

DESIGN-ADAPTIVE POINTWISE NONPARAMETRIC REGRESSION ESTIMATION FOR RECURRENT MARKOV TIME SERIES

BY EMMANUEL GUERRE

LSTA-Université Paris 6 and CREST

Abstract: A general framework is proposed for (auto)regression nonparametric estimation of recurrent time series in a class of Hilbert Markov processes with a Lipschitz conditional mean. This includes various nonstationarities by relaxing usual dependence assumptions as mixing or ergodicity, which are replaced with recurrence. The cornerstone of design-adaptation is a data-driven bandwidth choice based on an empirical bias variance tradeoff, giving rise to a random consistency rate for a uniform kernel estimator. The estimator converges with this random rate, which is the optimal minimax random rate over the considered class of recurrent time series. Extensions to general kernel estimators are investigated. For weak dependent time-series, the order of the random rate coincides with the deterministic minimax rate previously derived. New deterministic estimation rates are obtained for modified Box-Cox transformations of Random Walks.

Résumé : On propose un cadre général pour l'estimation nonparamétrique ponctuelle de la fonction d'autorégression dans une classe de processus de Markov récurrent à valeurs Hilbert, et de moyenne conditionnelle Lipschitz. Ce nouveau cadre incorpore de nombreux type de nonstationnarité, en relaxant les conditions usuelles de mélangeance ou d'ergodicité au profit d'une hypothèse de récurrence. La clé de l'approche proposée est un équilibre biais-variance empirique qui permet de choisir une fenêtre optimale aléatoire, donnant lieu à une vitesse de convergence aléatoire pour un estimateur à noyau uniforme. Cette vitesse s'adapte automatiquement à chacun des processus considérés, est optimale dans un cadre minimax, et est aussi atteinte par des estimateurs à noyau plus généraux. Dans le cas de processus faiblement dépendants, l'ordre de la vitesse aléatoire coïncide avec les taux minimax connus. De nouvelles vitesses sont obtenues pour une famille de transformations Box-Cox modifiées appliquées à la marche aléatoire.

September 2004

The author acknowledges stimulating discussions with Delphine Blanke, Sylvain Delattre and Marc Hoffmann.

All errors remain mine.

AMS 2000 subject classifications. Primary 62M10 ; Secondary 62G08.

Key words and phrases. Nonparametric regression estimation, Recurrent time series, Design-adaptation, Optimal random estimation rate.

1. Introduction Since Roussas (1969), the nonparametric literature for dependent data has considerably grown, due in particular to an increasing interest on nonlinear modelling. See Tjøstheim (1994) for a review of such subtle interplays. A vast majority of work deals with the mixing stationary framework, see Bosq (1998), Fan and Yao (2003), or Györfi, Härdle, Sarda and Vieu (1989) and the references therein. Some, as Delecroix (1987), Morvai, Yakowitz and Györfi (1996), Yakowitz, Györfi, Kieffer and Morvai (1999) among others have pushed the limits to stationary ergodic time series. The recurrence properties of ergodic processes, as formalized by the Law of Large Numbers, provide indeed an intuitively appealing sufficient condition for consistency of local smoothers. However, as argued in Karlsen and Tjøstheim (2001) who impuled new directions for nonparametric analysis of time series, this is too restrictive for active research areas as long range dependence or unit root processes, see Robinson (1997) and Phillips and Park (1998) for nonparametric approaches. The ergodic assumption imposes that the number of visits to the estimation domain, over which nonparametric regression is performed, must remain proportional to the sample size, a condition that does not hold for many nonstationary models of interest. An alternative is to weaken ergodicity by assuming that the process is recurrent over the estimation domain. An important difficulty is that it becomes practically relevant to view recurrence as an unknown characteristic of the observations. As a matter of fact, articles investigating nonparametric estimation under recurrence remain exceptions. The retained framework deals with Harris-recurrent Markov processes, as in Yakowitz (1993) for a sequential nearest-neighbor regression estimator. Karlsen and Tjøstheim (2001) established consistency and asymptotic normality of kernel estimators for β -null recurrent time series. Closer to our approach but in the context of diffusion models is Delattre, Hoffmann and Kessler (2002) who considered adaptation to the unknown recurrence rate. See also Blanke (2004) for the potentially related issue of adapting to sample path smoothness.

In the present paper, nonparametric pointwise estimation of the Lipschitz conditional mean given the past of a recurrent Hilbert-valued Markov time series is considered. The point of view developed here differs from Karlsen and Tjøstheim (2001) whose approach leads in particular to extend standard bias variance analysis to β -null recurrent processes. However, their results hold for bandwidths restricted with the recurrence rate of the observations, which should be partially known. By contrast, our design-adaptive approach replaces bandwidth choices based on asymptotic expansion of the mean squared error with an empirical bias variance tradeoff which does not require such *a priori* information. The expression “design-adaptation” was coined out in Fan (1992) for local polynomial estimators which converge under weak conditions on the i.i.d. covariate distribution. Such an approach was extended in Guerre (2000) to cover arbitrary designs. Guerre (2000) considered, to capture the unknown recurrence features of the covariates, data-driven bandwidths and random rates as standardization of nonparametric

estimators. In the context of chaotic data, Guerre and Maës (1998) built on a data-driven standardization, see also Berlinet and Biau (2001). The dynamic-adaptive approach of Delattre et al. (2002) is crucially based on data-driven bandwidth and random standardization, see also Spokoiny (2000) for smoothness adaptation under ergodic paths. The interest of data-driven standardization has also been acknowledged in the context of smoothness adaptation, see Hoffmann and Lepski (2002), as well as in the probability literature, see Chen (1999a) for the Law of Iterated Logarithm under Harris-recurrence and the discussion therein.

More specifically, the distinctive merits of the design-adaptive approach for recurrent time series are as follows. First, a general framework is proposed for Markov processes valued in Hilbert spaces which avoid mixing conditions, allowing so for a wide range of nonstationarities. The focus is set on recurrence over the estimation domain, with a recurrence rate which is considered as an unknown characteristic of the observations. Second, an empirical bias variance tradeoff for a uniform kernel estimator allows to propose a baseline data-driven design-adaptive bandwidth, which gives rise to a random consistency rate. The resulting regression estimator converges with this rate and automatically adapts to a wide range of local recurrence behaviors. Extensions to a more general radial kernel estimators is also investigated. Third, the random rate derived from the design-adaptive bandwidth is shown to be optimal in a minimax sense. In case of weak dependent time series, the deterministic exact order of the random rate coincides with the minimax optimal order previously derived. How design-adaptation deals with nonstationary processes is illustrated here with the example of a modified Box-Cox transformation of the Gaussian Random Walk.

The rest of the paper is divided in three sections and two appendices. Section 2 presents the design-adaptive approach, our baseline nonparametric regression estimator as well as potential improvements, with a non technical overview of our main results. Section 3 groups our assumptions and states the main results. Section 4 gives the deterministic order of the optimal data-driven rate for some examples of time series. Proofs of our main results (i.e. Theorems 1 to 4 and Corollary 1) are given in Appendix A, while Appendix B gathers proofs of more illustrative results as Propositions 1 and 2.

2. Design-adaptive nonparametric regression estimation: an overview Let $(\mathcal{X}, \|\cdot\|)$ be an Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. Consider $T + 1$ observations X_0, \dots, X_T from a time series valued in \mathcal{X} , with

$$X_t = m(X_{t-1}) + e_t, \quad \mathbb{E}[e_t | X_{t-1}, \dots, X_0] = 0 \quad \text{and} \quad \mathbb{E}[\|e_t\|^2 | X_{t-1}, \dots, X_0] \leq \sigma^2$$

for all $t \geq 1$. Assume that the (auto)regression function $m(\cdot)$ is Lipschitz over a bounded open subset \mathcal{D} of \mathcal{X} , i.e. that $\|m(x) - m(x')\| \leq L\|x - x'\|$ for all x, x' in \mathcal{D} . The purpose is to estimate

$m(x)$ for x in \mathcal{D} , assuming that the number of visits of the X_{t-1} 's to the estimation domain \mathcal{D} ,

$$N_T(\mathcal{D}) = \sum_{t=1}^T \mathbb{I}(X_{t-1} \in \mathcal{D}) ,$$

diverges with the sample size, a condition referred as a Recurrence Condition in what follows. A solution to that issue is to average the X_t 's over the X_{t-1} close enough to x , as done by the radial uniform kernel estimate

$$(2.1) \quad m_T(x; b) = \frac{\sum_{t=1}^T X_t \mathbb{I}(\|X_{t-1} - x\| \leq b)}{N_T(b)} \quad \text{where } N_T(b) = \sum_{t=1}^T \mathbb{I}(\|X_{t-1} - x\| \leq b) ,$$

setting $m_T(x; b) = 0$ if $N_T(b) = 0$. Before introducing our choice of the bandwidth b , we present some examples of time series to motivate design-adaptation and the framework of Section 3.

EXAMPLE 1: LINEAR AUTOREGRESSIVE MODELS. A baseline example to motivate our approach is the simple AR (1) model

$$X_t = aX_{t-1} + e_t , \quad |a| \leq 1 ,$$

with i.i.d. e_t such that $\mathbb{E}e_t = 0$ and $\text{Var}(e_t) = \sigma^2$. The dependence structure of $\{X_t, t \geq 0\}$ drastically differs following $|a| < 1$ or $a = 1$. In the former, the process is asymptotically stationary with an invariant probability and the order of $N_T(\mathcal{D})$ is T by the Law of Large Numbers. This model fits the framework of nonparametric estimation for dependent variables. On the other hand, the AR(1) model is nonstationary with $X_t = X_0 + \sum_{i=1}^t e_i$ when $a = 1$, the Lebesgue measure is invariant, and $N_T(\mathcal{D})$ has the smaller order \sqrt{T} . This is outside the scope of the vast majority of the nonparametric literature. Since the recurrence features of this model are highly sensitive to the unknown parameter a , it is therefore desirable to design nonparametric methods which can adapt to the time series at hand. Such features are shared by many parametric models as linear or threshold autoregressive models of higher order models which may generate more sophisticated seasonal nonstationarities, see Tong (1990).

EXAMPLE 2: NONLINEAR AUTOREGRESSIVE MODELS. More complex recurrent time series include nonlinear Markov processes of order d . Consider an univariate d -Markov time series $\{Y_t, t \geq 0\}$. Let $\{\varepsilon_t, t \geq 1\}$ be the associated innovation process $\varepsilon_t = Y_t - \mathbb{E}[Y_t | Y_{t-1}, \dots, Y_{t-d}]$ assuming that $\text{Var}[\varepsilon_t | Y_{t-1}, \dots, Y_{t-d}] \leq \sigma^2$. Under time homogeneity, the process admits an autoregressive representation $Y_t = \mu(Y_{t-1}, \dots, Y_{t-d}) + \varepsilon_t$, where the conditional mean $\mu(Y_{t-1}, \dots, Y_{t-d}) = \mathbb{E}[Y_t | Y_{t-1}, \dots, Y_{t-d}]$ is the parameter of interest. Taking $\mathcal{X} = \mathbb{R}^d$ equipped with the usual Euclidean norm and $X_t = (Y_t, \dots, Y_{t-d+1})'$ yield that $\{X_t, t \geq 0\}$ is Markov with

$$X_t = m(X_{t-1}) + e_t \quad \text{where } m(X_{t-1}) = (\mu(Y_{t-1}, \dots, Y_{t-d}), Y_{t-1}, \dots, Y_{t-d+1})' ,$$

and $e_t = (\varepsilon_t, 0, \dots, 0)'$. Note that $m_T(x; b)$ could be changed into its first entry for such representation since the $d - 1$ last components of $m(\cdot)$ are known. As it can be expected from Example 1, the probability structure of such nonparametric models is extremely complex and developing practical tests for hypothesis as mixing decay conditions seems out of reach in practical applications. A more reasonable strategy is to postulate a weaker Recurrence Condition over \mathcal{D} . Conditions ensuring recurrence of such processes have been investigated, see Meyn and Tweedie (1993) and the references therein, and Proposition 1 below.

EXAMPLE 3: NONLINEAR TRANSFORMATION OF A RECURRENT PROCESS. As mentioned in Karlsen and Tjøstheim (2001), a nonparametric class of recurrent time series can be obtained by transforming a baseline recurrent process $\{Y_t, t \geq 0\}$, as for instance the AR(1) considered in Example 1. Indeed, if $X_t = H(Y_t)$ where $H(\cdot)$ is one to one and $H(\mathcal{D})$ is a subset of \mathcal{D} , then the number of visits of the X_{t-1} 's to \mathcal{D} diverges if the one of the Y_{t-1} 's does. A family of transformations $H(\cdot)$ in the spirit of the Box-Cox transformation will be applied to the Gaussian Random Walk in Section 4 to exemplify the capability of the design-adaptive approach to cope with nonstationary time series given by ill-conditioned transformation $H(\cdot)$.

In the expression (2.1) of the nonparametric estimator $\mu_T(x; b)$, the parameter b is a bandwidth which indicates the closeness of the X_{t-1} 's to x . A crucial issue is the choice of b . For the general class of nonparametric autoregressive models under consideration, usual standard bandwidth choice using asymptotic expansion of the mean squared error of $m_T(x; b)$ are unlikely to apply since the time series at hand can be nonstationary. On the other hand, it is still possible to obtain a random order for the estimation error $m_T(x; b) - m(x)$ which decomposes into the two following terms,

$$m_T(x; b) - m(x) = \frac{\sum_{t=1}^T (m(X_{t-1}) - m(x)) \mathbb{I}(\|X_{t-1} - x\| \leq b)}{N_T(b)} + \frac{\sum_{t=1}^T e_t \mathbb{I}(\|X_{t-1} - x\| \leq b)}{N_T(b)},$$

where the first sum is viewed as a bias term while the second corresponds to a stochastic error. Since $m(\cdot)$ is Lipschitz over \mathcal{D} , the bias term is bounded by Lb for any x in \mathcal{D} provided that b is small enough. Since $\sum_{t=1}^T e_t \mathbb{I}(\|X_{t-1} - x\| \leq b)$ is a sum of martingale differences, we have, for the stochastic error term times $N_T(b)$,

$$\begin{aligned} \mathbb{E} \left\| \sum_{t=1}^T e_t \mathbb{I}(\|X_{t-1} - x\| \leq b) \right\|^2 &= \sum_{t=1}^T \mathbb{E} \left(\mathbb{I}(\|X_{t-1} - x\| \leq b) \mathbb{E}[\|e_t\|^2 | X_{t-1}, \dots, X_0] \right) \\ &+ 2 \sum_{t=1}^T \sum_{t'=t+1}^T \mathbb{E} \left(\mathbb{I}(\|X_{t-1} - x\| \leq b) \mathbb{I}(\|X_{t'-1} - x\| \leq b) \langle e_t, \mathbb{E}[e_{t'} | X_{t'-1}, \dots, X_0] \rangle \right) \end{aligned}$$

$$\leq \sigma^2 \mathbb{E} N_T(b) .$$

It then follows, from the Markov inequality, that the stochastic term times $N_T(b)$ is of order $\sigma N_T^{1/2}(b)$, so that the resulting order of the stochastic term is $\sigma/N_T^{1/2}(b)$. Combining the bounds for the squared bias and stochastic terms gives, for the regression estimator

$$(2.2) \quad m_T(x; b) - m(x) = O_{\mathbb{P}} \left(\left(L^2 b^2 + \frac{\sigma^2}{N_T(b)} \right)^{1/2} \right) .$$

Therefore, $m_T(x; b_T)$ is a consistent estimator of $m(x)$ as soon as the deterministic b_T asymptotically vanishes and $N_T(b_T)$ diverges in probability. However, achieving a divergent $N_T(b_T)$ supposes some *a priori* information on the recurrence of the process $\{X_t, t \geq 0\}$ which may not be available in practice. An alternative to such an *ad hoc* deterministic choice is to let the sample X_0, \dots, X_T suggests a proper bandwidth according to an empirical bias variance tradeoff. As in Guerre (2000), it is technically convenient in a first step to bound the right-hand side of (2.2) with $\max(L^2 b^2, \sigma^2/N_T(b))$ and to propose

$$\widehat{b}_T = \widehat{b}_T(x; L, \sigma) = \arg \min_{b \geq 0} \max \left(L^2 \widehat{b}_T^2, \frac{\sigma^2}{N_T(\widehat{b}_T)} \right) \text{ i.e., for instance,}$$

$$(2.3) \quad \widehat{b}_T = \min \{ b \geq 0; L^2 b^2 N_T(b) \geq \sigma^2 \} \text{ so that } \min_{b \geq 0} \max \left(L^2 \widehat{b}_T^2, \frac{\sigma^2}{N_T(\widehat{b}_T)} \right) = L^2 \widehat{b}_T^2 .$$

Note that $N_T(\widehat{b}_T) > 0$ so that $m_T(x; \widehat{b}_T)$ is defined without ambiguity. That \widehat{b}_T in (2.3) achieves $\min_{b \geq 0} \max(L^2 b^2, \sigma^2/N_T(b))$, equal to $L^2 \widehat{b}_T^2$, is easily seen from a graph, since $\sigma^2/N_T(\cdot)$ is cadlag (right-continuous, left-limit) and decreases from $+\infty$ to σ^2/T , while $L^2 b^2$ continuously increases from 0 to $+\infty$. The bandwidth \widehat{b}_T accounts for the local recurrence properties of the process at x , which is the purpose of design-adaptation, since \widehat{b}_T decreases when the number of X_{t-1} 's in small vicinities of x increases.

For the choice (2.3) of \widehat{b}_T , one would expect that the order for the estimation error (2.2) is

$$(2.4) \quad R_T = R_T(x; L, \sigma) = \frac{1}{\left(\min_{b \geq 0} \max \left(L^2 b^2, \frac{\sigma^2}{N_T(b)} \right) \right)^{1/2}} = \frac{1}{L \widehat{b}_T} .$$

However, the order $m(x; \widehat{b}_T) - m(x) = O_{\mathbb{P}}(1/R_T)$ cannot be directly derived from (2.2) which does not hold for a data-driven bandwidth as \widehat{b}_T . However, empirical processes techniques can be helpful to achieve such a result. This necessitates to restrict to a class of Markov processes as done with Definition 1 in Section 3. Theorem 2 then shows that the estimation error $m(x; \widehat{b}_T) - m(x)$ is of order $1/R_T$ uniformly over the considered class, without mixing or ergodicity conditions.

A pleasant feature is that R_T can be interpreted as a random convergence rate. Since $N_T(b)$ increases with b and T , it is easily seen from (2.4) that R_T increases with the sample size, as any of the power function of T which are usually considered as a normalization in nonparametric

inference. However, choosing such a deterministic standardization necessitates some *a priori* information on the dynamic of the process at hand. For instance, the optimal minimax rate to estimate a Lipschitz conditional mean is $T^{1/3}$ under weak dependence and becomes $T^{1/6}$ for a recurrent Random Walk. By contrast, Theorem 4 shows that R_T is an optimal minimax random rate over a class including these two kind of time series. On the other hand, R_T may have a slower deterministic equivalent than the minimax rates derived for more specific classes of processes, due to an excess of generality. As addressed in Section 4, the exact order of R_T coincides with the minimax rates for weak dependent time series or recurrent Random Walks. This suggests that the design-adaptive approach is not affected by a loss due to its capability to deal with a large class of recurrent time series.

