Design-Adaptive Pointwise Nonparametric Regression Estimation for Recurrent Markov Time Series

E. GUERRE

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

1 LSTA-Université Paris 6 and CREST.
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 coincide 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 impelled 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 \( \beta \)-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 \( \beta \)-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 \(<\cdot,\cdot>\) and norm \(\|\cdot\|\). Consider \(T+1\) observations \(X_0,\ldots,X_T\) from a time series valued in \(\mathcal{X}\), with

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

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

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

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 $\varepsilon_t = Y_t - \mathbb{E}[Y_t|Y_{t-1}, \ldots, Y_{t-d}]$ assuming that $\text{Var}(\varepsilon_t|Y_{t-1}, \ldots, Y_{t-d}) \leq \sigma^2$. Under time homogeneity, the process admits an autoregressive representation $Y_t = \mu(Y_{t-1}, \ldots, Y_{t-d}) + \varepsilon_t$, where the conditional mean $\mu(Y_{t-1}, \ldots, Y_{t-d}) = \mathbb{E}[Y_t|Y_{t-1}, \ldots, Y_{t-d}]$ is the parameter of interest. Taking $\mathcal{X} = \mathbb{R}^d$ equipped with the usual Euclidean norm and $X_t = (Y_{t}, \ldots, Y_{t-d+1})'$ yield that $\{X_t, t \geq 0\}$ is Markov with

$$X_t = m(X_{t-1}) + e_t$$

where $m(X_{t-1}) = (\mu(Y_{t-1}, \ldots, Y_{t-d}), Y_{t-1}, \ldots, Y_{t-d+1})'$,
and $e_t = (e_t, 0, \ldots, 0)'$. 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 $N_T(b)$,

$$\mathbb{E} \left[ \sum_{t=1}^{T} e_t \mathbb{I}(\|X_{t-1} - x\| \leq b) \right]^2 = \sum_{t=1}^{T} \mathbb{E} \left[ \mathbb{I}(\|X_{t-1} - x\| \leq b) \mathbb{E}[\|e_t\|^2|X_{t-1}, \ldots, X_0] \right] + 2 \sum_{t=1}^{T} \sum_{t' = t+1}^{T} \mathbb{E} \left[ \mathbb{I}(\|X_{t-1} - x\| \leq b) \mathbb{I}(\|X_{t'-1} - x\| \leq b) \langle e_t, \mathbb{E}[e_{t'}|X_{t'-1}, \ldots, X_0] \rangle \right].$$
\[ \leq \sigma^2 \mathbb{E} N_T(b). \]

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
\[
m_T(x; b) - m(x) = O_p \left( \left( L^2b^2 + \frac{\sigma^2}{N_T(b)} \right)^{1/2} \right).
\]

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 \( X_0, \ldots, X_T \) 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\((L^2b^2, \sigma^2/N_T(b))\) and to propose
\[
\tilde{b}_T = \hat{b}_T(x; L, \sigma) = \arg\min_{b \geq 0} \left( L^2\tilde{b}_T^2, \frac{\sigma^2}{N_T(b)} \right) \text{ i.e., for instance,}
\]
\[
(2.3) \quad \tilde{b}_T = \min \left\{ b \geq 0; L^2b^2 N_T(b) \geq \sigma^2 \right\} \text{ so that } \min_{b \geq 0} \left( L^2\tilde{b}_T^2, \frac{\sigma^2}{N_T(b)} \right) = L^2\tilde{b}_T^2.
\]

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

For the choice (2.3) of \( \tilde{b}_T \), one would expect that the order for the estimation error (2.2) is
\[
(2.4) \quad R_T = R_T(x; L, \sigma) = \frac{1}{\left( \min_{b \geq 0} \max \left( L^2b^2, \frac{\sigma^2}{N_T(b)} \right) \right)^{1/2}} = \frac{1}{L\tilde{b}_T}.
\]

However, the order \( m(x; \tilde{b}_T) - m(x) = O_p(1/R_T) \) cannot be directly derived from (2.2) which does not hold for a data-driven bandwidth as \( \tilde{b}_T \). 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 \( m(x; \tilde{b}_T) - m(x) \) 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

\begin{equation}
\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)}.
\end{equation}

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{align*}
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) = \left( \frac{\sum_{t=1}^{T} K \left( \frac{\|X_{t-1} - x\|}{h} \right) \left( \frac{\|X_{t-1} - x\|}{h} \right)}{\sum_{t=1}^{T} K^2 \left( \frac{\|X_{t-1} - x\|}{h} \right)} \right)^2 = \left( \frac{\sum_{t=1}^{T} K \left( \frac{\|X_{t-1} - x\|}{h} \right) \left( \frac{\|X_{t-1} - x\|}{h} \right)}{N_T(h; K)} \right)^2,
\end{align*}

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

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

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

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

\[
R_T(K) = R_T(x; L, \sigma, \kappa_0, K) = \frac{1}{(L^2 B_T^2(\hat{h}T) + \sigma^2/V_T(\hat{h}T))^{1/2}}.
\]

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 \( N_T(h; K) \leq N_T(h; K) \), \( \mu_T(x; \hat{h}_{0T}) \) 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

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, \( (v_1, \ldots, v_d) \) is the canonical basis and \( v_{i+j} = 0, j \geq 1 \). 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 \( \mathcal{P}_D(L, \sigma) \) of admissible time series involves some additional quantities. Consider a probability measure \( \mathcal{P} \) over \( \mathcal{X} \), with \( \mathcal{P}(\mathcal{D}) = 1 \) and

\[
\mathcal{F}(h) = \inf_{x \in \mathcal{D}} \mathcal{P} \left( \{ y \in \mathcal{D} : \| y - x \| \leq h \} \right) \text{ is continuous,}
\]

\[
\text{with } \mathcal{F}(h) > 0 \text{ for } h > 0 \text{ and } \lim_{h \to \infty} \mathcal{F}(h) = 1.
\]
Recall that \( e_t = \sum_{j=1}^{\infty} e_t, v_j > v_j \). The standard deviation \( \sigma \) 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 \( \sigma \), consider \( \rho \) in \([0, 1]\), an integer number \( d \geq 1 \) as in Example 2, and a positive sequence \( n_T \) which typically diverges with a slow rate as \( \ln T \).

**Definition 1.** The class \( \mathcal{P}_D(L, \sigma) = \mathcal{P}_D(L, \sigma; \rho, P_d, n_T, \kappa_2) \) 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

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

\[
\mathbb{E}_x N_T(\mathcal{D}) \geq n_T \text{ for all } T \geq 0.
\]

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

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

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

The following condition on \( P(\cdot) \) and \( \rho \) allows to consider a subclass \( \mathcal{P}_D(L, \sigma) \subset \mathcal{P}_D(L, \sigma) \) 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 \( P(\cdot) \) and \( \rho \) in Definition 1-(iv) are such that, for any Borel subset \( A \) of \( \mathcal{D} \), \( \inf_{[m] \leq \kappa_{n+1}} P_g (m + e_t \in A) \geq (1 - \rho) P(A) \).

Under Assumption P, a choice of \( P(\cdot) \) is a uniform distribution since the Gaussian density is positive continuous. The quantities \( \rho, P, P_d, n_T \) and \( \kappa_2 \) in Definition 1 need not to be known and are only used to obtain rate-consistency uniformly over \( \mathcal{P}_D(L, \sigma) \). 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 \( Y_t = \mu(Y_{t-1}, \ldots, Y_{t-d}) + \varepsilon_t \) for \( t \geq d \). Let \( \{X_t, t \geq 0\} \) with \( X_t = (Y_t, \ldots, Y_{t-d+1})' \) and \( m(\cdot) \) be as in Example 2 with \( X = \mathbb{R}^d \). Assume that

(a) The distribution of \( \varepsilon_t \) given \( (Y_{t-1}, \ldots, Y_{t-d}) = X_{t-1} = x \) has a probability density function \( f(\varepsilon_t|x) \) which is bounded away from 0 and continuous with respect to \( (\varepsilon_t, x) \) 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 \( \lim_{||x|| \to \infty} U(x) = \infty \), an integer number \( d' > 0 \) and a compact subset \( C \) of \( \mathbb{R}^d \) such that \( \mathbb{E}_D[U(X_{t+d'})|X_t = x] \leq U(x) \) for all \( x \) in \( \mathbb{R}^d \setminus C \).

Let \( D \) be any bounded subset of \( \mathbb{R}^d \). Then the Continuity Condition (3.2) holds under (a) and (b), and there exists a distribution \( \mathcal{P} \) satisfying (3.1) and a \( \rho \) in \( [0,1) \) such that \( \{X_t, t \geq 0\} \) verifies the Minorization Condition of Definition 1-(iv). Under (a), (b) and (c), \( \mathbb{E}_F N_T(D) \) 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 \( ||m(x)||^2 + \text{Var}_\mathcal{P}[X_t | X_{t-1} = x] \leq ||x||^2 \), 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 \( \mathbb{P}(\lim_{T \to \infty} N_T(D) = \infty|X_0 = x) = 1 \) 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 \( N_T(D) \) has a deterministic exact order which diverges, as done in Theorem 1.

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

\[
\lim_{T, z \to +\infty} \inf_{\mathcal{P} \in \mathcal{P}(L, s)} \mathbb{P}\left( \frac{n_T(\mathcal{P})}{z} \leq N_T(D) \leq n_T(\mathcal{P}) z \right) = 1.
\]

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

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

THEOREM 3. Assume that the kernel function \( K(\cdot) \) satisfies Assumption K. Let \( \mu_T(x; h) = \mu_T(x; h, K) \), \( \hat{b}_T = \hat{b}_T(x; L, \sigma, \kappa_0, K) \) 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 \( D \), \( R_T(K) \left( \mu_T(x; \hat{b}_T) - m(x) \right) \) is bounded in probability uniformly over \( \mathcal{P}_D(L, \sigma) \), and the random rates \( R_T \) and \( R_T(K) \) are of the same order i.e.

\[
\lim_{T; z \to +\infty} \sup_{P \in \mathcal{P}_D(L, \sigma)} \mathbb{P} \left( R_T(K) \left\| \mu_T(x; \hat{b}_T) - m(x) \right\| \geq z \right) = 0
\]

(3.4)

\[
\text{and } \frac{\kappa_1}{2\sqrt{2}} R_T \leq R_T(K) \leq \sqrt{\kappa_0} R_T \text{ for any } x \in D.
\]

(3.5)

