Confidence Regions for Trends in Time Series: a Simultaneous Approach with a Sieve Bootstrap

Peter Bühlmann *
University of California
Berkeley

January 1996

Abstract

We study a sieve bootstrap procedure for time series with a deterministic trend. The sieve for constructing the bootstrap is based on autoregressive approximation. Given time series data, one would first use a preliminary estimate of the trend of the underlying time series and then approximate the noise process by a large autoregressive model of increasing order as the sample size grows. The bootstrap scheme is based on resampling estimated innovations of fitted autoregressive models.

We show the validity of such sieve bootstrap approximations for the limiting distribution of linear trend estimators, such as general regression predictors or kernel smoothers. This bootstrap scheme can then be used to construct simultaneous confidence intervals for the trend, where the simultaneity can be achieved over a range of points which can be chosen by the user.

The time series context is substantially different from the independent set-up: methods from the independent, adapted to the dependent case, seem to loose much of their accuracy. Our resampling procedure yields satisfactory results in a simulation study for finite sample sizes.

Key words and phrases. AIC, AR(∞), ARMA, autoregressive approximation, linear regression, kernel smoothing, MA(∞), nonparametric regression, second order correctness, wavelets.

Short title: Confidence Regions for Time Series Trends

*Research supported in part by the Swiss National Science Foundation.

1 Introduction

We are considering the problem of making confidence statements for a trend in a time series. More precisely, we consider the model $Y_t = s(t) + Z_t,\ t \in \mathbb{Z}$, where $\{s(t)\}_{t\in\mathbb{Z}}$ is a deterministic trend (or signal) and $\{Z_t\}_{t\in\mathbb{Z}}$ a stationary noise 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.

In the independent set-up, where the noise is white, several researchers proposed bootstrap methods to construct more reliable interval estimators, cf. Freedman (1981, 1984), Bickel and Freedman (1983), Wu (1986), Härdle and Bowman (1988), Hall (1989, 1992), Härdle and Marron (1991). There are several reasons for this. The bootstrap, when used correctly, exhibits a second order property and hence would usually yield a better coverage probability for finite sample size. The bootstrap is able to correct for bias (for biased estimation procedures such as smoothers), cf. Härdle and Marron (1991), and it has the potential to yield simultaneous confidence bands, cf. Härdle and Marron (1991).

In the time series context, things become inherently more complex. An estimate for the asymptotic variance in a normal approximation for a trend estimator typically involves an estimate of the noise spectral density, often at frequency zero. It is exactly here, where the difference between independence and dependence becomes crucial. In the independent case a normal approximation requires only estimation of the variance of an observation, for this one can use $\sqrt{n}$-consistent estimators which are based directly on the observed values. However, for estimating the spectral density of the underlying noise process the $\sqrt{n}$-consistency is lost, and one seems to be forced to use estimated residuals and. This methodological difference is found to be serious enough so that methods, which work well in the independent set-up, cf. Eubank and Speckman (1993), break down when adapted to the time series case.

Our approach relies on a bootstrap procedure which does not rely on an explicit estimate for the spectral density of the noise process. Unfortunately, there is not such a unique bootstrap procedure for time series as Efron's (1979) bootstrap for the independent set-up. A procedure that works for the rich sub-class of linear stationary processes $Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ which can be inverted and represented as $\sum_{j=0}^{\infty} \phi_j Z_{t-j} = \varepsilon_t$, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ being an i.i.d. sequence with mean zero, is based on bootstrapping autoregressive processes of order infinity (AR(∞)), cf. Kreiss (1988), Bühlmann (1995b). Given data, the idea is to select an autoregressive order, increasing with the sample size, fit an autoregressive model and use Efron's bootstrap on estimated residuals. In the limit for sample size n tending to infinity, we would fit an AR(∞). Think of this as an approximation with an increasing family of parametric models, this then explains the terminology 'sieve bootstrap', cf. Bühlmann (1995b). This approach has the advantage that no particular finite parameter model for the noise process is assumed. It is shown in Bühlmann (1995b) that for many linear processes, a sieve bootstrap for AR(∞) models has generally a better performance than some nonparametric block-based bootstrap technique as proposed by Künsch (1989).

We extend here the sieve bootstrap for stationary AR(∞) processes to non-stationary time series with a deterministic trend and an AR(∞) noise process. We will argue, also by results from a simulation study, that the sieve bootstrap is superior over a normal approximation or over an extreme value approximation for constructing simultaneous confidence intervals for the underlying trend, and we will prove how bias in kernel smoothing can be corrected in an automatic way by the sieve bootstrap. Our extension is a new contribution in model free resampling for non-stationary time series and can serve as a new tool in time series analysis.

The article is organized as follows. In section 2 we describe our bootstrap scheme, in section 3 we give asymptotic results for bootstrapping parametric linear trend estimators as well as for nonparametric smoothers, in section 4 we present some alternatives, which all suffer from the fact that the spectral density of the noise process is estimated via estimated residuals, in section 5 we show results from a simulation study and in section 6 we give some concluding remarks. The theoretical arguments and proofs are given in an appendix.

2 The model and the resampling procedure

Consider the model

$$Y_t = s(t) + Z_t,\ t \in \mathbb{Z}, \tag{2.1}$$

with $\{s(t)\}_{t\in\mathbb{Z}}$ a deterministic trend (or signal) and $\{Z_t\}_{t\in\mathbb{Z}}$ a zero mean stationary noise process. We restrict ourselves to the case where $\{Z_t\}_{t\in\mathbb{Z}}$ is an autoregressive process of order infinity (AR(∞)), i.e.,

$$\sum_{j=0}^{\infty} \phi_j Z_{t-j} = \varepsilon_t, \phi_0 = 1, \tag{2.2}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. sequence with expectation $\mathbf{E}[\varepsilon_t] = 0$.
The object to recover is the deterministic trend which is thought to be a function $s(.) : \mathbb{Z} \to \mathbb{R}$. Given observations $Y_1, \ldots, Y_n$, there are various approaches to do so, we mention now two generic examples which are discussed through the whole article.

Example 2.1 (Finite parameter linear model).
Consider the trend function

$$s(x) = s_n(x) = \sum_{j=1}^{J} \beta_j g_{j;n}(x),\ x \in \mathbb{R},$$

for some real-valued regressors $g_{j;n}(.)$ $(j = 1, \ldots, J)$ which can depend on the sample size n, and for some coefficients $\beta_1, \ldots \beta_J$ in $\mathbb{R}$.
The least-squares estimates $\hat{\beta}_1, \ldots, \hat{\beta}_J$ are in a broad range of situations asymptotically BLUE, some sufficient conditions are known as 'Grenander conditions', cf. Grenander and Rosenblatt (1984, Ch. 7.3), Hannan (1971,Ch. IV.3).

Example 2.2 (Nonparametric kernel smoothing).
Assume that the trend function is $s(t) = m(t/n),\ t = 1, \ldots, n$ for some function m :

$[0,1] \to \mathbb{R}$, where n denotes the sample size. Consider estimation of $m(.)$ by kernel smoothing

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

where $h = h(n) = o(1),\ h(n)^{-1} = o(n)\ (n \to \infty)$ is a bandwidth parameter and $K(.)$ a kernel function. This is the so called Nadaraya-Watson kernel estimator, cf. Nadaraya (1964) and Watson (1964), other kernel smoothers like the Gasser-Müller type (Gasser and Müller, 1979) could also be considered. Kernel smoothing with dependent errors has been studied by Härdle and Tuan (1986), Hall and Hart (1990), Altman (1990), Truong (1991), Hart (1991, 1994), Herrmann et al. (1992). An estimate of the trend function $s(.)$ is then derived by setting $\hat{s}(t) = \hat{m}(t/n),\ t \in \mathbb{N}$ with $1 \le t \le n$.

2.1 A sieve bootstrap scheme

We develop here a general approach for constructing confidence bands of the unknown trend $s(.)$. It is based on an extension of the bootstrap for stationary autoregressive processes of order infinity (AR(∞)), cf. Kreiss (1988), Bühlmann (1995b), Bickel and Bühlmann (1995). We define now our bootstrap scheme. Assume that we have observations $Y_1, \ldots, Y_n$ being realizations of the model as given in (2.1) and (2.2).

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

$$Z_{t,n} = Y_t - \tilde{s}(t),\ t = 1, \ldots, n.$$

Step 2: Assume $p = p(n) \to \infty,\ p(n) = o(n)\ (n \to \infty)$. Fit an autoregressive model of order $p = p(n)$ to the residuals $Z_{t,n},\ t = 1, \ldots, n$, i.e., compute $\hat{\phi}_{1,n}, \ldots \hat{\phi}_{p,n}$ based on $\{Z_{t,n}\}_{t=1}^{n}$ for the autoregressive coefficients $\phi_1, \ldots \phi_p$ in formula (2.2). Then form another set of residuals

$$\varepsilon_{t,n} = \sum_{j=0}^{p} \hat{\phi}_{j,n} Z_{t-j,n},\ t = p+1, \ldots, n,\ \hat{\phi}_{0,n} = 1.$$

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

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

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

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

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

Then generate bootstrap observations by setting

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

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

The estimates $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$ in step 2 can be obtained by the Yule-Walker method. They are defined by

$$\hat{\Gamma}_{\hat{Z}} \hat{\phi}_p = -\hat{\gamma}_{\hat{Z}}, \tag{2.3}$$

where $\hat{\Gamma}_{\hat{Z}} = [\hat{R}_{\hat{Z}}(i-j)]_{i,j=1}^{p}$, $\hat{\gamma}_{\hat{Z}} = (\hat{R}_{\hat{Z}}(1), \ldots, \hat{R}_{\hat{Z}}(p))'$, $\hat{\phi}_p = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n})'$, $\hat{R}_{\hat{Z}}(j) = n^{-1} \sum_{t=1}^{n-|j|} Z_{t,n} Z_{t+|j|,n}$, $|j| \leq n-1$.

