bias expansions are used to bound the expectation of expressions which include $X_{t,n}$, stochastic bias expansions can be used to replace the whole localized sum by a localized sum of the stationary process $\tilde{X}_t(u)$ which is much easier to analyze with known tools. We also prove results on the spectrum and the empirical distribution function.

3.1. Taylor expansions and expansions of localized sums

Proposition 3.1 (Taylor expansion). *Suppose that Assumption* 2.6 *holds. Suppose that Assumption* 2.1 *holds for some $q > 0$. Then we have for all $u \in [0, 1]$ and $t = 1, \ldots, n$:*

$$\tilde{X}_t\left(\frac{t}{n}\right) = \tilde{X}_t(u) + \left(\frac{t}{n} - u\right) \cdot \partial_u \tilde{X}_t(u) + R_{t,n} \qquad a.s. \tag{11}$$

If $|\frac{t}{n} - u| = o(1)$, then it holds that $R_{t,n} = o_{a.s.}(|\frac{t}{n} - u|)$. Furthermore,

$$
\begin{aligned}
&\frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \tilde{X}_t\left(\frac{t}{n}\right) \\
&= \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \tilde{X}_t(u) + \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right)\left(\frac{t}{n} - u\right) \partial_u \tilde{X}_t(u) + R_n \qquad a.s.
\end{aligned}
\tag{12}
$$

holds and $|R_n| \leq |K|_\infty \cdot b \cdot \sup_{|v-u| \leq b} |\partial_u X_t(v) - \partial_u X_t(u)| = o_{a.s.}(b)$.

If one is interested in the approximation of moments of localized sums (as it may be the case in bias expansions in nonparametric frameworks), the following theorem is appropriate. It also closes the gap between the locally stationary process $X_{t,n}$ and its approximation $\tilde{X}_t(t/n)$.

Corollary 3.2 (Almost sure and L^q expansion of localized sums). *Suppose that Assumption* 2.6 *holds. Suppose that Assumption* 2.1 *holds for some $q \geq 1$. Then for each fixed $u \in (0, 1)$,*

$$\left\| \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \cdot X_{t,n} - \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \cdot \tilde{X}_t\left(\frac{t}{n}\right) \right\|_q = O(n^{-1}),$$

and the expansion (12) *is valid with $\|R_n\|_q = o(b)$.*

It is also possible to use expansions similar to above if the sum is not localized by a kernel. An example can be found in the proof of Theorem 2.7.

3.2. Bias expansions

In nonparametric statistics, bias expansions play an important role to control the mean squared error (MSE) of estimators. Here we give an approach to estimate the deterministic bias term involving locally stationary processes. In recent years (for instance due to model selection via contrast minimization) a more careful analysis of the stochastic part in the calculation of the MSE became important. We make a contribution to this topic via a stochastic bias expansion which allows us to remove the bias from a localized sum $\frac{1}{n} \sum_{t=1}^{n} K_b(\frac{t}{n} - u) X_{t,n}$ such that only a localized sum over a stationary process $\frac{1}{n} \sum_{t=1}^{n} K_b(\frac{t}{n} - u) \tilde{X}_t(u)$ remains for which much more theoretical work was done. To obtain bias expansions of smaller order than $O(b)$ we have to assume differentiability of the upcoming expectations which is ensured by assuming Assumption 2.1(S3). To emphasize some differences that occur when Assumption 2.1(S3) is changed to differentiability in L^q, we state the deterministic bias expansion for both settings and comment in Remark 3.5.

Proposition 3.3 (Deterministic bias expansion). *Suppose that Assumption* 2.6 *holds. Let* $q \geq 1$. *Suppose that Assumption* 2.1(S1) *is fulfilled with some* $1 \geq \alpha > 0$, *then we have uniformly in* $u \in [0, 1]$:

$$\frac{1}{n} \sum_{t=1}^{n} K_b\left(\frac{t}{n} - u\right) \{\mathbb{E} X_{t,n} - \mathbb{E}\tilde{X}_t(t/n)\} = O(n^{-\alpha}), \tag{13}$$

and

$$\frac{1}{n} \sum_{t=1}^{n} K_b\left(\frac{t}{n} - u\right) \{\mathbb{E}\tilde{X}_t(t/n) - \mathbb{E}\tilde{X}_0(u)\} = O(b^\alpha) + O(n^{-1}). \tag{14}$$

Now assume additionally that K is symmetric.

(a) *If Assumption* 2.1(S3) *holds, then* (13) *is valid with* $\alpha = 1$ *and we have uniformly in* $u \in [\frac{b}{2}, 1 - \frac{b}{2}]$

$$\frac{1}{n} \sum_{t=1}^{n} K_b\left(\frac{t}{n} - u\right) \{\mathbb{E}\tilde{X}_t(t/n) - \mathbb{E}\tilde{X}_0(u)\} = o(b) + O(n^{-1}). \tag{15}$$

(b) *Suppose that* $[0, 1] \to L^q$, $u \mapsto \tilde{X}_0(u)$ *is Fréchet differentiable with derivative* $\tilde{D}_t(u)$, *i.e. for all* $u \in [0, 1]$,

$$\lim_{h \to 0} \left\| \frac{\tilde{X}_0(u+h) - \tilde{X}_0(u)}{h} - \tilde{D}_0(u) \right\|_q = 0. \tag{16}$$

Then the statement of (a) *holds.*

The proof of (13) and (14) follows from the Hoelder inequality and the fact that K has bounded variation and thus is bounded. To prove (15), note that

$$\tilde{X}_t(t/n) = \tilde{X}_t(u) + \left(\frac{t}{n} - u\right) \cdot \partial_u \tilde{X}_t(u) + \int_u^{t/n} \left\{\partial_u \tilde{X}_t(s) - \partial_u \tilde{X}_t(u)\right\} ds.$$

As long as $|\frac{t}{n} - u| \le b$, we have

$$\left| \mathbb{E} \int_u^{t/n} \left\{\partial_u \tilde{X}_t(s) - \partial_u \tilde{X}_t(u)\right\} ds \right| \le b \cdot \sup_{|u-s| \le b} \left\| \partial_u \tilde{X}_t(s) - \partial_u \tilde{X}_t(u) \right\|_1 = o(b),$$

since $u \mapsto \partial_u \tilde{X}_t(u)$ is continuous and $\| \sup_u |\partial_u \tilde{X}_t(u)| \|_1 < \infty$. Finally, because K has bounded variation and is symmetric,

$$\frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \left\{\mathbb{E}\tilde{X}_t(t/n) - \mathbb{E}\tilde{X}_t(u)\right\}$$

$$= \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \cdot \left(\frac{t}{n} - u\right) \cdot \mathbb{E}\left[\partial_u \tilde{X}_t(u)\right] + o(b)$$

$$= O\left(n^{-1}\right) + o(b).$$

The proof of (b) follows by using

$$\left| \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right) \left\{\mathbb{E}\tilde{X}_t(t/n) - \mathbb{E}\tilde{X}_t(u)\right\} \right|$$

$$\le \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right)\left(\frac{t}{n} - u\right) \left\| \frac{\tilde{X}_t(t/n) - \tilde{X}_t(u)}{\frac{t}{n} - u} - \tilde{D}_t(u) \right\|_1$$

$$+ \frac{1}{n} \sum_{t=1}^n K_b\left(\frac{t}{n} - u\right)\left(\frac{t}{n} - u\right) \mathbb{E}\tilde{D}_t(u) = o(b) + O\left(n^{-1}\right).$$

Remark 3.4. Note that in the situation of Proposition 3.3, derivative processes were used to get $o(b)$ instead of $O(b)$ in (14). Even smaller rates can be obtained by using higher order derivative processes together with higher order kernels. If we assume that $u \mapsto \tilde{X}_t(u)$ has a twice continuously differentiable modification and K is symmetric, we obtain a bias decomposition whose structure is well-known:

$$\frac{1}{nb} \sum_{t=1}^n K\left(\frac{t/n - u}{b}\right) \left\{\mathbb{E}X_{t,n} - \mathbb{E}\tilde{X}_0(u)\right\}$$

$$= \int x^2 K(x)\,dx \cdot \mathbb{E}\left[\partial_u^2 \tilde{X}_t(u)\right] \cdot b^2 + o\left(b^2\right) + O\left(n^{-1}\right).$$

Remark 3.5 (Almost sure v.s. L^q differentiability). Let us briefly comment on the different conditions in Proposition 3.3(a), (b). In (a) we ask for a.s. differentiability of $u \mapsto \tilde{X}_0(u)$ while (b) asks for differentiability in L^q which is weaker. If we want to apply Proposition 3.3 to $g(Z_{t,n})$ with some function $g \in \mathcal{L}_r(M, C)$ which is continuously differentiable, then there occur some differences due to the different natures of the conditions:

- If Assumption 2.1(S3) is fulfilled for $q' = q(M + 1)$, then Assumption 2.1(S3) is fulfilled for the process $g(Z_{t,n})$ with q (cf. Proposition 2.5(ii) and the comment afterwards) and we obtain (15) by Proposition 3.3(a).
- If (16) holds, we have to assume additionally that all the derivatives $\partial_j g$ ($j = 1, \ldots, r$) are Hoelder continuous with polynomially growing Hoelder constant, that is, with some $\gamma > 0$,

$$\sup_{y \neq y'} \frac{|\partial_j g(y) - \partial_j g(y')|}{|y - y'|_1^{\gamma} \cdot (1 + |y|_1^{M-1} + |y'|_1^{M-1})} < \infty,$$

to obtain (16) for $\tilde{X}_t^{\circ}(u) = g(\tilde{Z}_t(u))$ with derivative $D_t^{\circ}(u) = \partial_z g(\tilde{Z}_t(u)) \cdot D_t(u)$.

Note that we have to ask g to be slightly more smooth when using differentiability in L^q.

