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 TABLES

LIST OF FIGURES

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 $\delta > 0$, $\lvert a_n/b_n \rvert \le \delta$ for n sufficiently large.
$a_n = O(b_n)$	For some $C > 0$, $\lvert a_n/b_n \rvert \le C$ for n sufficiently large.
$X_n = o_p(a_n)$	For any $\delta, \epsilon > 0$, $P(\lvert X_n/a_n \rvert \ge \delta) \le \epsilon$ for n sufficiently large.
$X_n = O_p(a_n)$	For any $\epsilon > 0$, there is $C > 0$ such that $P(\lvert X_n/a_n \rvert \ge C) \le \epsilon$ 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 $C > 0$, $P(\limsup_{n\to\infty} \lvert X_n/a_n \rvert \le C) = 1$.
$\overset{d}{\to}$	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 $\lVert f \rVert_r = (\int_\Omega \lvert f \rvert^r dP)^{1/r} < +\infty, 1 \le r < +\infty$, 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 $x_{t,T} = t/T,\ t \in \{1, \ldots, T\},\ T \in \mathbb{N}$. 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

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

where $j_{\max} \in \mathbb{N}$ is fixed, $K_h(u) := K(u/h)/h$ with $K : \mathbb{R} \to \mathbb{R}$ being a kernel-like function and $h := h_T$ 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

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

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,

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

and

$$\beta > \frac{2s - 2}{s - 2}. \tag{2.4}$$

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

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

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

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

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*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Along the above lines,

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

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,

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

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 $|Z_{i,T}(x)| \leq 2\bar{K}\tau_T \leq C_1\tau_T \coloneqq b_T$ 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

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

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

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

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

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

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

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

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

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 $||s_j| - |\mu_j|| \leq |s_j - \mu_j|$ guarantees that $|s_j| = |\mu_j| + O(1/(Th))$ holds in Lemma 2.6. In addition, for any $x \in [0,1]$, we have $0 < \mu_j \leq C$ for $j \in \{0,2\}$ and $|\mu_1| \leq C$ 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

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

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 $\mu_0\mu_2 - \mu_1^2 \neq 0$. Then

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

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

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

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

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

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

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

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

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 $\alpha_{2,T}(j) \leq \alpha_{1,T}(j)$, for all $0 \leq j < T$.*

Proof. Fix $j : 0 \leq j < T$. Denote $\mathcal{G}_{T,i}^k = \sigma((f(Y_{l,T})) : i \leq l \leq k)$ and $\mathcal{F}_{T,i}^k = \sigma((Y_{l,T}) : i \leq l \leq k)$ for $1 \leq i \leq k \leq T$. 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 $\alpha_{2,T}(j) \leq \alpha_{1,T}(j)$. 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

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

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

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

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 $0 \le c \le 1$.

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,*

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

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

$$z'S_x z = \int_{G_x} z'XX'zKd\mu \ge 0,$$

for $X := X(w) = (1,w)'$. To get a contradiction, suppose $\exists y \ne 0 : \int_{[-c,c]} y'XX'yKd\mu = 0$. 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 $y'XX'y \overset{a.e.}{\ne} 0$ 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 $T_0 : \forall T \ge T_0$ 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 $S_{T,x}y = S_{T,x}(c_1 u + c_2 v) = c_1 \lambda_u u + c_2 \lambda_v v \ge \lambda_0 y$ when T is large enough. It implies $(1/\lambda_0)\|y\| \ge \|S_{T,x}^{-1} y\|$.

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 $\forall T \ge T^*, Th \ge 1/2$. Then for any $T \ge \max(T^*, T_0)$ and any $x \in [0,1]$, the weights of the local linear estimator defined in (2.19) satisfy*

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

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

(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 $t \in \{1, \dots, T\}$. By Lemma 2.7,

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

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

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

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

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

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 $C_1, C_2 \in \mathbb{R} : C_2 \ne 0$, 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 $Y_T \le X_T$ 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, $\exists T_0, M : P(|Y_T| \geq Ma_T) \leq \epsilon$ 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

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

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

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

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

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

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,

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

□

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

3 NONPARAMETRIC ESTIMATION OF A SMOOTH TREND IN THE PRESENCE OF A PERIODIC SEQUENCE

Abstract. We develop the asymptotic theory for the estimators derived from reversing the three-step procedure of Vogt and Linton (2014). We provide the uniform weak convergence rates of the trend function and periodic sequence estimators. We establish the asymptotic normality for the trend estimator. We also show that the period estimator is consistent.

