

SIEVE BOOTSTRAP FOR SMOOTHING IN NONSTATIONARY TIME SERIES¹

BY PETER BÜHLMANN

ETH Zürich

We propose a sieve bootstrap procedure for time series with a deterministic trend. The sieve for constructing the bootstrap is based on nonparametric trend estimation and autoregressive approximation for some noise process. The bootstrap scheme itself does i.i.d. resampling of estimated innovations from fitted autoregressive models.

We show the validity and indicate second-order correctness of such sieve bootstrap approximations for the limiting distribution of nonparametric linear smoothers. The resampling can then be used to construct nonparametric confidence intervals for the underlying trend. In particular, we show asymptotic validity for constructing confidence bands which are simultaneous within a neighborhood of size of the order of the smoothing bandwidth.

Our resampling procedure yields satisfactory results in a simulation study for finite sample sizes. We also apply it to the longest series of total ozone measurements from Arosa (Switzerland) and find a significant decreasing trend.

1. Introduction. We are considering the problem of interval estimation for a trend in a time series. More precisely, we consider the model

$$(1.1) \quad Y_t = s(t) + Z_t, \quad t \in \mathbb{Z},$$

where $\{s(t)\}_{t \in \mathbb{Z}}$ is a deterministic trend (or signal) and $\{Z_t\}_{t \in \mathbb{Z}}$ a stationary noise process with mean zero. Various estimators, such as least squares in a parametric trend model or nonparametric smoothers, are known for recovering such trends. For constructing confidence intervals one usually relies on asymptotic normal theory.

We focus here on nonparametric trend estimation, where $s(t) = m(t/n)$, $t = 1, \dots, n$, with $m: [0, 1] \rightarrow \mathbb{R}$ a smooth function and n denoting the sample size. In the independent setup, where the noise is white, bootstrap methods have been proposed for interval estimation of $m(\cdot)$ [cf. Härdle and Bowman (1988), Härdle and Marron (1991) and Hall (1992)]. There are several reasons for this.

REASON 1. The bootstrap approximation, when used correctly for asymptotically pivotal quantities, is second order correct yielding better coverage probability for finite sample size [cf. Hall (1992)]. More than that, we argue

Received January 1996; revised October 1997.

¹Supported in part by the Swiss National Science Foundation. Most of the work was done while staying at the University of California, Berkeley.

AMS 1991 subject classifications. Primary 62G09; secondary 62G07, 62M10.

Key words and phrases. AIC, $AR(\infty)$, ARMA, autoregressive approximation, kernel smoothing, $MA(\infty)$, nonparametric regression, second-order correctness, simultaneous confidence band.

for a moment why and when the skewness of the simpler, but asymptotically nonpivotal, quantity $(nh)^{1/2}(\hat{m}(x) - \mathbb{E}[\hat{m}(x)])$ plays a dominant role and hence should be taken into account as well. Here, $\hat{m}(x)$ is a nonparametric smoother and h a bandwidth. Formally we can write

$$\begin{aligned} F(u) &= \mathbb{P}[(nh)^{1/2}(\hat{m}(x) - \mathbb{E}[\hat{m}(x)]) \leq u] \\ &= \Phi_{\mathcal{N}}\left(\frac{u}{\sigma_{as}}\right) + \varphi_{\mathcal{N}}\left(\frac{u}{\sigma_{as}}\right) \frac{1}{6} \left(\left(\frac{u}{\sigma_{as}}\right)^2 - 1 \right) (nh)^{-1/2} \mu_{3;as} + o((nh)^{-1/2}), \end{aligned}$$

$n \rightarrow \infty,$

where $\sigma_{as}^2, \mu_{3;as}$ are the asymptotic variance and (unstandardized) skewness of $(nh)^{1/2}\hat{m}(x)$. Let us ignore here the bias term $(nh)^{1/2}(\mathbb{E}[\hat{m}(x)] - m(x))$ which comes into play for constructing confidence regions for $m(x)$. From the formula above we see that an estimated normal approximation deals with an error

$$F(u) - \Phi_{\mathcal{N}}\left(\frac{u}{\hat{\sigma}_{as}}\right) = O_P(|\hat{\sigma}_{as}^2 - \sigma_{as}^2|) + O((nh)^{-1/2}),$$

where $\hat{\sigma}_{as}^2$ is an estimate of σ_{as}^2 . If the error in variance estimation $\hat{\sigma}_{as}^2 - \sigma_{as}^2 = o_P((nh)^{-1/2})$ is of smaller order than the $O((nh)^{-1/2})$ skewness term (this is typically achievable if the observations are independent or with a sufficiently fast decay of dependence, e.g., sufficiently fast decay of mixing coefficients), then, assuming $\mu_{3;as} \neq 0$,

$$F(u) - \Phi_{\mathcal{N}}\left(\frac{u}{\hat{\sigma}_{as}}\right) = O_P((nh)^{-1/2}),$$

and not $O_P(|\hat{\sigma}_{as}^2 - \sigma_{as}^2|)$. In this sense, the estimated normal approximation is not even first order correct, meaning not of the order $O_P(|\hat{\sigma}_{as}^2 - \sigma_{as}^2|)$: the leading error term is the skewness which is not considered by the estimated normal approximation. This is different from many classical problems, for example, the mean for estimating the expected value of i.i.d. observations. In our setup, it thus makes sense (even for the asymptotically nonpivotal quantity) to take the leading skewness term into account as well. This is automatically the case for the bootstrap approximation which is asymptotically correct for the variance σ_{as}^2 and the skewness term $(nh)^{-1/2}\mu_{3;as}$, too. One should also note that, in nonparametric smoothing, the problem of skewness becomes more severe, when estimating a “rough” underlying trend. The heuristics goes as follows: to calibrate bias against variance, a smaller smoothing bandwidth should be used which in turn would make the skewness term of the order $O((nh)^{-1/2})$ bigger, by the small value of h .

REASON 2. The bootstrap has the potential to yield some kinds of simultaneous confidence bands [cf. Härdle and Marron (1991)].

REASON 3. The bootstrap is able to correct for bias [cf. Härdle and Marron (1991)].

The time series context is more complex and there is even more need for re-sampling procedures. Already the normal approximation requires estimation of the spectral density at zero $f_Z(0)$ of the error process in (1.1), because the limiting variance of a trend smoother, properly standardized, is proportional to $f_Z(0) = (1/(2\pi)) \sum_{k=-\infty}^{\infty} \text{Cov}(Z_0, Z_k)$. Generally, this cannot be estimated in a \sqrt{n} -consistent way, implying that the variance estimate of a trend smoother will not be \sqrt{n} -consistent. The skewness of a trend smoother is asymptotically proportional to the bispectral density of the noise process at zero, which is a more complicated functional of the noise than just the skewness $\mathbb{E}[Z_t^3]$ of a one-dimensional marginal. Also, the issue about selection of smoothing parameters becomes more delicate: Härdle and Bowman (1988) and also Eubank and Speckman (1993) use data-driven bandwidths for trend smoothing, but such automatic choices for smoothing parameters are more difficult and less straightforward in the time series case. As another problem we mention the construction of simultaneous confidence bands: one could try to adapt the extreme value approximation, which was developed by Bickel and Rosenblatt (1973) and refined and applied by Eubank and Speckman (1993) for independent observations. A plausible justification for this approach in the time series case can be deduced from Aldous' (1989) Poisson clumping heuristics. However, the rate of convergence to the limiting extreme value approximation is very slow. This is an unsatisfactory feature, also present in the independent case, in particular if simultaneity is of interest over a subset of the full range. Again, we would have to estimate the spectral density at zero $f_Z(0)$.

For coping with the difficulties above we propose a resampling technique for distribution estimation of trend smoothers. We construct a scheme which is asymptotically valid for the rich subclass of linear stationary processes $Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ which can be inverted and represented as an autoregressive process of order infinity [AR(∞)], that is, $\sum_{j=0}^{\infty} \phi_j Z_{t-j} = \varepsilon_t$, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ being an i.i.d. sequence with mean zero. We extend the sieve bootstrap for stationary AR(∞) processes [cf. Kreiss (1988) and Bühlmann (1997)] to nonstationary time series as in (1.1) with an AR(∞) noise process. The extension is based on a preliminary trend estimation yielding a residual process on which we then apply the stationary AR(∞) sieve bootstrap. The name originates from the idea of sieve estimation, here as autoregressive approximation for AR(∞) processes. Unlike Freedman's (1984) pioneering work for bootstrapping stationary finite order AR processes which has been further developed by Bose (1988), our approach has the advantage that no particular finite (semi)parametric model for the noise process is assumed. It is shown in Bühlmann (1997) that, for many linear processes, the stationary sieve bootstrap for AR(∞) models has generally a better performance than some nonparametric block-based bootstrap technique as proposed by Künsch (1989).

We develop a theory which reflects the three advantageous reasons of the bootstrap mentioned above. The resampling approximation reflects nonnormal features like skewness, which is of special interest when estimating a "rough" trend. Plug-in estimates for bias and variance are not necessary. We also focus on construction of simultaneous interval estimation and prove asymptotic

correctness of our bootstrap scheme for constructing simultaneous confidence bands within neighborhoods whose size is of the order of the bandwidth for smoothing. To our knowledge, this result is also new for the case of independent observations. The concept of neighborhood simultaneity is often more appropriate than simultaneity everywhere. In many applications, the aim is to construct simultaneous confidence regions over some particular range of interest. For any specified set G from a large class of sets, the bootstrap approach allows a straightforward construction for asymptotically correct (not conservative) confidence regions which are simultaneous over G . Finally, a simulation study indicates that the procedure is less sensitive to the choice of tuning parameters than approximations based on limiting normal distributions.

The article is organized as follows. In Section 2 we describe our general sieve bootstrap scheme; in Section 3 we give asymptotic results for bootstrapping nonparametric kernel smoothers; in Section 4 we focus on the construction of confidence bands; in Section 5 we report results from a simulation study and from an analysis of the longest series of total ozone measurements from Arosa, Switzerland. The theoretical arguments and proofs are given in Section 6.

2. The model and the resampling procedure. Consider the model (1.1), where $\{Z_t\}_{t \in \mathbb{Z}}$ is an $\text{AR}(\infty)$ process, that is,

$$(2.1) \quad \sum_{j=0}^{\infty} \phi_j Z_{t-j} = \varepsilon_t, \quad \phi_0 = 1, t \in \mathbb{Z},$$

where $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is an i.i.d. sequence with expectation $\mathbb{E}[\varepsilon_t] = 0$.

The object to recover is the deterministic trend, which is thought to be a function $s(\cdot): \mathbb{Z} \rightarrow \mathbb{R}$. Given observations Y_1, \dots, Y_n , there are various approaches.

2.1. A sieve bootstrap scheme. We develop here a general resampling method which is based on an extension of the bootstrap for stationary autoregressive processes of order infinity [$\text{AR}(\infty)$], [cf. Kreiss (1988), Bühlmann (1997) and Bickel and Bühlmann (1995)]. Assume that we have observations Y_1, \dots, Y_n , realizations of the model as given in (1.1) and (2.1).

STEP 1. Compute an estimate $\tilde{s}(t)$ for the unknown trend values $s(t)$, $t = 1, \dots, n$. Then form residuals

$$\hat{Z}_{t,n} = Y_t - \tilde{s}(t), \quad t = 1, \dots, n.$$

STEP 2. Assume $p = p(n) \rightarrow \infty$, $p(n) = o(n)$, $n \rightarrow \infty$. Fit an autoregressive model of order $p = p(n)$ to the residuals $\hat{Z}_{t,n}$, $t = 1, \dots, n$; that is, compute $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}$ based on $\{\hat{Z}_{t,n}\}_{t=1}^n$ for the autoregressive coefficients

ϕ_1, \dots, ϕ_p in (2.1). Then form another set of residuals

$$\hat{\varepsilon}_{t,n} = \sum_{j=0}^p \hat{\phi}_{j,n} \hat{Z}_{t-j,n}, \quad t = p+1, \dots, n, \quad \hat{\phi}_{0,n} = 1.$$

Denote by $\tilde{\varepsilon}_{t,n} = \hat{\varepsilon}_{t,n} - \bar{\varepsilon}_{\cdot,n}$, where $\bar{\varepsilon}_{\cdot,n} = (n-p)^{-1} \sum_{t=p+1}^n \hat{\varepsilon}_{t,n}$.

STEP 3. Denote by $F_{\tilde{\varepsilon},n}(\cdot) = (n-p)^{-1} \sum_{t=p+1}^n 1_{[\tilde{\varepsilon}_{t,n} \leq \cdot]}$ the empirical cumulative distribution function of $\{\tilde{\varepsilon}_{t,n}\}_{t=p+1}^n$. Now resample

$$\varepsilon_t^* \text{ i.i.d. } \sim F_{\tilde{\varepsilon},n}, \quad t \in \mathbb{Z} \text{ (or in a subset of } \mathbb{Z}\text{)}.$$

STEP 4. Generate the bootstrap error process $\{Z_t^*\}_t$, defined by

$$\sum_{j=0}^p \hat{\phi}_{j,n} Z_{t-j}^* = \varepsilon_t^*, \quad \hat{\phi}_{0,n} = 1, \quad t \in \mathbb{Z} \text{ (or in a subset of } \mathbb{Z}\text{)}.$$

Then generate bootstrap observations by setting

$$Y_t^* = \tilde{s}(t) + Z_t^*, \quad t = 1, \dots, n,$$

with $\{\tilde{s}(t)\}_{t=1}^n$ the same estimates as in Step 1.

A guideline for the approximating order $p = p(n)$ in Step 2 is given by the Akaike information criterion (AIC) in an increasing range $[0, C_n]$ with C_n growing as n increases. In the case of no deterministic trend, such an order selection has been studied by Shibata (1980): there, the increasing range is of the above form with $C_n = o(n^{1/2})$, $C_n^{-1} = o(1)$. As a practical guideline for the increasing range we suggest $C_n = 10 \log_{10}(n)$, which is also the S-PLUS default. The estimates $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}$ in Step 2 can be obtained by the Yule–Walker method. They are defined by

$$(2.2) \quad \hat{\Gamma}_{\hat{Z}} \hat{\Phi}_p = -\hat{\gamma}_{\hat{Z}},$$

where $\hat{\Gamma}_{\hat{Z}} = [\hat{R}_{\hat{Z}}(i-j)]_{i,j=1}^p$, $\hat{\gamma}_{\hat{Z}} = (\hat{R}_{\hat{Z}}(1), \dots, \hat{R}_{\hat{Z}}(p))'$, $\hat{\Phi}_p = (\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n})'$, $\hat{R}_{\hat{Z}}(j) = n^{-1} \sum_{t=1}^{n-|j|} \hat{Z}_{t,n} \hat{Z}_{t+|j|,n}$, $|j| \leq n-1$. For generating the bootstrap error process $\{Z_t^*\}_t$ in Step 4, we start the recursion with some starting values and wait until stationarity is reached. We implemented the algorithm with starting values being equal to some resampled innovations ε_t^* .