We now prove the stochastic bias expansion. It turns out that we have to bound moments of sums of the upcoming derivative processes $\partial_u \tilde{X}_t(u)$ which means that we have to pose dependence conditions on $\partial_u \tilde{X}_t(u)$. This is done via Assumption 2.3(M3). Using the projection operator $P_j \cdot := \mathbb{E}[\cdot | \mathcal{F}_j] - \mathbb{E}[\cdot | \mathcal{F}_{j-1}]$, we can bound moments of sums of $\partial_u \tilde{X}_t(u)$ by moments of martingales which can then be bounded with results from [18]. It can be shown (similar to [29], Theorem 1(i) and (ii)) that for some shift process $W_t = H_t(\mathcal{F}_t)$ with measurable H_t it holds for $q \geq 1$:

$$\|P_{t-k} W_t\|_q \leq \delta_q^W(k). \tag{17}$$

Proposition 3.6 (Stochastic bias expansion). *Suppose that Assumptions 2.1 and 2.3(M3) are fulfilled for some $q \geq 2$. Assume that K is symmetric. Then we have*

$$\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} \left\| \frac{1}{n} \sum_{t=1}^{n} K_b\left(\frac{t}{n} - u\right) \{X_{t,n} - \tilde{X}_t(u)\} \right\|_q = o(b) + O(n^{-1}). \tag{18}$$

Proof of Proposition 3.6. To prove (18), we can show similarly as in (13) that

$$\frac{1}{n} \sum_{t=1}^{n} \left| K_b\left(\frac{t}{n} - u\right) \right| \cdot \left\| X_{t,n} - \tilde{X}_t(t/n) \right\|_q = O(n^{-1}).$$

To deal with $\frac{1}{n}\|\sum_{t=1}^n K_b(\frac{t}{n}-u)(\tilde{X}_t(t/n)-\tilde{X}_t(u))\|_q$, we use the expansion (12) together with the result $\|R_n\|_q = o(b)$ from Corollary 3.2. It therefore remains to analyze

$$\frac{1}{n}\left\|\sum_{t=1}^n K_b\left(\frac{t}{n}-u\right)\left(\frac{t}{n}-u\right)\cdot\left\{\partial_u\tilde{X}_t(u)-\mathbb{E}\partial_u\tilde{X}_t(u)\right\}\right\|_q$$

$$+\frac{1}{n}\sum_{t=1}^n K_b\left(\frac{t}{n}-u\right)\left(\frac{t}{n}-u\right)\cdot\mathbb{E}\partial_u\tilde{X}_t(u).$$

While the second term is $O(n^{-1})$ by stationarity, the first term is bounded by

$$\frac{1}{n}\sum_{k=0}^\infty\left\|\sum_{t=1}^n K_b\left(\frac{t}{n}-u\right)\left(\frac{t}{n}-u\right)P_{t-k}\partial_u\tilde{X}_t(u)\right\|_q$$

$$\leq\frac{q^{1/2}}{nb}\sum_{k=0}^\infty\left(\sum_{t=1}^n\left(K_b\left(\frac{t}{n}-u\right)\cdot\left(\frac{t}{n}-u\right)\right)^2\|P_{t-k}\partial_u\tilde{X}_t(u)\|_q^2\right)^{1/2}$$

$$\leq\frac{q^{1/2}|K|_\infty b^{1/2}}{n}\sum_{k=0}^\infty\delta_q^{\partial_u\tilde{X}(u)}(k)=\frac{q^{1/2}|K|_\infty b^{1/2}}{n}\Delta_{0,q}^{\tilde{X}}$$

by Theorem 2.1 in [18] and (17). □

The main advantage of a stochastic bias expansion is that we can reduce a sum over locally stationary processes to a sum over stationary processes by keeping the terms stochastic. This allows for instance to apply large deviation results for stochastic processes which usually have a simpler and more closed form.

Remark 3.7 (An application of the stochastic bias expansion: Deviation inequalities). Suppose that the assumptions of Proposition 3.6 are fulfilled with $q = 2$. Assume that $\mathbb{E}\tilde{X}_t(u) = 0$. Then we have for $\gamma > 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^n K_b(u-t/n)X_{t,n}\right|>\gamma\right)$$

$$\leq\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^n K_b(u-t/n)\tilde{X}_t(u)\right|>\frac{\gamma}{2}\right)$$

$$+\frac{1}{(\gamma/2)^2}\left\|\frac{1}{n}\sum_{t=1}^n K_b(u-t/n)\left(X_{t,n}-\tilde{X}_t(u)\right)\right\|_2^2.$$

By Proposition 3.6, the second term on the right-hand side is $o(b) + O(n^{-1})$, that is, has bias order. For the first term, one can use deviation results for stationary processes.

3.3. Differentiability of functionals

The existence of derivative processes allows an expansion of the corresponding mean $\mathbb{E}g(Z_{t,n})$ into the mean of the corresponding stationary version $\mathbb{E}g(\tilde{Z}_t(u))$. This can be applied to various functionals such as expectations, covariances, the Wigner-Ville spectrum and the distribution function. The following result is an immediate Corollary from Lemma 2.5 applied to some $g \in \mathcal{L}_r(M, C)$. Recall $Z_{t,n} = (X_{t,n}, \ldots, X_{t-r+1,n})'$ and $\tilde{Z}_t(u) = (\tilde{X}_t(u), \ldots, \tilde{X}_{t-r+1}(u))'$.

Proposition 3.8. *Assume that $g \in \mathcal{L}_r(M, C)$. Suppose that Assumption 2.1(S1) is fulfilled for some $1 \geq \alpha > 0$ and $q = M + 1$. Then we have uniformly for $t = 1, \ldots, n$:*

$$\mathbb{E}g(Z_{t,n}) = \mathbb{E}g\left(\tilde{Z}_t\left(\frac{t}{n}\right)\right) + O\left(n^{-\alpha}\right) = \mathbb{E}g\left(\tilde{Z}_t(u)\right) + O\left(n^{-\alpha} + \left|\frac{t}{n} - u\right|^\alpha\right). \tag{19}$$

If additionally Assumption 2.1(S3) is fulfilled and g is continuously differentiable with partial derivatives $\partial_j g \in \mathcal{L}_r(M - 1, C')$, $j = 1, \ldots, r$, then $u \mapsto \mathbb{E}g(\tilde{Z}_t(u))$ is continuously differentiable with derivative

$$\partial_u \mathbb{E}g\left(\tilde{Z}_t(u)\right) = \sum_{j=1}^r \mathbb{E}\left[\partial_j g\left(\tilde{X}_t(u), \ldots, \tilde{X}_{t-r+1}(u)\right) \cdot \partial_u \tilde{X}_{t-j+1}(u)\right]. \tag{20}$$

The result of Proposition 3.8 enables us to get expansions of the mean, the covariance and the distribution function of $X_{t,n}$. Suppose in the following that Assumption 2.1 holds for some $q \geq M + 1$.

Corollary 3.9 (Mean expansion, $M = 0$). *Choosing $g : \mathbb{R} \to \mathbb{R}$, $g(y) = y$ yields*

$$\mathbb{E}X_{t,n} = \mathbb{E}\tilde{X}_t(t/n) + O\left(n^{-1}\right),$$

where $\mu(u) := \mathbb{E}\tilde{X}_0(u)$ is continuously differentiable with derivative $\partial_u \mu(u) = \mathbb{E}\partial_u \tilde{X}_0(u)$.

Corollary 3.10 (Covariance expansion, $M = 1$). *Fix $r > 0$. Define the covariances $\gamma(u, r) := \text{Cov}(\tilde{X}_t(u), \tilde{X}_{t-r}(u))$. Choosing $g : \mathbb{R}^{r+1} \to \mathbb{R}$, $g(y) = y_1 y_{r+1}$, we obtain uniformly for $t = 1, \ldots, n$:*

$$\gamma_{t,n}(r) := \text{Cov}(X_{t,n}, X_{t-r,n}) = \gamma\left(\frac{t}{n}, r\right) + O\left(n^{-1}\right) \tag{21}$$

and $\gamma(u, r)$ is continuously differentiable with derivative

$$\partial_u \gamma(u, r) = \text{Cov}\left(\partial_u \tilde{X}_0(u), \tilde{X}_r(u)\right) + \text{Cov}\left(\tilde{X}_0(u), \partial_u \tilde{X}_r(u)\right).$$

Similar expansions can be derived for higher-order cumulants and also for the Wigner-Ville spectrum (cf. [14]).

As a last application of Proposition 3.8, we present an expansion of the distribution function of $X_{t,n}$ which may also be used to approximate quantiles of locally stationary processes.

Example 3.11 (Expansion of the distribution function). Suppose that Assumption 2.1 holds with $q = 1$. Assume that the i.i.d. random variables ε_t, $t \in \mathbb{Z}$ have a Lipschitz continuous and continuously differentiable distribution function F_ε with Lipschitz constant L_ε and derivative f_ε. Let the processes $X_{t,n}$ and $\tilde{X}_t(u)$ obey the recursion equations (3) and (4).

Assume that $(\varepsilon, y, u) \mapsto G_\varepsilon(y, u)$ is continuously differentiable and that the derivative $\partial_\varepsilon G_\varepsilon(y, u) \geq \delta_G > 0$ is uniformly bounded from below by some positive constant $\delta_G > 0$. By the inverse function theorem we know that there exists a continuously differentiable inverse $x \mapsto H(x, y, u)$ of $\varepsilon \mapsto G_\varepsilon(y, u)$. Finally, assume that for all $x \in \mathbb{R}$, the expressions

$$C(x) := \sup_{u \in [0,1]} \sup_{y \neq y'} \frac{|H(x, y, u) - H(x, y', u)|}{|y - y'|_1}$$

are finite.

