UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

DANILO HIROSHI MATSUOKA

ESSAYS ON ASYMPTOTIC ANALYSIS OF NONPARAMETRIC REGRESSION

ENSAIOS EM ANÁLISE ASSINTÓTICA DE REGRESSÃO NÃO-PARAMÉTRICA

Porto Alegre

2020

DANILO HIROSHI MATSUOKA

ESSAYS ON ASYMPTOTIC ANALYSIS OF NONPARAMETRIC REGRESSION

ENSAIOS EM ANÁLISE ASSINTÓTICA DE REGRESSÃO NÃO-PARAMÉTRICA

Tese submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Doutor em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. Hudson da Silva Torrent

Porto Alegre

2020

DANILO HIROSHI MATSUOKA

ESSAYS ON ASYMPTOTIC ANALYSIS OF NONPARAMETRIC REGRESSION

ENSAIOS EM ANÁLISE ASSINTÓTICA DE REGRESSÃO NÃO-PARAMÉTRICA

Tese submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Doutor em Economia, com ênfase em Economia Aplicada.

Aprovada em:

BANCA EXAMINADORA:

Prof. Dr. Hudson da Silva Torrent – Orientador
UFRGS

Prof. Dr. Carlos Brunet Martins-Filho
University of Colorado Boulder

Prof. Dr. Eduardo de Oliveira Horta
UFRGS

Prof. Dr. Marcelo Brutti Righi
UFRGS

Prof. Dr. Helton Saulo Bezerra os Santos
UFRGS

Aos meus pais.

AGRADECIMENTOS

Dedico esta tese ao professor José Carlos Fernandes Rodrigues por ter me feito despertar o interesse pela Matemática de forma pioneira. Agradeço pela sua enorme paciência e amizade.

Ao meu orientador, Hudson, que tem me acompanhado por vários anos seguidos e que inspirou, em definitivo, o meu direcionamento na busca por "fazer Ciência", mesmo em ambientes tão nebulosos. Ainda sobre meu orientador, fui surpreendido com uma grande facilidade de convivência e troca de ideias. Agradeço pela sua paciência, pela sua amizade, pelo seu incentivo e por acompanhar este meu início na trajetória acadêmica.

Aos meus pais, por me darem todo o suporte necessário e por terem me feito com um grande espírito.

RESUMO

Este trabalho é composto por três ensaios na área de inferência não-paramétrica, bastante inter-relacionados. O primeiro ensaio visa estabelecer ordens de convergência uniforme sob condições *mixing* para o estimador linear local quando a estrutura de pontos é fixa e da forma $t/T, t \in \{1, \ldots, T\}, T \in \mathbb{N}$. A ordem encontrada para as convergências uniforme, em probabilidade e quase certa, é a mesma daquela estabelecida por Hansen (2008) e Kristensen (2009) para o caso de estrutura de pontos aleatórios. O segundo ensaio estuda as propriedades assintóticas de estimadores obtidos ao se inverter o esquema de estimação em três etapas de Vogt e Linton (2014). Foram fornecidas as ordens de convergência uniforme em probabilidade para os estimadores da função de tendência e da sequência periódica. Além disso, a consistência do estimador do período fundamental e a normalidade assintótica do estimador de tendência também foram estabelecidas. O último estudo investiga o comportamento em amostras finitas dos estimadores considerados no segundo ensaio. Foram propostas janelas para o estimador de tendência do tipo plug-in. Para as simulações realizadas, a janela plug-in mostrou bom desempenho e o estimador do período revelou-se bastante robusto em resposta à diferentes escolhas de janelas. O estudo foi complementado com duas aplicações, uma em climatologia e outra em economia.

Palavras chave: Econometria Não-paramétrica. Regressão Local. Teoria Assintótica. Séries Temporais. Convergência Uniforme.

ABSTRACT

This work is composed of three essays in the field of nonparametric inference, all closely inter-related. The first essay aims to stablish uniform convergence rates under mixing conditions for the local linear estimator under a fixed-design setting of the form $t/T,\ t \in \{1, \ldots, T\},\ T \in \mathbb{N}$. It was found that the order of the weak and the strong uniform convergence is the same as that of stablished by Hansen (2008) and Kristensen (2009) for the random design setting. The second essay studies the asymptotic properties of the estimators derived from reversing the three-step procedure of Vogt and Linton (2014). Weak uniform convergence rates was given to the trend and the periodic sequence estimators. Furthermore, the consistency of the fundamental period estimator and the asymptotic normality of the trend estimator was also stablished. The last study investigates the finite sample behavior of the estimators considered in the second essay. A plug-in type bandwith was proposed for the trend estimator. From our simulation results, the plug-in bandwidth performed well and the period estimator showed to be quite robust with respect to different bandwidth choices. The study was complemented with two applications, one in climatology and the other in economics.

Keywords: Nonparametric Econometrics. Local Regression. Asymptotic Theory. Time Series. Uniform Convergence.

LIST OF NOTATIONS

$(\Omega, \mathcal{F}, P)$	Probability space: Ω nonempty set, $\mathcal{F}$ σ-algebra of subsets of Ω, P probability measure on $\mathcal{F}$.
$\sigma(X_i, i \in A)$	σ-algebra generated by the random variables $X_i, i \in A$.
$\mathcal{B}_{\mathbb{R}^d}$	σ-algebra of Borel sets on $\mathbb{R}^d$.
i.i.d.	Independent and identically distributed
$N(m, \sigma^2)$	Normal distribution with mean m and variance σ^2.
$[T]^d$	The dth Cartesian power of $\{1, \ldots, T\}$.
$\lfloor \cdot \rfloor, \lceil \cdot \rceil$	Floor and ceiling functions.
$a_n \overset{a}{\approx} b_n$	$a_n/b_n \overset{n \to \infty}{\longrightarrow} 1$.
$a_n = o(b_n)$	For any $\delta > 0$, $\lvert a_n/b_n \rvert \leq \delta$ for n sufficiently large.
$a_n = O(b_n)$	For some $C > 0$, $\lvert a_n/b_n \rvert \leq C$ for n sufficiently large.
$X_n = o_p(a_n)$	For any $\delta, \epsilon > 0$, $P(\lvert X_n/a_n \rvert \geq \delta) \leq \epsilon$ for n sufficiently large.
$X_n = O_p(a_n)$	For any $\epsilon > 0$, there is $C > 0$ such that $P(\lvert X_n/a_n \rvert \geq C) \leq \epsilon$ for n sufficiently large.
$X_n = o(a_n)$a.s.	For any $\delta > 0$, $P(\limsup_{n \to \infty} \lvert X_n/a_n \rvert > \delta) = 0$.
$X_n = O(a_n)$a.s.	For some $C > 0$, $P(\limsup_{n \to \infty} \lvert X_n/a_n \rvert \leq C) = 1$.
$\overset{d}{\to}$	Convergence in distribution.
$\overset{p}{\to}$	Convergence in probability.
$\#A$	Cardinal of A.
$L^r(\Omega, \mathcal{F}, P)$	Space of classes of real $\mathcal{F} - \mathcal{B}_{\mathbb{R}}$ measurable functions f such that $\lVert f \rVert_r = (\int_\Omega \lvert f \rvert^r dP)^{1/r} < +\infty, 1 \leq r < +\infty$, and $\lVert f \rVert_\infty = \inf\{a : P(f > a) = 0\} < +\infty, r = +\infty$.

SUMMARY

1 INTRODUCTION

The first essay of this thesis develops uniform consistency results for the local linear estimator under mixing conditions in order to be directly applied in the next essays. The weak and strong uniform convergence rates were provided for general kernel averages from which we obtained the uniform rates for the local linear estimator. We restricted our attention to equally-spaced design points of the form $x_{t,T} = t/T,\ t \in \{1, \dots, T\},\ T \in \mathbb{N}$. This setting is quite common in the literature of nonparametric time series regression (ROBINSON, 1989; EL MACHKOURI, 2007; VOGT;LINTON, 2014; among others). Furthermore, it also appears in the literature of nonparametric time-varying models (DALHAUS et al., 1999; CAI, 2007) and in situations where a continuous-time process is sampled at discrete time points (BANDI; PHILLIPS, 2003; KRISTENSEN, 2010). The convergences were stablished uniformly over $[0, 1]$ under arithmetically strong mixing conditions. The kernel function was restricted to be compactly supported and Lipschitz continuous, and inlcudes the popular Epanechnikov kernel. The uniform convergence in probability was provided without imposing stationarity while the almost sure uniform convergence was proved only for the stationary case.

Hansen (2008) provided a set of results on uniform convergence rates for kernel based estimators under stationary and strongly mixing conditions. Kristensen (2009) extended the results of Hansen (2008) by allowing the data to be heterogeneously dependent as well as parameter dependent. A simple situation where the results of Kristensen (2009) could be applied relates to local linear regression models where the error process is strongly mixing without the stationarity restriction. In the literature, one can find the direct application of the results of Kristensen (2009), originally for random design, done for fixed design settings (see KRISTENSEN, 2009; VOGT; LINTON, 2014). While it is unclear, we believe that providing explicit results would not only justifies such application but also creates a background for further theoretical developments.

The second essay is the main study of this thesis. We investigated the asymptotic properties of the estimators obtained by reversing the three-step procedure of Vogt and Linton (2014), for time series modelled as the sum of a periodic and a trend deterministic components plus a stochastic error process. In the first step, the trend function is estimated; given the trend estimate, an estimate of the period is provided in the second step; the last step consists in estimating the periodic sequence. The weak uniform convergence rates of the estimators of the trend function and the periodic sequence were provided.

The asymptotic normality for the trend estimator was also stablished. Furthermore, it was shown that the period estimator is consistent.

When the data has only the slowly varying component (plus an error term), its nonparametric estimation is popularly done by using a local polynomial fitting (WATSON, 1964; NADARAYA, 1964; CLEVELAND, 1979; FAN, 1992) or a spline smoothing (WAHBA, 1990; GREEN; SILVERMAN, 1993; EUBANK, 1999). On the other hand, for models where the data is written as a periodic component plus an error term, the nonparametric estimation of the period and values of the periodic component was investigated by Sun et al. (2012) for evenly spaced fixed design points and by Hall et al. (2000) for a random design setting. A few nonparametric methods are available to address the problem of estimating models where both periodic and trend components are taken into account. As an example, there is the Singular Spectrum Analysis (BROOMHEAD; KING, 1986; BROOMHEAD et al., 1987) that have been applied in natural sciences as well as in social sciences such as economics. A more recent nonparametric method is the three-step estimation procedure proposed by Vogt and Linton (2014). In their supplementary material, they suggested that reversing the order of the estimation scheme was possible in principle. In other words, one could estimate the trend function first and subsequently estimate the period and the periodic sequence. We aimed to investigate this reversed estimation version more deeply.

The third essay exploits the bandwidth selection problem and the finite sample performance of the period estimator studied in the second essay. A plug-in type bandwidth is proposed in order to estimate the trend function and a simulation exercise showed good performance for the proposed bandwidth. Although we do not provide an optimal bandwidth selection for the period estimator, we employ another simulation exercise to evaluate the sensitivity of the estimator for different bandwidth choices having the plug-in bandwidth, as a baseline. The motivation is simple, if the performance of the period estimator along different bandwidths is roughly the same as that obtained using the first-step's bandwidth, then we would not be far worse off by choosing the plug-in bandwidth again in the second step of the reversed estimation procedure. In our simulation, the period estimator had a robust behaviour along different bandwidths. To evaluate how the estimators behave for real data, we made two applications: one for climatological data and the other for economic data. In the former, we used global temperture anomalies data which is exactly the same as that in Vogt and Linton (2014). The latter application consists in providing central estimates for the australian non-accelerating inflation rate of unemployment by means of the reversed estimation procedure.

2 UNIFORM CONVERGENCE OF LOCAL LINEAR REGRESSION FOR STRONGLY MIXING ERRORS UNDER A FIXED DESIGN SETTING

Abstract. We provide the uniform convergence rates for the local linear estimator on $[0, 1]$, under equally-spaced fixed design points of the form $x_{t,T} = t/T,\ t \in \{1, \ldots, T\},\ T \in \mathbb{N}$. The rates of weak uniform consistency are given without imposing stationarity, while the rates of strong uniform consistency are given only for stationary data. Both rates are stablished assuming the data is strongly mixing. These results explicitly show that the result of Kristensen (2009) also hold for the mentioned fixed design setting.

Keywords: Uniform convergence. Convergence in probability. Almost sure convergence. Local linear regression. Mixing process

JEL Codes. C1,C10, C14

2.1 Introduction

The uniform consistency of kernel-based estimators in discrete-time has been widely investigated under various mixing conditions (BIERENS, 1983; PELIGRAD, 1992; ANDREWS, 1995; MASRY, 1996; NZE; DOUKHAN, 2004; FAN; YAO, 2008; HANSEN, 2008; KRISTENSEN, 2009; BOSQ, 2012; KONG et al., 2010; LI et al., 2016; HIRUKAWA et al., 2019). In particular, Hansen (2008) provided a set of results on uniform convergence rates for stationary and strongly mixing data. More recently, Kristensen (2009) extended the results of Hansen (2008) by allowing the data to be heterogeneously dependent as well as parameter dependent. While the latter extension has an special relevance for some semiparametric problems (see LI; WOOLDRIDGE, 2002; XIA; HÄRDLE, 2006), the former is useful in situations where data are allowed to be nonstationary but strongly mixing, for example, in Markov-Chains that have not been initialized at their stationary distribution (YU, 1993; KIM; LEE, 2005). A simple situation where the results of Kristensen (2009) could be applied relates to local linear regression models where the error process is strongly mixing without the stationarity restriction.

In the literature, one can find the direct application of the results of Kristensen (2009), originally for random design, done for fixed design settings (see KRISTENSEN, 2009; VOGT; LINTON, 2014). While it is unclear, we believe that providing explicit results would not only justify such application but also creates a background for further theoretical developments.

In this study, we provide the weak and strong uniform convergence rates for kernel averages under fixed design and its application to the local linear estimator. We restrict our attention to equally-spaced design points of the form $x_{t,T} = t/T,\ t \in \{1, \ldots, T\},\ T \in \mathbb{N}$. This setting is quite common in the literature of nonparametric time series regression (ROBINSON, 1989; HALL; HART, 2012; EL MACHKOURI, 2007; VOGT; LINTON, 2014; among others). Furthermore, it also appears in the literature of nonparametric time-varying models (DALHAUS et al., 1999; CAI, 2007) and in situations where a continuous-time process is sampled at discrete time points (BANDI; PHILLIPS, 2003; KRISTENSEN, 2010).

The convergence is stablished uniformly over $[0, 1]$ under arithmetically strong mixing conditions. The kernel function is restricted to be compactly supported and Lipschitz continuous, and inlcudes the popular Epanechnikov kernel. The uniform convergence in probability is provided without imposing stationarity while the almost sure uniform convergence is proved only for the stationary case.

2.2 General results for kernel averages

Let $\{\epsilon_{t,T} : 1 \leq t \leq T, 1 \leq T\}$ be a triangular array of random variables on $(\Omega, \mathcal{F}, P)$. In this section, we aim to provide uniform bounds for kernel averages of the form

$$\hat{\Psi}(x) = T^{-1} \sum_{i=1}^{T} \epsilon_{i,T} K_h(i/T - x) \left(\frac{i/T - x}{h} \right)^j, \; j \in \{0, 1, \ldots, j_{\max}\}, \; x \in [0,1], \qquad (2.1)$$

where $j_{\max} \in \mathbb{N}$ is fixed, $K_h(u) \coloneqq K(u/h)/h$ with $K : \mathbb{R} \to \mathbb{R}$ being a kernel-like function and $h \coloneqq h_T$ is a positive sequence satisfying $h \to 0$ and $Th \to \infty$ as $T \to \infty$. Since the local polynomial regression estimators can be computed from simpler terms of the form (2.1), we firstly focus on providing bounds for the latter.

