Bootstrap of kernel smoothing in nonlinear time series

Jürgen Franke
Universität Kaiserslautern

Jens-Peter Kreiss
Technische Universität Braunschweig

Enno Mammen
Ruprecht-Karls-Universität Heidelberg

July 30, 1997

Abstract

Kernel smoothing in nonparametric autoregressive schemes offers a powerful tool in modelling time series. In this paper it is shown that the bootstrap can be used for estimating the distribution of kernel smoothers. This can be done by mimicking the stochastic nature of the whole process in the bootstrap resampling or by generating a simple regression model. Consistency of these bootstrap procedures will be shown.

1 Introduction

Nonlinear modelling of time series has appeared as a promising approach in applied time series analysis. A lot of parametric models can be found in the books of Priestley (1988) and Tong (1990). In this paper we consider nonparametric models of nonlinear autoregression. Motivated by econometric applications, we allow for heteroschedastic errors:

$$X_t = m(X_{t-1}, \dots , X_{t-p}) + \sigma(X_{t-1}, \dots , X_{t-q})\, \varepsilon_t, \quad t = 0, 1, 2, \dots . \tag{1.1}$$

Here (ε_t) are i.i.d. random variables with mean 0 and variance 1. Furthermore, m and σ are unknown smooth functions. Ergodicity and mixing properties of such processes have been discussed in Diebolt and Guegan (1990). For simplicity, in this paper we consider only the case $p = q = 1$. In this particular case, (1.1) can be interpreted as discrete versions of the general Black-Scholes formula with arbitrary (nonlinear) trend m and volatility function σ

$$dS_t = m(S_t) + \sigma(S_t)\; dW_t\ ,$$

where W_t is a standard Wiener process. The class of processes (1.1) contains also as a special case the QTARCH processes. These processes were proposed by Gourieroux and Montfort (1990) as models for financial time series.
Estimation of m and σ can be done by kernel smoothing of Nadaraya-Watson type:

$$(1.2) \qquad \hat{m}_h(x) = \frac{1}{T-1} \sum_{t=1}^{T-1} K_h(x - X_t)\ X_{t+1}\ /\ \hat{p}_h(x)\ ,$$

$$(1.3) \qquad \hat{\sigma}_h^2(x) = \frac{1}{T-1} \sum_{t=1}^{T-1} K_h(x - X_t)\ X_{t+1}^2\ /\ \hat{p}_h(x) - \hat{m}_h^2(x)\ .$$

Here $K_h(\cdot)$ denotes $h^{-1}K(\cdot/h)$ for a kernel K. The estimate $\hat{p}_h$ is a kernel estimate of the univariate stationary density p of the time series $\{X_t\}$

$$(1.4) \qquad \hat{p}_h(x) = \frac{1}{T-1} \sum_{t=1}^{T-1} K_h(x - X_t)\ .$$

Asymptotic normality of $\hat{m}_h$, $\hat{\sigma}_h$ and $\hat{p}_h$ has been shown in Robinson (1983). Uniform consistency results have been given in Collomb and Härdle (1986), Härdle and Vieu (1992), Delecroix (1987) and Ango Nze and Portier (1994). Asymptotic expansions for bias and variance have been derived in Auestad and Tjøstheim (1990) and Masry and Tjøstheim (1994). Tests for parametric models based on the comparison of these estimates and parametric estimates have been proposed in Hjellvik and Tjøstheim (1993), compare also Yao and Tong (1995).
Recently, so-called local polynomial estimators for m and σ have attracted much interest in the literature. For nonparametric regression these estimators have been studied in Stone (1977), Tsybakov (1986), and Fan (1992, 1993) [see also Fan and Gijbels (1992, 1995)]. Härdle and Tsybakov (1995) applied the idea of local polynomial fitting to autoregressive models. As an example consider a r-th order local polynomial estimator of m, which is given as $\hat{a}_o$, where $(\hat{a}_o, \dots, \hat{a}_{r-1})^T$ minimizes

$$\sum_{t=1}^{T-1} K_h\left(x - X_t\right) \left(X_{t+1} - \sum_{j=0}^{r-1} a_j \left(\frac{x - X_t}{h} \right)^j \right)^2 .$$

In particular for $r = 2$, a local linear estimator $\hat{m}_h^{loclin}$ of m can be written as a modified Nadaraya-Watson type estimator:

$$(1.5) \qquad \hat{m}_h^{loclin}(x) = \hat{m}_h(x) + \frac{\sum_t X_{t+1} \left(X_t - \hat{\xi}(x) \right) K_h\left(x - X_t\right)}{\sum_t \left(X_t - \hat{\xi}(x) \right)^2 K_h\left(x - X_t\right)} \left(x - \hat{\xi}(x) \right)\ ,$$

where $\hat{\xi}(x) = \sum_t X_t K_h(x - X_t) / \sum_t K_h(x - X_t)$ denotes the center of the design points around x. All bootstrap results presented in this paper also hold true for local polynomials. It is only for the sake of simplicity that we restrict our attention in the following to the case $r = 1$, i.e. to kernel estimates $\hat{m}_h$ and $\hat{\sigma}_h$, cf. (1.2) and (1.3).

In this paper several bootstrap procedures will be considered which approximate the laws of $\hat{m}_h$ and $\hat{\sigma}_h^2$. The first resampling scheme (*autoregression bootstrap*) follows a proposal of Franke and Wenzel (1992) and Kreutzberger (1993). This approach is similar to residual-based resampling of linear autoregressions as discussed by Kreiß and Franke (1992). It is based on generating a bootstrap process

$$X_t^* = \tilde{m}(X_{t-1}^*) + \tilde{\sigma}(X_{t-1}^*)\ \varepsilon_t^*\ ,$$

where $\tilde{m}$ and $\tilde{\sigma}$ are some estimates of m and σ and where $\varepsilon_1^*, \dots, \varepsilon_T^*$ is an i.i.d. resample. In our second bootstrap approach (*regression bootstrap*), a regression model is generated with (conditionally) fixed design $(X_0, \dots, X_T)$

$$X_t^* = \tilde{m}(X_{t-1}) + \tilde{\sigma}(X_{t-1})\ \varepsilon_t^*\ ,$$

where, again, an i.i.d. resample of residuals $\varepsilon_1^*, \dots, \varepsilon_T^*$ is used. In the third bootstrap, again a regression model is generated with (conditionally) fixed design $(X_0, \dots, X_T)$

$$X_t^* = \tilde{m}(X_{t-1}) + \eta_t^*\ .$$

Here $\eta_1^*, \dots, \eta_T^*$ is an independent resample where η_t^* has (conditional) mean zero and variance $(X_t - \hat{m}_h(X_{t-1}))^2$. This procedure has been called *wild bootstrap* by Mammen (1992), Härdle and Mammen (1994). Mathematics for autoregression bootstrap will turn out as the most difficult one. Note that in this bootstrap proposal a complicated resampling structure has to be generated.

The paper is organized as follows. An explicit description of the three bootstrap procedures can be found in the next section. In the third section we state our main results on consistency of the bootstrap procedures. Simulation results will be given in Section 4. Section 5 contains some auxiliary results on uniform convergence of $\hat{m}_h$ and $\hat{\sigma}_h^2$ on increasing subsets of the real line (cf. Lemma 5.1 and 5.3) which may be of some interest of its own. The proofs are defered to Section 6.

2 How to Bootstrap

We consider a stationary and geometrically ergodic process of the form

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})\ \varepsilon_t. \tag{2.6}$$

The unique stationary distribution is denoted by π. Simple sufficient conditions for stationarity and geometric ergodicity are the following

- The distribution of the i.i.d. innovations ε_t possesses a Lebesgue density p_ε, which fulfills $\inf_{x \in C} p_\varepsilon(x) > 0$ for all compact sets C

- m, σ and σ^{-1} are bounded on compact sets and $\limsup_{|x|\to\infty} \frac{\mathcal{E}|m(x)+\sigma(x)\varepsilon_1|}{|x|} < 1$.

This is a direct consequence of Theorems 1 and 2 in Diebolt and Guegan (1990), compare also Meyn and Tweedie (1993) or Doukhan (1995, p. 106/107). The assumptions ensure that the stationary distribution π of the time series $\{X_t\}$ possesses a strictly positive Lebesgue density, which we denote by p. From (2.6) we obtain

$$p(x) = \int_{\mathbb{R}} \frac{1}{\sigma(u)} p_\varepsilon \left(\frac{x - m(u)}{\sigma(u)} \right) \, d\pi(u). \tag{2.7}$$

For a stationary solution of (2.6), geometric ergodicity implies that the process is strongly mixing (α-mixing) with geometrically decreasing mixing coefficients (cf. Doukhan, 1995, chapter 2.4 and 1.3). Moreover this property carries over to processes of the type $Y_t = f_t(X_t)$.
To keep our proofs simple, we need somewhat stronger assumptions

(**A1**) m is Lipschitz continuous with constant L_m.
(**A2**) σ is Lipschitz continuous with constant L_σ.
(**A3**) $\sigma(x) > 0$ for all $x \in \mathbb{R}$.
(**A4**) The innovations ε_t are i.i.d. random variables with mean 0, variance 1 and a density p_ε satisfying $\inf_{x\in C} p_\varepsilon(x) > 0$ for all compact sets C.
(**A5**) $L_m + L_\sigma \, \mathcal{E}|\varepsilon_1| < 1$.

For the sake of simplicity we assume that the observed data $X_1, \ldots, X_T$ are realizations of the stationary version of (2.6).

2.1 Autoregression Bootstrap

Let $I = [-\gamma_T, \gamma_T]$ be a growing interval with $\gamma_T \longrightarrow \infty$ for $T \longrightarrow \infty$. More detailed assumptions on γ_T will be given later. We define

$$\tilde{m}_h(x) = \hat{m}_h(x) \cdot 1\{|x| \le \gamma_T\} \tag{2.8}$$

$$\tilde{\sigma}_h(x) = \hat{\sigma}_h(x) \cdot 1\{|x| \le \gamma_T\} + 1\{|x| > \gamma_T\} \ . \tag{2.9}$$

Outside of I the estimates $\hat{m}_h$ and $\hat{\sigma}_h$ are replaced by constants. This is done because $\hat{m}_h(x)$ and $\hat{\sigma}_h(x)$ are no reliable estimates for $|x|$ large. Other definitions of $\tilde{m}_h$ and $\tilde{\sigma}_h$ outside of I would work, too.
The bootstrap procedure requires calculation of residuals

$$\hat{\varepsilon}_j = \frac{X_j - \hat{m}_g(X_{j-1})}{\hat{\sigma}_g(X_{j-1})}, \ \ j = 1, \ldots, T \ ,$$

where $g > 0$ denotes a bandwidth possibly different to the bandwidth $h > 0$ used for the kernel smoother of interest. We remove those $\hat{\varepsilon}_j$ corresponding to the X_{j-1} outside of $[-\gamma_T, \gamma_T]$. Let $A = \{j = 1, \ldots, T \mid |X_{j-1}| \leq \gamma_T\}$. Then, we recenter the remaining residuals

$$\tilde{\varepsilon}_j = \hat{\varepsilon}_j - \frac{1}{|A|} \sum_{k \in A} \hat{\varepsilon}_k,$$

and define $\hat{F}_T$ as the empirical distribution given by $\tilde{\varepsilon}_j,\ j \in A$. Then, we smooth this distribution by convoluting it with some probability density $H_b(u) = \frac{1}{b}\ H(\frac{u}{b})$, where H is a probability density with mean 0 and variance 1. Let $\hat{F}_{T,b} = \hat{F}_T * H_b$ be this smoothed empirical law. Let us denote the density of $\hat{F}_{T,b}$ by $\hat{f}_{T,b}$. We draw the bootstrap residuals $\varepsilon_t^*,\ t = 1, \ldots, T$, as i.i.d. variables from $\hat{F}_{T,b}$. Then, we get the bootstrap sample $X_1^*, \ldots, X_T^*$ by

$$X_t^* = \tilde{m}_g(X_{t-1}^*) + \tilde{\sigma}_g(X_{t-1}^*)\ \varepsilon_t^*$$

with, for sake of simplicity, $X_0^* = X_0$.

Analogously to (1.2) the bootstrap sample $X_1^*, \ldots, X_T^*$ defines for each point x a kernel estimate $\hat{m}_h^*(x)$. The conditional distribution of $\sqrt{Th}\{\hat{m}_h^*(x) - \tilde{m}_g(x)\}$ given $X_1, \ldots, X_T$ is denoted by $\mathcal{L}_B(x)$. This is the bootstrap estimate of $\mathcal{L}(x)$, the distribution of $\sqrt{Th}\{\hat{m}_h(x) - m(x)\}$.

The distribution of $\sqrt{Th}\{\hat{\sigma}_h^2(x) - \sigma^2(x)\}$ is denoted by $\mathcal{L}^\sigma(x)$, its bootstrap estimate by $\mathcal{L}_B^\sigma(x)$. Consistency of these estimates will be shown in the next section.