Put $Y_{t-1,n} = (X_{t-1,n}, \ldots, X_{t-p,n})'$, $\tilde{Y}_{t-1}(u) = (\tilde{X}_{t-1}(u), \ldots, \tilde{X}_{t-p}(u))'$. In this situation, it holds that the distribution function of $X_{t,n}$,

$$F_{X_{t,n}}(x) = \mathbb{E}\big[\mathbb{P}\big(G_{\varepsilon_t}(Y_{t-1,n}, t/n) \leq x | \mathcal{F}_{t-1}\big)\big] = \mathbb{E}\big[F_\varepsilon\big(H(x, Y_{t-1,n}, t/n)\big)\big]$$

can be approximated by the distribution function $F_{\tilde{X}_t(u)}(x) := \mathbb{P}(\tilde{X}_t(u) \leq x)$ by

$$\big|F_{X_{t,n}}(x) - F_{\tilde{X}_t(t/n)}(x)\big|$$

$$\leq L_\varepsilon \big\| H(x, Y_{t-1,n}, t/n) - H\big(x, \tilde{Y}_{t-1}(t/n), t/n\big) \big\|_1$$

$$\leq L_\varepsilon C(x) \sum_{j=1}^{p} \big\| X_{t-j-1,n} - \tilde{X}_{t-j-1}(t/n) \big\|_1 \leq pC_B L_\varepsilon \cdot C(x) \cdot n^{-1}$$

Furthermore $u \mapsto F_{\tilde{X}_t(u)}(x)$ is differentiable with derivative

$$\partial_u F_{\tilde{X}_t(u)}(x)$$

$$= \mathbb{E}\big[f_\varepsilon\big(H\big(x, \tilde{Y}_{t-1}(u), u\big)\big) \cdot \big(\big\langle \partial_2 H\big(x, \tilde{Y}_{t-1}(u), u\big), \partial_u \tilde{Y}_{t-1}(u) \big\rangle + \partial_3 H\big(x, \tilde{Y}_{t-1}(u), u\big)\big)\big].$$

4. Nonlinear locally stationary processes

In this section, we show in a sequence of theorems that the Markov processes given by (3) and (4) fulfill Assumption 2.1 and the mixing conditions of Assumption 2.3. Furthermore, we prove that the derivative process can be obtained as the solution of a functional equation. The existence of these processes and their properties have previously been derived for tvAR models (cf. [4]), tvARCH models (cf. [8]) and random coefficient models (cf. [24]). The situation in the present case is however different since the process (3) is only defined by a recursion and the explicit solution is usually not available for the calculations. To prove the results, we state the following elementary assumptions on the recursion function $G_\varepsilon(y, u)$. Let $\partial_1 G$, $\partial_2 G$ denote the derivatives of G w.r.t. y and u, respectively.

Assumption 4.1. In the model (3), (4) we assume with $Y_{t-1,n} = (X_{t-1,n}, \ldots, X_{t-p,n})'$ and $\tilde{Y}_{t-1}(u) = (\tilde{X}_{t-1}(u), \ldots, \tilde{X}_{t-p}(u))'$ that there exists $q > 0$, $\chi = (\chi_1, \ldots, \chi_p) \in \mathbb{R}^p_{\geq 0}$ with $|\chi|_1 = \sum_{i=1}^{p} \chi_i < 1$ and $y_0 \in \mathbb{R}^p$ such that with $q' := \min\{q, 1\}$:

(L1) $\sup_{u \in [0,1]} \|G_{\varepsilon_0}(y_0, u)\|_q < \infty$, and (with $|z|_{\chi,q'} := (\sum_{i=1}^{p} |z_i|^{q'} \cdot \chi_i)^{1/q'}$ the weighted q'-norm)

$$\sup_{u \in [0,1]} \sup_{y \neq y'} \frac{\|G_{\varepsilon_0}(y, u) - G_{\varepsilon_0}(y', u)\|_q}{|y - y'|_{\chi,q'}} \leq 1. \tag{22}$$

(L2) $(y, u) \mapsto G_\varepsilon(y, u)$ is continuous for all ε, $\| \sup_{u \in [0,1]} |G_{\varepsilon_0}(y_0, u)| \|_q < \infty$, and

$$\left\| \sup_{u \in [0,1]} \sup_{y \neq y'} \frac{|G_{\varepsilon_0}(y, u) - G_{\varepsilon_0}(y', u)|}{|y - y'|_{\chi,q'}} \right\|_q \leq 1. \tag{23}$$

(L3) $(y, u) \mapsto G_\varepsilon(y, u)$ is continuously differentiable for all ε, $\| \sup_{u \in [0,1]} |\partial_2 G_{\varepsilon_0}(y_0, u)| \|_q < \infty$, and

$$C_i := \left\| \sup_{u \in [0,1]} \sup_{y \neq y'} \frac{|\partial_i G_{\varepsilon_0}(y, u) - \partial_i G_{\varepsilon_0}(y', u)|_1}{|y - y'|_{1,q'}} \right\|_q < \infty, \qquad i = 1, 2. \tag{24}$$

Furthermore, assume that either (a) (23) holds for $q/2$ instead of q or (b) $y \mapsto \partial_1 G_\varepsilon(y, u)$ is constant for all ε, u.

(L4) For some $0 < \alpha \leq 1$, it holds that

$$C := \sup_{u \in [0,1]} \left\| C\big(\tilde{Y}_t(u)\big) \right\|_q < \infty \qquad \text{where}$$

$$C(y) := \sup_{u \neq u'} \frac{\|G_{\varepsilon_0}(y, u) - G_{\varepsilon_0}(y, u')\|_q}{|u - u'|^\alpha}. \tag{25}$$

Let us briefly discuss the conditions in Assumption 4.1.

Remark 4.2.

(i) Note that (L1)–(L3) impose increasingly strong smoothness assumptions on the recursion function $G_\varepsilon(y, u)$. While (L1)–(L3) are directly verifiable, (L4) includes conditions on the stationary approximation $\tilde{X}_t(u)$. Note that the upcoming theorems also state properties of $\tilde{X}_t(u)$. Their results can be used to verify (L4).

(ii) The condition (L2) means that the mapping $y \mapsto G_\varepsilon(y, u)$ can be viewed as a contraction in the space of continuous functions $C[0, 1]$ which in turn implies the a.s. continuity of the limit. (L3) is necessary to ensure that $y \mapsto G_\varepsilon(y, u)$ is a contraction in $C^1[0, 1]$.

(iii) Condition (L3)(a) or (b) is necessary due to the product in (28). Note that by Hoelder's inequality, (L3)(a) follows from (23) if $q \geq 2$. The inequality $|z|_q \leq |z|_{q'}$ for $0 < q' \leq q$, $z \in \mathbb{R}^p$ implies that (L3)(a) is fulfilled if (23) holds with $\sum_{i=1}^{p} \chi_i^{1/2} < 1$.

(iv) For $p > 1$, the conditions stated in Assumption 4.1 may lead to non-optimal restrictions on G which is due to the general formulation. One way to circumvent this is by posing conditions on the mth iteration of G instead of G itself. Some models like tvAR(p) or tvARCH(p) also allow a reformulation to a p-dimensional recursion with only one lag. Since our aim is to cover a wide range of models with simple conditions, we will not discuss these approaches in detail.

Remark 4.3 (Almost sure calculus v.s. L^q calculus). As mentioned in the beginning of this paper, many of the statistical applications in Section 3 can be proved by only assuming differentiability of $u \mapsto \tilde{X}_t(u)$ in L^q, an example was given in Proposition 3.3(b). As pointed out by a referee, to obtain L^q differentiability, Assumptions 4.1(L2), (L3) can be weakened. Technically, proving differentiability of $u \mapsto \tilde{X}_t(u)$ then corresponds to analyzing smoothness properties of a fixed point of the iteration (4). As analytic results in [20] suggest, the main difference is a decrease in smoothness assumptions that have to be posed on G, namely G is no longer needed to be continuously differentiable but only differentiable a.e. and the suprema in (23) and (24) can be taken outside. This theory would also include $tvTAR$ processes (cf. Example 2.2).

There are some drawbacks when using only L^q calculus and no a.s. statements. As already mentioned in Remark 3.5, one has to pose slightly more smoothness conditions on g if one wants to apply the theory to processes $g(Z_{t,n})$. Moreover, it seems that a Lindeberg-type condition which is used in the proof of the global CLT Theorem 2.9 can only be shown under Assumption 2.1(S2) when only second moments are available. To ensure 2.1(S2), we have to ask for 4.1(L2) in view of Theorem 4.6.

Existence and uniqueness of $X_{t,n}$ and $\tilde{X}_t(u)$

We now establish existence and uniqueness under mild contraction conditions.

Proposition 4.4.

(i) *Existence of a stationary approximation*: *Suppose that Assumption* 4.1(L1) *holds. Then for all $u \in [0, 1]$, the recursion* (4) *has an a.s. unique $\mathcal{F}_t$-measurable, stationary and ergodic solution $\tilde{X}_t(u) = H(u, \mathcal{F}_t)$ and we have with some $C > 0$ and $0 < \rho < 1$:*

$$\sup_{u \in [0,1]} \delta_q^{\tilde{X}(u)}(k) \leq C\rho^k, \qquad \sup_{u \in [0,1]} \left\| \tilde{X}_0(u) \right\|_q < \infty.$$

(ii) *Existence of the nonstationary process*: *Under the above conditions, there exists an a.s. unique $\mathcal{F}_t$-measurable solution of* (3) *with $X_{t,n} = H_{t,n}(\mathcal{F}_t)$, where $H_{t,n}$ are measurable functions. Furthermore, $\sup_{n \in \mathbb{N}} \sup_{t=1,\ldots,n} \|X_{t,n}\|_q < \infty$ and with some $C > 0$ and $0 < \rho < 1$:*

$$\sup_{n \in \mathbb{N}} \delta_q^{X_{\cdot,n}}(k) \leq C\rho^k.$$

The proof of (i) for fixed $u \in [0, 1]$ is similar to the proof in [23], Theorem 5.1. Since we state the results uniformly in $u \in [0, 1]$, we will give the proof in the Supplementary Material

[7] for completeness. Since the definition of $X_{t,n}$ and $\tilde{X}_t(0)$ coincide for $t \leq 0$, existence and uniqueness of $X_{t,n}$ follow from the existence and uniqueness of $\tilde{X}_t(0)$. Therefore, the existence statement in (ii) is an immediate corollary of (i).

A uniform L^q approximation

We now prove that $X_{t,n}$ can be approximated by the stationary process $\tilde{X}_t(u)$ uniformly in a L^q-sense.