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 \( m_T(x; \hat{b}_T) \) averages over values of \( X_{t-1} \) in \( D \) for \( \hat{b}_T \) small enough. Hence the proof of Theorems 2 and 3 rely on the so-called “process on \( D \)” of Orey (1959) given by the successive values of \( \{X_{t-1}, t \geq 1\} \) in \( D \). As recalled in Lemma A.1 in Appendix A, the process on \( D \) is \( \Phi \)-mixing with exponential decay. This allows to find the exact order of the bandwidth \( \hat{b}_T \) when the bound for \( N_T(D) \) in Theorem 1 holds. This gives the order of \( R_T = 1/(L\hat{b}_T) \) which is used to bound the stochastic error term of \( m_T(x; \hat{b}_T) - m(x) \) using a maximal inequality, leading to Theorem 2. Theorem 3 is proven similarly using the restrictions in (2.7) for \( \hat{h}_T \). Compared to Karlsen and Tjøstheim (2001), Theorems 2 and 3 do not restrict the dynamics of the time series to \( \beta \)-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 \( \mathcal{P}_D(L, \sigma) \), 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 \( D \), 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 \( m_T(x; \hat{b}_T) \) and \( \mu_T(x; \hat{b}_T) \) 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, $m_T(x; \hat{h}_T)$ and $\mu_T(x; \hat{h}_T)$ converge to $m(x)$ with the same rate $R_T$. But the upper bound $R_T(K) \leq \sqrt{n} R_T$ 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 next two results show that $R_T$ is the minimax optimal random estimation rate of $m(x)$ over $\mathcal{P}_D(L, \sigma)$. Theorem 4 below adopts a local minimax framework. Define, for any $P_0$ in $\mathcal{P}_D(L, \sigma)$, $x$ in $D$ and $\epsilon > 0$,

$$V_x(P_0; \epsilon) = \{ P \in \mathcal{P}_D(L, \sigma); \| m(x) - m_0(x) \| \leq \epsilon \},$$

where $m(\cdot) = \mathbb{E}_P[X_t | X_{t-1} = \cdot]$ and $m_0(\cdot) = \mathbb{E}_{P_0}[X_t | X_{t-1} = \cdot]$. 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. $N(0, \sigma^2)$ innovation terms $\{ \epsilon_t, t \geq 0 \}$. Let $P_0$ be the distribution of $\{ X^0_t, t \geq 0 \}$ with $X^0_t = \epsilon_0$ and $X^0_t = m_0(X^0_{t-1}) + \epsilon_t$, where the regression function $m_0(\cdot)$ is continuous over $\mathbb{R}$ and $L/2$-Lipschitz over $D$, with $\sup_{x \in D} |m_0(x)| \leq \kappa_3$ for $\kappa_3$ as in Assumption $P$. Assume also that there is a real continuous function $U(\cdot) \geq 0$ with $\lim_{|x| \to \infty} U(x) = +\infty$ such that

$$\mathbb{E}_{P_0} [U(X^0_{t+1}) | X^0_t = x] \leq U(x) \text{ for all } x \in \mathbb{R} \setminus D',$$

where $D'$ is a finite closed interval containing $D$.

Then there exists a diverging sequence $n_T > 0$ such that $P_0$ is in $\mathcal{P}_D(L, \sigma)$, and there is a $z > 0$ such that, for any $x$ in $D$,

$$\lim_{T \to +\infty} \inf \sup_{P \in V_x(P_0; \epsilon)} \mathbb{P} (R_T \| \hat{m}_T(x) - m(x) \| \geq z) > 0 \text{ for any } \epsilon > 0,$$

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

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 $V_x(P_0; \epsilon)$ 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 $\mathcal{P}_D(L, \sigma)$.

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

$$\lim_{T \to +\infty} \inf \sup_{P \in \mathcal{P}_D(L, \sigma)} \mathbb{P} (R_T \| \hat{m}_T(x) - m(x) \| \geq z) > 0 \text{ for all } x \in D.$$
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 $D = (-1, 1)$. We first recall some results of Guerre (2000) for the regression model $Y_t = m(X_t) + \epsilon_t$. Consider the modified Box-Cox transformation

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

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 $(x/\theta)^{1/\alpha-1}I_{[0,\theta]}(x)/(\theta\alpha)$ 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 $\alpha$. Both the expression of $H_{\alpha,\theta}(\cdot)$ and of the design density are helpful to understand how $\alpha$ and $\theta$ affect the design repartition in the vicinity of 0. Indeed, for $x \in D$, $\lim_{\alpha \to \infty} H_{\alpha,\theta}(x) = 0$ while $\lim_{\alpha \to 0} H_{\alpha,\theta}(x) = \theta \text{sgn}(x)$. Therefore \{X_t, t \geq 0\} visits small neighborhoods of 0 more frequently for large $\alpha$ and small $\theta$. 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

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

(4.1) $X_t = H_{\alpha,\theta}(Y_t)$ 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

$$m(x) = \mathbb{E}[X_t | X_{t-1} = x] = \mathbb{E}\left[H_{\alpha, \theta}\left(H_{1/\alpha, \theta^{-1/\alpha}}(x) + g_t\right)\right]$$

$$= \int_0^x H'_{1/\alpha, \theta^{-1/\alpha}}(z)\mathbb{E}\left[H'_{\alpha, \theta}\left(H_{1/\alpha, \theta^{-1/\alpha}}(x) + g_t\right)\right] dz$$

$$m'(x) = H'_{1/\alpha, \theta^{-1/\alpha}}(x)\mathbb{E}\left[H'_{\alpha, \theta}\left(H_{1/\alpha, \theta^{-1/\alpha}}(x) + g_t\right)\right],$$

which gives $\sup_{x \in \mathcal{D}} |m'(x)| < \infty$ for $\alpha \leq 1$, and,

$$\sup_{x \in \mathcal{D}} \mathbb{E}[|e_t|^2 | X_{t-1} = x] \leq 2^{2p-1} \left(\sup_{x \in \mathcal{D}} \mathbb{E}[|X_t|^2 | X_{t-1} = x] + \sup_{x \in \mathcal{D}} |m(x)|^{2p}\right) < \infty,$$
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 \( T^{1/6} \), which is slower than the rate \( T^{1/3} \) achieved for weak dependent processes due to a lower recurrence rate \( \sqrt{T} \) for \( N_T(D) \) 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 \( N_T(D) \), 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 \( 1/(N_T(D)b)^{1/2} \) with a bias bounded by \( b \) time a constant. Balancing this two terms suggests a bandwidth of order \( N_T^{-1/3}(D) \) 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 \( 1/(N_T(D)b^{1/\alpha})^{1/2} \) for small \( b \), as seen from (B.2.3) in Appendix B, instead of \( 1/(N_T(D)b)^{1/2} \). It follows that a bandwidth of order \( N_T^{-1/3}(D) \) gives a rate \( T^{-1/6} \) which is, for \( \alpha < 1 \), smaller than the rate \( T^{(3\alpha-1)/12\alpha} \) 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 \( \hat{b}_T \), 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 \( N_T^{-1/3}(D) \) 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 \( D \), i.e.

\[
\tau_i = \inf \{ t \geq 1, X_{t-1} \in D \}, \ldots, \tau_{i+1} = \inf \{ t > \tau_i; X_{t-1} \in D \}.
\]
and let \( x_{i-1} = X_{r_{i-1}} \) 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 \( \eta_i = e_{r_{i-1}} \) and let \( \mathcal{F}_i \) be the Borel field generated by \( x_{i-1}, \eta_{i-1}, \ldots, x_0, \eta_0 \). Set

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

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

\[
\Phi_k = \Phi_k(\mathbb{P}) = \max_{n \geq 0} \mathbb{P}(x_{n+k} \in B|x_n \in A) - \mathbb{P}(x_{n+k} \in B) .
\]

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 \( K(z) = I(z \leq 1) \). For \( A \subset \mathbb{R} \), define

\[
K_A^2(x_{i-1}) = K_A^2(x_{i-1}; x) = -\int_0^1 I(\frac{\|x_{i-1} - x\|}{\beta}) dK^2(z)
\]

where the negative differential term \( dK^2(z) \) is minus the Dirac mass at 1 if \( K(z) = I(z \leq 1) \), in which case \( K_A(x_{i-1}) = \mathbb{I}(\|x_{i-1} - x\| \in A) \). Note that \( K^2(||x_{i-1} - x||/h) = K_t^2(||x_{i-1} - x||) = -\int_0^1 I(\|x_{i-1} - x\| \leq z h) dK^2(z) \), which increases with \( h \) under Assumption K. Since \( x \) is in \( \mathcal{D} \) which is open, the regression estimators \( m_T(x; \hat{b}_T), \mu_T(x; \hat{h}_T) \) and \( m_T(x; \hat{b}_T) \) averages over the \( X_{i-1}'s \) in \( \mathcal{D} \), that is over the \( x_{i-1}'s \), when \( \hat{b}_T, \hat{h}_T \) 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 \( \hat{h}_T \) defined with respect to \( \{ x_i, i \geq 0 \} \). For \( \beta \in \mathbb{R} \) and \( A \subset \mathbb{R} \), define

\[
u_\beta(N; x, K) = \sum_{i=1}^n K^2(\frac{\|x_{i-1} - x\|}{\beta}) , \quad \nu_\beta(A) = \sum_{i=1}^n K^2(\|x_{i-1} - x\|)
\]

noticing that \( \nu_\beta(-\infty, \beta) = \nu_\beta(\beta) \). Observe that \( N_T(\beta; K) = \nu_{N_T,D}(\beta) \) for \( \beta \) small enough. The counterpart of the stochastic term is \( \Sigma_\beta(A)/\nu_\beta(\beta) \) with

\[
A.1.2 \quad \Sigma_\beta(A) = \sum_{i=1}^n K(\frac{\|x_{i-1} - x\|}{\beta}) \eta_i .
\]

The counterpart of the bandwidths \( \hat{b}_T \) and \( \hat{h}_T \) is

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

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 \( K(z) = I(z \leq 1) \), 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 \( \hat{h}_T \) yields that \( \hat{b}_T = \hat{b}_{N_T,D} \) if \( K(z) = I(z \leq 0) \) and \( \hat{h}_T = \hat{h}_{N_T,D} \) as soon as \( \hat{b}_{N_T,D}, \) or \( \hat{b}_T \) and \( \hat{h}_T \), are small enough.