Keywords: Nonparametric Regression. Periodic sequence. Asymptotic analysis

JEL Codes. C13, C14, C22;

3.1 Introduction

One way to deal with time series presenting a periodic and a trend behavior is to model them additively. That is, the series is written as the sum of a periodic and a trend components plus a stochastic error process. Although the nonparametric estimation of such model seems to be appealing due to its flexibility, in most studies the data are modeled as having only the trend or only the periodic component, and rarely both components are considered together. When the data has only the slowly varying component (plus an error term), its nonparametric estimation is popularly done by using a local polynomial fit (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, we can mention 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 the first step, the fundamental period of the periodic sequence is estimated. Given the period estimate, an estimate of the periodic sequence is provided in the second step. The last step consists in estimating the trend function using the local linear regression. Their asymptotic analysis investigated the uniform weak convergence rates and the asymptotic normality for the estimators of the trend function and the periodic sequence. In addition, the period estimator was proved to be consistent. 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 aim to investigate this reversed estimation version more deeply.

In this section, we develop the asymptotic theory for the estimators involved in the reversed procedure of Vogt and Linton (2014). We provide the uniform weak convergence of the estimators of the trend function and of the periodic sequence. The asymptotic normality for the trend estimator is also stablished. We also show that the period estimator is consistent.

3.2 The model

Let $T \in \mathbb{N}$ and assume the time series $\{Y_{t,T} : t = 1, \cdots, T\}$ is observed and follows the model

$$Y_{t,T} = g(t/T) + m(t) + \epsilon_{t,T}, \quad t = 1, 2, \ldots, T, \tag{3.1}$$

where g is a function of deterministic trend, $\{m(t)\}_{t\in\mathbb{N}}$ is a deterministic periodic sequence with unknown period $\theta_0 \in \mathbb{N}$ and $E(\epsilon_{t,T}) = 0$. By definition, the periodic sequence must satisfy $m(s) = m(s + k\theta_0)$ for any $s \in [\theta_0]$ and any $k \in \mathbb{N}$. Implicitly, θ_0 is assumed to be the smallest period of the sequence $m(t)$. For the asymptotic analysis, model (3.1) offers a framework such that as T grows we get additional information on the value of $g(t/T)$, at a given neighborhood of t/T, and on the value $m(s)$, for a given $s \in [\theta_0]$, due to its periodic property.

The assumption on $m(t)$ allows us to represent the values of the sequence as $m(t) = \sum_{s=1}^{\theta_0} \beta_s I_s(t)$ where $I_s(t) = I(t = s + k\theta_0 : k \in \mathbb{N})$ and I the indicator function. Note that this representation comes naturally from the periodicity of the sequence without having to make any additional parametric restriction.

In matrix notation, model (3.1) becomes

$$Y = g + X_{\theta_0}\beta + \epsilon, \tag{3.2}$$

where $Y = (Y_{1,T}, \ldots, Y_{T,T})'$ is the vector of observations, $g = \{g(1/T), \ldots, g(T/T)\}'$ is the trend component, $X_{\theta_0} = [I_{\theta_0} \quad I_{\theta_0} \quad \ldots]'$ is the design matrix with I_{θ_0} being the $\theta_0 \times \theta_0$ identity matrix and $\epsilon = (\epsilon_{1,T}, \ldots, \epsilon_{T,T})'$ is the error vector.

3.3 Estimation

The estimation procedure is done by reversing the steps of Vogt and Linton (2014) as they suggested in their supplementary material. We first estimate the trend function and then proceed by estimating the periodic sequence.

For the asymptotic analysis, we assume that m and g are normalized to satisfy $\sum_{s=1}^{\theta_0} m(s) = 0$. From now on, we denote by C a generic positive constant which may take different values at different appearances.