Lemma 4.5. *Suppose that Assumption* 4.1(L1), (L4) *hold. Then*

$$\sup_{u \neq u'} \frac{\|\tilde{X}_t(u) - \tilde{X}_t(u')\|_q}{|u - u'|^\alpha} \leq \frac{C}{(1 - |\chi|_1)^{1/q'}}. \tag{26}$$

Furthermore, we have:

$$\sup_{t=1,\ldots,n} \left\| X_{t,n} - \tilde{X}_t(t/n) \right\|_q \leq C p^\alpha \left(\frac{|\chi|_1}{(1 - |\chi|_1)^2} \right)^{1/q'} \cdot n^{-\alpha}. \tag{27}$$

Note that the approximation error in (27) cannot be avoided – cf. [4], (49), for the tvAR(1) case (with a different error due to different assumptions).

Existence of continuous modifications and derivative processes

Proposition 4.4 gives the almost sure uniqueness of $\tilde{X}_t(u)$ for each $u \in [0, 1]$, but not continuity of $u \mapsto \tilde{X}_t(u)$ since this involves uncountably many points $u \in [0, 1]$. In order to guarantee the existence of a continuous or even differentiable modification $\hat{X}_t(u)$ of $\tilde{X}_t(u)$ we have to impose stronger conditions on the recursion function G in (3) ($\hat{X}_t(u)$ is a modification of $\tilde{X}_t(u)$ if for all $u \in [0, 1]$, $\hat{X}_t(u) = \tilde{X}_t(u)$ a.s.). A natural way would be to apply extensions of the Kolmogorov–Chentzov theorem, but they usually contain tradeoffs in their conditions between moment assumptions and smoothness of the process which usually leads to either strong moment or smoothness assumptions which may not be useful in practice. Furthermore it does not use the specific structure of the process which is known and we could not give a bound for moments of $\sup_{u \in [0,1]} |\hat{X}_t(u)|$. We therefore use a different approach.

Theorem 4.6 (Existence of a continuous modification). *Suppose that Assumption* 4.1(L2) *holds. Then for each* $t \in \mathbb{Z}$, *there exists a continuous modification* $(\hat{X}_t(u))_{u \in [0,1]}$ *of* $(\tilde{X}_t(u))_{u \in [0,1]}$ *from Proposition* 4.4 *with* $\sup_{u \in [0,1]} |\hat{X}_t(u)| \in L^q$.

Remark 4.7. In the case $G_\varepsilon(y, u) = \tilde{G}_\varepsilon(y, \theta_0(u))$ with some parameter curve $\theta_0 : [0, 1] \to \Theta$ (cf. Section 5), the supremum taken over $u \in [0, 1]$ in (23) restricts the parameter space Θ.

If additionally Assumption 4.1(L3) is fulfilled, Theorem 4.6 also holds under the weaker condition $\sup_{u \in [0,1]} \| \sup_{y \neq y'} \frac{|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y',u)|}{|y - y'|_{\chi,q'}} \|_q \leq 1$ which leads to larger admissible parameter spaces Θ. For details, see Proposition A.3 in Supplement A.

In the following, we assume that $(y, u) \mapsto G_\varepsilon(y, u)$ is differentiable in both components. For the moment, assume that there exists a modification $(\hat{X}_t(u))_{u \in [0,1]}$ of the process $(\tilde{X}_t(u))_{u \in [0,1]}$ with differentiable paths and denote the derivative by $\partial_u \hat{X}_t(u)$. Then the following recursion equation for $D_t(u) = \partial_u \hat{X}_t(u)$, obtained by differentiating (4) *should hold* a.s.:

$$D_t(u) = \langle \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u), (D_{t-1}(u), \dots, D_{t-p}(u))' \rangle + \partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u). \quad (28)$$

This is shown in the next theorem. The first part is devoted to the existence of a solution $D_t(u)$ of the recursion (28) given the existence of the process $\tilde{X}_t(u)$ from Theorem 4.4; in the second part we prove that $\tilde{X}_t(u)$ has a differentiable modification with respect to u and that the derivative coincides with $D_t(u)$. Both $\tilde{X}_t(u)$ and $D_t(u)$ are uniquely determined by (4) and (28).

Theorem 4.8 (Existence of derivative processes). *Suppose that Assumptions* 4.1(L2), (L3) *hold. Then the following statements are true.*

(i) *Existence of the first derivative process*: *For all $u \in [0, 1]$, the recursion (28) has a unique stationary and ergodic solution $D_t(u) = \tilde{H}(u, \mathcal{F}_t)$ with some measurable H and it holds that*

$$\delta_q^{D(u)}(k) \leq C\rho^k, \qquad \sup_{u \in [0,1]} \| D_t(u) \|_q < \infty$$

with some $C > 0, 0 < \rho < 1$.

(ii) *Differentiability*:
   (a) *There exists a continuously differentiable modification $(\hat{X}_t(u))_{u \in [0,1]}$ of the process $(\tilde{X}_t(u))_{u \in [0,1]}$ from Proposition 4.4 such that for all $u \in [0, 1]$ it holds that $\partial_u \hat{X}_t(u) = D_t(u)$ a.s.*
   (b) $\sup_{u \in [0,1]} |\partial_u \hat{X}_t(u)| \in L^q$.

Finally, let us summarize the results from this section in the following corollary.

Corollary 4.9. *Let Assumption* 4.1 *be fulfilled. Then modifications of the a.s. unique solutions of* (3) *and* (4) *fulfill Assumptions* 2.1 *and* 2.3.

For some models, it is possible to obtain explicit expressions for the corresponding derivative processes.

Example 4.10 (Explicit representations for derivative processes).

(i) The tvAR(p) process $X_{t,n} = \sum_{j=1}^p a_j(\frac{t}{n}) X_{t-j,n} + \varepsilon_t$ has the corresponding stationary approximation $\tilde{X}_t(u) = \sum_{j=1}^p a_j(u) \tilde{X}_{t-j}(u) + \varepsilon_t$ which has an explicit representation $\tilde{X}_t(u) = \sum_{j=0}^\infty \psi_j(u) \cdot \varepsilon_{t-j}$ with differentiable $\psi_j : [0, 1] \to \mathbb{R}$ ($j = 0, 1, 2, \dots$). It is

easy to see that $\partial_u \tilde{X}_t(u) = \sum_{j=0}^{\infty} \partial_u \psi_j(u) \cdot \varepsilon_{t-j}$ is the a.s. uniquely determined derivative process.

(ii) Similarly to (i), it is easy to see that general linear processes $\tilde{X}_t(u) = \sum_{j=0}^{\infty} \psi_j(u) \cdot \varepsilon_{t-j}$ with differentiable $\psi_j : [0, 1] \to \mathbb{R}$ ($j = 0, 1, 2, \ldots$) have derivative process $\partial_u \tilde{X}_t(u) = \sum_{j=1}^{\infty} \partial_u \psi_j(u) \cdot \varepsilon_{t-j}$ under appropriate summability conditions.

(iii) For tvARCH(p) processes, explicit expressions for the derivative processes were obtained in [8].

In the following, we will write $\tilde{X}_t(u)$ even if we mean the differentiable modification to keep notation simple. Since all our results only involve countably many observations, this will not cause any problems.

Higher order derivative processes

Under additional assumptions, one can show uniform L^q Hoelder properties of the first derivative process:

Proposition 4.11 (Hoelder property of the first derivative process). *Suppose that Assumption* 4.1(L2), (L3) *hold. Additionally assume that for some* $1 \geq \alpha_2 > 0$ *and* $i = 1, 2$ *it holds component-wise*:

$$D_i := \sup_u \big\| D_i\big(\tilde{Y}_t(u)\big) \big\|_q < \infty, \qquad D_i(y) := \sup_{u \neq u'} \frac{\|\partial_i G_{\varepsilon_0}(y, u) - \partial_i G_{\varepsilon_0}(y, u')\|_q}{|u - u'|^{\alpha_2}}. \qquad (29)$$

Then

$$\sup_{u \neq u'} \frac{\|\partial_u \tilde{X}_t(u) - \partial_u \tilde{X}_t(u')\|_{q/2}}{|u - u'|^{\alpha_2}} \leq C,$$

with some constant $C > 0$.

If $\tilde{X}_t(u)$ has a twice continuously differentiable modification and $(y, u) \mapsto G_\varepsilon(y, u)$ is twice continuously differentiable, then the following recursion equation for $\partial_u^2 \tilde{X}_t(u)$ should hold:

$$\begin{aligned}
\partial_u^2 \tilde{X}_t(u) = {} & \big\langle \partial_1 G_{\varepsilon_t}\big(\tilde{Y}_{t-1}(u), u\big), \partial_u^2 \tilde{Y}_{t-1}(u) \big\rangle \\
& + \big\langle \partial_1^2 G_{\varepsilon_t}\big(\tilde{Y}_{t-1}(u), u\big) \partial_u \tilde{Y}_{t-1}(u), \partial_u \tilde{Y}_{t-1}(u) \big\rangle \\
& + 2 \big\langle \partial_1 \partial_2 G_{\varepsilon_t}\big(\tilde{Y}_{t-1}(u), u\big), \partial_u \tilde{Y}_{t-1}(u) \big\rangle + \partial_2^2 G_{\varepsilon_t}\big(\tilde{Y}_{t-1}(u), u\big).
\end{aligned} \qquad (30)$$

Using the same techniques as in Theorem 4.8, one can find similar conditions as in Assumption 4.1 such that a second (or even higher) order derivative process $\partial_u^2 \tilde{X}_t(u)$ exists. Let us point out an interesting anomaly in the case of second order derivatives that is also existent for higher order derivatives: Due to the additional products in (30) it turns out that, in general, one has to assume $2q$th moments of $\tilde{X}_t(u)$ to guarantee the existence of the qth moment of $\partial_u^2 \tilde{X}_t(u)$. The formalization of this is beyond the scope of this paper, but in Proposition 4.11 one already can see the imbalance of moments in the assumption and the obtained result.