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

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

\[
A.1.5 \quad \mathbb{F}_n(\beta) = \frac{1}{n} \sum_{i=1}^n F_i(\beta) = \mathbb{E}_x [\nu_\beta(\beta)/n] , \quad F_i(A) = \mathbb{E}_x K^2_A(\|x_{i-1}\|) , \quad \mathbb{F}_n(A) = \frac{1}{n} \sum_{i=1}^n F_i(A) .
\]
Under Assumption K or if $K(z) = I(z \leq 1)$, $F_i(\cdot)$ and $\mathcal{F}_n(\cdot)$ are cumulative distribution functions. With the same convention as in (A.1.3), set

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

In what follows, $\lfloor z \rfloor$ 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 $\mathcal{P}_D(L, \sigma)$. 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 $K(z) = I(z \leq 1)$ or satisfies Assumption K. Let \{ $F_i(\cdot) = F_i(\cdot; x, \mathbb{P}, K), i \geq 1$ \} and \{ $\mathcal{F}_n(\cdot) = \mathcal{F}_n(\cdot; x, \mathbb{P}, K), n \geq 1$ \} be as in (A.1.4)-(A.1.5). For any distribution $\mathbb{P}$ in $\mathcal{P}_D(L, \sigma)$, we have:

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

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

ii. The $\Phi$-mixing coefficients of $\{x, i \geq 0\}$ 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 $F(\beta) \geq (\rho^d - 1) \int_0^1 F(\beta z) dK^2(z)$. The $\mathcal{F}_n(\cdot)$’s are continuous with $\mathcal{F}_n(\beta) > 0$ for all $\beta > 0$ and $n \geq d + 1$. Moreover

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

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) \[ \text{Var}_P \left( \sum_{i=n_1+1}^{n_2} K_i^2(x_{i-1}) \right) \leq \left( 1 + \frac{4 \rho}{1-\rho} \right) (n_2 \mathcal{F}_{n_2}(A) - n_1 \mathcal{F}_{n_1}(A)) , \]

(A.2.2) \[ \mathbb{E}_P^{1/3} \left( \sum_{i=n_1+1}^{n_2} K_i^2(x_{i-1}) \right)^3 \leq C_1 \left( (1 + n_2 \mathcal{F}_{n_2}(A) - n_1 \mathcal{F}_{n_1}(A))^3 - 1 \right)^{1/3} , \]

(A.2.3) \[ \mathbb{E}_P^{2} \left( \sum_{i=n_1+1}^{n_2} K_i(x_{i-1}) \eta_i \right)^6 \leq C_1 \sigma \left( (1 + n_2 \mathcal{F}_{n_2}(A) - n_1 \mathcal{F}_{n_1}(A))^3 - 1 \right)^{\frac{3}{2}} . \]

**Proof of Lemma A.1:** (i) follows from Definition 1-(iii), the Strong Markov Property, $\mathbb{E}[\omega_i^2(j)|\mathcal{F}_i] \leq \mathbb{E}[\omega_i^2(j)|\mathcal{F}_1], \mathbb{E}[\zeta_i^2(j)|\mathcal{F}_i] \leq \mathbb{E}[\eta_i^2(j)|\mathcal{F}_1] \leq \mathbb{E}[\eta_i^2(j)|\mathcal{F}_1] + 6 \mathbb{E}[\eta_i^2(j)|\mathcal{F}_1] \mathbb{E}[\eta_i^2(j)|\mathcal{F}_1] + \mathbb{E}[\eta_i^2(j)|\mathcal{F}_1].$

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

$$
P_{\mathcal{D}}(x, A) = P(x, A) + \int_{X \setminus \mathcal{D}} P(x, dx_1) P(x_1, A) + \int_{X \setminus \mathcal{D}} \int_{X \setminus \mathcal{D}} P(x, dx_1) P(x_1, dx_2) P(x_2, A) + \cdots \geq P(x, A) ,
$$
which implies $P_D(x, dy) \geq P(x, dy)$ for (almost) all $x, y$ in $\mathcal{D}$. Hence

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

and then, by Definition 1-(iv)

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

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

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

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

$$I(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_D(x_{n+k} \in B) + \mathbb{P}(x_{n+k} \in B) - \pi_D(x_{n+k} \in B) |$

$$= \mathbb{P}(x_{n+k} \in B | x_n = x) - \pi_D(x_{n+k} \in B) \mathbb{P}(x_n \in dx) |$

$$+ \int_{x \in \mathcal{D}} (\mathbb{P}(x_{n+k} \in B | x_n = x) - \pi_D(x_{n+k} \in B) \mathbb{P}(x_n \in dx) | \leq 2\rho^k ,$$

and then $\Phi_k \leq 2\rho^k$. Set $p_i(\beta) = \mathbb{P}(\|x_{i-1} \| \leq \beta)$, $p_n(\beta) = \sum_{i=1}^{n} p_i(\beta)/n$ and $p(\beta) = \mathbb{P}(\|x_{i-1} \| \leq \beta),$

so that $F_i(\beta) = -\int_0^1 p_i(\beta) dK^2(z)$ and $\mathcal{F}_n = -\int_0^1 p_n(\beta) dK^2(z)$ by (A.1.4). We have $p(\beta) = \int_{\mathcal{D}} \mathbb{P}(\|x_{i+d} \| \leq \beta | x_1 = x) \mathbb{P}(x_1 \in dx) \geq (1 - \rho^d)\mathbb{P}(\|x_1 \| \leq \beta) \geq (1 - \rho^d)\mathbb{P}(\|x_1 \| \leq \beta) \geq (1 - \rho^d)\mathbb{P}(\|x_{i+d} \| \leq \beta)$ by (A.2.4) and (3.1), which gives $F(\beta) = -\int_0^1 p(\beta) dK^2(z) \geq (\rho^d - 1)\int_0^1 \mathcal{F} dK^2(z)$. (3.2) yields that the $p_i(\cdot)$’s are continuous. Hence the $F_i(\cdot)$’s and $\mathcal{F}_n(\cdot)$’s are continuous. (A.2.4) gives

$$p_{i+d+1}(\beta) = \mathbb{P}(\|x_{i+d+1} \| \leq \beta) = \int_{\mathcal{D}} \mathbb{P}(\|x_{i+d+1} \| \leq \beta | x_1 = x) \mathbb{P}(x_1 \in dx) \geq (1 - \rho^d)\mathbb{P}(\|x_{i+d+1} \| \leq \beta) ,$$

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

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

(iiv) 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

$$K^2_i(x_{i-1}) = -\int_0^1 y_i(z) dK^2(z) , \quad F_i(A) = -\int_0^1 p_i(z) dK^2(z) ,$$

$$\left| \text{Cov} \left( K^2_i(x_{i-1}), K^2_j(x_{j-1}) \right) \right| \leq \int_0^1 \int_0^1 \left| \text{Cov} \left( y_i(z), y_{j-1}(z') \right) \right| dK^2(z) dK^2(z') ,$$

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

$$\left| \text{Cov} \left( y_i(z), y_{j}(z') \right) \right| = \left| \mathbb{P}(y_i(z') = 1 | y_i(z) = 1) - p_j(z') \right| p_i(z) \leq \Phi(j - i)p_i(z) .$$
Since \( \int_0^1 dK^2(z) = 1 \), it follows that
\[
\operatorname{Var}(\sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1})) \leq \sum_{i=n_1+1}^{n_2} \operatorname{Var}(K_A^2(x_{i-1})) + 2 \sum_{i=n_1+1}^{n_2} \sum_{j=i+1}^{n_2} |\operatorname{Cov}(K_A^2(x_{i-1}), K_A^2(x_{j-1}))|
\leq - \sum_{i=n_1+1}^{n_2} \int_0^1 p_i(z)dK^2(z) + 2 \sum_{i=n_1+1}^{n_2} \int_0^1 \int_0^1 p_i(z)dK^2(z)dK^2(z') \sum_{k=1}^{\infty} \Phi(k)
\leq \left(1 + \frac{4\rho}{1 - \rho}\right) \sum_{i=n_1+1}^{n_2} F_i(A),
\]
which gives (A.2.1). For (A.2.2), set \( I_3 = \{(i_1, i_2, i_3); n_1 + 1 \leq i_1 < i_2 < i_3 \leq n_2\} \) and \( I_2 = \{(i_1, i_2); n_1 + 1 \leq i_1 < i_2 \leq n_2\} \). This gives
\[
\mathbb{E}_p \left| \sum_{i=n_1+1}^{n_2} K_A^2(x_{i-1}) \right|^3 = \mathbb{E}_p \left| \int_0^1 \left( \sum_{i=n_1+1}^{n_2} y_i(z) \right) dK^2(z) \right|^3 \leq C \left[ - \sum_{j=1}^{n_2} \int_{(0,1]^3} \mathbb{E}_p \left( \prod_{j=1}^{3} y_{i_j}(z_j) dK^2(z_j) \right) \right.
\left. + \sum_{j=1}^{n_2} \int_{(0,1]^2} \mathbb{E}_p \left( \prod_{j=1}^{2} y_{i_j}(z_j) dK^2(z_j) \right) - \sum_{i=n_1+1}^{n_2} \int_0^1 \mathbb{E}_p (y_i(z)) dK^2(z) \right].
\]
We have, for \( i_1 < i_2 < i_3 \),
\[
\mathbb{E}_p(y_1(z_1)y_2(z_2)y_3(z_3)) = P(y_{i_1}(z_1) = 1|y_{i_2}(z_2) = 1, y_{i_3}(z_3) = 1)P(y_{i_2}(z_2) = 1|y_{i_1}(z_1) = 1)p_{i_1}(z_1)
\leq (p_{i_3}(z_3) + \Phi(i_3 - i_2)) (p_{i_2}(z_2) + \Phi(i_2 - i_1)) p_{i_1}(z_1)
= p_{i_1}(z_1)p_{i_2}(z_2)p_{i_3}(z_3) + p_{i_1}(z_1)p_{i_2}(z_2)\Phi(i_3 - i_2)
+ p_{i_1}(z_1)p_{i_3}(z_3)\Phi(i_2 - i_1) + p_{i_1}(z_1)\Phi(i_3 - i_2)\Phi(i_2 - i_1),
\]
\[
\mathbb{E}_p(y_1(z_1)y_2(z_2)) = P(y_{i_2}(z_2) = 1|y_{i_1}(z_1) = 1)p_{i_1}(z_1) \leq p_{i_2}(z_2)p_{i_1}(z_1) + p_{i_1}(z_1)\Phi(i_2 - i_1).
\]
Summing gives
\[
\sum_{i=1}^{n_2} p_{i_1}(z_1)p_{i_2}(z_2)\Phi(i_3 - i_2) \leq 2 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} p_{i_2}(z_2) \sum_{i_3=i_2+1}^{n_2} \rho^{i_3-i_2}
\leq \frac{2\rho}{1 - \rho} \sum_{i=n_1+1}^{n_2} p_i(z_1) \sum_{j=i+1}^{n_2} p_j(z_2),
\]
\[
\sum_{i=1}^{n_2} p_{i_1}(z_1)p_{i_2}(z_2)\Phi(i_2 - i_1) \leq 2 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} \rho^{i_2-i_1} \sum_{i_3=i_2+1}^{n_2} p_{i_3}(z_3)
\leq \frac{2\rho}{1 - \rho} \sum_{i=n_1+1}^{n_2} p_i(z_1) \sum_{j=i+1}^{n_2} p_j(z_3),
\]
\[
\sum_{i=1}^{n_2} p_{i_1}(z_1)\Phi(i_2 - i_1)\Phi(i_3 - i_1) \leq 4 \sum_{i_1=n_1+1}^{n_2} p_{i_1}(z_1) \sum_{i_2=i_1+1}^{n_2} \rho^{i_2-i_1} \sum_{i_3=i_2+1}^{n_2} \rho^{i_3-i_2}
\leq \frac{4\rho^2}{(1 - \rho)^2} \sum_{i=n_1+1}^{n_2} p_i(z_1),
\]
\[
\sum_{i_2} p_{i_1}(z_2) \Phi(i_2 - i_1) \leq 2 \sum_{i_2 = n_1 + 1}^{n_2} p_{i_1}(z_1) \sum_{i_2 = n_1 + 1}^{n_2} \rho^{i_2 - i_1} \leq \frac{2\rho}{1 - \rho} \sum_{i_2 = n_1 + 1}^{n_2} p_{i_1}(z_1).
\]