3.3.1 Step 1: Estimation of the Trend Function

If the periodic sequence m in equation (3.1) is known, then the local linear estimator for the trend g and its first derivative $g^{(1)}h$, at $x \in [0, 1]$, is given by

$$\tilde{P}(x) = \begin{bmatrix} \tilde{g}(x) \\ \tilde{g}^{(1)}(x)h \end{bmatrix} := S_{T,x}^{-1} A_{T,x} \tag{3.3}$$

where

$$S_{T,x} = \frac{1}{Th}\begin{bmatrix} \sum_{t=1}^{T} K\left(\frac{\frac{t}{T}-x}{h}\right) & \sum_{t=1}^{T}\left(\frac{\frac{t}{T}-x}{h}\right)K\left(\frac{\frac{t}{T}-x}{h}\right) \\ \sum_{t=1}^{T}\left(\frac{\frac{t}{T}-x}{h}\right)K\left(\frac{\frac{t}{T}-x}{h}\right) & \sum_{t=1}^{T}\left(\frac{\frac{t}{T}-x}{h}\right)^2 K\left(\frac{\frac{t}{T}-x}{h}\right) \end{bmatrix}, \; Z_{t,T} = Y_{t,T} - m(t)$$

$$A_{T,x} = \frac{1}{Th}\begin{bmatrix} \sum_{t=1}^{T} K\left(\frac{\frac{t}{T}-x}{h}\right) Z_{t,T} \\ \sum_{t=1}^{T}\left(\frac{\frac{t}{T}-x}{h}\right)K\left(\frac{\frac{t}{T}-x}{h}\right) Z_{t,T} \end{bmatrix},$$

with $S_{T,x}$ being an invertible matrix, $h_T := h$ a bandwidth sequence and K a kernel-like function. Straighforward calculations shows that we can write

$$\tilde{g}(x) = \sum_{t=1}^{T} W_{t,T}(x) Z_{t,T}, \tag{3.4}$$

where $W_{t,T}(x) = \frac{1}{Th} e_1' S_{T,x}^{-1} X\left(\frac{t/T-x}{h}\right) K\left(\frac{\frac{t}{T}-x}{h}\right)$ for $e_1 = (1,0)'$ and $X(u) = (1,u)'$. However, the estimator $\tilde{P}$ is infeasible since we do not observe $m(t)$. One could try to estimate g by simply ignoring the periodic component, i.e., using

$$\hat{g}(x) = \sum_{t=1}^{T} W_{t,T}(x) Y_{t,T}. \tag{3.5}$$

The local linear weights $W_{t,T}(x)$ can be readily replaced by Nadaraya-Watson's weights. Although the latter is simpler, it suffers from boundary bias (WAND; JONES, 1994, p. 126).

3.3.2 Step 2: Estimation of the Period

The period estimation is carried out by means of a penalized residual sum of squares minimization.

Let $S_{t,T} = Y_{t,T} - g(t/T)$. If the trend function were known, the period θ_0 could be estimated from

$$S = X_{\theta_0}\beta + \epsilon, \tag{3.6}$$

where $S = [S_{1,T}, \ldots, S_{T,T}]'$. For each $\theta \in \{1, \ldots, \Theta_T\}$ with $\Theta_T < T$, define the least squares estimate of model (3.6) with period θ by

$$\hat{\beta}_\theta = (X_\theta' X_\theta)^{-1} X_\theta' S, \tag{3.7}$$

where $X_\theta = [I_\theta \quad I_\theta \quad \ldots]'$ with I_θ being the $\theta \times \theta$ identity matrix. In addition, let the associated penalized residual sum of squares be given by

$$Q(\theta, \lambda_T) = \mathrm{RSS}(\theta) + \lambda_T \theta, \tag{3.8}$$

where λ_T is a divergent real sequence and $\text{RSS}(\theta) = \|S - X_\theta \hat{\beta}_\theta\|^2$ with $\|\cdot\|$ denoting the usual Euclidean norm on $\mathbb{R}^T$. The estimator of the period θ_0 is the minimizer

$$\hat{\theta} = \arg\min_{1 \leq \theta \leq \Theta_T} Q(\theta, \lambda_T). \tag{3.9}$$

The rates with which the sequences λ_T and Θ_T are allowed to diverge will be specified later on. The estimator in (3.9) is infeasible though. We approximate $\hat{\theta}$ using $\tilde{S}_{t,T} = Y_{t,T} - \hat{g}(t/T)$ through

$$\tilde{\theta} = \arg\min_{1 \leq \theta \leq \Theta_T} \tilde{Q}(\theta, \lambda_T) \tag{3.10}$$

where

$$\tilde{Q}(\theta, \lambda_T) = \overline{\text{RSS}}(\theta) + \lambda_T \theta \quad \text{and} \quad \overline{\text{RSS}}(\theta) = \|\tilde{S} - X_\theta \tilde{\beta}_\theta\|^2,$$

with $\tilde{\beta}_\theta = (X_\theta' X_\theta)^{-1} X_\theta' \tilde{S}$ and $\tilde{S} = [\tilde{S}_{1,T}, \ldots, \tilde{S}_{T,T}]'$.