2.2 Regression Bootstrap

With an i.i.d. resample $\varepsilon_1^*, \ldots, \varepsilon_T^*$, generated as in the last subsection, we put

$$X_t^* = \hat{m}_g(X_{t-1}) + \hat{\sigma}_g(X_{t-1})\ \varepsilon_t^*\ .$$

Here $\hat{m}_g$ and $\hat{\sigma}_g$ are kernel smoothing estimates (cf. (1.2), (1.3)) with bandwidth g. The original sample $X_1, \ldots, X_T$ acts in the resampling as a fixed design. We now define

$$\begin{aligned}
\hat{m}_h^*(x) &= \frac{1}{T-1} \sum_{t=1}^{T-1} K_h(x - X_t)\ X_{t+1}^*\ /\ \hat{p}_h(x)\ , \\
\hat{\sigma}_h^{*^2}(x) &= \frac{1}{T-1} \sum_{t=1}^{T-1} K_h(x - X_t)\ X_{t+1}^{*^2}\ /\ \hat{p}_h(x) - \hat{m}_h^{*^2}(x)\ .
\end{aligned}$$

The conditional distribution of $\sqrt{Th}\{\hat{m}_h^*(x) - \hat{m}_g(x)\}$ is denoted by $\mathcal{L}_{RB}(x)$ and the conditional distribution of $\sqrt{Th}\{\hat{\sigma}_h^{*^2}(x) - \hat{\sigma}_g^2(x)\}$ is denoted by $\mathcal{L}_{RB}^\sigma(x)$. These are our second type of bootstrap estimates for $\mathcal{L}(x)$ and $\mathcal{L}^\sigma(x)$.

2.3 Wild Bootstrap

This procedure starts by generating an i.i.d. sample $\overline{\eta}_1, \ldots, \overline{\eta}_T$ with mean 0 and variance 1. [Often, for a higher order performance, the distribution of $\overline{\eta}_t$ is chosen such that additionally $E\ \overline{\eta}_t^3 = 1$; for a discussion of this point and for choices of the distribution of $\overline{\eta}_t$, compare Mammen (1992).] Put now $\eta_t^* = (X_t - \hat{m}_h(X_{t-1}))\ \overline{\eta}_t$. The Wild Bootstrap resample is defined as

$$X_t^* = \hat{m}_g(X_{t-1}) + \eta_t^* \ .$$

As in the last subsection, this resample can be used for calculating $\hat{m}_h^*(x)$. The conditional distribution of $\sqrt{Th}\{\hat{m}_h^*(x) - \hat{m}_g(x)\}$ is denoted by $\mathcal{L}_{WB}(x)$. In particular, Wild Bootstrap is appropriate in cases of irregular variance functions $\sigma(x)$. Such models may arise when $\sigma(x)$ acts only as a nuisance parameter and the main interest lies in estimating m.

3 Bootstrap Works

In this section we present our main results. We give assumptions under which the three Bootstrap procedures of the last section are consistent. We start with the first Bootstrap procedure. Here and in the following, C denotes a positive generic constant.

(**B1**) There exists $\sigma_o > 0$ such that $\sigma(x) \geq \sigma_o$ for all $x \in \mathbb{R}$.

(**B2**) m and σ are twice continuously differentiable with bounded derivatives.

(**B3**) $E\varepsilon_1^6 < \infty$. p_ε is twice continuously differentiable. $p_\varepsilon, p_\varepsilon'$ and p_ε'' are bounded and $\sup_{x\in\mathbb{R}} |x\, p_\varepsilon'(x)| < \infty$

(**B4**) $g, h \to 0,\ Th^5 \to B^2 \geq 0$ and $g \sim T^{-\alpha}$ with $0 < \alpha \leq \frac{2}{15}$ for $T \to \infty$,

(**B5**) $b \to 0$ and $g/b^{12} \to 0$ as $T \to \infty$.

(**B6**) $\gamma_T \to \infty,\ \inf_{|x|\leq 2\ \gamma_T/\sigma_0} p_\varepsilon(x) \geq (g \log T)^2$ and $\gamma_T / \log T$ is bounded.

(**B7**) H is a probability density, twice continuously differentiable with bounded derivatives and satisfies $\int v^4 H(v) dv < \infty,\ \int v^2 |H'(v)| dv < \infty$.

(**B8**) K has compact support $[-1, 1]$, say. K is symmetric, nonnegative and three times continuously differentiable with $K(1) = K'(1) = 0$ and $\int K(v)\ dv = 1$.

Assumption (B4) allows for the rate $h \sim T^{-1/5}$ as well as for faster rates of convergence. Bandwidths of order $O(T^{-1/5})$ have been motivated by optimality considerations. For bandwidths of order $o(T^{-1/5})$ the variances of $\hat{m}_h, \hat{\sigma}_h$ dominate the bias parts. By comparison with bootstrapping nonparametric statistics in other simpler situations oversmoothing of the reference estimates $\tilde{m}_g, \tilde{\sigma}_g$ in the sense that $Tg^5 \to \infty$ seems to be necessary. We require a bit more due to technical reasons.

Condition (B5) is needed for purely technical reasons in the proof of Lemma 6.5. It implies together with (B4) a very slow convergence of b to 0. In simulations the bootstrap seems to work even without any smoothing (corresponding to $b \equiv 0$ for finite T).

We are now ready to state our first theorem.

Theorem 1: *Assume (A1) - (A5), (B1) - (B8). Then for all* $x \in \mathbb{R}$:

$$d_K(\mathcal{L}_B(x),\ \mathcal{L}(x)) \longrightarrow 0 \quad \textit{(in probability)},$$

$$d_K(\mathcal{L}_B^\sigma(x),\ \mathcal{L}^\sigma(x)) \longrightarrow 0 \quad \textit{(in probability)} .$$

Here d_K *denotes the Kolmogorov distance, i.e. for two distributions* P *and* Q *the distance* $d_K(P,Q)$ *is defined as* $\sup_{x\in\mathbb{R}} |P(X \le x) - Q(X \le x)|$.

We come now to the discussion of regression bootstrap. We assume

(**RB**) Assume (B3), (B4), and (B8). Furthermore, suppose that σ is continuously differentiable and that m is twice continuously differentiable with bounded derivatives.

Theorem 2: *Assume (A1) - (A5), (RB). Then for all* $x \in \mathbb{R}$:

$$d_K(\mathcal{L}_{RB}(x),\ \mathcal{L}(x)) \longrightarrow 0 \quad \textit{(in probability)}$$

$$d_K(\mathcal{L}_{RB}^\sigma(x),\ \mathcal{L}^\sigma(x)) \longrightarrow 0 \quad \textit{(in probability)}.$$

We come now to the Wild Bootstrap. We assume

(**WB**) Assume (B3), (B4), (B8), that m is twice continuously differentiable with bounded derivatives and that σ is continuous.

Theorem 3: *Assume (A1) - (A5), (WB). Then for all* $x \in \mathbb{R}$:

$$d_K(\mathcal{L}_{WB}(x),\ \mathcal{L}(x)) \longrightarrow 0 \quad \textit{(in probability)}.$$

Remark. Note that less smoothness assumptions on σ are made for wild bootstrap compared with regression bootstrap. Furthermore, autoregression bootstrap requires even more smoothness assumptions as regression bootstrap.

4 Simulations

In this section we intend to demonstrate the finite sample size performance of the bootstrap and wild bootstrap proposal of the paper. For this purpose we consider the processes ($t = 1, \ldots, T$)

$$X_t = 4 \cdot \sin(X_{t-1}) + \varepsilon_t, \tag{4.10}$$

$$X_t = \sqrt{1 + 0.8\, X_{t-1}^2} \cdot \varepsilon_t, \tag{4.11}$$

$$X_t = 0.9\, X_{t-1} + \sqrt{0.5 + 0.25 X_{t-1}^2}\, \varepsilon_t. \tag{4.12}$$

Here $\varepsilon_t : t = 1, \ldots, T$ are i.i.d. error variables with standard normal law. Equation (4.11) is a model of ARCH(1)-type, and (4.12) is a discrete version of the Black-Scholes formula for stock prices. It has been modified by assuming a nonconstant volatility. In both cases, $\sigma(x)$ grows proportional to x.

Figure 1a and 1b show typical realizations of size $T = 250$ of the models (4.10) and (4.11).

At first we consider the local linear estimator $\hat{m}_h^{loclin}$ for m in the first model with bandwidth $h = 0.4$. Based on a Monte Carlo simulation of size $M = 2000$, Figure 2a and 2b show the simulated density of $\sqrt{Th}(\hat{m}_h^{loclin}(x) - m(x))$ for $x = 0$ and $x = -\pi/2$ (thick lines) together with three bootstrap estimates of this quantity (thin lines) based upon different original time series. Here we make use of the bootstrap proposal of Section 2.1. The pilot bandwidth g is chosen to be equal to 1, and the size of the resample is 2000.

Figures 3a and 3b are devoted to the behaviour of the usual kernel estimator $\hat{\sigma}_h$ of the volatility function $\sigma(x) = \sqrt{1 + 0.8\,x^2}$ in model (4.11). In this case all bootstrap estimates are again obtained by using the first bootstrap proposal (cf. Section 2.1). The plots show again three different bootstrap approximations together with the simulated true distribution of $\sqrt{Th}(\hat{\sigma}_h(x) - \sigma(x))$ for $x = 0$ and $x = 1$, respectively.

In both cases, the bootstrap provides a reasonable approximation of the densities of the estimators of interest.
Finally Figure 4a (for model (4.10)) and Figure 4b (for model (4.11)) give us an impression of the density of the stationary distribution π of the corresponding processes (X_t).

Considering model (4.12), we illustrate how the bootstrap can be used to get approximative confidence intervals and to select an appropriate bandwidth. Figure 5 shows the data, i.e. a sample of size $T = 500$ from (4.12). Figure 6a-c show the kernel estimates with bandwidth $h = 0.8$ of the trend function $m(x) = 0.9\,x$, the volatility function $\sigma(x) = \sqrt{0.5 + 0.25\,x^2}$ and the stationary density of (4.12). As our sample is essentially contained in the interval $[-4, 6]$, the estimates are of course quite poor outside of this interval.

Figure 7a shows a pointwise 90%-confidence band for $m(x)$ based on a Monte Carlo simulation of size $M = 500$, whereas Figure 7b provides the bootstrap approximation of this confidence band based on the sample of Figure 5 and using $g = 1$. Here, as in the above cases too, we use the unsmoothed law of the sample residuals for the resample, i.e. $b = 0$. This case is not covered by our theoretical results, but it works in practice quite well. The two confidence bands are quite close in the central part around 0 where we have enough data in the sample of Figure 5.

Analogously, Figure 8a-b and 9a-b show pointwise 90%-confidence bands for $\sigma(x)$ and for the stationary density $p(x)$. In the interval $[-2.5, 4.5]$, the bootstrap provides a good approximation of the confidence band for $p(x)$ apart from a slight shift to the left near 0 - for $p(0)$, e.g., the 90%-bootstrap confidence interval is $[0.19, 0.28]$, compared to the Monte Carlo result of $[0.20, 0.30]$. The bootstrap confidence band for $\sigma(x)$ has a similar shape as the Monte Carlo band, but it is considerably shifted to the right for x around 0. This is not surprising because variance function estimates are not very reliable even for sample sizes of T of order 500. From Figure 6b we see that for our particular sample the estimate $\hat{\sigma}_h(x)$ lies by chance considerably above $\sigma(x)$. This cannot be caused by smoothing bias alone, as can be seen by looking at other kernel estimates with smaller h.

Finally, Figure 10a-b for $m(x)$ and Figure 11a-b for $\sigma(x)$ show Monte Carlo estimates and the corresponding bootstrap approximations for the root mean-square (rms) error of $\hat{m}_h(x)$ and $\hat{\sigma}_h(x)$ as function of x. Between -4 and 4 the bootstrap approximation comes very close to the "true" rms-curves [only for $\hat{\sigma}_h(x)$ near 0 the bootstrap-rms is a bit too small.] It is also possible to consider the rms as function of h for fixed x. Then its bootstrap approximation can be used for local bandwidth selection.

5 Auxiliary results: Uniform Convergence of the Kernel Smoothers

In this section we collect some results on uniform convergence of our estimates $\hat{m}_h$ and $\hat{\sigma}_h$ on slowly growing intervals of the form $[-\gamma_T, \gamma_T]$, $\gamma_T \to \infty$ as $T \to \infty$. These results are essential for our proof of consistency of the bootstrap proposals of Section 2. For all bootstrap procedures it is not sufficient to consider behaviour of $\hat{m}_h$ and $\hat{\sigma}_h$ only on fixed compact sets.