Estimating a trend $\tilde{s}(\cdot)$ in Step 1 and fitting an autoregressive model of growing order $p = p(n)$ in Step 2 is a sieve procedure for the true underlying trend $s(\cdot)$ and AR(∞) process; that is why we call this resampling scheme *sieve bootstrap* [cf. Bühlmann (1997)]. In contrast to the sieve bootstrap for stationary time series, the sieve here involves also the estimation of the trend.

All the bootstrap quantities are denoted by an asterisk. This sieve bootstrap allows us to resample observations Y_1^*, \dots, Y_n^* for various purposes. The bootstrapped trend estimates $\hat{s}^*(\cdot)$ are defined by the plug-in rule, that is, if $\hat{s}(t) = T_n(Y_1, \dots, Y_n)$ is a measurable function of the observations, then the bootstrap estimate is defined by $\hat{s}^*(t) = T_n(Y_1^*, \dots, Y_n^*)$. Note that the

estimate $\hat{s}(t)$ above is not necessarily the same as in Step 1 and Step 4 of our sieve bootstrap. For nonparametric smoothers, we rather would take for $\tilde{s}(t)$ oversmoothed estimates (see Section 3). Whenever (1.1) and (2.1) hold, the given bootstrap scheme has the potential to approximate the distribution of a variety of estimators.

3. Bootstrapping kernel smoothers for interval trend estimation.

We focus here on a nonparametric technique, given by kernel smoothing. Assume that the trend function is

$$(3.1) \quad s(t) = m(t/n), \quad t = 1, \dots, n,$$

for some function $m: [0, 1] \rightarrow \mathbb{R}$, where n denotes the sample size. Consider estimation of $m(\cdot)$ by kernel smoothing

$$(3.2) \quad \hat{m}(x) = (nh)^{-1} \sum_{t=1}^n K\left(\frac{x - t/n}{h}\right) Y_t, \quad x \in [0, 1],$$

where $h = h(n) = o(1)$, $h(n)^{-1} = o(n)$ $n \rightarrow \infty$, is a bandwidth parameter and $K(\cdot)$ a kernel function. This is the so-called Nadaraya–Watson kernel estimator [cf. Nadaraya (1964) and Watson (1964)]. Our results and procedures also hold for other linear smoothers

$$\hat{m}(x) = \sum_{t=1}^n w_{t,n}(x) Y_t, \quad \lim_{n \rightarrow \infty} \sum_{t=1}^n w_{t,n}(x) = 1,$$

such as local linear regression smoothers [cf. Fan (1993)], Gasser–Müller type kernel smoothing [Gasser and Müller (1979)] or linear wavelet smoothing [cf. Brillinger (1996)]. Linear kernel smoothing with dependent errors has been studied by, among others, Härdle and Tuan (1986), Hall and Hart (1990), Altman (1990), Truong (1991), Hart (1991, 1994) and Herrmann, Gasser and Kneip (1992).

We are going to discuss now the validity of the sieve bootstrap scheme for kernel estimators $\hat{s}(t) = \hat{m}(t/n)$, $t \in \mathbb{N}$, with $1 \leq t \leq n$. These estimates are known to exhibit edge effects. They can cause problems in the sieve bootstrap scheme, namely, in Steps 1 and 4 where a pilot estimate $\tilde{s}(t) = \tilde{m}(t/n)$ has to be computed. We will use for $\tilde{m}(\cdot)$ again a kernel smoother as in (3.2) but only in a region $[\delta, 1 - \delta]$ for some small $\delta > 0$. Thus, in Step 1 we will use pilot estimates

$$(3.3) \quad \tilde{s}(t) = \tilde{m}(t/n), \quad t = [\delta n] + 1, \dots, [(1 - \delta)n], \quad 0 < \delta < 1/2,$$

and apply the bootstrap scheme in the same way but now with the smaller number of Z -residuals $\hat{Z}_{[\delta n] + 1, n}, \dots, \hat{Z}_{[(1 - \delta)n], n}$.

By choosing the bandwidth of a mean square error optimal order $n^{-1/5}$, we have to deal with a nonnegligible bias,

$$(3.4) \quad (nh)^{1/2}(\mathbb{E}[\hat{m}(x)] - m(x)) \rightarrow B_{as}(x), \quad n \rightarrow \infty, \\ \text{with } 0 \leq B_{as}(x) = \lim_{n \rightarrow \infty} (nh)^{1/2} h^2 m^{(2)}(x) \int_{\mathbb{R}} x^2 K(x) dx / 2 < \infty$$

where $m^{(j)}(x) = (d^j/dx^j)m(x)$ for $j \in \mathbb{N}$.

Under some regularity conditions, asymptotic normality holds,

$$(3.5) \quad (nh)^{1/2}(\hat{m}(x) - m(x)) \rightarrow_d \mathcal{N}(B_{as}(x), \sigma_{as}^2), \quad n \rightarrow \infty,$$

with $0 < \sigma_{as}^2 = 2\pi f_Z(0) \int_{\mathbb{R}} K^2(x) dx < \infty$,

where $f_Z(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \text{Cov}(Z_0, Z_k) e^{-i\lambda k}$ is the spectral density of the noise process $\{Z_t\}_{t \in \mathbb{Z}}$.

For a successful approximation of this limiting normal distribution, the bootstrap should be able to estimate the asymptotic bias $B_{as}(x)$ as well. This can be achieved by using in Steps 1 and 4 of our sieve bootstrap scheme an oversmoothed estimate $\tilde{s}(t) = \tilde{m}(t/n)$, $t = [\delta n] + 1, \dots, [(1 - \delta)n]$, with $\tilde{m}(\cdot)$ of the same form as in (3.2), with the same kernel $K(\cdot)$, but with a larger bandwidth \tilde{h} than h . Then it will be possible that

$$(nh)^{1/2}(\mathbb{E}^*[\hat{m}^*(x)] - \tilde{m}(x)) - B_{as}(x) = o_P(1).$$

The intuitive reason for this is that, for a second order kernel K ,

$$\begin{aligned} \mathbb{E}[\hat{m}(x)] - m(x) &\sim h^2 m^{(2)}(x) C_K, \\ \mathbb{E}^*[\hat{m}^*(x)] - \tilde{m}(x) &\sim h^2 \tilde{m}^{(2)}(x) C_K, \end{aligned}$$

where $C_K = \int_{\mathbb{R}} x^2 K(x) dx / 2$. The convergence $\tilde{m}^{(2)}(x) - m^{(2)}(x) = o_P(1)$ is only possible for a bandwidth \tilde{h} with $\tilde{h} n^{1/5} \rightarrow \infty$ [cf. Gasser and Müller (1984)].

For the noise process and its estimation we make the following assumptions.

(A1) Model (2.1) holds with $\mathbb{E}|\varepsilon_t|^4 < \infty$.

(A2) The AR(∞) transfer function $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$, $z \in \mathbb{C}$, of model (2.1) satisfies: $\Phi(z)$ is bounded away from zero for $|z| \leq 1$ and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$.

(A3) In Step 2 of the sieve bootstrap scheme,

$$p(n) \rightarrow \infty, \quad p(n) = o(\min\{(n/\log(n))^{1/4}, (n\tilde{h})^{1/4}, \tilde{h}^{-1}\}), \quad n \rightarrow \infty,$$

with \tilde{h} as in assumption (K) below.

Moreover, $\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n}$ are defined by (2.2).

The conditions are met by a broad subclass of linear processes, including many ARMA models. A quite standard assumption for ARMA processes is that the defining finite AR- and MA-polynomials have their roots outside the unit circle $|z| \leq 1$, $z \in \mathbb{C}$ [cf. Brockwell and Davis (1987), Chapter 3.1]. Since the polynomials are of finite degrees, their roots are outside $|z| \leq 1 + \kappa$, $\kappa > 0$, and the corresponding AR(∞) representation of such ARMA processes has exponentially decaying coefficients $|\phi_j| \leq \text{const. } \rho^j$ for some $0 < \rho < 1$. Note that our conditions require only summability of $j|\phi_j|$, which thus include also processes with much stronger dependencies than finite parameter ARMA processes as above.

Our conditions are nice in the sense that the nonstationary nature of the process $\{Y_t\}_{t \in \mathbb{Z}}$, due to the deterministic trend $\{s(t)\}_{t \in \mathbb{Z}}$, does not add any additional restrictions for the stationary noise process $\{Z_t\}_{t \in \mathbb{Z}}$: the same, rather mild, conditions are inherited from the stationary sieve bootstrap as in

Bühlmann (1997). The only additional condition for our nonstationary sieve bootstrap is in (A3) on the growth for the approximating order $p = p(n)$ in Step 2, the additional restriction is $p(n) = o(\min\{(n\tilde{h})^{1/4}, \tilde{h}^{-1}\})$. This reflects the fact that we deal here with a sieve approximation which also estimates the trend: the order $p(n)$ of the autoregressive sieve approximation for the noise process should now be smaller.

For the trend function in (3.1) and its estimation in (3.2) we make the following assumptions.

(K) The function $m(\cdot)$ is twice continuously differentiable in the open interval $(0, 1)$ with $\sup_{0 < x < 1} |m^{(j)}(x)| < \infty$ for $j = 0, 1, 2$.

The estimator $\hat{m}(\cdot)$ is as given in (3.2) with $K(\cdot)$ a probability density, symmetric, twice continuously differentiable, compactly supported and the order of the bandwidth is $h(n) \sim \text{const. } n^{-1/5}$.

The estimator $\tilde{m}(\cdot)$ is as given in (3.2) with the same kernel $K(\cdot)$ as for $\hat{m}(x)$ but with bandwidth $\tilde{h}(n)$ such that $\tilde{h}n^{1/5} \rightarrow \infty$.

Moreover, the sieve bootstrap scheme is modified as described in (3.3).

THEOREM 3.1. *Assume that (A1)–(A3) and (K) hold. Then, for $0 < x < 1$,*

$$(nh)^{1/2}(\hat{m}(x) - m(x)) \rightarrow_d \mathcal{N}(B_{as}(x), \sigma_{as}^2),$$

$$n \rightarrow \infty, \text{ with } 0 \leq B_{as}(x), \sigma_{as}^2 < \infty,$$

and

$$\begin{aligned} & \sup_{u \in \mathbb{R}} |\mathbb{P}^*[(nh)^{1/2}(\hat{m}^*(x) - \tilde{m}(x)) \leq u] - \mathbb{P}[(nh)^{1/2}(\hat{m}(x) - m(x)) \leq u]| \\ & = o_p(1), \quad n \rightarrow \infty. \end{aligned}$$

A proof is given in Section 6.

REMARK 3.1. Theorem 3.1 can be extended to $\text{AR}(\infty)$ processes with stationary, ergodic innovation processes $\{\varepsilon_t\}_{t \in \mathbb{Z}}$, satisfying $\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] \equiv 0$, $\mathbb{E}[\varepsilon_t^2 | \mathcal{F}_{t-1}] \equiv \sigma^2$, $\mathbb{E}|\varepsilon_t|^4 < \infty$, where $\mathcal{F}_t = \sigma(\{\varepsilon_s\}; \{s \leq t\})$ denotes the σ -field generated up to time t . This then establishes a result for nonlinear processes, although the nonlinearity enters only in a rather weak form.

3.1. Behavior within $h(n)$ -neighborhoods. For two distinct points $x \neq y$, the estimates $\hat{m}(x)$ and $\hat{m}(y)$ become asymptotically independent. However, for finite sample size and $|x - y|$ of order $h = h(n)$, there is non-negligible correlation. We show that the sieve bootstrap approximation reflects this neighborhood correlation correctly. This property is useful for constructing simultaneous sieve bootstrap confidence intervals over a finite number of $h(n)$ -neighborhoods.

For $0 < x_0 < 1$, let

$$Z_{x_0;n}(x) = (nh)^{1/2}(\hat{m}(x_0 + xh) - m(x_0 + xh)), \quad -1 \leq x \leq 1,$$

$$Z_{x_0;n}^*(x) = (nh)^{1/2}(\hat{m}^*(x_0 + xh) - \tilde{m}(x_0 + xh)), \quad -1 \leq x \leq 1.$$

We assume here and in the sequel without loss of generality that $m(x_0 + xh)$ is always defined, since $h = h(n) = o(1)$. Denote by $\mathcal{C}[-1, 1]$ the space of real-valued continuous functions on $[-1, 1]$ and by \Rightarrow weak convergence in $\mathcal{C}[-1, 1]$ with respect to the sup-norm.

THEOREM 3.2. *Assume the conditions of Theorem 3.1. Then, for any $0 < x_0 < 1$,*

$$\begin{aligned} \{Z_{x_0;n}(x) - B_{as}(x_0)\}_{x \in [-1, 1]} &\Rightarrow \{W(x)\}_{x \in [-1, 1]}, \\ \{Z_{x_0;n}^*(x) - B_{as}(x_0)\}_{x \in [-1, 1]} &\Rightarrow \{W(x)\}_{x \in [-1, 1]} \quad \text{in probability,} \end{aligned}$$

where $\{W(x)\}_{x \in [-1, 1]}$ is a Gaussian process with $\mathbb{E}[W(x)] = 0$ and $\text{Cov}(W(x), W(y)) = 2\pi f_Z(0) \int_{\mathbb{R}} K(w)K(w + y - x) dw$.

REMARK 3.2. The result in Theorem 3.2 is also new for the case with independent observations and extends Theorem 2 in Härdle and Marron (1991). The statement holds for processes in $\mathcal{C}(C)$, where the index set C is a compact subset of \mathbb{R} , too.

A proof of Theorem 3.2 is given in Section 6. For constructing a simultaneous confidence band over the whole open interval $(0, 1)$ we need to know the distribution of quantities like $\sup_{x \in (0, 1)} (nh)^{1/2}(\hat{m}(x) - m(x))$. We do not know whether the sieve bootstrap yields an asymptotically correct approximation for the distribution of this random variable; note that the limit is an extreme value distribution (cf. Section 5.1.2), which makes the consistency of the bootstrap a delicate problem. However, Theorem 3.2 still contributes to understanding the uniform behavior of the sieve bootstrap. As described by the corollary below, it implies the asymptotic validity of the bootstrap for simultaneous confidence intervals over a finite number of $h(n)$ -neighborhoods, compare also with the approach of Härdle and Marron (1991). In some practical applications, this is of great interest (see Section 4.2).