It follows that
\[
\mathbb{E}_P \left| \sum_{i = n_1 + 1}^{n_2} K^2_A(x_{i-1}) \right|^3 \leq C \left( \sum_{n_1 + 1 \leq i_1, i_2 \leq n_2} F_{i_1}(A) F_{i_2}(A) F_{i_2}(A) + \sum_{n_1 + 1 \leq i_1, i_2 \leq n_2} F_{i_1}(A) F_{i_2}(A) + \sum_{i = n_1 + 1}^{n_2} F_i(A) \right)
\]
\[
= C \left[ (n_2 F_{n_2}(A) - n_1 F_{n_1}(A))^3 + (n_2 F_{n_2}(A) - n_1 F_{n_1}(A))^2 + n_2 F_{n_2}(A) - n_1 F_{n_1}(A) \right]
\]
\[
\leq C \left[ (1 + n_2 F_{n_2}(A) - n_1 F_{n_1}(A))^3 - 1 \right],
\]

which is (A.2.2). For (A.2.3), set \( S_j = \sum_{i = n_1 + 1}^{n_2} K_A(x_{i-1}) \eta_i(j) \), \( S = \sum_{i = n_1 + 1}^{n_2} K_A(x_{i-1}) \eta_i \), so that we have to bound \( \mathbb{E}_P^{1/6} ||S||^6 \). The triangular inequality gives
\[
\mathbb{E}_P^{1/6} ||S||^6 = \mathbb{E}_P^{1/6} \left( \sum_{j=1}^{n_1} S_j \right)^6 \leq \mathbb{E}_P^{1/6} \left( \sum_{j=1}^{n_1} S_j^2 \right)^3 \leq \mathbb{E}_P^{1/6} \left( \sum_{j=1}^{n_1} |S_j| \right)^6 \leq \mathbb{E}_P^{1/6} |S|^6,
\]

and we now bound the \( \mathbb{E}_P^{1/6} S_j^6 \)'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
\[
\mathbb{E}_P S_j^6 = \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K_A(x_{i-1}) \eta_i(j) \right)^6 \leq C \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^3_A(x_{i-1}) \eta_i(j) \right)^6
\]
\[
\leq C \left[ \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^2_A(x_{i-1}) \mathbb{E}_P[|\eta_i(j)| |F_i] \right)^3 + \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^2_A(x_{i-1}) \omega_i(j) \right)^3 \right]^{\frac{1}{3}}
\]
\[
\leq C \left[ \sigma^6 P \left( \sum_{i = n_1 + 1}^{n_2} K^2_A(x_{i-1}) \right)^3 + \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^4_A(x_{i-1}) \left( \mathbb{E}_P[|\omega_i(j)| |F_i] + |\zeta_i(j)| \right) \right)^{\frac{1}{3}} \right]
\]
\[
\leq C \left[ \sigma^6 P \left( \sum_{i = n_1 + 1}^{n_2} K_A(x_{i-1}) \right)^3 + \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^4_A(x_{i-1}) \right)^{\frac{1}{3}} + \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^4_A(x_{i-1}) \zeta_i(j) \right)^{\frac{1}{3}} \right]
\]
\[
\leq C \left[ \sigma^6 P \left( \sum_{i = n_1 + 1}^{n_2} K^2_A(x_{i-1}) \right)^3 + \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^4_A(x_{i-1}) \right)^{\frac{1}{3}} + \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^6_A(x_{i-1}) \zeta_i(j) \right)^{\frac{1}{3}} \right].
\]

Observe that \( 0 \leq K^2_A(\cdot) \leq 1 \) gives \( K^p_A(x_{i-1}) \leq K^3_A(x_{i-1}) \) for \( p = 2, 4 \), that \( \mathbb{E}[|Z|^{3/2}] = \mathbb{E}[|Z|^{3/2}(|Z| \leq 1) + \mathbb{E}[|Z|^{3/2}(|Z| > 1)] \leq \mathbb{E}[|Z|] + \mathbb{E}[|Z|^3] \), and that \( \mathbb{E}[|Z|] \leq \mathbb{E}[|Z|] + \mathbb{E}[|Z|^2] \). This together the Hölder inequality yield
\[
\mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K_A^2(x_{i-1}) \right) \leq \mathbb{E}_P \left( \sum_{i = n_1 + 1}^{n_2} K^2_A(x_{i-1}) \right)^{\frac{1}{2}} + \mathbb{E}_P K_A^2(x_{i-1}),
\]
Substituting then gives, using (A.2.2),
\[ \beta \text{ consequence of (3.1) which gives } \]
\[ \text{in } \{ \text{such that} \}
\[ \text{continuously increases from 0 to } 1 \text{ by (3.1). Lemma A.1-(iii) gives}
\]
\[ \text{then }
\]
\[ \text{E} \geq (8\kappa_2 \sigma_j)^6 \]  

\[ \text{over } j \text{ gives (A.2.3), using the bound of } E^{1/6} |S|^6, \text{ since } \sum_{j=1}^\infty \sigma_j \leq \sigma. \]

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

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

i. For any \( P \in P_D(L, \sigma) \) and \( x \in D \), \( \beta_n \) decreases with \( n \), \( L^2 n^{2\beta_n} F_n(\beta_n) = \sigma^2 \) for \( n \geq 1 \), and \( n F_n(\beta_n) \) is increasing.

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

\[ \sup_{P \in P_D(L, \sigma)} \beta_n \leq \overline{\beta}_n \text{ and } \inf_{P \in P_D(L, \sigma)} n F_n(\beta_n) \geq \frac{\sigma^2}{L^2 \overline{\beta}_n}. \]

iii. For \( z_0 > 0 \), set \( n_1 = \lfloor n/z_0 \rfloor \) and \( n_2 = \lfloor nz_0 \rfloor \). Then there exists a constant \( C_2 = C_2(z_0) > 0 \) such that

\[ \frac{C_2}{n_0} \leq \frac{F_n(\beta_1)}{n_2 F_n(\beta_2)} \leq 1, \quad \ell = n_1, \ldots, n_2, \]

for all \( n_1, n_2 \in P_D(L, \sigma) \) and \( x \in D \).

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

(ii). It is sufficient to prove the first inequality by (i). Set \( F_K(\beta) = (\rho^4 - 1) \int_0^1 K(\beta z) dK^2(z) \), which continuously increases from 0 to 1 by (3.1). Lemma A.1-(iii) gives \( F_n(\cdot) \geq F_K(\cdot) - 1/n(1-\rho) \) for all \( x \) in \( D \) and \( P \) in \( P_D(L, \sigma) \). Hence \( L^2 n^{2\beta_n} F_n(\beta) \geq L^2 n^{2\beta} (F_K(\beta) - 1/n(1-\rho)) \). Note that there is a \( \overline{\beta}_n \), which does not depend on \( P \), such that

\[ L^2 n^{2\overline{\beta}_n} \left( F_K(\overline{\beta}_n) - \frac{1}{n(1-\rho)} \right) = \sigma^2. \]

Then \( \beta_n \leq \overline{\beta}_n \) for all \( P \in P_D(L, \sigma) \) and \( L^2 n^{2\overline{\beta}_n} F_n(\overline{\beta}_n) \geq \sigma^2 \). That \( \lim_{n \to \infty} \overline{\beta}_n = 0 \) is a direct consequence of (3.1) which gives \( F_K(\cdot) > 0 \) over \( \mathbb{R}^+ \).
(iii). It is sufficient to bound \( n_1 \mathcal{F}_{n_1}(\beta_{n_1})/(n_2 \mathcal{F}_{n_2}(\beta_{n_2})) \) 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

\[
\frac{n_1 \mathcal{F}_{n_1}(\beta_{n_1})}{n_2 \mathcal{F}_{n_2}(\beta_{n_2})} = \frac{1}{\frac{1}{\beta}} \mathcal{F}_{n_1}(\beta_{n_1}) \geq \frac{1}{\frac{1}{\beta}} \mathcal{F}(\beta_{n_1}) \geq \frac{1}{\frac{1}{\beta}} \mathcal{F}_{n_1}(\beta_{n_1}) + \frac{1}{n_1(1-\rho)} \geq \frac{1}{\frac{1}{\beta}} \mathcal{F}_{n_1}(\beta_{n_1}) + \frac{1}{n_1(1-\rho)}
\]

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

The next lemma shows in particular that \( \hat{\beta}_{n}/\beta_{n} \) goes to 1 in probability.

**Lemma A.3.** Assume that \( K(z) = I(z \leq 1) \) or satisfies Assumption K, and let \( \{\beta_n = \beta_n(x, L, \sigma, \mathbb{P}, K) \mid n \geq 1\} \) be as in (A.1.6) and \( \{\hat{\beta}_n = \hat{\beta}_n(x, L, \sigma, K) \mid n \geq 1\} \) as in (A.1.3). Define \( n_1, n_2 \) as in Lemma A.2, with \( n_1 = \lfloor n/2 \rfloor \) and \( n_2 = \lfloor n/2 \rfloor \) for some \( \epsilon_0 > 1 \). Consider the event

\[
\mathcal{E}_{1n}(\epsilon) = \left\{ (1-\epsilon)\beta < \hat{\beta}_n \leq (1+\epsilon)\beta, \mathcal{F}(\beta_{n_2}) \leq \mathcal{F}_{n_1}(\beta_{n_1}) \leq \frac{\mathcal{F}(\beta_{n_2})}{(1-\epsilon)^2} \right\}.
\]

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

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

\[
\mathcal{E}_{1n}(\delta) = \left\{ 1 - \delta \leq \frac{L^2 \beta^2 \nu L}{L^2 \beta^2 \mathcal{F}_L} \leq 1 + \delta, \ell = n_1, \ldots, n_2 \right\}, \delta > 0.
\]

Since \( \beta_{n_1} \) decreases and \( L^2 \beta^2 \mathcal{F}_L(\beta_{n_1}) = \sigma^2 \) by Lemma A.2-(i), we have

\[
\frac{L^2 \beta^2 \nu L}{L^2 \beta^2 \mathcal{F}_L} \geq \frac{L^2 \beta^2 \nu L}{L^2 \beta^2 \mathcal{F}_L} \geq \frac{L^2 \beta^2 \nu L}{L^2 \beta^2 \mathcal{F}_L} = \frac{L^2 \beta^2 \nu L}{L^2 \beta^2 \mathcal{F}_L} = \frac{\nu_{n_2}(\beta_{n_2})}{n_2 \mathcal{F}_{n_2}(\beta_{n_2})},
\]

Let \( \mathcal{E}_{1n}^c(\delta) \) be the event “\( \mathcal{E}_{1n}(\delta) \) 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

\[
\mathbb{P}(\mathcal{E}_{1n}^c(\delta)) \leq \mathbb{P} \left( \frac{\nu_{n_2}(\beta_{n_2})}{n_1 \mathcal{F}_{n_1}(\beta_{n_1})} - 1 \geq \delta \right) + \mathbb{P} \left( \frac{\nu_{n_2}(\beta_{n_2})}{n_2 \mathcal{F}_{n_2}(\beta_{n_2})} - 1 \leq -\delta \right)
\]

with

\[
\mathbb{P}(\mathcal{E}_{1n}^c(\delta)) \leq \frac{\text{Var} \left( \frac{\nu_{n_2}(\beta_{n_2})}{n_1 \mathcal{F}_{n_1}(\beta_{n_1})} \right)}{(\mathcal{F}_{n_2}(\beta_{n_2}))^2} \leq \frac{\text{Var} \left( \frac{\nu_{n_2}(\beta_{n_2})}{n_2 \mathcal{F}_{n_2}(\beta_{n_2})} \right)}{(\mathcal{F}_{n_2}(\beta_{n_2}))^2}
\]

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

\[
\mathcal{E}_{2n}(\epsilon) = \left\{ (1-\epsilon)\beta < \hat{\beta}_n \leq (1+\epsilon)\beta, \ell = n_1, \ldots, n_2 \right\}, 0 < \epsilon < 1.
\]
Lemma A.4.