For each $T > 1$, the *α-mixing coefficients* of $\epsilon_{1,T}, \ldots, \epsilon_{T,T}$ is defined by

$$\alpha_T(t) = \sup_{1 \leq k \leq T-t} \sup\{|P(A \cap B) - P(A)P(B)| : B \in \mathcal{F}_{T,1}^k, A \in \mathcal{F}_{T,k+t}^T\}, \quad 0 \leq t < T,$$

where $\mathcal{F}_{T,i}^k = \sigma(\epsilon_{T,l} : i \leq l \leq k)$. By convention, set $\alpha_T(t) = 1/4$ for $t \leq 0$ and $\alpha_T(t) = 0$ for $t \geq T$. This definition is in line with Francq and Zakoïan (2005) and Withers (1981). We say that $\{\epsilon_{i,T} : 1 \leq i \leq T, 1 < T\}$ is *α-mixing* (or *strong mixing*) if the sequence

$$\alpha(t) = \sup_{T : 0 \leq t < T} \alpha_T(t), \quad 0 \leq t < \infty,$$

satisfies $\alpha(t) \to 0$ as $t \to \infty$.

Assumptions Throughout the text, we make the following assumptions:

A.1 [**Strong Mixing Conditions**] The triangular array $\{\epsilon_{i,T} : 1 \leq i \leq T, T \geq 1\}$ is strongly mixing with mixing coefficients satisfying

$$\alpha_T(i) \leq A i^{-\beta} \qquad (2.2)$$

for some finite constants $\beta, A > 0$. In addition, there exist universal constants $s > 2$ and $C > 0$ such that, uniformly over T and i,

$$E[|\epsilon_{i,T}|^s] \leq C < \infty \qquad (2.3)$$

and

$$\beta > \frac{2s - 2}{s - 2}. \qquad (2.4)$$

A.2 [**Kernel Function Conditions**] The real function K is Lipschitz continuous and has compact support, i.e., for every $u \in \mathbb{R}$, there are $L, \Lambda_1 > 0$ such that

$$K(u) = 0 \text{ for } |u| > L, \text{ and } |K(u) - K(u')| \leq \Lambda_1 |u - u'|, \forall u' \in \mathbb{R}.$$

Note that $A.2$ implies that K is bounded and integrable[1]:

$$|K(u)| \leq \bar{K} < \infty, \quad \int_{\operatorname{supp} K} |K(u)| du \leq \bar{\mu} < \infty, \tag{2.5}$$

for some constants $\bar{K}, \bar{\mu} > 0$. Furthermore, there is $\bar{C} > 0$ such that[2]

$$\int_{\operatorname{supp} K} |K(u) u^j| du \leq \bar{C} < \infty, \quad j \in \mathbb{N}. \tag{2.6}$$

Assumption A.1 specifies that the triangular array is arithmetically strong mixing. The mixing rate in (2.2) is related to the uniform moment bound in (2.3) by the condition (2.4). Clearly the parameter β, which controls the decay rate of mixing coefficients, must be greater than 2.

The boundedness and finiteness in (2.5) and (2.6) show that assumption A.2 is strong enough so that we do not need to make extra assumptions on the integrability of the Kernel function.

In what follows, we assume $L = 1$ and $\int K(w) dw = 1$ for the sake of simplicity. In addition, we will denote by $C > 0$ a generic constant which may assume different values at each appearance and does not depend on any limit variables.

2.2.1 Uniform convergence in probability

As the data is assumed to be dependent, the following variance bound involves nonzero covariances. The proof strategy of Hansen (2008) and Kristensen (2009) consists of bounding the covariances of short, medium and long lag lengths, separately. Due to our fixed design setting, this splitting procedure is unnecessary and we are able to prove the result more straightforwardly.

Theorem 2.1. *Under A.1–A.2, for all sufficiently large T, we have*

$$\operatorname{Var}(\hat{\Psi}(x)) \leq \frac{C}{Th}, \quad \forall x \in [0, 1].$$

[1]Since $|K|$ has compact support and is continuous, its image is compact, and thus bounded. Since $|K|$ is continuous, it is Lebesgue-measurable. Then $\int_{\operatorname{supp} K} |K| d\mu \leq C \int_{\operatorname{supp} K} d\mu \leq C$ as $\operatorname{supp} K$ has finite (Lebesgue) measure.

[2]Denote $f(u) := K(u) u^j$. Note that f is a compactly supported continuous real function. Then $f(\mathbb{R}) = \{0\} \cup f(\operatorname{supp} f)$ which is compact, and thus bounded. Since the functions u^j, $I(|u| \leq L)$ and K are (Lebesgue) measurable, $f(u) = K(u) u^j I(|u| \leq L)$ is also a measurable function, as well as its absolute value. Then $\int_{\mathbb{R}} |f| d\mu = \int_{-L}^{L} |f(u)| du \leq 2CL < \infty$, for some $C > 0$.

Observe that, given $\delta > 0$, Theorem 2.1 and Chebyshev's inequality imply

$$P\left(\left|\frac{\hat{\Psi}(x) - E\hat{\Psi}(x)}{1/\sqrt{Th}}\right| > \delta\right) \leq \frac{Th\,\mathrm{Var}(\hat{\Psi}(x))}{\delta^2} \leq \frac{C}{\delta^2},$$

which is sufficient to conclude that $|\hat{\Psi}(x) - E\hat{\Psi}(x)| = O_p(1/\sqrt{Th})$, pointwise, in $x \in [0,1]$.

Besides establishing a variance bound, we will also need an exponential type inequality. We state a triangular version of Theorem 2.1 of Liebscher (1996), which is derived from Theorem 5 of Rio et al. (1995).

Lemma 2.1 (Liebscher-Rio). *Let $\{Z_{i,T}\}$ be a zero-mean triangular array such that $|Z_{i,T}| \leq b_T$, with strongly mixing sequence α_T. Then for any $\epsilon > 0$ and $m_T \leq T$ such that $4b_T m_T < \epsilon$, it holds that*

$$P\left(\left|\sum_{i=1}^{T} Z_{i,T}\right| > \epsilon\right) \leq 4\exp\left[-\frac{\epsilon^2}{64\sigma^2_{T,m_T}T/m_T + \epsilon b_T m_T 8/3}\right] + 4\alpha_T(m_T)\frac{T}{m_T},$$

where $\sigma^2_{T,m_T} = \sup_{0\leq j\leq T-1} E[(\sum_{i=j+1}^{\min(j+m_T,T)} Z_{i,T})^2]$.

Now we give the uniform convergence in probability over the interval $[0,1]$. This is an adaptation of Theorem 2 of Hansen (2008).

Theorem 2.2. *Assume that A.1–A.2 hold and that, for*

$$\beta > \frac{2+2s}{s-2} \tag{2.7}$$

and

$$\theta = \frac{\beta(1-2/s) - 2 - 2/s}{\beta+2}, \tag{2.8}$$

the bandwidth satisfies

$$\frac{\phi_T \ln T}{T^\theta h} = o(1), \tag{2.9}$$

where ϕ_T is a positive slowly divergent sequence. Then, for

$$a_T = \left(\frac{\ln T}{Th}\right)^{1/2}, \tag{2.10}$$

we have $\sup_{x\in[0,1]}|\hat{\Psi}(x) - E\hat{\Psi}(x)| = O_p(a_T)$.

Theorem 2.2 establishes the rate for uniform convergence in probability. Note that (2.7) is a strengthening of (2.4). Furthermore, (2.7) together with (2.8) implies $\theta \in (0,1)$. In particular, when $\beta = +\infty$, we have $\theta = 1 - 2/s$. Therefore condition (2.9) strengthens of the conventional assumption that $Th \to \infty$.

2.2.2 Almost sure uniform convergence

In this section we establish the almost sure convergence under strict stationarity.

Theorem 2.3. *Assume that for any T, $\{\epsilon_{t,T}\}_{t=1}^T$ have the same joint distribution as $\{u_t\}_{t=1}^T$ with $\{u_t : t \in \mathbb{Z}\}$ being a strictly stationary stochastic process. Furthermore, assume that A.1–A.2 are satisfied with*

$$\beta > \frac{4s+2}{s-2} \tag{2.11}$$

and that, for

$$\theta = \frac{\beta(1-2/s) - 4 - 2/s}{\beta+2}, \tag{2.12}$$

the bandwidth satisfies

$$\frac{\phi_T^2}{T^\theta h} = O(1), \tag{2.13}$$

with $\phi_T = \ln T(\ln\ln T)^2$. Then, for

$$a_T = \left(\frac{\ln T}{Th}\right)^{1/2}, \tag{2.14}$$

we have $\sup_{x\in[0,1]}|\hat{\Psi}(x) - E\hat{\Psi}(x)| = O(a_T)$ almost surely.

2.3 Application to local linear regression

Assume that the univariate data $Y_{1,T}, Y_{2,T}, \ldots, Y_{T,T}$ are observed and that

$$Y_{t,T} = g(t/T) + \epsilon_{t,T}, \quad t \in \{1,\ldots,T\} \tag{2.15}$$

where g is a smooth continuous function on $[0,1]$ and $\{\epsilon_{t,T}\}$ is a strongly mixing triangular array of zero mean random variables.

The local linear estimator for g can be defined[3] as $\hat{g}(x) = e_1' S_T^{-1} D_T$, where

$$S_{T,x} = \frac{1}{T}\begin{bmatrix} \sum_{t=1}^T K_h(x_t - x) & \sum_{t=1}^T K_h(x_t-x)(x_t-x)/h \\ \sum_{t=1}^T K_h(x_t-x)(x_t-x)/h & \sum_{t=1}^T K_h(x_t-x)((x_t-x)/h)^2 \end{bmatrix}, \tag{2.16}$$

$$D_{T,x} = \frac{1}{T}\begin{bmatrix} \sum_{t=1}^T Y_{t,T}K_h(x_t-x) \\ \sum_{t=1}^T Y_{t,T}K_h(x_t-x)(x_t-x)/h \end{bmatrix} \text{ and } e_1 = (1,0)'. \tag{2.17}$$

For simplicity, the dependence of the design points, $x_t = t/T$, on T was omitted. It follows

[3] See Chapter 5 of Wand and Jones (1994) or Section 1.6 of Tsybakov (2008).

from this representation that

$$(S_{T,x})_{i,j} = s_{T,i+j-2}(x): \; s_{T,k}(x) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t - x}{h}\right)^k K_h(x_t - x), \; k \in \{0,1,2\}, \qquad (2.18)$$

Simple calculations show that we can also write the local linear estimator as

$$\hat{g}(x) = \sum_{t=1}^{T} W_{t,T}(x) Y_{t,T}, \qquad (2.19)$$

where $W_{t,T}(x) = T^{-1}e_1' S_{T,x}^{-1} X\left(\frac{t/T-x}{h}\right) K_h(t/T - x)$ for $X(u) = (1,u)'$. The weights $W_{t,T}$ have an useful reproducing property (see Lemma 2.6). We now give the uniform convergence rates of the local linear estimator for the model (2.15).

Theorem 2.4. *Assume the conditions of Theorem 2.2 hold. In addition, let the function g be twice continuously differentiable on $[0,1]$ and let K be nonnegative and symmetric. Then*

$$\sup_{x\in[0,1]} |\hat{g}(x) - g(x)| = O_p(a_T + h^2). \qquad (2.20)$$

If the conditions were strengthen to that of Theorem 2.3, then we have

$$\sup_{x\in[0,1]} |\hat{g}(x) - g(x)| = O(a_T + h^2) \; a.s. \qquad (2.21)$$

2.4 Proofs

Appendix A contains several lemmas (from 2.2 to 2.11) which are used in the proofs of this section.

Proof of Theorem 2.1 Let $x \in [0,1]$ and let T be large enough so that J_x, defined by (2.33) and (2.34), is well-defined. By assumptions A.1-A.2, Lemma 2.2 and Dadvydov's inequality, it follows that

$$\begin{aligned}
\mathrm{Var}(\hat{\Psi}(x)) &\le \frac{1}{T^2}\sum_{i,t\in J_x}\left| K_h(i/T - x)K_h(t/T - x)\left(\frac{i/T - x}{h}\right)^j \left(\frac{t/T - x}{h}\right)^j \mathrm{Cov}(\epsilon_{i,T}\epsilon_{t,T})\right| \\
&\le \frac{C}{(Th)^2}\sum_{i,t\in J_x} |\mathrm{Cov}(\epsilon_{i,T}\epsilon_{t,T})| \\
&\le \frac{C}{(Th)^2}\sum_{i,t\in J_x} 6\alpha_T(|i-t|)^{((s-2)/s)}(E|\epsilon_{i,T}^s|)^{1/s}(E|\epsilon_{t,T}^s|)^{1/s} \\
&\le \frac{C}{(Th)^2}\sum_{i\in J_x}\sum_{t=1}^{T}|i-t|^{-\beta((s-2)/s)} \le \frac{C}{(Th)^2}\sum_{i\in J_x}\sum_{t=1}^{T}|i-t|^{2/s-2}
\end{aligned}$$

$$\leq \frac{C}{(Th)^2} \sum_{i \in J_x} 2 \sum_{l=0}^{\infty} l^{2/s-2} \leq \sum_{i \in J_x} \frac{C}{(Th)^2} = O\left(\frac{1}{Th}\right).$$

Proof of Theorem 2.2 For the sake of brevity, denote $k_{i,T}(x) = K((i/T - x)/h)$ and $\xi_{i,T}(x) = ((i/T - x)/h)^j$, for any $x \in [0,1], T \in \mathbb{N}$ and $i \in [T]$. Further, let T be sufficiently large so that the set J_x, given by (2.33) and (2.34), is well-defined. Write

$$\hat{\Psi}(x) = \frac{1}{Th} \sum_{i=1}^{T} \epsilon_{i,T} k_{i,T}(x) \xi_{i,T}(x) I(|\epsilon_{i,T}| > \tau_T) + \frac{1}{Th} \sum_{i=1}^{T} \epsilon_{i,T} k_{i,T}(x) \xi_{i,T}(x) I(|\epsilon_{i,T}| \leq \tau_T)$$
$$:= R_{1,T}(x) + R_{2,T}(x), \tag{2.22}$$

where I is the indicator function and $\tau_T = \rho_T (Th)^{1/s}$ with $\rho_T = (\ln T)^{1/(1+\beta)} \phi_T^{(1+\beta/2)/(1+\beta)}$. Using Holder's and Markov's inequalities, we have that

$$\begin{aligned} E(|\epsilon_{i,T}| I(|\epsilon_{i,T}| > \tau_T)) &\leq [E(|\epsilon_{i,T}|^s)]^{1/s} [E(I(|\epsilon_{i,T}| > \tau_T))]^{1-1/s} \\ &= [E(|\epsilon_{i,T}|^s)]^{1/s} [P(|\epsilon_{i,T}| > \tau_T)]^{1-1/s} \\ &\leq [E(|\epsilon_{i,T}|^s)]^{1/s} \left[\frac{E(|\epsilon_{i,T}|^s)}{\tau_T^s}\right]^{1-1/s} = E(|\epsilon_{i,T}|^s) \tau_T^{1-s}. \end{aligned} \tag{2.23}$$

It follows by (2.23), Assumption A.2 and Lemma 2.2 that

$$\begin{aligned} |ER_{1,T}(x)| \leq E|R_{1,T}(x)| &\leq \frac{1}{Th} \sum_{i \in J_x} |k_{i,T}(x) \xi_{i,T}(x)| E(|\epsilon_{i,T}|^s) \tau_T^{1-s} \\ &\leq \sum_{i \in J_x} \frac{C \tau_T^{1-s}}{Th} = O(\tau_T^{1-s}) = o(a_T), \end{aligned} \tag{2.24}$$

since, for $s > 2$,

$$\frac{\tau_T^{1-s}}{a_T} = \rho_T^{1-s} T^{1/s-1/2} \left(\frac{h}{\ln T}\right)^{1/2} = o(1).$$

Hence $\sup_{x \in [0,1]} |ER_{1,T}(x)| = o(a_T)$. From this, we cannot say much about the order of $\sup_{x \in [0,1]} |R_{1,T}(x)|$. Note that

$$\begin{aligned} & w \in \left\{ w : \sup_x \left| \sum_{i \in J_x} k_{i,T}(x) \xi_{i,T}(x) \epsilon_{i,T}(w) I(|\epsilon_{i,T}|(w) > \tau_T) \right| > C a_T \right\} \\ & \implies \exists i \in J_x : w \in \{|\epsilon_{i,T}|(w) > \tau_T\} \\ & \implies w \in \bigcup_{i \in J_x} \{|\epsilon_{i,T}|(w) > \tau_T\}. \end{aligned}$$

By the monotonicity and subadditivity of the measure, and using Markov's inequality, we have

$$P\Big(\sup_x |R_{1,T}| > Ca_T\Big) \leq \sum_{i\in J_x} P(|\epsilon_{i,T}| > \tau_T) \leq \sum_{i\in J_x} \frac{E(|\epsilon_{i,T}|^s)}{\tau_T^s}$$
$$\leq C\frac{Th}{\tau_T^s} \leq \frac{C}{\phi_T} = o(1). \tag{2.25}$$

From expressions (2.24), (2.25), Lemma 2.9(v) and the triangle inequality,

$$\sup_{x\in[0,1]} |R_{1,T}(x) - ER_{1,T}(x)| \leq \sup_{x\in[0,1]} |R_{1,T}(x)| + \sup_{x\in[0,1]} |ER_{1,T}(x)|$$
$$= O_p(a_T) + o(a_T) = O_p(a_T).$$

Lemma 2.9(iv) implies that $\sup_x |R_{1,T}(x) - ER_{1,T}(x)| = O_p(a_T)$. The replacement of $\epsilon_{i,T}$ by the bounded variable $\epsilon_{i,T} I(|\epsilon_{i,T}| \leq \tau_T)$ produce an error of order $O_p(a_T)$, uniformly in x.