COROLLARY 3.3. *Assume the conditions of Theorem 3.1. Let $0 < x_1 < \dots < x_d < 1$ and $U_i = [x_i - h, x_i + h]$, $i = 1, \dots, d < \infty$. Then, for $u \in \mathbb{R}$,*

$$\begin{aligned} &\mathbb{P}^* \left[\max_{1 \leq i \leq d} \sup_{x \in U_i} (nh)^{1/2}(\hat{m}^*(x) - \tilde{m}(x)) \leq u \right] \\ &- \mathbb{P} \left[\max_{1 \leq i \leq d} \sup_{x \in U_i} (nh)^{1/2}(\hat{m}(x) - m(x)) \leq u \right] = o_P(1), \quad n \rightarrow \infty. \end{aligned}$$

PROOF. The statement is a straightforward extension of Theorem 3.2: one can use a multivariate version of Theorem 3.1 and then proceed as in the proof of Theorem 3.2. \square

REMARK 3.3. The statement of Corollary 3.3 is also true for neighborhoods of the form $U_i = [x_i - ah, x_i + bh]$, $0 \leq a, b < \infty$.

3.2. *Skewness, bispectrum of noise and second order accuracy.* Often, the bootstrap becomes more powerful when resampling an asymptotically pivotal quantity, such as a studentized statistic [cf. Hall (1992)]. The key to second order accuracy is the correct skewness of the bootstrap distribution.

THEOREM 3.4. *Assume the conditions of Theorem 3.1. Let $\mu_{3;n}(x) = \mathbb{E}[(nh)^{1/2}(\hat{m}(x) - \mathbb{E}[\hat{m}(x)])^3]$ and $\mu_{3;n}^*(x) = \mathbb{E}^*[(nh)^{1/2}(\hat{m}^*(x) - \mathbb{E}^*[\hat{m}(x)])^3]$. Then, for $0 < x < 1$,*

$$\begin{aligned} \lim_{n \rightarrow \infty} (nh)^{1/2} \mu_{3;n}(x) &= \mu_{3;as}, \\ (nh)^{1/2} \mu_{3;n}^*(x) - \mu_{3;as} &= o_P(1), \quad n \rightarrow \infty, \end{aligned}$$

where $\mu_{3;as} = \sum_{s,t=-\infty}^{\infty} \mathbb{E}[Z_0 Z_s Z_t] \int_{\mathbb{R}} K^3(x) dx$.

A proof is given in Section 6. The asymptotic skewness $\mu_{3;as}$ is proportional to $\sum_{s,t=-\infty}^{\infty} \mathbb{E}[Z_0 Z_s Z_t]$, $(2\pi)^2$ times the bispectral density of the noise process at zero. This immediately implies that all Gaussian noise processes yield asymptotic skewness equal to zero [cf. Subba Rao and Gabr (1980)]. Also, we see that the asymptotic skewness is a more complicated functional of the noise than in the i.i.d. case: the bispectral density at zero is not only influenced by the skewness $\mathbb{E}[Z_t^3]$ of the one-dimensional marginal but also by the cumulants $\text{cum}_3(Z_0, Z_s, Z_t) = \mathbb{E}[Z_0 Z_s Z_t]$ of order 3 with arbitrary lags s and t .

The leading error term in an estimated normal approximation for $(nh)^{1/2}(\hat{m}(x) - \mathbb{E}[m(x)])$ is typically given by the skewness (see also Section 1),

$$\mu_{3;n}(x) \sim (nh)^{-1/2} \mu_{3;as}.$$

Thus, small bandwidths h and large values of the bispectral density of $\{Z_t\}_{t \in \mathbb{Z}}$ at zero yield a poor behavior of the normal approximation. A small bandwidth is preferred with a “rough” underlying trend by accounting for the bias $(nh)^{1/2}(\mathbb{E}[\hat{m}(x) - m(x)])$, which is asymptotically proportional to $(nh^5)^{1/2} m^{(2)}(x)$ [$m(\cdot)$ still twice continuously differentiable but with a large value $|m^{(2)}(x)|$]. A large bispectral density at zero occurs if the noise is strongly non-Gaussian. In these “rough and non-Gaussian” situations, the sieve bootstrap approximation is expected to bring a substantial improvement, as indicated by Theorem 3.4.

We now add some formal heuristics indicating how Theorem 3.4 implies second order accuracy of the bootstrap approximation. For $0 < x < 1$, denote by

$$\begin{aligned} S_n &= \frac{\hat{m}(x) - \mathbb{E}[\hat{m}(x)]}{\sigma_n(x)}, \\ U_n &= \frac{\hat{m}(x) - \mathbb{E}[\hat{m}(x)]}{\hat{\sigma}_n(x)}, \end{aligned}$$

where

$$\begin{aligned}
\sigma_n^2(x) &= \text{Var}(\hat{m}(x)), \\
\hat{\sigma}_n^2(x) &= 2\pi \hat{f}_{AR(p)}(0)(nh)^{-2} \sum_{t=1}^n K\left(\frac{x-t/n}{h}\right)^2, \\
(3.6) \quad \hat{f}_{AR(p)}(0) &= \frac{(n-p)^{-1} \sum_{t=p+1}^n \tilde{\varepsilon}_{t,n}^2}{2\pi |\hat{\Phi}_n(e^{-i\lambda})|^2}, \\
\hat{\Phi}_n(z) &= \sum_{j=0}^p \hat{\phi}_{j,n} z^j, \quad \hat{\Phi}_{0,n} = 1, \quad z \in \mathbb{C}, \\
&\quad (\hat{\phi}_{j,n})_{j=1}^p, \tilde{\varepsilon}_{t,n} \text{ as defined in Step 2 of the sieve bootstrap scheme} \\
&\quad \text{(with the same } p\text{)}.
\end{aligned}$$

Note that if $\sum_{k=-\infty}^{\infty} |kR_Z(k)| < \infty$, then

$$nh\sigma_n^2(x) - 2\pi f_Z(0)(nh)^{-1} \sum_{t=1}^n K\left(\frac{x-t/n}{h}\right)^2 = o(n^{-1}).$$

An Edgeworth expansion for U_n can typically be related to an expansion for S_n . By a Taylor expansion,

$$U_n = S_n \left(1 - \frac{\hat{\sigma}_n^2(x) - \sigma_n^2(x)}{2\sigma_n^3(x)}\right) + O_P\left(\frac{(\hat{\sigma}_n^2(x) - \sigma_n^2(x))^2}{\sigma_n^5(x)}\right).$$

If the autoregressive coefficients $\{\phi_j\}_{j=0}^{\infty}$ of the noise process $\{Z_t\}_{t \in \mathbb{Z}}$ decay sufficiently fast, we expect $(\hat{\sigma}_n^2(x) - \sigma_n^2(x))/\sigma_n^3(x) = o_P(1)$ and the expectation, variance and skewness of

$$S_n \left(1 - \frac{\hat{\sigma}_n^2(x) - \sigma_n^2(x)}{2\sigma_n^3(x)}\right)$$

should be reasonably close to 0, 1 and the skewness of S_n , respectively. This would indicate

$$(3.7) \quad \mathbb{P}[U_n \leq u] = \Phi_{\mathcal{N}}(u) - \varphi_{\mathcal{N}}(u) \frac{1}{6}(u^2 - 1)(nh)^{-1/2} \tilde{\mu}_{3;n}(x) + o((nh)^{-1/2}),$$

where $\tilde{\mu}_{3;n}(x) = (nh)^{-1} \mu_{3;n}(x)/\sigma_n^3(x)$. Since the bootstrap process $\{Z_t^*\}_{t \in \mathbb{Z}}$ is an $AR(p(n))$ process whose probabilistic law is close to the law of $\{Z_t\}_{t \in \mathbb{Z}}$, the expansion in (3.7) is also expected to hold for the sieve bootstrap,

$$(3.8) \quad \mathbb{P}^*[U_n^* \leq u] = \Phi_{\mathcal{N}}(u) - \varphi_{\mathcal{N}}(u) \frac{1}{6}(u^2 - 1)(nh)^{-1/2} \tilde{\mu}_{3;n}^*(x) + o_P((nh)^{-1/2}),$$

where

$$\begin{aligned}
U_n^* &= \frac{\hat{m}^*(x) - \mathbb{E}^*[\hat{m}^*(x)]}{\hat{\sigma}_n^*(x)}, \\
\hat{\sigma}_n^*(x)^2 &= 2\pi \hat{f}_{AR(p)}^*(0)(nh)^{-2} \sum_{t=1}^n K\left(\frac{x-t/n}{h}\right)^2, \\
\hat{f}_{AR(p)}^*(0) &= H_n(Z_{[\delta n]+1}^*, \dots, Z_{[(1-\delta)n]}^*), \\
(3.9) \quad \text{where } H_n(\hat{Z}_{[\delta n]+1,n}, \dots, \hat{Z}_{[(1-\delta)n],n}) &= \hat{f}_{AR(p)}(0) \text{ as in (3.6),}
\end{aligned}$$

$$\begin{aligned}\tilde{\mu}_{3;n}^*(x) &= (nh)^{-1} \frac{\mu_{3;n}^*(x)}{\sigma_n^*(x)^3}, \\ \sigma_n^*(x)^2 &= 2\pi f_{Z^*}(0)(nh)^{-2} \sum_{t=1}^n K\left(\frac{x-t/n}{h}\right)^2, \\ f_{Z^*}(0) &\text{ is the spectral density of } \{Z_t^*\}_{t \in \mathbb{Z}} \text{ at zero.}\end{aligned}$$

By Theorem 3.1, $nh(\sigma_n^*(x)^2 - \sigma_n^2(x)) = o_P(1)$ and thus

$$(3.10) \quad \tilde{\mu}_{3;n}(x) - \tilde{\mu}_{3;n}^*(x) = O_P((nh)^{1/2}(\mu_{3;n}(x) - \mu_{3;n}^*(x))).$$

A rigorous justification for (3.7) and (3.8) is postponed as a research problem for the future. Accepting the validity of these expansions, it then follows from (3.10) and Theorem 3.4.

COROLLARY 3.5. *Under the assumptions of Theorem 3.1 and if (3.7) and (3.8) are valid,*

$$\mathbb{P}^*[U_n^* \leq u] - \mathbb{P}[U_n \leq u] = o_P((nh)^{-1/2}), \quad n \rightarrow \infty.$$

Corollary 3.5 improves upon the $(nh)^{-1/2}$ rate in the estimated normal approximation.

In contrast to the automatic bootstrap bias correction of $\hat{m}^*(x) - \tilde{m}(x)$ for the nonpivotal quantity $\hat{m}(x) - m(x)$, the bootstrap approximation for the studentized statistic U_n will yield a (second order accurate) confidence interval for $\mathbb{E}[\hat{m}(x)]$, rather than for $m(x)$ itself. With a more subtle choice of the bandwidth h , second order properties are expected to hold when applying an explicit bias correction or undersmoothing. A better theoretical understanding about the choice of such h can be borrowed from the independent case, studied by Hall (1992). A more practical choice of an undersmoothing bandwidth h , now not in mean square error optimal sense, has been less explored.

4. Constructing confidence regions.

4.1. Pointwise sieve bootstrap confidence regions. The sieve bootstrap can be applied for constructing pointwise confidence bands in the same fashion as in the independent setup.

Based on the nonpivotal quantity $\hat{m}(x) - m(x)$, the ordinary sieve bootstrap confidence interval is given by

$$(4.1) \quad I_n^*(x; 1 - \alpha) = [\hat{m}(x) - \hat{q}_{1-\alpha/2}, \hat{m}(x) - \hat{q}_{\alpha/2}], \quad 0 < \alpha < 1,$$

where $\hat{q}_\alpha = \inf\{u; \mathbb{P}^*[\hat{m}^*(x) - \tilde{m}(x) \leq u] \geq \alpha\}$.

By Theorem 3.1, this construction is asymptotically valid.

Based on the pivotal Studentized statistic $(\hat{m}(x) - m(x))/\hat{\sigma}_n(x)$, the Studentized sieve bootstrap confidence interval is given by

$$(4.2) \quad I_n^{*, \text{stud}}(x; 1 - \alpha) = [\hat{m}(x) - \hat{q}_{1-\alpha/2}\hat{\sigma}_n(x), \hat{m}(x) - \hat{q}_{\alpha/2}\hat{\sigma}_n(x)],$$

$$0 < \alpha < 1,$$

where

$$\hat{q}_\alpha = \inf \left\{ u; \mathbb{P}^* \left[\frac{\hat{m}^*(x) - \tilde{m}(x)}{\hat{\sigma}_n^*(x)} \leq u \right] \geq \alpha \right\}.$$

Note that this is slightly different than the situation in Section 3.2, where we neglected the problems with bias and just considered the centered quantities U_n and U_n^* , respectively.

4.2. Simultaneous within neighborhoods sieve bootstrap confidence bands. Due to Corollary 3.3, the sieve bootstrap can be applied for constructing simultaneous confidence bands within a finite number of neighborhoods U_1, \dots, U_d with diameters $ch(n)$, $0 < c < \infty$. Simultaneity within a finite number of neighborhoods can be of interest in many applications. As examples, we mention ecological time series, such as river water flow and deforestation of the Amazon basin [cf. Brillinger (1988)] or total ozone measurements from the longest series at Arosa (Switzerland) [cf. Staehelin et al. (1997) and Staehelin, Kegel and Harris (1997)]. In both cases, one aim is to construct confidence bands which are simultaneous over early (left end, x close to zero) and late times (right end, x close to 1). In particular, the question about increase within the given time stretch can now be answered via simultaneous interval estimates at the beginning and end.

The shape of a simultaneous band $\{I_n(x)\}_{x \in G}$ over the set $G = \bigcup_{i=1}^d U_i$ is not canonical, in an asymptotic sense we only require that $\lim_{n \rightarrow \infty} \mathbb{P}[m(x) \ni I_n(x) \text{ for all } x \in G] \geq 1 - \alpha$, $0 \leq \alpha \leq 1$. One does not have to stick with Kolmogorov–Smirnov type bands, which possess some mathematical elegance but seem not to be the most natural in practice. We rather want to construct simultaneous confidence bands with nonequal width. It seems evident that at some points there is less variability (at least for finite sample sizes) than at others.

For a practical implementation we present here a bisection search algorithm. The algorithm, based on the nonpivotal quantity $\hat{m}(x)$ to construct simultaneous confidence bands over the set G to the level $1 - \alpha$, proceeds as follows:

1. Compute pointwise quantiles

$$\hat{q}_{\alpha_P/2}(x), \hat{q}_{1-\alpha_P/2}(x), \quad x \in G,$$

where $\hat{q}_{\alpha_P} = \inf \{ u; \mathbb{P}^*[\hat{m}^*(x) - \tilde{m}(x) \leq u] \geq \alpha_P \}$ is a pointwise bootstrap quantile.

2. Vary the pointwise error α_p until

$$\mathbb{P}^*[\hat{q}_{\alpha_p/2}(x) \leq \hat{m}^*(x) - \tilde{m}(x) \leq \hat{q}_{1-\alpha_p/2}(x) \forall x \in G] \approx 1 - \alpha$$

(the ratio of bootstrap curves around $\tilde{m}(x)$ within the band $[\hat{q}_{\alpha_p/2}(x), \hat{q}_{1-\alpha_p/2}(x)]$ for all $x \in G$ is approximately $1 - \alpha$). Denote this α_p by α_S .