A simulation study

To quantify the quality of the approximations given in Lemma 4.5 and Proposition 3.1, we consider the tvARCH(1) model

$$X_{t,n} := \left(a_0 + a_1 \left(\frac{t}{n} \right) X_{t-1,n}^2 \right)^{1/2} \varepsilon_t$$

with $a_0 := 0.2$, $a_1(u) = 0.95u^2$ and $\varepsilon_0 \sim N(0, 1)$. Note that if t/n tends to 1, the values of $X_{t,n}$ are more dependent to each other than for smaller values of t/n. We generated realizations of $X_{t,n}$, $\tilde{X}_t(\frac{t}{n})$ with $n = 500$ (see Figure 1(a), (b) for a realization of $X_{t,n}$ and $X_{t,n} - \tilde{X}_t(\frac{t}{n})$). In Figure 1(c), we have the plotted empirical 5%- and 95%-quantile curves of the difference $X_{t,n} - \tilde{X}_t(\frac{t}{n})$ for $N = 1000$ replications. It can be seen that with stronger dependence, the quality of the approximation $X_{t,n} \approx \tilde{X}_t(\frac{t}{n})$ gets worse as it is suggested by the bound in Lemma 4.5.

Second, we consider the approximation quality of $\tilde{X}_t(t/n)$ by $\tilde{X}_t(u)$ and $\tilde{X}_t(u) + (\frac{t}{n} - u)\partial_u \tilde{X}_t(u)$, respectively. Since these approximations are only working locally (for $|t/n - u| \ll 1$), we compare them by dividing the whole time line $t = 1, \ldots, n$ into subsets $(u_i - b, u_i + b]$, where $b = 25$ and $u_i = (2i - 1)b$ for $i = 1, \ldots, 10$. In Figure 1(d), empirical 5%- and 95%-quantile curves obtained from $N = 1000$ replications for the differences $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i)$ and $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i) - (\frac{t}{n} - u_i)\partial_u \tilde{X}_t(u_i)$ (where $t \in (u_i - b, u_i + b]$) are depicted, respectively. We emphasize that the improvement of the (pointwise) approximation $\tilde{X}_t(\frac{t}{n})$ by taking into account the derivative process is remarkable. However, both approximations again get worse if the dependence of $X_{t,n}$ to earlier values increases.

5. Application to maximum likelihood estimation

In this section, we investigate the asymptotic properties of maximum likelihood estimates for parameter curves of locally stationary models which can be written in the form (3). The results are in particular derived by using the asymptotic results and the differential calculus of Sections 2 and 3. More precisely, we investigate the recursively defined model

$$X_{t,n} = \tilde{G}_{\varepsilon_t} \left(X_{t-1,n}, \ldots, X_{t-p,n}, \theta_0 \left(\frac{t}{n} \right) \right), \qquad t = 1, \ldots, n, \tag{31}$$

where now the function $G_\varepsilon(y, u)$ from (3) has been replaced by $\tilde{G}_\varepsilon(y, \theta_0(u))$ with the unknown parameter curve $\theta_0 : [0, 1] \to \Theta \subset \mathbb{R}^d$ which is to be estimated. Our goal is to obtain estimators for $\theta_0(\cdot)$ based on $X_{t,n}$, $t = 1, \ldots, n$ with a quasi maximum likelihood approach.

Suppose for the moment that $\varepsilon \mapsto G_\varepsilon(y, \theta)$ is continuously differentiable for all ε, y, u and that the derivative $\partial_\varepsilon \tilde{G}_\varepsilon(y, \theta) \geq \delta_G > 0$ is bounded uniformly from below with some constant $\delta_G > 0$. This ensures that the new innovation ε_t has an impact on the value of $X_{t,n}$ which is not too small. Under these conditions, there exists a continuously differentiable inverse $x \mapsto H(x, y, \theta)$ of $\varepsilon \mapsto G_\varepsilon(y, \theta)$ (see also Example 3.11).
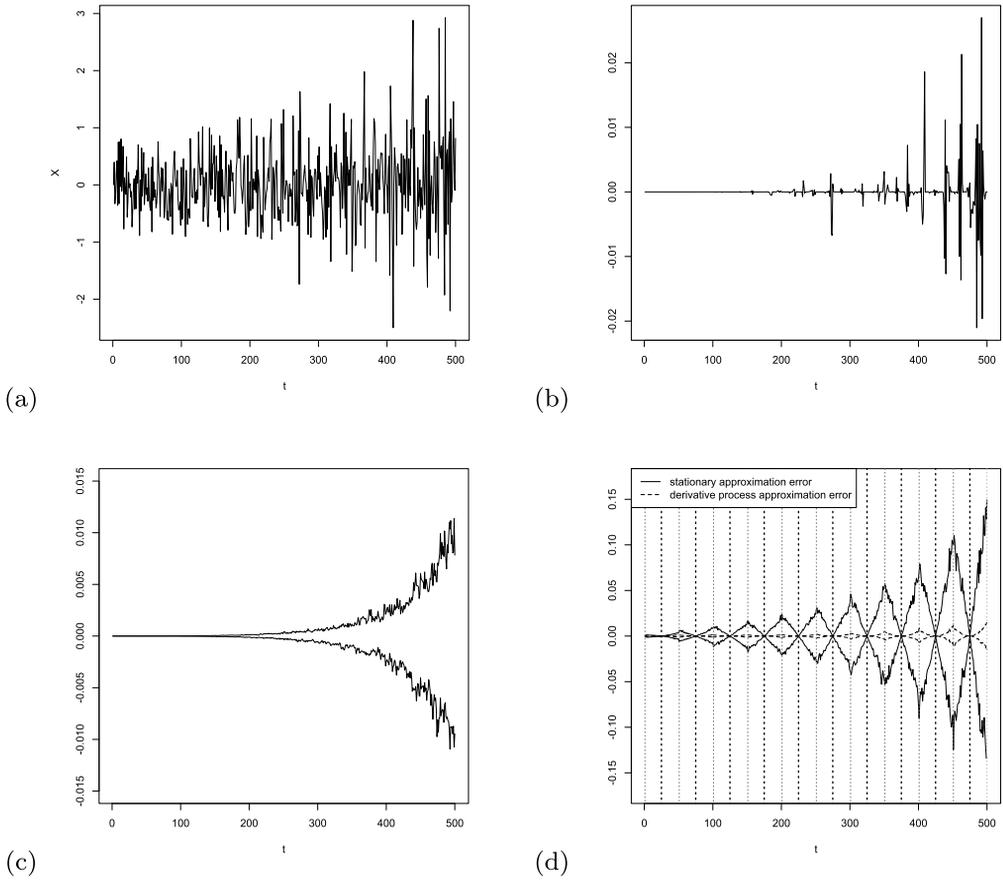
Figure 1. Top: (a) Realization of one $X_{t,n}$, $t = 1, \ldots, n$. (b) Difference $X_{t,n} - \tilde{X}_t(\frac{t}{n})$ for one realization. Bottom: (c) empirical 5%- and 95%-quantile curves of $X_{t,n} - \tilde{X}_t(\frac{t}{n})$ for $N = 1000$ replications. (d) Solid and Dashed: empirical 5%- and 95%-quantile curves of $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i)$ and $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i) - (\frac{t}{n} - u_i)\partial_u \tilde{X}_t(u_i)$ for $t \in (u_i - b, u_i + b]$ (grey thin vertical dotted lines) and $N = 1000$ replications, respectively. Here, $b = 25$ and $u_i = (2i - 1)b$ (black thick vertical dotted lines), $i = 1, \ldots, 10$.

Suppose that ε_0 has a continuous density f_ε. The negative conditional log likelihood of $X_{t,n} = x$ given $(X_{t-1,n}, \ldots, X_{t-p,n}) = y$ and $\theta_0(\cdot) \equiv \theta$ is then

$$\ell(x, y, \theta) = -\log f_\varepsilon\big(H(x, y, \theta)\big) - \log \partial_x H(x, y, \theta). \tag{32}$$

In the following derivations, we do not make use of the specific structure of ℓ. This means especially that we allow for model misspecifications due to a false density f_ε. Many authors prefer the case of a Gaussian density $f_\varepsilon(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ because then a minimizer θ

of ℓ can be interpreted as a minimum (quadratic) distance estimator (see [5] in the tvAR case, [8] in the tvARCH case).

Based on this we define $\ell_{t,n}(\theta) := \ell(X_{t,n}, Y_{t-1,n}, \theta)$. Let $b \in (0, 1)$ be a bandwidth K a kernel function as considered in Assumption 2.6. We define the local negative log conditional likelihood

$$L_{n,b}(u, \theta) := \frac{1}{n} \sum_{t=p+1}^{n} K_b\left(\frac{t}{n} - u\right) \cdot \ell_{t,n}(\theta).$$

For $u \in [0, 1]$, the estimator of $\theta_0(u)$ is defined via

$$\hat{\theta}_b(u) := \arg\min_{\theta \in \Theta} L_{n,b}(u, \theta). \tag{33}$$

Asymptotic results

We will now discuss conditions such that $\hat{\theta}_b(\cdot)$ is consistent and asymptotically normal. A convenient way to formulate these results is to make a structural assumption on ℓ: We suppose that ℓ is Lipschitz continuous in its components with at most polynomially increasing Lipschitz constant. To make this more precise, we introduce the class $\tilde{\mathcal{L}}_{p+1}(M, C)$ of functions using the definition of $\mathcal{L}_r(M, C)$ from Definition 2.4.

Definition 5.1 (The class $\tilde{\mathcal{L}}_{p+1}(M, C)$). We say that a function $g : \mathbb{R}^{p+1} \times \Theta \to \mathbb{R}$ is in the class $\tilde{\mathcal{L}}_{p+1}(M, C)$ with $C = (C_z, C_\theta)$ and constants $C_z, C_\theta \geq 0$ and $M \geq 0$ if for all $z \in \mathbb{R}^{p+1}$, $\theta \in \Theta$ it holds that $g(\cdot, \theta) \in \mathcal{L}_{p+1}(M, C_z)$ and $g(z, \cdot) \in \mathcal{L}_d(0, C_\theta(1 + |z|_1^{M+1}))$.