Now, we focus on the term $R_{2,T}(x)$. We shall construct a grid of N points on $A = [0,1]$. Let $A_j = \{x \in \mathbb{R} : |x - x_j| \leq a_T h\},\ j \in \mathbb{N}$. For $N = \lceil 1/(a_T h) \rceil$, it is easy to see that there is at least one set E such that $E = \cup_{j=1}^{N} A_j$ and $A \subseteq E$. The grid is obtained by selecting each $x_j \in E$ as grid points.

Make the following definitions

$$\tilde{\Psi}(x) = (Th)^{-1} \sum_{i=1}^{T} |k_{i,T}^*(x) \epsilon_{i,T}^*|;$$
$$\bar{\Psi}(x) = (Th)^{-1} \sum_{i=1}^{T} |k_{i,T}(x) \epsilon_{i,T}^*|;$$

where $\epsilon_{i,T}^* = \epsilon_{i,T} I\{|\epsilon_{i,T}| \leq \tau_T\}$ and $k_{i,T}^*(x) = K^*((i/T - x)/h)$ with $K^*(x) = \Lambda_1 I(|x| \leq 2L)$. By our convention (and without loss of generality), $L = 1$. From assumption A.1, it follows that

$$E|\tilde{\Psi}(x)| \leq \frac{C}{Th} \sum_{i\in G_x} E|\epsilon_{i,T}^*| \leq \frac{C}{Th} \sum_{i\in G_x} E|\epsilon_{i,T}| \leq C, \tag{2.26}$$

for some $C > 0$ and all T large enough, where $G_x = \{i \in [T] : i/T \in C_x\}$ with C_x given by (2.36). Analogously, we can show that $E|\bar{\Psi}(x)| = O(1)$.

If $x \in A_l$, then $|x - x_l|/h \leq a_T$ by definition. Also, as $a_T = o(1)$, we eventually have $a_T \leq 1$. Thus, for each $A_l,\ l \in \{1, \ldots, N\}$, for $x \in A_l$ and T sufficiently large, Lemma 2.3 with $\delta = a_T$ gives

$$|R_{2,T}(x) - R_{2,T}(x_l)| \leq \frac{1}{Th} \sum_{i=1}^{T} |\epsilon_{i,T}^*| |\xi_{i,T}(x) k_{i,T}(x) - \xi_{i,T}(x_l) k_{i,T}(x_l)| I(i \in D_x \cup D_{x_l})$$

$$
\begin{aligned}
&\leq \frac{1}{Th}\sum_{i\in D_x\cup D_{x_l}} |\epsilon^*_{i,T}|\{|k_{i,T}(x)||\xi_{i,T}(x)-\xi_{i,T}(x_l)| \\
&+ |\xi_{i,T}(x_l)||k_{i,T}(x)-k_{i,T}(x_l)|\} \\
&\leq \frac{1}{Th}\sum_{i\in D_x\cup D_{x_l}} |\epsilon^*_{i,T}|\Bigg\{|k_{i,T}(x)|\left|\frac{x_l-x}{h}\right|\sum_{l=0}^{j-1}\left|\frac{i/T-x}{h}\right|^{l}\left|\frac{i/T-x_l}{h}\right|^{j-1-l} \\
&+ \left|\frac{i/T-x_l}{h}\right|^{j} a_T k^*_{i,T}(x_l)\Bigg\} \\
&\leq \frac{1}{Th}\sum_{i\in D_x\cup D_{x_l}} |\epsilon^*_{i,T}|\Big\{|k_{i,T}(x)|a_T j + a_T k^*_{i,T}(x_l)\Big\} \\
&\leq \frac{a_T j}{Th}\sum_{i=1}^{T}|k_{i,T}(x)\epsilon^*_{i,T}| + \frac{a_T}{Th}\sum_{i=1}^{T}|k^*_{i,T}(x_l)\epsilon^*_{i,T}| \\
&= a_T j\bar{\Psi}(x) + a_T\tilde{\Psi}(x_l), \qquad (2.27)
\end{aligned}
$$

where $D_x = \{i \in [T] : |(i/T - x)/h| \leq 1\}$ for any $x \in \mathbb{R}$. By applying the same arguments used in expression (2.27), for $j = 0$, we obtain that $|\bar{\Psi}(x) - \bar{\Psi}(x_l)| \leq a_T\tilde{\Psi}(x_l)$. Using expressions (2.26)-(2.27), for each $l = 1, \ldots, N$, and for all sufficiently large T, we have

$$
\begin{aligned}
&\sup_{x\in A_l}|R_{2,T}(x) - ER_{2,T}(x)| \leq \sup_{x\in A_j}\{|R_{2,T}(x_l) - ER_{2,T}(x_l)| \\
&+ |R_{2,T}(x) - R_{2,T}(x_l)| + E|R_{2,T}(x_l) - R_{2,T}(x)|\} \\
&\leq \sup_{x\in A_l}\{|R_{2,T}(x_l) - ER_{2,T}(x_l)| + a_T j\bar{\Psi}(x) + a_T\tilde{\Psi}(x_l) + E(a_T j\bar{\Psi}(x) + a_T\tilde{\Psi}(x_l))\} \\
&= |R_{2,T}(x_l) - ER_{2,T}(x_l)| + a_T[\tilde{\Psi}(x_l) + E\tilde{\Psi}(x_l)] + a_T j\sup_{x\in A_l}[\bar{\Psi}(x) + E\bar{\Psi}(x)] \\
&\leq |R_{2,T}(x_l) - ER_{2,T}(x_l)| + a_T(|\tilde{\Psi}(x_l) - E\tilde{\Psi}(x_l)| + 2|E\tilde{\Psi}(x_l)|) + a_T j\sup_{x\in A_l}[\bar{\Psi}(x) + E\bar{\Psi}(x)] \\
&\leq |R_{2,T}(x_l) - ER_{2,T}(x_l)| + |\tilde{\Psi}(x_l) - E\tilde{\Psi}(x_l)| + Ca_T + j\sup_{x\in A_l}[\bar{\Psi}(x) + E\bar{\Psi}(x)] \\
&:= B_{1,l} + B_{2,l} + Ca_T + j\sup_{x\in A_l}[\bar{\Psi}(x) + E\bar{\Psi}(x)].
\end{aligned}
$$

Along the above lines,

$$
\begin{aligned}
&\sup_{x\in A_l}|\bar{\Psi}(x) + E\bar{\Psi}(x)| \leq \sup_{x\in A_l}\{|\bar{\Psi}(x) - E\bar{\Psi}(x)| + 2|E\bar{\Psi}(x)|\} \\
&\leq \sup_{x\in A_j}\{|\bar{\Psi}(x_l) - E\bar{\Psi}(x_l)| + |\bar{\Psi}(x) - \bar{\Psi}(x_l)| + E|\bar{\Psi}(x_l) - \bar{\Psi}(x)|\} + C \\
&\leq |\bar{\Psi}(x_l) - E\bar{\Psi}(x_l)| + a_T(\tilde{\Psi}(x_j) + E\tilde{\Psi}(x_j)) + C \\
&\leq |\bar{\Psi}(x_l) - E\bar{\Psi}(x_l)| + |\tilde{\Psi}(x_j) - E\tilde{\Psi}(x_j)| + C \\
&:= B_{3,l} + B_{2,l} + C
\end{aligned}
$$

for T sufficiently large. Therefore, when T is large enough, we have

$$\sup_{x\in A_l}|R_{2,T}(x) - ER_{2,T}(x)| \leq \gamma(B_{1,l} + B_{2,l} + B_{3,l} + Ca_T), \quad l \in \{1,\dots,N\} \tag{2.28}$$

where $\gamma = 1 + j_{\max}$.

Define $e(x) = |R_{2,T}(x) - ER_{2,T}(x)|$. Since $A = [0,1] \subseteq \bigcup_{l=1}^{N} A_l$, it follows that $\sup_{x\in A} e(x) \leq \sup_{x\in\cup A_l} e(x)$ which implies

$$\Big\{\sup_{x\in A} e(x) > 4\gamma Ca_T\Big\} \subseteq \Big\{\sup_{x\in\cup A_l} e(x) > 4\gamma Ca_T\Big\}.$$

In addition,

$$\begin{aligned} w \in \Big\{\sup_{x\in\cup A_i} e(x) > 4\gamma Ca_T\Big\} &\implies \exists i : 1 \leq i \leq N : w \in \Big\{\sup_{x\in A_i} e(x) > 4\gamma Ca_T\Big\} \\ &\implies w \in \bigcup_i \Big\{\sup_{x\in A_i} e(x) > 4\gamma Ca_T\Big\}. \end{aligned}$$

Thus, from inequality (2.28), Lemma 2.11, the monotonicity and subadditivity of the measure,

$$\begin{aligned} &P\Big(\sup_{x\in A}|R_{2,T}(x) - ER_{2,T}(x)| > 4\gamma Ca_T\Big) \leq P\Big(\sup_{x\in\cup A_l}|R_{2,T}(x) - ER_{2,T}(x)| > 4\gamma Ca_T\Big) \\ &\leq \sum_{l=1}^{N} P\Big(\sup_{x\in A_l} e(x) > 4\gamma Ca_T\Big) \leq N \max_{1\leq l\leq N} P\Big(\sup_{x\in A_l} e(x) > 4\gamma Ca_T\Big) \\ &\leq N \max_{1\leq l\leq N} P\Big(\gamma B_{1,l} + \gamma B_{2,l} + \gamma B_{3,l} > 4\gamma Ca_T\Big) \\ &\leq N \max_{1\leq l\leq N} P\Big(B_{1,l} > a_T C\Big) + N \max_{1\leq l\leq N} P\Big(B_{2,l} > a_T C\Big) + N \max_{1\leq l\leq N} P\Big(B_{3,l} > a_T C\Big) \\ &:= T_1 + T_2 + T_3, \end{aligned} \tag{2.29}$$

for sufficiently large T.

We start bounding the term T_1. Let $Z_{i,T}(x) = \epsilon^*_{i,T} k_{i,T}(x)\xi_{i,T}(x) - E(\epsilon^*_{i,T} k_{i,T}(x)\xi_{i,T}(x))$. It is clear that $|Z_{i,T}(x)| \leq 2\bar{K}\tau_T \leq C_1\tau_T := b_T$ for some $C_1 > 0$, since $|\epsilon^*_{i,T}| \leq \tau_T$ and $|k_{i,T}(x)| \leq \bar{K}$. Set $m_t = (a_T\tau_T)^{-1}$ and $\epsilon = Ma_T Th$. Following the proof of Theorem 2.1, we can obtain that the sequence σ^2_{T,m_T} defined in Lemma 2.1 is $O(m_T h)$. Also, note that

$$m_T \leq \frac{1}{a_T} \leq T^{1/2}\left(\frac{h}{\ln T}\right)^{1/2} \leq T^{1/2} \leq T$$

for all sufficiently large T, and

$$\frac{m_T b_T}{a_T Th} = \frac{C_1}{a_T^2 Th} = \frac{C_1}{\ln T} \to 0.$$

These facts show that the conditions of Liebscher-Rio's Lemma are satisfied whenever T is large enough. Therefore, for any x, and T sufficiently large, we apply Liebscher-Rio's Lemma to obtain

$$\begin{aligned} P(|R_{2,T}(x) - ER_{2,T}(x)| > Ca_T) &= P\left(\left|\sum_{i=1}^{T} Z_{i,T}(x)\right| > Ca_T Th\right) \\ &\leq 4\exp\left[-\frac{(Ca_T Th)^2}{64\sigma^2_{T,m_T} T/m_T + (Ca_T Th) b_T m_T 8/3}\right] \\ &+ 4\alpha_T(m_T)\frac{T}{m_T} \\ &\leq 4\exp\left[-\frac{(Ca_T Th)^2}{64CTh + 6C_1 CTh}\right] + 4(Am_T^{-\beta})\frac{T}{m_T} \\ &\leq 4\exp\left[-\frac{(Ca_T)^2 Th}{64C + 6C_1 C}\right] + 4Am_T^{-1-\beta} T \\ &= 4\exp\left[-\frac{Ca_T^2 Th}{64 + 6C_1}\right] + 4Am_T^{-1-\beta} T \\ &= 4\exp\left[-\frac{C}{64 + 6C_1}\ln T\right] + 4Am_T^{-1-\beta} T \\ &= 4T^{-C/(64+6C_1)} + 4AT(a_T\tau_T)^{1+\beta}. \end{aligned} \tag{2.30}$$

The bound (2.30) holds for T_2 and T_3, which can be checked by the same arguments used for T_1. Recalling that N is asymptotically equivalent to $1/(a_T h)$, it follows from (2.29) that

$$\begin{aligned} T_1 + T_2 + T_3 &= O(T^{-C/(64+6C_1)}/(a_T h)) + O(T(a_T\tau_T)^{1+\beta}/(a_T h)) \\ &:= O(S_1) + O(S_2). \end{aligned} \tag{2.31}$$

Now we show that S_1 and S_2 are $o(1)$. Since $C > 0$ can be arbitrarily large, $\forall\eta > 0 : \exists C^* : \forall C > C^* : S_1 \leq T^{-\eta}$. Therefore $S_1 = o(1)$ for any $C > 0$ large enough. On the other hand, we have

$$\begin{aligned} S_2 &= \frac{h^{(1+\beta)/s}}{h^{\beta/2}}\frac{h}{h}(\ln T\phi_T)^{1+\beta/2} T^{1-\beta/2+(1+\beta)/s} = o\left[\left(\frac{\ln T\phi_T}{h}\right)^{1+\frac{\beta}{2}}\right] T^{1-\beta/2+(1+\beta)/s} \\ &= o(T^{\theta(2+\beta)/2+1-\beta/2+(1+\beta)/s}) = o(1), \end{aligned}$$

since $\phi_T \ln T/h = o(T^\theta)$ and

$$\theta\left(\frac{2+\beta}{2}\right) = -1 + \frac{\beta}{2} - \frac{\beta+1}{s},$$

by hypothesis. This shows that $\sup_{x\in[0,1]}|R_{2,T}(x) - ER_{2,T}(x)| = O_P(a_T)$. It completes the proof.