3. Build

$$(4.3) \quad I_n(x) = \hat{m}(x) - \hat{q}_{1-\alpha_S/2}(x), \hat{m}(x) - \hat{q}_{\alpha_S/2}(x), \quad x \in G.$$

Then, by Corollary 3.3, $\{I_n(x)\}_{x \in G}$ yields a consistent confidence band to the level $1 - \alpha$, which is simultaneous for G being a finite union of local neighborhoods.

By equivariance of the sample quantiles, it is faster to compute quantiles of $\hat{m}^*(x)$, rather than the centered version around $\tilde{m}(x)$, and correct at the stage when constructing the confidence band $I_n(x)$.

5. Numerical study.

5.1. *Simulations.* We study the accuracy of the sieve bootstrap approximation in terms of coverage probabilities and compare it with alternative methods. We consider the following trend models:

$$(T1) \quad m(x) = 2 - 5x + 5 \exp(-100(x - 0.5)^2), \quad x \in [0, 1];$$

$$(T2) \quad m(x) = 2 - 5x + 5 \exp(-1000(x - 0.5)^2), \quad x \in [0, 1].$$

The trend model (T1) is the same as in Herrmann, Gasser and Kneip (1992); the model (T2) is a “rougher” version of (T1), having 10 times higher second derivative at the peak at $x = \frac{1}{2}$. To this trend we add ARMA(1, 1) noise,

$$(5.1) \quad Z_t = 0.8Z_{t-1} - 0.5\varepsilon_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

where

$$(N1) \quad \sqrt{1.8}\varepsilon_t \text{ i.i.d. } \sim t_6,$$

$$(N2) \quad \varepsilon_t = (\eta_t - \mathbb{E}[\eta_t])/\sqrt{1.2}, \quad \eta_t \sim \text{Exp}(1).$$

The innovations are scaled so that $\text{Var}(Z_t) \approx 1$. We point out that our noise level (inverse of signal to noise ratio) is considerably higher than in most other simulation studies. The noise process [(5.1), (N1)] has been studied in connection with the bootstrap in Bühlmann (1997). The innovations in (N2) are skewed, yielding a nonvanishing bispectral density at zero.

We generate data sets of sample size n , according to

$$Y_t = m(t/n) + Z_t, \quad t = 1, \dots, n.$$

All our simulation results are based on 100 realizations of such processes. We aim to achieve a coverage probability $1 - \alpha = 0.9$.

5.1.1. *Pointwise coverage.* We consider here the coverage at $x = 1/2$ (corresponding to $t = n/2$) at the peak. Note that model (T2, N2) then yields a value for $(nh)^{-1/2}\mu_{3;as}$, which is substantially different from zero, due the non-Gaussian noise process and a small bandwidth for estimating the sharp peak at $x = 1/2$, compare with Section 3.2. The pointwise sieve bootstrap confidence bands are constructed as described in Section 4.1. The precise specifications of the methods are as follows:

- (*ord) \hat{m} as in (3.2), bandwidth h , $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$,
 \tilde{m} as in (3.2), bandwidth \tilde{h} , $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$,
 $\delta = 0.05$ for the pilot estimates in (3.3)
 $p = \hat{p}_{\text{AIC}}$ within the range $[0, 10 \log_{10}(n)]$ (S-PLUS default value), based on the variables $\hat{Z}_{t,n}$ ($t = [0.05n] + 1, \dots, [0.95n]$),
confidence band as in (4.1), 500 bootstrap replicates;
- (*stud) $\hat{m}, \tilde{m}, \delta, p$ as in (*ord),
 $\hat{\sigma}_n^2(x)$ as in (3.6), $\hat{\sigma}_n^*(x)^2$ as in (3.9),
confidence interval as in (4.2), 500 bootstrap replicates.

We study method (*stud) only for the case (T2, N2), where the skewness of $\hat{m}(x)$ is substantial. We compare these methods with the normal approximation, whose validity is justified by Theorem 3.1. The normal approximation confidence band for a fixed x is

$$(5.2) \quad \hat{m}(x) \pm \hat{\sigma}_n(x) \Phi_{\mathcal{N}}^{-1}(1 - \alpha/2),$$

or with a bias correction

$$(5.3) \quad \hat{m}(x) - h^2 \hat{m}^{(2)}(x) \int_{\mathbb{R}} x^2 K(x) dx / 2 \pm \hat{\sigma}_n(x) \Phi_{\mathcal{N}}^{-1}(1 - \alpha/2),$$

where $\hat{\sigma}_n^2(x)$ is given in (3.6) and $\hat{m}^{(2)}(x)$ is a kernel estimate for estimating the second derivative $m^{(2)}(x)$. One proposal is

$$(5.4) \quad \hat{m}^{(2)}(x) = n^{-1} \tilde{h}^{-3} \sum_{t=1}^n K_2\left(\frac{x - t/n}{\tilde{h}}\right) Y_t,$$

with $K_2(x) = 105/32(-5x^4 + 6x^2 - 1)1_{[-1 \leq x \leq 1]}$ [see Gasser, Müller and Mammitzsch (1985)].

The precise specifications of the methods are as follows:

- (\mathcal{N}) \hat{m} as in (*ord),
 $\hat{\sigma}_n^2(x)$ as in (3.6), based on \tilde{m}, δ, p as in (*stud),
confidence band as in (5.2);
- (\mathcal{N} BC) $\hat{m}, \hat{\sigma}_n^2(x)$ as in (\mathcal{N}),
 $\hat{m}^{(2)}$ as in (5.4), bandwidth \tilde{h} ,
confidence band as in (5.3).

TABLE 1

	$n = 128, (T1, N1)$	$n = 128, (T2, N2)$	$n = 512, (T1, N1)$	$n = 512, (T2, N2)$
$h_{\text{opt}}(1/2)$	0.0258	0.0103	0.0196	0.0078

We vary the bandwidths h and \tilde{h} according to the following rule. Given a bandwidth h we consider the pilot bandwidths

$$(5.5) \quad \tilde{h} = h \quad \text{and} \quad \tilde{h} = Ch^{5/9}, \quad C = 1/2, 1, 2.$$

The order $n^{-1/9}$ for \tilde{h} is expected to be mean square error optimal for estimating the bias $\mathbb{E}[\hat{m}(x)] - m(x)$ by the sieve bootstrap [cf. Härdle and Marron (1991)]. We then consider

$$(5.6) \quad h = C0.044n^{-1/5}, \quad C \in \{1/2, 1, 2, 4\}.$$

This gives reasonable bandwidths for our sample sizes: $h \in \{0.0083, 0.0167, 0.0333, 0.0677\}$ for $n = 128$ and $h \in \{0.0063, 0.0126, 0.0253, 0.0505\}$ for $n = 512$, respectively. As a guideline, the asymptotic mean square error optimal local bandwidths $h_{\text{opt}}(1/2)$ for estimating $m(1/2)$ are given in Table 1. Tables 2–5 show coverage percentages with estimated standard errors in parentheses (in percentages), which were derived by simulating over 100 different model realizations, that is, realizations of the noise processes in (5.1).

For (*ord) and (\mathcal{N} BC) we know from the theory for bias correction that \tilde{h} has to be larger than h . This is also seen in the simulation.

The sieve bootstrap methods (*ord) and (*stud) are overall less sensitive to the choice of the smoothing parameters than the normal approximations

TABLE 2

Pointwise coverages, model (T1, N1), $n = 128$; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6), $\tilde{h} = h$ or $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	*ord	\mathcal{N}	\mathcal{N} BC
$C_1 = 1, \tilde{h} = h$	21% (4.1)	41% (4.9)	34% (4.7)
$C_1 = 1, C_2 = 0.5$	76% (4.3)	—	36% (4.8)
$C_1 = 1, C_2 = 1$	96% (2.0)	—	44% (5.0)
$C_1 = 1, C_2 = 2$	100% (—)	—	43% (5.0)
$C_1 = 2, \tilde{h} = h$	46% (5.0)	78% (4.1)	65% (4.8)
$C_1 = 2, C_2 = 0.5$	83% (3.8)	—	80% (4.0)
$C_1 = 2, C_2 = 1$	96% (2.0)	—	84% (3.7)
$C_1 = 2, C_2 = 2$	100% (—)	—	80% (4.0)
$C_1 = 4, \tilde{h} = h$	30% (4.6)	48% (5.0)	84% (3.7)
$C_1 = 4, C_2 = 0.5$	43% (5.0)	—	90% (3.0)
$C_1 = 4, C_2 = 1$	83% (3.8)	—	72% (4.5)
$C_1 = 4, C_2 = 2$	95% (2.2)	—	59% (4.9)

TABLE 3

Pointwise coverages, model (T1, N1), $n = 512$; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6), $\tilde{h} = h$ or $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	*ord	\mathcal{N}	\mathcal{N} BC
$C_1 = 1, \tilde{h} = h$	36% (4.8)	62% (4.9)	35% (4.8)
$C_1 = 1, C_2 = 0.5$	83% (3.8)	—	59% (4.9)
$C_1 = 1, C_2 = 1$	98% (1.4)	—	61% (4.9)
$C_1 = 1, C_2 = 2$	100% (—)	—	62% (4.8)
$C_1 = 2, \tilde{h} = h$	64% (4.8)	88% (3.2)	64% (4.8)
$C_1 = 2, C_2 = 0.5$	86% (3.5)	—	90% (3.0)
$C_1 = 2, C_2 = 1$	97% (1.7)	—	89% (3.1)
$C_1 = 2, C_2 = 2$	100% (—)	—	88% (3.2)
$C_1 = 4, \tilde{h} = h$	44% (5.0)	52% (5.0)	79% (4.1)
$C_1 = 4, C_2 = 0.5$	63% (4.8)	—	87% (3.4)
$C_1 = 4, C_2 = 1$	96% (2.0)	—	70% (4.6)
$C_1 = 4, C_2 = 2$	100% (—)	—	58% (4.9)

(\mathcal{N}) and (\mathcal{N} BC). This is a very valuable and important feature. The sieve bootstrap outperforms the normal approximation for the case (T2, N2) with “rough” trend and strongly non-Gaussian noise. The theoretical advantage of (*stud) over (*ord) [for bandwidths reasonable for method (*ord)] is only slightly visible for the smaller sample size $n = 128$. A reasonable choice for the interplay between h and \tilde{h} in (*ord) is here $\tilde{h} = Ch^{5/9}$, $C \in [1/2, 1]$. We observe for (*ord) and (*stud) that the confidence intervals become more conservative as \tilde{h} increases with h fixed. Overall, the sieve bootstrap yields quite satisfactory results.

TABLE 4

Pointwise coverages, model (T2, N2), $n = 128$; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6), $\tilde{h} = h$ or $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	*ord	*stud	\mathcal{N}	\mathcal{N} BC
$C_1 = 0.5, \tilde{h} = h$	27% (4.6)	—	17% (3.8)	17% (3.8)
$C_1 = 0.5, C_2 = 0.5$	74% (4.4)	77% (4.2)	—	21% (4.1)
$C_1 = 0.5, C_2 = 1$	90% (3.0)	93% (2.6)	—	17% (3.8)
$C_1 = 0.5, C_2 = 2$	97% (1.7)	99% (1.0)	—	17% (3.8)
$C_1 = 1, \tilde{h} = h$	27% (4.6)	—	27% (4.6)	34% (4.7)
$C_1 = 1, C_2 = 0.5$	48% (5.0)	54% (5.0)	—	33% (4.7)
$C_1 = 1, C_2 = 1$	67% (4.7)	74% (4.4)	—	32% (4.7)
$C_1 = 1, C_2 = 2$	76% (4.3)	91% (2.9)	—	27% (4.6)
$C_1 = 2, \tilde{h} = h$	3% (1.7)	—	6% (2.4)	67% (4.7)
$C_1 = 2, C_2 = 0.5$	2% (1.4)	—	—	23% (4.2)
$C_1 = 2, C_2 = 1$	5% (2.2)	—	—	11% (3.1)
$C_1 = 2, C_2 = 2$	16% (3.7)	—	—	7% (2.6)

TABLE 5

Pointwise coverages, model (T2, N2), $n = 512$; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6), $\tilde{h} = h$ or $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	*ord	*stud	\mathcal{N}	\mathcal{N} BC
$C_1 = 0.5, \tilde{h} = h$	31% (4.6)	—	31% (4.6)	13% (3.4)
$C_1 = 0.5, C_2 = 0.5$	86% (3.5)	85% (3.6)	—	30% (4.6)
$C_1 = 0.5, C_2 = 1$	98% (1.4)	100% (—)	—	27% (4.4)
$C_1 = 0.5, C_2 = 2$	100% (—)	100% (—)	—	30% (4.6)
$C_1 = 1, \tilde{h} = h$	35% (4.8)	—	47% (5.0)	47% (5.0)
$C_1 = 1, C_2 = 0.5$	67% (4.7)	68% (4.7)	—	64% (4.8)
$C_1 = 1, C_2 = 1$	92% (2.7)	94% (2.4)	—	55% (5.0)
$C_1 = 1, C_2 = 2$	98% (1.4)	100% (—)	—	47% (5.0)
$C_1 = 2, \tilde{h} = h$	2% (1.4)	—	6% (2.4)	79% (4.1)
$C_1 = 2, C_2 = 0.5$	5% (2.2)	—	—	37% (4.8)
$C_1 = 2, C_2 = 1$	15% (3.6)	—	—	10% (3.0)
$C_1 = 2, C_2 = 2$	27% (4.6)	—	—	8% (2.7)

The methods (\mathcal{N}) and (\mathcal{N} BC) are very poor for the “rough” trend and strongly non-Gaussian noise (T2, N2). For case (T1, N1), the method (\mathcal{N}) is only reasonable with the midsize bandwidth h , the bias-corrected version (\mathcal{N} BC) can yield useful results for the two larger bandwidths h , but yields a poor result for the small bandwidth h . As said above, the choice of bandwidth in the normal approximation is more subtle.

The estimation of the spectral density $f_Z(0)$ in methods (\mathcal{N}), (\mathcal{N} BC) and (*stud) could actually be based on residuals, which are obtained by pilot-estimation of the trend with a third bandwidth $\tilde{\tilde{h}}$. However, this would make the procedures more complicated.

We compared also the lengths of confidence intervals and give in Table 6 the empirical results for some interval estimators with reasonable coverage accuracy, based on 100 simulations. Given $\tilde{h} = C_2 h^{5/9}$ with C_2 fixed, small bandwidths h yield short intervals. For the sieve bootstrap, the length of the confidence interval is growing as \tilde{h} increases with h fixed. In case (T1, N1) the best intervals with method (*ord) and (\mathcal{N} BC) are similar.