As pointed out by Vogt and Linton (2014), this period estimation can also be regarded as a model selection problem. Also, the presence of the l_0-regularization parameter λ_T can prevent the period estimator from choosing large periods (multiples of θ_0). The selection of λ_T will be discussed in the next chapter.

3.3.3 Step 3: Estimation of the Periodic Sequence

If $S_{t,T} = Y_{t,T} - g(t/T)$ and θ_0 were known, we could estimate β using

$$\hat{\beta} = (X_{\theta_0}' X_{\theta_0})^{-1} X_{\theta_0} S. \tag{3.11}$$

We propose to estimate β by the feasible estimator

$$\tilde{\beta} = (X_{\tilde{\theta}}' X_{\tilde{\theta}})^{-1} X_{\tilde{\theta}} \tilde{S}. \tag{3.12}$$

3.4 Asymptotics

For the asymptotic analysis, the following conditions are made.

1. (**Condition 1**) The triangular array $\{\epsilon_{t,T}\}$ is strongly mixing with coefficients $\alpha(k)$ satisfying $\alpha(k) \leq C a^k$ for some positive constants $a < 1$ and C.
2. (**Condition 2**) $E(|\epsilon_{t,T}|^{(4+\delta)}) \leq C$ and $E(\epsilon_{t,T}^4 (\ln(1 + \epsilon_{t,T}))^3) \leq C$ for some constants $0 < \delta$ and $0 < C < \infty$.
3. (**Condition 3**) g is twice continuously differentiable on $[0, 1]$
4. (**Condition 4**) The kernel function K is nonnegative, symmetric around zero, Lipschitz continuous and has compact support.
5. (**Condition 5**) The bandwidth $h > 0$ satisfies $h \to 0$ and $Th^2 \to \infty$ as $T \to \infty$.

Condition 1 says that the error array is α-mixing with geometrically mixing rates.

Condition 2 gives uniform moment bounds for the error random variables. The bound on $\mathrm{E}(\epsilon_{t,T}^4(\ln(1+\epsilon_{t,T}))^3)$ will be shown to be important since we are allowing for nonstationarity. The conditions on g and K are standard to derive the properties of the local linear estimator.

Without loss of generality, assume $\operatorname{supp} K = [-1,1]$ and $\int_{\operatorname{supp} K} K(u)du = 1$.

To derive the asymptotic properties of $\hat{g}(x)$, define

$$V_{T,x} = \frac{1}{k_T}\sum_{t,j=1}^{T} K\left(\frac{t/T-x}{h}\right)K\left(\frac{j/T-x}{h}\right)E(\epsilon_{t,T}\epsilon_{j,T}), \quad x \in (h, 1-h),$$

where k_T is the cardinality of the set $J_x = \{i \in [T] : i/T \in (x-h, x+h)\}$.

Theorem 3.1. *Suppose Conditions 1-4 hold. If* $\ln T/(T^\theta h) = o(1)$ *for some* $\theta \in (0,1]$, *then it holds that*

$$\sup_{x\in[0,1]} |\hat{g}(x) - g(x)| = O_p\left(\sqrt{\frac{\ln T}{Th}} + h^2\right), \quad T \to \infty.$$

Moreover, if $V_x = \lim_{T\to\infty} V_{T,x}$ *exists and* $Th^5 = O(1)$, *then*

$$\sqrt{Th}\big(\hat{g}(x) - g(x) - J_x\big) \xrightarrow{d} 2N(0, V_x), \quad T \to \infty, \quad \forall x \in (h, 1-h),$$

where $J_x = 2^{-1}h^2 g''(x)\int u^2 K(u)du$.