Lemma 5.1: *Assume (A1)-(A5), (B1)-(B4), (B6) and (B8). Then we have*

$$\sup_{|x|\le\gamma_T} |\hat{m}_g(x) - m(x)| \;=\; o_P\left(g^{1/6}\right) \;.$$

Proof: We use the decomposition

$$\hat{m}_g(x) - m(x)$$

$$= \frac{\sum_t K_g\left(x - X_t\right)\sigma\left(X_t\right)\varepsilon_{t+1}}{\sum_t K_g\left(x - X_t\right)} + \frac{\sum_t K_g\left(x - X_t\right)\left(m(X_t) - m(x)\right)}{\sum_t K_g\left(x - X_t\right)}.$$

By our assumption on g, it suffices to show

$$\text{(5.13)} \qquad \sup_{|x|\le\gamma_T} \left| \frac{1}{T}\sum_t K_g\left(x - X_t\right)\sigma\left(X_t\right)\varepsilon_{t+1} \right| \;=\; \mathcal{O}_P\left(\left(Tg\right)^{-1/3}\right),$$

$$\text{(5.14)} \qquad \sup_{|x|\le\gamma_T} \left| \frac{1}{T}\sum_t K_g\left(x - X_t\right) - p(x) \right| \;=\; \mathcal{O}_P\left(g^2\right),$$

$$\text{(5.15)} \qquad \inf_{|x|\le\gamma_T} p(x) \;\ge\; Cg^2 \log T$$

and

$$\text{(5.16)} \qquad \sup_{|x|\le\gamma_T} \left| \frac{\sum_t K_g\left(x - X_t\right)\left(m\left(X_t\right) - m(x)\right)}{\sum_t K_g\left(x - X_t\right)} \right| = \mathcal{O}_P(g) \;.$$

Claim (5.16) is an easy consequence of the differentiability of m. Note that the lefthand side of (5.16) is bounded by

$$\sup_x \frac{\sum_t K_g\left(x - X_t\right)|x - X_t|}{\sum_t K_g\left(x - X_t\right)} \cdot \sup_x \left|m'(x)\right| .$$

This is of order $\mathcal{O}(g)$ due to the compactness of the support of K. A proof of (5.13) is a bit more involved. Since we will make repeatedly use of the following argument we present it here in detail. In a first step we divide the interval $[-\gamma_T, \gamma_T]$ into equidistant subintervals of length $\delta = \left(g^5/T\right)^{1/3}$. We get

$$\frac{1}{T} \sup_{|x|\leq\gamma_T} \left| \sum_t K_g\left(x - X_t\right)\sigma\left(X_t\right)\varepsilon_{t+1} \right| \leq \max_i \sup_x \frac{1}{T} \left| \sum_t K_g\left(x - X_t\right)\sigma\left(X_t\right)\varepsilon_{t+1} \right|$$

where the suprema on the right hand side are taken over all $x \in [-\gamma_T + (i-1)\delta, -\gamma_T + i\delta]$ and where the maximum is taken over all $i \in \{1, ..., [2\gamma_T/\delta] + 1\}$. Let us denote $x_i = -\gamma_T + (i-1)\delta$. By the mean value theorem we get the following upper bound for the right hand side of the last inequality

$$\max_i \frac{1}{T} \left| \sum_t K_g\left(x_i - X_t\right)\sigma(X_t)\varepsilon_{t+1} \right| + \frac{\delta}{g^2}\frac{C}{T} \sum_t \sigma(X_t)|\varepsilon_{t+1}| \ ,$$

where C is some upper bound of $|K'|$. Since $\sum_t \sigma(X_t)|\varepsilon_{t+1}| = \mathcal{O}_P(T)$ we get with our choice of δ that the second term is of order $\mathcal{O}_P((Tg)^{-1/3})$. It remains to show that the first term is of order $\mathcal{O}_P((Tg)^{-1/3})$. For this purpose, we consider

$$\begin{aligned}
& P\left\{ \max_i \left| \sum_t K_g\left(x_i - X_t\right)\sigma(X_t)\varepsilon_{t+1} \right| \geq \left(T^2/g\right)^{1/3} \right\} \\
\leq\ & \sum_i P\left\{ \left| \sum_t K_g\left(x_i - X_t\right)\sigma(X_t)\varepsilon_{t+1} \right| \geq \left(T^2/g\right)^{1/3} \right\} \\
\leq\ & \frac{g^2}{T^4} \sum_i \mathcal{E} \left| \sum_t K_g\left(x_i - X_t\right)\sigma(X_t)\varepsilon_{t+1} \right|^6 \\
\leq\ & \mathcal{O}(1) \cdot \frac{g^2}{T} \sum_i \mathcal{E}\left[K_g^6\left(x_i - X_1\right)\sigma^6\left(X_1\right)\right]\mathcal{E}\varepsilon_1^6 \ ,
\end{aligned}$$

by Burkholder's inequality (cf. Hall and Heyde (1980), Theorem 2.10). We obtain that the last expression is of order $\mathcal{O}\left((\log T)^7/(Tg^7)^{2/3}\right)$, which is $o(1)$ by the assumption on g, since

$$\mathcal{E}K_g^6\left(x_i - X_1\right)\sigma^6\left(X_1\right) \leq \sup_{|x|\leq\gamma_T} \sigma^6(x) \cdot \mathcal{O}(g^{-5}) = \mathcal{O}\left(\frac{\gamma_T^6}{g^5}\right)$$

(5.14) is an immediate consequence of

$$\text{(5.17)} \qquad \sup_x \left| \frac{1}{T} \sum_t K_g\left(x - X_t\right) - \mathcal{E}K_g\left(x - X_1\right) \right| = \mathcal{O}_P\left(\frac{(\log T)^3}{\sqrt{Tg^{1+\varepsilon}}}\right)$$

$\varepsilon > 0$ arbitrary, and

$$\sup_x |\mathcal{E}K_g(x - X_1) - p(x)| \;=\; \mathcal{O}(g^2) \; . \tag{5.18}$$

To see (5.18) observe that $\mathcal{E}K_g(x - X_1) = \int K(v)p(x - gv)\,dv$. A Taylor expansion for p together with the fact that $\int vK(v)\,dv = 0$ (K is symmetric!) yields the desired result.

In order to prove (5.17) we make use of an exponential inequality for strong mixing processes (cf. Doukhan (1995), Proposition 1, p. 33). Before doing so we apply the splitting device for the supremum over x,$|x| \le \gamma_T$, discussed above. It turns out that it suffices to consider

$$\max_i \left| \frac{1}{T}\sum_t K_g(x_i - X_t) - \mathcal{E}K_g(x_i - X_1) \right| \;+\; \mathcal{O}\left(\frac{\delta}{g^2}\right) \; .$$

For the choice $\delta/g^2 = (\log T)^3 / \sqrt{Tg^{1+\varepsilon}} with arbitrary \varepsilon > 0$ the second term is of the desired order. For the first rerm, the above mentioned exponential inequality gives us that

$$\begin{aligned}
& P\left\{ \max_i \left| \frac{1}{T}\sum_t K_g(x_i - X_t) - \mathcal{E}K_g(x_i - X_t) \right| \ge \frac{M^2 (\log T)^3}{\sqrt{Tg^{1+\varepsilon}}} \right\} \\
\le \;& \sum_i P\left\{ \left| \sum_t \{K_g(x_i - X_t) - \mathcal{E}K_g(x_i - X_t)\}\, g \right| \ge M^2 \sqrt{Tg^{1-\varepsilon}} (\log T)^3 \right\} \\
\le \;& \mathcal{O}\left(\frac{\gamma_T}{\delta}\right) \cdot \exp\left(-b\sqrt{M}\log T\right) ,
\end{aligned}$$

for some constant $b > 0$. This is of the order $=\; o(1)$ for M large enough.

It remains to verify (5.15). With (2.7) we obtain

$$\begin{aligned}
\inf_{|x| \le \gamma_T} p(x) \ge \quad & \inf_x \int_{[-\gamma_T, \gamma_T]} \frac{1}{\sigma(u)} p_\varepsilon\left(\frac{x - m(u)}{\sigma(u)}\right) \, d\pi(u) \\
\ge \quad & \int_{[-\gamma_T, \gamma_T]} \frac{1}{C\gamma_T^\beta} \inf_{|v| \le 2\gamma_T/\sigma_o} p_\varepsilon(v) \; d\pi(u) \; ,
\end{aligned}$$

since, for T large enough, $|(x - m(u))/\sigma(u)| \le (\gamma_T + L_m\gamma_T + |m(0)|)/\sigma_o \le 2\gamma_T/\sigma_o$ for all $x, u \in [-\gamma_T, \gamma_T]$. Assumption (B6) together with $\pi[-\gamma_T, \gamma_T] \to 1$ yields the desired result. ∎

Lemma 5.2: *Under the assumptions of Lemma 5.1 we have on every compact interval B*

$$\sup_{x \in B} |\hat{m}_g(x) - m(x)| \;=\; \mathcal{O}_P\left(g^2\right)$$

Proof: As B is a fixed interval, p is bounded away from 0 by a fixed constant on B. Therefore, by the same type of argument used in the proof of Lemma 5.1

$$\frac{\sum_t K_g(x-X_t)\sigma(X_t)\varepsilon_{t+1}}{\sum_t K_g(x-X_t)} = \mathcal{O}_P(g^2)$$

uniformly on C under the assumption on g. Therefore, it remains to show

$$\sup_{x\in B}\left|\frac{\sum_t K_g\,(x-X_t)\,(m(X_t)-m(x))}{\sum_t K_g\,(x-X_t)}\right| \;=\; \mathcal{O}_P(g^2)\ .$$

A Taylor expansion of $m(X_t)-m(x)$ up to second order terms yields for the numerator

$$\frac{1}{T}\sum_t K_g\,(x-X_t)\,(X_t-x)\,m^{'}(x)+\frac{1}{2T}\sum_t K_g\,(x-X_t)\,(X_t-x)^2\,m^{''}(\hat{x}_t)\ .$$

The second term divided by $1/T\sum_t K_g\,(x-X_t)$ is obviously of order g^2 (recall that $m^{''}$ is bounded). For the first term, application of the exponential inequality [cited in the proof of Lemma 5.3] and of the same splitting device for the supremum over x as above concludes the proof. ∎

Remark. Under stronger assumptions (including the assumption that the Laplace transform $\int\exp(\delta u)p_\varepsilon(u)\;du$ of p_ε exists for $|\delta|$ small enough) we are able to show that the following stronger result holds.

$$\sup_{|x|\le\gamma_T}\left|\frac{1}{T}\sum_t K_g\,(x-X_t)\,\sigma(X_t)\varepsilon_{t+1}\right| \;=\; \mathcal{O}_P\left(\frac{\log T}{\sqrt{Tg}}\right)$$

Together with Lemma 5.2, this implies a known uniform convergence result for m on compact sets, cf. Masry and Tjøstheim (1994). Since we don't need better rates, we don't give more details here.

Additionally, we need uniform convergence of $\hat{\sigma}_g$ on the growing interval $[-\gamma_T,\gamma_T]$. This is the content of the following lemma.

Lemma 5.3: *Under the assumptions of Lemma 5.1, we have*

$$\sup_{|x|\le\gamma_T}|\hat{\sigma}_g(x)-\sigma(x)| \;=\; o_P(g^{1/6}\gamma_T)\ .$$

Proof: From (B1) we have $\sigma(x)\ge\sigma_o>0$ for all $x\in\mathbb{R}$. $\hat{\sigma}_g$ satisfies

$$\hat{\sigma}_g^2(x) \;=\; \frac{\sum_t K_g(x-X_t)X_{t+1}^2}{\sum_t K_g(x-X_t)}-\hat{m}_g^2(x) \;\ge\; 0\ .$$

Since $\sigma^2(x) = \mathcal{E}\left[X_{t+1}^2 | X_t = x\right] - m^2(x)$ we obtain

$$\sup_{|x|\le\gamma_T} |\hat\sigma_g(x) - \sigma(x)| \le \sup_{|x|\le\gamma_T} |\hat\sigma_g^2(x) - \sigma^2(x)| \cdot \sup_x |\hat\sigma_g(x) + \sigma(x)|^{-1}$$

$$\le \sigma_o^{-1}\left[\sup_{|x|\le\gamma_T}\left|\frac{\sum_t K_g\left(x - X_t\right) X_{t+1}^2}{\sum_t K_g\left(x - X_t\right)} - \mathcal{E}\left[X_{t+1}^2 | X_t = x\right]\right| + \sup_{|x|\le\gamma_T}\left|\hat m_g^2(x) - m^2(x)\right|\right] .$$

From Lemma 5.1 and from Lipschitz-continuity of m

$$\begin{aligned}
&\sup_{|x|\le\gamma_T}\left|\hat m_g^2(x) - m^2(x)\right| \\
&\le \sup_{|x|\le\gamma_T} |\hat m_g(x) - m(x)| \left[\sup_{|x|\le\gamma_T} |\hat m_g(x) - m(x)| + 2\sup_{|x|\le\gamma_T} |m(x)|\right] \\
&= o_P(g^{1/6}\gamma_T).
\end{aligned}$$

It therefore suffices to deal with

$$\frac{\sum_t K_g(x - X_t)X_{t+1}^2}{\sum_t K_g(x - X_t)} - \mathcal{E}\left[X_{t+1}^2 | X_t = x\right] = \frac{\sum_t K_g(x - X_t)\left(X_{t+1}^2 - m^2(x) - \sigma^2(x)\right)}{\sum_t K_g(x - X_t)} .$$