5.1.2. *Simultaneous coverage.* We consider now simultaneous coverage. We only consider the “smooth” Gaussian model (T1, N1) to have a reasonable competitor based on normal limit approximations. The sets of points, for which the confidence band is simultaneous, are

$$G = G(h) = \bigcup_{i=1}^4 U_{x_i}(h),$$

$$U_{x_i}(h) = \{x_i - h + j/100; j = 0, \dots, [200h]\}, \quad x_i = i/5$$

and

$$G_{\text{sub}} = U_{x_1}(h) \cup U_{x_4}(h).$$

TABLE 6

Lengths of pointwise confidence intervals and their coverages [for (\mathcal{N} BC) with maximal coverage among different \tilde{h}]; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6) and $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	Method	Ave[length]	S.E.(length)	Coverage
(T1, N1, $n = 128$)				
$C_1 = 1, C_2 = 1$	*ord	2.85	0.42	96%
$C_1 = 2, C_2 = 0.5$	*ord	1.73	0.31	83%
$C_1 = 2, C_2 = 1$	*ord	3.39	0.48	96%
$C_1 = 4, C_2 = 1$	*ord	3.76	0.54	83%
$C_1 = 4, C_2 = 2$	*ord	4.96	0.58	95%
$C_1 = 2$	\mathcal{N} BC	1.73	0.30	84%
$C_1 = 4$	\mathcal{N} BC	2.75	0.61	90%
(T1, N1, $n = 512$)				
$C_1 = 1, C_2 = 0.5$	*ord	1.33	0.16	83%
$C_1 = 2, C_2 = 0.5$	*ord	1.37	0.19	86%
$C_1 = 2, C_2 = 1$	*ord	2.90	0.35	97%
$C_1 = 4, C_2 = 1$	*ord	3.39	0.39	96%
$C_1 = 2$	\mathcal{N} BC	1.55	0.26	90%
$C_1 = 4$	\mathcal{N} BC	1.89	0.26	87%
(T2, N2, $n = 128$)				
$C_1 = 0.5, C_2 = 1$	*ord	2.94	0.36	90%
$C_1 = 0.5, C_2 = 1$	*stud	3.21	0.44	93%
(T2, N2, $n = 512$)				
$C_1 = 0.5, C_2 = 0.5$	*ord	1.83	0.20	86%
$C_1 = 0.5, C_2 = 0.5$	*stud	1.91	0.21	85%
$C_1 = 0.5, C_2 = 1$	*ord	2.67	0.27	98%
$C_1 = 0.5, C_2 = 1$	*stud	2.93	0.31	100%
$C_1 = 1, C_2 = 1$	*ord	2.67	0.29	92%
$C_1 = 1, C_2 = 1$	*stud	2.92	0.35	94%

The sets G and G_{sub} mimic four and two (in our cases disjoint) intervals of approximate lengths $2h$. Note that G and G_{sub} change with h , the bandwidth of the estimator \hat{m} .

The simultaneous sieve bootstrap band is constructed as described in Section 4.2. The precise specification is

(*simult) $\hat{m}, \tilde{m}, \delta, p$ as in (*ord),
confidence band as in (4.3), 500 bootstrap replicates.

For comparison we use an extreme value approximation, modified for the time series case, for the distribution of

$$\sup_{0 < x < 1} (nh)^{1/2} (\hat{m}(x) - m(x)).$$

To our knowledge, there exists no rigorous proof for the limiting distribution of the random variable above in the time series case. In the independent setup,

TABLE 7

Simultaneous coverages over G and G_{sub} , respectively: model (T1, N1), $n = 128$; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6) and $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	*simult (G)	\mathcal{E} BC. (G)	*simult (G_{sub})	\mathcal{E} BC. (G_{sub})
$C_1 = 1, C_2 = 0.5$	53% (5.0)	33% (4.7)	54% (5.0)	51% (5.0)
$C_1 = 1, C_2 = 1$	98% (1.4)	36% (4.8)	98% (1.4)	53% (5.0)
$C_1 = 2, C_2 = 0.5$	75% (4.3)	98% (1.4)	77% (4.2)	99% (0.1)
$C_1 = 2, C_2 = 1$	100% (—)	99% (0.1)	100% (—)	100% (—)
$C_1 = 4, C_2 = 0.5$	81% (3.9)	100% (—)	96% (2.0)	100% (—)
$C_1 = 4, C_2 = 1$	99% (0.1)	98% (1.4)	100% (—)	100% (—)

the distribution has been given in Eubank and Speckman (1993). In general, the Poisson clumping heuristics [Aldous (1989)] can be used to obtain formally the following simultaneous confidence band:

$$I_n(x) = \hat{m}(x) - h^2 \hat{m}^{(2)}(x) \int_{\mathbb{R}} \frac{x^2 K(x) dx}{2} \pm \hat{\sigma}_n(x) v_{1-\alpha/2},$$

$$(5.7) \quad v_{1-\alpha/2} = (-2 \log(h))^{1/2} + (-2 \log(h))^{-1/2} \left(C - \log \left(-\log \left(1 - \frac{\alpha}{2} \right) \right) \right),$$

$$C = \log \left(\frac{\left\{ -\int_{\mathbb{R}} K^{(2)}(x) K(x) dx / \int_{\mathbb{R}} K^2(x) dx \right\}^{1/2}}{2\pi} \right),$$

where $\hat{\sigma}_n(x)$ and $\hat{m}^{(2)}(x)$ are the same as in (3.6) and (5.4).

This band should be simultaneous over the whole open interval $(0, 1)$. The precise specification of this method is

$$(\mathcal{E} \text{ BC}) \quad \hat{m}, \hat{\sigma}_n^2(x), \hat{m}^{(2)} \text{ as in } (\mathcal{N} \text{ BC}), \text{ confidence band as in (5.7).}$$

We consider the combinations of bandwidths which yielded reasonably accurate results in Section 5.1.1. The aimed simultaneous coverage level is again

TABLE 8

Simultaneous coverages over G and G_{sub} , respectively: model (T1, N1), $n = 512$; bandwidths of the form $h = C_1 0.044n^{-1/5}$ as in (5.6) and $\tilde{h} = C_2 h^{5/9}$ as in (5.5)

Bandwidths	*simult (G)	\mathcal{E} BC. (G)	*simult (G_{sub})	\mathcal{E} BC. (G_{sub})
$C_1 = 1, C_2 = 0.5$	64% (4.8)	58% (4.9)	71% (4.5)	76% (4.3)
$C_1 = 1, C_2 = 1$	100% (—)	63% (4.8)	100% (—)	82% (3.8)
$C_1 = 2, C_2 = 0.5$	93% (2.6)	100% (—)	93% (2.6)	100% (—)
$C_1 = 2, C_2 = 1$	100% (—)	100% (—)	100% (—)	100% (—)
$C_1 = 4, C_2 = 0.5$	97% (1.7)	100% (—)	100% (—)	100% (—)
$C_1 = 4, C_2 = 1$	100% (—)	100% (—)	100% (—)	100% (—)

TABLE 9

Lengths of simultaneous bands, method (\mathcal{E} BC): model (T1, N1), $n = 128$; bandwidth of the form $h = C_1 0.044n^{-1/5}$ as in (5.6); \tilde{h} is irrelevant on length

Bandwidth	Ave[length]	S.E.(length)	2.5% and 97.5% sample quantiles of length
$C_1 = 1$	1.85	0.56	0.86, 3.02
$C_1 = 2$	3.98	0.69	2.84, 5.39
$C_1 = 4$	6.10	1.36	4.08, 9.45

$1 - \alpha = 0.9$. Tables 7 and 8 show coverage percentages with estimated standard errors in parentheses (in percentages), which were derived by simulating over 100 different realizations of the noise process in (5.1).

We also show in Tables 9 and 10 the lengths of confidence bands with method (\mathcal{E} BC) with empirical moments and quantiles, based on the 100 simulations. Figures 1 and 2 show the 2.5% and 97.5% sample quantiles of $\text{length}(x)$ of the simultaneous sieve bootstrap confidence interval at all x in the set $G = G(h)$ and $G_{\text{sub}} = G_{\text{sub}}(h)$, respectively.

For simultaneous confidence intervals, the sieve bootstrap method (*simult) yields a reasonable overall performance in terms of coverage accuracy. It compares very favorably to the results in Härdle and Marron (1991) for the independent case. The improvement by having a larger sample size is visible in coverage accuracy as well as in size of the confidence band. Although we are looking only at the better range of smoothing parameters for the analytical (nonbootstrap) pointwise approximations in Section 5.1.1, the method (*simult) is again less sensitive to the selection of bandwidth parameters than method (\mathcal{E} BC).

For every given bandwidth h , the smaller pilot-bandwidth \tilde{h} yields consistently shorter bands. They are sometimes too progressive, as some of the (\mathcal{E} BC)-bands, too. By informally calibrating coverage probability against size of confidence bands, method (*simult) is in most cases favorable to method (\mathcal{E} BC). The only case where method (*simult) did not yield a “winner” is case $[G, n = 128, C_1 = 2, C_2 = 0.5]$ ($h = C_1 0.044n^{-1/5}$, $\tilde{h} = C_2 h^{5/9}$). As expected, the advantage of method (*simult) in this comparison is more clearly visible for G_{sub} .

TABLE 10

Lengths of simultaneous bands, method (\mathcal{E} BC): model (T1, N1), $n = 512$; bandwidth of the form $h = C_1 0.044n^{-1/5}$ as in (5.6); \tilde{h} is irrelevant on length

Bandwidth	Ave[length]	S.E.(length)	2.5% and 97.5% sample quantiles of length
$C_1 = 1$	1.78	0.45	0.96, 2.67
$C_1 = 2$	3.64	0.62	2.34, 4.72
$C_1 = 4$	4.22	0.62	3.19, 5.44

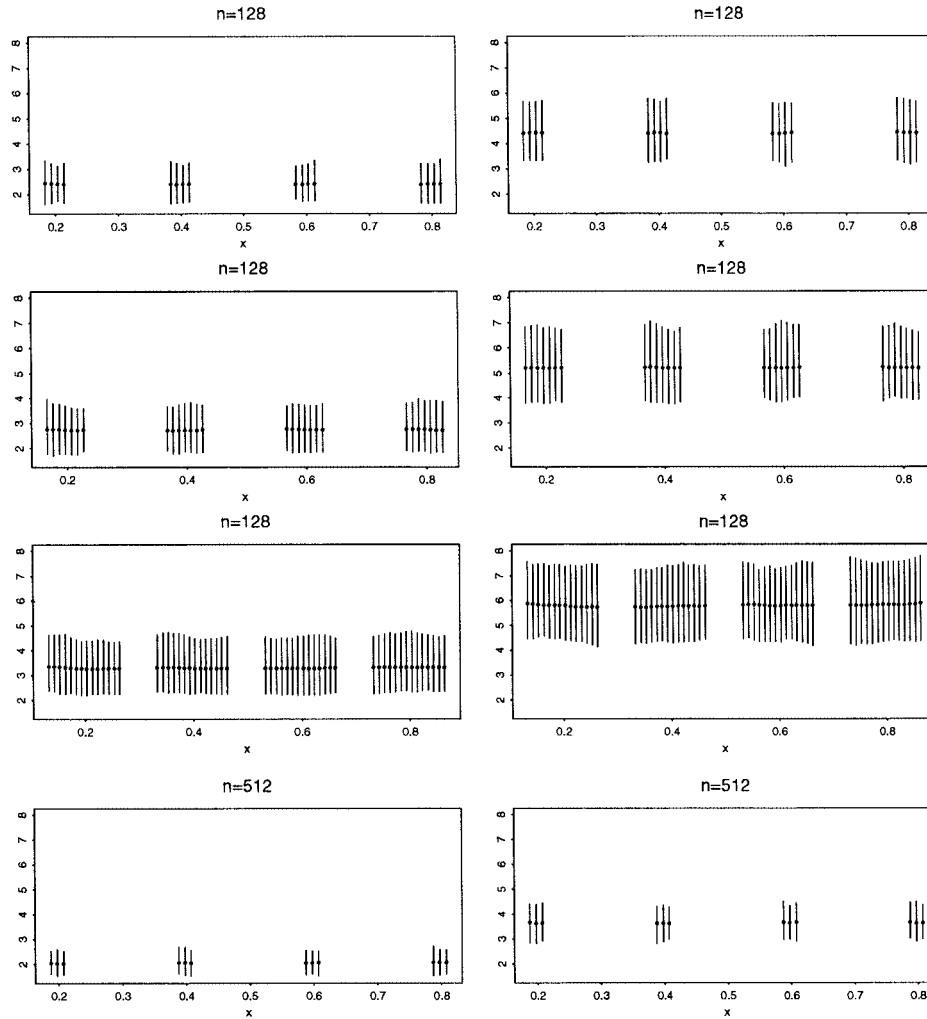


FIG. 1. $Length(x)$ of G -simultaneous confidence bands at all $x \in G$, method (*simult); the vertical bars indicate the range of the 2.5% and 97.5% sample quantiles of $length(x)$, the dots represent the average length $Ave[length(x)]$. The bandwidths, from top left to bottom right for each sample size n , are $(C_1 = 1, C_2 = 0.5)$, $(C_1 = 1, C_2 = 1)$, $(C_1 = 2, C_2 = 0.5)$, $(C_1 = 2, C_2 = 1)$, $(C_1 = 4, C_2 = 0.5)$, $(C_1 = 4, C_2 = 1)$.

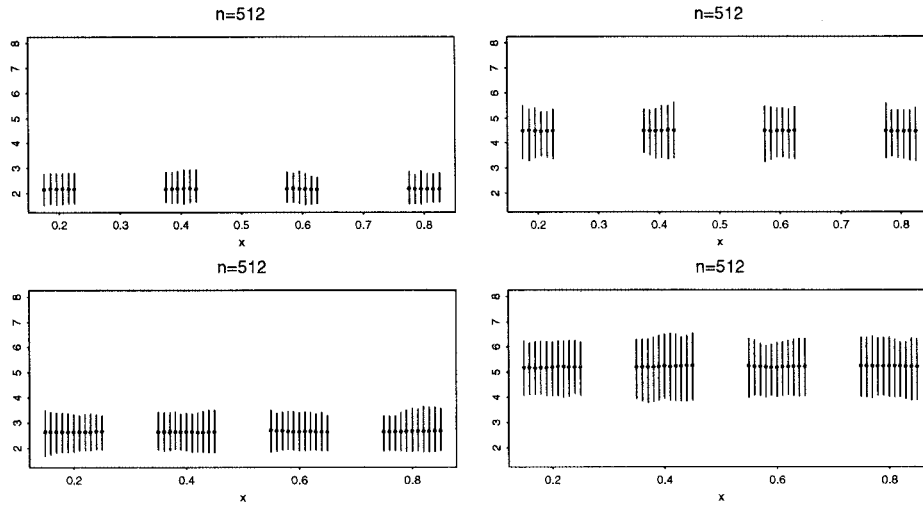


FIG. 1. (Continued).