Theorem 3.1 says that the local linear estimator still has good asymptotic properties if we ignore the presence of the periodic component. The uniform convergence rate is the same as that obtained in the oracle case. Inspecting the proof of the theorem, we can conclude that the naive estimator $\hat{g}$ has the oracle property, i.e., $\hat{g}$ has the same limiting distribution as that of estimator $\tilde{g}$, defined in (3.4), which is obtained assuming that m is known. We can also find that the replacement of $\tilde{g}(x)$ by $\hat{g}(x)$ results in an error of asymptotically negligible order $O(T^{-1})$, uniformly on x and h. This implies that the bandwidth for $\hat{g}$ could be selected using the same techniques as used for the estimator $\tilde{g}$. In the next chapter, however, we will see that employing asymptotic bandwidth selection rules for $\hat{g}$ may lead to poor performance on finite samples. [This is another theoretical result suggesting that the optimal plugin h for $\hat{g}$ is of order $T^{-1/5}$.]

Note that Theorem 3.1 can be applied to cases where the aim is only to estimate the trend function nonparametrically. If a correct examination detects the presence of a periodic component in the time series, then the direct application of the local linear estimator is acceptable, under certain circumstances.

Say that a real sequence a_T is $\Theta(b_T)$ if there are constants $m, M > 0$ such that $b_T m \leq a_T \leq M b_T$ for all sufficiently large T.

Theorem 3.2. *Let Conditions* $1-4$ *be fulfilled. Assume that the bandwidth satisfies* $h=\Theta(T^{-1/4})$ *and that* $\Theta_T \leq CT^{2/5-\omega}$ *, for some small* $\omega>0$*. Moreover, choose the regularization parameter* λ_T *to satisfy* $T^{1/4}\Theta_T\rho_T^{1/2}=o(\lambda_T)$ *and* $\lambda_T=o(T)$ *for some positive sequence* ρ_T *slowly diverging to infinity (e.g.* $\rho_T=\ln\ln T$*). Then* $\tilde{\theta}=\theta_0+o_p(1)$*.*

Theorem 3.3. *Let the conditions of Theorem 3.2 be satisfied. Then*

$$\max_{1\leq t\leq T}|\tilde{m}(t)-m(t)|=O_p\left(\left(\frac{\rho_T}{Th}\right)^{1/2}\right),\quad T\to\infty,$$

where ρ_T *is a positive sequence slowly diverging to infinity.*

3.5 Proofs

Appendix C contains several lemmas (from 3.1 to 3.11) which are used in the proofs of this section.

Proof of Theorem 3.1. Write

$$\begin{aligned}|\hat{g}(x)-g(x)| &\leq |\hat{g}(x)-\tilde{g}(x)|+|\tilde{g}(x)-g(x)|\\ &:= A_1+A_2, \end{aligned} \tag{3.13}$$

where $\tilde{g}$ is the estimator in the oracle case, defined in (3.4).

From Theorem 2.4 of Chapter 2, we have that $\sup_{x\in[0,1]}A_2=O_p(\sqrt{\ln T/(Th)}+h^2)$. Now, we show that $\sup_{x\in[0,1]}A_1$ is dominated by $\sup_{x\in[0,1]}A_2$. For this, we go along the lines of the proof of Theorem 2.4.

We have that $A_1=|e_1'S_{T,x}^{-1}M_{T,x}|$ where

$$M_{T,x}=\frac{1}{T}\begin{bmatrix}\sum_{t=1}^T K_h(t/T-x)m(t)\\ \sum_{t=1}^T(\frac{t/T-x}{h})K_h(t/T-x)m(t)\end{bmatrix}:=\begin{bmatrix}m_0\\ m_1\end{bmatrix}.$$

Then, rewrite

$$\sup_{x\in[0,1]}A_1=\sup_{x\in[0,1]}\left|e_1'\begin{bmatrix}s_0 & s_1\\ s_1 & s_2\end{bmatrix}^{-1}\begin{bmatrix}m_0\\ m_1\end{bmatrix}\right|=\sup_{x\in[0,1]}\left|\frac{m_0-s_1^2s_2^{-1}m_1}{s_0-s_1^2s_2^{-1}}\right|:=\sup_x\frac{V_n}{V_d},\tag{3.14}$$

where the dependencies of the entries on T and x were omitted, for brevity's sake. Consider the quantity $\mu_j, j\in\{0,1,2\}$, defined in Lemma 2.6 of Chapter 2. The fact $||s_j|-|\mu_j||\leq|s_j-\mu_j|$ guarantees that $|s_j|=|\mu_j|+O(1/(Th))$ also holds. In addition, given $x\in[0,1]$, we have $0<\mu_j\leq C$ for $j\in\{0,2\}$ and $|\mu_1|\leq C$, by Condition 4. It implies $\mu_1^2/\mu_2=O(1)$.