Therefore, potential improvements of the simple estimator $m_T(x, \hat{b}_T)$ are limited to decrease R_T with a constant multiplicative factor. An alternative design-adaptive estimator builds on the more general radial kernel smoother

$$(2.5) \quad \mu_T(x; h) = \mu_T(x; h, K) = \frac{\sum_{t=1}^T X_t K\left(\frac{\|X_{t-1} - x\|}{h}\right)}{\sum_{t=1}^T K\left(\frac{\|X_{t-1} - x\|}{h}\right)}.$$

The kernel $K(\cdot)$ is taken nonnegative, so that $0 \leq K(\cdot) \leq 1$ can be assumed without loss of generality. As $m_T(x; b)$, the estimation error $\mu_T(x; h) - m(x)$ decomposes into a bias a stochastic terms which can be used to propose a design-adaptive h . To achieve a better performance than $m_T(x; \hat{b}_T)$, a more precise bound than Lh for the bias term of $\mu_T(x; h)$ is considered. Define

$$\begin{aligned} B_T(h) &= B_T(h; x, K) = \frac{\sum_{t=1}^T \|X_{t-1} - x\| K\left(\frac{\|X_{t-1} - x\|}{h}\right)}{\sum_{t=1}^T K\left(\frac{\|X_{t-1} - x\|}{h}\right)}, \\ N_T(h; K) &= N_T(h; x, K) = \sum_{t=1}^T K^2\left(\frac{\|X_{t-1} - x\|}{h}\right), \\ V_T(h) &= V_T(h; x, K) = \frac{\left(\sum_{t=1}^T K\left(\frac{\|X_{t-1} - x\|}{h}\right)\right)^2}{\sum_{t=1}^T K^2\left(\frac{\|X_{t-1} - x\|}{h}\right)} = \frac{\left(\sum_{t=1}^T K\left(\frac{\|X_{t-1} - x\|}{h}\right)\right)^2}{N_T(h; K)}, \end{aligned}$$

where $LB_T(h)$ is the improved bias bound while $\sigma/V_T^{1/2}(h)$ gives the order of the stochastic error of $\hat{\mu}_T(x; h)$. A major difference with the bounds Lb and $\sigma/N_T^{1/2}(b)$ used for $m_T(x; b)$ is that $LB_T(h)$ and $\sigma/V_T^{1/2}(h)$ are not necessarily monotone. This complicates the study of a bandwidth achieving the minimum of $((LB_T(h))^2 + \sigma^2/V_T(h))^{1/2}$ which replaces the right-hand side of (2.2). To overcome this technicality, we have introduced the “ K -number of visits” $N_T(h; K)$ which parallels $N_T(b)$, and increases with h if $K(\cdot)$ decreases over \mathbb{R}^+ . The choice of

an optimal bandwidth \hat{h}_T for $\mu_T(x; h)$ is based upon a pilot bandwidth \hat{h}_{0T} as follows:

$$(2.6) \quad \hat{h}_{0T} = \hat{h}_{0T}(x; L, \sigma^2, K) = \min \{h \geq 0; L^2 h^2 N_T(h; K) \geq \sigma^2\} ,$$

$$(2.7) \quad \hat{h}_T = \hat{h}_T(x; L, \sigma, \kappa_0, K) = \arg \min_{h \in \mathcal{H}_T} \left(L^2 B_T^2(h) + \frac{\sigma^2}{V_T(h)} \right) \text{ with, for some } \kappa_0 > 1,$$

$$\mathcal{H}_T = \{h \geq 0; h \leq \kappa_0 \hat{h}_{0T}, N_T(\hat{h}_{0T}; K)/\kappa_0 \leq N_T(h; K), V_T(h) \leq \kappa_0 N_T(\hat{h}_{0T}; K)\} ,$$

$$(2.8) \quad R_T(K) = R_T(x; L, \sigma, \kappa_0, K) = \frac{1}{\left(L^2 B_T^2(\hat{h}_T) + \sigma^2/V_T(\hat{h}_T) \right)^{1/2}} .$$

The pilot bandwidth \hat{h}_{0T} parallels the optimal bandwidth \hat{b}_T of $m_T(x; b)$ with $N_T(b)$ changed into $N_T(h; K)$. However, since $0 \leq K(\cdot) \leq 1$ implies that $N_T(h; K) \leq N_T(h)$, $\mu_T(x; \hat{h}_{0T})$ does not have a better order than $m_T(x; \hat{b}_T)$, while $\mu_T(x; \hat{h}_T)$ can improve on $m_T(x; \hat{b}_T)$ and $\mu_T(x; \hat{h}_{0T})$, up to a limitation due to the constraints of \mathcal{H}_T . The variable $R_T(K)$ is viewed as the random rate of $\mu_T(x; \hat{h}_T)$, see Theorem 3 which parallels Theorem 2.

3. Assumptions and main results

3.1. *Assumptions* We assume that the kernel function $K(\cdot)$ in (2.5) satisfies:

ASSUMPTION K. *The kernel $K(\cdot)$ is continuously decreasing from \mathbb{R}^+ to $[0, 1]$, with compact support $[0, 1]$ and $K(0) = 1$. $K(\cdot)$ has bounded variations and $K(z) \geq \kappa_1 > 0$ for z in $[0, 1/2]$.*

Assumption K ensures in particular that $N_T(h; K)$ is continuously increasing in h . Let us now introduce our assumptions for $\{X_t, t \geq 0\}$. The reader is referred to Meyn and Tweedie (1993) or Nummelin (1984) for a general exposition of the theory of Markov processes, see also Karlsen and Tjøstheim (2001) for an overview covering most of the results needed here. The process $\{X_t, t \geq 0\}$ is Markov of order 1, with values in $(\mathcal{X}, \mathcal{S})$ where the Borel field \mathcal{S} is countably generated. The space \mathcal{X} with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$ is a real Hilbert space with finite or infinite dimension. The system $\{v_j, j \geq 1\}$ generates \mathcal{X} , assuming that the v_j 's with j less than the dimension of \mathcal{X} form an orthonormal basis while the other vanish, i.e. for $\mathcal{X} = \mathbb{R}^d$ as in Example 2, (v_1, \dots, v_d) is the canonical basis and $v_{d+j} = 0$, $j \geq 1$. The estimation set \mathcal{D} is a bounded open subset of \mathcal{X} . The distribution of $\{X_t, t \geq 0\}$ is denoted \mathbb{P} .

The definition of the class $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ of admissible time series involves some additional quantities. Consider a probability measure \underline{P} over \mathcal{X} , with $\underline{P}(\mathcal{D}) = 1$ and

$$(3.1) \quad \underline{F}(h) = \inf_{x \in \mathcal{D}} \underline{P}(\{y \in \mathcal{D} : \|y - x\| \leq h\}) \text{ is continuous,}$$

$$\text{with } \underline{F}(h) > 0 \text{ for } h > 0 \text{ and } \lim_{h \rightarrow \infty} \underline{F}(h) = 1.$$

Recall that $e_t = \sum_{j=1}^{\infty} \langle e_t, v_j \rangle v_j$. The standard deviation σ is identified with a sequence $\{\sigma_j, j \geq 1\}$ of nonnegative numbers with $\sigma_1 > 0$, $\sum_{j=1}^{\infty} \sigma_j^2 = \sigma^2$ and $\sum_{j=1}^{\infty} \sigma_j \leq \kappa_2 \sigma$ for some $\kappa_2 \geq 1$. In addition to $L > 0$ and σ , consider ρ in $[0, 1)$, an integer number $d \geq 1$ as in Example 2, and a positive sequence \underline{n}_T which typically diverges with a slow rate as $\ln T$.

DEFINITION 1. *The class $\mathcal{P}_{\mathcal{D}}(L, \sigma) = \mathcal{P}_{\mathcal{D}}(L, \sigma; \rho, \underline{P}, d, \underline{n}_T, \kappa_2)$ is the family of distributions \mathbb{P} such that*

i. \mathbb{P} is the distribution of a time-homogeneous Markov process $\{X_t, t \geq 0\}$, with

$$(3.2) \quad \mathbb{P}(\|X_t - x\| = h \text{ for some } t \geq 0) = 0 \text{ for all } x \text{ in } \mathcal{D} \text{ and all } h \geq 0,$$

$$(3.3) \quad \mathbb{E}_{\mathbb{P}} N_T(\mathcal{D}) \geq \underline{n}_T \text{ for all } T \geq 0.$$

ii. The regression $m(\cdot) = \mathbb{E}_{\mathbb{P}}[X_{t+1}|X_t = \cdot]$ is L -Lipschitz over \mathcal{D} , i.e. $\|m(x) - m(x')\| \leq L\|x - x'\|$ for all x, x' in \mathcal{D} .

iii. The innovation term $e_t = X_t - m(X_{t-1})$ satisfies $\sup_{x \in \mathcal{D}} \mathbb{E}[\langle e_t, v_j \rangle^2 | X_{t-1} = x] \leq \sigma_j^2$ and $\sup_{x \in \mathcal{D}} \mathbb{E}^{1/8}[\langle e_t, v_j \rangle^{>8} | X_{t-1} = x] \leq \kappa_2 \sigma_j$ for all $j \geq 1$.

iv. \mathbb{P} satisfies the Minorization Condition $\mathbb{P}(X_{t+d} \in A | X_t = x) \geq (1 - \rho^d) \underline{P}(A) \mathbb{I}(x \in \mathcal{D})$ for any Borel subset A of \mathcal{D} .

The following condition on $\underline{P}(\cdot)$ and ρ allows to consider a subclass $\mathcal{P}'_{\mathcal{D}}(L, \sigma) \subset \mathcal{P}_{\mathcal{D}}(L, \sigma)$ of univariate Markov processes with Gaussian innovations.

ASSUMPTION P. *Assume that $\mathcal{X} = \mathbb{R}$ and $d = 1$. Let P_g be the $\mathcal{N}(0, \sigma^2)$ distribution for the innovation term. Then the distribution $\underline{P}(\cdot)$ and ρ in Definition 1-(iv) are such that, for any Borel subset A of \mathcal{D} , $\inf_{|m| \leq \kappa_3 + 1} P_g(m + e_t \in A) \geq (1 - \rho) \underline{P}(A)$.*

Under Assumption P, a choice of $\underline{P}(\cdot)$ is a uniform distribution since the Gaussian density is positive continuous. The quantities $\rho, \underline{P}, d, \underline{n}_T$ and κ_2 in Definition 1 need not to be known and are only used to obtain rate-consistency uniformly over $\mathcal{P}_{\mathcal{D}}(L, \sigma)$. The Condition (3.2) ensures that the $\|X_{t-1} - x\|$'s has a continuous distribution. The condition (3.3) imposes a minimal recurrence rate over \mathcal{D} . The Minorization Condition (iv) is used in Markov processes Theory in the so-called splitting technique, see Meyn and Tweedie (1993), Nummelin (1984), and Karlsen and Tjøstheim (2001) for nonparametric statistical applications. The conditions (3.3) and (iv) impose some important qualitative restrictions on the evolution equation $X_t = m(X_{t-1}) + e_t$. The next Proposition briefly recalls some simple assumptions ensuring recurrence and a Minorization Condition in the setup of Example 2.

PROPOSITION 1. Let $\{Y_t, t \geq 0\}$ be a time-homogeneous Markov process of order d with the autoregressive representation $Y_t = \mu(Y_{t-1}, \dots, Y_{t-d}) + \varepsilon_t$ for $t \geq d$. Let $\{X_t, t \geq 0\}$ with $X_t = (Y_t, \dots, Y_{t-d+1})'$ and $m(\cdot)$ be as in Example 2 with $\mathcal{X} = \mathbb{R}^d$. Assume that

- (a) The distribution of ε_t given $(Y_{t-1}, \dots, Y_{t-d}) = X_{t-1} = x$ has a probability density function $f(\varepsilon|x)$ which is bounded away from 0 and continuous with respect to (ε, x) in $\mathbb{R} \times \mathbb{R}^d$.
- (b) The regression function $\mu(\cdot)$ is continuous over \mathbb{R}^d .
- (c) There is a continuous $U(\cdot)$ from \mathbb{R}^d to \mathbb{R}^+ with $\lim_{\|x\| \rightarrow \infty} U(x) = \infty$, an integer number $d' > 0$ and a compact subset \mathcal{C} of \mathbb{R}^d such that $\mathbb{E}_{\mathbb{P}}[U(X_{t+d'})|X_t = x] \leq U(x)$ for all x in $\mathbb{R}^d \setminus \mathcal{C}$.

Let \mathcal{D} be any bounded subset of \mathbb{R}^d . Then the Continuity Condition (3.2) holds under (a) and (b), and there exists a distribution \underline{P} satisfying (3.1) and a ρ in $[0, 1)$ such that $\{X_t, t \geq 0\}$ verifies the Minorization Condition of Definition 1-(iv). Under (a), (b) and (c), $\mathbb{E}_{\mathbb{P}}N_T(\mathcal{D})$ diverges.

Assumption (c) is the so-called drift condition, see e.g. Meyn and Tweedie (1993) and Tong (1990). The choice $U(x) = \|x\|^2$ yields (c) if $\|m(x)\|^2 + \text{Var}_{\mathbb{P}}[X_t|X_{t-1} = x] \leq \|x\|^2$, so that $\|m(x)\|$ can behave as $\|x\|$ when $\|x\|$ grows, as in the case of the Random Walk. Better choices of $U(\cdot)$ gives weaker restrictions on $m(\cdot)$, see Meyn and Tweedie (1993, Theorem 9.5.6). Meyn and Tweedie (1993, Theorem 9.4.2) gives a converse to Proposition 1 which ensures existence of such $U(\cdot)$ provided $\mathbb{P}(\lim_{T \rightarrow \infty} N_T(\mathcal{D}) = \infty | X_0 = x) = 1$ for all x .

3.2. *Main results* A major difficulty in the study of $m_T(x; \hat{b}_T)$ and $\mu_T(x; \hat{h}_T)$ is due to the fact that the bandwidths \hat{b}_T and \hat{h}_T are data-driven. A first step to overcome this dependence is to show that $N_T(\mathcal{D})$ has a deterministic exact order which diverges, as done in Theorem 1.

THEOREM 1. Let $N_T(\mathcal{D}) = \sum_{t=1}^T \mathbb{I}(X_{t-1} \in \mathcal{D})$. Then, for each \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, s)$, there exists a deterministic diverging sequence $n_T = n_T(\mathbb{P})$ which gives the exact order of $N_T(\mathcal{D})$ in probability uniformly over $\mathcal{P}_{\mathcal{D}}(L, s)$, i.e.

$$\lim_{T, z \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}(L, \sigma)} \mathbb{P} \left(\frac{n_T(\mathbb{P})}{z} \leq N_T(\mathcal{D}) \leq n_T(\mathbb{P})z \right) = 1.$$

Theorem 1 is a uniform version of Theorem 2.1 in Chen (1999b), and is proven using the splitting technique. The next two results build on Theorem 1 to bound the stochastic fluctuations of \hat{b}_T and \hat{h}_T , which gives the rate-consistency of $m_T(x; \hat{b}_T)$ and $\mu_T(x; \hat{h}_T)$.

THEOREM 2. *Let $m_T(x; b)$, $\widehat{b}_T = \widehat{b}_T(x; L, \sigma)$ and $R_T = R_T(x; L, \sigma)$ be as in (2.1), (2.3) and (2.4). Then, for any x in \mathcal{D} , $R_T \left(m_T(x; \widehat{b}_T) - m(x) \right)$ is bounded in probability uniformly over $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, i.e.*

$$\lim_{T, z \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(R_T \left\| m_T(x; \widehat{b}_T) - m(x) \right\| \geq z \right) = 0 \text{ for any } x \text{ in } \mathcal{D}.$$

THEOREM 3. *Assume that the kernel function $K(\cdot)$ satisfies Assumption K. Let $\mu_T(x; h) = \mu_T(x; h, K)$, $\widehat{h}_T = \widehat{h}_T(x; L, \sigma, \kappa_0, K)$ and $R_T(K) = R_T(x; L, \sigma, \kappa_0, K)$ be as in (2.5), (2.7) and (2.8). Then, for any x in \mathcal{D} , $R_T(K) \left(\mu_T(x; \widehat{h}_T) - m(x) \right)$ is bounded in probability uniformly over $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, and the random rates R_T and $R_T(K)$ are of the same order i.e.*

$$(3.4) \quad \lim_{T, z \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(R_T(K) \left\| \mu_T(x; \widehat{h}_T) - m(x) \right\| \geq z \right) = 0$$

$$(3.5) \quad \text{and } \frac{\kappa_1}{2\sqrt{2}} R_T \leq R_T(K) \leq \sqrt{\kappa_0} R_T \text{ for any } x \text{ in } \mathcal{D}.$$

At the difference of Theorem 1, the proofs of Theorems 2 and 3 avoid using a splitting argument, which may forbid extension to continuous-time processes or random fields. The proofs of Theorems 2 and 3 build instead on results from Orey (1959) used in Yakowitz (1993). A first step notes that $m_T(x; \widehat{b}_T)$ averages over values of X_{t-1} in \mathcal{D} for \widehat{b}_T small enough. Hence the proof of Theorems 2 and 3 rely on the so-called ‘‘process on \mathcal{D} ’’ of Orey (1959) given by the successive values of $\{X_{t-1}, t \geq 1\}$ in \mathcal{D} . As recalled in Lemma A.1 in Appendix A, the process on \mathcal{D} is Φ -mixing with exponential decay. This allows to find the exact order of the bandwidth \widehat{b}_T when the bound for $N_T(\mathcal{D})$ in Theorem 1 holds. This gives the order of $R_T = 1/(L\widehat{b}_T)$ which is used to bound the stochastic error term of $m_T(x; \widehat{b}_T) - m(x)$ using a maximal inequality, leading to Theorem 2. Theorem 3 is proven similarly using the restrictions in (2.7) for \widehat{h}_T . Compared to Karlsen and Tjøstheim (2001), Theorems 2 and 3 do not restrict the dynamics of the time series to β -null recurrence and avoid technical smoothness conditions on its invariant measure.

Achieving uniform results in Theorems 1 to 3 necessitates the uniform Recurrence Condition (3.3) over the class $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, together with (3.1) which bounds from below the transition probability of X_t given X_{t-d} in Definition 1-(iv). Without these restrictions the lagged process $\{X_{t-1}, t \geq 1\}$ may not visit the vicinity of some x in \mathcal{D} , so that it can be conjectured that consistent estimators of $m(x)$ do not exist. Under these conditions, the rate R_T diverges in probability with the sample size, showing that $m_T(x; \widehat{b}_T)$ and $\mu_T(x; \widehat{h}_T)$ are consistent estimators of $m(x)$. However R_T can diverge slowly, with an order ranging from 1 to \sqrt{T} as seen from (2.4). Section 4 shows that all these orders can be achieved.

In view of the bound (3.5) in Theorem 3, $m_T(x; \widehat{b}_T)$ and $\mu_T(x; \widehat{h}_T)$ converge to $m(x)$ with the same rate R_T . But the upper bound $R_T(K) \leq \sqrt{\kappa_0} R_T$ in (3.5) suggests that it can be an artifact due to choice of the set $\mathcal{H}_T = \mathcal{H}_T(\kappa_0)$ of admissible bandwidths in (2.7), since taking $\kappa_0 = +\infty$ may give a better kernel estimator. This raises the issue of the rate optimality, i.e. of the existence of estimator converging to $m(x)$ faster than R_T . The two next results show that R_T is the minimax optimal random estimation rate of $m(x)$ over $\mathcal{P}_{\mathcal{D}}(L, \sigma)$. Theorem 4 below adopts a local minimax framework. Define, for any \mathbb{P}_0 in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, x in \mathcal{D} and $\epsilon > 0$,

$$\mathcal{V}_x(\mathbb{P}_0; \epsilon) = \{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma); \|m(x) - m_0(x)\| \leq \epsilon\},$$

where $m(\cdot) = \mathbb{E}_{\mathbb{P}}[X_t | X_{t-1} = \cdot]$ and $m_0(\cdot) = \mathbb{E}_{\mathbb{P}_0}[X_t | X_{t-1} = \cdot]$. Theorem 4 and Corollary 1 are stated for univariate Markov time series, but multivariate extension with higher Markov order can be proven similarly.

THEOREM 4. *Let $R_T = R_T(x; L, \sigma)$ be as in (2.4) and assume that Assumption P holds. Consider a sequence of i.i.d. $\mathcal{N}(0, \sigma^2)$ innovation terms $\{e_t, t \geq 0\}$. Let \mathbb{P}_0 be the distribution of $\{X_t^0, t \geq 0\}$ with $X_0^0 = e_0$ and $X_t^0 = m_0(X_{t-1}^0) + e_t$, where the regression function $m_0(\cdot)$ is continuous over \mathbb{R} and $L/2$ -Lipschitz over \mathcal{D} , with $\sup_{x \in \mathcal{D}} |m_0(x)| \leq \kappa_3$ for κ_3 as in Assumption P. Assume also that there is a real continuous function $U(\cdot) \geq 0$ with $\lim_{|x| \rightarrow \infty} U(x) = +\infty$ such that*

$$(3.6) \quad \mathbb{E}_{\mathbb{P}_0} [U(X_{t+1}^0) | X_t^0 = x] \leq U(x) \text{ for all } x \text{ in } \mathbb{R} \setminus \mathcal{D}',$$

where \mathcal{D}' is a finite closed interval containing \mathcal{D} .