Since

$$X_{t+1}^2 - m^2(x) - \sigma^2(x)$$

$$= m^2(X_t) - m^2(x) + 2m(X_t)\sigma(X_t)\varepsilon_{t+1} + \sigma^2(X_t) - \sigma^2(x) + \sigma^2(X_t)\left(\varepsilon_{t+1}^2 - 1\right)$$

the assertion of Lemma 5.3 follows from (5.19) - (5.22) together with (5.14) and (5.15).

$$\sup_{|x|\le\gamma_T}\left|\sum_t K_g\left(x - X_t\right)\sigma^2(X_t)(\varepsilon_{t+1}^2 - 1)\right| = \mathcal{O}_P\left(\left(T^2/g\right)^{1/3}\right), \tag{5.19}$$

$$\sup_{|x|\le\gamma_T}\left|\sum_t K_g\left(x - X_t\right) m(X_t)\sigma(X_t)\varepsilon_{t+1}\right| = \mathcal{O}_P\left(\left(T^2/g\right)^{1/3}\right), \tag{5.20}$$

$$\sup_{|x|\le\gamma_T}\left|\frac{\sum_t K_g\left(x - X_t\right)\left(m^2(X_t) - m^2(x)\right)}{\sum_t K_g\left(x - X_t\right)}\right| = \mathcal{O}_P\left(g\gamma_T\right), \tag{5.21}$$

$$\sup_{|x|\le\gamma_T}\left|\frac{\sum_t K_g\left(x - X_t\right)\left(\sigma^2(X_t) - \sigma^2(x)\right)}{\sum_t K_g\left(x - X_t\right)}\right| = \mathcal{O}_P\left(g\gamma_T\right). \tag{5.22}$$

Claims (5.21) and (5.22) follow from the equalities $\sup_{|x|\le\gamma_T} |m(x)m'(x)| = \mathcal{O}(\gamma_T)$ and $\sup_{|x|\le\gamma_T} |\sigma(x)\sigma'(x)| = \mathcal{O}(\gamma_T)$, see (B2). Equations (5.19) and (5.20) can be shown analogously to (5.13). In the proof $\sigma(X_t)\varepsilon_{t+1}$ is replaced by $\sigma^2(X_t)(\varepsilon_{t+1}^2 - 1)$ or $m(X_t)\sigma(X_t)\varepsilon_{t+1}$, respectively. ■

The next lemma describes performance of $\hat\sigma$ on fixed compact sets B.

Lemma 5.4: *Under the assumptions of Lemma 5.1 we have on every compact interval B*

$$\sup_{x\in B} |\hat{\sigma}_g(x) - \sigma(x)| = \mathcal{O}_P(g^2\gamma_T).$$

Remark. As for the conditional mean function m, we can achieve better rates for the uniform convergence in Lemma 5.4 under stricter conditions.

We conclude this chapter with some weak consistency results concerning the derivatives of $\tilde{m}_g$.

Lemma 5.5: *Assume (A1) - (A5), (B2) - (B3) and (B8), and let $g \sim T^{-\alpha}$, $0 < \alpha < \frac{1}{5}$. For all $x \in \mathbb{R}$*

$$\begin{array}{ll} (i) & \tilde{m}_g'(x) \longrightarrow m'(x) \;\; \textit{in probability} \\ (ii) & \sup\limits_{u\in[x-h,x+h]} |\tilde{m}_g''(u) - m''(u)| \longrightarrow 0 \;\; \textit{in probability.} \end{array}$$

Proof: It suffices to deal with $\hat{m}_g$ instead of $\tilde{m}_g$, cf. (2.8). We have, abbreviating $g^{-2}\,K'(\cdot/g)$ by $K_g'(\cdot)$,

$$\hat{m}_g'(x) = \frac{\frac{1}{T}\sum_t K_g'(x-X_t)X_{t+1}}{\frac{1}{T}\sum_t K_g(x-X_t)} - \frac{\frac{1}{T}\sum_t K_g(x-X_t)X_{t+1}\frac{1}{T}\sum_t K_g'(x-X_t)}{\left(\frac{1}{T}\sum_t K_g(x-X_t)\right)^2}\,.$$

In the proofs of Lemmas 6.3 and 6.4, it is shown that

$$\frac{1}{T}\sum_t K_g(x-X_t) \to p(x) \quad \text{in probability}$$

$$\frac{1}{T}\sum_t K_g(x-X_t)X_{t+1} \to m(x)p(x) \quad \text{in probability.}$$

We will show that

$$\frac{1}{T}\sum_t K_g'(x-X_t) \longrightarrow p'(x), \tag{5.23}$$

$$\frac{1}{T}\sum_t K_g'(x-X_t)X_{t+1} \longrightarrow (m(x)p(x))' \tag{5.24}$$

in probability as $T \to \infty$. To see (5.23) observe that, by direct computation,

$$\mathcal{E}\left(\frac{1}{T}\sum_t \left(K_g'(x-X_t) - \mathcal{E}\left[K_g'(x-X_t)|\mathcal{F}_{t-1}\right]\right)\right)^2 = \mathcal{O}(1/(Tg^3)) = o(1).$$

Furthermore, we get

$$\begin{aligned}
&\frac{1}{Tg^2}\sum_t \mathcal{E}\left[K'\left(\frac{x-X_t}{g}\right)\Big|\mathcal{F}_{t-1}\right] \\
&= \frac{1}{Tg}\sum_t \int K'(v)\; p_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}-\frac{gv}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\;dv \\
&= -\frac{1}{T}\sum_t \int v\;K'(v)\;dv\;p'_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})^2}+\mathcal{O}_P(g),
\end{aligned}$$

since, by symmetry of K, $\int K'(v)\;dv=0$. Because of $K(-1)=K(1)=0$ we have $\int v\;K'(v)\;dv=-1$. This implies that

$$\begin{aligned}
&\frac{1}{Tg^2}\sum_t \mathcal{E}\left[K'\left(\frac{x-X_t}{g}\right)\Big|\mathcal{F}_{t-1}\right] \\
&= \frac{1}{T}\sum_t p'_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma^2(X_{t-1})}+\mathcal{O}_P(g).
\end{aligned}$$

By (2.7) and the ergodic theorem this converges towards $\frac{d}{dx}\left(\mathcal{E}\;p_\varepsilon\left(\frac{x-m(X_1)}{\sigma(X_1)}\right)\frac{1}{\sigma(X_1)}\right)=p'(x)$.

To see (5.24) replace X_{t+1} by $m(X_t)+\sigma(X_t)\varepsilon_{t+1}$ and treat both terms separately. We have

$$\mathcal{E}\left(\frac{1}{T}\sum_t K'_g(x-X_t)\,\sigma(X_t)\,\varepsilon_{t+1}\right)^2=\mathcal{O}(1/(Tg^3))=o(1)$$

and

$$\mathcal{E}\left(\frac{1}{T}\sum_t K'_g(x-X_t)m(X_t)-\mathcal{E}\left[K'_g(x-X_t)m(X_t)|\mathcal{F}_{t-1}\right]\right)^2=\mathcal{O}(1/(Tg^3))\;.$$

The remaining conditional expectation equals

$$\frac{1}{Tg}\sum_t \int K'(v)\,m(x-gv)\,p_\varepsilon\left(\frac{x-m(X_{t-1})-gv}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\;dv\;.$$

Differentiability of m and p_ε together with the facts that $\int K'(v)\;dv\;=\;0$ and $\int v\,K'(v)\;dv\;=\;-1$ gives us that this expression is equal to (up to terms of order $\mathcal{O}_P(g)$)

$$\frac{1}{T}\sum_t\left(m'(x)\,p_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)+m(x)p'_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\;.$$

The ergodic theorem concludes the proof of (i).

For the proof of (ii) one can proceed as in (i) to show that

$$\tilde{m}''_g(u)-m''(u)\longrightarrow 0\;\text{ in probability.}$$

Here we make use of $\int K''(v)\ dv = 0$, $\int v\, K''(v)\ dv = 0$ and $\int v^2\, K''(v)\ dv = 2$, which are easy consequences of (B8). It remains to show that this convergence holds uniformly for $u \in [x-h, x+h]$. This can be done e.g. by calculation of higher order moments of $\tilde{m}_g''(u) - m''(u)$. ■

6 Auxiliary Results and Proofs

Before proving the validity of all three bootstrap procedures we show some preliminary results on the performance of the first bootstrap procedure which tries to capture the time series structure of the process $\{X_t\}$. In particular, we show that the bootstrap innovations ε_t^* approximate the true residuals ε_t in Mallows distance defined as

$$d_2^2(X,Y) = d_2^2(\mathcal{L}(X), \mathcal{L}(Y)) = \inf\left\{ \mathcal{E}(U-V)^2 \;\middle|\; \mathcal{L}(U) = \mathcal{L}(X), \mathcal{L}(V) = \mathcal{L}(Y) \right\} .$$

Corresponding to the definition of $\hat{F}_T$ and $\hat{F}_{T,b}$, let F_T denote the empirical distribution of the $\varepsilon_j,\ \ j \in A$, and let $F_{T,b} = F_T * H_b$ denote the smoothed version of this empirical law. Let F_ε denote the law of the innovations ε_t.

Proposition 6.1: *For $T \to \infty$, we have*

$$\begin{aligned} d_2^2(\varepsilon_t, \varepsilon_t^*) &= d_2^2(F_\varepsilon, \hat{F}_{T,b}) \le 2\ d_2^2(F_\varepsilon, F_{T,b}) + 4 \sup_{|x| \le \gamma_T} \frac{(m(x) - \hat{m}_g(x))^2}{\hat{\sigma}_g^2(x)} + \\ &+ 5\ \mathcal{E}|\varepsilon_t| \cdot \sup_{|x| \le \gamma_T} \frac{(\sigma(x) - \hat{\sigma}_g(x))^2}{\hat{\sigma}_g^2(x)} + \mathcal{O}_P\left(\frac{1}{T}\right) \end{aligned}$$

Proof: By the triangular inequality

$$d_2(\varepsilon_t, \varepsilon_t^*) \le d_2(F_\varepsilon, F_{T,b}) + d_2(F_{T,b}, \hat{F}_{T,b}) .$$

For the second term, let J be Laplace distributed on A, i.e. $J = j$ with probability $\frac{1}{|A|}$ for each $j \in A$. We consider the random variables ε_J and $\tilde{\varepsilon}_J$ which have marginals F_T or $\hat{F}_T$, respectively. Let χ be a random variable with density H_b. Then $\varepsilon_J + \chi$

and $\tilde{\varepsilon}_J + \chi$ have marginals $F_{T,b}$ or $\hat{F}_{T,b}$, respectively. Therefore,

$$\begin{aligned}
d_2^2(F_{T,b}, \hat{F}_{T,b}) &\leq \mathcal{E}^*(\varepsilon_J + \chi - \tilde{\varepsilon}_J - \chi)^2 \\
&= \frac{1}{|A|}\sum_{j\in A}(\hat{\varepsilon}_j - \varepsilon_j - \frac{1}{|A|}\sum_{i\in A}\hat{\varepsilon}_i)^2 \\
&\leq \frac{2}{|A|}\sum_{j\in A}(\hat{\varepsilon}_j - \varepsilon_j)^2 - 2\left(\frac{1}{|A|}\sum_{j\in A}(\hat{\varepsilon}_j - \varepsilon_j)\right)^2 + 2\left(\frac{1}{|A|}\sum_{i\in A}\varepsilon_i\right)^2 \\
&\leq \frac{2}{|A|}\sum_{j\in A}(\hat{\varepsilon}_j - \varepsilon_j)^2 + \frac{2}{|A|^2}\left(\sum_{i\in A}\varepsilon_i\right)^2 \\
&= \frac{2}{|A|}\sum_{j\in A}\left(\frac{X_j - \hat{m}_g(X_{j-1})}{\hat{\sigma}_g(X_{j-1})} - \frac{X_j - m(X_{j-1})}{\sigma(X_{j-1})}\right)^2 + \frac{2}{|A|^2}\left(\sum_{j\in A}\varepsilon_i\right)^2 .
\end{aligned}$$

The first term on the right-hand side is bounded by

$$4\sup_{|x|\leq\gamma_T}\frac{(m(x)-\hat{m}_g(x))^2}{\hat{\sigma}_g^2(x)} + 4\,\frac{1}{|A|}\sum_{j\in A}|\varepsilon_j|^2\sup_{|x|\leq\gamma_T}\frac{(\sigma(x)-\hat{\sigma}_g(x))^2}{\hat{\sigma}_g^2(x)}$$

as $|X_{j-1}| \leq \gamma_T$ for all $j \in A$. By the definition of A and by the law of large numbers for stationary processes we have for a suitable constant $\delta > 0$ that $\mathrm{P}(|A| \geq \delta T) \longrightarrow 1$ for $T \to \infty$, i.e. $|A|$ grows at the same rate as T. Therefore,

$$\frac{1}{|A|}\sum_{j\in A}|\varepsilon_j| \underset{\text{a.s.}}{\longrightarrow} \mathcal{E}|\varepsilon_t|\,,\quad \frac{1}{|A|^2}\left(\sum_{j\in A}\varepsilon_j\right)^2 = \mathcal{O}_P\left(\frac{1}{T}\right).$$

which implies the rest of the assertion. ∎

Corollary 6.1: *Under the assumptions of Lemma 5.1*

$$d_2(\varepsilon_t, \varepsilon_t^*) \longrightarrow 0 \qquad \textit{if } b \to 0 \textit{ for } T \to \infty\,.$$