For any natural numbers $T,\theta_0>0$, define $E=\{i\in[\theta_0]:K_{i,T}^{\theta_0}=\lfloor T/\theta_0\rfloor+1\}$. Then, for $j \in \{0,1\}$, $x \in [0,1]$ and T sufficiently large, Lemmas 3.4 and 3.10 gives

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} m(t)K_h(t/T-x)\Big(\frac{t/T-x}{h}\Big)^j \\
&= \frac{1}{T}\sum_{i=1}^{\theta_0} m(i)\sum_{k\in J_{x,i}} K_h\Big(\frac{i+(k-1)\theta_0}{T}-x\Big)\Big(\frac{(i+(k-1)\theta_0)/T-x}{h}\Big)^j \\
&= \frac{1}{T}\sum_{i=1}^{\theta_0} m(i)O(K_{i,T}^{\theta_0}) = O\Big(\frac{1}{T}\Big)\Big[\sum_{i\in E} m(i)K_{i,T}^{\theta_0} + \sum_{i\notin E} m(i)K_{i,T}^{\theta_0}\Big] \\
&= O\Big(\frac{1}{T}\Big)\Big[\sum_{i\in E} m(i)\Big(\Big\lfloor\frac{T}{\theta_0}\Big\rfloor+1\Big) + \sum_{i\notin E} m(i)\Big\lfloor\frac{T}{\theta_0}\Big\rfloor\Big] \\
&= O\Big(\frac{1}{T}\Big)\Big[\underbrace{\sum_{i\in E} m(i)}_{\le C} + \underbrace{\sum_{i=1}^{\theta_0} m(i)}_{=0}\Big\lfloor\frac{T}{\theta_0}\Big\rfloor\Big] = O\Big(\frac{1}{T}\Big).
\end{aligned}$$

It implies that m_0 and m_1 are $O(1/T)$ uniformly on $x \in [0,1]$. Then

$$\begin{aligned}
V_n &\le \sup_{x\in[0,1]}|m_0| + |s_1^2 s_2^{-1}|\sup_{x\in[0,1]}|m_1| = O(T^{-1})\Big\{1+\frac{|\mu_1^2|+O(1/(Th))}{|\mu_2|+O(1/(Th))}\Big\} \\
&= O(T^{-1})\Big\{1+\Big|\frac{\mu_1^2}{\mu_2}\Big| + O\Big(\frac{1}{Th}\Big)\Big\} = O(T^{-1})
\end{aligned}$$

and

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

Lemma 2.7 of Chapter 2 guarantees that the limiting matrix of $S_{T,x}$ is positive definite. It implies that $\mu_0\mu_2 - \mu_1^2 \neq 0$. Then

$$\begin{aligned}
A_1 &= O(T^{-1})\sup_{x\in[0,1]}\Big|\frac{1}{\mu_0-\mu_1^2/\mu_2+O(1/(Th))}\Big| = O(T^{-1})\sup_{x\in[0,1]}\Big|\frac{\mu_2}{\mu_0\mu_2-\mu_1^2}+O\Big(\frac{1}{Th}\Big)\Big| \\
&= O(T^{-1}).
\end{aligned}$$

Hence $\sup_{x\in[0,1]}|\hat{g}(x)-g(x)| = O_p(\sqrt{\ln T/(Th)}+h^2+T^{-1}) = O_p(\sqrt{\ln T/(Th)}+h^2)$. To make the second equality clear, note that

$$Ta_T = \Big(\frac{T\ln T}{h}\Big)^{1/2} \to \infty \text{ and } Th^2 \to \infty$$

and so $\sqrt{\ln T/(Th)}+h^2+T^{-1} = O(\sqrt{\ln T/(Th)}+h^2)$.