Assume that $K(z) = \mathbb{E}(z \leq 1)$ or satisfies Assumption K, and consider $\{\beta_n = \beta_n(x, L, \sigma, P, K), n \geq 1\}$, $\{\tilde{\beta}_n = \tilde{\beta}_n(x, L, \sigma, K), n \geq 1\}$ as in (A.1.6), (A.1.3). Let $n_1, n_2$ be as in Lemma A.2, with $n_1 = [n/20]$ and $n_2 = \lfloor n/20 \rfloor$ for some $z_0 > 1$, and consider a random measurable sequence $\{\tilde{\gamma}_n, n \geq 1\}$ such that there is a $\kappa_5 > 1$, with for all $x$ in $D$,

\begin{equation}
\label{eq:A.2.6}
\inf_{n \to +\infty} \mathbb{E}_{P \in \mathcal{P}(L, \sigma)} \mathbb{P}(n_1 \leq \tilde{\gamma}_n \leq \kappa_5 \tilde{\beta}_n) \text{ for all } \ell = n_1, \ldots, n_2 = 1.
\end{equation}

Then there exist some deterministic sequences $\{\beta_{n_1, n \geq 1}\}$, $\{\beta_{n_2, n \geq 1}\}$, which depend upon $z_0, \mathbb{P}, x$ in $D, L$ and $\sigma^2$, such that, for

\begin{equation}
\label{eq:A.2.7}
nr_n = n_2 \mathcal{F}_{n_1}(\beta_{n_2}) = \sigma^2/(L^2\beta_{n_2}) \geq \sigma^2/(L^2\tilde{\beta}_n^2),
\end{equation}

we have

\begin{equation}
\label{eq:A.2.8}
\lim_{n \to +\infty} \inf_{n \to +\infty} \mathbb{E}_{P \in \mathcal{P}(L, \sigma)} \mathbb{P}(\beta_n \leq \tilde{\gamma}_n \leq \beta_{n_2} \text{ for all } \ell = n_1, \ldots, n_2 = 1,
\end{equation}

\begin{equation}
\label{eq:A.2.9}
\frac{1}{C_4} \leq \frac{\mathcal{F}_\ell(\tilde{\beta}_n)}{nr_n} \leq C_4 \text{ for all } \beta \in [\beta_{n_1}, \beta_{n_2}] \text{ and } \ell \in [n_1, n_2],
\end{equation}

for $n_1 \geq 1$, and where $\mathcal{F}_n$ 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 $\inf_{n \in \mathcal{P}(L, \sigma)} nr_n$ diverges when $n$ grows. Let $C_2 = C_2(z)$ 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 $n_2 \mathcal{F}_{n_1}(\cdot) \leq n_2 \mathcal{F}_{n_2}(\cdot)$ give that there are $\beta_{n_1} \leq \beta_{n_2}$ with

\begin{equation}
\label{eq:A.2.10}
n_2 \mathcal{F}_{n_2}(\beta_{n_1}) = \frac{C_2}{\kappa_5 z_0^2} nr_n \quad \text{and} \quad n_1 \mathcal{F}_{n_1}(\beta_{n_2}) = 3\kappa_5 nr_n.
\end{equation}
It follows that (A.2.9) holds by definition of \( r_n \) and Lemma A.1-(iii) which gives for some \( C_4 \)
\[
\mathbb{E} \ell_F(\beta) \geq n_1 \mathbb{P}(\beta_{n1}) + n_1 \mathbb{P}(\beta_{n2}) - \frac{n_1}{1 - \rho} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)
= \frac{C_2(1 + o(1)) n r_n}{3 - \frac{n_1}{1 - \rho} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \geq n r_n, 
\]
\[
\mathbb{E} \ell_F(\beta) \leq n_2 \mathbb{P}(\beta_{n2}) \leq n_2 \mathbb{P}(\beta_{n1}) + \frac{n_2}{1 - \rho} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)
= 3 \frac{\kappa_5}{\kappa_5} (1 + o(1)) n r_n + \frac{n_2}{1 - \rho} \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \leq C_4 n r_n. 
\]

(A.2.8) follows from (A.2.6), definition of \( r_n \), Lemmas A.3 and A.2-(iii). (A.2.7) holds if
\[
\lim_{n \to \infty} \inf_{p \in \mathcal{P}_D(L, \sigma)} \mathbb{P}(\nu(\beta_{n1}) + 1 < \nu(\gamma_{\ell}) < \nu(\beta_{n2}) - 1 \text{ for all } \ell \in [n_1, n_2]) = 1.
\]

But (A.2.6), Lemmas A.3 and A.2-(iii,ii) show that
\[
\lim_{n \to \infty} \inf_{p \in \mathcal{P}_D(L, \sigma)} \mathbb{P} \left( \frac{C_2 n r_n}{(2 \kappa_5^2) - 1} < \nu(\gamma_{\ell}) \leq 2 \kappa_5 n r_n \text{ for all } \ell \in [n_1, n_2] \right) = 1.
\]
Therefore, since \( \nu(\beta_{n1}) \leq \nu(\beta_{n2}) \) and \( \nu(\beta_{n2}) \geq \nu(\beta_{n1}) \), it is sufficient to show that
\[
\lim_{n \to \infty} \inf_{p \in \mathcal{P}_D(L, \sigma)} \mathbb{P} \left( \nu(\beta_{n2}) + 1 < \frac{C_2 n r_n}{(2 \kappa_5^2) - 1}, 2 \kappa_5 n r_n < \nu(\beta_{n1}) - 1 \text{ for } \ell \in [n_1, n_2] \right) = 1.
\]

But the definition of \( \beta_{n1} \) and \( \beta_{n2} \) and Lemma A.1-(A.2.1) yield
\[
\mathbb{E} \nu(\beta_{n1}) = \frac{C_2 n r_n}{\kappa_5^2 - 3} \quad \text{with } \text{Var} \left( \nu(\beta_{n1}) \right) \leq \left( 1 + \frac{4 \rho}{1 - \rho} \right) \frac{C_2 n r_n}{\kappa_5^2 - 3},
\]
\[
\mathbb{E} \nu(\beta_{n2}) = \frac{3 \kappa_5 n r_n}{2} \quad \text{with } \text{Var} \left( \nu(\beta_{n2}) \right) \leq \frac{3}{1 + \frac{4 \rho}{1 - \rho}} n r_n.
\]

with \( n r_n \geq \sigma^2 / (L^2 \kappa_5^2) \) for all \( p \in \mathcal{P}_D(L, \sigma) \), 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 \( K(z) = I(z \leq 1) \) 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 \( n_2 = \lfloor n z_0 \rfloor \) for some \( z_0 > 1 \). Consider a random measurable sequence \( \{ \gamma_n, n \geq 1 \} \) satisfying the condition (A.2.6) of Lemma A.4. Then, for any \( x \in D \),
\[
\lim_{n \to \infty} \sup_{p \in \mathcal{P}_D(L, \sigma)} \mathbb{P} \left( \max_{n_1 \leq n \leq n_2} \frac{\Sigma_{\ell}(\gamma_{\ell})}{\sqrt{r_n}} \geq z \right) = 0,
\]

**Proof of Lemma A.5:** Assume \( x = 0 \in D \) for brevity. By Lemma A.4-(A.2.8), it is sufficient to prove
\[
\lim_{n \to \infty} \sup_{p \in \mathcal{P}_D(L, \sigma)} \mathbb{P} \left( \max_{n_1 \leq n \leq n_2} \left\| \frac{\Sigma_{\ell}(\gamma_{\ell})}{\sqrt{r_n}} \right\| \geq z \right) = 0.
\]
Let $\beta_{1n}$ and $\beta_{2n}$ be as in Lemma A.4. Define, for $\lambda = (\lambda_1, \lambda_2)$ in $[0, 1]^2$,

$\beta_n(\lambda_1) = \beta_{1n} + \lambda_1(\beta_{2n} - \beta_{1n})$, $n(\lambda_2) = n_1 + (n_2 - n_1)\lambda_2$, $S_n(\lambda) = \frac{1}{\sqrt{n'}} \sum_{1 \leq n(\lambda_2)} K_{\beta_n(\lambda_1)}(x_{i-1}) \eta_i$.

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

$\sup_{p \in P_2(L_\lambda)} \mathbb{E}^{1/6} \max_{\lambda \in [0,1]^2} \|S_n(\lambda)\|^6 \leq C$.

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 $\lambda, \lambda'$ in $[0,1]^2$,

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

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

$\mathbb{E}_{\phi}^{1/6} \|S_n(\lambda_1, \lambda_2) - S_n(\lambda'_1, \lambda'_2)\|^6 = \frac{1}{\sqrt{n'}} \mathbb{E}_{\phi}^{1/6} \sum_{1 \leq n(\lambda_2)} K_{\beta_n(\lambda_1), \beta_n(\lambda'_1)}(x_{i-1}) \eta_i$,

(A.2.11)

$\leq C \left[ \left( 1 + n_2 (F_{n_2}(\beta_n(\lambda_1)) - F_{n_2}(\beta_n(\lambda'_1))) \right)^3 - 1 \right]^{1/6} = C d_{2n}(\lambda_1, \lambda'_1)$,

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

$\mathbb{E}_{\phi}^{1/6} \|S_n(\lambda'_1, \lambda'_2) - S_n(\lambda'_1, \lambda'_2')\|^6 = \frac{1}{\sqrt{n'}} \mathbb{E}_{\phi}^{1/6} \sum_{n(\lambda_2) < i \leq n(\lambda'_2)} K_{\beta_n(\lambda'_1)}(x_{i-1}) \eta_i$,

$\leq C' \left[ \left( 1 + n_2 (F_{n_2}(\beta_n(\lambda'_1))) \right)^3 - 1 \right]^{1/6} \leq \frac{C'}{\sqrt{n'}} \left[ \left( 1 + n_2 n(\lambda_2) \right)^3 - 1 \right]^{1/6} \leq C n r_2(\lambda_2 - \lambda'_2)$,

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

$\sum_{n(\lambda_2) < i \leq n(\lambda'_2)} F_{n_2}(\beta_n(\lambda'_1)) \leq C n r_2(\lambda_2 - \lambda'_2)$,

and then,

(A.2.12) $\mathbb{E}_{\phi} \|S_n(\lambda_1, \lambda_2) - S_n(\lambda'_1, \lambda'_2')\|^6 \leq C d_{2n}(\lambda_2, \lambda'_2) = C \left[ \left( 1 + n r_2(\lambda_2 - \lambda'_2) \right)^3 - 1 \right]^{1/6}$.