Proof of Theorem 2.3 We will use the same notation as for the proof of Theorem 2.2. Also, use the shorthand, $\sup_x := \sup_{x\in[0,1]}$. Let $\tau_T = (T\phi_T)^{1/s}$. As in (2.24), it follows that $|ER_{1,T}(x)| = O(a_T)$, or equivalently, for some $M_1 > 0$ and $T^* \in \mathbb{N}$, $T \geq T^*$ implies $|ER_{1,T}(x)| \leq M_1 a_T$. Therefore, for any $T > T^*$,

$$
\begin{aligned}
&P(\sup_x |R_{1,T}(x) - ER_{1,T}(x)| > M_1 a_T) \leq P(\sup_x |R_{1,T}(x)| + M_1 a_T > M_1 a_T) \\
&= P(\sup_x |R_{1,T}(x)| > 0) \leq P(|\epsilon_{i,T}| > \tau_T \text{ for some } i \in \{1,\ldots,T\}) \\
&= P(|u_T| > \tau_T),
\end{aligned}
$$

using the triangle inequality, the monotonicity of the measure and the strict stationarity assumption. Further, Markov's inequality gives[4]

$$
\sum_{T=1}^{\infty} P(|u_T| > \tau_T) \leq 2 + \sum_{T=3}^{\infty} \frac{C}{\tau_T^s} \leq 2 + \sum_{T=3}^{\infty} \frac{1}{T \ln T (\ln\ln T)^2} < \infty. \tag{2.32}
$$

Hence

$$
\begin{aligned}
&\sum_{T=1}^{\infty} P(\sup_x |R_{1,T}(x) - ER_{1,T}(x)| > M_1 a_T) \leq T^* + \sum_{T=T^*+1}^{\infty} P(|u_T| > \tau_T) \\
&\leq T^* + \sum_{T=1}^{\infty} P(|u_T| > \tau_T) < \infty.
\end{aligned}
$$

The application of Borel-Cantelli's Lemma yields,

$$
\begin{aligned}
&P(\limsup_T \{\sup_x |R_{1,T}(x) - ER_{1,T}(x)| > M_1 a_T\}) = 0 \\
&\iff P(\liminf_T \{\sup_x |R_{1,T}(x) - ER_{1,T}(x)| \leq M_1 a_T\}) = 1 \\
&\implies P(\limsup_T \{\sup_x |R_{1,T}(x) - ER_{1,T}(x)| \leq M_1 a_T\}) = 1,
\end{aligned}
$$

that is, $\sup_x |R_{1,T}(x) - ER_{1,T}(x)| = O(a_T)$ almost surely (a.s.).

Next, one can check that (2.30) and (2.31) hold for $\tau_T = (T\phi_T)^{1/s}$. Setting $A_j = \{x \in \mathbb{R} : |x - x_j| \leq a_T h \ln\ln T\}$, then $N \overset{a}{\approx} (a_T h \ln\ln T)^{-1}$. By hypothesis, it follows that

$$
\begin{aligned}
S_1 &= \frac{T^{-C/(64+6C_1)+1/2}}{(\phi_T h)^{1/2}} = T^{-C/(64+6C_1)+1/2} O\left(\frac{T^\theta}{\phi_T^{3/2}}\right) = T^{-C/(64+6C_1)+(1+\beta)/2} O\left(\frac{1}{\phi_T^{3/2}}\right) \\
&= o(T^{-1}) o(\phi_T^{-1}) = o((T\phi_T)^{-1}),
\end{aligned}
$$

---

[4] See page 63 of Rudin (1976).

for M sufficiently large, and that

$$\begin{aligned}S_2 &= T\left(\frac{\ln T}{Th}\right)^{\beta/2}\frac{(T\phi_T)^{(1+\beta)/s}}{h\ln\ln T} = \frac{T^{1-\beta/2+(1+\beta)/s}}{h^{1+\beta/2}}o(\phi_T^{\beta/2+(1+\beta)/s})\\ &= O(T^{1-\beta/2+(1+\beta)/s+\theta(1+\beta/2)})o(\phi_T^{\beta/2+(1+\beta)/s-2-\beta})\\ &= o(T^{1-\beta/2+(1+\beta)/s+\theta(1+\beta/2)}\phi_T^{-1+[(1+\beta)/s-1-\beta/2]})\\ &= o((T\phi_T)^{-1}).\end{aligned}$$

To see the last inequality, note that conditions (2.11) and (2.12) imply

$$\theta\left(\frac{2+\beta}{2}\right) = -2 + \frac{\beta}{2} - \frac{\beta+1}{s}$$

and

$$\begin{aligned}\frac{4s+2}{s-2} < \beta \iff 4s+2 < \beta(s-2) \iff \frac{4-\beta}{2} < -\frac{\beta+1}{s} \iff \frac{\beta}{2} - 2 > \frac{\beta+1}{s}\\ \implies \frac{\beta}{2} + 1 > \frac{\beta+1}{s},\end{aligned}$$

respectively. Since the series $\sum_T (T\phi_T)^{-1}$ converges, Borel-Cantelli's Lemma implies

$$P\Big(\limsup_{T\to\infty}\{\sup_{x\in[0,1]}|R_{2,T}(x) - ER_{2,T}(x)| > 4\gamma C a_T\}\Big) = 1$$

as desired.

Proof of Theorem 2.4 Write

$$|\hat{g}(x) - g(x)| \le |\hat{g}(x) - E\hat{g}(x)| + |E\hat{g}(x) - g(x)| := A_1 + A_2, \quad \forall x \in [0,1].$$

We start with the bias term A_2. Using Lemmas 2.5 and 2.8, and Taylor expansion with Lagrange reminder, we have that for any $x \in [0,1]$ and any T sufficiently large

$$\begin{aligned}A_2 &= \left|\sum_{t=1}^T W_{t,T}(x)\{g(t/T) - g(x)\}\right|\\ &= \left|\sum_{t=1}^T W_{t,T}(x)\{g(x) + g'[x+\tau_t(t/T-x)](t/T-x) - g(x)\}\right|\\ &= \left|\sum_{t=1}^T W_{t,T}(x)\{g'[x+\tau_t(t/T-x)](t/T-x)\} - \sum_{t=1}^T W_{t,T}(x)(t/T-x)g'(x)\right|\\ &\le \sum_{t=1}^T |W_{t,T}(x)||t/T-x|\big|g'(x+\tau_t(t/T-x)) - g'(x)\big|\end{aligned}$$

$$\le C\sum_{t=1}^{T}|W_{t,T}(x)||t/T-x|^2 = C\sum_{t=1}^{T}|W_{t,T}(x)||t/T-x|^2 I\left(\left|\frac{t/T-x}{h}\right|\le 1\right)$$
$$\le C\sum_{t=1}^{T}\sup_x|W_{t,T}(x)|h^2 \le Ch^2,$$

with $\tau_t \in (0,1)$. The second inequality above holds since $g \in C^2[0,1]$ implies g' is Lipschitz continuous on $[0,1]$. Thus $\sup_{x\in[0,1]} A_2 = O(h^2)$.

Turning to the next term, we have

$$A_1 = |e_1' S_{T,x}^{-1} D_{T,x}^{\epsilon}|$$

where $D_{T,x}^{\epsilon} = T^{-1}\begin{bmatrix} \sum_{i=1}^{T}\epsilon_{i,T}K_h(t/T-x) \\ \sum_{i=1}^{T}\epsilon_{i,T}K_h(t/T-x)((t/T-x)/h)\end{bmatrix} := \begin{bmatrix} d_{T,0}(x) \\ d_{T,1}(x)\end{bmatrix}$.

Therefore, we can write

$$A_1 = \left| e_1' \begin{bmatrix} s_0 & s_1 \\ s_1 & s_2 \end{bmatrix}^{-1} \begin{bmatrix} d_0 \\ d_1 \end{bmatrix} \right| = \left|\frac{d_0 - s_1^2 s_2^{-1} d_1}{s_0 - s_1^2 s_2^{-1}}\right| := \frac{V_n}{V_d},$$

omitting the dependence of the entries on x and T, for brevity's sake. Note that the fact $||s_j| - |\mu_j|| \le |s_j - \mu_j|$ guarantees that $|s_j| = |\mu_j| + O(1/(Th))$ holds in Lemma 2.6. In addition, for any $x \in [0,1]$, we have $0 < \mu_j \le C$ for $j \in \{0,2\}$ and $|\mu_1| \le C$ by hypothesis. It implies $\mu_1^2/\mu_2 = O(1)$. Thus, from Lemma 2.6, Lemma 2.9, and Theorem 2.2, we have

$$\sup_{x\in[0,1]} V_n \le \sup_{x\in[0,1]}|d_0| + \sup_{x\in[0,1]}|s_1^2 s_2^{-1}| \sup_{x\in[0,1]}|d_1| = O_p(a_T)\left\{1 + \sup_{x\in[0,1]}\frac{|\mu_1^2| + O(1/(Th))}{|\mu_2| + O(1/(Th))}\right\}$$
$$= O_p(a_T)\left\{1 + \sup_{x\in[0,1]}\left|\frac{\mu_1^2}{\mu_2}\right| + O\left(\frac{1}{Th}\right)\right\} = O_p(a_T)\left\{O(1) + O\left(\frac{1}{Th}\right)\right\}$$
$$= O_p(a_T)O(1) = O_p(a_T),$$

and

$$V_d = \left|\mu_0 + O\left(\frac{1}{Th}\right) - \frac{\mu_1^2 + O(1/(Th))}{\mu_2 + O(1/(Th))}\right| = \left|\mu_0 - \frac{\mu_1^2}{\mu_2} + O\left(\frac{1}{Th}\right)\right|.$$

Lemma 2.7 states that $S_{T,x}$ has a positive definite limiting matrix, implying that $\mu_0\mu_2 - \mu_1^2 \ne 0$. Then

$$\sup_{x\in[0,1]} A_1 \le O_p(a_T)\sup_{x\in[0,1]}\left|\frac{1}{\mu_0 - \mu_1^2/\mu_2 + O(1/(Th))}\right| = O_p(a_T)\sup_{x\in[0,1]}\left|\frac{\mu_2}{\mu_0\mu_2 - \mu_1^2} + O\left(\frac{1}{Th}\right)\right|$$
$$= O_p(a_T)O(1) = O_p(a_T).$$

Lemma 2.9(v) implies

$$\sup_{x\in[0,1]} |\hat{g}(x) - g(x)| = O(h^2) + O_p(a_T) = O_p(h^2 + a_T),$$

as desired.

The almost sure uniform convergence rate can be shown using the same arguments and Lemma 2.10

2.5 References

ANDREWS, D. W. Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, v. 11, n. 3, p. 560-596, 1995. Available in <https://www.jstor.org/stable/3532948?seq=1>. Acessed on 25/08/2020.

BANDI, F. M.; PHILLIPS, P. Fully nonparametric estimation of scalar diffusion models. *Econometrica*, v. 71, n. 1, p. 241-283, 2003. Available in <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00395>. Acessed on 25/08/2020.

BIERENS, H. J. Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association*, v. 78, n. 383, p. 699-707, 1983. Available in <https://www.jstor.org/stable/2288140?seq=1>. Acessed on 25/08/2020.

BOSQ, D., *Nonparametric Statistics for Stochastic Processes*: Estimation and Prediction. Lecture Notes in Statistics. Springer New York, 2012.

CAI, Z. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, v. 136, n. 1, p. 163-188, 2007. Available in <https://www.sciencedirect.com/science/article/abs/pii/S0304407605002058>. Acessed on 25/08/2020.

DALHAUS, R.; NEUMANN, M. H.; VON SACHS, R. Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli*, v. 5, n. 5, p. 873-906, 1999. Available in <https://projecteuclid.org/euclid.bj/1171290403>. Acessed on 25/08/2020.

EL MACHKOURI, M. Nonparametric regression estimation for random fields in a fixed-design. *Statistical Inference for Stochastic Processes*, v. 10, n. 1, p. 29-47, 2007. Available in <https://link.springer.com/article/10.1007/s11203-005-7332-6>. Acessed on 25/08/2020.

FAN, J.; YAO, Q. *Nonlinear time series*: nonparametric and parametric methods. Springer Science & Business Media, 2008.

FERNÁNDEZ, M. F.; FERNÁNDEZ, J. M. V. Local polynomial estimation with correlated errors. *Communications in Statistics - Theory and Methods*, v. 30, n. 7, p. 1271-1293, 2001. Available in <https://www.tandfonline.com/doi/abs/10.1081/STA-100104745>. Acessed on 25/08/2020.

FRANCQ, C.; ZAKOÏAN, J. M. A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size. *Econometric Theory*, v. 21, n. 6, p. 1165-1171, 2005. Available in <https://www.jstor.org/stable/3533463?seq=1>. Acessed on 25/08/2020.

HALL, P.; HART, J. D. Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications*, v. 36, n. 2, p. 339-351, 1990. Available in <https://www.sciencedirect.com/science/article/pii/0304414990901007>. Acessed on 25/08/2020.

HANSEN, B. E. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, v. 24, n. 3, p. 726-748, 2008. Available in <https://www.jstor.org/stable/20142515?seq=1>. Acessed on 01/05/2020.

KONG, E.; LINTON, O.; XIA, Y. Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, v. 26, n. 5, p. 1529-1564, 2010. Available in <https://www.jstor.org/stable/40800891?seq=1>. Acessed on 25/08/2020.

KRISTENSEN, D. Uniform Convergence Rates of Kernel Estimators with Heterogeneous Dependent Data. *Econometric Theory*, v. 25, n. 5, p.1433-1445, 2009. Available in <https://www.jstor.org/stable/40388594?seq=1>. Acessed on 25/08/2020.

KRISTENSEN, D. Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory*, v. 26, n. 1, p 60-93, 2010. Available in <https://www.jstor.org/stable/40388620?seq=1>. Acessed on 25/08/2020.

LI, X.; YANG, W.; HU, S. Uniform convergence of estimator for nonparametric regression with dependent data. *Journal of Inequalities and Applications*, v. 2016, n. 142, p. 1-12, 2016. Available in <https://journalofinequalitiesandapplications.springeropen.com/articles/10.1186/s13660-016-1087-z>. Acessed on

25/08/2020.

LI, Q.; WOOLDRIDGE, J. M. Semiparametric estimation of partially linear models for dependent data with generated regressors. *Econometric Theory*, v. 18, n. 3, p. 625-645, 2002. Available in <https://www.jstor.org/stable/3533642?seq=1>. Acessed on 25/08/2020.

LIEBSCHER, E. Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications*, v. 65, n. 1, p. 69-80, 1996. Available in <https://www.sciencedirect.com/science/article/pii/S0304414996000968>. Acessed on 01/05/2020.

MASRY, E. Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, v. 17, n. 6, p. 571-599, 1996. Available in <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.1996.tb00294.x>. Acessed on 25/08/2020.

MÜLLER, Hans-Georg. Smooth optimum kernel estimators near endpoints. *Biometrika*, v. 78, n. 3, p. 521-530, 1991. Available in <https://academic.oup.com/biomet/article-abstract/78/3/521/255902?redirectedFrom=fulltext >. Acessed on 25/08/2020.

NZE, P. A.; DOUKHAN, P. Weak dependence: models and applications to econometrics. *Econometric Theory*, v. 20, n. 6, p. 995-1045, 2004. Available in <https://www.jstor.org/stable/3533446?seq=1>. Acessed on 25/08/2020.

PELIGRAD, M. Properties of uniform consistency of the kernel estimators of density and regression functions under dependence assumptions. *Stochastics and Stochastics Reports*, v. 40, n. 3-4, p. 147-168, 1992. Available in <https://www.tandfonline.com/doi/abs/10.1080/17442509208833786>. Acessed on 25/08/2020.

RIO,E. The functional law of the iterated logarithm for stationary strongly mixing sequences. *The Annals of Probability*, v. 23, n. 3, p. 1188-1203, 1995. Available in <https://www.jstor.org/stable/2244868?seq=1>. Acessed on 01/05/2020.

RIO, E. *Asymptotic theory of weakly dependent random processes*. Springer-Verlag Berlin Heidelberg, 2017.

ROBINSON, P. M. Nonparametric estimation of time-varying parameters. In: HACKL, P. (Org.). *Statistical Analysis and Forecasting of Economic Structural Change*. Springer Berlin, p. 253-264, 1989. Available in <https://www.jstor.org/stable/43305599?seq=1>. Acessed on 25/08/2020.