The method (\mathcal{E} BC) based on extreme value approximation is very sensitive to the choice of bandwidths. The confidence band is either too progressive or too conservative. The latter is expected but the former is quite surprising and unpleasant, because the method is designed to be simultaneous over the whole interval $(0, 1)$. The reduction for length of the confidence band by increasing sample size is slow, reflecting the behavior of the limiting extreme value approximation.

The advantage of the sieve bootstrap is even more convincing for the construction of neighborhood simultaneous confidence intervals. For the case with a “rough” trend and a noise process with large bispectral density at zero, the sieve bootstrap is expected to be far better than the extreme value approximation; compare with the result in Section 5.1.1. Similar results as in Eubank and Speckman (1993) for the independent case could not be closely replicated. This may be due to the more complex time series structure and to the much stronger presence of noise: we choose $\text{Var}(Z_t) \approx 1$, whereas Eubank and Speckman (1993) use equivalents of $\text{Var}(Z_t) \leq 0.12$ (scaled by a factor 3 to approximately adjust to the magnitude of our curve). Also, it should be noted that Eubank and Speckman (1993) consider only the case of a “smooth” underlying trend and Gaussian noise, which leads to zero asymptotic skewness for the smoother.

5.2. Ozone measurements from Arosa. We apply here our sieve bootstrap method to the world’s longest series of monthly total ozone measurements from Arosa, Switzerland (1932–1996): due to some missing values we do not use the first stretch of the years 1926–1931. The measurements are currently

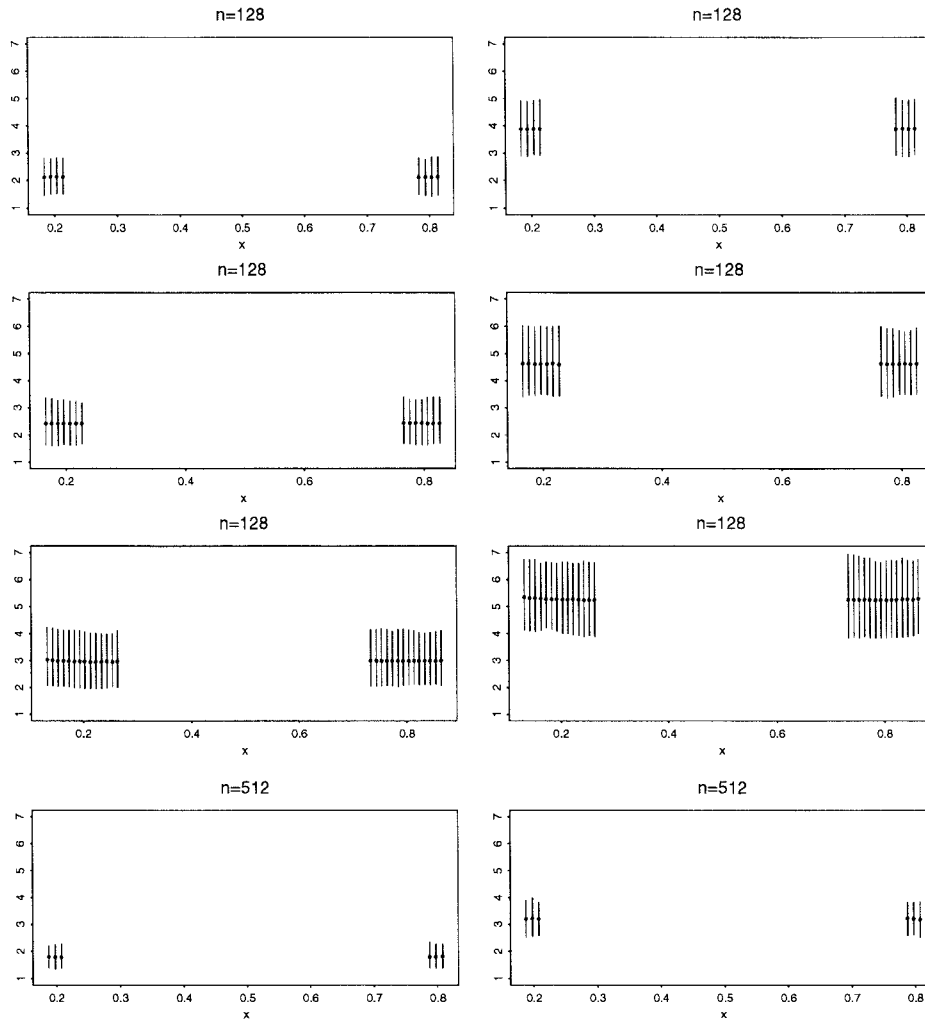


FIG. 2. Length(x) of G_{sub} -simultaneous confidence bands at all $x \in G_{\text{sub}}$, method (*simult). The vertical bars indicate the range of the 2.5% and 97.5% sample quantiles of length(x), the dots represent the average length Ave[length(x)]. The bandwidths, from top left to bottom right for each sample size n , are $(C_1 = 1, C_2 = 0.5)$, $(C_1 = 1, C_2 = 1)$, $(C_1 = 2, C_2 = 0.5)$, $(C_1 = 2, C_2 = 1)$, $(C_1 = 4, C_2 = 0.5)$, $(C_1 = 4, C_2 = 1)$.

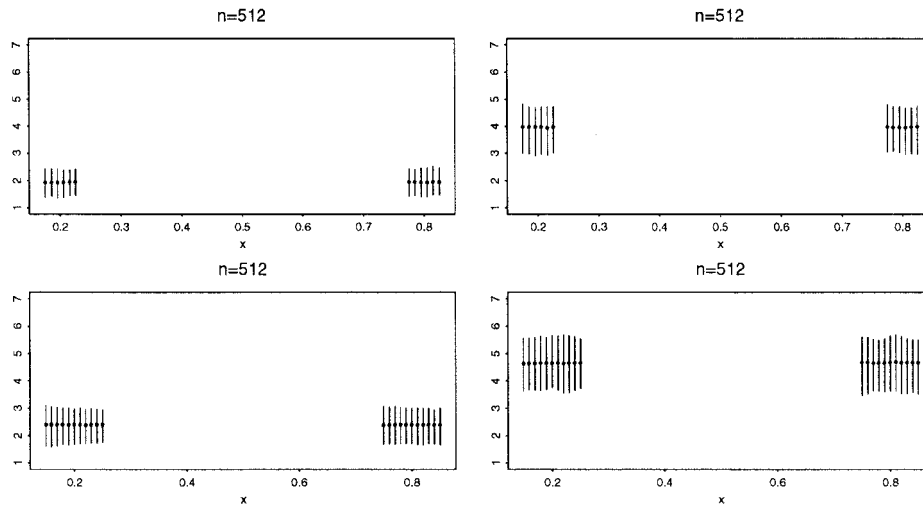


FIG. 2. (Continued).

performed by the Swiss Meteorological Institute. The homogenized data set is available from <http://www.umnw.ethz.ch/LAPETH/doc/totozon.html>.

Recently, a trend analysis has been given by Staehelin, Kegel and Harris (1997): based on a linear trend model, including other exogenous variables and an AR(2) noise process, a significantly decreasing trend has been found. It is now of interest, whether our nonparametric trend model in (3.1) exhibits a decreasing trend, even without using any other exogenous explanatory variables. The raw monthly measurements $\{O_t\}_t$ exhibit big seasonal effects, which can be explained very well. Assuming fixed monthly effects β_i , $i = 1, \dots, 12$, with $\sum_{i=1}^{12} \beta_i = 0$, we remove them by preliminary smoothing of the monthly data with a running mean, $Y_t = \sum_{i=-6}^6 c_i O_{t-i}$ with $c_i = 1/12$, $i = -5, \dots, 5$, and $c_i = 1/24$, $i = -6, 6$. Figure 3 displays the filtered data $\{Y_t\}_{t=1}^n$ with $n = 773$ on the Dobson scale, and its kernel smoother as in (3.2) with $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ and bandwidth $h = 2 \times 0.044n^{-1/5} = 0.024$, which is among the best in the simulations for simultaneous confidence bands [see also (5.6)]. The x -scale in Figure 3 is according to the trend model in (3.1): 12 months correspond to an amount of 0.016 on the x -scale.

We apply the sieve bootstrap as described in Section 2.1. For Step 1 we use as a pilot bandwidth the one which worked best in the simulations for simultaneous confidence bands, that is, $\tilde{h} = h^{5/9}/2 = 0.062$ [see also (3.3), in which we use $\delta = 0.05$]. In Step 2, the AIC optimal autoregressive order is $p = 15$, the coefficients are $(\hat{\phi}_{1,n}, \dots, \hat{\phi}_{15,n}) = (1.49, -0.59, 0.18, -0.11, 0.02, -0.05, 0.04, -0.01, 0.00, 0.09, -0.18, -0.14, 0.25, 0.08, -0.11)$. The decay is not very fast; our assumption about an AR(∞) noise process seems more realistic than any low order AR model.

Ozone measurements from Arosa

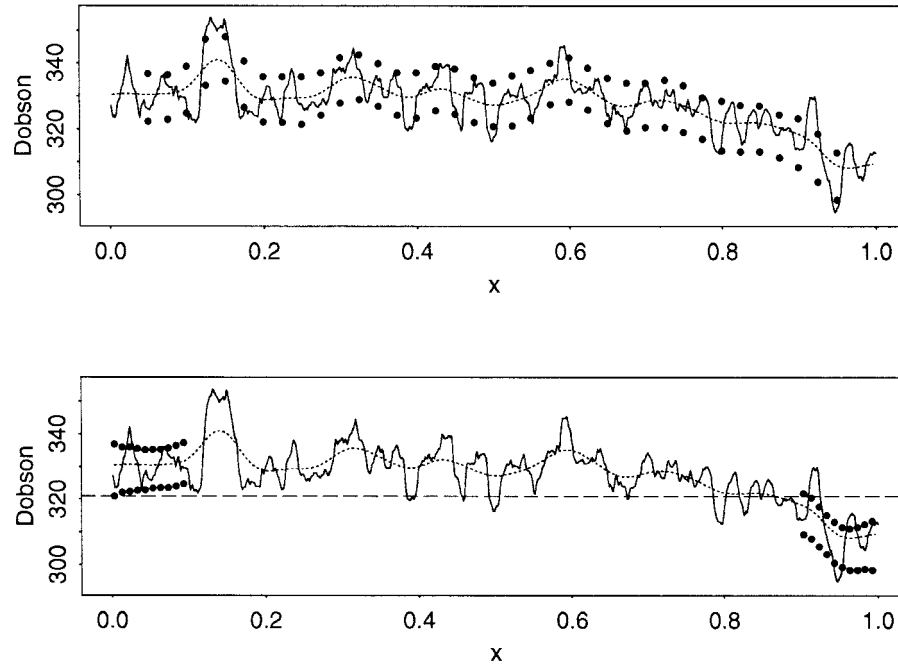


FIG. 3. The 95% simultaneous confidence regions for trend, based on the sieve bootstrap: (top) simultaneous over G_1 ; (bottom) simultaneous over G_2 . In both pictures (preliminary filtered) total ozone measurements are given by the solid line, the kernel smoother is given by the dotted line and the area of simultaneous confidence bands is indicated by dots at every x in G_1 and G_2 , respectively. At bottom is a dashed line showing the lowest value of the G_2 -simultaneous confidence region at early times.

We construct two types of simultaneous confidence regions, as described in Section 5.1.2, both on the level $1 - \alpha = 0.95$. One is over a set $G_1 = \{0.05 + j/40; j = 0, \dots, 36\}$, a set of equally spaced points over almost the whole range of $[0, 1]$; and another over a set $G_2 = U_{0.05} \cup U_{0.95}$, $U_x = \{x - 2h + j/100; j = 0, \dots, [400h] = 9\}$, which is an interesting set for comparing the beginning with the end of the series. Figure 3 displays both confidence regions.

The simultaneous region over G_1 in the top part of Figure 3 gives a reasonable idea about statistical variation over the whole range of time. From the simultaneous region over G_2 in the bottom part of Figure 3 we get that the confidence regions at $x \in U_{0.05} = \{0.003, 0.013, \dots, 0.093\}$ and $x \in \{0.95 - 2h + j/100; j = 1, \dots, 9\} = \{0.913, 0.923, \dots, 0.993\}$ do not intersect. These ranges of x correspond approximately to the early months $\{\text{April 1932}, \dots, \text{March 1938}\}$ and to the late months $\{\text{November 1990}, \dots, \text{January 1996}\}$, respectively. (Note that the simultaneous confidence region over the new set $G_3 = U_{0.05} \cup \{0.913, 0.923, \dots, 0.993\}$ is a subset of the

G_2 -simultaneous region at the corresponding x 's, implying that there is no overlap at all the early and all the late times of G_3 .) Our analysis supports the conclusion that there is a significant decreasing trend from the early times April 1932–March 1938 to the recent times November 1990–January 1996. Such a conclusion is consistent with the recent analysis in Staehelin, Kegel and Harris (1997). The novelty here is the significance within the framework of a nonparametric trend model with a nonparametric noise process in the form of an infinite-dimensional AR(∞) model.

6. Proofs. We first consider the effect of nonstationarity, that is, the effect of estimating the trend values $s(t)$, $t = 1, \dots, n$, in steps 1 and 4 of the sieve bootstrap scheme. We do not treat the case with fewer than n estimates as given in (3.3) separately, the modifications are obvious. We denote by $R_Z(j) = \text{Cov}(Z_0, Z_j)$ and, for $|j| \leq n-1$, $\hat{R}_Z(j) = n^{-1} \sum_{t=1}^{n-|j|} Z_t Z_{t+|j|}$, $\hat{R}_{\hat{Z}}(j)$ as in (2.2).

LEMMA 6.1. *Assume that (A1) and (A2) hold and $n^{-1} \sum_{t=1}^n (\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$, $n \rightarrow \infty$. If $p(n) = o((n/\log(n))^{1/2})$, then*

$$\max_{0 \leq j \leq p} |\hat{R}_{\hat{Z}}(j) - R_Z(j)| = O_P((\log(n)/n)^{1/2}) + O_P\left(\sum_{j=p+1}^{\infty} |\phi_j|\right) + O_P(b(n)^{1/2}).$$

PROOF. We write, for $|j| \leq n-1$,

$$\begin{aligned} \hat{R}_{\hat{Z}}(j) &= \hat{R}_Z(j) + n^{-1} \sum_{t=1}^{n-|j|} (\hat{Z}_{t,n} - Z_t) Z_{t+|j|} + n^{-1} \sum_{t=1}^{n-|j|} (\hat{Z}_{t+|j|,n} - Z_{t+|j|}) Z_t \\ &\quad + n^{-1} \sum_{t=1}^{n-|j|} (\hat{Z}_{t,n} - Z_t)(\hat{Z}_{t+|j|,n} - Z_{t+|j|}). \end{aligned}$$

Therefore, we get by the Cauchy–Schwarz inequality in a straightforward way,

$$\max_{0 \leq j \leq n-1} |\hat{R}_{\hat{Z}}(j) - \hat{R}_Z(j)| = O_P(b(n)^{1/2}).$$

Now we complete the proof by using the known bound

$$\max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)| = O((\log(n)/n)^{1/2}) + O\left(\sum_{j=p+1}^{\infty} |\phi_j|\right) \quad \text{almost surely}$$

[cf. Hannan and Kavalieris (1986), Theorem 2.1]. \square

Denote by ϕ_j the autoregressive coefficients as given in (2.1), by $\hat{\phi}_{j,n}$ the coefficients as defined in (2.2) and by $\phi_{j,n}$ the corresponding theoretical quantities with R_Z instead of $\hat{R}_{\hat{Z}}$, that is, $\Gamma_Z \Phi_p = -\gamma_Z$, $\Phi_p = (\phi_{1,n}, \dots, \phi_{p,n})'$.