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

$\mathbb{E}_{\phi}^{1/6} \|S_n(\lambda) - S_n(\lambda')\|^6 \leq C d_n(\lambda, \lambda')$ 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),

$\frac{n_2}{n r_2} \left| F_{n_2}(\beta_n(\lambda_1)) - F_{n_2}(\beta_n(\lambda'_1)) \right| \leq \frac{(n r_2)^{3} \epsilon + 1}{n r_2} - 1 \leq \epsilon^2$,

$|\lambda_2 - \lambda'_2| \leq \frac{(n r_2)^{3} \epsilon + 1}{n r_2} - 1 \leq \epsilon^2$. 

\( F_{u_2}(\cdot) \) is a cumulative distribution function and we used its quantiles to cover \([0, 1]\). It follows from (A.2.9) that the number of \( d_\epsilon \)-balls with radius \( \epsilon \) necessary to cover \([0, 1]^2\) is bounded by \( C/\epsilon^4 \).

Since (A.2.9), (A.2.11) and (A.2.12) yield that the \( d_\epsilon \)-diameter of \([0, 1]^2\) can be bounded by a \( \delta > 0 \) independently of \( n \) and \( P \). Corollary 2.2.5 of van der Vaart and Wellner (1996) shows that

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

This implies (A.2.10) since \( \sup_{P \in \mathcal{P}(L,\sigma)} \mathbb{E}_P^{1/6} \|S_n(0,1)\|^6 \) 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 \( X = \mathcal{X} \times [0,1] \). With a little abuse of language, we use \( P \) to denote the distribution of \( \{X^*_t, t \geq 0\} \).

Let \( \tau^*_1 \) be the first epoch at which \( \{X^*_{t-1}, t \geq 1\} \) visits \( a^* = \mathcal{D} \times \{1\} \). 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 \( P(Z_{\tau^*_1-1} = 0) = 0, P(X_{\tau^*_1-1} \in \mathcal{A}, Z_{\tau^*_1-1} = 1) = P(A) \).

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 \( N^*_T(a^*) = \sum_{t=1}^{T} I(Z_{t-1} = 1) = \sum_{t=\tau^*_1}^{T} I(Z_{t-1} = 1) \) to the atom \( a^* \). The last step concludes.

**Step 1. The visit time \( \tau^*_1 \).** (A.3.1) gives

\[
\mathbb{P}(\tau^*_1 > T) = \mathbb{P}(Z_{T-1} = 0, \ldots, Z_0 = 0) = \prod_{t=1}^{T-1} \mathbb{P}(Z_t = 0 | Z_{t-1} = 0, t \leq t - 1) \times \mathbb{P}(Z_0 = 0)
\]

\[
= \prod_{t=0}^{T-1} (1 - (1 - \rho))^{P(X_t \in \mathcal{D})} = \prod_{t=2}^{T} (1 - (1 - \rho))^{P(X_{t-1} \in \mathcal{D})} \text{ and then,}
\]

\[
\mathbb{P}(\tau^*_1 > T) \leq \exp \left(- (1 - \rho) \sum_{t=1}^{T} I(X_{t-1} \in \mathcal{D}) \right) = \exp \left(- (1 - \rho) \mathbb{E}_P N_T(D) \right),
\]
so that \( \sup_{\tau \in \mathcal{P}_\tau(L,\sigma)} \mathbb{P}(\tau^* > T) \leq \exp(- (1 - \rho) \mu_T) \rightarrow 0 \) when \( T \rightarrow +\infty \) by (3.3).

**Step 2. The number of visits \( N^*_T(a^*) \).** Set \( n_{1T} = n_{1T}(\mathbb{P}) = \mathbb{E}_c[N^*_T(a^*)\mid \tau^*_1 \leq T] \), observing that \( N^*_T(a^*) = 0 \) when \( \tau^*_1 > T \). (A.3.1) gives \((1 - \rho)\mathbb{E}_c N^*_T(D) = \mathbb{E}_c N^*_T(a^*) = n_{1T} (1 - \mathbb{P} (\tau^*_1 > T)) \). Therefore, \( \sup_{\tau \in \mathcal{P}_\tau(L,\sigma)} \mathbb{P}(\tau^*_1 > T) \leq \exp(- (1 - \rho) \mu_T) \) proven in Step 1 and (3.3) yield

\[
(A.3.2) \quad \sup_{\tau \in \mathcal{P}_\tau(L,\sigma)} \left| \mathbb{E}_c N^*_T(D) - \frac{1}{n_{1T}(\mathbb{P})} \right| = \sup_{\tau \in \mathcal{P}_\tau(L,\sigma)} \left| \mathbb{E}_c N^*_T(a^*) - \frac{1}{n_{1T}(\mathbb{P})} \right| \leq \exp(- (1 - \rho) \mu_T)
\]

so that \( \inf_{\tau \in \mathcal{P}_\tau(L,\sigma)} n_{1T}(\mathbb{P}) \geq \mathbb{C}_{\mathbb{H}_T} \).

The bound (2.4) of Chen (1999a) for \( \tilde{N}^*_T(a) = N^*_T(a^*) - 1 = \sum_{i=1}^{\tau^*_1} \mathbb{I}(Z_{t-1} = 1) \) yields, for any \( \delta > 0 \),

\[
\mathbb{E}_c \left[ \tilde{N}^*_T(a^*) \mathbb{I} \left( \tilde{N}^*_T(a^*) \geq \delta \right) \mid \tau^*_1 \leq T \right] \leq \mathbb{P} \left[ \tilde{N}^*_T(a^*) \geq \delta \mid \tau^*_1 \leq T \right] \left( \delta + 1 + \mathbb{E}_c \left[ \tilde{N}^*_T(a^*) \mid \tau^*_1 \leq T \right] \right)
\]

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

\[
(A.3.3) \quad \mathbb{P} \left( \tilde{N}^*_T(a^*) \geq 1 + \delta \right) \geq \frac{n_{1T} - 1 - \delta}{n_{1T} + \delta} \mathbb{P} (\tau^*_1 \leq T).
\]

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

\[
\text{lim}_{T,z \to +\infty} \inf_{\tau \in \mathcal{P}_\tau(L,\sigma)} \mathbb{P} \left( \frac{n_{1T}}{z} \leq \tilde{N}^*_T(a^*) \leq n_{1T} z \right) = 1,
\]

since the Markov inequality gives \( \sup_{T,z \to +\infty} \inf_{\tau \in \mathcal{P}_\tau(L,\sigma)} \mathbb{P} (n_{1T} z < \tilde{N}^*_T(a^*)) \leq 1/z \).

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

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

\[
(A.3.4) \quad R_T \left\| m_T(0, \mathbb{b}_T) - m(0) \right\| \leq \frac{1}{L_T} \left\| \sum_{i=1}^{T} (m(X_{t-1}) - m(0)) \mathbb{I} \left( \|X_{t-1}\| \leq \mathbb{b}_T \right) \right\| \frac{N_T(\mathbb{b}_T)}{N_T^1} + \frac{\sum_{i=1}^{T} \mathbb{I} \left( \|X_{t-1}\| \leq \mathbb{b}_T \right) \epsilon_i}{\sigma N_T^{3/4}(\mathbb{b}_T)}.
\]

Consider \( \epsilon > 0 \). By Theorem 1, there is a \( z_0 > 1 \) such that for \( n_{1T} = [n_T / z_0], n_T = [n_T z_0], \) we have

\[
(A.3.6) \quad \lim_{T \to \infty} \inf_{\tau \in \mathcal{P}_\tau(L,\sigma)} \mathbb{P} (E_{3T}) \geq 1 - \epsilon/2, \quad \text{where} \quad E_{3T} = \{ n_{1T} \leq N_T(D) \leq n_T \}.
\]

Observe that (3.3) and (A.3.2) yield that \( n_{1T} \geq \mathbb{N}_T \geq \mathbb{N}_T / z_0 \to +\infty \) for all \( \mathbb{P} \in \mathcal{P}_D(L,\sigma) \). Since the estimation point 0 is in the open \( \mathcal{D} \), there is a \( \mathbb{B} > 0 \) such that \( \{ x \in \mathcal{X} : \|x\| \leq \mathbb{B} \} \subset \mathcal{D} \). Let \( \mathbb{N}_u \) be as in Lemma A.2-(ii), so that \( \mathbb{N}_u \leq \mathbb{B}/2 \) for \( T \) large enough. Therefore, definitions (2.3) and (A.1.3) of \( \mathbb{b}_T \) and \( \mathbb{\beta}_u \), (A.3.6), Lemmas A.3 and A.2-(ii) yield that

\[
(A.3.7) \quad \lim_{T \to \infty} \inf_{\tau \in \mathcal{P}_\tau(L,\sigma)} \mathbb{P} (E_{4T}) \geq 1 - \epsilon \quad \text{where} \quad E_{4T} = \{ \mathbb{b}_T = \mathbb{\beta}_T(D) \leq \mathbb{B} \} \cap E_{3T}.
\]
On \( E_{xT} \), \( m_T(0; \bar{b}_T) \) averages over \( D \). In this case, (A.3.4)-(A.3.5) and Definition 1-(ii) yield

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

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

\[
\limsup_{T,x \to +\infty} \sup_{P \in P_T(L,\sigma)} P \left( R_T \left\| m_T(0; \bar{b}_T) - m(0) \right\| \geq z \right) = 0,
\]