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^*.

Fitting an autoregressive model of growing order $p = p(n)$ in step 2 is a sieve procedure for the true underlying AR(∞) process, that is why we call this resampling scheme 'sieve bootstrap', cf. Bühlmann (1995b). All the bootstrap quantities are denoted by an asterisk *. This sieve bootstrap allows to resample observations $Y_1^*, \ldots Y_n^*$ for various purposes. In this article we will only consider bootstrapped trend estimates $\hat{s}^*(.)$ which are defined by the plug-in rule, i.e., if $\hat{s}(x) = T_{n;x}(Y_1, \ldots, Y_n)$ is a measurable function of the observations, then the bootstrap estimate is defined by $\hat{s}^*(x) = T_{n;x}(Y_1^*, \ldots, Y_n^*)$. Note that the estimate $\hat{s}(x)$ above is not necessarily the same as in step 1 and step 4 of our sieve bootstrap. In the example 2.2 we rather would take for $\tilde{s}(t) = \tilde{m}(t/n)$ an oversmoothed kernel estimator, cf. Härdle and Marron (1991).

2.2 Choice of the approximating order p

The approximating order $p = p(n)$ in step 2 of our bootstrap scheme acts as some kind of smoothing parameter. We discuss now the question of a 'good' choice of p. Since the true underlying error process is assumed to be an AR(∞) as defined in (2.2) the choice of the order p determines the tradeoff between a variance and bias part in the autoregressive approximation.

We propose the choice via the AIC criterion in an increasing range such as $[-A_n, A_n]$ with $A_n \to \infty$, $A_n = o(n^{1/2})$. This is motivated by the fact that the AIC criterion leads to an asymptotically efficient selection if the underlying AR(∞) process is observed, cf. Shibata (1980). In our case, we first estimate a trend in step 1 of the procedure and then fit an autoregressive model in step 2. One expects that the estimation in step 1 causes additional variability so that for choosing an autoregressive model in step 2 we should add an additional penalty term for high order models, due to the variability from the preliminary estimation in step 1. By neglecting this additional variability we tend to choose an autoregressive model with too many parameters. But such kind of overfitting guarantees a resistance against a severe bias by paying a price for a larger variance part.

Another concept relies on the idea of prewhitening, as a graphical device. For some candidates p, we fit the autoregressive model in step 2, obtain the residuals $\varepsilon_{t,n}$ ($t = p+1, \ldots, n$) and compute some spectral density estimate based on these $\varepsilon_{t,n}$'s. We would choose p such that this estimated spectral density is close to a constant.

2.3 Confidence bands via bootstrap

The sieve bootstrap can be applied for constructing confidence bands $I_n(x)$, pointwise or, more remarkably, for simultaneous bands within a certain set K. There exist also analytical approximations for uniform confidence bands, we discuss them in section 4.3. In practice, we always draw simultaneous confidence bands for a finite set of points, to emphasize the finite-ness of such a band we sometimes use the terminology 'confidence grid' over a finite set G.

The shape of a simultaneous grid over a set G is not canonical, in an asymptotic sense we only require that $\lim_{n\to\infty} \mathbb{P}[s(x) \in I_n(x) \text{ for all } x \in G] = 1-\alpha,\ 0 \le \alpha \le 1$. This shows that one has not 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 non-equal width, it seems evident that at some grid points there is less variability (at least for finite sample sizes) than at others.

We propose to enlarge pointwise intervals until a certain probability level for a simultaneous grid over a finite set G is reached, cf. Härdle and Marron (1991). Say, we want to construct a simultaneous confidence grid over G to the level $1-\alpha$ by using B sieve bootstrap replicates $\hat{s}_1^*(.), \ldots, \hat{s}_B^*(.)$. Then we proceed as follows.

(i) Compute pointwise confidence intervals $I_n(x; 1-\alpha_P)$ to the pointwise level $\alpha_P < \alpha$ for every grid-point $x \in G$,

$$I_n(x; 1-\alpha_P) = [\hat{s}(x) - \hat{q}_{1-\alpha_P/2}, \hat{s}(x) - \hat{q}_{\alpha_P/2}],$$

where $\hat{q}_\alpha = \inf\{y; B^{-1}\sum_{i=1}^B 1_{[\hat{s}_i^*(x)-\tilde{s}(x)\le y]} \ge \alpha\}$ is a pointwise bootstrap quantile.

(ii) Vary the pointwise error α_P until

$$B^{-1}\sum_{i=1}^{B} 1_{[\hat{s}_i^*(x)\in I_n(x;1-\alpha_P) \text{ for all } x\in G]} \approx 1-\alpha$$

(the percentage of bootstrap curves within the confidence grid is approximately $1-\alpha$).

Denote such an α_P by α_S. Then $\{I_n(x, \alpha_S)\}_{x\in G}$ is an approximate simultaneous confidence grid over the set G to the level $1-\alpha$.

This conception of a bootstrap algorithm for simultaneous confidence grids is very attractive because the user can specify exactly in what range of the data simultaneity should be considered, i.e., one can specify a set G of own choice. Unlike with simultaneous intervals of the Scheffé-type (see section 4.2), one can avoid with the bootstrap a simultaneous region which is unreasonably large or dense, as for example the whole space of all theoretically possible (but not necessarily practically likely) observations.

3 Bootstrapping trend estimates

We are going to discuss now the validity of the sieve bootstrap scheme for linear trend estimators $\hat{s}(.)$. For the noise process and its estimation we make the following assumptions.

(A1) Model (2.2) holds with $\mathbb{E}|\varepsilon_t|^s < \infty$ for some $s \geq 4$.

(A2) The AR(∞) transfer function $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$ $(z \in \mathbb{C})$ of the model (2.2) satisfies: $\Phi(z)$ is bounded away from zero for $|z| \leq 1$ and $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$ for some integer $r > 0$.

(A3) In step 2 of the sieve bootstrap scheme, $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$ are defined by formula (2.3).

These are the same conditions as for the sieve bootstrap in the stationary case, cf. Bühlmann (1995b). The conditions are met by a broad subclass of linear processes, including many ARMA models. In particular, they also hold for models with a non-exponential decay of autoregressive coefficients, which represent a stronger dependence than usual ARMA-processes. The restriction in (A3) is convenient for computational reasons. To prevent against severe bias one could taper the time series first before calculating the Yule-Walker estimates, cf. Dahlhaus (1983).

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 (1995b). In a rough way we can say that the only additional condition for our nonstationary sieve bootstrap is on the growth for the approximating order $p = p(n)$ in step 2, which can now additionally depend on the behavior of

$$n^{-1} \sum_{t=1}^{n} (\tilde{s}(t) - s(t))^2,$$

cf. Lemma A.1. Typically this bound is negligible so that no real additional restrictions arise. Our theoretical arguments in the appendix give some good evidence that our sieve bootstrap scheme works for reasonably regular linear predictors $\hat{s}_n(x) = \sum_{t=1}^{n} w_{t;n}(x) Y_t$, if the approximating order $p = p(n)$ in step 2 satisfies

$$p(n) \to \infty,\ p(n) = o\left(min\{(n/log(n))^{1/4}, (n^{-1} \sum_{t=1}^{n} (\tilde{s}(t) - s(t))^2)^{1/2}\} \right)\ (n \to \infty)$$

and (A1)-(A3) hold.

To keep things simpler we focus here on some more special cases, we consider global and local trend estimators as given by example 2.1 and 2.2 respectively.

3.1 Linear regression predictors

Consider the model (2.1) where the trend function $s(.)$ is a linear regression model as in example 2.1, i.e.,

$$s(x) = s_n(x) = \sum_{j=1}^{J} \beta_j g_{j;n}(x), x \in \mathbb{R}$$

for some real-valued regressors $g_{j;n}(.)$ $(j = 1, \ldots, J)$, which can depend on the sample size n, and for some coefficients $\beta_1, \ldots \beta_J$ in $\mathbb{R}$.

We consider now the least-squares predictor for $s(.)$, that is, the trend estimate is given by

$$\hat{s}(x) = \sum_{j=1}^{J} \hat{\beta}_j g_{j;n}(x),\ x \in \mathbb{R}, \tag{3.1}$$

where $\hat{\beta}_1, \ldots, \hat{\beta}_J$ are the ordinary least-squares estimates of $\beta_1, \ldots, \beta_J$.
Under some conditions, known as Grenander's conditions, one has asymptotic normality for any point $x \in \mathbb{R}$,

$$a(n;x)(\hat{s}(x) - s(x)) = \sum_{j=1}^{J} d_j(n)(\hat{\beta}_j - \beta_j)g_{j;n}(x) \xrightarrow{d} \mathcal{N}(0, \sigma_{as}^2(x)),\ 0 < \sigma_{as}^2(x) < \infty,$$

where $d_j(n)^2 = \sum_{t=1}^{n} g_{j;n}(t)^2,\ j = 1, \ldots, J$. These normalizing constants then determine the normalizing constant $a(n;x)$. We remark here, that if $d_j(n)^2 = d(n)^2$ for all j, then the normalizing constant $a(n;x) = a(n) = d(n)$. A familiar example is trigonometric regression, where $g_{j;n}(x) = g_j(x) = cos(\omega_j t),\ -\pi \leq \omega_j < \pi,\ d_j(n)^2 = n$ and $a(n;x) = a(n) = \sqrt{n}$. We will describe now under which circumstances our bootstrapped estimates converge to the appropriate Gaussian limiting distribution as indicated above. We assume the following conditions.