RUDIN, W. *Principles of Mathematical Analysis*. International series in pure and applied mathematics. McGraw-Hill, 3rd edition, 1976.

TSYBAKOV, A. B. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2008.

VOGT, M.; LINTON, O. Nonparametric estimation of a periodic sequence in the presence of a smooth trend. *Biometrika*, v. 101, n. 1, p. 121-140, 2014. Available in <https://www.jstor.org/stable/43305599?seq=1>. Acessed on 25/08/2020.

WAND, M.; JONES, M. *Kernel smoothing*. Chapman & Hall/CRC, 1994.

WITHERS, C. S. Central limit theorems for dependent variables I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, v. 57, n. 4, p. 509-534, 1981. Available in <https://link.springer.com/article/10.1007/BF01025872>. Acessed on 25/08/2020.

XIA, Y.; HÄRDLE, W. Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, v. 97, n. 5, p. 1162-1184, 2006. Available in <https://www.sciencedirect.com/science/article/pii/S0047259X05001995>. Acessed on 25/08/2020.

Appendix A - Auxiliary results

The quantity $\hat{\Psi}(x)$ involves a sum over the set of indices $\{i\}_{i=1}^{T}$. Since the kernel function is assumed to be compactly supported, we only need to consider a subset of indices $J_x \subseteq \{1, \ldots, T\}$, which depends on the point $x \in [0,1]$. It is important to distinghish between x as an interior point and x as a boundary point of $[0,1]$ once the respective kernel averages may be related to different asymptotic equivalences. Analytically, we can examine the behaviour of the kernel average "near" the boundaries instead of its behaviour at the boundaries. Indeed, this approach is convenient when evaluating the boundary bias of kernel estimators (see MüLLER, 1991; WAND; JONES, 1994; among others). Inspired by these ideas, we will give a definition for the mentioned set of indices J_x and exploit various right Riemann sum approximations.

Let $T_0 \in \mathbb{N}$ be such that $h < 1/2$ for any $T \geq T_0$. For every $T \geq T_0$, define the set

$$J_x = \{i \in [T] : i/T \in C_x\} \tag{2.33}$$

with

$$C_x = \begin{cases} [0, x+h] & , \text{ if } x \in [0, h] \\ [x-h, x+h] & , \text{ if } x \in (h, 1-h) \\ [x-h, 1] & , \text{ if } x \in [1-h, 1] \end{cases} . \tag{2.34}$$

In this study, whenever we require T to be sufficiently large such that J_x is well defined, we will be implicity assuming that T is large enough to achieve $h < 1/2$.

Lemma 2.2. *Let $T \geq T_0$ and let k_T be the cardinality of J_x. Then $k_T = O(Th)$. In addition, suppose that the Kernel function K is Lipschitz continuous on its compact support. Then, for any $x \in [0,1]$ and any sufficiently large T, it holds that ,*

$$\left| \frac{1}{T} \sum_{i=1}^{T} \left| K\left(\frac{i/T - x}{h}\right) \right| \left| \frac{i/T - x}{h} \right|^j - \int_0^1 \left| K\left(\frac{u-x}{h}\right) \right| \left| \frac{u-x}{h} \right|^j du \right| \leq \frac{C}{T}.$$

Proof. Suppose $x \in (h, 1-h)$. Then $J_x = \{i \in [T] : i/T \in [x-h, x+h]\}$. Note that the length of $(x-h, x+h)$ shrinks to zero slower than $1/T$, that is, $2h/(1/T) = 2Th \to \infty$. It implies that $\exists T_1 \geq T_0 : \forall T \geq T_1 : J_x \neq \emptyset$. Then, for $T \geq T_1$, define $i_* = \min J_x$ and $i^* = \max J_x$. Since the design points are evenly spaced, we can write the elements of $\{i/T\}_{i \in [T]} \cap (x-h, x+h)$ as

$$i_*/T + (k-1)/T, \quad k \in \{1, \ldots, M_T\}, \quad T \geq T_1,$$

where M_T is a sequence of natural numbers. In order to provide an upper bound for k_T, it is sufficient to find an upper bound for M_T. But we clearly need

$$\frac{i_*}{T} + \frac{(M_T - 1)}{T} < \frac{i^*}{T} + 2h$$

which implies that $M_T < CTh$. Hence $k_T = O(Th)$. Analogous arguments show the same results for $x \in [0, h]$ and $x \in [1 - h, 1]$

Next, note that

$$\int_{[0,1]} I(|(u - x)/h| \leq 1) du = \int_{[0,1]} I(x - h \leq u \leq x + h) du = \int_{[0,1] \cap [x-h, x+h]} du.$$

For $x \in [0, 1]$ and $T \geq T_0$, we evaluate the following cases. If $h < x$ and $x < 1 - h$, then $0 < x - h$ and $x + h < 1$, and so $[x - h, x + h] \cap [0, 1] = [x - h, x + h]$. If $x \leq h$, then $x - h \leq 0$ and $0 < x + h \leq 2h < 1$, which gives $[x - h, x + h] \cap [0, 1] = [0, x + h]$. If $1 - h \leq x$, then $1 \leq x + h$ and $0 < 1 - 2h \leq x - h < 1$, which gives $[x - h, x + h] \cap [0, 1] = [x - h, 1]$. Therefore

$$\int_{[0,1]} I(|(u - x)/h| \leq 1) du = \int_{C_x} du, \quad x \in [0, 1], \; T \geq T_0.$$

Furthermore, given any $x \in [0, 1]$, we must have $i_*/T \leq \underaccent{\bar}{C}_x + 1/T$ and $\bar{C}_x - 1/T \leq i^*/T$, where $\underaccent{\bar}{C}_x = \inf C_x$ and $\bar{C}_x = \sup C_x$. Otherwise, if $i_*/T - 1/T > \underaccent{\bar}{C}_x$ or $\bar{C}_x > i^*/T + 1/T$, then we would find a contradiction with the fact that both i_* and i^* are the minimum and the maximum of J_x. These imply that

$$0 \leq i_*/T - \underaccent{\bar}{C}_x \leq 1/T \text{ and } 0 \leq \bar{C}_x - i^*/T \leq 1/T,$$

which will be used in the following.

Define $J_x^* = J_x \setminus \{i_*\}$ and let $x \in [0, 1]$ be arbitrary. Using the above observations, the triangle inequality and the Mean Value Theorem for integrals, we have

$$\begin{aligned}
&\left| \frac{1}{T} \sum_{i=1}^{T} \left| K\left(\frac{i/T - x}{h}\right) \right| \left| \frac{i/T - x}{h} \right|^j - \int_0^1 \left| K\left(\frac{u - x}{h}\right) \right| \left| \frac{u - x}{h} \right|^j du \right| \\
&= \left| \frac{1}{T} \sum_{i \in J_x} \left| K\left(\frac{i/T - x}{h}\right) \right| \left| \frac{i/T - x}{h} \right|^j - \int_{C_x} \left| K\left(\frac{u - x}{h}\right) \right| \left| \frac{u - x}{h} \right|^j du \right| \\
&\leq \left| \frac{1}{T} \sum_{i \in J_x^*} \left| K\left(\frac{i/T - x}{h}\right) \right| \left| \frac{i/T - x}{h} \right|^j - \sum_{i \in J_x^*} \int_{(i-1)/T}^{i/T} \left| K\left(\frac{u - x}{h}\right) \right| \left| \frac{u - x}{h} \right|^j du \right| \\
&+ \frac{1}{T} \left| K\left(\frac{i_*/T - x}{h}\right) \right| \left| \frac{i_*/T - x}{h} \right|^j + \int_{\underaccent{\bar}{C}_x}^{i_*/T} \left| K\left(\frac{u - x}{h}\right) \right| \left| \frac{u - x}{h} \right| du \\
&+ \int_{i^*/T}^{\bar{C}_x} \left| K\left(\frac{u - x}{h}\right) \right| \left| \frac{u - x}{h} \right| du
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{T}\sum_{i\in J_x^*}\left|\left|K\Big(\frac{i/T-x}{h}\Big)\right|\left|\frac{i/T-x}{h}\right|^j-\left|K\Big(\frac{\xi_i-x}{h}\Big)\right|\left|\frac{\xi_i-x}{h}\right|^j\right| \\
&+\frac{C}{T}+C\Big(\frac{i_*}{T}-\underline{C}_x\Big)+C\Big(\bar{C}_x-\frac{i^*}{T}\Big) \\
&\leq \frac{1}{T}\sum_{i\in J_x^*}\left|K\Big(\frac{i/T-x}{h}\Big)\Big(\frac{i/T-x}{h}\Big)^j-K\Big(\frac{\xi_i-x}{h}\Big)\Big(\frac{\xi_i-x}{h}\Big)^j\right|+\frac{C}{T} \\
&\leq \frac{1}{T}\sum_{i\in J_x^*}\left\{\left|K\Big(\frac{i/T-x}{h}\Big)\right|\left|\Big(\frac{i/T-x}{h}\Big)^j-\Big(\frac{\xi_i-x}{h}\Big)^j\right|\right. \\
&\left.+\left|\frac{\xi_i-x}{h}\right|^j\left|K\Big(\frac{i/T-x}{h}\Big)-K\Big(\frac{\xi_i-x}{h}\Big)\right|\right\}+\frac{C}{T} \\
&\leq \frac{C}{T}\sum_{i\in J_x^*}\left\{\left|\frac{i/T-\xi_i}{h}\right|\sum_{l=0}^{j-1}\left|\frac{i/T-x}{h}\right|^l\left|\frac{\xi_i-x}{h}\right|^{j-1-l}+\left|\frac{i/T-\xi_i}{h}\right|\right\}+\frac{C}{T} \\
&\leq \frac{C}{T}k_T\left\{\frac{j}{Th}+\frac{1}{Th}\right\}+\frac{C}{T}\leq\frac{C}{T},
\end{aligned}
$$

with $\xi_i \in ((i-1)/T, i/T)$ for each $i \in J_x^*$. □

One can easily check that Lemma 2.2 holds for the function $K(u)u^j$, i.e., the function without taking the absolute value. Also, note that the assumptions of the lemma are weaker than A.2 once K is allowed to not be continuous everywhere.

Lemma 2.3. *Let K be a kernel function satisfying Assumption A.2 and let $\delta > 0$. Then there is a function K^* and constants $\bar{K}^*$ and μ^* such that $|K^*| \leq \bar{K}^* < \infty$, $\int_{\mathbb{R}}|K^*(u)|du \leq \mu^* < \infty$ and*

$$
|x_1 - x_2| \leq \delta \leq L \implies |K(x_1) - K(x_2)| \leq \delta K^*(x_1), \quad \forall x_1, x_2 \in \mathbb{R}. \tag{2.35}
$$

Particularly, if $K^*(x) = \Lambda_1 I(|x| \leq 2L)$, then

$$
\left|\frac{1}{T}\sum_{i=1}^{T}K^*\Big(\frac{i/T-x}{h}\Big)\Big(\frac{i/T-x}{h}\Big)^j-\int_0^1K^*\Big(\frac{u-x}{h}\Big)\Big(\frac{u-x}{h}\Big)^jdu\right|\leq\frac{C}{T},
$$

for any $x \in [0, 1]$ and T large enough.

Proof. Fix $\delta > 0$ and let $x_1, x_2 : |x_1 - x_2| \leq \delta \leq L$. Indeed, if K is Lipschitz, then $|K(x_1)-K(x_2)| \leq \Lambda_1|x_1-x_2| = \Lambda_1|x_1-x_2|\{I(|x_1| \leq 2L)+I(|x_1| > 2L)\}$. But $|x_1| > 2L$ implies $2L-|x_2| < |x_1|-|x_2| \leq |x_1-x_2| \leq L$. So $|x_2| > L$, and then $K(x_1)-K(x_2) = 0$ since K has compact support. Therefore the term $I(|x_1| > 2L)$ is superfluous for the upper bound. Hence, we can take $K^*(x) = \Lambda_1 I(|x| \leq 2L)$ which satisfies $|K(x_1) - K(x_2)| \leq \delta K^*(x_1)$, $|K^*| \leq \Lambda_1$ and $\int_{\mathbb{R}}|K^*(u)|du \leq \Lambda_1(4L)$.

Next, let T be large enough so that the set $J_x = \{i : i/T \in C_x\}$ with

$$C_x = \begin{cases} [0, x+h^*] & , \text{ if } x \in [0, h^*] \\ [x-h^*, x+h^*] & , \text{ if } x \in (h^*, 1-h^*) \\ [x-h^*, 1] & , \text{ if } x \in [1-h^*, 1] \end{cases}, \tag{2.36}$$

where $h^* = 2Lh$, is well-defined and nonempty. Note that the arguments of Lemma 2.2's proof can be applied to K^* even though it is not continuous everywhere. Then, along the same lines of the proof of Lemma 2.2, for any T large enough and any $x \in [0,1]$, we have

$$\begin{aligned} &\left| \frac{1}{T} \sum_{i=1}^{T} K^*\left(\frac{i/T - x}{h}\right)\left(\frac{i/T - x}{h}\right)^j - \int_0^1 K^*\left(\frac{u-x}{h}\right)\left(\frac{u-x}{h}\right)^j du \right| \\ &= \left| \frac{1}{T} \sum_{i \in J_x} \Lambda_1 \left(\frac{i/T - x}{h}\right)^j - \int_{C_x} \Lambda_1 \left(\frac{u-x}{h}\right)^j du \right| \\ &\le \frac{\Lambda_1}{T} \sum_{i \in J_x^*} \left| \left(\frac{i/T - x}{h}\right)^j - \left(\frac{\xi_i - x}{h}\right)^j \right| + \frac{C}{T} \le \frac{C}{T}, \end{aligned}$$

where $J_x^* = J_x \setminus \{i_*\}$ with $i_* = \min J_x$, and $\xi_i \in ((i-1)/T, i/T), \forall i \in J_x^*$. □

Lemma 2.4. *Let $T \in \mathbb{N}$ and $f : (\mathbb{R}, \mathcal{B}_\mathbb{R}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ be a measurable function. Define $\alpha_{1,T}(j)$ and $\alpha_{2,T}(j)$ as the mixing coefficients of the processes $\{Y_{t,T}\}$ and $\{f(Y_{t,T})\}$, respectively. Then $\alpha_{2,T}(j) \le \alpha_{1,T}(j)$, for all $0 \le j < T$.*

Proof. Fix $j : 0 \le j < T$. Denote $\mathcal{G}_{T,i}^k = \sigma((f(Y_{l,T})) : i \le l \le k)$ and $\mathcal{F}_{T,i}^k = \sigma((Y_{l,T}) : i \le l \le k)$ for $1 \le i \le k \le T$. If $\sigma(f(Y_{t,T})) \subseteq \sigma(Y_{t,T})$, for any $t \in \{1, \ldots, T\}$, then $\mathcal{G}_{T,i}^k \subseteq \mathcal{F}_{T,i}^k$ for any i, k, which in turn implies that $\alpha_{2,T}(j) \le \alpha_{1,T}(j)$. But, $\sigma(f(Y_{t,T})) = \{(Y_{t,T}^{-1} \circ f^{-1})(A) : A \in \mathcal{B}_\mathbb{R}\} \subseteq \{Y_{t,T}^{-1}(B) : B \in \mathcal{B}_\mathbb{R}\} = \sigma(Y_{t,T})$, $\forall t \in [T]$, and so the result. □

A direct consequence of Lemma 2.4 is that if $\{\epsilon_{t,T}\}$ is strongly mixing triangular array of random variables on $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, then so is $\{|\epsilon_{t,T}|\}$, since the function $|\cdot|$ is $(\mathcal{B}_\mathbb{R}, \mathcal{B}_\mathbb{R})$-measurable.

Now we restate the Proposition 1.12 of Tsybakov (2008).