Then there exists a diverging sequence $\underline{n}_T > 0$ such that \mathbb{P}_0 is in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, and there is a $z > 0$ such that, for any x in \mathcal{D} ,

$$\liminf_{T \rightarrow +\infty} \inf_{\widehat{m}_T} \sup_{\mathbb{P} \in \mathcal{V}_x(\mathbb{P}_0; \epsilon)} \mathbb{P}(R_T \|\widehat{m}_T(x) - m(x)\| \geq z) > 0 \text{ for any } \epsilon > 0,$$

where $\inf_{\widehat{m}_T}$ is the infimum over all the possible estimators $\widehat{m}_T(\cdot)$ using X_0, \dots, X_T .

The lower bound of Theorem 4 shows that R_T is an optimal estimation rate in the local minimax sense, i.e. that estimator converging to $m(x)$ faster than R_T uniformly over $\mathcal{V}_x(\mathbb{P}_0; \epsilon)$ does not exist. The Drift Condition (3.6) is in line with Proposition 1. The next Corollary shows that R_T is a global minimax estimation rate over $\mathcal{P}_{\mathcal{D}}(L, \sigma)$.

COROLLARY 1. *Assume that Assumption P holds and that $\sup_{T \geq 1} (\underline{n}_T/T)$ is small enough. Then, for any x in \mathcal{D} and uniformly over $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, the fastest rate of convergence of any estimator $\widehat{m}_T(x)$ to $m(x)$ is R_T , i.e. there is a $z > 0$ with*

$$\liminf_{T \rightarrow +\infty} \inf_{\widehat{m}_T} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(R_T \|\widehat{m}_T(x) - m(x)\| \geq z) > 0 \text{ for all } x \text{ in } \mathcal{D}.$$

4. Examples of design-adaptation Finding the exact deterministic order of the random rate R_T can be useful to calibrate the sample size T to achieve a given precision for some specific class of time series. It also illustrates how the design-adaptive approach can cope with irregular recurrence behaviors of the observations at hand. For the sake of simplicity we consider univariate time series $\{X_t, t \geq 0\}$ and focus on the estimation of $m(0)$ with $\mathcal{D} = (-1, 1)$. We first recall some results of Guerre (2000) for the regression model $Y_t = m(X_t) + e_t$. Consider the modified Box-Cox transformation

$$H_{\alpha, \theta}(y) = \text{sgn}(y)\theta|y|^\alpha, \text{ with } \alpha, \theta > 0, \text{sgn}(y) = 1 \text{ if } y \geq 0, \text{sgn}(y) = -1 \text{ if } y < 0,$$

$$\text{so that } H_{\alpha, \theta}^{-1}(x) = \text{sgn}(x) \left(\frac{|x|}{\theta} \right)^{1/\alpha} = H_{1/\alpha, \theta^{-1/\alpha}}(x).$$

Guerre (2000) considered a design $X_t = H_{\alpha, \theta}(U_t)$ where the U_t are i.i.d. uniform random variables. The X_t 's have density $(x/\theta)^{1/\alpha-1} \mathbb{1}_{[0, \theta]}(x)/(\alpha\theta)$ which is ill-conditioned at 0, that is can vanish or diverge, except for $\alpha = 1$. Standard nonparametric methods are usually limited to $\alpha = 1$, see e.g. Stone (1980), while the design-adaptive approach deals with all values of the unknown α . Both the expression of $H_{\alpha, \theta}(\cdot)$ and of the design density are helpful to understand how α and θ affect the design repartition in the vicinity of 0. Indeed, for $x \in \mathcal{D}$, $\lim_{\alpha \rightarrow +\infty} H_{\alpha, \theta}(x) = 0$ while $\lim_{\alpha \rightarrow 0} H_{\alpha, \theta}(x) = \theta \text{sgn}(x)$. Therefore $\{X_t, t \geq 0\}$ visits small neighborhoods of 0 more frequently for large α and small θ . The counterpart of the random rate R_T in Guerre (2000) is equivalent to $(L^2/(\sigma^2\theta^{1/\alpha}))^{\alpha/(2\alpha+1)} T^{\alpha/(2\alpha+1)}/L$ in probability. These deterministic rates range from 1 for $\alpha = 0$ to \sqrt{T} for $\alpha = +\infty$, the rate 1 being achieved with the constant design $X_t = 1$ while \sqrt{T} corresponds to $X_t = 0$ for all t . For $\alpha = 1$, the order of R_T is $T^{1/3}$, the usual minimax rate for Lipschitz regression functions, see Stone (1980) among others.

Let us now return to the time series context. Such results carry over to transformations of weak dependent Markov time series and we consider now the case of transformations of a Gaussian Random Walk. As shown below, the Gaussian Random Walk is more spread out than a weak dependent process so that slower rates will be achieved. Set

$$Y_t = Y_{t-1} + g_t \text{ with } Y_0 = g_0, \text{ so that } Y_t = \sum_{i=0}^t g_i, \text{ where the } g_t\text{'s are i.i.d. } \mathcal{N}(0, 1),$$

$$(4.1) \quad X_t = H_{\alpha, \theta}(Y_t) \text{ for } t \geq 0.$$

Because $H_{\alpha, \theta}(\cdot)$ is one to one, $\{X_t = H_{\alpha, \theta}(Y_t), t \geq 0\}$ is also a Markov process. Since

$$X_t = H_{\alpha, \theta}(Y_t) = H_{\alpha, \theta}(H_{1/\alpha, \theta^{-1/\alpha}}(X_{t-1}) + g_t)$$

$$= H_{\alpha, \theta}(g_t) + \int_0^{X_{t-1}} H'_{\alpha, \theta}(H_{1/\alpha, \theta^{-1/\alpha}}(z) + g_t) H'_{1/\alpha, \theta^{-1/\alpha}}(z) dz,$$

with $\mathbb{E}H_{\alpha,\theta}(g_t) = 0$, we have

$$\begin{aligned} m(x) &= \mathbb{E}_{\mathbb{P}}[X_t | X_{t-1} = x] = \mathbb{E} [H_{\alpha,\theta} (H_{1/\alpha,\theta^{-1/\alpha}}(x) + g_t)] \\ &= \int_0^x H'_{1/\alpha,\theta^{-1/\alpha}}(z) \mathbb{E} [H'_{\alpha,\theta} (H_{1/\alpha,\theta^{-1/\alpha}}(z) + g_t)] dz \text{ so that} \\ m'(x) &= H'_{1/\alpha,\theta^{-1/\alpha}}(x) \mathbb{E} [H'_{\alpha,\theta} (H_{1/\alpha,\theta^{-1/\alpha}}(x) + g_t)] , \\ &\text{which gives } \sup_{x \in \mathcal{D}} |m'(x)| < \infty \text{ for } \alpha \leq 1, \text{ and,} \end{aligned}$$

$$\sup_{x \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[|e_t|^{2p} | X_{t-1} = x] \leq 2^{2p-1} \left(\sup_{x \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[|X_t|^{2p} | X_{t-1} = x] + \sup_{x \in \mathcal{D}} |m(x)|^{2p} \right) < \infty ,$$

for any $p \geq 1$, where e_t is the innovation term $X_t - m(X_{t-1})$. Therefore, for $\alpha \leq 1$, the time series $\{X_t, t \geq 0\}$ satisfies Conditions (ii) and (iii) of Definition 1 for some L and σ depending upon α and θ . The parameters α and θ drive the recurrence of the transformed process $\{X_t, t \geq 1\}$ as seen from its invariant measure. The Gaussian Random Walk $\{Y_t, t \geq 0\}$ is a null-recurrent process with the Lebesgue measure as invariant measure, so that $\{X_t = H_{\alpha,\theta}(Y_t), t \geq 0\}$ is also null-recurrent with invariant measure $(x/\theta)^{1/\alpha-1} dx / (\alpha\theta)$, which parallels the density of the design $H_{\alpha,\theta}(U_t)$ of Guerre (2000). The invariant density diverges at 0 for $\alpha > 1$, indicating clustering at 0, while it vanishes for $\alpha < 1$. This contrasts with the rate of $\mathbb{E}_{\mathbb{P}} N_T(\mathcal{D})$, which is \sqrt{T} for all α, θ . Indeed, since the Y_{t-1} 's are $\mathcal{N}(0, t)$, the expression of $H_{\alpha,\theta}(\cdot)$ yields

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} N_T(\mathcal{D}) &= \sum_{t=1}^T \mathbb{P} \left(|Y_{t-1}| \leq \frac{1}{\theta^{1/\alpha}} \right) = \sum_{t=1}^T \mathbb{P} \left(|\mathcal{N}(0, 1)| \leq \frac{1}{\theta^{1/\alpha} \sqrt{t}} \right) \\ &\sim \sum_{t=1}^T \frac{2}{\sqrt{2\pi} \theta^{1/\alpha} \sqrt{t}} \sim \frac{4\sqrt{T}}{\sqrt{2\pi} \theta^{1/\alpha}} . \end{aligned}$$

The order \sqrt{T} is smaller than the order T achieved by the number of visits of weak dependent Markov processes. A more precise asymptotic study of R_T and $N_T(\mathcal{D})$ necessitates to introduce the local time $\{\lambda(w, s), w \in \mathbb{R}, s \in \mathbb{R}^+\}$ of a standard Brownian Motion $\{W(s), s \in \mathbb{R}^+\}$. The local time is the density function of the occupation time $\int_0^t \mathbb{I}(W(s) \in A) ds$ with respect to Lebesgue measure, i.e.

$$\int \mathbb{I}(w \in A) \lambda(w, t) dw = \int_0^t \mathbb{I}(W(s) \in A) ds ,$$

see Revuz and Yor (1991) for a more detailed exposition. In particular, $\lambda(0, 1)$ has the distribution of the absolute value of a standard normal variables.

PROPOSITION 2. *Let $\{X_t, t \geq 0\}$ be as in (4.1), with $\alpha, \theta > 0$, $x = 0$ and $\mathcal{D} = (-1, 1)$. Define $R_T = R_T(0; L, \sigma)$ as in (2.4). Then, for any $L, \sigma > 0$, $(T^{-1/2} N_T(\mathcal{D}), T^{-\alpha/(4\alpha+2)} R_T)$*

converges in distribution to

$$\left(\frac{2\lambda(0,1)}{\theta^{1/\alpha}}, \frac{1}{L} \left(\frac{2L^2\lambda(0,1)}{\sigma^2\theta^{1/\alpha}} \right)^{\alpha/(2\alpha+1)} \right).$$

Compared with the weak dependence case, the order of R_T now increases from 1 to $T^{1/4}$ with an upper bound $T^{1/4}$ given by the order of $N_T^{1/2}(\mathcal{D})$. Combining Proposition 2 and Theorems 2 and 3 give, for $\alpha \leq 1$, new deterministic minimax rates $T^{\alpha/(4\alpha+2)}$ for the estimation of the regression function $m(0)$ over vicinities $\mathcal{V}_0(\mathbb{P}; \epsilon)$ of the model (4.1). In the regular case $\alpha = 1$, the exact order of R_T is $T^{1/6}$ which is slower than the rate $T^{1/3}$ achieved for weak dependent processes due to a lower recurrence rate \sqrt{T} for $N_T(\mathcal{D})$ in place of T . An estimation procedure in the spirit of Delattre et al. (2002), or a bias-variance tradeoff based on the asymptotic expansions of Karlsen and Tjøstheim (2001), give estimators which also achieved this optimal rate.

However, such procedures would be less efficient than the design-adaptive approach if $\alpha \neq 1$. From a heuristical viewpoint, these procedures parallel standard bias variance analysis, viewing $N_T(\mathcal{D})$, instead of the sample size, as the relevant number of observations to estimate $m(0)$. See also Spokoiny (2000) who imposes a condition corresponding to the ergodic case. According to such analogy, the order of the stochastic term is $1/(N_T(\mathcal{D})b)^{1/2}$ with a bias bounded by b time a constant. Balancing this two terms suggests a bandwidth of order $N_T^{-1/3}(\mathcal{D})$ in place of the usual $T^{-1/3}$ order for estimating Lipschitz functions. But the proof of Proposition 2 shows that the order of the stochastic error of $m_T(x; b)$ is $1/(N_T(\mathcal{D})b^{1/\alpha})^{1/2}$ for small b , as seen from (B.2.3) in Appendix B, instead of $1/(N_T(\mathcal{D})b)^{1/2}$. It follows that a bandwidth of order $N_T^{-1/3}(\mathcal{D}) \sim_{\mathbb{P}} T^{-1/6}$ gives a rate $\min((T^{1/2}T^{-1/6\alpha})^{1/2}, T^{1/6}) \sim T^{(3\alpha-1)/12\alpha}$ for $\alpha \leq 1$, which is, for $\alpha < 1$, smaller than the rate $T^{\alpha/(4\alpha+2)}$ of the design-adaptive approach. The superiority of the design-adaptive approach comes from the empirical bias-variance tradeoff leading to the choice (2.3) of \widehat{b}_T , while its competitors use a more theoretical bias-variance tradeoff which is not appropriate for the range $\alpha < 1$ of transformations of the Random Walk. Note that the gain can be important since a bandwidth of order $N_T^{-1/3}(\mathcal{D})$ give an estimator which is not consistent if $\alpha < 1/3$, showing the interest of design-adaptation which can automatically cope with irregular local recurrence behaviors of the sample.

APPENDIX A: PROOFS OF MAIN RESULTS.

A.1. Additional notations and conventions Let $\{\tau_i, i \geq 1\}$ be the successive epochs where the process $\{X_{t-1}, t \geq 1\}$ visits \mathcal{D} , i.e.

$$\tau_1 = \inf \{t \geq 1, X_{t-1} \in \mathcal{D}\}, \dots, \tau_{k+1} = \inf \{t > \tau_k; X_{t-1} \in \mathcal{D}\},$$

and let $x_{i-1} = X_{\tau_{i-1}}$ be the associated values of the process, the so-called ‘‘process on \mathcal{D} ’’ in the Orey (1959) terminology, which is also an homogeneous Markov process. Define $\eta_i = e_{\tau_i}$, and let \mathcal{F}_i be the Borel field generated by $x_{i-1}, \eta_{i-1}, \dots, x_0, \eta_0$. Set

$$(A.1.1) \quad \eta_i(j) = \langle \eta_i, v_j \rangle, \quad \omega_i(j) = \eta_i^2(j) - \mathbb{E}_{\mathbb{P}}[\eta_i^2(j)|\mathcal{F}_i], \quad \zeta_i(j) = \omega_i^2(j) - \mathbb{E}_{\mathbb{P}}[\omega_i^2(j)|\mathcal{F}_i].$$

Let $\{\Phi_k, k \geq 0\}$ be the Φ -mixing coefficients of the Markov process $\{x_i, i \geq 0\}$, i.e.

$$\Phi_k = \Phi_k(\mathbb{P}) = \sup_{n \geq 0} \sup_{A \text{ with } \mathbb{P}(x_n \in A) > 0, B} |\mathbb{P}(x_{n+k} \in B | x_n \in A) - \mathbb{P}(x_{n+k} \in B)|.$$

Consider now the estimation procedure. We aim to give, as far as possible, a unified treatment for $m_T(x; b)$ and $\mu_T(x; h)$ which coincide if $K(z) = \mathbb{I}(z \leq 1)$. For $A \subset \mathbb{R}$, define

$$K_A^2(x_{i-1}) = K_A^2(x_{i-1}; x) = - \int_0^1 \mathbb{I} \left(\frac{\|x_{i-1} - x\|}{z} \in A \right) dK^2(z)$$

where the negative differential term $dK^2(z)$ is minus the Dirac mass at 1 if $K(z) = \mathbb{I}(z \leq 1)$, in which case $K_A(x_{i-1}) = \mathbb{I}(\|x_{i-1} - x\| \in A)$. Note that $K^2(\|x_{i-1} - x\|/h) = K_{(-\infty, h]}^2(x_{i-1}) = - \int_0^1 \mathbb{I}(\|x_{i-1} - x\| \leq zh) dK^2(z)$, which increases with h under Assumption K. Since x is in \mathcal{D} which is open, the regression estimators $m_T(x; \hat{b}_T)$, $\mu_T(x; \hat{h}_{0T})$ and $m_T(x; \hat{h}_T)$ averages over the X_{t-1} 's in \mathcal{D} , that is over the x_{i-1} 's, when \hat{b}_T , \hat{h}_{0T} and \hat{h}_T are small enough. We therefore introduce some counterparts for $N_T(h; K)$, the stochastic errors of $m_T(x; b)$ and $\mu_T(x; h)$, \hat{b}_T , and \hat{h}_{0T} defined with respect to $\{x_i, i \geq 0\}$. For $\beta \in \mathbb{R}$ and $A \subset \mathbb{R}$, define

$$\nu_n(\beta) = \nu_n(\beta; x, K) = \sum_{i=1}^n K^2 \left(\frac{\|x_{i-1} - x\|}{\beta} \right), \quad \nu_n(A) = \sum_{i=1}^n K_A^2(\|x_{i-1} - x\|),$$

noticing that $\nu_n((-\infty, \beta]) = \nu_n(\beta)$. Observe that $N_T(\beta; K) = \nu_{N_T(\mathcal{D})}(\beta)$ for β small enough. The counterpart of the stochastic term is $\Sigma_n(\beta)/\nu_n(\beta)$ with

$$(A.1.2) \quad \Sigma_n(\beta) = \Sigma_n(\beta; x, K) = \sum_{i=1}^n K \left(\frac{\|x_{i-1} - x\|}{\beta} \right) \eta_i.$$

The counterpart of the bandwidths \hat{b}_T and \hat{h}_{0T} is

$$(A.1.3) \quad \hat{\beta}_n = \hat{\beta}_n(x; L, \sigma, K) = \arg \min_{\beta \geq 0} \max \left(L\beta, \frac{\sigma}{\sqrt{\nu_n(\beta)}} \right) = \min \{ \beta \geq 0; L^2 \beta^2 \nu_n(\beta) \geq \sigma^2 \},$$

using the same convention than in (2.3). It is easily seen that $\hat{\beta}_n$, as \hat{b}_T , decreases with the sample size under Assumption K or if $K(z) = \mathbb{I}(z \leq 1)$, see e.g. Lemma A.2-(i) for a proof of a similar statement. The definitions (2.3) and (2.6) of \hat{b}_T and \hat{h}_{0T} yields that $\hat{b}_T = \hat{\beta}_{N_T(\mathcal{D})}$ if $K(z) = \mathbb{I}(z \leq 0)$ and $\hat{h}_{0T} = \hat{\beta}_{N_T(\mathcal{D})}$ as soon as $\hat{\beta}_{N_T(\mathcal{D})}$, or \hat{b}_T and \hat{h}_{0T} , are small enough.

We now introduce some deterministic counterparts for $\nu_n(\cdot)$ and $\hat{\beta}_n$. Define, for $\beta \geq 0$ and $A \subset \mathbb{R}$

$$(A.1.4) \quad F_i(\beta) = F_i(\beta; x, \mathbb{P}, K) = \mathbb{E}_{\mathbb{P}} K^2 \left(\frac{\|x_{i-1} - x\|}{\beta} \right) = - \int_0^1 \mathbb{P}(\|x_{i-1} - x\| \leq z\beta) dK^2(z),$$

$$(A.1.5) \quad \bar{F}_n(\beta) = \frac{1}{n} \sum_{i=1}^n F_i(\beta) = \mathbb{E}_{\mathbb{P}}[\nu_n(\beta)/n], \quad F_i(A) = \mathbb{E}_{\mathbb{P}} K_A^2(x_{i-1}), \quad \bar{F}_n(A) = \frac{1}{n} \sum_{i=1}^n F_i(A).$$

Under Assumption K or if $K(z) = \mathbb{I}(z \leq 1)$, $F_i(\cdot)$ and $\bar{F}_n(\cdot)$ are cumulative distribution functions. With the same convention as in (A.1.3), set

$$(A.1.6) \quad \beta_n = \beta_n(x, L, \sigma, \mathbb{P}, K) = \min \{ \beta \geq 0; L^2 n \beta^2 \bar{F}_n(\beta; \mathbb{P}, x) \geq \sigma^2 \} ,$$

In what follows, $[z]$ is the integer part of $z \in \mathbb{R}$ and, C_i , C or C' stand for constants which does not depend upon \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$. The constant C or C' may vary from line to line.

A.2. Intermediate results We begin with a lemma for the process on \mathcal{D} .