(R1) The trend is of the form $s(x) = s_n(x) = \sum_{j=1}^{J} \beta_j g_{j;n}(x),\ x \in \mathbb{R}$, its estimate $\hat{s}(x)$ is given by ordinary least-squares as in formula (3.1) and the estimates $\{\tilde{s}(t)\}_{t=1}^{n}$ in steps 1 and 4 of the sieve bootstrap scheme are the same least-squares estimate $\tilde{s}(t) = \hat{s}(t), t = 1, \ldots, n$.
For the regressors we assume the 'Grenander conditions':

$$\lim_{n\to\infty} \sum_{t=1}^{n} g_{j;n}(t)^2 = \lim_{n\to\infty} d_j(n)^2 = \infty,$$
$$\lim_{n\to\infty} g_{j;n}(n)^2/d_j(n)^2 = 0$$
$$\lim_{n\to\infty} \sum_{t=1}^{n} g_{j;n}(t) g_{k;n}(t+m)/(d_j(n)d_k(n)) = \rho_{j,k}(m),\ m \in \mathbb{N}_0.$$

Then $\rho_{j,k}(m) = \int_{-\pi}^{\pi} e^{im\lambda} dM_{j,k}(\lambda)$ for some $J \times J$ matrix M, cf. Hannan (1970, Ch. II, Th 11). We assume that M is nonsingular and $M^{-1} \int_{-\pi}^{\pi} f_Z(\lambda) M(d\lambda) M^{-1}$ is not the null matrix. Furthermore, we assume the existence of some limiting regressors $g_j(.) : \mathbb{R} \to \mathbb{R},\ (j = 1, \ldots, J)$, at least for one $j \in \{1, \ldots, J\}$ the function $g_j(.)$ not being the null function, such that for every $x \in \mathbb{R}$,

$$g_{j;n}(x) \to g_j(x)\ (n \to \infty),\ j = 1, \ldots, J.$$

The assumptions on the regressors in (R1) are close to the almost minimal Grenander conditions, we essentially ensure in addition that there is no degenerate behavior in the limit.

We present now the consistency of the sieve bootstrap at fixed points $x \in \mathbb{R}$.

Theorem 3.1 *Assume that (A1)-(A3), (R1) hold and the approximating order $p(n)$ in step 2 of the sieve bootstrap scheme satisfies $p(n) = o((n/log(n))^{1/4})$. Then, for $x \in \mathbb{R}$, there exists a sequence $\{a(n;x)\}_{n \in \mathbb{N}}$ such that*

$$a(n;x)(\hat{s}(x) - s(x)) \stackrel{d}{\longrightarrow} \mathcal{N}(0, \sigma^2_{as}(x)) \ (n \to \infty) \ \textit{with} \ 0 < \sigma^2_{as}(x) < \infty,$$

and

$$\sup_{y \in \mathbb{R}} |\mathbb{P}^*[a(n;x)(\hat{s}^*(x) - \hat{s}(x)) \leq y] - \mathbb{P}[a(n;x)(\hat{s}(x) - s(x))]| = o_P(1) \ (n \to \infty).$$

A proof is given in the appendix.

We extend now Theorem 3.1 for constructing simultaneous confidence regions. The most elegant way to analyze such simultaneous properties is to consider for the estimator in (3.1) the process

$$\{Z_n(x)\}_{x \in K}, Z_n(x) = a(n;x)(\hat{s}(x) - s(x)), \tag{3.2}$$

and its sieve bootstrapped counterpart

$$\{Z_n^*(x)\}_{x \in K}, Z_n^*(x) = a(n;x)(\hat{s}^*(x) - \hat{s}(x)), \tag{3.3}$$

where K is a compact subset in $\mathbb{R}$.
If the bootstrapped process $Z_n^*(.)$ has the same limiting process as $Z_n(.)$, the construction of bootstrap confidence bands $\{I_n(x)\}_{x \in K}$ in some uniform or simultaneous sense is consistent. To achieve this we describe now more precisely the framework. We denote by $\Rightarrow$ weak convergence in the space of continuous functions $\mathcal{C}(K) = \{g : K \to \mathbb{R};\ g \text{ continuous}\}$ with respect to the sup-norm $\|.\|_K$, i.e., $\|g\|_K = \sup_{x \in K} |g(x)|$. We make the following additional assumptions about the regressors.

(R2) The regressors $g_{j,n}(.)\ j = 1, \ldots, J$ satisfy a Lipschitz condition: there exists a $\gamma > 0$ such that

$$\sup_{n \in \mathbb{N}} \sup_{x,y \in K} \{|g_{j;n}(x) - g_{j;n}(y)|/|x-y|^\gamma\} \leq C_j < \infty, \ j = 1, \ldots J.$$

Lemma 3.1 *Assume that (A1)-(A2) and (R1)-(R2) hold. Then there exists a sequence $\{a(n;x)\}_{n \in \mathbb{N}, x \in K}$ such that for $Z_n(.)$ as defined in formula (3.2),*

$$Z_n \Rightarrow G,$$

where $G(.)$ is a Gaussian process with continuous sample paths, mean zero and
$Cov(G(x_1), G(x_2)) = \sum_{j_1,j_2=1}^J g_{j_1}(x_1) g_{j_2}(x_2) \sigma_{j_1,j_2}$, where
$\sigma_{j_1,j_2} = lim_{n \to \infty} d_{j_1}(n) d_{j_2}(n) Cov(\hat{\beta}_{j_1}, \hat{\beta}_{j_2})\ (0 \leq |\sigma_{j_1,j_2}| < \infty),\ d_j(n)^2 = \sum_{t=1}^n g_{j;n}(t)^2$.

A proof is given in the appendix.

Theorem 3.2 *Assume that (A1)-(A3), (R1)-(R2) hold and the approximating order $p(n)$ in step 2 of the sieve bootstrap scheme satisfies $p(n) = o((n/log(n))^{1/4})$. Then for the process $Z_n^*(.)$, as defined in formula (3.3) with the same normalizing constants $\{a(n;x)\}_{n \in \mathbb{N}, x \in K}$ as in Lemma 3.1,*

$$Z_n^* \Rightarrow G \ \textit{in probability},$$

where G is the same Gaussian process as in Lemma 3.1.

The proof is given in the appendix.

3.2 Kernel smoothers

Consider now the general model (2.1), where the trend values are (deterministically) sampled from a function $m : [0,1] \to \mathbb{R}$ such that $s(t) = m(t/n)$ $t = 1, \dots, n$. As in example 2.2, we are looking at kernel smoothers for the function $m(.)$, such as

$$\hat{m}(x) = \sum_{t=1}^{n} K(\frac{x - t/n}{h})Y_t / \sum_{s=1}^{n} K(\frac{x - s/n}{h}),\ x \in [0,1], \tag{3.4}$$

where $K(.)$ is a kernel function and h is a bandwidth with $h = h(n) = o(1), h(n)^{-1} = o(n)$ $(n \to \infty)$.

These estimates are known to exhibit edge-effects. They can cause problems in the sieve bootstrap scheme, namely in step 1 and step 4 where a pilot estimate $\tilde{s}(t) = \tilde{m}(t/n)$ has to be computed. We will use for $\tilde{m}(.)$ again a kernel smoother as in (3.4) but only in a region $[\delta, 1-\delta]$ for some small $\delta > 0$. Thus we will use in step 1 pilot estimates

$$\tilde{s}(t),\ t = [n\delta] + 1, \dots, [(1-\delta)n],\ 0 < \delta < 1, \tag{3.5}$$

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

By choosing the bandwidth of a mean square error optimal order, we have to deal with a non-negligible bias,

$$\begin{aligned} &n^{1/2}h^{1/2}(\mathbb{E}[\hat{m}(x)] - m(x)) \to B_{as}(x)\ (n \to \infty) \text{ with} \\ &0 < B_{as}(x) = \int_{-\infty}^{\infty} x^2 K(x) dx m^{(2)}(x)/2 < \infty, \end{aligned} \tag{3.6}$$

and under some regularity conditions, asymptotic normality holds,

$$\begin{aligned} &n^{1/2}h^{1/2}(\hat{m}(x) - m(x)) \xrightarrow{d} \mathcal{N}(B_{as}(x), \sigma_{as}^2)\ (n \to \infty) \text{ with} \\ &0 < \sigma_{as}^2 = f_Z(0) \int_{\infty}^{\infty} K^2(x) dx 2\pi < \infty, \end{aligned} \tag{3.7}$$

where $f_Z(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} Cov(Z_0, Z_k) e^{-i\lambda k}$ 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 step 1 and step 4 of our sieve bootstrap scheme an oversmoothed estimate $\tilde{s}(t) = \tilde{m}(t/n)$, $t = [n\delta] + 1, \dots [(1-\delta)n]$, with $\tilde{m}(.)$ being of the same form as in (3.4), with the same kernel $K(.)$, but with a larger bandwidth $\tilde{h}$ than h. Then it will be possible that

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

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

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

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

We make now the following assumptions.

(K) The function $m(.)$ is two times continuously differentiable on the open interval $(0,1)$ with $\sup_{0<x<1} |m^{(j)}(x)| < \infty$ for $j = 0, 1, 2$.
The estimator $\hat{m}(.)$ is as given in formula (3.4) with $K(.)$ a probability density, symmetric, bounded, piecewise continuous, compactly supported and the order of the bandwidth is $h(n) \sim const.n^{-1/5}$.
The estimator $\tilde{m}(.)$ is as given in formula (3.4) with the same kernel $K(.)$ as for $\hat{m}(x)$ but with bandwidth $\tilde{h}$ such that $\tilde{h}n^{1/5} \to \infty$.
Moreover, the sieve bootstrap scheme is modified as described in formula (3.5).