Lemma 2.5 (Tsybakov)**.** *Let $x \in [0,1]$ such that $S_{T,x}$, defined in (2.16), is positive definite and let Q be a polynomial of degree at most 1. Then the local linear weights satisfy*

$$\sum_{t=1}^{T} Q(x_t) W_{t,T}(x) = Q(x),$$

for any sample $(x_1, \ldots, x_T)$. *In particular,*

$$\sum_{t=1}^{T} W_{t,T}(x) = 1 \text{ and } \sum_{t=1}^{T} (x_t - x) W_{t,T}(x) = 0. \tag{2.37}$$

Proof. By hypotesis $\partial^k Q(x_t)/\partial x_t^k = 0, \forall k \geq 2$, and then expanding $Q(x_t)$ around x gives

$$Q(x_t) = Q(x) + Q'(x)(x_t - x) := q'(x) \begin{bmatrix} 1 \\ (x_t - x)/h \end{bmatrix},$$

where $q(x) = (Q(x), Q'(x)h)'$. Since the local linear estimator is the solution of a weighted least squares, for $Z_t = Q(x_t)$ we have that

$$\begin{aligned}
\hat{\beta}_T(x) &= \arg\min_{\beta_x} (Z - X_x\beta_x)'W(Z - X_x\beta_x) = \arg\min_{\beta_x} (X_x q - X_x\beta_x)'W(X_x q - X_x\beta_x) \\
&= \arg\min_{\beta_x} (X_x(q - \beta_x))'W(X_x(q - \beta_x)) = \arg\min_{\beta_x} (q - \beta_x)'X_x'WX_x(q - \beta_x) \\
&= \arg\min_{\beta_x} (q - \beta_x)'S_{T,x}(q - \beta_x)
\end{aligned}$$

where $Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_T \end{bmatrix}$, $\quad X_x = \begin{bmatrix} 1 & (x_1 - x)/h \\ \vdots & \vdots \\ 1 & (x_T - x)/h \end{bmatrix}$, $\quad \beta_x = (g(x), g'(x)h)'$, $\quad q = q(x)$ and $W = \text{diag}(K((x_1 - x)/h), \cdots, K((x_T - x)/h))$. The necessary condition for $\hat{\beta}_T(x)$ is

$$\frac{\partial q'B_{T,x}q - 2q'B_{T,x}\beta_x + \beta_x'B_{T,x}\beta_x}{\partial \beta_x} = -2B_{T,x}'q + 2B_{T,x}\beta_x.$$

As $B_{T,x}$ is symmetric and positive definite, the unique solution is given by $\hat{\beta}_T(x) = q$. Then $\hat{g}(x) = e_1'\hat{\beta}_T(x) = Q(x)$. Hence $Q(x) = \sum_{T=1} Q(x_t)W_{t,T}(x)$ by (2.19). The results in (2.37) are immediate from the choices $Q(x_t) = 1$ and $Q(x_t) = x_t - x$. □

The following lemma is an extension of Proposition 1 of Fernández and Fernández (2001).

Lemma 2.6. *Under A.2, for any* $x \in [0, 1]$, *we have*

$$s_{T,j}(x) = \mu_j(x) + O(1/(Th)), \quad \forall j \in \{0, 1, 2, 3\}, \tag{2.38}$$

where $\mu_j(x) = \int_{G_x} u^j K(u)du$ *with*

$$G_x = \begin{cases} [-c, 1] & , \text{ if } x = ch \\ [-1, 1] & , \text{ if } x \in (h, 1-h) \\ [-1, c] & , \text{ if } x = 1 - ch \end{cases}$$

and $0 \leq c \leq 1$.

The proof of the above result follows directly from Lemma 2.2 and the definition of Big Oh, and thus is omitted. Lemma 2.6 implies that $S_{T,x} \to S_x$ as $T \to \infty$ where

$$S_x = \int_{G_x} \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix} K(u)du \tag{2.39}$$

Lemma 2.7. *Let K be nonnegative satisfying Assumption A.2. Suppose $\mu(\{K > 0\}) > 0$. Then the limiting matrix S_x in (2.39) is positive definite. Moreover,*

$$\exists \lambda_0, T_0 > 0 : \lambda_{min} \geq \lambda_0, \quad \forall T \geq T_0,\ \forall x \in [0,1],$$

where λ_{min} is the smallest eigenvalue of $S_{T,x}$.

Proof. Let $z \in \mathbb{R}^2$ be a nonzero vector. Since K is nonnegative, we have

$$z'S_x z = \int_{G_x} z'XX'zK d\mu \geq 0,$$

for $X := X(w) = (1, w)'$. To get a contradiction, suppose $\exists y \neq 0 : \int_{[-c,c]} y'XX'yKd\mu = 0$. Then $y'XX'y = 0$ μ-almost everywhere (a.e.) on $\{K > 0\} \cap G_x$ which has positive measure. However, $y'XX'y$ is a polynomial of degree at most 2 and cannot be equal to zero except on finitely many number of points. This means $y'XX'y \overset{a.e.}{\neq} 0$ on $\{K > 0\} \cap G_x$, a contradiction. Hence, we must have $z'S_x z > 0$.

To show the next result, note that $\det S_x, \operatorname{tr} S_x > 0$ as S_x is positive definite. Also, the trace and the determinant are continuous mappings. Since $S_{T,x} \to S_x$, the continuity implies $\operatorname{tr} S_{T,x} \to \operatorname{tr} S_x$ and $\det S_{T,x} \to \det S_x$. Therefore, there must be $T_0 : \forall T \geq T_0$ we have $\det S_{T,x} > 2^{-1} \det S_x > 0$ and $\operatorname{tr} S_{T,x} > 2^{-1} \operatorname{tr} S_x > 0$. Thus, the sum and the product of the two disctinct eigenvalues of $S_{T,x}$ are positive, implying a set of (strictly) positive eigenvalues, for all sufficiently large T. □

For any vector $y \in R^2$ and for an eigenpair $((\lambda_u, u), (\lambda_v, v))$ of $S_{T,x}$, it holds from Lemma 2.8 that there are $\lambda_0, c_1, c_2 > 0$ such that $S_{T,x}y = S_{T,x}(c_1 u + c_2 v) = c_1 \lambda_u u + c_2 \lambda_v v \geq \lambda_0 y$ when T is large enough. It implies $(1/\lambda_0)\|y\| \geq \|S_{T,x}^{-1} y\|$.

The following lemma is a restatement of Lemma 1.3 of Tsybakov (2008).

Lemma 2.8 (Tsybakov)**.** *Let Assumption A.2 hold, T_0 be as in Lemma 2.7 and $T^* \in \mathbb{N}$ is such that $\forall T \geq T^*, Th \geq 1/2$. Then for any $T \geq \max(T^*, T_0)$ and any $x \in [0,1]$, the weights of the local linear estimator defined in (2.19) satisfy*

(i) $\sup_{t,x} |W_{t,T}(x)| \leq \frac{C}{Th}$;

(ii) $\sum_{t=1}^{T} \sup_x |W_{t,T}(x)| \leq C$;

(iii) $W_{t,T}(x) = 0$ *if* $|\frac{X_t - x}{h}| \notin \operatorname{supp} K$.

for some constant $C > 0$.

Proof. (i) Denote $x_t = t/T$ for all $t \in \{1, \ldots, T\}$. By Lemma 2.7,

$$\begin{aligned}
|W_{t,T}(x)| = \|W_{t,T}(x)\| &= \left\| \frac{1}{Th} e_1' S_{T,x}^{-1} X\Big(\frac{x_t - x}{h}\Big) K\Big(\frac{x_t - x}{h}\Big) \right\| \\
&\leq \frac{1}{Th} \|e_1'\| \left\| S_{T,x}^{-1} X\Big(\frac{x_t - x}{h}\Big) \right\| \left| K\Big(\frac{x_t - x}{h}\Big) \right| \\
&\leq \frac{1}{Th} \frac{1}{\lambda_0} \left\| X\Big(\frac{x_t - x}{h}\Big) \right\| \left| K\Big(\frac{x_t - x}{h}\Big) \right| \\
&\leq \frac{1}{Th\lambda_0} \left\| X\Big(\frac{x_t - x}{h}\Big) \right\| \sup |K| I[(x_t - x)/h \in \operatorname{supp} K] \\
&\leq \frac{C}{Th} \left\| X\Big(\frac{x_t - x}{h}\Big) \right\| \leq \frac{C\sqrt{2}}{Th} \leq \frac{C}{Th}.
\end{aligned}$$

(ii) From the previous result, Lemma 2.7, it follows that

$$\sum_{t=1}^{T} \sup_x |W_{t,T}(x)| \leq \frac{C}{Th} \sum_{t=1}^{T} I\big[(x_t - x)/h \in \operatorname{supp} K\big] = \frac{C}{Th} \sum_{t \in J_x} 1 \leq C,$$

with J_x being as in Lemma 2.2, which has cardinality of order $O(Th)$.

(iii) From the proof of (i), we have $|W_{t,T}(x)| \leq \frac{C}{Th} I\big(|\frac{x_t - x}{h}| \in \operatorname{supp} K\big)$, and hence the result. $\square$

The next lemmas provide a list of results involving asymptotic notations.

Lemma 2.9. *Let a_t and b_t be positive sequences converging to zero. The following results hold:*

(i) If $C_1, C_2 \in \mathbb{R} : C_2 \neq 0$, then

$$\frac{C_1 + O(a_T)}{C_2 + O(b_T)} = \frac{C_1}{C_2} + O(a_T) + O(b_T);$$

In particular,

$$\frac{C_1}{C_2 + O(b_T)} = \frac{C_1}{C_2} + O(b_T);$$

(ii) If $Y_T = O_p(a_T)$ and $a_T = o(b_T)$, then $Y_T = o_p(b_t)$;

(iii) $O_p(a_T)O(b_T) = O_p(a_T b_T)$;

(iv) If $Y_T \leq X_T$ and $X_T = O_p(a_T)$, then $Y_T = O_p(a_T)$;

(v) If $c_T = o(b_T)$ and $X_T = O_p(a_T)$, then $c_T + X_T = O_p(a_T + b_T)$; if instead $c_T = O(b_T)$, then also $c_T + X_T = O_p(a_T + b_T)$.

Proof. (i) Denote $c_T = O(a_T)$ and $d_T = O(b_T)$. Then, using Taylor expansion,

$$\begin{aligned}
\frac{C_1 + c_T}{C_2 + d_T} &= \frac{C_1}{C_2} \frac{1}{1 + d_T/C_2} + \frac{c_T}{C_2} \frac{1}{1 + d_T/C_2} \\
&= \frac{C_1}{C_2} \Big\{ 1 - \frac{d_T}{C_2} + o(d_T) \Big\} + \frac{c_T}{C_2} \Big\{ 1 - \frac{d_T}{C_2} + o(d_T) \Big\}
\end{aligned}$$

$$= \frac{C_1}{C_2} + O(d_T) + O(c_T) + o(d_T) = \frac{C_1}{C_2} + O(a_T) + O(b_T).$$

The second result is obtained analogously by setting $c_T = 0$.

(*ii*) Let $\epsilon, \delta > 0$ be given. By the hypotheses, $\exists T_0, M : P(|Y_T| \geq M a_T) \leq \epsilon$ for all $T \geq T_0$. Further, $\exists T_1 : a_T \leq \delta b_T$ since $a_T = o(b_T)$, for all $T \geq T_1$. Take $\delta^* = M\delta$. Hence

$$P(|Y_T| \geq \delta^* b_T) \leq P(|X_T| \geq M a_T) \leq \epsilon,$$

for every $T \geq \max(T_0, T_1)$.

(*iii*) Let $X_t = O_p(a_T)$ and $c_T = O(b_T)$. Fix $\epsilon > 0$. Then $\exists T^*, M_1, C > 0 : \forall T \geq T^* :$ $P(|X_T| \geq M_1 a_T) \leq \epsilon$ and $|c_T/b_T| \leq C$. Take $M = M_1 C$. Then

$$\begin{aligned} P(|X_T c_T| \geq M a_T b_T) &= P(|X_T||c_T/b_T| \geq M a_T) \leq P(C|X_T| \geq M a_T) \\ &= P(|X_T| \geq M_1 a_T) \leq \epsilon. \end{aligned}$$

This shows that $X_T c_T = O_p(a_T b_T)$ as desired.

(*iv*) Clearly, $P(|Y_T| \geq M) \leq P(|X_T| \geq M)$ if $Y_T \leq X_T$, and this implies the result.

(*v*) Let $\epsilon > 0$ be fixed. By hypothesis, $\forall \delta > 0$, $\exists M_1 > 0 :$ $P(|X_T| \geq M_1 a_T) \leq \epsilon$ and $|c_T| \leq \delta b_T$, for sufficiently large T. Choose $M : M \geq \max(\delta, M_1)$. Then

$$\begin{aligned} P\big(|X_T + c_T| \geq M(a_T + b_T)\big) &\leq P\big(|X_T| \geq M(a_T + b_T) - |c_T|\big) \\ &\leq P\big(|X_T| \geq M(a_T + b_T) - \delta b_T\big) \\ &= P\big(|X_T| \geq M a_T + b_T(M - \delta)\big) \\ &\leq P\big(|X_T| \geq M a_T\big) \\ &\leq P\big(|X_T| \geq M_1 a_T\big) \leq \epsilon. \end{aligned}$$

The proof for $c_T = O(b_T)$ is analogous. □

The next lemma is Lemma 2.9's analogue for Big Oh and small oh almost surely.

Let $\{Y_n\}$ be a seqence of random variables on $(\Omega, \mathcal{F}, P)$. We say that $Y_n = O(1)$ almost surely, briefly $Y_n = O(1)$a.s., if $\exists M > 0$ such that $P(\limsup_{n\to\infty}\{|Y_n| \leq M\}) = 1$, and $Y_n = o(1)$ a.s. if $\forall \delta > 0$ we have $P(\limsup_{n\to\infty}\{|Y_n| > \delta\}) = 0$.

Lemma 2.10. *Let a_t and b_t be positive sequences converging to zero. The following results hold:*

(i) If $Y_T = O(a_T)$ a.s. and $a_T = o(b_T)$, then $Y_T = o(b_t)$ a.s.;

(ii) If $Y_T = O(a_T)$ a.s. and $c_T = O(b_T)$, then $Y_T c_T = O(a_T b_T)$ a.s.;

(iii) If $Y_T \leq X_T$ and $X_T = O(a_T)$ a.s., then $Y_T = O(a_T)$ a.s.;

(iv) If $c_T = O(b_T)$ and $X_T = O(a_T)$ a.s., then $c_T + X_T = O(a_T + b_T)$ a.s.;

(v) If $Y_T = O(1)$ a.s., then $a_T Y_T = O(a_T)$ a.s.; similarly, if $Y_T = o(1)$ a.s., then $a_T Y_T = o(a_T)$ a.s.;

(vi) If $Y_T = O(1)$ a.s. and $X_T = o(1)$ a.s., then $Y_T + X_T = O(1)$ a.s.

Proof. In what follows we will use the shorthand $\limsup_T$ for $\limsup_{T\to\infty}$.

(i) By hypothesis, $\exists M > 0 : P(\limsup_T\{|Y_T| \leq Ma_T\}) = 1$ and $a_T \leq \delta b_T$ for all $\delta > 0$ and all T sufficiently large. Let $\delta/M > 0$ be given. Then, for every T sufficiently large,

$$\{|Y_T| \leq Ma_T\} \subseteq \{|Y_T| \leq \delta b_T\}$$

Claim 1. Let A_T and B_T be two sequence of sets. Suppose that, for all sufficiently large T, $A_T \subseteq B_T$. Then $\limsup_T A_T \subseteq \limsup_T B_T$.