LEMMA 6.2. *Assume that (A1) and (A2) hold, $n^{-1} \sum_{t=1}^n (\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$, $n \rightarrow \infty$, and $p(n) = o(\min\{(n \times (\log(n))^{-1})^{1/2}, b(n)^{-1/2}\})$. Then*

$$\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}| = O_P((\log(n)/n)^{1/2}) + O_P\left(\sum_{j=p+1}^{\infty} |\phi_j|\right) + O_P(b(n)^{1/2}),$$

and the same bound holds for $\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_j|$.

PROOF. Denote by $\hat{\Phi}_p = (\hat{\phi}_{1,n}, \dots, \hat{\phi}_{p,n})'$, $\Phi_p = (\phi_{1,n}, \dots, \phi_{p,n})'$. According to An, Chen and Hannan [(1982), formula (25)],

$$\hat{\Phi}_p - \Phi_p = -\Gamma_Z^{-1}[(\hat{\gamma}_Z - \gamma_Z) + (\hat{\Gamma}_Z - \Gamma_Z)(\hat{\Phi}_p - \Phi_p) + (\hat{\Gamma}_Z - \Gamma_Z)\Phi_p].$$

By denoting

$$\|z\|_{\infty} = \begin{cases} \max_{1 \leq j \leq d} |z_j|, & \text{for } z \text{ a } d \times 1 \text{ vector,} \\ \max_{1 \leq j \leq d_1} \sum_{i=1}^{d_2} |z_{ji}|, & \text{for } z \text{ a } d_1 \times d_2 \text{ matrix,} \end{cases}$$

we arrive at

$$\|(\hat{\Gamma}_Z - \Gamma_Z)(\hat{\Phi}_p - \Phi_p)\|_{\infty} \leq p \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)| \|\hat{\Phi}_p - \Phi_p\|_{\infty},$$

$$\|(\hat{\Gamma}_Z - \Gamma_Z)\Phi_p\|_{\infty} \leq \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)| \sum_{j=1}^p |\phi_{j,n}|.$$

Therefore

$$\|\hat{\Phi}_p - \Phi_p\|_{\infty} \leq \|\Gamma_Z^{-1}\|_{\infty} \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)| \left(1 + p \|\hat{\Phi}_p - \Phi_p\|_{\infty} + \sum_{j=1}^p |\phi_{j,n}|\right),$$

and hence

$$\begin{aligned} \|\hat{\Phi}_p - \Phi_p\|_{\infty} &\left(1 - \|\Gamma_Z^{-1}\|_{\infty} p \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)|\right) \\ &\leq \|\Gamma_Z^{-1}\|_{\infty} \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)| \left(1 + \sum_{j=1}^p |\phi_{j,n}|\right). \end{aligned}$$

By the assumption and Lemma 6.1, $p \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)| = o_P(1)$; by An, Chen and Hannan [(1982), page 929, 1-4], $\|\Gamma_Z^{-1}\|_{\infty} < \infty$; by Baxter's inequality, $\sum_{j=1}^p |\phi_{j,n}| \leq \sum_{j=1}^p |\phi_j| + \text{const.}$ $\sum_{j=p+1}^{\infty} |\phi_j| = O(1)$. This, together with the bound for $\|\hat{\Phi}_p - \Phi_p\|_{\infty} (1 - \|\Gamma_Z^{-1}\|_{\infty} p \max_{0 \leq j \leq p} |\hat{R}_Z(j) - R_Z(j)|)$ above and Lemma 6.1, completes the proof. \square

It is very helpful to represent

$$(6.1) \quad Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1,$$

$$(6.2) \quad Z_t^* = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^*, \quad \hat{\psi}_{0,n} = 1,$$

where $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = 1/\Phi(z)$, $\Phi(\cdot)$ as in (A2), $\hat{\Psi}_n(z) = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j = 1/\hat{\Phi}_n(z)$, $\hat{\Phi}_n(z) = \sum_{j=0}^p \hat{\phi}_{j,n} z^j$. This representation is possible by assuming (A2).

LEMMA 6.3. *Assume that (A1) and (A2) hold, $n^{-1} \sum_{t=1}^n (\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$, $n \rightarrow \infty$, and $p(n) = o(\min\{(n \cdot (\log(n))^{-1})^{1/4}, b(n)^{-1/4}\})$. Then the following holds:*

(i) *there exists a random variable n_1 such that*

$$\sup_{n \geq n_1} \sum_{j=0}^{\infty} j |\hat{\psi}_{j,n}| < \infty \quad \text{in probability,}$$

(ii) $\sup_{j \in \mathbb{N}} |\hat{\psi}_{j,n} - \psi_j| = O_P((\log(n)/n)^{1/2}) + O_P(p^{-1}) + O_P(b(n)^{1/2})$.

PROOF. Both statements follow by using Lemma 6.2 and then proceeding as in the proofs of Theorems 3.1 and 3.2 in Bühlmann (1995). \square

LEMMA 6.4. *Assume that (A1) and (A2) hold, $n^{-1} \sum_{t=1}^n (\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$, $n \rightarrow \infty$, and $p(n) = o(\min\{(n \cdot (\log(n))^{-1})^{1/2}, b(n)^{-1/2}\})$. Then*

$$\mathbb{E}^* |\varepsilon_t^*|^{2w} = \mathbb{E} |\varepsilon_t|^{2w} + o_P(1), \quad w = 1, 2.$$

PROOF. We write

$$\hat{\varepsilon}_{t,n} = \varepsilon_t + Q_{t,n} + R_{t,n} + U_{t,n} + V_{t,n},$$

where $Q_{t,n} = \sum_{j=0}^p (\hat{\phi}_{j,n} - \phi_{j,n}) Z_{t-j}$, $R_{t,n} = \sum_{j=0}^p (\phi_{j,n} - \phi_j) Z_{t-j}$, $U_{t,n} = \sum_{j=0}^p \hat{\phi}_{j,n} (s(t-j) - \tilde{s}(t-j))$, $V_{t,n} = -\sum_{j=p+1}^{\infty} \phi_j Z_{t-j}$. Now the proof is straightforward as in Bühlmann [(1997), Lemma 5.3]. The only additional quantities to control are $U_{t,n}$ and $V_{t,n}$: by the Cauchy–Schwarz inequality and Lemma 6.2,

$$(n-p)^{-1} \sum_{t=p+1}^n |U_{t,n}|^{2w} \leq [p^2 (\|\Phi_p\|_{\infty} + \|\hat{\Phi}_p - \Phi_p\|_{\infty})^2 O_P(b(n))]^w = o_P(1),$$

$$(n-p)^{-1} \sum_{t=p+1}^n |V_{t,n}|^{2w} = O_P\left(\left(\sum_{j=p+1}^{\infty} |\phi_j|\right)^{2w}\right) = o_P(1).$$

With these bounds we complete the proof. \square

PROOF OF THEOREM 3.1. Let $\mathcal{J}_n = \{[\delta n] + 1, \dots, [(1 - \delta)n]\}$ and $n' = |\mathcal{J}_n|$. In Hall and Hart (1990) it is shown that, for $0 < \delta < 1/2$,

$$\begin{aligned} n'^{-1} \sum_{t \in \mathcal{J}_n} (\tilde{s}(t) - s(t))^2 &= n'^{-1} \sum_{t \in \mathcal{J}_n} (\tilde{m}(t/n) - m(t/n))^2 \\ &= O_P(\max\{\tilde{h}^4, n^{-1}\tilde{h}^{-1}\}). \end{aligned}$$

This explains that for $p(n) = o(\min\{(n/\log(n))^{1/4}, (n\tilde{h})^{1/4}, \tilde{h}^{-1}\})$ Lemmas 6.2–6.4 are valid. It is also known [cf. Hall and Hart (1990)] that

$$(6.3) \quad B_{as}(x) = \lim_{n \rightarrow \infty} (nh)^{1/2} h^2 m^{(2)}(x) \int_{\mathbb{R}} x^2 K(x) dx / 2, \quad 0 < x < 1.$$

Moreover,

$$\sigma_{as}^2 = \lim_{n \rightarrow \infty} (nh)^{1/2} \text{Var}(\hat{m}(x)) = \sum_{k=-\infty}^{\infty} R_Z(k) \int_{\mathbb{R}} K^2(x) dx,$$

where $R_Z(k) = \text{Cov}(Z_0, Z_k)$.

Now we write

$$\begin{aligned} (nh)^{1/2}(\hat{m}(x) - \mathbb{E}[\hat{m}(x)]) &= (nh)^{1/2} \sum_{t=1}^n w_{t;n}(x) Z_t \quad \text{with} \quad \sum_{t=1}^n w_{t;n}(x) \sim 1, \\ (nh)^{1/2}(\hat{m}^*(x) - \mathbb{E}^*[\hat{m}^*(x)]) &= (nh)^{1/2} \sum_{t=1}^n w_{t;n}(x) Z_t^* \\ &\quad \text{with the same weights } w_{t;n}(x). \end{aligned}$$

These expressions are similar to an arithmetic mean for stationary variables. The limiting normal distribution for the above quantities can be proved in the same way; we just outline the case for the bootstrap. We are reasoning as in the proof of Theorem 3.1 in Bühlmann (1997). The idea is to replace Z_t^* by $Z_{t,M}^* = \sum_{j=0}^M \hat{\psi}_{j,n} \varepsilon_{t-j}^*$; compare with (6.2). Using the M -dependence of $\{Z_{t,M}^*\}_{t \in \mathbb{Z}}$ we prove convergence in P^* -distribution for $\sum_{t=1}^n w_{t;n}(x) Z_{t,M}^*$ via blocking and the Lindeberg central limit theorem (a possible choice for the block sizes is $a(n) = n^{1/2}h$, $b(n) = n^{1/2}h^{2.4}$). Then we prove that the truncation error $\sum_{t=1}^n w_{t;n}(x)(Z_t^* - Z_{t,M}^*)$ is asymptotically negligible. For these steps we use Lemmas 6.3 and 6.4. Summarizing, with the same arguments as in the proof of Theorem 3.1 in Bühlmann (1997), we get

$$(nh)^{1/2}(\hat{m}^*(x) - \mathbb{E}^*[\hat{m}^*(x)]) \rightarrow_{d^*} \mathcal{N}(0, \sigma_{as}^2) \quad \text{in probability.}$$

What remains to show is

$$(nh)^{1/2}(\mathbb{E}^*[\hat{m}^*(x)] - \tilde{m}(x)) - B_{as}(x) = o_P(1),$$

but this follows from the more general result in (6.9). \square

PROOF OF THEOREM 3.2. It suffices to show finite-dimensional convergence to the same normal distribution and stochastic equicontinuity. Consider the centered versions

$$\begin{aligned} W_{x_0;n}(x) &= Z_{x_0;n}(x) - \mathbb{E}[Z_{x_0;n}(x)], \\ W_{x_0;n}^*(x) &= Z_{x_0;n}^*(x) - \mathbb{E}^*[Z_{x_0;n}^*(x)]. \end{aligned}$$

We will now show that the centered versions are asymptotically equivalent to bias corrected versions of $Z_{x_0;n}(\cdot)$ and $Z_{x_0;n}^*(\cdot)$, respectively. We first show

$$(6.4) \quad \sup_{x \in [-1, 1]} |Z_{x_0;n}(x) - W_{x_0;n}(x) - B_{as}(x_0)| = o(1).$$

To do so, the following, well-known integral approximation bound will be used. If $g(\cdot)$ is a continuous and Riemann-integrable function, then

$$(6.5) \quad \left| n^{-1} \sum_{t=1}^n g\left(\frac{t}{n}\right) - \int_0^1 g(z) dz \right| \leq \omega_g(n^{-1}),$$

$$\omega_g(r) = \sup_{|x-y| \leq r} |g(x) - g(y)|.$$

Thus, by using (6.5),

$$(6.6) \quad \begin{aligned} & \sup_{x \in [-1, 1]} \left| Z_{x_0;n}(x) - W_{x_0;n}(x) - B_{as}(x_0) \right. \\ & \quad \left. - \left((nh)^{1/2} \left(\int_{\mathbb{R}} K(w)m(x_0 + xh - wh) dw \right. \right. \right. \\ & \quad \quad \left. \left. \left. - m(x_0 + xh) \right) - B_{as}(x_0) \right) \right| \\ & \leq (nh)^{1/2} \sup_{x \in [-1, 1]} \omega_{x,h}(n^{-1}), \end{aligned}$$

where

$$\omega_{x,h}(r) = \sup_{|z_1 - z_2| \leq r} h^{-1} \left| K\left(\frac{x_0 + xh - z_1}{h}\right) m(z_1) - K\left(\frac{x_0 + xh - z_2}{h}\right) m(z_2) \right|.$$

Since $K(\cdot)$ and $m(\cdot)$ are Lipschitz, we get

$$\sup_{x \in [-1, 1]} \omega_{x,h}(n^{-1}) \leq \text{const. } n^{-1} h^{-2}.$$

This, together with (6.6), implies

$$(6.7) \quad \begin{aligned} & \sup_{x \in [-1, 1]} \left| Z_{x_0;n}(x) - W_{x_0;n}(x) - B_{as}(x_0) \right. \\ & \quad \left. - \left((nh)^{1/2} \left(\int_{\mathbb{R}} K(w)m(x_0 + xh - wh) dw - m(x_0 + xh) \right) - B_{as}(x_0) \right) \right| \\ & \leq \text{const. } (nh^3)^{-1/2} = o(1). \end{aligned}$$

Also, by the Lipschitz property of $m^{(2)}(\cdot)$ and using the boundedness of the support of $K(\cdot)$,

$$(6.8) \quad \sup_{x \in [-1, 1]} \left| (nh)^{1/2} \left(\int_{\mathbb{R}} K(w) m(x_0 + xh - wh) dw - m(x_0 + xh) \right) - B_{as}(x_0) \right| = o(1).$$

By (6.7) and (6.8) we have shown (6.4).