Theorem 3.3 *Assume that (A1)-(A3), (K) hold and the approximating order $p(n)$ in step 2 of the sieve bootstrap scheme satisfies $p(n) = o(\min\{(n/log(n))^{1/4}, n^{1/4}\tilde{h}^{1/4}\})$ with $\tilde{h}$ as in (K). Then for $0 < x < 1$,*

$$n^{1/2}h^{1/2}(\hat{m}(x) - m(x)) \xrightarrow{d} \mathcal{N}(B_{as}(x), \sigma_{as}^2) \ (n \to \infty) \ \textit{with} \ 0 < B_{as}(x), \sigma_{as}^2 < \infty,$$

and

$$\sup_{y \in \mathbf{R}} |\mathbb{P}^*[n^{1/2}h^{1/2}(\hat{m}^*(x) - \tilde{m}(x)) \le y] - \mathbb{P}[n^{1/2}h^{1/2}(\hat{m}(x) - m(x)) \le y]| = o_P(1) \ (n \to \infty).$$

A proof is given in the appendix.

To establish a similar result for uniform confidence bands as in Theorem 3.2 is a very difficult task. The processes

$$\{Z_n(x)\}_{x \in K}, \ Z_n(x) = n^{1/2}h^{1/2}(\hat{m}(x) - m(x)), \ K = [\delta, 1-\delta], \ 0 < \delta < 1/2,$$
$$\{Z_n^*(x)\}_{x \in K}, \ Z_n^*(x) = n^{1/2}h^{1/2}(\hat{m}^*(x) - \tilde{m}(x)), \ K = [\delta, 1-\delta], \ 0 < \delta < 1/2,$$

have now the property that neighboring values are asymptotically independent. One can renormalize and consider

$$Z_n^{renorm}(x) = n^{1/2}h^{1/2}(\hat{m}(hx) - m(hx)), \ x \in [h^{-1}\delta, h^{-1}(1-\delta)], \tag{3.8}$$

and analogously for the bootstrap.
A rigorous proof for weak convergence of such kind of processes in the context of density estimation for i.i.d. data can be found in Rosenblatt (1971), though under very restrictive assumptions. However, the arguments involve approximations by a Gaussian process which cannot be used directly for the bootstrap.

From a practical point of view we mention the work by Härdle and Marron (1991), which considers the bootstrap for kernel smoothers for independent observations to construct simultaneous confidence grids. They also do not present any theoretical result for processes as described above, but they show a simulation study which justifies to some extent the use of the bootstrap for simultaneous confidence grids.

3.2.1 Choice of tuning parameters for kernel smoothers

We address here the issue of selecting all the different tuning parameters. We do not give any exact theory here, this would be beyond the scope of the paper. All our proposals are derived in a heuristic sense. The sieve bootstrap procedure for kernel smoothers as in section 3.2 involves a pilot bandwidth $\tilde{h}$, the 'original' bandwidth h and also an approximating order p for the sieve bootstrap scheme.

We take the point of view that the practitioner comes up with a certain 'original' bandwidth h. It can be chosen automatically via a cross-validation procedure, cf. Altman (1990), Hart (1991, 1994), via an iterative plug-in approach, cf. Herrmann et al. (1992), or it can be chosen by eye or by some other reasons. Moreover, we keep our proposal for choosing the approximating order p via AIC as described in section 2.2, though the choice of the bandwidth $\tilde{h}$ for the pilot estimator $\tilde{m}$ could have a minor influence on the choice of the approximating order p. However, we believe that the procedure is not so sensitive to the choice of the order p which allows to choose p independent of other tuning parameters.

The main difficulty in this context is the choice of the pilot-bandwidth $\tilde{h}$. The choice of $\tilde{h}$ corresponds to the accuracy of estimation for the bias $\mathbb{E}[m(x)] - m(x)$, which is done here by $\mathbb{E}^*[\hat{m}^*(x)] - \tilde{m}(x)$. Härdle and Marron (1991) describe the problem very well in the independent set-up. Motivated by their Theorem 3, the rates for the optimal pilot-bandwidth $\tilde{h}$ and for the 'original' bandwidth h should be for second order kernels $K(.)$,

$$\tilde{h} \sim \tilde{C} n^{-1/9}, \quad h \sim C n^{-1/5} \quad \tilde{C} \text{ and } C \text{ constants.}$$

This gives a rough idea about the size of $\tilde{h}$ in comparison to the size of h, though the constants $\tilde{C}$ and C are unknown. Instead of trying to estimate these constants, which would complicate the procedure a lot, we suggest as a rough guide-line to take

$$\tilde{h} = h^{5/9}. \tag{3.9}$$

The problem of bias estimation in this context is a delicate issue, no satisfactory answers seem to be known. Härdle and Marron (1991) explain in greater details about this.

3.3 Higher order accuracy

Often, the bootstrap becomes even more powerful for constructing confidence regions when resampling a pivotal quantity, such as a studentized statistic, cf. Hall (1989, 1992). In our examples 2.1 and 2.2 a version of a studentized statistic is

for example 2.1:

$$\hat{t}_{student}(x) = a(n;x)(\sum_{j=1}^{J}(\hat{\beta}_j - \beta_j) g_{j;n}(x))/V_n, \ a(n;x) \text{ as in Theorem 3.1}, \ x \in \mathbb{R},$$

$$V_n^2 = \sum_{j_1,j_2=1}^{J} g_{j_1;n}(x) g_{j_2;n}(x) \hat{\sigma}_{j_1,j_2},$$

$$\hat{\sigma}_{i,j} = \left(M^{-1} 2\pi \int_{-\pi}^{\pi} \hat{f}_Z(\lambda) dM(\lambda) M^{-1}\right)_{i,j} / d_i(n) d_j(n),$$

where $M = [\rho_{i,j}(0)]_{i,j=1,\ldots,J}$ with $\rho_{.,.}(.)$ as defined by Grenander's condition in (R1).

for example 2.2:

$$\hat{t}_{student}(x) = n^{1/2} h^{1/2} (\hat{m}(x) - m(x))/V_n, \ x \in [\delta, 1-\delta], \ \delta > 0,$$

$$V_n^2 = \hat{f}_Z(0) m(x)^2 \int_{-\infty}^{\infty} K^2(x) dx 2\pi, \text{ see (3.7).}$$

In both cases we have to construct an estimate $\hat{f}_Z$ of f_Z, the spectral density of the underlying noise process $\{Z_t\}_{t\in\mathbb{Z}}$. In our case we have a 'natural estimate', based on the autoregressive sieve. Step 1 and step 2 of our sieve bootstrap scheme yield coefficient estimates $\hat{\phi}_{1,n},\ldots\hat{\phi}_{p,n}$ and estimates of the innovation variance based on the centered residuals $\{\tilde{\varepsilon}_{t,n}\}_{t=p+1}^{n}$. By using this set of estimates we propose the autoregressive spectral estimator

$$\hat{f}_Z(\lambda) = \frac{(n-p)^{-1}\sum_{t=p+1}^{n}\tilde{\varepsilon}_{t,n}^2}{2\pi|\hat{\Phi}_n(e^{-i\lambda})|^2},\quad \hat{\Phi}_n(z) = \sum_{j=0}^{p}\hat{\phi}_{j,n}z^j\ (z\in\mathbb{C}).$$

As one can see, the analytical approximation for V_n^2 in example 2.1 can become very complex. However, in some cases like trigonometric or polynomial regression the formula simplifies quite a bit. For sieve bootstrap versions we essentially use the plug-in principle,

$$\hat{t}^*_{student}(x) = a(n;x)(\hat{s}^*(x)-\mathbb{E}^*[\hat{s}^*(x)])/V_n^* \text{ for example 2.1,}$$
$$\hat{t}^*_{student}(x) = n^{1/2}h^{1/2}(\hat{m}^*(x)-\tilde{m}(x))/V_n^* \text{ for example 2.2.}$$

Based on such studentized estimators one would then construct (pointwise) confidence intervals

$$I_n(x;1-\alpha) = [\hat{s}(x)-\hat{q}_{1-\alpha/2}V_n, \hat{s}(x)-\hat{q}_{\alpha/2}V_n],$$

where $\hat{q}_\alpha = \inf\{y;\mathbb{P}^*[\hat{t}^*_{student}(x)\le y]\ge\alpha\}$.
Such confidence intervals lead in the independent case to better coverage properties in theory, cf. Hall (1989, 1992). We leave a theoretical investigation of higher order accuracy for the sieve bootstrap as an open research problem, but we include this method in our simulation study in section 5.

4 Classical approaches

We describe here some classical methods which could be applied to construct pointwise or simultaneous confidence bands. All these classical approaches rely on asymptotic arguments and we do not aim to give an exact justification for them, we rather see them as methods which have been used in practice as 'ad hoc' procedures.