Proof of claim: By definition, $\limsup_T A_T = \bigcap_{T=1}^{\infty}\bigcup_{k=T}^{\infty} A_k := \bigcap_{T=1}^{\infty} C_T$, where $C_T = \bigcup_{k=T}^{\infty} A_k$ is a decreasing sequence. Similarly, we can write $\limsup_T B_T := \bigcap_{T=1}^{\infty} D_T$, with $D_T = \bigcup_{k=T}^{\infty} B_k$. By hypothesis, there is some T_0 such that, for any $T > T_0$, we have $C_T \subseteq D_T$, which implies $\bigcap_{T>T_0} C_T \subseteq \bigcap_{T>T_0} D_T$. Since the sets C_T and D_T are decreasing,

$$\bigcap_{T}^{\infty} C_T = \bigcap_{T>T_0}^{\infty} C_T \subseteq \bigcap_{T>T_0}^{\infty} D_T = \bigcap_{T}^{\infty} D_T,$$

and hence the result. ■

By Claim 1 and using the monotonicity of the measure,

$$1 = P(\limsup_T\{|Y_T| \leq Ma_T\}) \leq P(\limsup_T\{|Y_T| \leq \delta b_T\}),$$

which implies that $P(\limsup_T\{|Y_T| \leq \delta b_T\}) = 1$. As δ is arbitrary, the result follows.

(ii) By hypothesis, $\exists M > 0 : P(\limsup_T\{|Y_T| \leq Ma_T\}) = 1$ and $|b_T/c_T| \geq 1/C$ for some constant $C > 0$ and all T sufficiently large. Take $M_1 = MC$. Then, for all T large enough,

$$\{|Y_T c_T| \leq M_1 a_T b_T\} = \{|Y_T| \leq M_1 a_T |b_T/c_T|\} \supseteq \{|Y_T| \leq Ma_T\}$$

From Claim 1 and the monotonicity of P,

$$P(\limsup_T\{|Y_T c_T| \leq M_1 a_T b_T\}) \geq P(\limsup_T\{|Y_T| \leq Ma_T\}) = 1$$

and thus the result.

(iii) By hypothesis and using Claim 1, there is $M > 0$ satisfying

$$P(\limsup_T\{|Y_T| \leq Ma_T\}) \geq P(\limsup_T\{|X_T| \leq Ma_T\}) = 1,$$

implying the result.

(iv) By hypothesis, $\exists M > 0 : P(\limsup_T\{|X_T| \leq Ma_T\}) = 1$ and $|c_T| \leq Cb_T$ for some constant $C > 0$ and all T sufficiently large. Choose $M_1 = \max(M, C)$. For this choice and all sufficiently large T,

$$\begin{aligned}\{|X_T + c_T| \le M_1(a_T + b_T)\} &\supseteq \{|X_T| \le M_1(a_T + b_T) - |c_T|\}\\ &\supseteq \{|X_T| \le M_1 a_T + b_T(M_1 - C)\}\\ &\supseteq \{|X_T| \le M_1 a_T + b_T(M_1 - M_1)\}\\ &\supseteq \{|X_T| \le M a_T\}\end{aligned}$$

Hence,

$$P(\limsup_T \{|X_T + c_T| \le M_1(a_T + b_T)\}) \ge P(\limsup_T \{|X_T| \le M a_T\}) = 1,$$

which gives the result.

(v) By hypothesis we clearly have, for some $M > 0$,

$$P(\limsup_T \{|Y_T a_T| \le M a_T\}) = P(\limsup_T \{|Y_T| \le M\}) = 1.$$

The proof for the small oh goes in the same lines.

(vi) Given any $c > 0$, note that

$$w \in \limsup_T \{|Y_T| \le c\} \iff |Y_T(w)| \le c \text{ for infinitely many } T$$

and

$$\begin{aligned}w \in \limsup_T \{|X_T| > c\} &\iff |X_T(w)| > c \text{ for infinitely many } T\\ &\iff |X_T(w)| \le c \text{ for all but finitely many } T.\end{aligned}$$

By hypothesis, for all $\delta > 0$ and for some $M > 0$, we have

$$\begin{aligned}&|Y_T(w)| \le M \text{ for infinitely many } T, \qquad \text{and}\\ &|X_T(w)| \le \delta \text{ for all but finitely many } T,\end{aligned}$$

with probability one. Then, with probability one, the triangle inequality gives

$$|X_T(w) + Y_T(w)| \le M + \delta \text{ for infinitely many } T,$$

and hence the result $X_T + Y_T = O(1)$a.s.

□

Lemma 2.11. *Let X and Y be two random variables and let $b \in \mathbb{R}$. Then*

$$P(|X + Y| > b) \le P(|X| > b/2) + P(|Y| > b/2).$$

Proof. Let $A = \{(x, y) : |x + y| \leq b\}$ and $B = \{(x, y) : |x| \leq b/2, |y| \leq b/2\}$. Note that A lies in the square of side b centered at the origin. Then $A \supseteq B$, which in turn implies that $\{(X, Y) \in A\} \supseteq \{(X, Y) \in B\}$. Using DeMorgan's Law, it follows that

$$\{(X, Y) \in A\}^c = \{|X + Y| > b\} \subseteq \{|X| > b/2\} \cup \{|Y| > b/2\} = \{(X, Y) \in B\}^c.$$

From the monotonicity and subadditivity of the measure,

$$P(|X + Y| > b) \leq P(\{|X| > b/2\} \cup \{|Y| > b/2\}) \leq P(|X| > b/2) + P(|Y| > b/2).$$

$\square$

Appendix B - The Davydov's inequality

The Davydov's inequality is a covariance inequality which will be extensively used in this study. Because it is our basic tool, we will review how it can be proved based on Bosq (2012) and Rio (2017). A good understanding of the results below can give us insights on how to bound covariances when we are faced with more complicated situations.

Define the *indicator function* of a subset $A \subseteq \mathbb{R}$ as

$$\chi_A(x) = \begin{cases} 1 & , \text{ if } x \in A \\ 0 & , \text{ if } x \notin A \end{cases}.$$

The following identity will be shown to be useful when dealing with covariances.

Lemma 2.12. *For any $a, b \in \mathbb{R}$, we have that $b - a = \int_{-\infty}^{\infty} \chi_{(-\infty,x]}(a) - \chi_{(-\infty,x]}(b)dx$.*

Proof. Clearly, $\chi_{(-\infty,x]}(a) - \chi_{(-\infty,x]}(b)$ is nonzero if, and only if, $a \leq x < b$ or $b \leq x < a$. Furthermore,

$$a \leq x < b \implies \int_{-\infty}^{\infty} \chi_{(-\infty,x]}(a) - \chi_{(-\infty,x]}(b)dx = \int_a^b 1dx = b - a$$

and

$$\begin{aligned} b \leq x < a \implies \int_{-\infty}^{\infty} \chi_{(-\infty,x]}(a) - \chi_{(-\infty,x]}(b)dx &= \int_b^a -1dx \\ &= \int_a^b 1dx = b - a. \end{aligned}$$

Hence, regardless the case, the desired equality holds. □

Given a measurable space $(\Omega, \mathcal{A})$, the above lemma shows that if $Z_1, Z_2 : \Omega \to \mathbb{R}$ are random variables, then $Z_2(w) - Z_1(w) = \int \chi_{(-\infty,x]}(Z_1(w)) - \chi_{(-\infty,x]}(Z_2(w))dx, \ \forall w \in \Omega$.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X, Y : \Omega \to \mathbb{R}$ be random variables. Define the *joint distribution function* as $F_{X,Y}(x, y) = P_{X,Y}((-\infty, x] \times (-\infty, y]) = P\{X(w) \leq x, Y(w) \leq y\}$, where $P_{X,Y} : \mathcal{B}_{\mathbb{R}^2} \to [0, 1]$ is the *joint probability distribution* (or the push-forward measure) of X and Y. Given the joint distribution function $F_{X,Y}$, the *marginal distribution function* of X is defined as $F_X(x) = P_{X,Y}((-\infty, x] \times \mathbb{R})$. We assume the notation $\{X(w) \in B\} = X^{-1}(B)$.

Lemma 2.13 (Hoeffding's Lemma)**.** *Let F_X and F_Y be the marginal distribution functions of X and Y, respectively, given their joint distribution function $F_{X,Y}$. Then*

$$Cov(XY) = E(XY) - E(X)E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{XY}(x, y) - F_X(x)F_Y(y)dxdy, \quad (2.40)$$

provided the expectations $E|XY|, E|X|$ and $E|Y|$ are finite.

Proof. Firstly, we need to show a few results. Let $(X, Y), (X_2, Y_2)$ be independent and identically distributed according to $F_{X,Y}$.

Claim 2. (i) $Cov(X, Y) = Cov(X_2, Y_2)$;

(ii) $EX = EX_2$;

(iii) $X \perp Y_2$ and $X_2 \perp Y$, where $\perp$ denotes the independence of random variables;

(iv) $Cov(\chi_{(-\infty,x]}(X), \chi_{(-\infty,x]}(Y)) = Cov(\chi_{(-\infty,x]}(X_2), \chi_{(-\infty,x]}(Y_2)), \forall x \in \mathbb{R}$;

(v) $E\chi_{(-\infty,x]}(X) = E\chi_{(-\infty,x]}(X_2), \forall x \in \mathbb{R}$;

(vi) $\chi_{(-\infty,x]}(X) \perp \chi_{(-\infty,x]}(Y_2)$ and $\chi_{(-\infty,x]}(X_2) \perp \chi_{(-\infty,x]}(Y), \forall x \in \mathbb{R}$;

(vii) $E[(\chi_{(-\infty,x]} \circ X)(\chi_{(-\infty,y]} \circ Y)] = P(\{X \le x, Y \le y\})$ and $E[(\chi_{(-\infty,x]} \circ X)] = P(\{X \le x\})$.

Proof of claim: (i) The first result is obvious. (ii) Since the probability distribution $P_{X,Y}$ is uniquely determined by the distribution function $F_{X,Y}$, it follows that $F_Y(y) = P_{X,Y}(\mathbb{R} \times (-\infty, y]) = P_{X_2Y_2}(\mathbb{R} \times (-\infty, y]) = F_{Y_2}(y)$, which in turn, implies that $P_Y = P_{Y_2}$. Hence $E(Y) = \int x P_Y(dx) = \int x P_{Y_2}(dx) = E(Y_2)$. (iii) To see the independence, $F_{X,Y_2}(x, y_2) = \lim_{y,x_2 \to \infty} F_{X,Y,X_2,Y_2}(x, y, x_2, y_2) = \lim_{x_2 \to \infty} F_{X_2,Y_2}(x_2, y_2) \lim_{y \to \infty} F_{X,Y}(x, y) = F_X(x) F_{Y_2}(y_2)$.

(vi) Since X is independent of Y_2, by definition, $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}_{\mathbb{R}}\}$ and $\sigma(Y_2)$ are independent, meaning that $P(A \cap B) = P(A)P(B)$, $\forall A \in \sigma(Y_2), B \in \sigma(X)$. It is well known that $\sigma(X), \sigma(Y_2)$ are sub-σ-algebras of $\mathcal{A}$. Given any $x, y \in \mathbb{R}$, let $f = \chi_{(-\infty,x]}$ and $g = \chi_{(-\infty,y]}$ be two $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}) - (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ measurable functions. Then $(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A)) \in \sigma(X), \forall A \in \mathcal{B}_{\mathbb{R}}$, since $f^{-1}(A) \in \mathcal{B}_{\mathbb{R}}$. The same holds for $g \circ Y_2$. It implies that $\sigma(f \circ X) = \{(f \circ X)^{-1}(A) : A \in \mathcal{B}_{\mathbb{R}}\} \subseteq \sigma(X)$ and $\sigma(g \circ Y_2) \subseteq \sigma(Y_2)$. As $\sigma(Y_2)$ and $\sigma(X)$ are independent, so are $\sigma(f \circ X)$ and $\sigma(g \circ Y_2)$. Therefore the measurable indicator functions preserve the independence of the random variables. (iv) Furthermore, $F_{f \circ X, g \circ Y}(x_1, y_1) = P\{f(X) \le x_1, g(Y) \le y_1\} = P\{X \in f^{-1}(-\infty, x_1], Y \in g^{-1}(-\infty, y_1]\} = P_{X,Y}(f^{-1}(-\infty, x_1] \times g^{-1}(-\infty, y_1]) = P_{X_2Y_2}(f^{-1}(-\infty, x_1] \times g^{-1}(-\infty, y_1]) = F_{f \circ X_2, g \circ Y_2}(x_1, y_1)$. This immediately implies $Cov(f \circ X_2, g \circ Y_2) = Cov(f \circ X, g \circ Y)$. (v) By assumption, it is clear that the marginal probability distributions must be the same ($P_X = P_{X_2}$). Therefore, $E(f \circ X) = \int_\Omega (f \circ X)(z) P(dz) = \int_{\mathbb{R}} f(w) P_X(dw) = \int_{\mathbb{R}} f(w) P_{X_2}(dw) = E(f \circ X_2)$, since the indicator function is a nonnegative measurable function. (vii) Finally,

$$\int_\Omega (\chi_{(-\infty,x]} \circ X)(w) P(dw) = \int_{\mathbb{R}} \chi_{(-\infty,x]}(w') P_X(dw') = P_X((-\infty, x]) = P(\{X \le x\})$$

and

$$\begin{aligned} \int_\Omega \chi_{(-\infty,x]\times(-\infty,y]}(X(w), Y(w)) P(dw) &= \int_{\mathbb{R}^2} \chi_{(-\infty,x]\times(-\infty,y]}(w') P_{X,Y}(dw') \\ &= P(\{X \le x, Y \le y\}). \end{aligned}$$

■

By Claim 2, Lemma 2.12 and the Fubini-Tonelli's theorem, it follows that

$$\begin{aligned}
2\,\mathrm{Cov}(X,Y) &= \mathrm{Cov}(X,Y) + \mathrm{Cov}(X_2,Y_2) \\
&= E(X,Y) + E(X_2,Y_2) - E(X)E(Y) - E(X_2)E(Y_2) \\
&= E(X,Y + X_2,Y_2) - E(X_2Y) - E(XY_2) \\
&= E((X_2 - X)(Y_2 - Y)) \\
&= \int_\Omega \int \int \Big[\chi_{(-\infty,x]}(X) - \chi_{(-\infty,x]}(X_2)\Big]\Big[\chi_{(-\infty,y]}(Y) - \chi_{(-\infty,y]}(Y_2)\Big] dxdydP \\
&= \int \int \int_\Omega \Big[\chi_{(-\infty,x]}(X) - \chi_{(-\infty,x]}(X_2)\Big]\Big[\chi_{(-\infty,y]}(Y) - \chi_{(-\infty,y]}(Y_2)\Big] dPdxdy \\
&= 2\int\int Cov(\chi_{(-\infty,x]}(X), \chi_{(-\infty,x]}(Y))dxdy \\
&= 2\int\int E\big[\chi_{(-\infty,x]}(X)\chi_{(-\infty,x]}(Y)\big] - E\big[\chi_{(-\infty,x]}(X)\big]E\big[\chi_{(-\infty,x]}(Y)\big]dxdy \\
&= 2\int\int F_{X,Y}(x,y) - F_X(x)F_Y(y)dxdy
\end{aligned}$$

since $E|X_2 - X||Y_2 - Y| \leq 2(E|XY| + E|X|E|Y|) < \infty$. □

Lemma 2.14. *Let F be the distribution function of random variable X and let $F^{-1} : [0,1] \to \overline{\mathbb{R}}$ be the generalized inverse distribution function defined by $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$. Moreover, define the quantile function of X by $Q(z) = \inf\{x \in \mathbb{R} : P(X > x) \leq z\}, z \in \mathbb{R}$. Then, for any $x \in \mathbb{R}$ and any $z \in (0,1)$*

$$z < P(X > x) \iff x < Q(z). \tag{2.41}$$

Proof. Let $x \in \mathbb{R}$ and $z \in (0,1)$. Then $x \in \{y : F(y) \geq F(x)\}$ and $F^{-1}(F(x)) = \inf\{y : F(y) \geq F(x)\}$, by definition. Thus $F^{-1}(F(x)) \leq x$, or equivalently, $Q(1 - F(x)) \leq x$, since $Q(1 - z) = \inf\{x : 1 - F(x) \leq 1 - z\} = F^{-1}(z)$. Also, $F(F^{-1}(z)) = F(\inf\{y : F(y) \geq z\}) \geq z$. It is clear that Q is nonincreasing since $z_1 \leq z_2$ implies $\{P(X > x) \leq z_1\} \subseteq \{P(X > x) \leq z_2\}$.