Proof: By Lemma 5.2 and 5.4 the second and third term in Proposition 6.1 vanish for $T \to \infty$. For the first term, we have

$$d_2(F_\varepsilon, F_{T,b}) \leq d_2(F_\varepsilon, F_T) + d_2(F_T, F_{T,b})\;.$$

As $|A| \to \infty$ for $T \to \infty$, the first term converges to 0 by Lemma 8.4 in Bickel and Freedman (1981). Let ε_J, χ be as in the proof of Proposition 6.1. Then,

$$d_2^2(F_T, F_{T,b}) \leq \mathcal{E}^*(\varepsilon_J - \varepsilon_J - \chi)^2 = \mathcal{E}^*\chi^2 = \mathcal{O}(b^2) = o(1).$$

$\blacksquare$

As next step, we show that X_t and X_t^* are not too far apart in a distributional sense. For this purpose, we consider samples of conditionally independent error variables $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T$ with the following properties: (i) $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T$ are condtionally i.i.d. [given the original data $X_0, \ldots, X_T$]. (ii) $\tilde{\varepsilon}_t$ has a conditional distribution [given the original data $X_0, \ldots, X_T$] which is identical with the unconditional distribution of $\{\varepsilon_t\}$ and (iii) $\mathcal{E}^*(\tilde{\varepsilon}_t - \varepsilon_t^*)^2 = d_2^2(\tilde{\varepsilon}_t, \varepsilon_t^*) = [d_2^2(\varepsilon_t, \varepsilon_t^*)]$. We define now a process $\tilde{X}_t$ by $\tilde{X}_0 = X_0$ and

$$\tilde{X}_t = m(\tilde{X}_{t-1}) + \sigma(\tilde{X}_{t-1})\tilde{\varepsilon}_t.$$

The process $\{\tilde{X}_t\}$ starts with the same value $\tilde{X}_0 = X_0$ at time 0 as $\{X_t\}$ and we suppose that the bootstrap process $\{X_t^*\}$ starts also with $X_0^* = X_0$. We will show that $|\tilde{X}_t - X_t^*| \to 0$ in mean for $T \to \infty$ [given the original data $X_0, \ldots, X_T$]. We use the common initial value for all 3 processes for convenience only. Note that for $T \to \infty$ the influence of this initial value vanishes exponentially fast. First, we show that under our assumptions $|X_t^*| \leq \gamma_T$ with high probability.

Lemma 6.1: *Under the assumptions of Lemma 5.1 and assuming $b \to 0$ for $T \to \infty$,*

$$P^*\left(|X_t^*| \geq \gamma_T\right) \to 0 \quad \textit{in probability.}$$

Proof: As a first step, we show that $\mathcal{E}^*|X_t^*|$ is bounded. By the definition of $\tilde{m}_g, \tilde{\sigma}_g$ for $|x| > \gamma_T$ we have

$$\begin{aligned} |X_{t+1}^*| &= |\varepsilon_{t+1}^*| \cdot 1_{\{|X_t^*| > \gamma_T\}} + |\hat{m}_g(X_t^*) + \hat{\sigma}_g(X_t^*) \cdot \varepsilon_{t+1}^*| \cdot 1_{\{|X_t^*| \leq \gamma_T\}} \\ &\leq |\varepsilon_{t+1}^*| + |m(X_t^*) + \sigma(X_t^*) \cdot \varepsilon_{t+1}^*| + o_P(1) \end{aligned}$$

using Lemma 5.1, 5.3 and the boundedness of $\mathcal{E}^*|\varepsilon_s^*|$ which follows from Corollary 6.1. Let $L < 1$ be an upper bound for $L_m + L_\sigma \mathcal{E}^*|\varepsilon_s^*|$ and C be an upper bound for $\mathcal{E}^*|\varepsilon_s^*| + |m(0)| + \sigma(0) \cdot \mathcal{E}^*|\varepsilon_s^*|$. Then, by Lipschitz continuity of m and σ, we have, iterating with respect to t,

$$\begin{aligned} \mathcal{E}^*|X_{t+1}^*| &\leq L \cdot \mathcal{E}^*|X_t^*| + C \leq \ldots \leq L^{t+1}\mathcal{E}^*|X_0^*| + \sum_{k=0}^{t} L^k \cdot C \\ &\leq \mathcal{E}|X_0| + \frac{C}{1-L}. \end{aligned}$$

Now, we consider $P^*\{X_t^* > \gamma_T\}$ since the arguments for $P^*\{X_t^* < -\gamma_T\}$ are completely similar. We use the abbreviations

$$q_T(x) = (\gamma_T - m(x))/\sigma(x) \quad \text{and} \quad \tilde{q}_T(x) = (\gamma_T - \tilde{m}_g(x))/\tilde{\sigma}_g(x).$$

Remark that $\tilde{q}_T(x) = \gamma_T$ for $|x| > \gamma_T$ and, by Lemma 5.1 and 5.3,

$$\tilde{q}_T(x) = q_T(x) + o_P(1) \text{ uniformly in } |x| \leq \gamma_T.$$

Therefore,

$$\begin{aligned} P^*\{X_t^* > \gamma_T\} &= P^*\{\varepsilon_t^* > \tilde{q}_T(X_{t-1}^*)\} \\ &= P^*\{\varepsilon_t^* > \gamma_T, |X_{t-1}^*| > \gamma_T\} \\ &\quad + P^*\{\varepsilon_t^* > q_T(X_{t-1}^*) + o_P(1), \ \ |X_{t-1}^*| \le \gamma_T\}. \end{aligned}$$

The first summand is bounded by $P^*\{\varepsilon_t^* > \gamma_T\}$ which converges to 0 for $\gamma_T \to \infty$, using Corollary 6.1. Denoting by p^* the probability density of X_{t-1}^*, the second summand is

$$\begin{aligned} \int_{-\gamma_T}^{\gamma_T} P^*\{\varepsilon_t^* > q_T(x) + o_P(1)\} p^*(x)dx &\le \int_{-\gamma_T}^{\gamma_T} \frac{\mathcal{E}^*|\varepsilon_t^*|}{q_T(x) + o_P(1)} p^*(x)dx \\ = \int_{-\gamma_T}^{\gamma_T} \frac{\mathcal{E}^*|\varepsilon_t^*|}{q_T(x)} p^*(x)dx\{1 + o_P(1)\} &= \int_{-\gamma_T}^{\gamma_T} \frac{\sigma(x)}{\gamma_T - m(x)} p^*(x)dx\{1 + o_P(1)\} \cdot \mathcal{E}^*|\varepsilon_t^*|. \end{aligned}$$

For the inequality, we have used Markov's inequality and the fact that $q_T(x)$ is positive and bounded away from 0 uniformly in $|x| \le \gamma_T$ for all γ_T large enough. Now, by Lipschitz continuity of m and σ, $\gamma_T - m(x) \ge \gamma_T - L_m \cdot |x| - |m(0)| \ge (1 - L_m)\gamma_T - |m(0)|$ for $|x| \le \gamma_T$, $\sigma(x) \le L_\sigma \cdot |x| + \sigma(0)$.
Therefore, for a suitable constant C^*, the last integral is bounded by

$$\frac{C^*}{(1 - L_m)\gamma_T - |m(0)|} \cdot \int_{-\gamma_T}^{\gamma_T} \{L_\sigma |x| + \sigma(0)\} p^*(x)dx \le C^* \frac{L_\sigma \mathcal{E}^*|X_{t-1}^*| + \sigma(0)}{(1 - L_m)\gamma_T - |m(0)|} \longrightarrow 0$$

for $\gamma_T \to \infty$ as, by the first part of the proof, $\mathcal{E}^*|X_{t-1}^*|$ remains bounded. ■

Proposition 6.2: *Assume $b \to 0$ for $T \to \infty$ and suppose that the assumptions of Lemma 5.1 hold. Then*

$$\sup_{1 \le t \le T} \mathcal{E}^* \left| X_t^* - \tilde{X}_t \right| = o_P(1).$$

Proof: Let $\tilde{m}(\cdot) = m(\cdot) 1\{|\cdot| \le \gamma_T\}$ and $\tilde{\sigma}(\cdot) = \sigma(\cdot) 1\{|\cdot| \le \gamma_T\} + 1\{|\cdot| > \gamma_T\}$.

$$\begin{aligned} \mathcal{E}^* \left| X_t^* - \tilde{X}_t \right| &= \mathcal{E}^* \Big| \tilde{m}_g(X_{t-1}^*) - m(\tilde{X}_{t-1}) + (\tilde{\sigma}_g(X_{t-1}^*) - \sigma(\tilde{X}_{t-1}))\, \varepsilon_t^* \\ &\quad + \sigma(\tilde{X}_{t-1})(\varepsilon_t^* - \tilde{\varepsilon}_t) \Big| \\ &\le \mathcal{E}^* |\tilde{m}_g(X_{t-1}^*) - \tilde{m}(X_{t-1}^*)| + \mathcal{E}^* |\tilde{m}(X_{t-1}^*) - m(\tilde{X}_{t-1})| \\ &+ \mathcal{E}^* \Big(|\tilde{\sigma}_g(X_{t-1}^*) - \tilde{\sigma}(X_{t-1}^*)| + |\tilde{\sigma}(X_{t-1}^*) - \sigma(\tilde{X}_{t-1})| \Big) |\varepsilon_t^*| \\ &\quad + \mathcal{E}^* \sigma(\tilde{X}_{t-1}) |\varepsilon_t^* - \tilde{\varepsilon}_t| \ . \end{aligned}$$

As, by Lemma 5.1 and 5.3, $\tilde{m}_g - \tilde{m}$ and $\tilde{\sigma}_g - \tilde{\sigma}$ converge to zero uniformly on $\mathbb{R}$, we have that the first and third term converge to zero in probability. For the second term we have from the definition of $\tilde{m}$ and from Lipschitz continuity of m

$$
\begin{aligned}
& \mathcal{E}^*|\tilde{m}(X^*_{t-1}) - m(\tilde{X}_{t-1})| \\
= \;& \mathcal{E}^*|m(X^*_{t-1}) - m(\tilde{X}_{t-1})| 1\{|X^*_{t-1}| \le \gamma_T\} + \mathcal{E}^*|m(\tilde{X}_{t-1})| 1\{|X^*_{t-1}| > \gamma_T\} \\
\le \;& L_m\, \mathcal{E}^*|X^*_{t-1} - \tilde{X}_{t-1}| + o_P(1), \quad \text{by Lemma 6.1}
\end{aligned}
$$

Exactly along the same lines we obtain

$$
\mathcal{E}^*|\tilde{\sigma}(X^*_{t-1}) - \sigma(\tilde{X}_{t-1})|\,|\varepsilon^*_t| \le \mathcal{E}^*|\varepsilon^*_1|[L_\sigma \mathcal{E}^*|X^*_{t-1} - \tilde{X}_{t-1}| + o_P(1)]\ .
$$

Finally, by Proposition 6.1,

$$
\mathcal{E}^*\,\sigma(\tilde{X}_{t-1})|\varepsilon^*_t - \tilde{\varepsilon}_t| = \mathcal{E}^*\,\sigma(\tilde{X}_1)\,\mathcal{E}^*|\varepsilon^*_1 - \tilde{\varepsilon}_1| = o_P(1)\ .
$$

Thus, we have shown that

$$
\begin{aligned}
\mathcal{E}^*|X^*_t - \tilde{X}_t| &\le (L_m + \mathcal{E}^*|\varepsilon^*_1|L_\sigma)\,\mathcal{E}^*|X^*_{t-1} - \tilde{X}_{t-1}| + o_P(1) \\
&\le \sum_{\nu=0}^{t-1} (L_m + \mathcal{E}^*|\varepsilon^*_1|L_\sigma)^\nu\, o_P(1)\ ,
\end{aligned}
$$

as we have chosen $X^*_0 = X_0$.
The result follows from (A5). Observe that $\mathcal{E}^*|\varepsilon^*_1| \le \mathcal{E}^*|\varepsilon^*_1 - \tilde{\varepsilon}_1| + \mathcal{E}^*|\tilde{\varepsilon}_1| \le o_P(1) + 1$, since $\mathcal{E}^*\tilde{\varepsilon}_1^2 = 1$, to obtain the assertion. ∎

Corollary 6.2: *Under the assumptions of Proposition 6.2, we have*

$$
\mathcal{E}^*\left\{\frac{1}{T}\sum_{t=1}^{T} |\tilde{X}_t - X^*_t|\right\} \to 0 \qquad \textit{for } T \to \infty\,.
$$

Proof: Let $(\tilde{\varepsilon}_t, \varepsilon^*_t),\ t = 1, \ldots, T$, be chosen such that $\mathcal{E}^*(\tilde{\varepsilon}_t - \varepsilon^*_t)^2 = d_2^2(\varepsilon_t, \varepsilon^*_t)$ for all t. Looking at the proof of Proposition 6.2, the $o_P(1)-$term converges to 0 uniformly in $t \le T$. Therefore, the Corollary follows immediately from the Proposition as, in particular, $L = L_m + L_\sigma \mathcal{E}\,|\varepsilon_t| < 1$ by (A5).
∎

Proof of Theorems 1-3: In all three cases we split the terms which have to be investigated into a variance and bias part. For the bootstrap of Theorem 1, e.g., this separation is as follows:

$$
\begin{aligned}
& \sqrt{Th}(\hat{m}_h(x) - m(x)) \\
= \;& \frac{\sqrt{Th}\sum_t K_h(x - X_t)\sigma(X_t)\varepsilon_{t+1}}{\sum_t K_h(x - X_t)} + \frac{\sqrt{Th}\sum_t K_h(x - X_t)(m(X_t) - m(x))}{\sum_t K_h(x - X_t)}
\end{aligned}
$$

and, correspondingly,

$$\sqrt{Th}(\hat{m}_h^*(x) - \tilde{m}_g(x))$$
$$= \frac{\sqrt{Th}\sum_t K_h(x - X_t^*)\tilde{\sigma}_g(X_t^*)\varepsilon_{t+1}^*}{\sum_t K_h(x - X_t^*)} + \frac{\sqrt{Th}\sum_t K_h(x - X_t^*)(\tilde{m}_g(X_t^*) - \tilde{m}_g(x))}{\sum_t K_h(x - X_t^*)}.$$

For the regression like and the wild bootstrap we obtain similar expressions, where X_t^* has to be replaced by the original observations X_t in the appropriate places.