4.1 Asymptotic normality for pointwise bands

To construct pointwise confidence bands one can use the normal approximation, this is justified by our Theorems 3.1 and 3.3. In particular, the confidence bands for a fixed x would be

for linear regression:

$$\hat{s}(x)\pm\left(\hat{Var}(\hat{s}(x))\right)^{1/2}z_{1-\alpha/2},\ z_{.}\text{ the standard normal quantile,}$$

for kernel smoothing:

$$\hat{m}(x)\pm\left(\hat{Var}(\hat{m}(x))\right)^{1/2}z_{1-\alpha/2},\ z_{.}\text{ the standard normal quantile,}$$

or with a bias correction

$$\hat{m}(x) - \hat{B}_{as}(x) \pm \left(\hat{Var}(\hat{m}(x))\right)^{1/2} z_{1-\alpha/2},\ z_{.} \text{ the standard normal quantile,}$$

where $\hat{B}_{as}(x) = \hat{m}^{(2)}(x) \int_{-\infty}^{\infty} x^2 K(x)dx/2$ is a bias estimate, based on a kernel estimate $\hat{m}^{(2)}(x)$ for estimating the second derivative $m^{(2)}(x)$. One proposal is

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

with $K_2(x) = 105/32(-5x^4 + 6x^2 - 1)1_{-1 \le x \le 1}$, see Gasser et al. (1985).

The estimates for the variances of the predictors can be obtained in the same way as in section 3.3.

4.2 Scheffé's method for simultaneous bands in linear regression

For constructing simultaneous confidence bands in the linear regression case as in (3.1), one can use Scheffé's approach, cf. Seber (1977, Ch. 5.1). Relying on the asymptotic normality of $\hat{\beta} - \beta$ as given in Theorem 3.1, the simultaneous confidence bands are then

$$\hat{s}(x) \pm \left(JF_{J,n-J;1-\alpha}\hat{Var}(\hat{s}(x))\right)^{1/2},\ x \in \mathbb{R},$$

where $\hat{Var}(\hat{s}(x))$ can be obtained as in section 3.3.
By construction, these confidence bands are conservative because one is usually not interested in simultaneity over the whole range of the design. Moreover, the asymptotic arguments are for quadratic forms and therefore more critical.

4.3 Poisson clumping heuristics for simultaneous bands via smoothing

For the case of kernel smoothing as in (3.4), one can make use of the very elegant and beautiful method of Poisson clumping heuristics, given by Aldous (1989). This approach is not widely known in statistics, therefore we give a few more details.

Our starting point is the renormalized process $Z_n^{renorm}(.)$ given in (3.8). Then one easily verifies that

$$Var(Z_n^{renorm}(x)) = f_Z(0)\int_{-\infty}^{\infty} K^2(x)dx2\pi\ (0 \le x \le h^{-1}),$$

$$Cov(Z_n^{renorm}(x), Z_n^{renorm}(x+y)) = f_Z(0)\int_{-\infty}^{\infty} K(x)K(x+y)dx2\pi\ (0 \le x, x+y \le h^{-1}),$$

compare also with (3.7).
Hence the self-standardized process

$$W_n(.) = Z_n^{renorm}(.)(f_Z(0)\int_{-\infty}^{\infty} K^2(x)dx2\pi)^{-1/2}$$

converges to a Gaussian process W with mean zero and variance one. Moreover, this limiting Gaussian process is stationary since its covariances are

$$Cov(W(x), W(y)) = \int_{-\infty}^{\infty} K(x)K(x+y)dx / \int_{-\infty}^{\infty} K^2(x)dx =: R_W(|x-y|).$$

Therefore Aldous' heuristics for smooth Gaussian processes applies, namely

$$R_W(y) \sim 1 + \frac{1}{2}\frac{\int_{-\infty}^{\infty} K(x)K^{(2)}(x)dx}{\int_{-\infty}^{\infty} K^2(x)dx} y^2 \text{ as } y \to 0,$$

yielding

$$\begin{aligned}
&\mathbb{P}[\max_{0 \le y \le h^{-1}} W_n(y) \le b] \approx exp(-\lambda_b h^{-1}),\\
&\lambda_b = (-\int_{-\infty}^{\infty} K(x)K^{(2)}(x)dx / \int_{-\infty}^{\infty} K^2(x)dx)^{1/2}(2\pi)^{-1} exp(-b^2/2),
\end{aligned}$$

see Aldous (1989, (C23d)). If we take into account that $h^{-1} \to \infty$, this is actually the same result as given in a long argument by Rosenblatt (1971, (62)) for density estimation. By symmetry one then can use the following simultaneous confidence band

$$\begin{aligned}
&\hat{m}(x) \pm n^{-1/2}h^{-1/2}\hat{f}_Z(0)\int_{\infty}^{\infty} K^2(x)dx 2\pi q_{1-\alpha/2},\\
&q_{1-\alpha/2} = (-2log(h) + 2log(C) - 2loglog(1-\alpha/2))^{1/2},\\
&C = (-\int_{-\infty}^{\infty} K(x)K^{(2)}(x)dx / \int_{-\infty}^{\infty} K^2(x)dx)^{1/2}(2\pi)^{-1}).
\end{aligned}$$

By Taylor expansions of $\sqrt{x}$ for large x and of $log(1-x)$ for small x, this is also the same result as in Eubank and Speckman (1993, (16a)-(16c)). Of course, we could also here apply a bias correction and use the band for $0 < x < 1$,

$$\hat{m}(x) - \hat{B}_{as}(x) \pm n^{-1/2}h^{-1/2}\left(\hat{f}_Z(0)\int_{\infty}^{\infty} K^2(x)dx 2\pi\right)^{1/2} q_{1-\alpha/2},$$

where the bias estimate $\hat{B}_{as}(x)$ can be derived as in section 4.1.
The estimate $\hat{f}_Z(0)$ can be obtained as in section 3.3.

4.4 A crucial step from independence to dependence

All of the methods described in sections 4.1-4.3 involve an estimate for the spectral density of the noise process. In our case, where the noise process is assumed to be an AR(∞) process, a natural way for estimation is given by the autoregressive spectral estimator, as given in section 3.3. But a crucial step has taken place here.

In the independent set-up, Eubank and Speckman (1993) use extreme value approximations for constructing simultaneous confidence bands via kernel smoothing and obtain surprisingly good results in a simulation study for finite sample sizes. Why? One needs an estimate of the asymptotic variance of the kernel smoother. In the independent case this is quite easy to achieve because one needs only to estimate the variance of an observation. This can be done in a direct way by a $\sqrt{n}$-consistent estimator, e.g.,

$$\hat{Var}(Y_t) = (n-2)^{-1}\sum_{i=1}^{n-2}(0.809Y_i - 0.5Y_{i+1} - 0.309Y_{i+2})^2,$$

as used by Eubank and Speckman (1993). Such estimators have the great advantage that they operate directly on the observed data rather than on estimated residuals. In the time series case, the relevant quantity to know is not the variance of an observation but the spectral density of the noise process. For estimating this quantity, the $\sqrt{n}$-consistency is lost and we see no other obvious way than estimating such a spectral density via estimated residuals. But it seems exactly here, where a lot of inaccuracy comes in, see also in section 5. This then explains that the methodological difference between the time series and the independent set-up is substantial. The lack of a good and simple method for variance estimation of a trend estimator in time series creates a new additional problem. The sieve bootstrap is successful to handle this problem in an implicit way, which makes explicit variance estimation unnecessary.

5 Numerical examples

We study our procedure on some simulated examples. We consider here the accuracy of the sieve bootstrap approximation in terms of coverage probabilities. We focus on the situation in example 2.1 and example 2.2 respectively and consider the two trend models

(T1) $s(x) = 3g(x),\ g(x) = cos(\pi x/256),\ x \in \mathbb{R}$,

(T2) $s(x) = m(x/n)$, $x \in [0, n]$ (n the sample size),
$m(y) = 2 - 5y + 5exp(-100(y - 0.5)^2),\ y \in [0, 1]$.

The trend model (T2) is the same as in Herrmann et al. (1992). The trigonometric trend in (T1) is scaled so that the 'variability' of (T1) and (T2) is approximately the same for $n = 512$, i.e., the quantities $(n-1)^{-1}\sum_{t=1}^{n}(s(t) - \bar{s})^2$ ($\bar{s} = n^{-1}\sum_{t=1}^{n} s(t)$) are approximately the same for both trend models. To both trend models we add the same ARMA(1,1) noise $Z_1, \ldots, Z_n$ (also the same realization of noise for both trend models),

$$Z_t = 0.8Z_{t-1} - 0.5\varepsilon_{t-1} + \varepsilon_t, \tag{5.1}$$

where ε_t i.i.d. $\sim t_6/\sqrt{1.8}$, i.e., $\sqrt{1.8}\varepsilon_t$ i.i.d $\sim t_6$.
This noise process has been studied in connection with the bootstrap in Bühlmann (1995b). It is scaled so that $Var(Z_t) \approx 1$. We generate data sets of sample size $n = 512$, according to

$$Y_t = s(t) + Z_t,\ t = 1, \ldots, n,$$

where $s(.)$ is as in (T1) or (T2) and $Z_1, \ldots Z_n$ follow the model in (5.1). Each trend model (T1) and (T2) with the same realization of noise is shown in figure 5.1. For the trend model (T1) and (T2) we use methods (M1) and (M2) respectively,

(M1) Least squares predictor $\hat{s}(.)$ as in (3.1) for the trigonometric regression model,

(M2) Kernel smoothing as in (3.4) with the Parzen-kernel (convolution of the triangular kernel on $[-0.5, 0.5]$) and $\tilde{h} = 0.19,\ h = 0.05$. (The bandwidths are standardized: the upper and lower quartiles of the Parzen-kernel are ± 0.25 when bandwidth is 1, this is the standardization in S-Plus).