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

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

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

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

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

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

\( N_T(\cdot; K) \leq N_T(\cdot) \) yields \( R_{OT} \leq R_T \) and \( N_T(\cdot; K) \geq \kappa_T^2 N_T(\cdot/2) \) implies

\[
\frac{1}{R_{OT}^2} \leq \min_{h \geq 0} \frac{L^2 h^4, \sigma^2}{N_T(h/2)} \leq 4 \min_{h \geq 0} \frac{L^2 h^2, \sigma^2}{N_T(h/2)} = \frac{4 \kappa_T^2 R_T^2}{\kappa_T^4},
\]

so that \( \kappa_T R_T/2 \leq R_{OT} \leq R_T \). We now bound \( R_T(K) \) with \( R_{OT} \). Because \( LB_T(h) \leq Lh, \sigma^2/V_T(h) \leq \sigma^2/N_T(h; K) \) and \( \hat{h}_{OT} \) is in \( H_T \), we have

\[
\frac{1}{R_T^2(K)} = \min_{h \in H_T} \left( \frac{L^2 B^2_T(h) + \sigma^2}{V_T(h)} \right) \leq 2 \max \left( \frac{L^2 \hat{h}_{OT}^2, \sigma^2}{N_T(\hat{h}_{OT}; K)} \right) = \frac{2}{R_{OT}^2},
\]

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

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

\[
R_T(K) \leq \frac{\sum_{t=1}^T K \left( \frac{\| X_{t-1} \|}{\hat{h}_{OT}} \right)^2}{\sigma^2 N_T^{1/2}(\hat{h}_{OT}; K)} \cdot \frac{\sum_{t=1}^T (m(X_{t-1}) - m(0)) K \left( \frac{\| X_{t-1} \|}{\hat{h}_{OT}} \right)}{\sum_{t=1}^T K \left( \frac{\| X_{t-1} \|}{\hat{h}_{OT}} \right)} + \frac{\sum_{t=1}^T K \left( \frac{\| X_{t-1} \|}{\hat{h}_{OT}} \right) \epsilon_t}{\sigma N_T^{1/2}(\hat{h}_{OT}; K)}.
\]
Now, (A.3.7) yields that \( \hat{h}_{\omega T} = \hat{\beta}_{N_T(D)} \) with a large probability, and \( \hat{h}_T \leq \kappa_0 \hat{h}_{\omega T} \) by definition of \( \mathcal{H}_T \). Hence \( \mu_T(0; \hat{h}_T) \) averages on \( D \) with large probability. Setting \( \hat{\gamma}_{N_T(D)} = \hat{h}_T \) gives in this case
\[
R_T(K) \left\| \mu_T(0; \hat{h}_T) - m(0) \right\| \leq 1 + \frac{\| \Sigma_{N_T(D)} \left( \hat{\gamma}_{N_T(D)} \right) \|}{\sigma_{N_T(D)}^2}.
\]
Since \( N_T(\cdot; K) \leq V_T(\cdot; K) \), the bounds in \( \mathcal{H}_T \) implies
\[
\frac{1}{\kappa_0} \nu_{N_T(D)} \left( \hat{\beta}_{N_T(D)} \right) \leq \nu_{N_T(D)} \left( \hat{\gamma}_{N_T(D)} \right) \leq \kappa_0 \nu_{N_T(D)} \left( \hat{\beta}_{N_T(D)} \right),
\]
that is Condition (A.2.6) in Lemmas A.4 and A.5, which gives (3.4).

A.3.4. Proof of Theorem 4 and Corollary 1. Take \( D = (-1, 1) \) and \( x = 0 \) without loss of generality. Recall that the \( \{e_i, t \geq 0\} \) 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 \( \mathbb{P}_T \). Consider a 1/2-Lipschitz function \( \mu(\cdot) : \mathbb{R} \to [0, 1/2] \) with support \([0, 1]\) and \( \mu(0) = 1/2 \), as \( \mu(x) = (1 - |x|)[[-1,1](x)]/2 \). Define
\[
(A.3.8) \quad m^b(x) = m_0(x) + L \mu \left( \frac{x}{b} \right), \quad b \in [0, \min(1,1/L)],
\]
with \( m^b(\cdot) = \lim_{b \to 0} m^b(\cdot) = m_0(\cdot) \). Note that \( m^b(x) \) is continuous over \( \mathbb{R} \), \( L \)-Lipschitz over \( D \), that \( \sup_{x \in D} |m^b(x)| \leq \kappa_3 + 1 \) and \( m^b(x) = m_0(x) \) for all \( x \in \mathbb{R} \setminus D \), for all \( b \in [0, \min(1,1/L)] \). Let \( \mathbb{P}^b \) be the distribution of \( \{X^b_t, t \geq 0\} \) where \( X^b_0 = e_0 \) and \( X^b_t = m(X^b_{t-1}) + e_t \), so that \( \mathbb{P}^b \) satisfies Definition 1-(i,iii). Since \( m^b(\cdot) \) and \( m_0(\cdot) \) coincide over \( \mathbb{R} \setminus D \), (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 \( \lim_{T \to \infty} \mathbb{E}_T N_T(D) = \infty \) for all \( b \in [0, \min(1,1/L)] \).

Set \( n_T = \inf_{b \in [0, \min(1,1/L)]} \mathbb{E}_T n_T(D) \) so that \( \mathbb{E}_{\mathbb{P}^b} N_T(D) \geq n_T \) for all \( b \in [0, \min(1,1/L)] \). Recall that \( n_T \) must diverge to define a class \( \mathcal{P}_T(L, \sigma) \). Observe that \( b \mapsto X^b_t \) is continuous for all \( t \geq 0 \) so that the Lebesgue Dominated Convergence Theorem yields that
\[
n_T(b) = \mathbb{E}_{\mathbb{P}^b} N_T(D) = \mathbb{E}_T \left[ \sum_{t=0}^{T-1} I(X^b_t \in D) \right] \geq P_\sigma(e_0 \in D) > 0
\]
is continuous for all \( T \geq 0 \). Since \( 1/n_T(\cdot) \) decreases with \( T \) and \( \lim_{T \to \infty} 1/n_T(b) = 0 \) for all \( b \) in \( [0, \min(1,1/L)] \), the Dini Theorem yields that \( 1/n_T = \sup_{b \in [0, \min(1,1/L)]} 1/n_T(b) \to 0 \). Hence \( \lim_{T \to \infty} n_T = \infty \) and, for this choice of \( n_T, \mathbb{P}^b \) is in \( \mathcal{P}_T(L, \sigma) \) for all \( b \in [0, \min(1,1/L)] \).

Step 2. A contiguous alternative \( \mathbb{P}_1 = \mathbb{P}_1^T \). 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 \( b_T = b_T(z) \to 0 \) such that
\[
(A.3.9) \quad b_T = \min \left\{ b \geq 0; L^2 b^2 \mathbb{E}_0 N_T(b) \geq \frac{\sigma^2}{2T} \right\}, \quad \text{or equivalently} \quad L^2 b^2 \sum_{t=1}^{T} \mathbb{P}(|X^0_{t-1}| \leq b_T) = \frac{\sigma^2}{z^2}.
\]
Abbreviate \( \mathbb{P}_T^T, m^{b_T}(\cdot), \{X^0_t, t \geq 0\} \in \mathbb{P}_1, m_1(\cdot), \{X^1_t, t \geq 0\} \) respectively. (A.3.8) yields that \( |m_1(0) - m_0(0)| \leq L b_T/2 \) 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 \( \mathcal{L}_{T} = \log(dP_{1}/dP_{0})(X_{0}, \ldots, X_{T}) \) writes

\[
\mathcal{L}_{T} = \frac{L_{bT}}{\sigma^{2}} \sum_{t=1}^{T} \mu \left( \frac{X_{t-1}}{b_{T}} \right) \epsilon_{t} - \frac{L^{2}_{bT}2}{2\sigma^{2}} \sum_{t=1}^{T} \mu^{2} \left( \frac{X_{t-1}}{b_{T}} \right)
\]

under \( P_{0} \).

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

\[
\mathbb{E}_{P_{0}} \left( \frac{L^{2}_{bT}2}{2\sigma^{2}} \sum_{t=1}^{T} \mu^{2} \left( \frac{X_{t-1}}{b_{T}} \right) \epsilon_{t} \right) \leq \frac{L^{2}_{bT}2}{2\sigma^{2}} \sum_{t=1}^{T} \mathbb{P}_{0} (|X_{t-1}| \leq b_{T}) \leq \frac{L^{2}_{bT}2\mathbb{E}_{P_{0}}N_{T}(b_{T})}{8\sigma^{2}} = \frac{1}{8z^{2}},
\]

and then, for any \( z \geq 1 \),

\[
\mathbb{E}_{P_{0}} \left( |\mathcal{L}_{T}| \right) \leq \mathbb{E}_{P_{0}}^{1/2} \left( \frac{L_{bT}}{\sigma^{2}} \sum_{t=1}^{T} \mu \left( \frac{X_{t-1}}{b_{T}} \right) \epsilon_{t} \right)^{2} + \mathbb{E}_{P_{0}} \left( \frac{L^{2}_{bT}2}{2\sigma^{2}} \sum_{t=1}^{T} \mu^{2} \left( \frac{X_{t-1}}{b_{T}} \right) \right) \leq \frac{1}{2z^{2}}.
\]

Therefore the Markov inequality gives, for any \( z' > 0 \),

\[
\sup_{T \geq 1} \mathbb{P}_{0} (|\mathcal{L}_{T}| \geq z') \leq \frac{z'}{2z^{2}}.
\]

**Step 3. The random rate \( R_{T} \).** Under \( P_{0} \) and \( P_{1} \), Recall that \( R_{T} = 1/(L_{bT}) \) with \( b_{T} \) as in (2.3).

Therefore, the definition (A.3.9) of \( b_{T} \) gives for \( z \geq 1 \) and under \( P_{0} \),

\[
\mathbb{P}_{0} \left( R_{T} \geq \frac{1}{2z^{2}L_{bT}} \right) = \mathbb{P}_{0} \left( \bar{b}_{T} \leq 2z^{2}b_{T} \right) = \mathbb{P}_{0} \left( \sigma^{2} \leq 4z^{2}L^{2}_{bT}N_{T}(2z^{2}b_{T}) \right)
\]

\[
\geq \mathbb{P}_{0} \left( \sigma^{2} \leq 4z^{2}L^{2}_{bT}b_{T}N_{T}(b_{T}) \right) = \mathbb{P}_{0} \left( \frac{\mathbb{E}_{P_{0}}N_{T}(b_{T})}{4} \leq N_{T}(b_{T}) \right).
\]

To find a lower bound for the probability above will be done using the split chain \( \{X_{t}, Z_{t}, t \geq 0\} = \{(X_{t}, Z_{t}), t \geq 0\} \) defined in the Proof of Theorem 1, and \( N_{T}^{*}(a^{*}, b_{T}) = \sum_{t=1}^{T} I(Z_{t-1} = 1, |X_{t}| \leq b_{T}) \leq N_{T}(b_{T}) \). 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 \( N_{T}^{*}(a^{*}), N_{T}(D) \), gives

\[
\mathbb{P}_{0} \left( \frac{\mathbb{E}_{P_{0}}N_{T}(b_{T})}{4} \leq N_{T}(b_{T}) \right) \geq \frac{(1 - \frac{1}{3})}{(1 + \frac{1}{3})} \mathbb{E}_{P_{0}}N_{T}(b_{T})(1 + o(1)) = \frac{3}{5}(1 + o(1)).
\]