We now turn to the asymptotic normality. Write

$$\begin{aligned}\hat{g}(x)-g(x) &= \big[\hat{g}(x)-\tilde{g}(x)\big]+\big[\tilde{g}(x)-g(x)\big] \\ &= \big[\hat{g}(x)-\tilde{g}(x)\big]+\left[\sum_{t=1}^{T} W_{t,T}(x)g(t/T)-g(x)\right]+\left[\sum_{t=1}^{T} W_{t,T}(x)\epsilon_{t,T}\right] \quad (3.15)\\ &:= g^O+g^B+g^V. \quad (3.16)\end{aligned}$$

From the previous part of the proof, $\sqrt{Th}g^O = O((h/T)^{1/2})$. Furthermore, standard calculations for the bias of the local linear estimator give $g^B = J_x + o(h^2)$ where $J_x = 2^{-1}h^2g''(x)\int u^2K(u)du$ (see Appendix G of Chapter 4). From (3.3), we can express the stochastic term by

$$g^V = e_1' S_{T,x}^{-1} V_{T,x}, \quad (3.17)$$

where

$$V_{T,x} = \frac{1}{T}\begin{bmatrix} \sum_{t=1}^{T} K_h(t/T-x)\epsilon_{t,T} \\ \sum_{t=1}^{T} (\frac{t/T-x}{h}) K_h(t/T-x)\epsilon_{t,T}\end{bmatrix} := \begin{bmatrix} v_0 \\ v_1 \end{bmatrix}.$$

As before, rewrite

$$g^V = \frac{v_0 - s_1^2 s_2^{-1} v_1}{s_0 - s_1^2 s_2^{-1}} = \frac{v_0}{s_0 - s_1^2 s_2^{-1}} - \frac{s_1^2 s_2^{-1} v_1}{s_0 - s_1^2 s_2^{-1}},$$

where

$$s_1^2 s_2^{-1} = \frac{\mu_1^2 + O(1/(Th))}{\mu_2 + O(1/(Th))} = \frac{\mu_1^2}{\mu_2} + O(1/(Th))$$

and

$$s_0 - s_1^2 s_2^{-1} = \frac{\mu_2\mu_0 - \mu_1^2}{\mu_2} + O(1/(Th)),$$

using Lemma 2.6 of Chapter 2. The assumption that $x \in (h, 1-h)$ implies $\mu_1 = 0$ and $\mu_0 = 1$ for T large enough. Also, we have that $0 < \mu_2 \leq 1$ and that $\mu_2\mu_0 - \mu_1^2 \neq 0$, where the latter is implied by Lemma 2.7 of Chapter 2 . Thus, from Theorem 2.2 and Lemma 2.9 of Chapter 2, it is easily seen that

$$\begin{aligned}g^V &= \frac{1}{1+O(1/(Th))}v_0 + \frac{O(1/(Th))}{1+O(1/(Th))}v_1 = (1+O(1/(Th)))v_0 + O(1/(Th))v_1 \\ &= v_0 + O(1/(Th))O_p(\sqrt{\ln T/(Th)}) = v_0 + o_p(1/\sqrt{Th}).\end{aligned}$$

By definition,

$$\sqrt{Th}v_0 = \sqrt{\frac{k_T}{Th}}\frac{1}{\sqrt{k_T}}\sum_{t=1}^{T} K\left(\frac{t/T-x}{h}\right)\epsilon_{t,T} I(t \in J_x) = \sqrt{\frac{k_T}{Th}}\frac{1}{\sqrt{k_T}}\sum_{t=1}^{T} X_{t,T} \quad (3.18)$$

where $X_{t,T} := K((t/T-x)/h)\epsilon_{t,T} I(t \in J_x)$. Lemma 3.11 implies that the triangle array

$\{X_{t,T}\}$ is also strong mixing with mixing coefficients bounded by the mixing coefficients of $\{\epsilon_{t,T}\}$. The problem of obtaining exact value of k_T is similar to that of counting the number of $1/T$-periodic points on the set $(0, 2h)$. Therefore, k_T equals $K^{1/T}_{s,2h} = \lfloor (2h - s)/(1/T) \rfloor$, for some $0 < s \leq 1/T$. Thus $k_T \overset{a}{\approx} 2Th - sT \overset{a}{\approx} 2Th$, since $0 < sT \leq 1$. It implies $\sqrt{k_T/(Th)} \overset{a}{\approx} \sqrt{2}$.