Now, we show that

i) the numerators of the variance parts of the original estimator $\hat{m}_h(x)$ and its three bootstrap versions have the same asymptotic behaviour (formulated precisely as Lemma 6.2),

ii) the rescaled numerators of the bias parts of the original estimator and its three bootstrap versions converge to the same limit (compare Lemma 6.3, where the bias components of the regression-like and the wild bootstrap are identical),

iii) and, for the bootstrap of section 2.1, the denominators of variance and bias parts coincide asymptotically for the original estimator and its bootstrap version (compare Lemma 6.3).

Lemmas 6.2 - 6.4 together prove the assertions of Theorems 1-3 concerning the estimate $\hat{m}_h(x)$. The validity of the three bootstrap procedures for approximating the law of $\hat{\sigma}_h^2(x)$ can be shown in a completely analogous manner.
We deal with the variance parts first.

Lemma 6.2:

(i) Assume (A1) - (A5), (B1)-(B4) and (B8). Then for all $x \in \mathbb{R}$

$$\sqrt{\frac{h}{T}}\sum_t K_h(x - X_t)\sigma(X_t)\varepsilon_{t+1} \Longrightarrow \mathcal{N}(0, \tau^2(x)),$$

where $\tau^2(x) = \sigma^2(x)p(x)\int K^2(v)dv$.

(ii) Assume (A1) - (A5) and (B1) - (B8). Then, for the bootstrap of section 2.1, for all $x \in \mathbb{R}$ in probability

$$\sqrt{\frac{h}{T}}\sum_t K_h(x - X_t^*)\tilde{\sigma}_g(X_t^*)\varepsilon_{t+1}^* \Longrightarrow \mathcal{N}(0, \tau^2(x))$$

(iii) Assume (A1) - (A5) and (RB). Then, for the bootstrap of section 2.2, for all $x \in \mathbb{R}$ in probability

$$\sqrt{\frac{h}{T}} \sum_t K_h(x - X_t)\hat{\sigma}_g(X_t)\varepsilon^*_{t+1} \Longrightarrow \mathcal{N}(0, \tau^2(x))$$

(iv) Assume (A1) - (A5) and (WB). Then, for the wild bootstrap of section 2.3, for all $x \in \mathbb{R}$ in probability

$$\sqrt{\frac{h}{T}} \sum_t K_h(x - X_t)\eta^*_{t+1} \Longrightarrow \mathcal{N}(0, \tau^2(x))$$

Proof: (i) It suffices to verify the assumptions of a version of the central limit theorem for martingale difference arrays (Brown, 1971), namely

$$\frac{h}{T} \sum_t \mathcal{E}\left[K_h^2(x - X_t)\sigma^2(X_t)\varepsilon^2_{t+1}|\mathcal{F}_t\right] \to \sigma^2(x)p(x) \int K^2(v)\, dv \tag{6.25}$$

in probability and for all $\delta > 0$, again in probability,

(6.26)
$$\frac{h}{T} \sum_t \mathcal{E}\left[K_h^2(x - X_t)\sigma^2(X_t)\varepsilon^2_{t+1}\mathbf{1}\{\frac{h}{T}K_h^2(x - X_t)\sigma^2(X_t)\varepsilon^2_{t+1} > \delta\}\Big|\mathcal{F}_t\right] \to 0$$

Here $\mathcal{F}_t = \sigma(X_1, \ldots, X_t) = \sigma(X_1, \varepsilon_2, \ldots, \varepsilon_t)$. Since K and σ are bounded in a neighborhood of x assertion (6.26) can be concluded from

$$\frac{1}{Th} \sum_t \mathcal{E}\left(\varepsilon^2_{t+1}\mathbf{1}\{\varepsilon^2_{t+1} > \delta h T\}\right) = \mathcal{O}\left(\frac{1}{Th^2}\right) = o(1)\ .$$

To see (6.25) consider that, by (A4), $\mathcal{E}\left[\varepsilon^2_{t+1}|\mathcal{F}_t\right] = 1$, and therefore the l.h.s. equals

$$\frac{1}{Th} \sum_t \left\{K^2\left(\frac{x - X_t}{h}\right)\sigma^2(X_t) - \mathcal{E}\left[K^2\left(\frac{x - X_t}{h}\right)\sigma^2(X_t)\Big|\mathcal{F}_{t-1}\right]\right\}\ +$$
$$+\frac{1}{Th} \sum_t \int K^2\left(\frac{x - m(X_{t-1}) - \sigma(X_{t-1})u}{h}\right)\sigma^2\left(m(X_{t-1}) + \sigma(X_{t-1})u\right)p_\varepsilon(u)\, du\ .$$

It is easy to see, that the expectation of the square of the first summand is $\mathcal{O}(1/(Th^2)) = o(1)$. The second summand equals, using the symmetry of K,

$$\frac{1}{T} \sum_t \int K^2(v)\sigma^2(x + hv)p_\varepsilon\left(\frac{x - m(X_{t-1})}{\sigma(X_{t-1})} + \frac{hv}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\, dv\ .$$

Differentiability of σ and p_ε and boundedness of the derivatives of p_ε and σ^{-1} imply that this term is equal to

$$\int K^2(v)dv \cdot \sigma^2(x) \cdot \frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})} + \mathcal{O}(h)$$
$$\longrightarrow \int K^2(v)dv \cdot \sigma^2(x) \cdot \mathcal{E}\ p_\varepsilon\left(\frac{x-m(X_1)}{\sigma(X_1)}\right)\frac{1}{\sigma(X_1)}\ ,\ \ T\to\infty\ ,$$

by the ergodic theorem. Observe that $\sigma(X_1)^{-1}p_\varepsilon\left([\cdot - m(X_1)]/\sigma(X_1)\right)$ is the transition density of the underlying process, i.e. by (2.7)

$$\mathcal{E}\, p_\varepsilon\left(\frac{x-m(X_1)}{\sigma(X_1)}\right)\frac{1}{\sigma(X_1)} = \int p_\varepsilon\left(\frac{x-m(u)}{\sigma(u)}\right)\frac{1}{\sigma(u)}\ d\pi(u) = p(x).$$

In order to verify (ii) we will verify conditions quite similar to (6.25) and (6.26). Since the Lindeberg condition can be obtained quite easily, we focus on ($\mathcal{F}_t^* = \sigma(X_1^*,\ldots,X_t^*)$)

$$\begin{aligned}
&\frac{h}{T}\sum_t \mathcal{E}^*[K_h^2(x-X_t^*)\tilde\sigma_g^2(X_t^*)(\varepsilon_{t+1}^*)^2|\mathcal{F}_t^*] \\
=\ & \frac{\mathcal{E}^*(\varepsilon_1^*)^2}{Th}\sum_t\left(K^2\left(\frac{x-X_t^*}{h}\right)\tilde\sigma_g^2(X_t^*) - \mathcal{E}^*\left[K^2\left(\frac{x-X_t^*}{h}\right)\tilde\sigma_g^2(X_t^*)\Big|\mathcal{F}_{t-1}^*\right]\right) \\
+\ & \frac{\mathcal{E}^*(\varepsilon_1^*)^2}{Th}\sum_t\int K^2\left(\frac{x-\tilde m_g(X_{t-1}^*)-\tilde\sigma_g(X_{t-1}^*)u}{h}\right) \\
& \tilde\sigma_g^2\left(\tilde m_g(X_{t-1}^*)+\tilde\sigma_g(X_{t-1}^*)u\right)\hat f_{T,b}(u)\ du\ .
\end{aligned}$$

Observe that $\mathcal{E}^*(\varepsilon_1^*)^2 \to \mathcal{E}\,\varepsilon_1^2 = 1$, by Corollary 6.1, and that the first summand is of order $\mathcal{O}_P(1/(Th^2))$, by the same arguments as above. It suffices to deal with the integral, which is equal to

$$\frac{1}{T}\sum_t\int_{[-1,1]} K^2(v)\tilde\sigma_g^2(x+hv)\hat f_{T,b}\left(\frac{x-\tilde m_g(X_{t-1}^*)}{\tilde\sigma_g(X_{t-1}^*)}+\frac{hv}{\tilde\sigma_g(X_{t-1}^*)}\right)\frac{1}{\tilde\sigma_g(X_{t-1}^*)}dv.$$

The argument of $\hat f_{T,b}$ is bounded in absolute value by $C\gamma_T$. This will be shown in the proof of Lemma 6.3 below. Therefore, using Lemma 6.5 to replace $\hat f_{T,b}$ by p_ε and Corollary 6.1 for the uniform convergence of $\tilde\sigma_g$ to σ on compact sets, we obtain that the last expression equals.

$$\begin{aligned}
&\frac{1}{T}\sum_t\int K^2(v)\sigma^2(x+hv)p_\varepsilon\left(\frac{x-\tilde m_g(X_{t-1}^*)}{\tilde\sigma_g(X_{t-1}^*)}+\frac{hv}{\tilde\sigma_g(X_{t-1}^*)}\right)\frac{1}{\tilde\sigma_g(X_{t-1}^*)}\ dv + o_p(1) \\
=\ & \int K^2(v)\ dv\cdot\sigma^2(x)\cdot\frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-\tilde m_g(X_{t-1}^*)}{\tilde\sigma_g(X_{t-1}^*)}\right)\frac{1}{\tilde\sigma_g(X_{t-1}^*)} + o_p(1),
\end{aligned}$$

using again the differentiability of σ and p_ε as in proving (i). Finally, we have to verify that this term converges in probability towards $\tau^2(x)$, i.e.

$$\frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-\tilde{m}_g(X^*_{t-1})}{\tilde{\sigma}_g(X^*_{t-1})}\right)\frac{1}{\tilde{\sigma}_g(X^*_{t-1})} \sim \frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}$$

where $\sim$ denotes asymptotic equivalence and where we know already from proving (i) that the r.h.s. converges to $p(x)$. Such a result, which means that the Bootstrap process has in some sense an ergodic behaviour, will be needed at several places later on. We present the arguments in some detail here. The proof can be splitted up into the following steps.

$$\frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-\tilde{m}_g(X^*_{t-1})}{\tilde{\sigma}_g(X^*_{t-1})}\right)\frac{1}{\tilde{\sigma}_g(X^*_{t-1})}\mathbf{1}\{|X^*_{t-1}|>\gamma_T\} = o_P(1)\ . \tag{6.27}$$

$$\frac{1}{T}\sum_t \left\{p_\varepsilon\left(\frac{x-\tilde{m}_g(X^*_{t-1})}{\tilde{\sigma}_g(X^*_{t-1})}\right)\frac{1}{\tilde{\sigma}_g(X^*_{t-1})} \quad - \quad p_\varepsilon\left(\frac{x-m(X^*_{t-1})}{\sigma(X^*_{t-1})}\right)\frac{1}{\sigma(X^*_{t-1})}\right\} \mathbf{1}\{|X^*_{t-1}|\le\gamma_T\} = o_P(1)\ . \tag{6.28}$$

$$\frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-m(X^*_{t-1})}{\sigma(X^*_{t-1})}\right)\frac{1}{\sigma(X^*_{t-1})}\mathbf{1}\{|X^*_{t-1}|>\gamma_T\} = o_P(1)\ . \tag{6.29}$$

$$\frac{1}{T}\sum_t \left|p_\varepsilon\left(\frac{x-m(X^*_{t-1})}{\sigma(X^*_{t-1})}\right)\frac{1}{\sigma(X^*_{t-1})} \quad - \quad p_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\right| = \mathcal{O}\left(\frac{1}{T}\sum_t |X^*_t - X_t|\right)\ . \tag{6.30}$$

With (6.27) - (6.30) the desired result follows from Corollary 6.2. This completes the proof of (ii). To see (6.27) observe that $\tilde{m}_g(X^*_{t-1}) = 0$ and $\tilde{\sigma}_g(X^*_{t-1}) = 1$ on $\{|X^*_{t-1}| > \gamma_T\}$, thus the lefthand side of (6.27) equals $\frac{1}{T}\sum_t p_\varepsilon(x)\,\mathbf{1}\{|X^*_{t-1}|>\gamma_T\}$ which is $o_P(1)$ by Lemma 6.1. Similarly, (6.29) follows from Lemma 6.1 together with boundedness of p_ε and σ^{-1}.
Recall that $\tilde{m}_g = \hat{m}_g$ and $\tilde{\sigma}_g = \hat{\sigma}_g$ on $[-\gamma_T, \gamma_T]$ such that $\tilde{m}_g$ and $\tilde{\sigma}_g$ converge uniformly on this growing interval by Lemma 5.1 and Lemma 5.3. Together with boundedness of p_ε and p'_ε, this implies (6.28).