Under \( P_{1} \), (A.3.10) gives

\[
\mathbb{P}_{1} \left( R_{T} \geq \frac{1}{2z^{2}L_{bT}} \right) = \mathbb{E}_{P_{1}} \left[ \mathbb{I} \left( R_{T} \geq \frac{1}{2z^{2}L_{bT}} \right) \exp(\mathcal{L}_{T}) \right] \geq \frac{1}{3} \mathbb{P}_{0} \left( R_{T} \geq \frac{1}{2z^{2}L_{bT}} \right), \mathcal{L}_{T} \geq - \ln 3
\]

\[
\geq \frac{1 + o(1)}{5} - \mathbb{P}_{0} (|\mathcal{L}_{T}| \geq \ln 3) \geq \frac{1 + o(1)}{5} - \frac{\ln 3}{z^{2}}.
\]

Hence

\[
(A.3.11) \liminf_{T \to +\infty} \frac{1}{\mathbb{P}_{0} \left( |\mathcal{L}_{T}(z)| \geq \frac{1}{6} \right) \mathbb{P}(E_{T}(z))} \geq \frac{1}{6} \text{ with } E_{T}(z) = \left\{ R_{T} \geq \frac{1}{2z^{2}L_{bT}} \right\}, \text{ for } z \text{ large enough.}
\]
Step 4. Proof of the lower bound. Because $\mathbb{P}_1$ is in $\mathcal{V}_0(\mathbb{P}_0; \varepsilon)$ for $T$ large enough, (A.3.11) gives, for any $z' > 0$,

$$
\lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \sup_{\tilde{R}_T} \mathbb{P} \left( |\tilde{m}_T(0) - m(0)| \geq z' \right)
\geq \lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \max_{\tilde{R}_T} \mathbb{P} \left( |\tilde{m}_T(0) - m(0)| \geq z', E_T(z) \right)
\geq \lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \max_{\tilde{R}_T} \mathbb{P} \left( |\tilde{m}_T(0) - m(0)| \geq 2Lz'^2b_T, E_T(z) \right).
$$

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

$$
\mathbb{P}_1 \left( |\tilde{m}_T(0) - m_i(0)| \geq 2Lz'^2b_T, E_T(z) \right) \geq \mathbb{P}_1 \left( Lb_T/2 - |\tilde{m}_T(0) - m_0(0)| \geq 2Lz'^2b_T, E_T(z) \right).
$$

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

$$
\lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \sup_{\tilde{R}_T} \mathbb{P} \left( |\tilde{m}_T(0) - m(0)| \geq 1/8z^2 \right)
\geq \lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \max_{\tilde{R}_T} \mathbb{P}_0 \left( |\tilde{m}_T(0) - m_0(0)| \geq Lb_T/4, E_T(z) \right), \mathbb{P}_1 \left( |\tilde{m}_T(0) - m_0(0)| < Lb_T/4, E_T(z) \right).
$$

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

$$
\lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \sup_{\tilde{R}_T} \mathbb{P} \left( |\tilde{m}_T(0) - m(0)| \geq 1/8z^2 \right)
\geq \lim \inf_{T \to \infty} \inf_{\chi_T} \max_{\tilde{R}_T} \left\{ \mathbb{P}_0(\chi_T = 1, E_T(z)), \mathbb{P}_1(\chi_T = 0, E_T(z)) \right\}.
$$

But, for any test $\chi_T$,

$$
\mathbb{P}_1(\chi_T = 0, E_T(z)) = \int I(\chi_T = 1, E_T(z)) e^{\mathcal{L}_T} d\mathbb{P}_0 \geq \exp(-z) \int I(\chi_T = 1, \mathcal{L}_T \geq -z, E_T(z)) e^{\mathcal{L}_T} d\mathbb{P}_0 \geq \exp(-z) (\mathbb{P}_0(\mathcal{L}_T \geq -z, E_T(z)) - \mathbb{P}_0(\mathcal{L}_T = 1, E_T(z))), \text{ and then}
$$

$$
\inf_{\chi_T} \max_{\mathbb{P}_0}(\mathbb{P}_0(\chi_T = 1, E_T(z)), \mathbb{P}_1(\chi_T = 0, E_T(z))) \geq \min_{\mathbb{P}_0} \max_{p \in \{0, 1\}} [p, \exp(-z)(\mathbb{P}_0(\mathcal{L}_T \geq -z, E_T(z)) - p)] = \frac{\exp(-2z)}{1 + \exp(-z)} \mathbb{P}_0(\mathcal{L}_T \geq -z, E_T(z)) \geq \frac{\exp(-2z)}{1 + \exp(-z)} \mathbb{P}_0(\mathcal{L}_T \geq \mathcal{L}_T(z)) - \mathbb{P}_0(|\mathcal{L}_T| \geq z).
$$

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

$$
\lim \inf_{T \to \infty} \inf_{\tilde{m}_T \in \mathcal{V}_0(\mathbb{P}_0; \varepsilon)} \sup_{\tilde{R}_T} \mathbb{P} \left( |\tilde{m}_T(0) - m(0)| \geq 1/8z^2 \right) \geq \frac{\exp(-2z)}{1 + \exp(-z)} \left( \frac{1}{6} - \frac{1}{z} \right),
$$

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

Corollary 1 follows from minor modifications of the proof of Theorem 4 detailed now. Let $\mathbb{P}_0$ be the distribution of $\{\epsilon_t, t \geq 0\}$ 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 $\mathbb{E}_0[\epsilon^2_{t+1} | \epsilon_t] = \sigma^2$. Assumption $P$ ensures that $\mathbb{P}_0$ is in $\mathcal{P}_T(L, \sigma)$ provided that $TP_0(\epsilon_t \in \mathcal{D}) \geq \mu_T$. The perturbations $\mathbb{P}_0^h$ of $\mathbb{P}_0$ defined through (A.3.8) satisfy a Minorization Condition $\mathbb{P}_0^h(X_{i+1} \in A | X_i = x) \geq (1 - \rho') \mathcal{P}'(A)$ for all $x$ in $\mathbb{R}$ and $A \subset \mathbb{R}$, where $\mathcal{P}'(dy)$ is proportional.
to \(\inf_{m} \geq n+1 \exp (-y - m)^2 / 2\sigma^2)dy > 0\). This ensures that \(\mathbb{E}_n N_T(D) \geq (1 - \rho') \mathbb{P}^\prime(D) > 0\), so that these \(\mathbb{P}^k\) are in \(\mathcal{P}_{\mathcal{D}}(L, \sigma)\) provided \(\sup_{\mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}^\prime(D) \geq (1 - \rho') \mathbb{P}^\prime(D)\). Corollary 1 then follows since \(\sup_{\mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(R_T | \hat{\mu}_T(0) - m(0) | \geq z) \geq \sup_{\mathcal{P}_{\mathcal{D}}(L, \sigma)} \mathbb{P}(R_T | \hat{\mu}_T(0) - m(0) | \geq z)\).

\[\square\]

**Appendix B: Proof of Propositions 1 and 2**

**B.1. Proof of Proposition 1.** Set \(x_t = (y_t, \ldots, y_{t-d+1})\) 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 \(f(y) = \inf_{x \in \mathcal{D}} f(y - m(x)|x)\) 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 [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 \(\mathbb{P}([\|X_t - x\| = h] \mathcal{X}_{t-d} = 0, \|X_t - x\| = h) = 0\) 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 \(x_{t-d} \in D\),

\[
\mathbb{P}(X_t \in A | X_{t-d} = x_{t-d}) \geq \int [I(x_t \in A \cap \mathcal{D}) \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) I(x_{t-k-1} \in \mathcal{D}')] dy_{t-k} = (1 - \rho_D) \int [I(x_t \in A) \mathcal{J}_D(x_t)]dx_t,
\]

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

\[
1 - \rho_D = \int [I(x_t \in \mathcal{D}) \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k}) I(y_{t-k} \in [a, b])] dy_{t-k} = \int [I(x_t \in \mathcal{D}) \prod_{k=0}^{d-1} f_{\mathcal{D}'}(y_{t-k})] dy_{t-k}.
\]

If \(\mathcal{D}\) is a bounded open subset of \(\mathbb{R}^d\), \(0 \leq \rho_D < 1\) and \(\mathcal{J}_D(\cdot)\) is a positive continuous density over \(\mathcal{D}\). Hence \(\mathcal{J}_D(\cdot)\) defines a \(\mathcal{P}_{\mathcal{D}}(\cdot)\) satisfying (3.1) and \(\mathbb{P}(X_{t+d} \in A | X_t = x) \geq (1 - \rho_D) \mathcal{J}_D(A) \mathcal{I}(x_t \in \mathcal{C})\) 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. \(\mathbb{E}_n |N_T(D)| |X_0 = x| \to \infty\) for almost all \(x\). Hence \(\lim_{T \to \infty} \mathbb{E}_n N_T(D) = \infty\). \(\square\)
Consider a sequence $b_T > 0$ going to 0 with $\lim_{T \to \infty} \sqrt{T} b_T = +\infty$. 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

$$N_T(b_T) = \sum_{t=1}^T \mathbb{I}(|Y_{t-1}| \leq \left(\frac{b_T}{\theta}\right)^{1/\alpha}) = N_T^{\gamma}(\left(\frac{b_T}{\theta}\right)^{1/\alpha})$$

where $N_T^{\gamma}(b) = \sum_{t=1}^T \mathbb{I}(|Y_{t-1}| \leq b)$.

Observe that

$$\frac{N_T^{\gamma}(b_T)}{b_T \sqrt{T}} \to 0$$

as $T \to \infty$, where $\gamma$ is such that

$$\gamma = \frac{\alpha}{\alpha + 1}$$

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

$$\lim_{T \to \infty} P(\frac{T^{-\alpha/(\alpha+2)}}{Lb_{K,T}}(1 + \epsilon)) \leq T^{-\alpha/(\alpha+2)} R_T \leq \frac{T^{-\alpha/(\alpha+2)}}{Lb_{K,T}(1 - \epsilon)}, E_K = 1$$

for any $\epsilon > 0$. 

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

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

with

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

and then, by the Skohorod device given 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,

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

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

$$\tilde{b}_{k,T} = b_{k,T} \text{ if } \left(1 + \frac{1}{\sqrt{k}}\right)^{k} \leq \lambda(0,1) < \left(1 + \frac{1}{\sqrt{k}}\right)^{k+1}, k = -K, \ldots, K - 1,$$

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

$$b_{K,T} \leq b_T \leq \tilde{b}_{K,T} \text{ on } E_K, \text{ with } \lim_{K \to \infty} \sup_{T \geq 1} \frac{b_T}{\tilde{b}_{K,T}} = \lim_{K \to \infty} \inf_{T \geq 1} \frac{\tilde{b}_{K,T}}{b_T} = 1.$$
Hence (B.2.4), (B.2.3), \( \lim_{T,K \to \infty} \mathbb{P}(E_K) = 1 \) and definition of \( b_T \) yields Proposition 2. \( \square \)

REFERENCES