As in Section 2, a generalization to Hoelder-type conditions (7) in the first component of g is possible. It turns out in Theorem 5.2 that the (pointwise) consistency of $\hat{\theta}_b$ can be obtained by posing conditions on the likelihood of the corresponding stationary process which is defined via $L(u, \theta) := \mathbb{E}[\tilde{\ell}_t(u, \theta)]$ with $\tilde{\ell}_t(u, \theta) := \ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta)$. Especially if ℓ is taken to be of the form (32) with f_ε the standard Gaussian density, the properties of $L(u, \theta)$ are usually well known from the maximum likelihood theory of the stationary process $X_t(\theta)$ and therefore are easy to verify (see also Example 5.5).

To prove consistency, we have to inspect $L_{n,b}(u, \theta)$ which consists of summands of the form $\ell(X_{t,n}, Y_{t-1,n}, \theta)$. Since $\ell \in \tilde{\mathcal{L}}_{p+1}(M, C)$, these terms behave like polynomials of degree $M + 1$ in $X_{t,n}$. We mainly need the law of large numbers Proposition 2.7(ii) and the statements about deterministic bias expansions Proposition 3.3. The conditions therein require Assumption 2.1(S1) with $q = M + 1$. Translated to the Markov process setting in this section, we have to assume Assumption 4.1(L1), (L4) with $q = M + 1$ by the results from Section 4.

Theorem 5.2 (Pointwise and uniform consistency of $\hat{\theta}_b$). *Let Assumption 2.6 hold. Assume that $\ell \in \tilde{\mathcal{L}}_{p+1}(M, C)$ for some $M \geq 0$. Suppose that Assumption 4.1(L1), (L4) holds with some $1 \geq \alpha > 0$ and $q = M + 1$.*

Furthermore suppose that for all $u \in [0, 1]$, $\theta_0(u) \in \text{int}(\Theta)$ is the unique minimizer of $L(u, \theta)$ over $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^d$ is a compact set. Then:

(i) *For all $u \in (0, 1)$ with $b \to 0$ and $bn \to \infty$:*

$$\hat{\theta}_b(u) \xrightarrow{p} \theta_0(u).$$

(ii) *If additionally $q > M + 1$ and $b = o(n^{1 - \frac{M+1}{q}})$ and $\theta_0(\cdot)$ is continuous, we have*

$$\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} \left| \hat{\theta}_b(u) - \theta_0(u) \right| \xrightarrow{p} 0.$$

Remark 5.3. Note that in nearly all cases, the conditions of Assumption 4.1(L4) assumed in Theorem 5.2 implicitly impose a Hoelder continuity condition on $\theta_0(\cdot)$; see also Example 5.5.

Proof of Theorem 5.2. (i) For fixed $u \in [0, 1]$ and $\theta \in \Theta$, it holds that $\ell(\cdot, \cdot, \theta) \in \mathcal{L}_{p+1}(M, C_z)$. Application of Theorem 2.7(ii) (see also Remark 2.8(ii)) leads to

$$L_{n,b}(u, \theta) = \frac{1}{n} \sum_{t=1}^{n} K_b\left(\frac{t}{n} - u\right) \cdot \ell(X_{t,n}, Y_{t-1,n}, \theta) \xrightarrow{p} \mathbb{E}\ell\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta\big) = L(u, \theta).$$

The function $\theta \mapsto L(u, \theta)$ is continuous since

$$\left| L(u, \theta) - L(u, \theta') \right| \le \left\| \ell\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta\big) - \ell\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta'\big) \right\|_1$$

$$\le C_\theta \cdot \left| \theta - \theta' \right|_1 \cdot \left(1 + \left(\sum_{j=0}^{p} \left\| \tilde{X}_t(u) \right\|_{M+1} \right)^{M+1} \right).$$

It remains to show stochastic equicontinuity of $L_{n,h}(u, \theta)$: Define $h : \mathbb{R}^{p+1} \to \mathbb{R}$, $h(z) = C_\theta(1 + |z|_1^{M+1})$. Fix $\eta > 0$. We have

$$\left| L_{n,b}(u, \theta) - L_{n,b}(u, \theta') \right| \le \left| \theta - \theta' \right|_1 \cdot \frac{1}{n} \sum_{t=1}^{n} \left| K_b\left(\frac{t}{n} - u\right) \right| \cdot h(X_{t,n}, Y_{t-1,n}).$$

Obviously, $h \in \mathcal{L}_{p+1}(M, C)$ with some constant $C > 0$. Application of Proposition 2.7(ii) to $K / \int K \, dx$ and h (see also Remark 2.8(ii)) yields for all $u \in (0, 1)$:

$$\frac{1}{n} \sum_{t=1}^{n} \left| K_b\left(\frac{t}{n} - u\right) \right| \cdot h(X_{t,n}, Y_{t-1,n}) \xrightarrow{p} \int |K| \, dx \cdot \mathbb{E}h\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u)\big) =: c(u). \qquad (34)$$

Choosing $\delta = \frac{\eta}{2c(u)}$ yields

$$\mathbb{P}\left(\sup_{|\theta - \theta'|_1 \le \delta} \left| L_{n,b}(u, \theta) - L_{n,b}(u, \theta') \right| > \eta \right)$$

$$\le \mathbb{P}\left(\left| \frac{1}{n} \sum_{t=1}^{n} \left| K_b\left(\frac{t}{n} - u\right) \right| \cdot h(X_{t,n}, Y_{t-1,n}) - c(u) \right| > c(u) \right) \to 0 \qquad (n \to \infty).$$

This gives $\sup_{\theta \in \Theta} |L_{n,b}(u, \theta) - L(u, \theta)| \overset{p}{\to} 0$. By standard arguments (cf. [26], Theorem 5.7), the proof is complete.

To prove (ii), we apply Theorem 2.7(iii) on $\ell(X_{t,n}, Y_{t-1,n}, \theta)$ with $\tilde{q} = \frac{q}{M+1} > 1$ (see also Remark 2.8(ii)) to obtain for each $\theta \in \Theta$ that

$$\sup_{u \in [0,1]} \left| L_{n,b}(u, \theta) - \mathbb{E} L_{n,b}(u, \theta) \right| = O_p\left(n^{\frac{M+1}{q} - 1} b^{-1} \right).$$

By Proposition 3.3 and the bounded variation of K, we have $\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} |\mathbb{E} L_{n,b}(u, \theta) - L(u, \theta)| = O(b^\alpha) + O((nb)^{-1})$, which yields $\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} |L_{n,b}(u, \theta) - L(u, \theta)| \overset{p}{\to} 0$. Similarly we can strengthen (34) to

$$\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} \left| \frac{1}{n} \sum_{t=1}^{n} \left| K_b\left(\frac{t}{n} - u \right) \right| \cdot h(X_{t,n}, Y_{t-1,n}) - c(u) \right| \overset{p}{\to} 0.$$

Now define $c := \inf_u c(u) > 0$ (by continuity of $c(\cdot)$). Choosing $\delta = \frac{\eta}{2c}$ yields

$$\mathbb{P}\left(\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} \sup_{|\theta - \theta'|_1 \leq \delta} \left| L_{n,b}(u, \theta) - L_{n,b}(u, \theta') \right| > \eta \right)$$

$$\leq \mathbb{P}\left(\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} \left| \frac{1}{n} \sum_{t=1}^{n} \left| K_b\left(\frac{t}{n} - u \right) \right| \cdot h(X_{t,n}, Y_{t-1,n}) - c(u) \right| > c \right) \to 0 \qquad (n \to \infty).$$

So we have seen that $\sup_{u \in [\frac{b}{2}, 1 - \frac{b}{2}]} \sup_{\theta \in \Theta} |L_{n,b}(u, \theta) - L(u, \theta)| \overset{p}{\to} 0$. Standard arguments give the result (see also the Supplementary Material [7]). □

We now provide a central limit theorem for $\hat{\theta}_b$ including a bias decomposition. Let ∇ denote the derivative with respect to θ. To use a standard Taylor expansion from M-estimation theory, we need the existence of $\nabla \ell \in \tilde{\mathcal{L}}_p(M', C')$ and $\nabla^2 \ell \in \tilde{\mathcal{L}}_p(M'', C'')$. To apply the local central limit theorem 2.10 to $\nabla L_{n,b}(u, \theta)$, we additionally need Assumption (S2) with $q = 2(M' + 1)$ which is fulfilled if Assumption 4.1(L2) is valid for $q = 2(M' + 1)$.

Theorem 5.4 (A central limit theorem for $\hat{\theta}_b$). *Additionally to Theorem 5.2(i), suppose that ℓ is twice continuously differentiable w.r.t. θ and*

- *$\nabla \ell \in \tilde{\mathcal{L}}_{p+1}(M', C')$ for some $M' \geq 0$, $\nabla^2 \ell \in \tilde{\mathcal{L}}_{p+1}(M'', C'')$ for some $M'' \geq 0$,*
- *Assumption 4.1(L1), (L4) is fulfilled with $q = \max\{2(M' + 1), M'' + 1\}$ and some $1 \geq \alpha' > 0$, Assumption 4.1(L2) is fulfilled with $q = 2(M' + 1)$.*

Assume that the model is correct in the weak sense that $\mathbb{E}[\nabla \tilde{\ell}(u, \theta_0(u)) | \mathcal{F}_{t-1}] = 0$, i.e. $\nabla \tilde{\ell}_t(u, \theta_0(u))$ is a martingale difference sequence with respect to $(\mathcal{F}_t)$. Let $b \to 0$, $nb \to \infty$ and $bn^{1-2\alpha} = o(1)$.

(i) *Then we have for $nb^{1+2\alpha'} = o(1)$:*

$$\sqrt{nb}\big(\hat{\theta}_b(u) - \theta_0(u)\big) \overset{d}{\to} N\left(0, \int K(x)^2 \, dx \cdot V(u)^{-1} I(u) V(u)^{-1}\right), \qquad (35)$$

where $I(u) := \mathbb{E}[\nabla\tilde{\ell}_t(u, \theta_0(u))\nabla\tilde{\ell}_t(u, \theta_0(u))']$ and $V(u) := \nabla^2 L(u, \theta_0(u))$ is assumed to be positive definite.