By boundedness of p'_ε and $\frac{1}{\sigma}$ and by the last part of assumption (B3), the derivative of $\frac{1}{\sigma(\cdot)}p_\varepsilon\left(\frac{x-m(\cdot)}{\sigma(\cdot)}\right)$ is bounded. This implies (6.30).

For a proof of (iii) and (iv) we make use of a central limit theorem for triangular arrays of independent observations. The Lindeberg condition can be verified in both cases by routine arguments and also for the variance the argument is quite simple. In case (iii) we have

$$
\begin{aligned}
& \mathcal{E}^*\left(\sqrt{\frac{h}{T}}\sum_t K_h(x-X_t)\hat{\sigma}_g(X_t)\varepsilon_{t+1}^*\right)^2 \\
= \;& \frac{1}{Th}\sum_t K^2\left(\frac{x-X_t}{h}\right)\left(\sigma^2(X_t)+o_P(1)\right)\;\mathcal{E}^*(\varepsilon_1^*)^2 \\
\to \;& \int K^2(v)\,dv\cdot p(x)\cdot\sigma^2(x)\;,\;\text{as } T\to\infty\;,\;\text{in probability}
\end{aligned}
$$

as $\mathcal{E}^*(\varepsilon_1^*)^2 \longrightarrow \mathcal{E}\,\varepsilon_1^2 = 1$ and the remaining term is a kernel estimate with kernel K^2. For replacing $\hat{\sigma}_g^2(X_t)$ by $\sigma^2(X_t)+o_P(1)$ we have used the fact that, by (B8), the sum extends only over those t with $|X_t - x|\le h$ and that $\hat{\sigma}_g^2(z)\longrightarrow\sigma^2(z)$ uniformly on $[x-h, x+h]$. The latter can be easily shown quite similar to Lemma 5.5.
For case (iv) we similarily obtain, using in particular $\mathcal{E}\,\varepsilon_t^2 = 1$,

$$
\begin{aligned}
& \mathcal{E}^*\left(\sqrt{\frac{h}{T}}\sum_t K_h(x-X_t)\eta_{t+1}^*\right)^2 \\
= \;& \frac{1}{Th}\sum_t K^2\left(\frac{x-X_t}{h}\right)\left(X_{t+1}-\hat{m}_g(X_t)\right)^2 \\
= \;& \frac{1}{Th}\sum_t K^2\left(\frac{x-X_t}{h}\right)\left(X_{t+1}-m(X_t)+o_P(1)\right)^2 \\
= \;& \frac{1}{Th}\sum_t K^2\left(\frac{x-X_t}{h}\right)\sigma^2(X_t)\varepsilon_{t+1}^2+o_P(1) \\
= \;& \frac{1}{Th}\sum_t K^2\left(\frac{x-X_t}{h}\right)\sigma^2(X_t)+\frac{1}{Th}\sum_t K^2\left(\frac{x-X_t}{h}\right)\sigma^2(X_t)(\varepsilon_{t+1}^2-1)+o_P(1) \\
\to \;& \sigma^2(x)p(x)\int K^2(v)\,dv\;,\;\text{as } T\to\infty\;.
\end{aligned}
$$

as the expected square of the second term of the previous line is of order $\mathcal{O}_P(\frac{1}{Th^2})$ due to boundedness of K and the compactness of its support. ■

In a next step we have to deal with the kernel estimates for the stationary density.

Lemma 6.3:

(i) Assume (A1) - (A5), (B1), (B3), (B8), and let $h \to 0$ such that $Th \to \infty$ for $T \to \infty$. . Then for all $x \in \mathbb{R}$

$$\frac{1}{T}\sum_t K_h(x - X_t) \to p(x) \ \text{ in probability}$$

(ii) Assume (A1) - (A5) and (B1) - (B8). Then for all $x \in \mathbb{R}$

$$\frac{1}{T}\sum_t K_h(x - X_t^*) \to p(x) \ \text{ in probability}$$

for the bootstrap process (X_t^*) of section 2.1.

Proof: Conclude (i) from

$$\mathcal{E}\left(\frac{1}{T}\sum_t \{K_h(x - X_t) - \mathcal{E}\left[K_h(x - X_t)|\mathcal{F}_{t-1}\right]\}\right)^2 = \mathcal{O}(\frac{1}{Th}) = o(1) \tag{6.31}$$

and

$$\frac{1}{T}\sum_t \mathcal{E}\left[K_h(x - X_t)|\mathcal{F}_{t-1}\right] \to p(x) \ \text{ in probability } \ . \tag{6.32}$$

(6.31) follows by direct computation, while (6.32) is a consequence of the ergodicity of the Markov chain (X_t), namely by (2.7)

$$\begin{aligned}
&\frac{1}{T}\sum_t \mathcal{E}\left[K_h(x - X_t)|\mathcal{F}_{t-1}\right] \\
&= \frac{1}{T}\sum_t \int K\left(\frac{x - m(X_{t-1}) - \sigma(X_{t-1})u}{h}\right) p_\varepsilon(u)\ du \\
&= \frac{1}{T}\sum_t \int K(v) p_\varepsilon\left(\frac{x - m(X_{t-1})}{\sigma(X_{t-1})} - \frac{vh}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\ dv \\
&= \frac{1}{T}\sum_t p_\varepsilon\left(\frac{x - m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})} + \mathcal{O}_P(h)\ ,
\end{aligned}$$

since p_ε is continuously differentiable with bounded derivative and $\int K(v)dv = 1$. The ergodic theorem gives us the result because of (2.7).

In order to prove (ii) observe that

$$\mathcal{E}^*\left(\frac{1}{T}\sum_t \{K_h(x - X_t^*) - \mathcal{E}^*[K_h(x - X_t^*)|\mathcal{F}_{t-1}^*]\}\right)^2 = \mathcal{O}_P(\frac{1}{Th}) = o_P(1)$$

and

$$\mathcal{E}^*[K_h(x-X_t^*)|\mathcal{F}_{t-1}^*] = \int_{[-1,1]} K(v)\hat{f}_{T,b}\left(\frac{x-\tilde{m}_g(X_{t-1}^*)-vh}{\tilde{\sigma}_g(X_{t-1}^*)}\right)\frac{1}{\tilde{\sigma}_g(X_{t-1}^*)}\,dv$$

where $\hat{f}_{T,b}$ denotes the density of the bootstrap residuals, c.f. section 2.1. The argument of $\hat{f}_{T,b}$ is bounded in absolut value by

$$\frac{|x| + \sup_{x\in\mathbb{R}}|\tilde{m}_g(x)-\tilde{m}(x)| + \sup_{x\in\mathbb{R}}|\tilde{m}(x)| + h}{\inf_{x\in\mathbb{R}}\tilde{\sigma}(x) - \sup_{x\in\mathbb{R}}|\tilde{\sigma}_g(x)-\tilde{\sigma}(x)|} \le C\cdot\gamma_T\ ,$$

using Lemma 5.1 and Lemma 5.3. By Lemma 6.5 below, $\hat{f}_{T,b}$ converges uniformly on $[-C\gamma_T, C\gamma_T]$ towards p_ε. Thus, it suffices to consider

$$\begin{aligned}
&\frac{1}{T}\sum_t \int K(v)\ p_\varepsilon\left(\frac{x-\tilde{m}_g(X_{t-1}^*)-vh}{\tilde{\sigma}_g(X_{t-1}^*)}\right)\frac{1}{\tilde{\sigma}_g(X_{t-1}^*)}\,dv \\
=\ &\mathcal{O}_P(h) + \int K(v)\ dv\ \frac{1}{T}\sum_t p_\varepsilon\left(\frac{x-\tilde{m}_g(X_{t-1}^*)}{\tilde{\sigma}_g(X_{t-1}^*)}\right)\frac{1}{\tilde{\sigma}_g(X_{t-1}^*)}\ .
\end{aligned}$$

We dealt with such an expression already in the proof of Lemma 6.2 (ii). This concludes the proof of Lemma 6.3 . ■

Finally, it remains to deal with the various bias parts.

Lemma 6.4:

(i) *Assume (A1) - (A5), (B2) - (B4) and (B8). Then for all* $x\in\mathbb{R}$

$$\sqrt{\frac{h}{T}}\sum K_h(x-X_t)(m(X_t)-m(x)) \longrightarrow b(x)\ \ \textit{in probability,}$$

where $b(x) = B\cdot\int v^2K(v)\ dv\cdot[p'(x)m'(x)+\frac{1}{2}p(x)m''(x)]$.

(ii) *Assume (A1) - (A5) and (B1) - (B8). Then for all* $x\in\mathbb{R}$

$$\sqrt{\frac{h}{T}}\sum_t K_h(x-X_t^*)(\tilde{m}_g(X_t^*)-\tilde{m}_g(x)) \longrightarrow b(x)\ \ \textit{in probability.}$$

(iii) *Assume (A1) - (A5) and (WB). Then for all* $x\in\mathbb{R}$

$$\sqrt{\frac{h}{T}}\sum_t K_h(x-X_t)(\hat{m}_g(X_t)-\hat{m}_g(x)) \longrightarrow b(x)\ \ \textit{in probability.}$$

Proof: A Taylor expansion for the left hand side of (i) yields

(6.33)
$$\sqrt{\frac{h}{T}}\sum_t K_h(x-X_t)(X_t-x)m'(x)+\frac{1}{2}\sqrt{\frac{h}{T}}\sum_t K_h(x-X_t)(X_t-x)^2m''(\hat{X}_t)$$

where $\hat{X}_t$ denotes a suitable value between x and X_t. For the first summand of (6.33) we have

$$\frac{h}{T}\mathcal{E}\left(\sum_t\{K_h(x-X_t)(X_t-x)-\mathcal{E}\left[K_h(x-X_t)(X_t-x)|\mathcal{F}_{t-1}\right]\}\right)^2=\mathcal{O}(h)$$

and the conditional expectation part is equal to

$$\sqrt{\frac{h^3}{T}}\sum_t\int v\ K(v)p_\varepsilon\left(\frac{x-m(X_{t-1})+vh}{\sigma(X_{t-1})}\right)\frac{1}{\sigma(X_{t-1})}\ dv\ .$$

A Taylor expansion for p_ε and the fact that $\int v\ K(v)\ dv=0$, by symmetry of K, leads to the following expression for the conditional expectation part

$$\begin{aligned} & o_P(1)+\sqrt{Th^5}\int v^2K(v)\ dv\ \frac{1}{T}\sum_t p'_\varepsilon\left(\frac{x-m(X_{t-1})}{\sigma(X_{t-1})}\right)\frac{1}{\sigma^2(X_{t-1})} \\ = &\ \ o_P(1)+B\int v^2K(v)\ dv\int p'_\varepsilon\left(\frac{x-m(u)}{\sigma(u)}\right)\frac{1}{\sigma^2(u)}\ d\pi(u)\ , \end{aligned}$$

by the ergodic theorem. Boundedness of p''_ε yields that

$$\int p'_\varepsilon(\frac{x-m(u)}{\sigma(u)})\frac{1}{\sigma^2(u)}d\pi(u)=\frac{d}{dx}\int p_\varepsilon(\frac{x-m(u)}{\sigma(u)})\frac{1}{\sigma(u)}d\pi(u)=p'(x)\ .$$

The second summand of (6.33) can be dealt with quite similar. It converges towards $\frac{1}{2}B\cdot\int v^2K(v)\ dv\cdot p(x)m''(x)$, but we omit the details.
Let us proceed with a proof of (ii). Differentiability of $\tilde{m}_g$ around x, a similar decomposition as in (6.33), the facts that $\tilde{m}'_g$ and $\tilde{m}''_g$ estimate m' and m'' consistently, by Lemma 5.5, and conditioning yield that it suffices to show (6.34) and (6.35).