Next we will show the analog of (6.4) for the bootstrap

$$(6.9) \quad \sup_{x \in [-1, 1]} |Z_{x_0; n}^*(x) - W_{x_0; n}^*(x) - B_{as}(x_0)| = o_P(1).$$

Write

$$(6.10) \quad \begin{aligned} & Z_{x_0; n}^*(x) - W_{x_0; n}^*(x) - B_{as}(x_0) \\ &= (nh)^{1/2} (R_1(x) - R_2(x)) + (nh)^{1/2} (D_1(x) - D_2(x)) - B_{as}(x_0), \end{aligned}$$

where $R_1(x) = \mathbb{E}^*[\hat{m}^*(x_0 + xh)] - \mathbb{E}[\mathbb{E}^*[\hat{m}^*(x_0 + xh)]]$, $R_2(x) = \tilde{m}(x_0 + xh) - \mathbb{E}[\tilde{m}(x_0 + xh)]$, $D_1(x) = \mathbb{E}[\mathbb{E}^*[\hat{m}^*(x_0 + xh)]]$, $D_2(x) = \mathbb{E}[\tilde{m}(x_0 + xh)]$.

By using (6.5) and similarly as in (6.6) and (6.7),

$$\begin{aligned} D_1(x) - D_2(x) &= \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} K(v) [m(x_0 + xh - uh - v\tilde{h}) \\ &\quad - m(x_0 + xh - v\tilde{h})] dv du + O(n^{-1}h^{-2}) \\ &= h^2 m^{(2)}(x_0) \int_{\mathbb{R}} u^2 K(u) du / 2 + O(h^2(h + \tilde{h})) + O(n^{-1}h^{-2}), \end{aligned}$$

where the O -terms are uniform in x [use the Lipschitz property of $K(\cdot)$, $m(\cdot)$ and $m^{(2)}(\cdot)$]. Thus, by the definition of $B_{as}(x_0)$,

$$(6.11) \quad \sup_{x \in [-1, 1]} |(nh)^{1/2} (D_1(x) - D_2(x)) - B_{as}(x_0)| = o(1).$$

Let us now consider the stochastic part $R_1(x) - R_2(x)$. For the fluctuations of $R_1(\cdot)$, we have, by the definition of $\tilde{m}(\cdot)$,

$$\begin{aligned} & \mathbb{E}|R_1(x) - R_1(y)|^2 \\ & \leq n^{-4} h^{-2} \tilde{h}^{-2} \sum_{t_1, t_2, s=1}^n \sum_{k=-n+1}^{n-1} \left| \left[K\left(\frac{x_0 + xh - t_1/n}{h}\right) - K\left(\frac{x_0 + yh - t_1/n}{h}\right) \right] \right. \\ & \quad \times \left[K\left(\frac{x_0 + xh - t_2/n}{h}\right) - K\left(\frac{x_0 + yh - t_2/n}{h}\right) \right] \\ & \quad \left. \times K\left(\frac{t_1/n - s/n}{\tilde{h}}\right) K\left(\frac{t_2/n - (s-k)/n}{\tilde{h}}\right) \right| |R_Z(k)|. \end{aligned}$$

Since the support of $K(\cdot)$ is bounded, say $K(\cdot) \subseteq [a, b]$, the indices t_1, t_2 range over a sub-set $I_{x_0, x, y, n, h}$ whose cardinality can be bounded by $\sup_{x, y \in [-1, 1]} |I_{x_0, x, y, n, h}| \leq (b-a)nh$. Likewise, s ranges over a subset $I_{x_0, x, y, n, \tilde{h}}$

whose cardinality is uniformly bounded by $(b - a)n\tilde{h}$. Therefore, by the Lipschitz property of $K(\cdot)$,

$$\sup_{x, y \in [-1, 1]} \frac{nh\mathbb{E}|R_1(x) - R_2(x)|^2}{|x - y|^2} = O\left(\frac{h}{\tilde{h}}\right),$$

and hence

$$(6.12) \quad \sup_{x, y \in [-1, 1]} n\tilde{h}\mathbb{E}|R_1(x) - R_2(x)|^2 \leq \text{const.}|x - y|^2.$$

A similar estimate applies for

$$\begin{aligned} & \mathbb{E}|R_2(x) - R_2(y)|^2 \\ & \leq n^{-2}\tilde{h}^{-2} \sum_{t=1}^n \sum_{k=-n+1}^{n-1} \left| \left[K\left(\frac{x_0 + xh - t/n}{\tilde{h}}\right) - K\left(\frac{x_0 + yh - t/n}{\tilde{h}}\right) \right] \right. \\ & \quad \times \left[K\left(\frac{x_0 + xh - (t-k)/n}{\tilde{h}}\right) \right. \\ & \quad \left. \left. - K\left(\frac{x_0 + yh - (t-k)/n}{\tilde{h}}\right) \right] \right| |R_Z(k)|. \end{aligned}$$

Again, the index t ranges over a set whose cardinality can be uniformly (in x, y) bounded by $\text{const.}n\tilde{h}$. Hence, by the Lipschitz property of $K(\cdot)$,

$$(6.13) \quad \sup_{x, y \in [-1, 1]} n\tilde{h}\mathbb{E}|R_2(x) - R_2(y)|^2 \leq \text{const.}|x - y|^2.$$

By (6.12) and (6.13), we have, for $R(x) = R_1(x) - R_2(x)$,

$$(6.14) \quad \sup_{x, y \in [-1, 1]} n\tilde{h}\mathbb{E}|R(x) - R(y)|^2 \leq \text{const.}|x - y|^2.$$

By (6.14) we can control the fluctuation of $R(\cdot)$ and get stochastic equicontinuity by Theorem 12.3 in Billingsley (1968),

$$(6.15) \quad \lim_{\gamma \searrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left[\sup_{|x-y| \leq \gamma} (n\tilde{h})^{1/2} |R(x) - R(y)| > \kappa \right] = 0, \quad \kappa > 0.$$

Moreover, it follows with some of the arguments for deriving (6.14): for $\kappa > 0$, there exists $M = M(\kappa)$ and $n_0 = n_0(\kappa)$ such that

$$(6.16) \quad \sup_{x \in [-1, 1]} \mathbb{P}[(n\tilde{h})^{1/2}|R(x)| > M] \leq \kappa \quad \text{for all } n \geq n_0.$$

Thus, a well-known discretization argument together with (6.15), (6.16) and the fact $h = o(\tilde{h})$ yields

$$(6.17) \quad \sup_{x \in [-1, 1]} (nh)^{1/2}|R(x)| = o_P(1).$$

Therefore, by (6.10), (6.11) and (6.17) we have shown (6.9).

Therefore, by (6.4) and (6.9) we can work with the centered processes $\{W_{x_0;n}(x)\}_{x \in [-1, 1]}$ and $\{W_{x_0;n}^*(x)\}_{x \in [-1, 1]}$. For the covariances of the process we get as in the proof of Theorem 3.1, by using the Lipschitz property of $K(\cdot)$,

$$(6.18) \quad \begin{aligned} \lim_{n \rightarrow \infty} \text{Cov}(W_{x_0;n}(x), W_{x_0;n}(y)) &= f_Z(0)2\pi \int_{\mathbb{R}} K(w)K(w+y-x) dw, \\ \text{Cov}(W_{x_0;n}^*(x), W_{x_0;n}^*(y)) &= f_Z(0)2\pi \int_{\mathbb{R}} K(w)K(w+y-x) dw \\ &\quad + o_P(1). \end{aligned}$$

Finite-dimensional convergence of $(W_{x_0;n}(x_1), \dots, W_{x_0;n}(x_m))'$ and $(W_{x_0;n}^*(x_1), \dots, W_{x_0;n}^*(x_m))'$, $x_1, \dots, x_m \in [-1, 1]$, $m \in \mathbb{N}$, to the same normal limiting distribution follows as in Theorem 3.1, by applying the Cramér–Wold device and using (6.18).

It remains to show stochastic equicontinuity. For $\kappa > 0$, $\eta > 0$, there exists $\gamma > 0$ and $n_0 \in \mathbb{N}$, such that

$$(6.19) \quad \begin{aligned} \mathbb{P} \left[\sup_{|x-y| \leq \gamma} |W_{x_0;n}(x) - W_{x_0;n}(y)| > \kappa \right] &< \eta \quad \text{for all } n \geq n_0, \\ \mathbb{P} \left[\mathbb{P}^* \left[\sup_{|x-y| \leq \gamma} |W_{x_0;n}^*(x) - W_{x_0;n}^*(y)| > \kappa \right] < \eta \right] &> 1 - \eta \quad \text{for all } n \geq n_0. \end{aligned}$$

By using the Lipschitz property of $K(\cdot)$, and similarly as in the derivation for (6.13) we get

$$\begin{aligned} \mathbb{E} |W_{x_0;n}(x) - W_{x_0;n}(y)|^2 &\leq \text{const.} |x - y|^2, \\ \mathbb{E}^* |W_{x_0;n}^*(x) - W_{x_0;n}^*(y)|^2 &\leq O_P(1) |x - y|^2. \end{aligned}$$

(For the bootstrap, we use Lemmas 6.3 and 6.4: the $O_P(1)$ term is uniformly over x, y .)

By applying Theorem 12.3 in Billingsley (1968), stochastic equicontinuity as stated in (6.19) follows. \square

PROOF OF THEOREM 3.4. Before proving this theorem we show the following.

LEMMA 6.5. *Assume that (A1) and (A2) hold, $n^{-1} \sum_{t=1}^n (\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$, $n \rightarrow \infty$, and $p(n) = o(\min\{(n \times (\log(n))^{-1})^{1/4}, b(n)^{-1/4}\})$. Then the following hold:*

- (i) $\sum_{k_1, k_2 = -\infty}^{\infty} |k_1 k_2 \mathbb{E}[Z_0 Z_{k_1} Z_{k_2}]| < \infty$;
- (ii) *there exists a random variable n_1 such that*

$$\sup_{n \geq n_1} \sum_{k_1, k_2 = -\infty}^{\infty} |k_1 k_2 \mathbb{E}^*[Z_0^* Z_{k_1}^* Z_{k_2}^*]| < \infty \quad \text{in probability};$$

- (iii) $\sum_{k_1, k_2 = -n+1}^{n-1} \mathbb{E}^*[Z_0^* Z_{k_1}^* Z_{k_2}^*] - \sum_{k_1, k_2 = -n+1}^{n-1} \mathbb{E}[Z_0 Z_{k_1} Z_{k_2}] = o_P(1)$, $n \rightarrow \infty$.

PROOF. Represent $Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, $Z_t^* = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^*$ [see (6.1) and (6.2)]. In the same spirit as for proving Lemma 6.4 one can show

$$\mathbb{E}^*[(\varepsilon_t^*)^3] = \mathbb{E}[\varepsilon_t^3] + o_P(1).$$

Then the conclusions follow straightforwardly by using Lemma 6.3.

Since the kernel $K(\cdot)$ is Lipschitz continuous,

$$\max_{1 \leq t \leq n} \left| K\left(\frac{x - (t+k)/n}{h}\right) - K\left(\frac{x - t/n}{h}\right) \right| \leq \text{const.} k(nh)^{-1}.$$

Using this estimate and Lemma 6.5, we immediately obtain the conclusions from Theorem 3.4. \square

Acknowledgments. I thank David Brillinger for many helpful discussions and a referee for constructive comments. I also thank Johannes Staehelin for providing the ozone data set.

REFERENCES

- ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristics*. Springer, New York.
- ALTMAN, N. S. (1990). Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.* **85** 749–759.
- AN, H. Z., CHEN, Z.-G. and HANNAN, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10** 926–936. (Correction: *Ann. Statist.* **11** 1018.)
- BICKEL, P. J. and BÜHLMANN, P. (1995). Mixing property and functional central limit theorems for a sieve bootstrap in time series. Technical Report 440, Dept. Statistics, Univ. California, Berkeley.
- BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- BOSE, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Ann. Statist.* **16** 1709–1722.
- BRILLINGER, D. R. (1988). An elementary trend analysis of Rio Negro levels at Manaus, 1903–1985. *Revista Brasileira de Probabilidade e Estatística* **2** 63–79.
- BRILLINGER, D. R. (1996). Some uses of cumulants in wavelet analysis. *J. Nonparametr. Statist.* **6** 93–114.
- BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.
- BÜHLMANN, P. (1995). Moving-average representation for autoregressive approximations. *Stochastic Process. Appl.* **60** 331–342.
- BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148.
- EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301.
- FAN, J. (1993). Local linear smoothers and their minimax efficiency. *Ann. Statist.* **21** 196–216.
- FREEDMAN, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.
- GASSER, T. and MÜLLER, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, eds.). Springer, Berlin.
- GASSER, T. and MÜLLER, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–185.
- GASSER, T., MÜLLER, H.-G. and MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

- HALL, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695–711.
- HALL, P. and HART, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351.
- HANNAN, E. J. and KAVALIERS, L. (1986). Regression, autoregression models. *J. Time Series Anal.* **7** 27–49.
- HÄRDLE, W. and BOWMAN, A. W. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 102–110.
- HÄRDLE, W. and MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796.
- HÄRDLE, W. and TUAN, P.-D. (1986). Some theory of M -smoothing of time series. *J. Time Series Anal.* **7** 191–204.
- HART, J. D. (1991). Kernel regression estimation with time series errors. *J. Roy. Statist. Soc. Ser. B* **53** 173–187.
- HART, J. D. (1994). Automated kernel smoothing of dependent data using time series cross-validation. *J. Roy. Statist. Soc. Ser. B* **56** 529–542.
- HERRMANN, E., GASSER, T. and KNEIP, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* **79** 783–795.
- KREISS, J.-P. (1988). Asymptotic statistical inference for a class of stochastic processes. Habilitationsschrift, Fachbereich Mathematik, Univ. Hamburg.
- KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241.
- NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **10** 186–190.
- SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164.
- STAEHELIN, J., RENAUD, A., BADER, J., MCPETERS, R., VIATTE, P., HOEGGER, B., BUGNION, V., GIROUD, M. and SCHILL, H. (1997). Total ozone series at Arosa (Switzerland): homogenization and data comparison. *Journal of Geophysical Research*. To appear.
- STAEHELIN, J., KEGEL, R. and HARRIS, N. R. P. (1997). Trend analysis of the homogenised total ozone series of Arosa (Switzerland), 1926–1996. *Journal of Geophysical Research*. To appear.
- SUBBA RAO, T. and GABR, M. M. (1980). A test for linearity of stationary time series. *J. Time Series Anal.* **1** 145–158.
- TRUONG, Y. K. (1991). Nonparametric curve estimation with time series errors. *J. Statist. Plann. Inference* **28** 167–183.
- WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

SEMINAR FÜR STATISTIK, SOL
ETH ZENTRUM
CH-8092 ZÜRICH
SWITZERLAND
E-MAIL: buhlmann@stat.math.ethz.ch