(ii) *If additionally $\nabla\ell$ is continuously differentiable and Assumption 4.1(L3) is fulfilled for $q = M' + 1$, then we have for $nb^3 = O(1)$:*

$$\sqrt{nb}\left(\hat{\theta}_b(u) - \theta_0(u) - b \cdot V^{-1}(u)\mathbb{E}\partial_u\nabla\ell\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta_0(u)\big) \cdot \int K(x)x \, dx\right)$$

$$\overset{d}{\to} N\left(0, \int K(x)^2 \, dx \cdot V(u)^{-1} I(u) V(u)^{-1}\right),$$

so the result (35) remains true if K is symmetric.

Proof of Theorem 5.4. The conditions on $\nabla^2\ell$ imply that the function $u \mapsto \nabla^2 L(u, \theta) = \mathbb{E}[\nabla^2\tilde{\ell}_t(u, \theta)]$ is continuous. Note that by Theorem 2.10, we have

$$\sqrt{nb}\nabla L_{n,b}\big(u, \theta_0(u)\big)$$

$$= \frac{1}{\sqrt{nb}}\sum_{t=p+1}^{n} K\left(\frac{t/n - u}{b}\right)\big(\nabla\ell\big(X_{t,n}, Y_{t-1,n}, \theta_0(u)\big) - \mathbb{E}\nabla\ell\big(X_{t,n}, Y_{t-1,n}, \theta_0(u)\big)\big)$$

$$\overset{d}{\to} N\left(0, \int K(x)^2 \, dx \cdot \sigma^2(u)\right),$$

where $\sigma^2(u) = \|\sum_{l=0}^{\infty} P_0\nabla\tilde{\ell}_t(u, \theta_0(u))\|_2^2 = I(u)$ by the martingale difference property. Furthermore,

$$\frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\left(\frac{t/n - u}{b}\right)\mathbb{E}\nabla\ell\big(X_{t,n}, Y_{t-1,n}, \theta_0(u)\big)$$

$$= \frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\left(\frac{t/n - u}{b}\right)\big(\mathbb{E}\nabla\ell\big(X_{t,n}, Y_{t-1,n}, \theta_0(u)\big) - \mathbb{E}\nabla\ell\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta\big)\big)$$

$$= \frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\left(\frac{t/n - u}{b}\right)\big(\mathbb{E}\nabla\ell\big(\tilde{X}_t(t/n), \tilde{Y}_{t-1}(t/n), \theta_0(u)\big) - \mathbb{E}\nabla\ell\big(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta\big)\big)$$

$$+ O\big(\sqrt{n^{1-2\alpha}b}\big).$$

Proposition 3.3 gives that the first term is $O(\sqrt{nb^{1+2\alpha'}})$ in the case of (i). In case of (ii), the first term has the form

$$\frac{1}{\sqrt{nb}} \sum_{t=1}^{n} K\left(\frac{t/n-u}{b}\right) \cdot \left(\frac{t}{n}-u\right) \cdot \mathbb{E}\left[\partial_u \nabla \ell\left(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta\right)\right]\Big|_{\theta=\theta_0(u)}$$

$$+ O\left((nb)^{-1/2}\right) + o\left(\sqrt{nb^3}\right)$$

and is $o(\sqrt{nb^3}) + O((nb)^{-1/2})$ if K is symmetric. Since $\nabla^2 \ell$ fulfills the same assumptions as ℓ in Theorem 5.2, we can mimic its proof and obtain

$$\sup_{\theta \in \Theta} \left|\nabla^2 L_{n,b}(u,\theta) - \nabla^2 L(u,\theta)\right| \xrightarrow{P} 0.$$

By continuity of $\theta \mapsto \nabla^2 L(u,\theta)$, we obtain for each sequence $\tilde{\theta}_n \xrightarrow{P} \theta_0(u)$ that

$$\left|\nabla^2 L_{n,b}(u,\tilde{\theta}_n) - V(u)\right| \leq \left|\nabla^2 L_{n,b}(u,\tilde{\theta}_n) - \nabla^2 L(u,\tilde{\theta}_n)\right| + \left|\nabla^2 L(u,\tilde{\theta}_n) - \nabla^2 L\left(u,\theta_0(u)\right)\right|$$

$$\xrightarrow{P} 0.$$

Standard arguments now give the result. $\square$

The results of Theorem 5.4(ii) show that under the existence of derivative processes, one can choose the MSE-optimal rate $b \sim n^{-1/3}$ for the bandwidth, keeping $\hat{\theta}_b(u)$ still asymptotically unbiased. This result can be used in several applications, for instance, for bootstrapping $X_{t,n}$ via the recursion (31) with estimated errors $\hat{\varepsilon}_t = H(X_{t,n}, Y_{t-1,n}, \hat{\theta}_b(t/n)), t = p+1, \ldots, n$.

An important special case is the case of Gaussian conditional likelihoods combined with non-linear autoregressive models. Specific examples for these are given in Example 2.2.

Example 5.5 (Nonlinear autoregressive models). In this example, we discuss the model $\tilde{G}_\varepsilon(y,\theta) = \mu(y,\theta) + \sigma(y,\theta)\varepsilon$, where $\mu, \sigma : \mathbb{R}^p \times \Theta \to \mathbb{R}$ satisfy

$$\sup_\theta \sup_{y \neq y'} \frac{|\mu(y,\theta) - \mu(y',\theta)|}{|y-y'|_{\chi,1}} + \sup_\theta \sup_{y \neq y'} \frac{|\sigma(y,\theta) - \sigma(y',\theta)|}{|y-y'|_{\chi,1}} \|\varepsilon_0\|_2 \leq 1 \tag{36}$$

with some $\chi \in \mathbb{R}_{\geq 0}^p$ with $|\chi|_1 < 1$. Assume that $\mathbb{E}\varepsilon_0 = 0$ and $\mathbb{E}\varepsilon_0^2 = 1$ and that θ_0 is Hoelder-continuous with exponent α. Then Assumption 4.1(L2) is fulfilled with $q = 2$.

If we choose f_ε to be the standard Gaussian density, we obtain from (32):

$$\ell(x,y,\theta) = \frac{1}{2}\left(\frac{x - \mu(y,\theta)}{\sigma(y,\theta)}\right)^2 - \frac{1}{2}\log \sigma^2(y,\theta) + \text{const.} \tag{37}$$

Furthermore assume that

$$\sup_y \sup_{\theta \neq \theta'} \frac{|\mu(y,\theta) - \mu(y,\theta')|}{|\theta-\theta'|_1 \cdot (1+|y|_1)} < \infty, \qquad \sup_y \sup_{\theta \neq \theta'} \frac{|\sigma(y,\theta) - \sigma(y,\theta')|}{|\theta-\theta'|_1 \cdot (1+|y|_1)} < \infty. \tag{38}$$

Let $\sigma(\cdot) \geq \delta_\sigma$ be uniformly bounded from below with some $\delta_\sigma > 0$. Then $\ell \in \tilde{\mathcal{L}}_{p+1}(1, C)$ with some $C > 0$, and Assumption 4.1(L1), (L4) is fulfilled with $q = 2$ and α from above.

Fix $u \in [0, 1]$. Suppose that

$$\mu\big(\tilde{Y}_{t-1}(u), \theta\big) = \mu\big(\tilde{Y}_{t-1}(u), \theta_0(u)\big) \quad \text{and} \quad \sigma\big(\tilde{Y}_{t-1}(u), \theta\big) = \sigma\big(\tilde{Y}_{t-1}(u), \theta_0(u)\big) \quad \text{a.s.}$$

implies $\theta = \theta_0(u)$. Then $\theta \mapsto L(u, \theta)$ has a unique minimum in $\theta = \theta_0(u)$ since $\log(x) \leq x - 1$ if and only if $x = 1$ and $x^2 \geq 0$ if and only if $x = 0$ and, omitting the argument $\tilde{Y}_{t-1}(u)$,

$$2\big(L(u, \theta) - L\big(u, \theta_0(u)\big)\big) = \mathbb{E}\left(\frac{\mu(\theta) - \mu(\theta_0(u))}{\sigma(\theta)}\right)^2 + \mathbb{E}\left[\log\frac{\sigma(\theta)^2}{\sigma(\theta_0(u))^2} - 1 + \frac{\sigma(\theta_0(u))^2}{\sigma(\theta)^2}\right] \geq 0.$$

If additionally Θ is compact and $\theta_0(u) \in \text{int}(\Theta)$, the assumptions of Theorem 5.2 are fulfilled and we obtain for $\hat{\theta}_b$ defined by (33):

$$\hat{\theta}_b(u) \xrightarrow{p} \theta_0(u).$$

We now will show asymptotic normality of $\hat{\theta}_b$. To keep the presentation simple, we will assume $\sigma(\cdot, \cdot) \equiv 1$, $\mathbb{E}\varepsilon_0^4 < \infty$ and replace $\mathbb{E}\varepsilon_0^2 = 1$ by $\mathbb{E}\varepsilon_0^2 = \sigma_0^2 > 0$. Note that Assumption 4.1(L2) is fulfilled with $q = 4$. Then, omitting the arguments (y, θ) of μ, we have

$$\nabla\ell(x, y, \theta) = -(x - \mu)\nabla\mu, \qquad \nabla^2\ell(x, y, \theta) = \nabla\mu \cdot \nabla\mu' - (x - \mu)\nabla^2\mu.$$

This shows $\mathbb{E}[\nabla\ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta_0(u))|\mathcal{F}_{t-1}] = 0$ and $I(u) = \mathbb{E}[\nabla\ell \cdot \nabla\ell'] = \sigma_0^2\mathbb{E}[\nabla\mu \cdot \nabla\mu'] = \sigma_0^2 V(u)$ with $V(u) := \nabla^2 L(u, \theta_0(u))$. If additionally

$$\sup_\theta \sup_{y \neq y'} \frac{|\nabla\mu(y, \theta) - \nabla\mu(y', \theta)|_1}{|y - y'|_1} < \infty, \qquad \sup_y \sup_{\theta \neq \theta'} \frac{|\nabla\mu(y, \theta) - \nabla\mu(y, \theta')|_1}{|\theta - \theta'|_1(1 + |y|_1)} < \infty \quad (39)$$

and similar assumptions are fulfilled for $\nabla^2\mu$, then we have $\nabla\ell, \nabla^2\ell \in \tilde{\mathcal{L}}_{p+1}(1, C')$ with some $C' > 0$. This shows that all conditions of the first part of Theorem 5.4 are fulfilled and we obtain for $b \to 0$, $nb \to \infty$ and $nb^3 = o(1)$:

$$\sqrt{nb}\big(\hat{\theta}_b(u) - \theta_0(u)\big) \xrightarrow{d} N\big(0, \sigma_0^2 \cdot V(u)^{-1}\big). \quad (40)$$

If additionally, μ, $\nabla\mu$ and θ_0 are continuously differentiable and

$$\sup_\theta \sup_{y \neq y'} \frac{|\partial_i\mu(y, \theta) - \partial_i\mu(y', \theta)|_1}{|y - y'|_1} < \infty \qquad (i = 1, 2), \quad (41)$$

then $\nabla\ell$ is continuously differentiable and Assumption 4.1(L3) is fulfilled with $q = 2$. If K is symmetric, all conditions of the second part of Theorem 5.4 are fulfilled and we obtain (40) even if $nb^3 = O(1)$.