(6.34)
$$\sqrt{\frac{h}{T}}\sum_t\mathcal{E}^*\left[K_h(x-X_t^*)(X_t^*-x)\Big|\mathcal{F}_{t-1}^*\right]\longrightarrow B\ p'(x)\ \int v^2K(v)\ dv$$

(6.35)
$$\sqrt{\frac{h}{T}}\sum_t\mathcal{E}^*\left[K_h(x-X_t^*)(X_t^*-x)^2\Big|\mathcal{F}_{t-1}^*\right]\longrightarrow B\ p(x)\int v^2K(v)\ dv$$

The conditional expectation in (6.34) is equal to

$$\sqrt{\frac{h^3}{T}}\sum_t \int v\ K(v)\hat f_{T,b}\Big(\frac{x-\tilde m_g(X^*_{t-1})+hv}{\tilde\sigma_g(X^*_{t-1})}\Big)\frac{1}{\tilde\sigma_g(X^*_{t-1})}\ dv\ .$$

A Taylor expansion for $\hat f_{T,b}$ together with the fact $\int v\ K(v)\ dv = 0$ yields for a suitable value $\hat Z^*_t$ between $(x-\tilde m_g(X^*_{t-1}))/\tilde\sigma_g(X^*_{t-1})$ and $(x-\tilde m_g(X^*_{t-1})+hv)/\tilde\sigma_g(X^*_{t-1})$ that the above expression equals

$$\sqrt{\frac{h^5}{T}}\sum_t \int v^2\ K(v)\ \hat f'_{T,b}\ (\hat Z^*_t)\frac{1}{\tilde\sigma_g(X^*_{t-1})}\ dv\ .$$

In the proof of Lemma 6.3 we have seen already that $|\hat Z^*_t| \le C\gamma_T$ for a suitable constant $C>0$. Since $\hat f'_{T,b}$ converges uniformly to p'_ε (cf. Lemma 6.5) on this growing interval and since p''_ε is bounded, the term under investigation is equal to

$$\sqrt{Th^5}\int v^2K(v)\ dv\ \frac{1}{T}\sum_t p'_\varepsilon\Big(\frac{x-\tilde m_g(X^*_{t-1})}{\tilde\sigma_g(X^*_{t-1})}\Big)\frac{1}{\tilde\sigma_g(X^*_{t-1})}\ .$$

Similar arguments as in the proof of Lemma 6.2 (ii) (cf. (6.27) - (6.30)) and Corollary 6.2) yield (6.34).
(6.35) can be obtained along the same lines. We omit the details.
Finally, a proof of (iii) can be obtained analogously since we have that $\hat m'_g$ and $\hat m''_g$ estimate m' and m'' consistently, by Lemma 5.5. ■

Lemma 6.5: *Assume (A1) - (A5), (B1), (B3) - (B7). Then for all $c>0$ and $j=0,1$*

$$\sup_{|x|\le c\gamma_T} |\hat f^{(j)}_{T,b}(x) - p^{(j)}_\varepsilon(x)| = o_P(1)\ .$$

Proof: Let $A=\{t;\ |X_{t-1}|\le\gamma_T\}$ as in section 2.1. In a first step we compare $\hat f_{T,b}$ with $f_{T,b}$ defined as

$$f_{T,b}(x) = \frac{1}{|A|}\sum_{t\in A} H_b(\varepsilon_t - x).$$

We have uniformly in $x\in\mathbb{R}$ for some constant $C(j=0,1)$.

$$|\hat f^{(j)}_{T,b}(x) - f^{(j)}_{T,b}(x)| \le \sup_u |H(u)^{(j+1)}|\frac{1}{|A|b^{j+2}}\sum_{t\in A}|\tilde\varepsilon_t - \varepsilon_t|$$

$$= \frac{C}{|A|b^{j+2}}\sum_{t\in A}|\hat\varepsilon_t - \varepsilon_t - \frac{1}{|A|}\sum_{k\in A}\hat\varepsilon_k|$$

$$\le \frac{2C}{b^2}\left\{\sup_{|x|\le\gamma_T}\frac{|m(x)-\hat m_g(x)|}{\hat\sigma_g(x)} + (\mathcal{E}\,|\varepsilon_t| + o_P(1))\sup_{|x|\le\gamma_T}\frac{|\sigma(x)-\hat\sigma_g(x)|}{\hat\sigma_g(x)}\right\} + \mathcal{O}_P(\frac{1}{\sqrt{T}b^2})$$

by the same type of argument as in the proof of Proposition 6.1. Together with Lemma 5.1, 5.3 and (B1) we conclude

$$\sup_{x\in\mathbb{R}} |\hat{f}_{T,b}(x) - f_{T,b}(x)| = o_P\left(\frac{g^{1/6}\gamma_T}{b^2}\right) + \mathcal{O}_P(\frac{1}{\sqrt{T}b^2}) = o_P(1)$$

by our assumptions on b.

In the second step we compare $f_{T,b}(x)$ with its expectation. We divide the interval $[-c\gamma_T, c\gamma_T]$ into subintervals of length $\delta > 0$. The number of this subintervals is of the same magnitude as $2c\gamma_T/\delta$. The supremum over all x belonging to such a subinterval can be bounded using the mean value theorem. If we denote the intersection points by x_i we obtain

$$\begin{aligned}
&\sup_{|x|\le c\gamma_T} \left| \frac{1}{T}\sum_{t\in A} \left\{ H_b^{(j)}(\varepsilon_t - x) - \mathcal{E}\ H_b^{(j)}(\varepsilon_1 - x)\right\}\right| \\
&\le\ \mathcal{O}\left(\frac{\delta}{b^{j+2}}\right) + \max_i \left|\frac{1}{T}\sum_{t\in A}\left\{ H_b^{(j)}(\varepsilon_t - x_i) - \mathcal{E}\ H_b^{(j)}(\varepsilon_1 - x_i)\right\}\right|,
\end{aligned}$$

where $H_b'(u-x) = -\frac{1}{b^2}H'(\frac{u-x}{b})$ denotes the derivative of $H_b(u-x)$ with respect to x.

Choose $\delta = b^{j+2}/\log T$ and obtain for the second summand ($\beta > 0$)

$$\begin{aligned}
&P\{\max_i |\frac{1}{T}\sum_{t\in A}\left\{ H_b^{(j)}(\varepsilon_t - x_i) - \mathcal{E}\ H_b^{(j)}(\varepsilon_1 - x_i)\right\}| > \beta\} \\
&\le\ \sum_i \frac{1}{\beta^2}\frac{1}{T^2}\sum_t \text{var}\ \ H_b^{(j)}(\varepsilon_t - x_i) \le \mathcal{O}\left(\frac{\gamma_T \log T}{T b^{3j+4}}\right) = o(1)\ .
\end{aligned}$$

Finally, since $\mathcal{E}\ H_b(\varepsilon_1 - x) = \int H(v)p_\varepsilon(x+bv)dv$

$$\sup_{x\in\mathbb{R}} |\mathcal{E}\ H_b(\varepsilon_1 - x) - p_\varepsilon(x)| \le \sup_x \left|\int H(v)\left\{p_\varepsilon(x+bv) - p_\varepsilon(x)\right\}\, dv\right| = \mathcal{O}(b)\ .$$

using (B3). As H is a probability density we have, using (B7), $\int H'(v)dv = 0$ and $\int v\, H'(v)dv = -1$. Since

$$\mathcal{E}\ H_b'(\varepsilon_1 - v) = -\int \frac{1}{b}H'(v)\{p_\varepsilon(x+bv) - p_\varepsilon(x)\}dv,$$

and since p_ε'' is bounded, we have

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} |\mathcal{E}\ H_b'(\varepsilon_1 - x) - p_\varepsilon'(x)| \\
&\le\ \sup_x \left|\int v\, K'(v)\,\{\frac{p_\varepsilon(x+bv) - p_\varepsilon(x)}{bv} - p_\varepsilon'(x)\}dv\right| = \mathcal{O}(b)\ ,
\end{aligned}$$

∎

References

[1] Ango Nze, P. and Portier, D. (1994). Estimation of the density and of the regression functions of an absolutely regular stationary process. Publ. Inst. Statist. Paris 38, 59-87.

[2] Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. Biometrika **77**, 669-687.

[3] Bickel, P.J. and Freedman, D.A. (1981). Some Asymptotic Theory for the Bootstrap. Ann. Statist. **9**, 1196-1217.

[4] Brown, B.M. (1971). Martingale Central Limit Theorems. Ann. Math. Statist. **42**, 54-66.

[5] Collomb, G. and Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. Stochastic Processes and Their Applications, **23**, 77-89.

[6] Delecroix, M. (1987). Sur l'estimation et la prévision non-paramétrique des processus ergodiques. Thesis at Univ. of Lille Flandres Artois, Lille, France.

[7] Dieboldt, J. and Guegan, D. (1990). Probabilistic properties of the general nonlinear Markovian process of order one and applications to time series modelling. Rapport Technique de L.S.T.A. 125, Université Paris XIII.

[8] Doukhan, P. and Ghindes, M. (1980). Etude du processus $X_{n+1} = f(X_n)\varepsilon_n$. C.R. Acad. Sci. Paris, **A290**, 921-923.

[9] Doukhan, P. (1995). Mixing: Properties and Examples. Springer Lecture Notes in Statistics 85, Heidelberg.

[10] Fan, J. (1992). Design-adaptive nonparametric regression. J. Amer. Statist. Assoc. **87**, 998-1004.

[11] Fan, J, (1993). Local linear regression smoothers and their minimax efficiencies. Ann. Statist. **21**, 196-216.

[12] Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. Ann. Statist. **20**, 2008-2036.

[13] Fan, J. and Gijbels, I. (1995). Local Polynomial Modelling and its Applications – Theory and Methodologies. Chapman and hall, New York.

[14] Franke, J. and Wendel, M. (1992). A bootstrap approach for nonlinear autoregressions - some preliminary results. In: Jäckel, K.-H., Rothe, G. and Sendler, W. eds.: Bootstrapping and Related Techniques. Lecture Notes in Economics and Mathematical Systems 376, Springer, Berlin-Heidelberg-New York.

[15] Gourieroux, C. and Montfort, A. (1992). Qualitative threshold ARCH models. J. Econometrics **52**, 159-199.

[16] Härdle, W. and Mammen, E. (1994). Comparing nonparametric versus parametric regression fits. Ann. Statist. **21**, 1926-1947.

[17] Härdle, W. and Tsybakov, A.B. (1995). Local polynomial estimators of the volatility function in nonparametric autoregression. Discussion Paper 42, SFB 373, Berlin.

[18] Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series. Journal of Time Series Analysis, **13**, 209-232.

[19] Hall, P. and Heyde, C.C (1980). Martingale Limit Theory and its Application. Academic Press, New York.

[20] Hjellvik, V. and Tjøstheim, D. (1993). Nonparametric Tests of Linearity for Time Series. Preprint.

[21] Kreiss, J.-P. and Franke, J. (1992). Bootstrapping stationary autoregressive-moving-average models. Journal of Time Series Analysis, **13**, 297-317.

[22] Kreutzberger, E. (1993). Bootstrap für nichtlineare AR(1)-Prozesse. Thesis, Univ. of Kaiserslautern, Germany.

[23] Mammen, E. (1992). When does the Bootstrap work? Asymptotic results and simulations. Springer Lecture Notes in Statistics 77, Springer, Heidelberg, Berlin.

[24] Masry, E. and Tjøstheim, D. (1994). Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong Convergence and Asymptotic Normality. Econometric Theory **11**, 258-289.

[25] Meyn, S.P. and Tweedie, R.L. (1992). Stability of Markovian Processes I: Criteria for Discrete-Time Chains. Adv. Appl. Prob. **24**, 542-574.

[26] Meyn, S.P. and Tweedie, R.L. (1993). Markov Chains and Stochastic Stability. Springer-Verlag, New York.

[27] Priestley, M.B. (1988). Nonlinear and Nonstationary Time Series Analysis. Academic Press, London.

[28] Robinson, P.M. (1983). Nonparametric Estimators for Time Series. Journal of Time Series Analysis 4, 185-207.

[29] Stone, C.J. (1977). Consistent nonparametric regression. Ann. Statist. **5**, 595-620.

[30] Tjøstheim, D. (1994). Non-linear Time Series: A Selective Review, Scand. J. Statist. **21**, 97-130.

[31] Tjøstheim, D. and Auestad, B. (1994). Nonparametric identification of nonlinear time series: projections. J. Amer. Statist. Assoc. **89**, 1398-1409.

[32] Tong, H. (1990). Non-linear Time Series. A Dynamic System Approach. Oxford University Press, Oxford.

[33] Tsybakov, A. B. (1986). Robust reconstruction of functions by the local approximation method. Prob. Inform. Transmission **22**, 133-146.

[34] Yao, Q. and Tong, H. (1995). A Bandwidth Selector and a Test for Operational Determinism. Preprint.

Acknowledgement: Financial support was received by the Deutsche Forschungsgemeinschaft, SFB 373 ("Quantifikation und Simulation Ökonomischer Prozesse"), Humboldt-Universität zu Berlin.