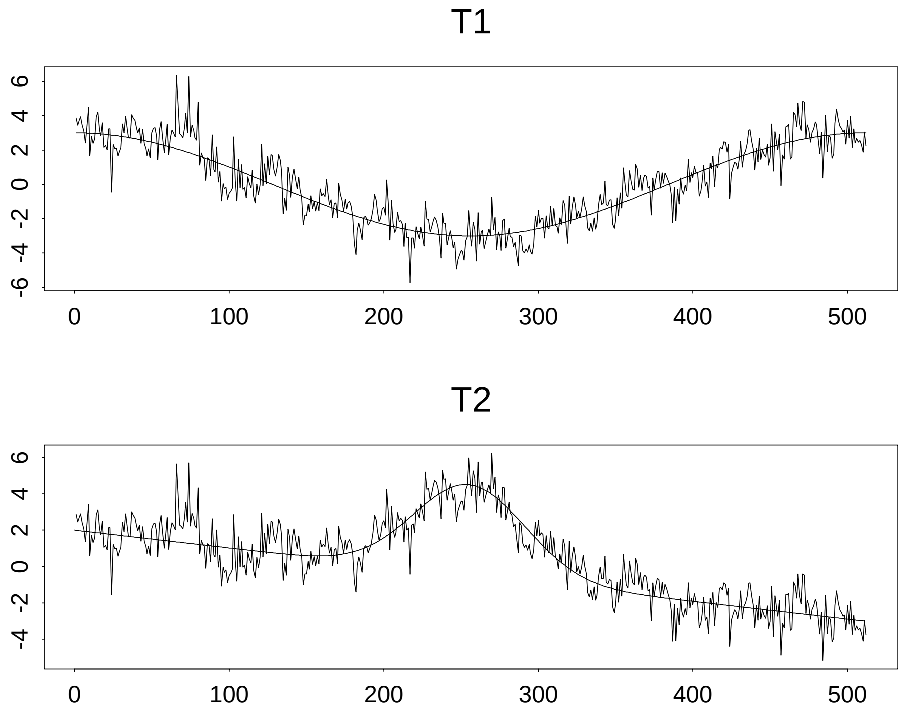

Figure 5.1: Models (T1) and (T2) overlaid with the same realization of ARMA(1,1) noise

We will consider the coverage for pointwise intervals at $x = 250$ and for simultaneous intervals over the grids

$$x_1, \dots, x_{50},\ x_i = 10i,\ i = 1, 2, \dots, 50 \text{ in case (T1,M1)},$$
$$x_1, \dots, x_{32},\ x_i = 10i,\ i = 10, 11, \dots, 41 \text{ in case (T2,M2)}.$$

Based on the sieve bootstrap, the confidence bands are constructed as described in section 2.3: the pointwise as in step (i), the simultaneous intervals as in steps (i)-(ii), both with $\alpha = 0.1$. The number of bootstrap replicates is 500. We construct also some intervals based on a studentized sieve bootstrap procedure as described in section 3.3.

For comparison, we consider in the case of pointwise intervals the method based on asymptotic normality described in section 4.1. The variances of the predictors are then

$$Var(\hat{s}(x)) \sim 4\pi cos(\pi x/256)^2 f_Z(\pi/256) n^{-1} \text{ in case (T1,M1)}$$
$$Var(\hat{m}(x)) \sim 2\pi \int_{-\infty}^{\infty} K^2(x)dx f_Z(0) n^{-1} h^{-1} = 1.53\pi f_Z(0) n^{-1} h^{-1} \text{ in case (T2,M2)},$$

where $f_Z(.)$ is the spectral density of the noise process as in (5.1).
An estimate of the spectral density $f_Z(.)$ can be obtained as in section 3.3. In the case (T2,M2) we used the original bandwidth $h = 0.05$ for constructing such an estimate; using the pilot bandwidth $\tilde{h}$ resulted in a poorer result.
The bias correction was made as proposed in section 4.1 with the same bandwidth $\tilde{h} = 0.19$ as in (M2). We did not achieve any better results by using other bandwidths for the bias correction.

The following table shows relative coverage frequencies with estimated standard deviations, which were derived by simulating over 100 different model realizations, i.e., realizations of the noise process in (5.1).

pointwise coverage	(T1,M1)	(T2,M2)
sieve bootstrap	0.88 (0.033)	0.95 (0.022)
studentized sieve bootstrap	0.90 (0.030)	0.96 (0.020)
as. normality without bias corr.	0.89 (0.031)	0.50 (0.050)
as. normality with bias corr.	–	0.51 (0.050)

For pointwise confidence intervals, the sieve bootstrap procedure yields a very satisfactory result for both trends and both methods. In the case (T1,M1) the asymptotic normality is a good competitor, but in the case (T2,M2) the sieve bootstrap outperforms both normal approximations, with and without bias correction. The bias correction with estimating the second derivative $m^{(2)}(x)$ directly did not enhance performance in this case, where the bias is already small, see also figure 5.2.

The sieve bootstrap for studentized statistics as given in section 3.3 seems to work but does not yield any significant gains in our examples.

The next table shows the empirical simultaneous coverage for sieve bootstrap simultaneous confidence grids as described above. For comparison we give the coverages of the following alternatives: pointwise confidence based on asymptotic normality as described in section 4.1 and based on the extreme value approximation via the Poisson clumping heuristics as described in section 4.3. Note that in case (T1,M1) the pointwise confidence intervals based on asymptotic normality give an asymptotically correct answer which is equivalent to Scheffé's method described in section 4.2, since

$$\max_x |(\hat{\beta} - \beta) cos(\pi/256x)| = |\hat{\beta} - \beta|, \tag{5.2}$$

which can be analyzed with a limiting normal distribution.
In the case (T2,M2) the pointwise approach based on asymptotic normality is conceptually wrong, nevertheless one might wonder about the price of doing such a wrong thing.

simultaneous coverage	(T1,M1)	(T2,M2)
sieve bootstrap	0.89 (0.031)	0.99 (0.001)
pointwise as. normality without bias corr.	0.79 (0.041)	0.00 (–)
pointwise as. normality with bias corr.	–	0.00 (–)
Poisson clumping without bias corr.	–	0.02 (0.014)
Poisson clumping with bias corr.	–	0.02 (0.014)

For simultaneous confidence intervals, the sieve bootstrap yields again an approximately correct coverage probability. The pointwise normal approximation in case (T1,M1), which is also a simultaneous approximation (see (5.2)), exhibits a drop of 12% of the required coverage level 0.9. Knowing that the pointwise normal approximation is conceptually wrong for simultaneous intervals in case (T2,M2), the pointwise normal approximation breaks down completely. The extreme value approximation yields also a very poor result, for an explanation see also section 4.4. The bias correction in all the analytic approximations does not yield any improvement. An explanation is given by figure 5.2, where one can easily see that the bias and its estimation are of a much smaller order than the

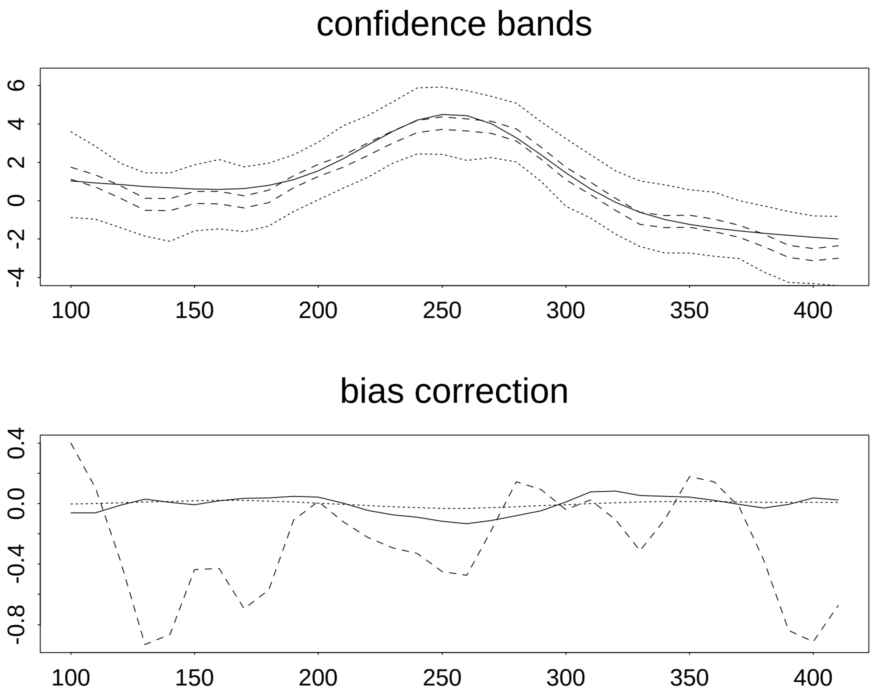

Figure 5.2: Confidence bands: true trend (line), sieve bootstrap simultaneous confidence band (small dashes), simultaneous band based on Poisson clumping heuristics, without bias correction (large dashes). Bias correction: estimated bias over 100 simulations (line), bias correction $\hat{B}_{as}(x)$ (small dashes), $\hat{m}(x) - m(x)$ (large dashes). Both based on a typical realization of (M2).

difference between the original estimator and the truth. Our original bandwidth $h = 0.05$, chosen bye eye, seems to undersmooth so that the variance part is dominating.

The big advantage of the sieve bootstrap is demonstrated here for the construction of simultaneous confidence grids. The method based on extreme value approximation exhibits a serious lack of accuracy, figure 5.2 illustrates in an impressive way that the bands from extreme value approximation are much too small and yield to a completely invalid confidence statement.

We also tried for the sieve bootstrap in case (T2,M2) some other pilot bandwidths $\tilde{h}$, resulting in a poorer performance. Our suggestion is to take the rule as in (3.8).