Supose $z \geq P(X > x) = 1 - F(x)$. Then $Q(z) \leq Q(1 - F(x)) \leq x$. Conversely, if $x \geq Q(z) = F^{-1}(1-z)$, then $F(x) \geq F(F^{-1}(1-z)) \geq 1 - z \iff z \geq 1 - F(x) = P(X > x)$. The result follows by contraposition. □

The next theorem can be found in Bosq (2012, Theorem 1.1).

Theorem 2.5 (Rio's Inequality)**.** *Let X and Y be two integrable random variables and let $Q_{|X|}, Q_{|Y|}$ be the quantile functions of $|X|, |Y|$, respectively. Then if $Q_{|X|}Q_{|Y|}$ is integrable over $(0,1)$,*

$$|\mathrm{Cov}(X,Y)| \leq 2\int_0^{2\alpha} Q_{|X|}(u)Q_{|Y|}(u)du \tag{2.42}$$

where $\alpha = \alpha(\sigma(X), \sigma(Y)) = \sup_{B \in \sigma(X), C \in \sigma(Y)} |P(B \cap C) - P(B)P(C)|$.

Proof. Let $X = X^+ - X^-$ and $Y = Y^+ - Y^-$. From the bilinearity of the covariance,

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(X^+, Y^+) + \mathrm{Cov}(X^-, Y^-) - \mathrm{Cov}(X^+, Y^-) - \mathrm{Cov}(X^-, Y^+) \\
&\leq \mathrm{Cov}(X^+, Y^+) + \mathrm{Cov}(X^-, Y^-) + \mathrm{Cov}(X^+, Y^-) + \mathrm{Cov}(X^-, Y^+) \\
&= \mathrm{Cov}(|X|, |Y|).
\end{aligned}$$

By the Hoeffding's Lemma, $\mathrm{Cov}(X^+, Y^+) = \int_0^\infty \int_0^\infty P(X \leq u, Y \leq v) - P(X \leq u)P(Y \leq v)dudv$. Note that, if $A_1 = \{X \leq u\}$ and $A_2 = \{Y \leq v\}$, then $P(A_1 \cap A_2) - P(A_1)P(A_2) = 1 - P(A_1^c \cup A_2^c) - [(1 - P(A_1^c))(1 - P(A_2^c))] = P(A_1^c \cap A_2^c) - P(A_1^c)P(A_2^c)$. Hence $\mathrm{Cov}(X^+, Y^+) = \int_0^\infty \int_0^\infty P(X > u, Y > v) - P(X > u)P(Y > v)dudv$. Apply the same argument to the other covariance's terms to obtain the following set of equalities

$$\begin{aligned}
\mathrm{Cov}(X^+, Y^+) &= \int_0^\infty \int_0^\infty P(X > u, Y > v) - P(X > u)P(Y > v)dudv \\
\mathrm{Cov}(X^-, Y^-) &= \int_0^\infty \int_0^\infty P(-X > u, -Y > v) - P(-X > u)P(-Y > v)dudv \\
\mathrm{Cov}(X^-, Y^+) &= \int_0^\infty \int_0^\infty P(-X > u, Y > v) - P(-X > u)P(Y > v)dudv \\
\mathrm{Cov}(X^+, Y^-) &= \int_0^\infty \int_0^\infty P(X > u, -Y > v) - P(X > u)P(-Y > v)dudv.
\end{aligned}$$

Put $a = P(X > u), b = P(-X > u), c = P(Y > v)$ and $d = P(-Y > v)$. Note that the integrand of any of the above equations are bounded by $\alpha \geq 0$ as well as by, at least, two elements of $\{a, b, c, d\}$, due to the monotonicity of the measure. Then

$$\begin{aligned}
|\mathrm{Cov}(X, Y)| &\leq |\mathrm{Cov}(|X|, |Y|)| \\
&\leq |\mathrm{Cov}(X^+, Y^+)| + |\mathrm{Cov}(X^-, Y^-)| + |\mathrm{Cov}(X^+, Y^-)| + |\mathrm{Cov}(X^-, Y^+)| \\
&= \int_0^\infty \int_0^\infty \big[\inf\{\alpha, a, c\} + \inf\{\alpha, a, d\} + \inf\{\alpha, b, c\} + \inf\{\alpha, b, d\}\big] dudv \\
&= \int_0^\infty \int_0^\infty \big[\inf\{2\alpha, 2a, c + d\} + \inf\{2\alpha, 2b, c + d\}\big] dudv \\
&= \int_0^\infty \int_0^\infty \inf\{4\alpha, 2(a + b), 2(c + d)\} dudv \\
&= 2 \int_0^\infty \int_0^\infty \inf\{2\alpha, P(|X| > u), P(|Y| > v)\} dudv, \qquad (2.43)
\end{aligned}$$

where the last equality follows from

$$\begin{aligned}
a + b &= P(X > u) + P(-X > u) = P(\{X > u\} \cup \{X < -u\}) + P(\{X > u\} \cap \{X < -u\}) \\
&= P(\{X > u\} \cup \{X < -u\}) + P(\emptyset) \\
&= P(|X| > u),
\end{aligned}$$

and, similarly, from $c+d = P(|Y| > v)$. Define $e = P(|X| > u)$ and $f = P(|Y| > v)$, and note that

$$\int_0^\alpha \chi_{(-\infty,\inf(e,f)]}(z)dz = \begin{cases} \alpha & , \text{ if } \alpha \leq \inf\{e,f\} \\ \inf\{e,f\} & , \text{ if } \alpha > \inf\{e,f\} \end{cases},$$

and that $z \in (-\infty, \inf(e,f)] \iff z \in (-\infty, e] \cap (-\infty, f]$. Then, by Lemma 2.14,

$$\inf(2\alpha, e, f) = \int_0^{2\alpha} \chi_{(-\infty,e]}(z)\chi_{(-\infty,f]}(z)dz = \int_0^{2\alpha} \chi_{(-\infty,Q_{|X|}(z)]}(u)\chi_{(-\infty,Q_{|Y|}(z)]}(v)dz,$$

since it holds that $0 \leq \alpha \leq 1/4$ (see Bradley, 2005). From Fubini-Tonelli's theorem and (2.43), we have that

$$\begin{aligned} |\text{Cov}(X,Y)| &\leq 2\int_0^\infty \int_0^\infty \left[\int_0^{2\alpha} \chi_{(-\infty,Q_{|X|}(z)]}(u)\chi_{(-\infty,Q_{|Y|}(z)]}(v)dz\right]dudv \\ &\leq 2\int_0^{2\alpha}\left[\int_0^{Q_{|X|}(z)} 1du \int_0^{Q_{|Y|}(z)} 1dv\right]dz \\ &= 2\int_0^{2\alpha} Q_{|X|}(z)Q_{|Y|}(z)dz. \end{aligned}$$

□

Corollary 2.5.1 (Davydov's Inequality). *Let X and Y be two random variables such that $X \in L^q(P), Y \in L^r(P)$ where $q > 1, r > 1$ are finite and $1/q + 1/r = 1 - 1/p$. Then*

$$|\text{Cov}(X,Y)| \leq 2p(2\alpha)^{1/p}\|X\|_q\|Y\|_r. \tag{2.44}$$

Proof. Let $X \in L^p(P), Y \in L^p(P)$, meaning that $\|X\|_q = (\int|X|^q dP)^{1/q} < \infty$ and that $\|Y\|_r = (\int|Y|^r dP)^{1/r} < \infty$, respectively. By the Markov's inequality, we have that

$$\begin{aligned} P\left[|X| > \frac{\|X\|_q}{u^{1/q}}\right] &= P\left[|X|^q > \left(\frac{\|X\|_q}{u^{1/q}}\right)^q\right] \leq P\left[|X|^q \geq \left(\frac{\|X\|_q}{u^{1/q}}\right)^q\right] \\ &\leq \frac{u}{\|X\|_q^q}\int_\Omega |X|^q dP = \frac{u}{\|X\|_q^q}\|X\|_q^q \\ &= u, \ \forall u \in (0,1). \end{aligned} \tag{2.45}$$

The inequality (2.45) is equivalent to $Q_{|X|}(u) \leq \|X\|_q/u^{1/q}, \forall u \in (0,1)$, by the contraposition of Lemma 2.14. These results hold analogously for Y. From Rio's inequality,

$$\begin{aligned} |\text{Cov}(X,Y)| &\leq 2\int_0^{2\alpha} Q_{|X|}(u)Q_{|Y|}(u)du \leq 2\int_0^{2\alpha} \frac{\|X\|_q\|Y\|_r}{u^{1/q}u^{1/r}}du \\ &= 2\|X\|_q\|Y\|_r\int_0^{2\alpha} u^{1/p-1}du = 2\|X\|_q\|Y\|_r(2\alpha)^{1/p}p. \end{aligned}$$

□

Assumption A.1 imposes that $\{\epsilon_{t,T}\}$ is strongly mixing on $(\Omega, \mathcal{F}, P)$. Remember that the α-mixing coefficients are defined as

$$\alpha_T(j) = \sup_{1 \le k \le T-j} \sup\{|P(A \cap B) - P(A)P(B)| : B \in \mathcal{F}_{T,1}^k, A \in \mathcal{F}_{T,k+j}^T\}, \quad 0 \le j < T,$$

where $\mathcal{F}_{T,i}^k = \sigma(\epsilon_{T,l} : i \le l \le k)$. Let $f(A, B) = |P(A \cap B) - P(A)P(B)|$ for any $A, B \in \mathcal{F}$. It holds that

$$\begin{aligned}
\alpha(\sigma(\epsilon_{t,T}), \sigma(\epsilon_{l,T})) &\stackrel{\text{def}}{=} \sup\{f(A, B) : A \in \sigma(\epsilon_{t,T}), B \in \sigma(\epsilon_{l,T})\} \\
&\in \{\sup\{f(A, B) : A \in \sigma(\epsilon_{j,T}), B \in \sigma(\epsilon_{j+|t-l|,T})\} : 0 \le j < T\} \\
&\subseteq \{\sup\{f(A, B) : A \in \sigma(\cup_{i=1}^j \sigma(\epsilon_{i,T})), B \in \sigma(\cup_{i=j+|t-l|}^\infty \sigma(\epsilon_{i,T}))\} : 0 \le j < T\} \\
&= \{\sup\{f(A, B) : A \in \mathcal{F}_1^j, B \in \mathcal{F}_{j+|t-l|}^\infty\} : 0 \le j < T\}.
\end{aligned}$$

Taking the supremum over j yields $\alpha(\sigma(\epsilon_t), \sigma(\epsilon_l)) \le \alpha(|l-t|)$. We shall use this fact when applying Davydov's inequality.

If X and Y are essentially bounded random variables ($X, Y \in L^\infty(P)$), where we define $\|Z\|_\infty = \inf\{a : P(Z > a) = 0\} < +\infty$, $\forall Z \in L^\infty(P)$, then Rio's inequality implies

$$|\text{Cov}(X, Y)| \le 2Q_{|X|}(0)Q_{|(Y)|}(0) \int_0^{2\alpha} du = 4\alpha \|X\|_\infty \|Y\|_\infty.$$

This result is also known as Billingsley's inequality. From Corollary 2.5.1, we immediately see that

$$|\text{Cov}(X, Y)| \le 4\alpha^{1-1/q} \|X\|_q \|Y\|_\infty,$$

if $X \in L^q(P)$ and $Y \in L^\infty(P)$. It is then possible to derive another version of Davydov's inequality.

Corollary 2.5.2 (Davydov's Inequality 2)**.** *Let X and Y be two random variables such that $X \in L^q(P), Y \in L^r(P)$ where $q > 1, r > 1$ are finite and $1/q + 1/r = 1 - 1/p$. Then*

$$|\text{Cov}(X, Y)| \le 6\alpha^{1/p} \|X\|_q \|Y\|_r. \tag{2.46}$$

Proof. Put $M = \alpha^{-1/r} \|Y\|_r$, $Y_1 = Y\chi_{\{|Y| \le M\}}$ and $Y_2 = Y - Y_1$. Then $Y = Y_1 + Y_2$ and $|Y_1| \le M$. Therefore, applying Corollary 2.5.1 and Holder's inequality,

$$\begin{aligned}
|\text{Cov}(X, Y)| = |\text{Cov}(X, Y_1 + Y_2)| &\le |\text{Cov}(X, Y_1)| + |\text{Cov}(X, Y_2)| \\
&\le 4\alpha^{1-1/q} \|X\|_q \|Y_1\|_\infty + 2\|X\|_q \|Y_2\|_{q/(q-1)} \\
&\le 2\|X\|_q (2M\alpha^{1-1/q} + \|Y_2\|_{q/(q-1)}).
\end{aligned}$$

Let $s = q/(q-1)$ for simplicity. By Holder's and Markov's inequalities, it follows that

$$\begin{aligned} E(|Y|^s \chi_{\{|Y|>M\}}) &\le [E|Y|^r]^{s/r} (P(|Y| > M))^{1-s/r} \le [E|Y|^r]^{s/r} [E(|Y|^r/M^r)]^{1-s/r} \\ &= E|Y|^r M^{s-r}, \end{aligned}$$

and then

$$\begin{aligned} \|Y_2\|_s &= \left\{E\left|Y(1-\chi_{\{|Y|\le M\}})\right|^s\right\}^{1/s} = \left\{E\left(|Y|^s \chi_{\{|Y|>M\}}\right)\right\}^{1/s} = \left\{E|Y|^r M^{s-r}\right\}^{1/s} \\ &= \{E|Y|^r (\alpha^{-1} E|Y|^r)^{(s-r)/r}\}^{1/s} = (E|Y|^r)^{\frac{1}{r}\left(1-\frac{r}{s}\right)+\frac{1}{s}} \alpha^{-\frac{1}{r}\left(1-\frac{r}{s}\right)} \\ &= (E|Y|^r)^{1/r} \alpha^{1/p}. \end{aligned}$$

From this, $|\mathrm{Cov}(X,Y)| \le 2\|X\|_q (2\alpha^{1/p}\|Y\|_r + \|Y\|_r \alpha^{1/p}) = 6\alpha^{1/p}\|X\|_q \|Y\|_r$. $\square$

3 CONCLUDING REMARKS

The first essay of this thesis develops uniform consistency results for the local linear estimator under mixing conditions in order to be directly applied in the next essays. The weak and strong uniform convergence rates were provided for general kernel averages from which we obtained the uniform rates for the local linear estimator. We restricted our attention to equally-spaced design points of the form $x_{t,T} = t/T,\ t \in \{1, \ldots, T\},\ T \in \mathbb{N}$. The convergences were stablished uniformly over $[0, 1]$ under arithmetically strong mixing conditions. The kernel function was restricted to be compactly supported and Lipschitz continuous, and inlcudes the popular Epanechnikov kernel. The uniform convergence in probability was provided without imposing stationarity while the almost sure uniform convergence was proved only for the stationary case.

The second essay is the main study of this thesis. We investigated the asymptotic properties of the estimators obtained by reversing the three-step procedure of Vogt and Linton (2014), for time series modelled as the sum of a periodic and a trend deterministic components plus a stochastic error process. In the first step, the trend function is estimated; given the trend estimate, an estimate of the period is provided in the second step; the last step consists in estimating the periodic sequence. The weak uniform convergence rates of the estimators of the trend function and the periodic sequence were provided. The asymptotic normality for the trend estimator was also stablished. Furthermore, it was shown that the period estimator is consistent.

The third essay exploits the bandwidth selection problem and the finite sample performance of the period estimator studied in the second essay. A plug-in type bandwidth is proposed in order to estimate the trend function and a simulation exercise showed good performance for the proposed bandwidth. We also employed another simulation where the period estimator behaved robustly in response to different bandwidth choices. As a complement, two applications applications were made: one for climatological data and the other for economic data. In the former, we used global temperture anomalies data which is exactly the same as that in Vogt and Linton (2014). The latter application consists in providing central estimates for the australian non-accelerating inflation rate of unemployment by means of the reversed estimation procedure.