LEMMA A.1. *Assume that $K(z) = \mathbb{I}(z \leq 1)$ or satisfies Assumption K. Let $\{F_i(\cdot) = F_i(\cdot; x, \mathbb{P}, K), i \geq 1\}$ and $\{\bar{F}_n(\cdot) = \bar{F}_n(\cdot; x, \mathbb{P}, K), n \geq 1\}$ be as in (A.1.4)-(A.1.5). For any distribution \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, we have:*

i. For any $j \geq 1$, the sequences $\{\eta_i(j), i \geq 1\}$, $\{\omega_i(j), i \geq 1\}$ and $\{\zeta_i(j), i \geq 1\}$ defined in (A.1.1) are \mathcal{F}_i centered martingale differences, with

$$\mathbb{E}_{\mathbb{P}}[\eta_i^2(j)|\mathcal{F}_i] \leq \sigma_j^2, \quad \mathbb{E}_{\mathbb{P}}[\omega_i^2(j)|\mathcal{F}_i] \leq \kappa_2^4 \sigma_j^4 \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}[\zeta_i^2(j)|\mathcal{F}_i] \leq 8\kappa_2^8 \sigma_j^8 .$$

ii. The Φ -mixing coefficients of $\{x_i, i \geq 0\}$ are such that $\Phi_k(\mathbb{P}) \leq 2\rho^k$ for all $k \geq 0$.

iii. For any x in \mathcal{D} , the $F_i(\cdot)$'s are continuous, have a limit $F(\cdot) = F(\cdot; x, \mathbb{P}, K)$ when i grows which is such that $F(\beta) \geq (\rho^d - 1) \int_0^1 \underline{F}(\beta z) dK^2(z)$. The $\bar{F}_n(\cdot)$'s are continuous with $\bar{F}_n(\beta) > 0$ for all $\beta > 0$ and $n \geq d + 1$. Moreover

$$\sup_{\beta \geq 0} |F_i(\beta) - F(\beta)| \leq \rho^{i-1} \quad \text{and} \quad \sup_{\beta \geq 0} |\bar{F}_n(\beta; \mathbb{P}, x) - F(\beta; \mathbb{P}, x)| \leq \frac{1}{n(1-\rho)} .$$

iv. There exists a constant $C_1 = C_1(\rho, \kappa_2, K(\cdot)) > 0$ such that, for any $1 \leq n_1 < n_2$ and any $A \subset \mathbb{R}$,

$$(A.2.1) \quad \text{Var}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right) \leq \left(1 + \frac{4\rho}{1-\rho} \right) (n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A)) ,$$

$$(A.2.2) \quad \mathbb{E}_{\mathbb{P}}^{1/3} \left| \sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right|^3 \leq C_1 \left((1 + n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A))^3 - 1 \right)^{1/3} ,$$

$$(A.2.3) \quad \mathbb{E}_{\mathbb{P}}^{1/6} \left\| \sum_{i=n_1+1}^{n_2} K_A(x_{i-1}) \eta_i \right\|^6 \leq C_1 \sigma \left((1 + n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A))^3 - 1 \right)^{1/6} .$$

Proof of Lemma A.1 : (i) follows from Definition 1-(iii), the Strong Markov Property, $\mathbb{E}_{\mathbb{P}}[\omega_i^2(j)|\mathcal{F}_i] \leq \mathbb{E}_{\mathbb{P}}[\eta_i^4(j)|\mathcal{F}_i]$, $\mathbb{E}_{\mathbb{P}}[\zeta_i^2(j)|\mathcal{F}_i] \leq \mathbb{E}_{\mathbb{P}}[\omega_i^4(j)|\mathcal{F}_i] \leq \mathbb{E}_{\mathbb{P}}[\eta_i^8(j)|\mathcal{F}_i] + 6\mathbb{E}_{\mathbb{P}}[\eta_i^4(j)|\mathcal{F}_i]\mathbb{E}_{\mathbb{P}}^2[\eta_i^2(j)|\mathcal{F}_i] + \mathbb{E}_{\mathbb{P}}^4[\eta_i^2(j)|\mathcal{F}_i]$.

(ii) and (iii). Take $x = 0 \in \mathcal{D}$ without loss of generality, and let x denote now a variable. Set $P(x, A) = \mathbb{P}(X_1 \in A | X_0 = x)$, $P_{\mathcal{D}}(x, A) = \mathbb{P}(x_1 \in A | x_0 = x)$. We have, for all $A \subset \mathcal{D}$,

$$\begin{aligned} P_{\mathcal{D}}(x, A) &= P(x, A) + \int_{\mathcal{X} \setminus \mathcal{D}} P(x, dx_1) P(x_1, A) + \int_{\mathcal{X} \setminus \mathcal{D}} \int_{\mathcal{X} \setminus \mathcal{D}} P(x, dx_1) P(x_1, dx_2) P(x_2, A) + \dots \\ &\geq P(x, A) , \end{aligned}$$

which implies $P_{\mathcal{D}}(x, dy) \geq P(x, dy)$ for (almost) all x, y in \mathcal{D} . Hence

$$\mathbb{P}(x_{i+d} \in A | x_i = x) = \int P_{\mathcal{D}}(x_{d-1}, A) P_{\mathcal{D}}(x_{d-2}, dx_{d-1}) \times \cdots \times P_{\mathcal{D}}(x, dx_1) \geq \mathbb{P}(X_{t+d} \in A | X_t = x),$$

and then, by Definition 1-(iv)

$$(A.2.4) \quad \mathbb{P}(x_{i+d} \in A | x_i = x) \geq (1 - \rho^d) \underline{P}(A) \text{ for any } A \subset \mathcal{D} \text{ and } x \in \mathcal{D}.$$

This implies by Theorem 16.2.4 of Meyn and Tweedie (1993) that $\{x_i, i \geq 0\}$ has an invariant distribution $\pi_{\mathcal{D}}(\cdot)$ with

$$(A.2.5) \quad \sup_{A, x \in \mathcal{D}} |\mathbb{P}(x_{i-1} \in A | x_0 = x) - \pi_{\mathcal{D}}(A)| \leq \rho^{i-1}.$$

This gives, for all A with $\mathbb{P}(x_n \in A) > 0$,

$$\begin{aligned} & |\mathbb{P}(x_{n+k} \in B | x_n \in A) - \mathbb{P}(x_{n+k} \in B)| \\ & \leq |\mathbb{P}(x_{n+k} \in B | x_n \in A) - \pi_{\mathcal{D}}(x_{n+k} \in B)| + |\mathbb{P}(x_{n+k} \in B) - \pi_{\mathcal{D}}(x_{n+k} \in B)| \\ & = \frac{1}{\mathbb{P}(x_n \in A)} \left| \int_{x \in A} (\mathbb{P}(x_{n+k} \in B | x_n = x) - \pi_{\mathcal{D}}(x_{n+k} \in B)) \mathbb{P}(x_n \in dx) \right| \\ & \quad + \left| \int_{x \in \mathcal{D}} (\mathbb{P}(x_{n+k} \in B | x_n = x) - \pi_{\mathcal{D}}(x_{n+k} \in B)) \mathbb{P}(x_n \in dx) \right| \leq 2\rho^k, \end{aligned}$$

and then $\Phi_k \leq 2\rho^k$. Set $p_i(\beta) = \mathbb{P}(\|x_{i-1}\| \leq \beta)$, $\bar{p}_n(\beta) = \sum_{i=1}^n p_i(\beta)/n$ and $p(\beta) = \pi_{\mathcal{D}}(\|x_{i-1}\| \leq \beta)$, so that $F_i(\beta) = -\int_0^1 p_i(z\beta) dK^2(z)$ and $\bar{F}_n(\beta) = -\int_0^1 \bar{p}_n(z\beta) dK^2(z)$ by (A.1.4). We have $p(\beta) = \int_{\mathcal{D}} \mathbb{P}(\|x_{i+d}\| \leq \beta | x_i = x) \pi_{\mathcal{D}}(dx) \geq (1 - \rho^d) \underline{P}(\|x_i\| \leq \beta) \geq (1 - \rho^d) \underline{F}(\beta)$ by (A.2.4) and (3.1), which gives $F(\beta) = -\int_0^1 p(z\beta) dK^2(z) \geq (\rho^d - 1) \int_0^1 \underline{F}(z\beta) dK^2(z)$. (3.2) yields that the $p_i(\cdot)$'s are continuous. Hence the $F_i(\cdot)$'s and $\bar{F}_n(\cdot)$'s are continuous. (A.2.4) gives

$$p_{i+d+1}(\beta) = \mathbb{P}(\|x_{i+d}\| \leq \beta) = \int_{\mathcal{D}} \mathbb{P}(\|x_{i+d}\| \leq \beta | x_i = x) \mathbb{P}(x_i \in dx) \geq (1 - \rho^d) \underline{P}(\|x_{i+d}\| \leq \beta),$$

so that $p_{i+d+1}(\beta) > 0$ for $\beta > 0$ by (3.1). Hence $F_i(\beta) > 0$ and $\bar{F}_n(\beta) > 0$ for $\beta > 0$ and $i, n \geq d+1$.

(A.2.5) gives $\sup_{\beta \geq 0} |p_i(\beta) - p(\beta)| \leq \rho^{i-1}$. Hence integrating with respect to $dK^2(z)$ yields

$$\sup_{\beta \geq 0} |F_i(\beta) - F(\beta)| \leq \rho^{i-1}, \quad \sup_{\beta \geq 0} |\bar{F}_n(\beta) - F(\beta)| \leq \frac{1}{n} \sum_{i=1}^n \rho^{i-1} \leq \frac{1}{n(1-\rho)}.$$

(iv). Set $y_i(z) = \mathbb{I}\left(\frac{\|x_{i-1}\|}{z} \in A\right)$ and define now $p_i(\cdot)$ as $p_i(z) = \mathbb{P}(y_i(z) = 1)$, so that

$$\begin{aligned} K_A^2(x_{i-1}) &= -\int_0^1 y_i(z) dK^2(z), \quad F_i(A) = -\int_0^1 p_i(z) dK^2(z), \\ |\text{Cov}_{\mathbb{P}}(K_A^2(x_{i-1}), K_A^2(x_{j-1}))| &\leq \int_0^1 \int_0^1 |\text{Cov}_{\mathbb{P}}(y_i(z), y_{j-1}(z'))| dK^2(z) dK^2(z'), \end{aligned}$$

with $\text{Var}_{\mathbb{P}}(y_i(z)) = p_i(z)(1 - p_i(z)) \leq p_i(z)$ and for $i < j$,

$$|\text{Cov}_{\mathbb{P}}(y_i(z), y_j(z'))| = |\mathbb{P}(y_j(z') = 1 | y_i(z) = 1) - p_j(z')| p_i(z) \leq \Phi(j-i) p_i(z).$$

Since $-\int_0^1 dK^2(z) = 1$, it follows that

$$\begin{aligned} \text{Var}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right) &\leq \sum_{i=n_1+1}^{n_2} \text{Var}_{\mathbb{P}} (K_A^2(x_{i-1})) + 2 \sum_{i=n_1+1}^{n_2} \sum_{j=i+1}^{n_2} |\text{Cov}_{\mathbb{P}} (K_A^2(x_{i-1}), K_A^2(x_{j-1}))| \\ &\leq - \sum_{i=n_1+1}^{n_2} \int_0^1 p_i(z) dK^2(z) + 2 \sum_{i=n_1+1}^{n_2} \int_0^1 \int_0^1 p_i(z) dK^2(z) dK^2(z') \sum_{k=1}^{\infty} \Phi(k) \\ &\leq \left(1 + \frac{4\rho}{1-\rho} \right) \sum_{i=n_1+1}^{n_2} F_i(A), \end{aligned}$$

which gives (A.2.1). For (A.2.2), set $\mathcal{I}_3 = \{(i_1, i_2, i_3); n_1+1 \leq i_1 < i_2 < i_3 \leq n_2\}$ and $\mathcal{I}_2 = \{(i_1, i_2); n_1+1 \leq i_1 < i_2 \leq n_2\}$. This gives

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left| \sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right|^3 &= \mathbb{E}_{\mathbb{P}} \left| \int_0^1 \left(\sum_{i=n_1+1}^{n_2} y_i(z) \right) dK^2(z) \right|^3 \leq C \left[- \sum_{\mathcal{I}_3} \int_{[0,1]^3} \mathbb{E}_{\mathbb{P}} \left(\prod_{j=1}^3 y_{i_j}(z_j) dK^2(z_j) \right) \right. \\ &\quad \left. + \sum_{\mathcal{I}_2} \int_{[0,1]^2} \mathbb{E}_{\mathbb{P}} \left(\prod_{j=1}^2 y_{i_j}(z_j) dK^2(z_j) \right) - \sum_{i=n_1+1}^{n_2} \int_0^1 \mathbb{E}_{\mathbb{P}} (y_i(z)) dK^2(z) \right]. \end{aligned}$$

We have, for $i_1 < i_2 < i_3$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}(y_{i_1}(z_1)y_{i_2}(z_2)y_{i_3}(z_3)) &= \mathbb{P}(y_{i_3}(z_3) = 1 | y_{i_2}(z_2) = 1, y_{i_1}(z_1) = 1) \mathbb{P}(y_{i_2}(z_2) = 1 | y_{i_1}(z_1) = 1) p_{i_1}(z_1) \\ &\leq (p_{i_3}(z_3) + \Phi(i_3 - i_2)) (p_{i_2}(z_2) + \Phi(i_2 - i_1)) p_{i_1}(z_1) \\ &= p_{i_1}(z_1) p_{i_2}(z_2) p_{i_3}(z_3) + p_{i_1}(z_1) p_{i_2}(z_2) \Phi(i_3 - i_2) \\ &\quad + p_{i_1}(z_1) p_{i_3}(z_3) \Phi(i_2 - i_1) + p_{i_1}(z_1) \Phi(i_3 - i_2) \Phi(i_2 - i_1), \end{aligned}$$

$$\mathbb{E}_{\mathbb{P}}(y_{i_1}(z_1)y_{i_2}(z_2)) = \mathbb{P}(y_{i_2}(z_2) = 1 | y_{i_1}(z_1) = 1) p_{i_1}(z_1) \leq p_{i_2}(z_2) p_{i_1}(z_1) + p_{i_1}(z_1) \Phi(i_2 - i_1).$$

Summing gives

$$\begin{aligned} \sum_{\mathcal{I}_3} p_{i_1}(z_1) p_{i_2}(z_2) \Phi(i_3 - i_2) &\leq 2 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} p_{i_2}(z_2) \sum_{i_3=i_2+1}^{n_2} \rho^{i_3-i_2} \\ &\leq \frac{2\rho}{1-\rho} \sum_{i=n_1+1}^{n_2} p_i(z_1) \sum_{j=i+1}^{n_2} p_j(z_2), \\ \sum_{\mathcal{I}_3} p_{i_1}(z_1) p_{i_3}(z_3) \Phi(i_2 - i_1) &\leq 2 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} \rho^{i_2-i_1} \sum_{i_3=i_2+1}^{n_2} p_{i_3}(z_3) \\ &\leq \frac{2\rho}{1-\rho} \sum_{i=n_1+1}^{n_2} p_i(z_1) \sum_{j=i+1}^{n_2} p_j(z_3), \\ \sum_{\mathcal{I}_3} p_{i_1}(z_1) \Phi(i_2 - i_1) \Phi(i_3 - i_1) &\leq 4 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} \rho^{i_2-i_1} \sum_{i_3=i_2+1}^{n_2} \rho^{i_3-i_2} \\ &\leq \frac{4\rho^2}{(1-\rho)^2} \sum_{i=n_1+1}^{n_2} p_i(z_1), \end{aligned}$$

$$\sum_{\mathcal{I}_2} p_{i_1}(z_1) \Phi(i_2 - i_1) \leq 2 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} \rho^{i_2-i_1} \leq \frac{2\rho}{1-\rho} \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1).$$

It follows that

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} \left| \sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right|^3 \\ & \leq C \left[\sum_{n_1+1 \leq i_1, i_2, i_3 \leq n_2} F_{i_1}(A) F_{i_2}(A) F_{i_3}(A) + \sum_{n_1+1 \leq i_1, i_2 \leq n_2} F_{i_1}(A) F_{i_2}(A) + \sum_{i=n_1+1}^{n_2} F_i(A) \right] \\ & = C \left[(n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A))^3 + (n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A))^2 + n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A) \right] \\ & \leq C \left[(1 + n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A))^3 - 1 \right], \end{aligned}$$

which is (A.2.2). For (A.2.3), set $S_j = \sum_{i=n_1+1}^{n_2} K_A(x_{i-1}) \eta_i(j)$, $S = \sum_{i=n_1+1}^{n_2} K_A(x_{i-1}) \eta_i = \sum_{j=1}^{\infty} S_j v_j$, so that we have to bound $\mathbb{E}_{\mathbb{P}}^{1/6} \|S\|^6$. The triangular inequality gives

$$\mathbb{E}_{\mathbb{P}}^{1/6} \|S\|^6 = \mathbb{E}_{\mathbb{P}}^{1/6} \left\| \sum_{j=1}^{\infty} S_j v_j \right\|^6 = \mathbb{E}_{\mathbb{P}}^{1/6} \left(\sqrt{\sum_{j=1}^{\infty} S_j^2} \right)^6 \leq \mathbb{E}_{\mathbb{P}}^{1/6} \left(\sum_{j=1}^{\infty} |S_j| \right)^6 \leq \sum_{j=1}^{\infty} \mathbb{E}_{\mathbb{P}}^{1/6} S_j^6,$$

and we now bound the $\mathbb{E}_{\mathbb{P}}^{1/6} S_j^6$'s. The Burkholder inequality (see e.g. Chow and Teicher (1998), Theorem 1 p. 396), convexity inequality, and the definition of $\omega_i(j)$, $\zeta_i(j)$ in (A.1.1) give that there are some constants $C > 0$ such that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}^{1/6} S_j^6 & = \mathbb{E}_{\mathbb{P}}^{1/6} \left(\sum_{i=n_1+1}^{n_2} K_A(x_{i-1}) \eta_i(j) \right)^6 \leq C \mathbb{E}_{\mathbb{P}}^{1/6} \left(\sqrt{\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \eta_i^2(j)} \right)^6 \\ & \leq C \left[\mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \mathbb{E}_{\mathbb{P}}[\eta_i^2(j) | \mathcal{F}_i] \right)^3 + \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \omega_i(j) \right)^3 \right]^{1/6} \\ & \leq C \left[\sigma_j^6 \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right)^3 + \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^4(x_{i-1}) (\mathbb{E}_{\mathbb{P}}[\omega_i^2(j) | \mathcal{F}_i] + \zeta_i(j)) \right)^{3/2} \right]^{1/6} \\ & \leq C \left[\sigma_j^6 \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right)^3 + \kappa_2^6 \sigma_j^6 \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^4(x_{i-1}) \right)^{3/2} + \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^4(x_{i-1}) \zeta_i(j) \right)^{3/2} \right]^{1/6} \\ & \leq C \left[\sigma_j^6 \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right)^3 + \kappa_2^6 \sigma_j^6 \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^4(x_{i-1}) \right)^{3/2} + \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^8(x_{i-1}) \zeta_i^2(j) \right)^{3/4} \right]^{1/6}. \end{aligned}$$

Observe that $0 \leq K_A^2(\cdot) \leq 1$ gives $K_A^{2p}(x_{i-1}) \leq K_A^2(x_{i-1})$ for $p = 2, 4$, that $\mathbb{E}|Z|^{3/2} = \mathbb{E}|Z|^{3/2} \mathbb{I}(|Z| \leq 1) + \mathbb{E}|Z|^{3/2} \mathbb{I}(Z > 1) \leq \mathbb{E}|Z| + \mathbb{E}|Z|^3$, and that $\mathbb{E}^{4/3}|Z| \leq \mathbb{E}|Z| + \mathbb{E}^2|Z|$. This together the Hölder inequality yield

$$\mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^4(x_{i-1}) \right)^{3/2} \leq \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right)^3 + \sum_{i=n_1+1}^{n_2} \mathbb{E}_{\mathbb{P}} K_A^2(x_{i-1}),$$

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^8(x_{i-1}) \zeta_i^2(j) \right)^{\frac{3}{4}} &\leq \mathbb{E}_{\mathbb{P}}^{\frac{4}{3}} \left(\sum_{i=n_1+1}^{n_2} K_A^8(x_{i-1}) \mathbb{E}[\zeta_i^2(j) | \mathcal{F}_i] \right) \leq (8\kappa_2\sigma_j)^6 \mathbb{E}_{\mathbb{P}}^{\frac{4}{3}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right) \\ &\leq (8\kappa_2\sigma_j)^6 \left[\mathbb{E}_{\mathbb{P}} \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right) + \mathbb{E}_{\mathbb{P}}^2 \left(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right) \right]. \end{aligned}$$

Substituting then gives, using (A.2.2), $\mathbb{E}_{\mathbb{P}}^{1/6} S_j^6 \leq C\sigma_j \left[(1 + n_2 \bar{F}_{n_2}(A) - n_1 \bar{F}_{n_1}(A))^3 - 1 \right]^{1/6}$. Summing over j gives (A.2.3), using the bound of $\mathbb{E}^{1/6} \|S\|^6$, since $\sum_{j=1}^{\infty} \sigma_j \leq \sigma$. \square

We now turn to the theoretical bandwidths of (A.1.6).