Finally we want to reflect the drastic effect of having no accurate variance estimator, see section 4.4. When plugging in the true spectral density at zero for the noise process we found an asymptotic variance $\sigma_{as}^2 = 0.1520$ for the kernel smoother in (M2) , whereas the mean over the 100 simulations of the estimated variances was $ave[\hat{\sigma}_{as}^2] = 0.0260$, which is smaller by about a factor 6. By using the correct σ_{as}^2 (which in practice would be unknown) we achieved in the case (T2,M2) a simultaneous coverage of 0.94, based on Poisson clumping heuristics (with and without bias correction).

6 Summary and discussion

We have given a method for constructing confidence bands for a deterministic trend in time series. We focus on the case where the noise process is stationary and restricted to be in a large subclass of linear processes, examples would be many of the well known ARMA processes. Our method relies on a sieve bootstrap scheme from stationary time series which is fairly simple to apply and computationally feasible. We propose a bootstrap technology because we believe that it comes close to the aim for constructing a simultaneous confidence band over some arbitrary set of points, which can be specified by the user. In this sense, such intervals do not suffer from the conservative approach of a Scheffé-type band, see section 4.2. Moreover, the theory does not rely on extreme value approximation as discussed in section 4.3.

We have also argued why known methods from the independent case cannot be directly adapted to the time series case. The extreme value approximation, as given in Eubank and Speckman (1993) performs very poor, when adapted to our example.

The theory for the sieve bootstrap in the case of time series with deterministic trends works almost under the same conditions as in the case for stationary time series, this justifies the method as a good procedure from a theoretical point of view. From a practical point of view we report satisfactory finite sample results from our simulation study with a 'middle sample size' $n = 512$. In particular, in a time series set-up such sample sizes are not considered as large.

Though the tuning parameters in the procedure were selected by heuristical rules, our method performed well in the simulation study. There is hope that this sieve bootstrap procedure is not very sensitive when varying the different tuning parameters.

Open problems in this area are refined choices of tuning parameters, see section 3.2.1, and higher order accuracy as discussed in section 3.3 which would rule out classical analytical approximations from a theoretical point of view. There is also a considerable interest in using wavelet methods for estimating a trend function in time series. Brillinger (1995) proves asymptotic normality of linear wavelet estimators in the time series set-up. Our sieve bootstrap yields again a consistent procedure for approximating the distribution of such estimators. The validity of the sieve bootstrap for nonlinear wavelet estimators which are based on shrinking coefficients, see for example Donoho et al. (1995), is not straightforward. A quite different problem would be the extension from linear to nonlinear noise processes. Block bootstrap techniques, see Künsch (1989), might be used, by paying a higher price for the 'nonparametric' character of such a general nonlinear stochastic error process, see also Bühlmann (1995b).

Appendix

We first consider the effect of non-stationarity, i.e., the effect of estimating the trend values $s(t)$ $(t = 1, \ldots, n)$ in step 1 and 4 of the sieve bootstrap scheme. We do not treat the case as given in formula (3.5) separately, the modifications are obvious. For $|j| \leq n-1$, we denote by $R_Z(j) = Cov(Z_0, Z_j)$, $\hat{R}_Z(j) = n^{-1} \sum_{t=1}^{n-|j|} Z_t Z_{t+|j|}$, and for $\hat{R}_{\hat{Z}}(j)$ see (2.3).

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

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

Remark: If $b(n)^{1/2} = O(\max\{(log(n)/n)^{1/2}, \sum_{j=p+1}^{\infty}|\phi_j|\})$, then Lemma A.1 yields the same bound as in the stationary case for $\max_{0\le j\le p}|\hat{R}_Z(j) - R_Z(j)|$, cf. Hannan and Kavalieris (1986, Th. 2.1).

Proof: We write for $|j| \le n-1$,

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

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

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

Now we complete the proof by using the known bound

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

cf. Hannan and Kavalieris (1986, Th. 2.1). □

Denote by ϕ_j the autoregressive coefficients as given in (2.2), by $\hat{\phi}_{j,n}$ the coefficients as defined in (2.3) and by $\phi_{j,n}$ the corresponding theoretical quantities with R_Z instead of $\hat{R}_Z$, i.e., $,_Z\phi_p = -\gamma_Z$, $\phi_p = (\phi_{1,n}, \ldots, \phi_{p,n})'$.

Lemma A.2 *Assume that (A1) with $s = 4$, (A2) with $r = 1$,(A3) hold, $n^{-1}\sum_{t=1}^{n}(\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$ $(n \to \infty)$ and $p(n) = o(min\{(n/log(n))^{1/2}, b(n)^{-1/2}\})$. Then*

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

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

Proof: Denote by $\hat{\phi}_p = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n})'$, $\phi_p = (\phi_{1,n}, \ldots, \phi_{p,n})'$. According to An et al. (1982, formula (25)),

$$\hat{\phi}_p - \phi_p = -,_Z^{-1}[(\hat{\gamma}_{\hat{Z}} - \gamma_Z) + (\hat{,}_{\hat{Z}} - ,_Z)(\hat{\phi}_p - \phi_p) + (\hat{,}_{\hat{Z}} - ,_Z)\phi_p].$$

By denoting

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

we arrive at

$$\|(\hat{\Gamma}_{\hat{Z}} - \Gamma_Z)(\hat{\phi}_p - \phi_p)\|_\infty \leq p \max_{0\leq j\leq p} |\hat{R}_{\hat{Z}}(j) - R_Z(j)| \|\hat{\phi}_p - \phi_p\|_\infty,$$

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

Therefore

$$\|\hat{\phi}_p - \phi_p\|_\infty \leq \|\Gamma_Z^{-1}\|_\infty \max_{0\leq j\leq p} |\hat{R}_{\hat{Z}}(j) - R_Z(j)|(1 + p\|\hat{\phi}_p - \phi_p\|_\infty + \sum_{j=1}^{p} |\phi_{j,n}|),$$

and hence

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

By the assumption and Lemma A.1, $p \max_{0\leq j\leq p} |\hat{R}_{\hat{Z}}(j) - R_Z(j)| = o_P(1)$; by An et al. (1982, p. 929, l-4), $\|\Gamma_Z^{-1}\|_\infty < \infty$; by Baxter's inequality, $\sum_{j=1}^{p} |\phi_{j,n}| \leq \sum_{j=1}^{p} |\phi_j| + const. \sum_{j=p+1}^{\infty} |\phi_j| = O(1)$. This, together with the bound for $\|\hat{\phi}_p - \phi_p\|_\infty(1 - \|\Gamma_Z^{-1}\|_\infty p \max_{0\leq j\leq p} |\hat{R}_{\hat{Z}}(j) - R_Z(j)|)$ above and Lemma A.1 completes the proof. □

It is very helpful to represent

$$Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},\ \psi_0 = 1 \tag{A.1}$$

$$Z_t^* = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}^*,\ \hat{\psi}_{0,n} = 1, \tag{A.2}$$

where $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = 1/\Phi(z)$, $\Phi(.)$ 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 A.3 *Assume that (A1) with $s = 4$, (A2) with $r = 1$, (A3) hold, $n^{-1}\sum_{t=1}^{n}(\tilde{s}(t) - s(t))^2 = O_P(b(n))$ for some sequence $b(n) = o(1)$ $(n \to \infty)$ and $p(n) = o(min\{(n/log(n))^{1/4}, b(n)^{-1/4}\})$. Then the following holds:*

(i) there exists a random variable n_1 such that

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

(ii)

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

Proof: Both statements follow by using Lemma A.2 and then proceeding as in the proofs of Theorem 3.1 and 3.2 in Bühlmann (1995a). □

Lemma A.4 *Assume that (A1) with* $s = \max\{2w, 4\}$, $w \in \mathbb{N}$, *(A2) with* $r = 1$, *(A3) hold,* $n^{-1}\sum_{t=1}^{n}(\tilde{s}(t) - s(t))^2 = O_P(b(n))$ *for some sequence* $b(n) = o(1)$ $(n \to \infty)$ *and* $p(n) = o(min\{(n/log(n))^{1/2}, b(n)^{-1/2}\})$. *Then*

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

Proof: We write

$$\varepsilon_{t,n} = \varepsilon_t + Q_{t,n} + R_{t,n} + U_{t,n} + V_{t,n},$$

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 (1995b, Lem.5.3). The only additional quantities to control are $U_{t,n}$, $V_{t,n}$: by the Cauchy-Schwarz inequality and Lemma A.2,

$$(n-p)^{-1}\sum_{t=p+1}^{n}|U_{t,n}|^{2w} \le [p^2(\|\phi_p\|_\infty + \|\hat{\phi}_p - \phi_p\|_\infty)^2 O_P(b(n))]^w = o_P(1),$$
$$(n-p)^{-1}\sum_{t=p+1}^{n}|V_{t,n}|^{2w} = O_P((\sum_{j=p+1}^{\infty}|\phi_j|)^{2w}) = o_P(1).$$

With these bounds we complete the proof. □

Proof of Theorem 3.1. The first statement follows directly from Theorem 10, Ch. IV.4 in Hannan (1971), our conditions imply the conditions for this theorem.
By (A.2) we know that the bootstrap noise $\{Z_t^*\}_{t\in\mathbb{Z}}$ is a linear process and by Lemma A.3 and A.4 we are able to control its coefficients and its innovations. The main key is to write

$$a(n;x)(\hat{s}(x) - s(x)) = \sum_{t=1}^{n} w_{t;n}(x)Z_t,$$
$$a(n;x)(\hat{s}^*(x) - \hat{s}(x)) = \sum_{t=1}^{n} w_{t;n}(x)Z_t^*. \tag{A.3}$$