We close this section by using the results of Example 5.5 in a more specific example of the tvExpAR(1) process which is a locally stationary version of the ExpAR(1) process discussed in [10]. Up to now, there is no asymptotic theory available for parameter estimators in this model; we show that our theory immediately provides consistency and asymptotic normality of the corresponding maximum likelihood estimator.

Example 5.6 (**Maximum likelihood estimation in the** tvExpAR(1) **process**)**.** Assume that there exists $\theta_0 : [0, 1] \to \Theta$ (where the image of θ_0 is in the interior of Θ) with $\Theta := \{\theta \in \mathbb{R} : 0 \leq \theta \leq \rho\}$ and some fixed $\rho > 0$, $0 < |a_0| < 1$ such that

$$X_{t,n} = a_0 \exp\left(-\theta_0\left(\frac{t}{n}\right)X_{t-1,n}^2\right)X_{t-1,n} + \varepsilon_t, \qquad t = 1, \ldots, n.$$

Assume that $\mathbb{E}\varepsilon_0 = 1$, $\mathbb{E}\varepsilon_0^2 = \sigma_0^2 > 0$ and $\mathbb{E}\varepsilon_0^4 < \infty$. It is easily seen that this model fulfills the smoothness assumptions (36), (38), (39) and (41) with $\mu(y, \theta) := a_0 \exp(-\theta y^2)y$ and $\sigma(\cdot, \cdot) \equiv 1$. Let $\tilde{X}_t(u)$ denote the corresponding stationary approximation of $X_{t,n}$. Identifiability of θ is obtained due to

$$\mathbb{E}\left[\left(\mu\left(\tilde{X}_t(u), \theta\right) - \mu\left(\tilde{X}_t(u), \theta'\right)\right)^2\right] \geq a_0^2 \mathbb{E}\left[\exp\left(-2\rho\tilde{X}_0(u)^2\right)\tilde{X}_0(u)^6\right] \cdot \left|\theta - \theta'\right|^2,$$

since $\mathbb{E}[\exp(-2\rho\tilde{X}_t(u)^2)\tilde{X}_t(u)^6] = 0$ would imply $\tilde{X}_t(u) = 0$ a.s. which is a contradiction to $\mathbb{E}[\tilde{X}_t(u)^2] \geq \sigma_0^2$ which follows from the recursion of $\tilde{X}_t(u)$. Let $\hat{\theta}_b(u)$ be defined by (33) based on the likelihood (37) and let Assumption 2.6 hold. We obtain for $b \to 0$, $bn \to \infty$:

$$\hat{\theta}_b(u) \xrightarrow{p} \theta_0(u),$$

and for $nb^3 = O(1)$:

$$\sqrt{nb}\left(\hat{\theta}_b(u) - \theta_0(u)\right) \xrightarrow{d} N\left(0, \sigma_0^2 V(u)^{-1}\right),$$

where $V(u) = a_0^2 \mathbb{E}[\exp(-2\theta_0(u)\tilde{X}_0(u)^2)\tilde{X}_0(u)^6]$.

6. Concluding remarks

In this paper, we have made some steps towards a general asymptotic theory for nonlinear locally stationary processes. A key role in our derivations is played by the local stationary approximation, the derivative process, the corresponding Taylor-expansion and the resulting differential calculus.

Just based on this local approximation we were able to prove laws of large numbers, a central limit theorem, and stochastic and deterministic bias approximations – results which have not been proved so far for general locally stationary processes. For example, for the global strong law of large numbers we need only the existence of the first order moment of the process. It should be noted that for these results we concluded from local assumptions to global results such as the strong law of large numbers and the central limit theorem. A simulation displayed in Figure 1 shows that the pointwise approximation of $X_{t,n}$ by $\tilde{X}_t(u)$ and $\partial_u \tilde{X}_t(u)$ works quite well.

We also showed that these results can be applied to a general nonlinear time series model with a nonstationary Markov structure which includes several nonlinear models. As another application we derived the asymptotic properties of the maximum likelihood estimator for such processes. The result is proved by applying the differential calculus of the derivative process.

Acknowledgements

We are very grateful to the associate editor and a referee whose comments lead to a considerable improvement of the paper. In particular the observation that in many cases differentiability in L^q is sufficient (see Remark 4.3) was pointed out by the referee. We also gratefully acknowledge support by Deutsche Forschungsgemeinschaft through the Research Training Group RTG 1653.

Supplementary Material

Supplement to "Towards a general theory for nonlinear locally stationary processes" (DOI: 10.3150/17-BEJ1011SUPP; .pdf). Supplement A: Proofs of Section 2, 4 and 5. This supplement contains the remaining proofs for Sections 2, 4 and 5. Supplement B: Discussion of the Assumptions 2.3(M1), (M2). This supplement contains another counterexample where Assumption 2.3(M1) is satisfied but not (M2).

References

[1] Chen, X., Xu, M. and Wu, W.B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist*. **41** 2994–3021. MR3161455

[2] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist*. **25** 1–37. MR1429916

[3] Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist*. **28** 1762–1794. MR1835040

[4] Dahlhaus, R. (2012). Locally stationary processes. In *Time Series Analysis*: *Methods and Applications* (T.S. Rao, S.S. Rao and C.R. Rao, eds.). *Handbook of Statistics* **30** 351–413. Amsterdam: Elsevier.

[5] Dahlhaus, R. and Giraitis, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Series Anal*. **19** 629–655. MR1665941

[6] Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797

[7] Dahlhaus, R., Richter, S. and Wu, W.B. (2019). Supplement to "Towards a general theory for nonlinear locally stationary processes." DOI:10.3150/17-BEJ1011SUPP.

[8] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist*. **34** 1075–1114. MR2278352

[9] Eichler, M., Motta, G. and von Sachs, R. (2011). Fitting dynamic factor models to non-stationary time series. *J. Econometrics* **163** 51–70. MR2803666

[10] Jones, D.A. (1978). Nonlinear autoregressive processes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci*. **360** 71–95. MR0501672

[11] Koo, B. and Linton, O. (2012). Estimation of semiparametric locally stationary diffusion models. *J. Econometrics* **170** 210–233. MR2955950

[12] Kreiss, J.-P. and Paparoditis, E. (2015). Bootstrapping locally stationary processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 267–290. MR3299408

[13] Liu, W., Xiao, H. and Wu, W.B. (2013). Probability and moment inequalities under dependence. *Statist. Sinica* **23** 1257–1272. MR3114713

[14] Martin, W. and Flandrin, P. (1985). Wigner-Ville spectral analysis of nonstationary processes. *IEEE Trans. Acoust. Speech Signal Process.* **33** 1461–1470.

[15] Motta, G., Hafner, C.M. and von Sachs, R. (2011). Locally stationary factor models: Identification and nonparametric estimation. *Econometric Theory* **27** 1279–1319. MR2868840

[16] Palma, W. and Olea, R. (2010). An efficient estimator for locally stationary Gaussian long-memory processes. *Ann. Statist.* **38** 2958–2997. MR2722461

[17] Preuß, P., Vetter, M. and Dette, H. (2013). A test for stationarity based on empirical processes. *Bernoulli* **19** 2715–2749. MR3160569

[18] Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163. MR2472010

[19] Roueff, F. and von Sachs, R. (2011). Locally stationary long memory estimation. *Stochastic Process. Appl.* **121** 813–844. MR2770908

[20] Sedro, J. (2017). A regularity result for fixed points, with applications to linear response. Preprint. Available at arXiv:1705.04078.

[21] Sergides, M. and Paparoditis, E. (2008). Bootstrapping the local periodogram of locally stationary processes. *J. Time Series Anal.* **29** 264–299. MR2392774

[22] Sergides, M. and Paparoditis, E. (2009). Frequency domain tests of semi-parametric hypotheses for locally stationary processes. *Scand. J. Stat.* **36** 800–821. MR2573309

[23] Shao, X. and Wu, W.B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105

[24] Subba Rao, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. MR2285698

[25] Truquet, L. (2016). Local stationarity and time-inhomogeneous Markov chains. Preprint. Available at arXiv:1610.01290.

[26] van der Vaart, A.W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247

[27] Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MR3097614

[28] Wu, W. and Zhou, Z. (2017). Nonparametric inference for time-varying coefficient quantile regression. *J. Bus. Econom. Statist.* **35** 98–109. MR3591532

[29] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

[30] Wu, W.B. (2007). *M*-Estimation of linear models with dependent errors. *Ann. Statist.* **35** 495–521. MR2336857

[31] Wu, W.B. (2008). Empirical processes of stationary sequences. *Statist. Sinica* **18** 313–333. MR2384990

[32] Wu, W.B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226. MR2812816

[33] Wu, W.B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528

[34] Zhou, Z. (2014). Inference of weighted V-statistics for nonstationary time series and its applications. *Ann. Statist.* **42** 87–114. MR3161462

[35] Zhou, Z. (2014). Nonparametric specification for non-stationary time series regression. *Bernoulli* **20** 78–108. MR3160574

[36] Zhou, Z. and Wu, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444