LEMMA A.2. *Assume that $K(z) = \mathbb{I}(z \leq 1)$ or satisfies Assumption K, and let $\{\beta_n = \beta_n(x, L, \sigma, \mathbb{P}, K), n \geq 1\}$ be as in (A.1.6). Then*

i. For any $\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)$ and $x \in \mathcal{D}$, β_n decreases with n , $L^2 n \beta_n^2 \bar{F}_n(\beta_n) = \sigma^2$ for $n \geq 1$, and $n \bar{F}_n(\beta_n)$ is increasing.

ii. There exists a positive sequence $\{\bar{\beta}_n = \bar{\beta}_n(L, \sigma, K), n \geq 1\}$ which does depend upon x in \mathcal{D} and \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$, such that $\lim_{n \rightarrow \infty} \bar{\beta}_n = 0$, and, for any x in \mathcal{D} and $n \geq 1$,

$$\sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \beta_n \leq \bar{\beta}_n \text{ and } \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} n \bar{F}_n(\beta_n) \geq \frac{\sigma^2}{L^2 \bar{\beta}_n^2}.$$

iii. For $z_0 > 1$, set $n_1 = \lfloor n/z_0 \rfloor$ and $n_2 = \lfloor n z_0 \rfloor$. Then there exists a constant $C_2 = C_2(z_0) > 0$ such that

$$\frac{C_2}{z_0^2} \leq \frac{\ell \bar{F}_\ell(\beta_\ell)}{n_2 \bar{F}_{n_2}(\beta_{n_2})} = \frac{\beta_{n_2}^2}{\beta_\ell^2} \leq 1, \ell = n_1, \dots, n_2,$$

for all $n_1 \geq 1$, \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ and x in \mathcal{D} .

Proof of Lemma A.2 : (i). Since $n \bar{F}_n(\beta) = F_1(\beta) + \dots + F_n(\beta)$, $L^2(n+1)\beta^2 \bar{F}_{n+1}(\beta) \geq L^2 n \beta^2 \bar{F}_n(\beta)$ and then $L^2(n+1)\beta_n^2 \bar{F}_{n+1}(\beta_n) \geq \sigma^2$ by definition of β_n , so that $\beta_{n+1} \leq \beta_n$ by (A.1.6). The continuity of the $\bar{F}_n(\cdot)$'s stated in Lemma A.1-(iii) gives $L^2 n \beta_n^2 \bar{F}_n(\beta_n) = \sigma^2$ since $\beta \in \mathbb{R}^+ \rightarrow L^2 \beta^2 \bar{F}_n(\beta)$ continuously increases from 0 to infinity. It follows that $n \bar{F}_n(\beta_n) = \sigma^2 / (L^2 \beta_n^2)$ decreases.

(ii). It is sufficient to prove the first inequality by (i). Set $\underline{F}_K(\beta) = (\rho^d - 1) \int_0^1 \underline{F}(\beta z) dK^2(z)$, which continuously increases from 0 to 1 by (3.1). Lemma A.1-(iii) gives $\bar{F}_n(\cdot) \geq \underline{F}_K(\cdot) - 1/n(1 - \rho)$ for all x in \mathcal{D} and \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$. Hence $L^2 n \beta^2 \bar{F}_n(\beta) \geq L^2 n \beta^2 (\underline{F}_K(\beta) - 1/n(1 - \rho))$. Note that there is a $\bar{\beta}_n$, which does not depend on \mathbb{P} , such that

$$L^2 n \bar{\beta}_n^2 \left(\underline{F}_K(\bar{\beta}_n) - \frac{1}{n(1 - \rho)} \right) = \sigma^2.$$

Then $\beta_n \leq \bar{\beta}_n$ for all \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ since $L^2 n \bar{\beta}_n^2 \bar{F}_n(\bar{\beta}_n) \geq \sigma^2$. That $\lim_{n \rightarrow \infty} \bar{\beta}_n = 0$ is a direct consequence of (3.1) which gives $\underline{F}_K(\cdot) > 0$ over \mathbb{R}_*^+ .

(iii). It is sufficient to bound $n_1 \bar{F}_{n_1}(\beta_{n_1}) / (n_2 \bar{F}_{n_2}(\beta_{n_2}))$ by (i), which implies that this ratio is $(\beta_{n_2} / \beta_{n_1})^2 \leq 1$. For the lower bound, $\beta_{n_1} \geq \beta_{n_2}$ and Lemma A.1-(iii) yield

$$\begin{aligned} \frac{n_1 \bar{F}_{n_1}(\beta_{n_1})}{n_2 \bar{F}_{n_2}(\beta_{n_2})} &\geq \frac{1}{z_0^2} \frac{\bar{F}_{n_1}(\beta_{n_1})}{\bar{F}_{n_2}(\beta_{n_1})} \geq \frac{1}{z_0^2} \frac{\bar{F}_{n_1}(\beta_{n_1})}{F(\beta_{n_1}) + \frac{1}{n_2(1-\rho)}} \geq \frac{1}{z_0^2} \frac{\bar{F}_{n_1}(\beta_{n_1})}{\bar{F}_{n_1}(\beta_{n_1}) + \frac{1}{n_2(1-\rho)} + \frac{1}{n_1(1-\rho)}} \\ &= \frac{1}{z_0^2} \left(1 - \frac{\frac{1}{1-\rho} \left(1 + \frac{1}{z_0^2}\right)}{n_1 \bar{F}_{n_1}(\beta_{n_1}) + \frac{1}{1-\rho} \left(1 + \frac{1}{z_0^2}\right)} \right) \geq \frac{1}{z_0^2} \left(1 - \frac{\frac{1}{1-\rho} \left(1 + \frac{1}{z_0^2}\right)}{\sigma^2 / (L^2 \bar{\beta}_{n_1}^2) + \frac{1}{1-\rho} \left(1 + \frac{1}{z_0^2}\right)} \right) \end{aligned}$$

for all $\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)$ and $x \in \mathcal{D}$, where the last lower bound goes to $1/z_0^2$ when n grows. \square

The next lemma shows in particular that $\hat{\beta}_n / \beta_n$ goes to 1 in probability.

LEMMA A.3. Assume that $K(z) = \mathbb{I}(z \leq 1)$ or satisfies Assumption K, and let $\{\beta_n = \beta_n(x, L, \sigma, \mathbb{P}, K), n \geq 1\}$ be as in (A.1.6) and $\{\hat{\beta}_n = \hat{\beta}_n(x, L, \sigma, K), n \geq 1\}$ as in (A.1.3). Define n_1, n_2 as in Lemma A.2, with $n_1 = \lfloor n/z_0 \rfloor$ and $n_2 = \lfloor n z_0 \rfloor$ for some $z_0 > 1$. Consider the event

$$\mathcal{E}_{1n}(\epsilon) = \left\{ (1-\epsilon)\beta_\ell < \hat{\beta}_\ell \leq (1+\epsilon)\beta_\ell, \frac{\ell \bar{F}_\ell(\beta_\ell)}{(1+\epsilon)^2} \leq \nu_\ell(\hat{\beta}_\ell) \leq \frac{\ell \bar{F}_\ell(\beta_\ell)}{(1-\epsilon)^2} \text{ for } \ell = n_1, \dots, n_2 \right\}.$$

Then, for any x in \mathcal{D} , $\lim_{n \rightarrow \infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(\mathcal{E}_{1n}(\epsilon)) = 1$ for all ϵ in $(0, 1)$.

Proof of Lemma A.3 : For brevity, assume $x = 0 \in \mathcal{D}$. Consider the event

$$E_{1n}(\delta) = \left\{ 1 - \delta \leq \frac{L^2 \beta_\ell^2 \nu_\ell(\beta_\ell)}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} \leq 1 + \delta, \ell = n_1, \dots, n_2 \right\}, \delta > 0.$$

Since β_ℓ decreases and $L^2 \beta_\ell^2 \ell \bar{F}_\ell(\beta_\ell) = \sigma^2$ by Lemma A.2-(i), we have

$$\begin{aligned} \frac{L^2 \beta_\ell^2 \nu_\ell(\beta_\ell)}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} &\geq \frac{L^2 \beta_{n_2}^2 \nu_\ell(\beta_{n_2})}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} \geq \frac{L^2 \beta_{n_2}^2 \nu_{n_1}(\beta_{n_2})}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} = \frac{L^2 \beta_{n_2}^2 \nu_{n_1}(\beta_{n_2})}{L^2 n_2 \beta_{n_2}^2 \bar{F}_{n_2}(\beta_{n_2})} = \frac{\nu_{n_1}(\beta_{n_2})}{n_2 \bar{F}_{n_2}(\beta_{n_2})}, \\ \frac{L^2 \beta_\ell^2 \nu_\ell(\beta_\ell)}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} &\leq \frac{L^2 \beta_{n_1}^2 \nu_\ell(\beta_{n_1})}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} \leq \frac{L^2 \beta_{n_1}^2 \nu_{n_2}(\beta_{n_1})}{L^2 \ell \beta_\ell^2 \bar{F}_\ell(\beta_\ell)} = \frac{L^2 \beta_{n_1}^2 \nu_{n_2}(\beta_{n_1})}{L^2 n_1 \beta_{n_1}^2 \bar{F}_{n_1}(\beta_{n_1})} = \frac{\nu_{n_2}(\beta_{n_1})}{n_1 \bar{F}_{n_1}(\beta_{n_1})}. \end{aligned}$$

Let $E_{1n}^c(\delta)$ be the event “ $E_{1n}(\delta)$ is false”. The upper and lower bounds above, the Markov inequality, (A.2.1) in Lemma A.1, and Lemmas A.1-(iii) and A.2-(i,ii) give

$$\begin{aligned} \mathbb{P}(E_{1n}^c) &\leq \mathbb{P}\left(\frac{\nu_{n_2}(\beta_{n_1})}{n_1 \bar{F}_{n_1}(\beta_{n_1})} - 1 \geq \delta\right) + \mathbb{P}\left(\frac{\nu_{n_1}(\beta_{n_2})}{n_2 \bar{F}_{n_2}(\beta_{n_2})} - 1 \leq -\delta\right) \\ &\leq \frac{1}{\delta^2} \left(\frac{\text{Var}_{\mathbb{P}}(\nu_{n_2}(\beta_{n_1}))}{(n_1 \bar{F}_{n_1}(\beta_{n_1}))^2} + \frac{\text{Var}_{\mathbb{P}}(\nu_{n_1}(\beta_{n_2}))}{(n_2 \bar{F}_{n_2}(\beta_{n_2}))^2} \right), \end{aligned}$$

$$\text{with } \frac{\text{Var}_{\mathbb{P}}(\nu_{n_1}(\beta_{n_2}))}{(n_2 \bar{F}_{n_2}(\beta_{n_2}))^2} \leq C \frac{n_1 \bar{F}_{n_1}(\beta_{n_2})}{(n_2 \bar{F}_{n_2}(\beta_{n_2}))^2} \leq \frac{C}{n_2 \bar{F}_{n_2}(\beta_{n_2})} \leq C \bar{\beta}_{n_2}^2,$$

$$\frac{\text{Var}_{\mathbb{P}}(\nu_{n_2}(\beta_{n_1}))}{(n_1 \bar{F}_{n_1}(\beta_{n_1}))^2} \leq C \frac{n_2 \bar{F}_{n_2}(\beta_{n_1})}{(n_1 \bar{F}_{n_1}(\beta_{n_1}))^2} \leq C \frac{n_1 \bar{F}_{n_1}(\beta_{n_1}) + \frac{n_2}{1-\rho} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{(n_1 \bar{F}_{n_1}(\beta_{n_1}))^2} \leq \frac{C}{n_1 \bar{F}_{n_1}(\beta_{n_1})} \leq C \bar{\beta}_{n_1}^2.$$

Hence $\lim_{n \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(E_{1n}(\delta)) = 1$ for any $\delta > 0$. Consider now the event

$$E_{2n}(\epsilon) = \{(1-\epsilon)\beta_\ell < \hat{\beta}_\ell \leq (1+\epsilon)\beta_\ell, \ell = n_1, \dots, n_2\}, 0 < \epsilon < 1.$$

(A.1.3) gives, since $\nu_\ell(\beta)$ increases with β and $\sigma^2 = L^2\beta_\ell^2\ell\bar{F}_\ell(\beta_\ell)$,

$$\begin{aligned} E_{2n}(\epsilon) &= \{L^2(1-\epsilon)^2\beta_\ell^2\nu_\ell((1-\epsilon)\beta_\ell) < \sigma^2, L^2(1+\epsilon)^2\beta_\ell^2\nu_\ell((1+\epsilon)\beta_\ell) \geq \sigma^2, \ell \in [n_1, n_2]\} \\ &\supset \{L^2(1-\epsilon)^2\beta_\ell^2\nu_\ell(\beta_\ell) < \sigma^2, L^2(1+\epsilon)^2\beta_\ell^2\nu_\ell(\beta_\ell) \geq \sigma^2, \ell \in [n_1, n_2]\} \\ &= \left\{ \frac{1}{(1+\epsilon)^2} \leq \frac{L^2\beta_\ell^2\nu_\ell(\beta_\ell)}{L^2\beta_\ell^2\ell\bar{F}_\ell(\beta_\ell)} < \frac{1}{(1-\epsilon)^2}, \ell \in [n_1, n_2] \right\} \supset E_{1n}(\delta) \end{aligned}$$

for some $\delta \in (0, 1)$. Hence $\lim_{n \rightarrow \infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(E_{2n}(\epsilon)) = 1$. The definition (A.1.3) of $\hat{\beta}_\ell$ gives

$$\nu_\ell(\hat{\beta}_\ell) \geq \frac{\sigma^2}{L^2\hat{\beta}_\ell^2} \geq \frac{\sigma^2}{L^2(1+\epsilon)^2\beta_\ell^2} = \frac{\ell\bar{F}_\ell(\beta_\ell)}{(1+\epsilon)^2}, \ell = n_1, \dots, n_2 \text{ on } E_{2n}(\epsilon).$$

For the upper bound of $\nu_\ell(\hat{\beta}_\ell)$, note that (3.2) implies that there is no ties among the x_i 's \mathbb{P} -almost surely. Consequently, there is a $\beta < \hat{\beta}_\ell$ close enough to $\hat{\beta}_\ell$ such that $\nu_\ell(\beta) = \nu_\ell(\hat{\beta}_\ell) - 1$ and with $L^2\beta^2(\nu_\ell(\hat{\beta}_\ell) - 1) < \sigma^2$. Taking $\beta \rightarrow \hat{\beta}_\ell$ in this inequality gives, on $E_{2n}(\epsilon)$,

$$\nu_n(\hat{\beta}_\ell) \leq \frac{\sigma^2}{L^2\hat{\beta}_\ell^2} + 1 \leq \frac{\sigma^2}{L^2(1-\epsilon)^2\beta_\ell^2} + 1 = \frac{\ell\bar{F}_\ell(\beta_\ell)}{(1-\epsilon)^2} + 1, \ell = n_1, \dots, n_2 \text{ on } E_{2n}(\epsilon).$$

This ends the proof of the Lemma since $\inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \inf_{n_1 \leq \ell \leq n_2} \ell\bar{F}_\ell(\beta_\ell) \rightarrow +\infty$ by Lemma A.2-(ii). \square

LEMMA A.4. Assume that $K(z) = \mathbb{I}(z \leq 1)$ or satisfies Assumption K, and consider $\{\beta_n = \beta_n(x, L, \sigma, \mathbb{P}, K), n \geq 1\}$, $\{\hat{\beta}_n = \hat{\beta}_n(x, L, \sigma, K), n \geq 1\}$ as in (A.1.6), (A.1.3). Let n_1, n_2 be as in Lemma A.2, with $n_1 = \lfloor n/z_0 \rfloor$ and $n_2 = \lfloor nz_0 \rfloor$ for some $z_0 > 1$, and consider a random measurable sequence $\{\hat{\gamma}_n, n \geq 1\}$ such that there is a $\kappa_5 > 1$, with for all x in \mathcal{D} ,

$$(A.2.6) \quad \lim_{n \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\frac{\nu_\ell(\hat{\beta}_\ell)}{\kappa_5} \leq \nu_\ell(\hat{\gamma}_\ell) \leq \kappa_5 \nu_\ell(\hat{\beta}_\ell) \text{ for all } \ell = n_1, \dots, n_2 \right) = 1.$$

Then there exists some deterministic sequences $\{\beta_{1n}, n \geq 1\}$, $\{\beta_{2n}, n \geq 1\}$, which depend upon z_0, \mathbb{P}, x in \mathcal{D}, L and σ^2 , such that, for

$$nr_n = n_2\bar{F}_{n_2}(\beta_{n_2}) = \sigma^2/(L^2\beta_{n_2}^2) \geq \sigma^2/(L^2\bar{\beta}_n^2),$$

we have

$$(A.2.7) \quad \lim_{n \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(\beta_{1n} \leq \hat{\gamma}_\ell \leq \beta_{2n} \text{ for all } \ell = n_1, \dots, n_2) = 1,$$

$$(A.2.8) \quad \lim_{n, z \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\frac{1}{z} \leq \frac{\nu_\ell(\hat{\gamma}_\ell)}{nr_n} \leq z \text{ for all } \ell = n_1, \dots, n_2 \right) = 1,$$

$$(A.2.9) \quad \frac{1}{C_4} \leq \frac{\ell\bar{F}_\ell(\beta)}{nr_n} \leq C_4 \text{ for all } \beta \in [\beta_{1n}, \beta_{2n}] \text{ and } \ell \in [n_1, n_2],$$

for $n_1 \geq 1$, and where $\bar{\beta}_n$ is as in Lemma A.2-(ii), with a constant $C_4 = C_4(z_0, \kappa_5, L, \sigma) > 1$.

Proof of Lemma A.4 : recall that $\inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} nr_n$ diverges when n grows. Let $C_2 = C_2(z_0)$ with $C_2/z_0^2 \leq 1$ be as in Lemma A.2-(iii). The continuity of the $F_\ell(\cdot)$'s from Lemma A.1-(iii) and $n_1\bar{F}_{n_1}(\cdot) \leq n_2\bar{F}_{n_2}(\cdot)$ give that there are $\beta_{1n} \leq \beta_{2n}$ with

$$n_2\bar{F}_{n_2}(\beta_{1n}) = \frac{C_2}{\kappa_5 z_0^2} \frac{nr_n}{3} \text{ and } n_1\bar{F}_{n_1}(\beta_{2n}) = 3\kappa_5 nr_n.$$

It follows that (A.2.9) holds by definition of r_n and Lemma A.1-(iii) which gives for some C_4

$$\begin{aligned}\ell\bar{F}_\ell(\beta) &\geq n_1\bar{F}_{n_1}(\beta_{1n}) \geq n_1\bar{F}_{n_2}(\beta_{1n}) - \frac{n_1}{1-\rho} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \\ &= \frac{C_2(1+o(1))nr_n}{3\kappa_5z_0^4} \frac{nr_n}{3} - \frac{n_1}{1-\rho} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \geq \frac{nr_n}{C_4}, \\ \ell\bar{F}_\ell(\beta) &\leq n_2\bar{F}_{n_2}(\beta_{2n}) \leq n_2\bar{F}_{n_1}(\beta_{2n}) + \frac{n_2}{1-\rho} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \\ &= 3z_0^2\kappa_5(1+o(1))nr_n + \frac{n_2}{1-\rho} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \leq C_4nr_n.\end{aligned}$$

(A.2.8) follows from (A.2.6), definition of r_n , Lemmas A.3 and A.2-(iii). (A.2.7) holds if

$$\lim_{n \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(\nu_\ell(\beta_{1n}) + 1 < \nu_\ell(\hat{\gamma}_\ell) < \nu_\ell(\beta_{2n}) - 1 \text{ for all } \ell \in [n_1, n_2]) = 1.$$

But (A.2.6), Lemmas A.3 and A.2-(iii,ii) show that

$$\lim_{n \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\frac{C_2nr_n/(2z_0^2) - 1}{\kappa_5} \leq \nu_\ell(\hat{\gamma}_\ell) \leq 2\kappa_5nr_n \text{ for all } \ell \in [n_1, n_2] \right) = 1.$$

Therefore, since $\nu_\ell(\beta_{1n}) \leq \nu_{n_2}(\beta_{1n})$ and $\nu_\ell(\beta_{2n}) \geq \nu_{n_1}(\beta_{2n})$, it is sufficient to show that

$$\lim_{n \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\nu_{n_2}(\beta_{1n}) + 1 < \frac{C_2nr_n/(2z_0^2) - 1}{\kappa_5}, 2\kappa_5nr_n < \nu_{n_1}(\beta_{2n}) - 1 \text{ for } \ell \in [n_1, n_2] \right) = 1.$$

But the definition of β_{1n} and β_{2n} and Lemma A.1-(A.2.1) yield

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\nu_{n_2}(\beta_{1n}) &= \frac{C_2}{\kappa_5z_0^2} \frac{nr_n}{3} \text{ with } \text{Var}_{\mathbb{P}}(\nu_{n_2}(\beta_{1n})) \leq \left(1 + \frac{4\rho}{1-\rho}\right) \frac{C_2}{\kappa_5z_0^2} \frac{nr_n}{3}, \\ \mathbb{E}_{\mathbb{P}}\nu_{n_1}(\beta_{2n}) &= 3\kappa_5nr_n \text{ with } \text{Var}_{\mathbb{P}}(\nu_{n_1}(\beta_{2n})) \leq 3\kappa_5 \left(1 + \frac{4\rho}{1-\rho}\right) nr_n,\end{aligned}$$

with $nr_n \geq \sigma^2/(L^2\bar{\beta}_n^2)$ for all $\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)$, so that the Markov inequality shows that the latter limit holds. It follows that (A.2.7) is true. \square

The next lemma is used for the stochastic error terms of the nonparametric regression estimators.