The weights $w_{t;n}(x)$ are such that

$$\sup_{n\in\mathbb{N}}\sum_{t=1}^{n}|w_{t;n}(x)| < \infty. \tag{A.4}$$

To see this, write

$$a(n;x)(\hat{s}(x) - s(x)) = \sum_{j=1}^{J} d_j(n)(\hat{\beta}_j - \beta_j)g_{j;n}(x) = \sum_{j=1}^{J}((D_n^{-1}G'GD_n^{-1})^{-1}D_n^{-1}G'\mathbf{Z}_n)_j g_{j;n}(x),$$

where $D_n = diag(d_1(n), \ldots, d_J(n))$, G the $n \times J$ design matrix with element $G_{t,j} = g_{j;n}(t)$ and $\mathbf{Z}_n = (Z_1, \ldots, Z_n)'$.
By assumption (R1),

$$(D_n^{-1}G'GD_n^{-1})^{-1} \to ([\rho_{j,k}(0)]_{j,k=1}^{J})^{-1} \ (n \to \infty).$$

Therefore, by changing the order of summation and by (R2), (A.4) follows.
By (A.3)-(A.4) and Lemma A.3 and A.4 we get

$$Var^*(a(n;x)\hat{s}^*(x)) - Var(a(n;x)\hat{s}(x)) = o_P(1).$$

Now, again by using Lemma A.3 and A.4, we proceed as in the proof of Theorem 10, Ch. IV.4 in Hannan (1971). This completes the proof. □

Proof of Lemma 3.1. As in Theorem 3.1 we get asymptotic normality for a finite collection, i.e., for $Z_n(x_1), \ldots, Z_n(x_d)$, where $x_1, \ldots x_d \in K$, $d \in \mathbb{N}$. What remains to show is stochastic equicontinuity, i.e., for all $\kappa > 0$, for all $\eta > 0$ exists $\delta > 0$ and exists $n_0 = n_0(\eta)$ such that

$$\mathbb{P}[\sup_{x,y \in K, |x-y| \leq \delta} |Z_n(x) - Z_n(y)| > \kappa] < \eta \text{ for all } n \geq n_0, \tag{A.5}$$

cf.Billingsley (1968, Th. 8.2).
But by the Lipschitz condition for the regressors (R2),

$$\sup_{x,y \in K, |x-y| \leq \delta} |Z_n(x) - Z_n(y)| \leq \sum_{j=1}^{J} d_j(n) |\hat{\beta}_j - \beta_j| C_j \delta^{\gamma},$$

which implies (A.5) since $d_j(n)(\hat{\beta}_j - \beta_j) = O_P(1)$, $j = 1, \ldots, J$. □

Proof of Theorem 3.2. The proof is analogous to the proof of Lemma 3.1 above. □

Proof of Theorem 3.3. In Hall and Hart (1990) it is shown that for $0 < \delta < 1/2$,

$$((1-2\delta)n)^{-1} \sum_{t=[n\delta]+1}^{[(1-\delta)n]} (\tilde{s}(t) - s(t))^2 = ((1-2\delta)n)^{-1} \sum_{t=[n\delta]+1}^{[(1-\delta)n]} (\tilde{m}(t/n) - m(t/n))^2 = O_P(n^{-1}\tilde{h}^1).$$

This explains that for $p(n) = o(\min\{(n/log(n))^{1/4}, n^{1/4}\tilde{h}^{1/4}\})$ Lemma A.2 - A.4 are valid. It is also known, cf. Hall and Hart (1990), that

$$B_{as}(x) = \lim_{n\to\infty} n^{1/2} h^{1/2} h^2 m^{(2)}(x) \int_{-\infty}^{\infty} x^2 K(x) dx/2, \; 0 < x < 1. \tag{A.6}$$

Moreover

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

where $R_Z(k) = Cov(Z_0, Z_k)$.
Now we write

$$n^{1/2} h^{1/2} (\hat{m}(x) - \mathbb{E}[\hat{m}(x)]) = n^{1/2} h^{1/2} \sum_{t=1}^{n} w_{t;n}(x) Z_t, \text{ with } \sum_{t=1}^{n} w_{t;n}(x) = 1,$$

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

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 (1995b). The idea is to replace Z_t^* by $Z_{t,M}^* = \sum_{j=0}^{M} \hat{\psi}_{j,n}\varepsilon_{t-j}^*$, compare with (A.2). Using the M-dependence of $\{Z_{t,M}^*\}_{t\in\mathbb{Z}}$ we prove convergence in d^*-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 Lemma A.3 and A.4. Summarizing, with the same arguments as in the proof of Theorem 3.1 in Bühlmann (1995b) we get

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

What remains to show is

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

But by (A.6) and the same argument for the bootstrap,

$$n^{1/2}h^{1/2}(\mathbb{E}^*[\hat{m}^*(x)] - \tilde{m}(x)) - B_{as}(x) \sim n^{1/2}h^{1/2}h^2 C_K(\tilde{m}^{(2)}(x) - m^{(2)}(x)),$$

hence it suffices to show

$$\tilde{m}^{(2)}(x) - m^{(2)}(x) = o_P(1).$$

But this holds true because $\tilde{h}n^{1/5} \to \infty$, cf. Gasser and Müller (1984) (our assumptions are stronger than their conditions). □

Acknowledgments: I would like to thank David Brillinger for many helpful comments and discussions.

References

[1] Aldous, D. (1989). Probability Approximations via the Poisson Clumping Heuristics. Appl. Math. Science **77**. Springer, New York.

[2] Altman, N.S. (1990). Kernel smoothing of data with correlated errors. J. Amer. Statist. Assoc. **85** 749-759.

[3] Bickel, P.J. and Bühlmann, P. (1995). Mixing property and functional central limit theorems for a sieve bootstrap in time series. Tech. Rep. 440, Dept. of Stat., U.C. Berkeley.

[4] Bickel, P.J. and Freedman, D.A. (1983). Bootstrapping regression models with many parameters. Festschrift for Erich L. Lehmann 28-48.

[5] Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

[6] Brillinger, D.R. (1995). Uses of cumulants in wavelet analysis. To appear in J. Nonparam. Statist.

[7] Bühlmann, P. (1995a). Moving-average representation for autoregressive approximations. To appear in Stoch. Proc. and Appl.

[8] Bühlmann, P. (1995b). Sieve bootstrap for time series. Tech. Rep. 431, Dept. of Stat., U.C. Berkeley.

[9] Dahlhaus, R. (1983). Spectral analysis with tapered data. J. Time Series Anal. **4** 163-175.

[10] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (With discussion). J. R. Statist. Soc. B **57** 301-337.

[11] Efron, B. (1979). Bootstrap methods: another look at the jackknife. Ann. Statist. **7** 1-26.

[12] Eubank, R.L. and Speckman, P.L. (1993). Confidence bands in nonparametric regression. J. Amer. Statist. Assoc. **88** 1287-1301.

[13] Freedman, D.A. (1981). Bootstrapping regression models. Ann. Statist. **9** 1218-1228.

[14] Freedman, D.A. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist. **12** 827-842.

[15] Gasser, Th. and Müller, H.-G. (1979). Kernel estimation of regression functions. In Smoothing Techniques for Curve Estimation (ed. Th. Gasser and M. Rosenblatt). Springer, Heidelberg.

[16] Gasser, Th. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. Scand. J. Statist. **11** 171-185.

[17] Gasser, Th., Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. J. R. Statist. Soc. B **47** 238-252.

[18] Hall, P. (1989). Unusual properties of bootstrap confidence intervals in regression problems. Probab. Th. and Rel. Fields **81** 247-273.

[19] Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. Ann. Statist. **20** 695-711.

[20] Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. Stoch. Proc. and Appl. **36** 339-351.

[21] Hannan, E.J. and Kavalieris, L. (1986). Regression, autoregression models. J. Time Series Anal. **7** 27-49.

[22] Härdle, W. and Bowman, A.W. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. J. Amer. Statist. Assoc. **83** 102-110.

[23] Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. Ann. Statist. **19** 778-796.

[24] Härdle, W. and Tuan, P.-D. (1986). Some theory of M-smoothing of time series. J. Time Series Anal. **7** 191-204.

[25] Hart, J.D. (1991). Kernel regression estimation with time series errors. J. R. Statist. Soc. B **53** 173-187.

[26] Hart, J.D. (1994). Automated kernel smoothing of dependent data using time series cross-validation. J. R. Statist. Soc. B **56** 529-542.

[27] Herrmann, E., Gasser, Th. and Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. Biometrika **79** 783-795.

[28] Kreiss, J.-P. (1988). Asymptotic Statistical Inference for a Class of Stochastic Processes. Habilitationsschrift, Universität Hamburg.

[29] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. Ann. Statist. **17** 1217-1241.

[30] Nadaraya, E.A. (1964). On estimating regression. Theory Probab. Appl. **10** 186-190.

[31] Rosenblatt, M. (1971). Curve estimates. Ann. Math. Statist. **42** 1815-1842.

[32] Seber, G.A.F. (1977). Linear Regression Analysis. Wiley, New York.

[33] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Statist. **8** 147-164.

[34] Truong, Y.K. (1991). Nonparametric curve estimation with time series errors. J. Statist. Plan. Inf. **28** 167-183.

[35] Watson, G.S. (1964). Smooth regression analysis. Sankya A **26** 359-372.

[36] Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (including discussion). Ann. Statist. **14** 1261-1295. (Corr: V**16** p479).

Department of Statistics
367 Evans Hall #3860
University of California
Berkeley, CA 94720-3860
USA