LEMMA A.5. *Take $K(z) = \mathbb{I}(z \leq 1)$ or assume that Assumption K holds. Let n_1, n_2 be as in Lemma A.2, with $n_1 = \lfloor n/z_0 \rfloor$ and $n_2 = \lfloor nz_0 \rfloor$ for some $z_0 > 1$. Consider a random measurable sequence $\{\hat{\gamma}_n, n \geq 1\}$ satisfying the condition (A.2.6) of Lemma A.4. Then, for any x in \mathcal{D} ,*

$$\lim_{n, z \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\max_{n_1 \leq \ell \leq n_2} \left\| \frac{\Sigma_\ell(\hat{\gamma}_\ell)}{\nu_\ell^{1/2}(\hat{\gamma}_\ell)} \right\| \geq z \right) = 0.$$

Proof of Lemma A.5 : Assume $x = 0 \in \mathcal{D}$ for brevity. By Lemma A.4-(A.2.8), it is sufficient to prove

$$\lim_{n, z \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\max_{n_1 \leq \ell \leq n_2} \left\| \frac{\Sigma_\ell(\hat{\gamma}_\ell)}{\sqrt{nr_n}} \right\| \geq z \right) = 0.$$

Let β_{1n} and β_{2n} be as in Lemma A.4. Define, for $\lambda = (\lambda_1, \lambda_2)$ in $[0, 1]^2$,

$$\beta_n(\lambda_1) = \beta_{1n} + \lambda_1(\beta_{2n} - \beta_{1n}), \quad n(\lambda_2) = n_1 + (n_2 - n_1)\lambda_2, \quad S_n(\lambda) = \frac{1}{\sqrt{nr_n}} \sum_{1 \leq i \leq n(\lambda_2)} K\left(\frac{\|x_{i-1}\|}{\beta_n(\lambda_1)}\right) \eta_i.$$

(A.2.7) and the Markov inequality show that the Lemma is proven if

$$(A.2.10) \quad \sup_{\mathbb{P} \in \mathcal{P}_D(L, \sigma)} \mathbb{E}_{\mathbb{P}}^{1/6} \max_{\lambda \in [0, 1]^2} \|S_n(\lambda)\|^6 \leq C.$$

The proof of (A.2.10) builds on a maximal inequality given in van der Vaart and Wellner (1996, Corollary 2.2.5). A first step is to bound the increments of $\|S_n(\cdot)\|$. We have, for any λ, λ' in $[0, 1]^2$,

$$\mathbb{E}_{\mathbb{P}}^{1/6} \left| \|S_n(\lambda)\| - \|S_n(\lambda')\| \right|^6 \leq \mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(\lambda_1, \lambda_2) - S_n(\lambda'_1, \lambda_2)\|^6 + \mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(\lambda'_1, \lambda_2) - S_n(\lambda'_1, \lambda'_2)\|^6.$$

The bound (A.2.3) of Lemma A.1-(iv) yields for the first item, assuming $\lambda_1 < \lambda'_1$

$$(A.2.11) \quad \begin{aligned} \mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(\lambda_1, \lambda_2) - S_n(\lambda'_1, \lambda_2)\|^6 &= \frac{1}{\sqrt{nr_n}} \mathbb{E}_{\mathbb{P}}^{1/6} \left\| \sum_{1 \leq i \leq n(\lambda_2)} K_{(\beta_n(\lambda_1), \beta_n(\lambda'_1))} (x_{i-1}) \eta_i \right\|^6 \\ &\leq \frac{C \left[(1 + n_2 (\bar{F}_{n_2}(\beta_n(\lambda_1)) - \bar{F}_{n_2}(\beta_n(\lambda'_1))))^3 - 1 \right]^{1/6}}{\sqrt{nr_n}} = Cd_{1n}(\lambda_1, \lambda'_1), \end{aligned}$$

For the increment with respect to λ_2 , (A.2.3) yields, for $\lambda_2 \leq \lambda'_2$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(\lambda'_1, \lambda_2) - S_n(\lambda'_1, \lambda'_2)\|^6 &= \frac{1}{\sqrt{nr_n}} \mathbb{E}_{\mathbb{P}}^{1/6} \left\| \sum_{n(\lambda_2) < i \leq n(\lambda'_2)} K\left(\frac{\|x_{i-1}\|}{\beta_n(\lambda'_1)}\right) \eta_i \right\|^6 \\ &\leq \frac{C' \left[\left(1 + \sum_{n(\lambda_2) < i \leq n(\lambda'_2)} F_i(\beta_n(\lambda'_1))\right)^3 - 1 \right]^{1/6}}{\sqrt{nr_n}} \leq \frac{C' \left[\left(1 + \sum_{n(\lambda_2) < i \leq n(\lambda'_2)} F_i(\beta_{2n})\right)^3 - 1 \right]^{1/6}}{\sqrt{nr_n}}. \end{aligned}$$

Now, Lemma A.1-(iii), the definition of r_n together with Lemma A.2-(ii), give

$$\sum_{n(\lambda_2) < i \leq n(\lambda'_2)} F_i(\beta_{2n}) \leq \sum_{n(\lambda_2) < i \leq n(\lambda'_2)} \left(\bar{F}_{n_2}(\beta_{2n}) + \frac{1}{n_1(1-\rho)} + \rho^{n_1-1} \right) \leq Cnr_n |\lambda_2 - \lambda'_2|,$$

and then,

$$(A.2.12) \quad \mathbb{E}_{\mathbb{P}} \|S_n(\lambda'_1, \lambda_2) - S_n(\lambda'_1, \lambda'_2)\|^6 \leq Cd_{2n}(\lambda_2, \lambda'_2) = C \frac{\left[(1 + nr_n |\lambda_2 - \lambda'_2|)^3 - 1 \right]^{1/6}}{\sqrt{nr_n}}.$$

Substituting (A.2.11)-(A.2.12) into the bound for $\mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(\lambda) - S_n(\lambda')\|^6$ yields

$$\mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(\lambda) - S_n(\lambda')\|^6 \leq Cd_n(\lambda, \lambda') \text{ where } d_n(\lambda, \lambda') = d_{1n}(\lambda_1, \lambda'_1) + d_{2n}(\lambda_2, \lambda'_2).$$

Note that $d_n(\cdot, \cdot)$ is a semimetric on $[0, 1]^2$. We now bound the covering numbers of $[0, 1]^2$ with respect to $d_n(\cdot, \cdot)$. Observe that $d_{1n}(\lambda_1, \lambda'_1) \leq \epsilon$ and $d_{2n}(\lambda_2, \lambda'_2) \leq \epsilon$ are equivalent to, by (A.2.11)-(A.2.12),

$$\begin{aligned} \frac{n_2 |\bar{F}_{n_2}(\beta_n(\lambda_1)) - \bar{F}_{n_2}(\beta_n(\lambda'_1))|}{nr_n} &\leq \frac{((nr_n)^3 \epsilon^6 + 1)^{1/3} - 1}{nr_n} \leq \epsilon^2, \\ |\lambda_2 - \lambda'_2| &\leq \frac{((nr_n)^3 \epsilon^6 + 1)^{1/3} - 1}{nr_n} \leq \epsilon^2. \end{aligned}$$

$\bar{F}_{n_2}(\cdot)$ is a cumulative distribution function and we used its quantiles to cover $[0, 1]$. It follows from (A.2.9) that the number of d_n -balls with radius ϵ necessary to cover $[0, 1]^2$ is bounded by C/ϵ^4 . Since (A.2.9), (A.2.11) and (A.2.12) yield that the d_n -diameter of $[0, 1]^2$ can be bounded by a $\delta > 0$ independently of n and \mathbb{P} , Corollary 2.2.5 of van der Vaart and Wellner (1996) shows that

$$\sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{E}_{\mathbb{P}}^{1/6} \max_{\lambda, \lambda' \in [0, 1]^2} \left| \|S_n(\lambda)\| - \|S_n(\lambda')\| \right|^6 \leq C \int_0^\delta \frac{d\epsilon}{\epsilon^{2/3}} < \infty.$$

This implies (A.2.10) since $\sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{E}_{\mathbb{P}}^{1/6} \|S_n(0, 1)\|^6$ is finite by (A.2.3) and (A.2.9). \square

A.3. Proofs of main results

A.3.1. Proof of Theorem 1. We proceed with the so called split chain technique, see e.g. Meyn and Tweedie (1993, Section 5.1) and Nummelin (1984, Section 4.4). We briefly recall some results of the split chain technique useful for the proof, and refer the reader to the references above for a broader overview. We begin with the case $d = 1$. Under the Minorization Condition of Definition 1-(iv), one can embed $\{X_t, t \geq 0\}$ into a larger probability space on which a sequence $\{Z_t, t \geq 0\}$ of $\{0, 1\}$ -valued random variables is defined such that $\{X_t^*, t \geq 0\} = \{(X_t, Z_t), t \geq 0\}$ is a Markov process on $\mathcal{X}^* = \mathcal{X} \times \{0, 1\}$. With a little abuse of language, we use \mathbb{P} to denote the distribution of $\{X_t^*, t \geq 0\}$. The process $\{X_t^*, t \geq 0\}$ has an atom $a^* = \mathcal{D} \times \{1\}$, with

$$\begin{aligned} \{X_t \in a^*\} &= \{Z_t = 1\}, \quad \mathbb{P}(X_t \in A | Z_t = 1, X_{t-1}^*, \dots, X_0^*) = \underline{P}(A) \text{ and} \\ \mathbb{P}(Z_t = 1 | X_t, X_{t-1}^*, \dots, X_0^*) &= (1 - \rho)\mathbb{I}(X_t \in \mathcal{D}), \end{aligned} \tag{A.3.1}$$

see (4.16b,c) in Nummelin (1984). This extends to the case $d > 1$ using (17.21) in Meyn and Tweedie (1993) to the d -step Markov chains $\{X_{dt+k}, t \geq 0\}$, $k = 0, \dots, d-1$. In the proof, we restrict to $d = 1$ since the case $d > 1$ can be similarly dealt with by considering each of the d -step Markov chains, because, for each T , at least one of them has an expected number of visits larger than n_T/d .

Let τ_1^* be the first epoch at which $\{X_{t-1}^*, t \geq 1\}$ visits a^* , i.e. $\tau_1^* = \inf\{t \geq 1; Z_{t-1} = 1\}$. By the strong Markov property, $\{X_{t-1}^*, t \geq \tau_1^*\}$, as $\{X_{t-1}^*, t \geq \tau_1^*\}$ given $\tau_1^* \leq T$, is a Markov chain with the transition of $\{X_t^*, t \geq 0\}$, initialized in the atom distribution given by $\mathbb{P}(Z_{\tau_1^*-1} = 0) = 0$, $\mathbb{P}(X_{\tau_1^*-1} \in A, Z_{\tau_1^*-1} = 1) = \underline{P}(A)$. The proof is divided in three steps. In the first step, we bound $\mathbb{P}(\tau_1^* > T)$. In the second step, we find a deterministic order for the number of visits $N_T^*(a^*) = \sum_{t=1}^T \mathbb{I}(Z_{t-1} = 1) = \sum_{t=\tau_1^*}^T \mathbb{I}(Z_{t-1} = 1)$ to the atom a^* . The last step concludes.

Step 1. The visit time τ_1^ .* (A.3.1) gives

$$\begin{aligned} \mathbb{P}(\tau_1^* > T) &= \mathbb{P}(Z_{T-1} = 0, \dots, Z_0 = 0) = \prod_{t=1}^{T-1} \mathbb{P}(Z_t = 0 | Z_i = 0, i \leq t-1) \times \mathbb{P}(Z_0 = 0) \\ &= \prod_{t=0}^{T-1} (1 - (1 - \rho)\mathbb{P}(X_t \in \mathcal{D})) = \prod_{t=1}^T (1 - (1 - \rho)\mathbb{P}(X_{t-1} \in \mathcal{D})) \text{ and then,} \\ \mathbb{P}(\tau_1^* > T) &\leq \exp\left(- (1 - \rho) \sum_{t=1}^T \mathbb{I}(X_{t-1} \in \mathcal{D})\right) = \exp(-(1 - \rho)\mathbb{E}_{\mathbb{P}} N_T(\mathcal{D})), \end{aligned}$$

so that $\sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(\tau_1^* > T) \leq \exp(-(1-\rho)\underline{n}_T) \rightarrow 0$ when $T \rightarrow +\infty$ by (3.3).

Step 2. The number of visits $N_T^(a^*)$.* Set $n_{1T} = n_{1T}(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}[N_T^*(a^*) | \tau_1^* \leq T]$, observing that $N_T^*(a^*) = 0$ when $\tau_1^* > T$. (A.3.1) gives $(1-\rho)\mathbb{E}_{\mathbb{P}}N_T(\mathcal{D}) = \mathbb{E}_{\mathbb{P}}N_T^*(a^*) = n_{1T}(1 - \mathbb{P}(\tau_1^* > T))$. Therefore, $\sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(\tau_1^* > T) \leq \exp(-(1-\rho)\underline{n}_T)$ proven in Step 1 and (3.3) yield

$$(A.3.2) \quad \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \left| \frac{(1-\rho)\mathbb{E}_{\mathbb{P}}N_T(\mathcal{D})}{n_{1T}(\mathbb{P})} - 1 \right| = \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \left| \frac{\mathbb{E}_{\mathbb{P}}N_T^*(a^*)}{n_{1T}(\mathbb{P})} - 1 \right| \leq \exp(-(1-\rho)\underline{n}_T)$$

so that $\inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} n_{1T}(\mathbb{P}) \geq C\underline{n}_T$.

The bound (2.4) of Chen (1999a) for $\tilde{N}_T^*(\alpha) = N_T^*(a^*) - 1 = \sum_{t=\tau_1^*+1}^T \mathbb{I}(Z_{t-1} = 1)$ yields, for any $\delta > 0$,

$$\mathbb{E}_{\mathbb{P}} \left[\tilde{N}_T^*(a^*) \mathbb{I}(\tilde{N}_T^*(a^*) \geq \delta) | \tau_1^* \leq T \right] \leq \mathbb{P} \left[\tilde{N}_T^*(a^*) \geq \delta | \tau_1^* \leq T \right] \left(\delta + 1 + \mathbb{E}_{\mathbb{P}} \left[\tilde{N}_T^*(a^*) | \tau_1^* \leq T \right] \right)$$

$$= \mathbb{P} [N_T^*(a^*) \geq 1 + \delta | \tau_1^* \leq T] (\delta + n_{1T}) \text{ with}$$

$$\mathbb{E}_{\mathbb{P}} \left[\tilde{N}_T^*(a^*) \mathbb{I}(\tilde{N}_T^*(a^*) \geq \delta) | \tau_1^* \leq T \right] = n_{1T} - 1 - \mathbb{E}_{\mathbb{P}} \left[\tilde{N}_T^*(a^*) \mathbb{I}(\tilde{N}_T^*(a^*) < \delta) | \tau_1^* \leq T \right] \geq n_{1T} - 1 - \delta,$$

where the last bound is, for T large, positive for all \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ by (A.3.2). Rearranging gives

$$(A.3.3) \quad \mathbb{P}(N_T^*(a^*) \geq 1 + \delta) \geq \frac{n_{1T} - 1 - \delta}{n_{1T} + \delta} \mathbb{P}(\tau_1^* \leq T).$$

Take $\delta = n_{1T}/z - 1 > 0$ for T large enough and all $\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)$ by (A.3.2). Substituting yields

$$\lim_{T, z \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(\frac{n_{1T}}{z} \leq N_T^*(a^*) \leq n_{1T}z \right) = 1,$$

since the Markov inequality gives $\limsup_{T \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(n_{1T}z < N_T^*(a^*)) \leq 1/z$.

Step 3. The number of visits $N_T(\mathcal{D})$. Set $n_T = (1-\rho)n_{1T}$ and observe that $N_T^*(a^*) \leq N_T(\mathcal{D})$ so that $\mathbb{P}(n_T/z \leq N_T^*(a^*)) \leq \mathbb{P}(n_T/z \leq N_T(\mathcal{D}))$ and then $\lim_{T, z \rightarrow +\infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(n_T/z \leq N_T(\mathcal{D})) = 1$. The Markov inequality and (A.3.2) give $\limsup_{T \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(N_T(\mathcal{D}) > n_Tz) = 1/z$. \square

A.3.2. Proof of Theorem 2. Assume for brevity that $x = 0 \in \mathcal{D}$. Since $R_T = 1/L\hat{b}_T$ and $R_T \leq N_T^{1/2}(\hat{b}_T)/\sigma$ by (2.3), the bias variance decomposition of $m_T(0; \hat{b}_T)$ yields

$$(A.3.4) \quad R_T \left\| m_T(0; \hat{b}_T) - m(0) \right\| \leq \frac{1}{L\hat{b}_T} \frac{\left\| \sum_{t=1}^T (m(X_{t-1}) - m(0)) \mathbb{I}(\|X_{t-1}\| \leq \hat{b}_T) \right\|}{N_T(\hat{b}_T)}$$

$$(A.3.5) \quad + \frac{\left\| \sum_{t=1}^T \mathbb{I}(\|X_{t-1}\| \leq \hat{b}_T) e_t \right\|}{\sigma N_T^{1/2}(\hat{b}_T)}.$$

Consider $\epsilon > 0$. By Theorem 1, there is a $z_0 > 1$ such that for $n_{1T} = \lfloor n_T/z_0 \rfloor$, $n_{2T} = \lfloor n_T z_0 \rfloor$, we have,

$$(A.3.6) \quad \liminf_{T \rightarrow \infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(E_{3T}) \geq 1 - \epsilon/2, \text{ where } E_{3T} = \{n_{1T} \leq N_T(\mathcal{D}) \leq n_{2T}\}.$$

Observe that (3.3) and (A.3.2) yield that $n_{1T} \geq \underline{n}'_T = \lfloor C\underline{n}_T/z_0 \rfloor \rightarrow +\infty$ for all \mathbb{P} in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$. Since the estimation point 0 is in the open \mathcal{D} , there is a $\bar{\beta} > 0$ such that $\{x \in \mathcal{X}; \|x\| \leq \bar{\beta}\} \subset \mathcal{D}$. Let $\bar{\beta}_n$ be as in Lemma A.2-(ii), so that $\bar{\beta}_{\underline{n}'_T} \leq \bar{\beta}/2$ for T large enough. Therefore, definitions (2.3) and (A.1.3) of \hat{b}_T and $\hat{\beta}_n$, (A.3.6), Lemmas A.3 and A.2-(ii) yield that

$$(A.3.7) \quad \liminf_{T \rightarrow \infty} \inf_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(E_{4T}) \geq 1 - \epsilon \text{ where } E_{4T} = \left\{ \hat{b}_T = \hat{\beta}_{N_T(\mathcal{D})} \leq \bar{\beta} \right\} \cap E_{3T}.$$

On E_{4T} , $m_T(0; \hat{b}_T)$ averages over \mathcal{D} . In this case, (A.3.4)-(A.3.5) and Definition 1-(ii) yield

$$R_T \left\| m_T(0; \hat{b}_T) - m(0) \right\| \leq 1 + \frac{\left\| \Sigma_{N_T(\mathcal{D})}(\hat{\beta}_{N_T(\mathcal{D})}) \right\|}{\sigma \nu_{N_T(\mathcal{D})}^{1/2}(\hat{\beta}_{N_T(\mathcal{D})})}.$$

Therefore, taking $\hat{\gamma}_n = \hat{\beta}_n$ in Condition (A.2.6) of Lemmas A.4 and A.5 yields that

$$\limsup_{T, z \rightarrow +\infty} \sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P} \left(R_T \left\| m_T(0; \hat{b}_T) - m(0) \right\| \geq z \text{ on } E_{4T} \right) = 0,$$

and Theorem 2 is proven since $\epsilon > 0$ can be arbitrarily small in (A.3.6)-(A.3.7). \square

A.3.3. Proof of Theorem 3. Take $x = 0 \in \mathcal{D}$ for brevity. We begin with (3.5). Under Assumption K, $N_T(\cdot; K)$ is continuous increasing, and

$$B_T(h) \leq h, \quad N_T(h) \geq N_T(h; K) \geq \sum_{t=1}^T K^2 \left(\frac{\|X_{t-1}\|}{h} \right) \mathbb{I} \left(\|X_{t-1}\| \leq \frac{h}{2} \right) \geq \kappa_1^2 N_T(h/2),$$

$$V_T(h) = \frac{\left(\sum_{t=1}^T K \left(\frac{\|X_{t-1}\|}{h} \right) \right)^2}{\sum_{t=1}^T K^2 \left(\frac{\|X_{t-1}\|}{h} \right)} \geq \frac{\left(\sum_{t=1}^T K^2 \left(\frac{\|X_{t-1}\|}{h} \right) \right)^2}{\sum_{t=1}^T K^2 \left(\frac{\|X_{t-1}\|}{h} \right)} = N_T(h; K).$$

Observe that the definition (2.6) of \hat{h}_{0T} gives $L^2 \hat{h}_{0T}^2 = \sigma^2 / N_T(\hat{h}_{0T}; K)$. Define

$$R_{0T} = \frac{1}{L \hat{h}_{0T}} = N_T^{1/2}(\hat{h}_{0T}; K) / \sigma = \left(\min_{h \geq 0} \max \left(L^2 h^2, \frac{\sigma^2}{N_T(h; K)} \right) \right)^{-1/2}.$$

$N_T(\cdot; K) \leq N_T(\cdot)$ yields $R_{0T} \leq R_T$ and $N_T(\cdot; K) \geq \kappa_1^2 N_T(\cdot/2)$ implies

$$\frac{1}{R_{0T}^2} \leq \min_{h \geq 0} \max \left(L^2 h^2, \frac{\sigma^2}{\kappa_1^2 N_T(h/2)} \right) \leq \frac{4}{\kappa_1^2} \min_{h \geq 0} \max \left(L^2 (h/2)^2, \frac{\sigma^2}{N_T(h/2)} \right) = \frac{4}{\kappa_1^2 R_T^2},$$

so that $\kappa_1 R_T/2 \leq R_{0T} \leq R_T$. We now bound $R_T(K)$ with R_{0T} . Because $LB_T(h) \leq Lh$, $\sigma^2/V_T(h) \leq \sigma^2/N_T(h; K)$ and \hat{h}_{0T} is in \mathcal{H}_T , we have

$$\frac{1}{R_T^2(K)} = \min_{h \in \mathcal{H}_T} \left(L^2 B_T^2(h) + \frac{\sigma^2}{V_T(h)} \right) \leq 2 \max \left(L^2 \hat{h}_{0T}^2, \frac{\sigma^2}{N_T(\hat{h}_{0T}; K)} \right) = \frac{2}{R_{0T}^2},$$

and $R_{0T}/\sqrt{2} \leq R_T(K)$. For the upper bound, observe that $V_T(\hat{h}_T) \leq \kappa_0 N_T(\hat{h}_T; K) = \kappa_0 \sigma^2 R_{0T}^2$ by definition of \mathcal{H}_T and (2.7). This gives $R_T(K) \leq V_T^{1/2}(\hat{h}_T)/\sigma \leq \sqrt{\kappa_0} R_{0T}$. Hence $R_{0T}/\sqrt{2} \leq R_T(K) \leq \sqrt{\kappa_0} R_{0T}$ with $\kappa_1 R_T/2 \leq R_{0T} \leq R_T$, and then (3.5).

The proof of (3.4) follows the same steps than the Proof of Theorem 2 and is briefly detailed. Since $R_T(K) \leq 1/LB_T(\hat{h}_T)$ and

$$R_T(K) \leq \frac{V_T^{1/2}(\hat{h}_T)}{\sigma} = \frac{1}{\sigma} \frac{\sum_{t=1}^T K \left(\|X_{t-1}\| / \hat{h}_T \right)}{N_T^{1/2}(\hat{h}_T; K)},$$

the bias-variance decomposition gives that $R_T(K) \left\| \mu_T(0; \hat{h}_T) - m(0) \right\|$ is bounded by

$$\frac{1}{LB_T(\hat{h}_T)} \left\| \frac{\sum_{t=1}^T (m(X_{t-1}) - m(0)) K \left(\frac{\|X_{t-1}\|}{\hat{h}_T} \right)}{\sum_{t=1}^T K \left(\frac{\|X_{t-1}\|}{\hat{h}_T} \right)} \right\| + \frac{\left\| \sum_{t=1}^T K \left(\frac{\|X_{t-1}\|}{\hat{h}_T} \right) e_t \right\|}{\sigma N_T^{1/2}(\hat{h}_T; K)}.$$

Now, (A.3.7) yields that $\widehat{h}_{0T} = \widehat{\beta}_{N_T(\mathcal{D})}$ with a large probability, and $\widehat{h}_T \leq \kappa_0 \widehat{h}_{0T}$ by definition of \mathcal{H}_T . Hence $\mu_T(0; \widehat{h}_T)$ averages on \mathcal{D} with large probability. Setting $\widehat{\gamma}_{N_T(\mathcal{D})} = \widehat{h}_T$ gives in this case

$$R_T(K) \left\| \mu_T(0; \widehat{h}_T) - m(0) \right\| \leq 1 + \frac{\left\| \Sigma_{N_T(\mathcal{D})}(\widehat{\gamma}_{N_T(\mathcal{D})}) \right\|}{\sigma \nu_{N_T(\mathcal{D})}^{1/2}(\widehat{\gamma}_{N_T(\mathcal{D})})}.$$

Since $N_T(\cdot; K) \leq V_T(\cdot; K)$, the bounds in \mathcal{H}_T implies

$$\frac{1}{\kappa_0} \nu_{N_T(\mathcal{D})}(\widehat{\beta}_{N_T(\mathcal{D})}) \leq \nu_{N_T(\mathcal{D})}(\widehat{\gamma}_{N_T(\mathcal{D})}) \leq \kappa_0 \nu_{N_T(\mathcal{D})}(\widehat{\beta}_{N_T(\mathcal{D})}),$$

that is Condition (A.2.6) in Lemmas A.4 and A.5, which gives (3.4). \square

A.3.4. Proof of Theorem 4 and Corollary 1. Take $\mathcal{D} = (-1, 1)$ and $x = 0$ without loss of generality. Recall that the $\{e_t, t \geq 0\}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$, and let \mathbb{P} be the associated distribution. The proof of Theorem 4 is divided in four steps.

Step 1. Construction of a family of alternatives and choice of \underline{n}_T . Consider a 1/2-Lipschitz function $\mu(\cdot) : \mathbb{R} \mapsto [0, 1/2]$ with support $[0, 1]$ and $\mu(0) = 1/2$, as $\mu(x) = (1 - |x|)\mathbb{I}_{[-1, 1]}(x)/2$. Define

$$(A.3.8) \quad m^b(x) = m_0(x) + Lb\mu\left(\frac{x}{b}\right), \quad b \in [0, \min(1, 1/L)],$$

with $m^0(\cdot) = \lim_{b \rightarrow 0} m^b(\cdot) = m_0(\cdot)$. Note that $m^b(x)$ is continuous over \mathbb{R} , L -Lipschitz over \mathcal{D} , that $\sup_{x \in \mathcal{D}} |m^b(x)| \leq \kappa_3 + 1$ and $m^b(x) = m_0(x)$ for all x in $\mathbb{R} \setminus \mathcal{D}$, for all b in $[0, \min(1, 1/L)]$. Let \mathbb{P}^b be the distribution of $\{X_t^b, t \geq 0\}$ where $X_0^b = e_0$ and $X_t^b = m(X_{t-1}^b) + e_t$, so that \mathbb{P}^b satisfies Definition 1-(ii,iii). Since $m^b(\cdot)$ and $m_0(\cdot)$ coincide over $\mathbb{R} \setminus \mathcal{D}$, (3.6) yields that the \mathbb{P}^b 's, b in $[0, \min(1, 1/L)]$, satisfy the Drift Condition (c) of Proposition 1, as well as (a) and (b) since $m^b(\cdot)$ is continuous and the e_t 's are i.i.d. normal variables. Hence Proposition 1 and Assumption P yield that (3.2) and Definition 1-(iv) hold with $\lim_{T \rightarrow \infty} \mathbb{E}_{\mathbb{P}^b} N_T(\mathcal{D}) = \infty$ for all b in $[0, \min(1, 1/L)]$.

Set $\underline{n}_T = \inf_{b \in [0, \min(1, 1/L)]} \mathbb{E}_{\mathbb{P}^b} N_T(\mathcal{D})$ so that $\mathbb{E}_{\mathbb{P}^b} N_T(\mathcal{D}) \geq \underline{n}_T$ for all b in $[0, \min(1, 1/L)]$. Recall that \underline{n}_T must diverge to define a class $\mathcal{P}_{\mathcal{D}}(L, \sigma)$. Observe that $b \mapsto X_t^b$ is continuous for all $t \geq 0$ so that the Lebesgue Dominated Convergence Theorem yields that

$$n_T(b) = \mathbb{E}_{\mathbb{P}^b} N_T(\mathcal{D}) = \mathbb{E}_{\mathbb{P}} \left[\sum_{t=0}^{T-1} \mathbb{I}(X_t^b \in \mathcal{D}) \right] \geq P_g(e_0 \in \mathcal{D}) > 0$$

is continuous for all $T \geq 0$. Since $1/n_T(\cdot)$ decreases with T and $\lim_{T \rightarrow \infty} 1/n_T(b) = 0$ for all b in $[0, \min(1, 1/L)]$, the Dini Theorem yields that $1/\underline{n}_T = \sup_{b \in [0, \min(1, 1/L)]} 1/n_T(b) \rightarrow 0$. Hence $\lim_{T \rightarrow \infty} \underline{n}_T = \infty$ and, for this choice of \underline{n}_T , \mathbb{P}^b is in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ for all b in $[0, \min(1, 1/L)]$.

Step 2. A contiguous alternative $\mathbb{P}_1 = \mathbb{P}_{1T}$. Consider a $z > 1$ to be chosen large enough. Under (3.2) and Definition 1-(iv) with (3.1), arguing as for Lemma A.2-(i) shows that there is a $b_T = b_T(z) \searrow 0$ such that

$$(A.3.9) \quad b_T = \min \left\{ b \geq 0; L^2 b^2 \mathbb{E}_{\mathbb{P}_0} N_T(b) \geq \frac{\sigma^2}{z^4} \right\}, \quad \text{or equivalently } L^2 b^2 \sum_{t=1}^T \mathbb{P}(|X_{t-1}^0| \leq b_T) = \frac{\sigma^2}{z^4}.$$

Abbreviate \mathbb{P}^{b_T} , $m^{b_T}(\cdot)$, $\{X_t^{b_T}, t \geq 0\}$ in \mathbb{P}_1 , $m_1(\cdot)$, $\{X_t^1, t \geq 0\}$ respectively. (A.3.8) yields that $|m_1(0) - m_0(0)| \leq Lb_T/2$ so that \mathbb{P}_1 is in $\mathcal{V}_0(\mathbb{P}_0, \epsilon)$ for T large enough. Normality of the e_t 's and

(A.3.8) show that the loglikelihood ratio $\mathcal{L}_T = \ln(d\mathbb{P}_1/d\mathbb{P}_0)(X_0, \dots, X_T)$ writes

$$\mathcal{L}_T = \frac{Lb_T}{\sigma^2} \sum_{t=1}^T \mu \left(\frac{X_{t-1}}{b_T} \right) e_t - \frac{L^2 b_T^2}{2\sigma^2} \sum_{t=1}^T \mu^2 \left(\frac{X_{t-1}}{b_T} \right) \text{ under } \mathbb{P}_0.$$

Note that the choice of $\mu(\cdot)$ in (A.3.8), and (A.3.9) give

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_0} \left(\frac{L^2 b_T^2}{2\sigma^2} \sum_{t=1}^T \mu^2 \left(\frac{X_{t-1}}{b_T} \right) \right) &\leq \frac{L^2 b_T^2}{8\sigma^2} \sum_{t=1}^T \mathbb{P}_0(|X_{t-1}| \leq b_T) \leq \frac{L^2 b_T^2 \mathbb{E}_{\mathbb{P}_0} N_T(b_T)}{8\sigma^2} = \frac{1}{8z^4}, \\ \mathbb{E}_{\mathbb{P}_0} \left(\frac{Lb_T}{\sigma^2} \sum_{t=1}^T \mu \left(\frac{X_{t-1}}{b_T} \right) e_t \right)^2 &= \frac{L^2 b_T^2}{\sigma^2} \sum_{t=1}^T \mathbb{E}_{\mathbb{P}_0} \mu^2 \left(\frac{X_{t-1}}{b_T} \right) \leq \frac{1}{4z^4}, \end{aligned}$$

and then, for any $z \geq 1$,

$$\mathbb{E}_{\mathbb{P}_0} |\mathcal{L}_T| \leq \mathbb{E}_{\mathbb{P}_0}^{1/2} \left(\frac{Lb_T}{\sigma^2} \sum_{t=1}^T \mu \left(\frac{X_{t-1}}{b_T} \right) e_t \right)^2 + \mathbb{E}_{\mathbb{P}_0} \left(\frac{L^2 b_T^2}{2\sigma^2} \sum_{t=1}^T \mu^2 \left(\frac{X_{t-1}}{b_T} \right) \right) \leq \frac{1}{z^2}.$$

Therefore the Markov inequality gives, for any $z' > 0$,

$$(A.3.10) \quad \sup_{T \geq 1} \mathbb{P}_0(|\mathcal{L}_T| \geq z') \leq \frac{z'}{z^2}.$$

Step 3. The random rate R_T under \mathbb{P}_0 and \mathbb{P}_1 . Recall that $R_T = 1/(L\hat{b}_T)$ with \hat{b}_T as in (2.3). Therefore, the definition (A.3.9) of b_T gives for $z \geq 1$ and under \mathbb{P}_0 ,

$$\begin{aligned} \mathbb{P}_0 \left(R_T \geq \frac{1}{2z^2 L b_T} \right) &= \mathbb{P}_0 \left(\hat{b}_T \leq 2z^2 b_T \right) = \mathbb{P}_0 \left(\sigma^2 \leq 4z^2 L^2 b_T^2 N_T(2z^2 b_T) \right) \\ &\geq \mathbb{P}_0 \left(\sigma^2 \leq 4z^2 L^2 b_T^2 N_T(b_T) \right) = \mathbb{P}_0 \left(\frac{\mathbb{E}_{\mathbb{P}_0} N_T(b_T)}{4} \leq N_T(b_T) \right). \end{aligned}$$

To find a lower bound for the probability above will be done using the split chain $\{X_t^*, t \geq 0\} = \{(X_t, Z_t), t \geq 0\}$ defined in the Proof of Theorem 1, and $N_T^*(a^*, b_T) = \sum_{t=1}^T \mathbb{I}(Z_{t-1} = 1, |X_t| \leq b_T) \leq N_T(b_T)$. Observe that $\{Z_{t-1} = 1, |X_t| \leq b_T\} \subset a^*$ is an atom of the split chain for T large enough. Therefore, arguing as for (A.3.3) in the proof of Theorem 1, with $N_T^*(a^*, b_T)$, $N_T(b_T)$ in place of $N_T^*(a^*)$, $N_T(\mathcal{D})$, gives

$$\mathbb{P}_0 \left(\frac{\mathbb{E}_{\mathbb{P}_0} N_T(b_T)}{4} \leq N_T(b_T) \right) \geq \frac{(1 - \frac{1}{4}) \mathbb{E}_{\mathbb{P}_0} N_T(b_T)}{(1 + \frac{1}{4}) \mathbb{E}_{\mathbb{P}_0} N_T(b_T)} (1 + o(1)) = \frac{3}{5} (1 + o(1)).$$

Under \mathbb{P}_1 , (A.3.10) gives

$$\begin{aligned} \mathbb{P}_1 \left(R_T \geq \frac{1}{2z^2 L b_T} \right) &= \mathbb{E}_{\mathbb{P}_0} \left[\mathbb{I} \left(R_T \geq \frac{1}{2z^2 L b_T} \right) \exp(\mathcal{L}_T) \right] \geq \frac{1}{3} \mathbb{P}_0 \left(R_T \geq \frac{1}{2z^2 L b_T}, \mathcal{L}_T \geq -\ln 3 \right) \\ &\geq \frac{1 + o(1)}{5} - \mathbb{P}_0(|\mathcal{L}_T| \geq \ln 3) \geq \frac{1 + o(1)}{5} - \frac{\ln 3}{z^2}. \end{aligned}$$

Hence

$$(A.3.11) \quad \liminf_{T \rightarrow +\infty} \max_{\mathbb{P} \in \{\mathbb{P}_0, \mathbb{P}_1\}} \mathbb{P}(E_T(z)) \geq \frac{1}{6} \text{ with } E_T(z) = \left\{ R_T \geq \frac{1}{2z^2 L b_T} \right\}, \text{ for } z \text{ large enough.}$$

Step 4. Proof of the lower bound. Because \mathbb{P}_1 is in $\mathcal{V}_0(\mathbb{P}_0; \epsilon)$ for T large enough, (A.3.11) gives, for any $z' > 0$,

$$\begin{aligned} & \liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \sup_{\mathbb{P} \in \mathcal{V}_0(\mathbb{P}_0; \epsilon)} \mathbb{P} \left(R_T |\hat{m}_T(0) - m(0)| \geq z' \right) \\ & \geq \liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \max_{\mathbb{P} \in \{\mathbb{P}_0, \mathbb{P}_1\}} \mathbb{P} \left(R_T |\hat{m}_T(0) - m(0)| \geq z', E_T(z) \right) \\ & \geq \liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \max_{\mathbb{P} \in \{\mathbb{P}_0, \mathbb{P}_1\}} \mathbb{P} \left(|\hat{m}_T(0) - m(0)| \geq 2Lz' z^2 b_T, E_T(z) \right). \end{aligned}$$

Since $m_1(0) = m_0(0) + Lb_T/2$ by choice of $\mu(\cdot)$ in (A.3.8), the triangular inequality gives

$$\begin{aligned} \mathbb{P}_1 \left(|\hat{m}_T(0) - m_1(0)| \geq 2Lz' z^2 b_T, E_T(z) \right) & \geq \mathbb{P}_1 \left(\frac{Lb_T}{2} - |\hat{m}_T(0) - m_0(0)| \geq 2Lz' z^2 b_T, E_T(z) \right) \\ & \geq \mathbb{P}_1 \left(|\hat{m}_T(0) - m_0(0)| < \frac{Lb_T}{2} - 2Lz' z^2 b_T, E_T(z) \right). \end{aligned}$$

Take z' with $Lb_T/2 - 2Lz' z^2 b_T = 2Lz' z^2 b_T$, i.e. $z' = 1/(8z^2)$, $2Lz' z^2 b_T = Lb_T/4$. Substituting gives

$$\begin{aligned} & \liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \sup_{\mathbb{P} \in \mathcal{V}_0(\mathbb{P}_0; \epsilon)} \mathbb{P} \left(R_T |\hat{m}_T(0) - m(0)| \geq \frac{1}{8z^2} \right) \\ & \geq \liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \max \left\{ \mathbb{P}_0 \left(|\hat{m}_T(0) - m_0(0)| \geq \frac{Lb_T}{4}, E_T(z) \right), \mathbb{P}_1 \left(|\hat{m}_T(0) - m_0(0)| < \frac{Lb_T}{4}, E_T(z) \right) \right\}. \end{aligned}$$

The maximum above is the maximum error on $E_T(z)$ of a test χ_T which accepts $\mathbb{P} = \mathbb{P}_0$, i.e. $\chi_T = 0$, if and only if $|\hat{m}_T(0) - m_0(0)| < Lb_T/4$. It follows that

$$\begin{aligned} & \liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \sup_{\mathbb{P} \in \mathcal{V}_0(\mathbb{P}_0; \epsilon)} \mathbb{P} \left(R_T |\hat{m}_T(0) - m(0)| \geq \frac{1}{8z^2} \right) \\ & \geq \liminf_{T \rightarrow \infty} \inf_{\chi_T} \max \{ \mathbb{P}_0(\chi_T = 1, E_T(z)), \mathbb{P}_1(\chi_T = 0, E_T(z)) \}. \end{aligned}$$

But, for any test χ_T ,

$$\begin{aligned} \mathbb{P}_1(\chi_T = 0, E_T(z)) & = \int \mathbb{I}(\chi_T = 1, E_T(z)) \exp(\mathcal{L}_T) d\mathbb{P}_0 \geq \exp(-z) \int \mathbb{I}(\chi_T = 1, \mathcal{L}_T \geq -z, E_T(z)) d\mathbb{P}_0 \\ & \geq \exp(-z) (\mathbb{P}_0(\mathcal{L}_T \geq -z, E_T(z)) - \mathbb{P}_0(\chi_T = 1, E_T(z))), \text{ and then} \\ \inf_{\chi_T} \max [\mathbb{P}_0(\chi_T = 1, E_T(z)), \mathbb{P}_1(\chi_T = 0, E_T(z))] & \geq \min_{p \in [0, 1]} \max [p, \exp(-z) (\mathbb{P}_0(\mathcal{L}_T \geq -z, E_T(z)) - p)] \\ & = \frac{\exp(-2z)}{1 + \exp(-z)} \mathbb{P}_0(\mathcal{L}_T \geq -z, E_T(z)) \geq \frac{\exp(-2z)}{1 + \exp(-z)} (\mathbb{P}_0(E_T(z)) - \mathbb{P}_0(|\mathcal{L}_T| \geq z)). \end{aligned}$$

Hence (A.3.10) and (A.3.11) show that

$$\liminf_{T \rightarrow \infty} \inf_{\hat{m}_T} \sup_{\mathbb{P} \in \mathcal{V}_0(\mathbb{P}_0; \epsilon)} \mathbb{P} \left(R_T |\hat{m}_T(0) - m(0)| \geq \frac{1}{8z^2} \right) \geq \frac{\exp(-2z)}{1 + \exp(-z)} \left(\frac{1}{6} - \frac{1}{z} \right),$$

with a lower bound which is strictly positive for z large enough, hence Theorem 4. \square

Corollary 1 follows from minor modifications of the proof of Theorem 4 detailed now. Let \mathbb{P}_0 be the distribution of $\{e_t, t \geq 0\}$ with $m_0(\cdot) = 0$, which is such that (3.6) holds for $U(x) = x^2$ and any \mathcal{D}' with $[-\sigma, \sigma] \subset \mathcal{D}'$ since $\mathbb{E}_{\mathbb{P}}[e_{t+1}^2 | e_t] = \sigma^2$. Assumption P ensures that \mathbb{P}_0 is in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ provided that $TP_g(e_t \in \mathcal{D}) \geq \underline{n}_T$. The perturbations \mathbb{P}^b of \mathbb{P}_0 defined through (A.3.8) satisfy a Minorization Condition $\mathbb{P}^b(X_{t+1} \in A | X_t = x) \geq (1 - \rho') \underline{P}'(A)$ for all x in \mathbb{R} and $A \subset \mathbb{R}$, where $\underline{P}'(dy)$ is proportional

to $\inf_{|m| \leq \kappa_3 + 1} \exp(-(y-m)^2/2\sigma^2) dy > 0$. This ensures that $\mathbb{E}_{\mathbb{P}^b} N_T(\mathcal{D}) \geq T(1-\rho') \underline{P}'(\mathcal{D}) > 0$, so that these \mathbb{P}^b are in $\mathcal{P}_{\mathcal{D}}(L, \sigma)$ provided $\sup_{T \geq 1} (\underline{n}_T/T) \geq (1-\rho') \underline{P}'(\mathcal{D})$. Corollary 1 then follows since $\sup_{\mathbb{P} \in \mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(R_T | \hat{m}_T(0) - m(0) \geq z) \geq \sup_{\mathbb{P} \in \mathcal{V}_0(\mathbb{P}_0; \epsilon)} \mathbb{P}(R_T | \hat{m}_T(0) - m(0) \geq z)$. \square

APPENDIX B: PROOF OF PROPOSITIONS 1 AND 2

B.1. Proof of Proposition 1. Set $x_t = (y_t, \dots, y_{t-d+1})$ and note that Y_t given $X_{t-1} = x_{t-1}$ has density $f(y - m(x_{t-1}) | x_{t-1})$ which is positive continuous in (y, x) under (a) and (b). Define $f_{\mathcal{D}}(y) = \inf_{x \in \mathcal{D}} f(y - m(x) | x)$ which is positive continuous for bounded \mathcal{D} . We have for any subset A of \mathbb{R}^d and integer number $n \geq 1$,

$$(B.1.1) \quad \mathbb{P}(X_t \in A | X_{t-n} = x_{t-n}) = \int \mathbb{I}(x_t \in A) \prod_{k=0}^{n-1} f(y_{t-k} - m(x_{t-k-1}) | x_{t-k-1}) dy_{t-k} .$$

Since the sphere $\{y \in \mathbb{R}^d; \|y - x\| = h\}$ has vanishing Lebesgue measure, (B.1.1) yields $\mathbb{P}(\|X_t - x\| = h | X_{t-d} = \cdot) = 0$, so that $\mathbb{P}(\|X_t - x\| = h) = 0$ for all t , which implies (3.2). We now check the Minorization Condition. Let $\mathcal{D}' = \prod_{i=1}^d [a, b]$ be such that $\mathcal{D} \subset \mathcal{D}'$. (B.1.1) yields, for any x_{t-d} in \mathcal{D} ,

$$\begin{aligned} \mathbb{P}(X_t \in A | X_{t-d} = x_{t-d}) &\geq \int \mathbb{I}(x_t \in A \cap \mathcal{D}) \prod_{k=0}^{d-1} f(y_{t-k} - m(x_{t-k-1}) | x_{t-k-1}) \mathbb{I}(x_{t-k-1} \in \mathcal{D}') dy_{t-k} \\ &\geq \int \mathbb{I}(x_t \in A \cap \mathcal{D}) \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) \mathbb{I}(x_{t-k-1} \in \mathcal{D}') dy_{t-k} = (1 - \rho_{\mathcal{D}}) \int \mathbb{I}(x_t \in A) \underline{f}_{\mathcal{D}}(x_t) dx_t , \end{aligned}$$

where, by construction of $\mathcal{D}' = \prod_{i=1}^d [a, b]$ and since x_{t-d} is in $\mathcal{D} \subset \mathcal{D}'$,

$$\begin{aligned} 1 - \rho_{\mathcal{D}} &= \int \mathbb{I}(x_t \in \mathcal{D}) \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) \mathbb{I}(y_{t-k} \in [a, b]) dy_{t-k} = \int \mathbb{I}(x_t \in \mathcal{D}) \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) dy_{t-k} , \\ \underline{f}_{\mathcal{D}}(x_t) &= \frac{\mathbb{I}(x_t \in \mathcal{D})}{1 - \rho_{\mathcal{D}}} \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) \mathbb{I}(y_{t-k} \in [a, b]) = \frac{\mathbb{I}(x_t \in \mathcal{D})}{1 - \rho_{\mathcal{D}}} \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) . \end{aligned}$$

If \mathcal{D} is a bounded open subset of \mathbb{R}^d , $0 \leq \rho_{\mathcal{D}} < 1$ and $\underline{f}_{\mathcal{D}}(\cdot)$ is a positive continuous density over \mathcal{D} . Hence $\underline{f}_{\mathcal{D}}(\cdot)$ defines a $\underline{P}_{\mathcal{D}}(\cdot)$ satisfying (3.1) and $\mathbb{P}(X_{t+d} \in A | X_t = x_t) \geq (1 - \rho_{\mathcal{D}}) \underline{P}_{\mathcal{D}}(A) \mathbb{I}(x_t \in \mathcal{D})$ for any Borel set $A \subset \mathcal{D}$, which is the Minorization Condition of Definition 1-(iv).

We now show that $\mathbb{E}_{\mathbb{P}} N_T(\mathcal{D})$ diverges for any \mathcal{D} with positive Lebesgue measure. Under (a) and (b), arguing as above yields a Minorization Condition $\mathbb{P}(X_{t+d} \in A | X_t = x_t) \geq (1 - \rho_{\mathcal{C}}) \underline{P}_{\mathcal{C}}(A) \mathbb{I}(x_t \in \mathcal{C})$ for any compact \mathcal{C} with positive Lebesgue measure, showing that the level sets of $U(\cdot)$, $\{x \in \mathbb{R}^d; U(x) \leq r\}$ are small for r large enough. (B.1.1), (a) and (b) implies that $\{X_t, t \geq 0\}$ is Lebesgue-irreducible, so that (c) and Theorem 8.4.3 in Meyn and Tweedie (1993) yield that $\{X_t, t \geq 0\}$ is Lebesgue recurrent, i.e. $\mathbb{E}_{\mathbb{P}}[N_T(\mathcal{D}) | X_0 = x] \rightarrow \infty$ for almost all x . Hence $\lim_{T \rightarrow \infty} \mathbb{E}_{\mathbb{P}} N_T(\mathcal{D}) = \infty$. \square

B.2. *Proof of Proposition 2.* (4.1) gives

$$(B.2.2) \quad N_T(b) = \sum_{t=1}^T \mathbb{I} \left(|Y_{t-1}| \leq \left(\frac{b}{\theta} \right)^{1/\alpha} \right) = N_T^Y \left(\left(\frac{b}{\theta} \right)^{1/\alpha} \right) \quad \text{where } N_T^Y(b) = \sum_{t=1}^T \mathbb{I}(|Y_{t-1}| \leq b).$$

Consider a sequence $b_T > 0$ going to 0 with $\lim_{T \rightarrow \infty} \sqrt{T}b_T = +\infty$. Theorem 2 of Akonom (1993, p. 70) gives that $\{Y_t, t \geq 0\}$ can be reconstructed on a larger probability space jointly with a standard Brownian motion $\{W(s), s \in \mathbb{R}^+\}$ such that

$$\frac{N_T^Y(b_T)}{b_T \sqrt{T}} - \frac{1}{b_T \sqrt{T}} \int_0^T \mathbb{I}(|W(s)| \leq b_T) ds = o(1), \quad \text{a.s.}$$

Observe that

$$\begin{aligned} \frac{1}{b_T \sqrt{T}} \int_0^T \mathbb{I}(|W(s)| \leq b_T) ds &\stackrel{d}{=} \frac{\sqrt{T}}{b_T} \int_0^1 \mathbb{I}(\sqrt{T}|W(s)| \leq b_T) ds = \frac{\sqrt{T}}{b_T} \int_{-b_T/\sqrt{T}}^{b_T/\sqrt{T}} \lambda(w, 1) dw \\ &\rightarrow 2\lambda(0, 1) \quad \text{a.s.} \end{aligned}$$

by continuity of the local time. Define now, for $k = -K, \dots, K$,

$$b_T = \frac{1}{(2\lambda(0, 1))^{\frac{\alpha}{2\alpha+1}}} \left(\frac{\sigma^2 \theta^{1/\alpha}}{L^2 \sqrt{T}} \right)^{\frac{\alpha}{2\alpha+1}}, \quad b_{k,T} = \frac{1}{\left(2 \left(1 + \frac{1}{\sqrt{K}} \right)^k \right)^{\frac{\alpha}{2\alpha+1}}} \left(\frac{\sigma^2 \theta^{1/\alpha}}{L^2 \sqrt{T}} \right)^{\frac{\alpha}{2\alpha+1}},$$

$$E_K = \left\{ \left(1 + \frac{1}{\sqrt{K}} \right)^{-K} \leq \lambda(0, 1) < \left(1 + \frac{1}{\sqrt{K}} \right)^K \right\},$$

noticing that $\lim_{T \rightarrow +\infty} \sqrt{T}b_{k,T}^{1/\alpha} = \infty$. Hence the Skohorod device gives that $\{Y_t, t \geq 0\}$ can be reconstructed jointly with a variable $\lambda(0, 1)$ such that, by (B.2.2) and the limits above,

$$(B.2.3) \quad \frac{N_T(\mathcal{D})}{\sqrt{T}} = \frac{2\lambda(0, 1)}{\theta^{1/\alpha}} + o(1), \quad N_T(b_{k,T}) = 2\sqrt{T}\lambda(0, 1) \left(\frac{b_{k,T}}{\theta} \right)^{\frac{1}{\alpha}} (1 + o(1)) \quad \text{a.s.},$$

$k = -K, \dots, K$. On the event E_K , set

$$\begin{aligned} \bar{b}_{K,T} &= b_{k,T} \text{ if } \left(1 + \frac{1}{\sqrt{K}} \right)^k \leq \lambda(0, 1) < \left(1 + \frac{1}{\sqrt{K}} \right)^{k+1}, \quad k = -K, \dots, K-1, \\ \underline{b}_{K,T} &= b_{k,T} \text{ if } \left(1 + \frac{1}{\sqrt{K}} \right)^{k-1} \leq \lambda(0, 1) < \left(1 + \frac{1}{\sqrt{K}} \right)^k, \quad k = -K+1, \dots, K, \end{aligned}$$

setting $\underline{b}_{K,T} = \bar{b}_{K,T} = 1$ outside E_K . It follows that

$$(B.2.4) \quad \underline{b}_{K,T} \leq b_T \leq \bar{b}_{K,T} \text{ on } E_K, \quad \text{with } \limsup_{K \rightarrow \infty} \sup_{T \geq 1} \frac{b_T}{\underline{b}_{K,T}} = \liminf_{K \rightarrow \infty} \inf_{T \geq 1} \frac{\bar{b}_{K,T}}{b_T} = 1.$$

(B.2.3) gives, on E_K

$$\begin{aligned} L^2 \underline{b}_{K,T}^2 N_T(\underline{b}_{K,T}) &= 2\sqrt{T}L^2 \lambda(0, 1) \frac{\underline{b}_{K,T}^{(2\alpha+1)/\alpha}}{\theta^{1/\alpha}} (1 + o(1)) \leq \sigma^2 (1 + o(1)) \quad \text{a.s.}, \\ L^2 \bar{b}_{K,T}^2 N_T(\bar{b}_{K,T}) &= 2\sqrt{T}L^2 \lambda(0, 1) \frac{\bar{b}_{K,T}^{(2\alpha+1)/\alpha}}{\theta^{1/\alpha}} (1 + o(1)) \geq \sigma^2 (1 + o(1)) \quad \text{a.s.}, \end{aligned}$$

and then, since $R_T = 1/(L\hat{b}_T)$ with \hat{b}_T as in (2.3),

$$\lim_{T \rightarrow \infty} \mathbb{P} \left(\frac{T^{-\alpha/(4\alpha+2)}}{L\underline{b}_{K,T}(1+\epsilon)} \leq T^{-\alpha/(4\alpha+2)} R_T \leq \frac{T^{-\alpha/(4\alpha+2)}}{L\bar{b}_{K,T}(1-\epsilon)}, E_K \right) = 1 \quad \text{for any } \epsilon > 0.$$

Hence (B.2.4), (B.2.3), $\lim_{T,K \rightarrow \infty} \mathbb{P}(E_K) = 1$ and definition of b_T yields Proposition 2. \square

REFERENCES

- AKONOM, J. (1993). Comportement asymptotique du temps d'occupation du processus des sommes partielles. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **29**, 57–81.
- BERLINET, A. and G. BIAU (2001). Minimax bounds in nonparametric estimation of multidimensional deterministic dynamical systems. *Statistical Inference for Stochastic Processes* **4**, 229–248.
- BLANKE, D. (2004). Sample paths adaptive density estimation. Forthcoming in *Mathematical Methods of Statistics*.
- BOSQ, D. (1998). *Nonparametric statistics for stochastic processes. Estimation and prediction. Lecture Notes in Statistics* **110**. Springer-Verlag, second edition.
- CHEN, X. (1999a). The law of the iterated logarithm for functionals of Harris recurrent Markov chains: self normalization. *Journal of Theoretical Probability* **12**, 421–445.
- CHEN, X. (1999b). How often does a Harris recurrent Markov chain recur ? *The Annals of Probability* **27**, 1327–1346.
- CHOW, Y.S. and H. TEICHER (1988). *Probability theory. Independence, interchangeability, martingales*. Second edition, Springer-Verlag.
- DELATTRE, S, M. HOFFMANN and M. KESSLER (2002). Dynamics adaptive estimation of a scalar diffusion. Prépublication n762, LPMA, Universités Paris 6 et 7.
- DELECROIX, M. (1987). Sur l'estimation et la prévision non-paramétrique des processus ergodiques. Doctorat d'Etat, Université de Lille-Flandres-Artois.
- FAN, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998–1004.
- FAN, J. and Q. YAO (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer-Verlag.
- GUERRE, E. (2000). Design adaptive nearest neighbor regression estimation. *Journal of Multivariate Analysis* **75**, 219–244.
- GUERRE, E. and J. MAËS (1998). Optimal rate for nonparametric estimation in deterministic dynamical systems. *Statistical Inference for Stochastic Processes* **1** 157–173.
- GYÖRFI, L., W. HÄRDLE, P. SARDA and P. VIEU (1989). *Nonparametric curve estimation from time series. Lecture Notes in Statistics* **60**. Springer-Verlag.
- HOFFMANN, M. and O. LEPSKI (2002). Random rates in anisotropic regression. With discussions and a rejoinder by the authors. *The Annals of Statistics* **30**, 325–396.
- KARLSEN, H.A. and D. TJØSTHEIM (2001). Nonparametric estimation in null recurrent time series. *The Annals of Statistics* **29**, 372–416.
- MEYN, S.P. and R.L. TWEEDIE (1993). *Markov chains and stochastic stability*. Springer-Verlag.
- MORVAI, G., S. YAKOWITZ and L. GYÖRFI (1996). Nonparametric inference for ergodic, stationary time series. *The Annals of Statistics* **24**, 370–379.
- NUMMELIN, E. (1984). *General irreducible Markov chains and non-negative operators*. Cambridge University Press.
- OREY, S. (1959). Recurrent Markov chains. *Pacific Journal of Mathematics* **9**, 805–827.
- PHILLIPS, P.C.B. and J.Y. PARK (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation discussion paper 1181, University of Yale.

- REVUZ, D. and M. YOR (1991). *Continuous martingales and Brownian motion*. Springer-Verlag.
- ROBINSON, P.M. (1997). Large sample inference for nonparametric regression with dependent errors. *The Annals of Statistics* **25**, 2054–2083.
- ROUSSAS, G.G. (1969). Non-parametric estimation of the transition distribution of a Markov process. *Annals of the Institute of Statistical Mathematics* **21**, 73–87.
- SPOKOINY, V.G. (2000). Adaptive drift estimation for nonparametric diffusion model. *The Annals of Statistics* **28**, 815–836.
- STONE, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* **8**, 1348–1360.
- TJØSTHEIM, D. (1994). Non-linear time series: A selective review. *The Scandinavian Journal of Statistics* **21**, 97–130.
- TONG, H. (1990). *Non-linear time series. A dynamical system approach*. Oxford University Press.
- YAKOWITZ, S. (1993). Nearest neighbor regression estimation for null-recurrent Markov time series. *Stochastic Processes and their Applications* **48**, 311–318.
- YAKOWITZ, S., L. GYÖRFI, J. KIEFFER and G. MORVAI (1999). Strongly consistent nonparametric forecasting and regression for stationary ergodic sequences. *Journal of Multivariate Analysis* **71**, 24–41.