The application of Politis-Ekstrom's CLT (Theorem 3.6 of Appendix D) and Corollary 3.6.1 gives

$$\sqrt{Th}v_0 \xrightarrow{d} 2N(0, V_x). \tag{3.19}$$

Hence, by Slutsky's Theorem,

$$\sqrt{T}h\big(\hat{g}(x) - g(x) - J_x\big) = o(1) + o(\sqrt{Th^5}) + o_p(1) + \sqrt{Th}v_0 \xrightarrow{d} 2N(0, V_x).$$

Proof of Theorem 3.2 By Lemmas 3.2 ,3.8 and 3.9, we obtain that $P(\tilde{\theta} \neq \theta_0) = o(1)$. Note that, for any $\delta > 0$, $\{|\tilde{\theta} - \theta_0| \geq \delta\} \subseteq \{|\tilde{\theta} - \theta_0| > 0\} = \{\tilde{\theta} \neq \theta_0\}$. Using the monotonicity of the measure and taking limits, it follows that $\lim_{T\to\infty} P(|\tilde{\theta} - \theta_0| \geq \delta) = 0,\ \forall \delta > 0$, that is, $\tilde{\theta} - \theta_0 = o_p(1)$.

Proof of Theorem 3.3 Denote

$$\begin{aligned}
\tilde{m}(s) &= e_s'\tilde{\beta} = e_s'(X_{\tilde{\theta}}'X_{\tilde{\theta}})^{-1}X_{\tilde{\theta}}'\tilde{S}, \quad s \in \{1, \dots, \tilde{\theta}\} \\
\bar{m}(s) &= e_s'\bar{\beta} = e_s'(X_{\theta_0}'X_{\theta_0})^{-1}X_{\theta_0}'\tilde{S}, \quad s \in \{1, \dots, \theta_0\} \\
\hat{m}(s) &= e_s'\hat{\beta} = e_s'(X_{\theta_0}'X_{\theta_0})^{-1}X_{\theta_0}'S, \quad s \in \{1, \dots, \theta_0\}.
\end{aligned}$$

Then

$$\begin{aligned}
\max_{1\leq t\leq T}|\tilde{m}(t) - m(t)| \leq& \max_{1\leq t\leq T}|\tilde{m}(t) - \bar{m}(t)| + \max_{1\leq t\leq T}|\bar{m}(t) - \hat{m}(t)| + \max_{1\leq t\leq T}|\hat{m}(t) - m(t)| \\
:=& M_1 + M_2 + M_3.
\end{aligned}$$

By monotonicity ans subadditivity of the measure and by Theorem 3.2, we have that

$$\begin{aligned}
P(\sqrt{Th}M_1 > \delta) &\leq P(\sqrt{Th}\max_{1\leq t\leq T}|\tilde{m}(t) - \bar{m}(t)| > \delta, \tilde{\theta} = \theta_0) + P(\tilde{\theta} \neq \theta_0) \\
&\leq \sum_{t=1}^{\theta_0} P(\sqrt{Th}|\tilde{m}(t) - \bar{m}(t)| > \delta, \tilde{\theta} = \theta_0) + o(1) \\
&= \theta_0 P(\emptyset) + o(1) = o(1),
\end{aligned}$$

for every $\delta > 0$, since $\bar{\beta} = \tilde{\beta}$ when $\tilde{\theta} = \theta_0$. Thus

$$\sqrt{Th}M_1 = o_p(1). \tag{3.20}$$

For the term M_2, observe that

$$\bar{\beta}-\hat{\beta}=D_{\theta_0}X'_{\theta_0}(\hat{g}-g)=\begin{bmatrix}\frac{1}{K^{\theta_0}_{1,T}} & & \\ & \ddots & \\ & & \frac{1}{K^{\theta_0}_{\theta_0,T}}\end{bmatrix}\begin{bmatrix}\sum_{k=1}^{K^{\theta_0}_{1,T}}\hat{g}\left(\frac{1+(k-1)\theta_0}{T}\right)-g\left(\frac{1+(k-1)\theta_0}{T}\right)\\ \vdots \\ \sum_{k=1}^{K^{\theta_0}_{\theta_0,T}}\hat{g}\left(\frac{\theta_0+(k-1)\theta_0}{T}\right)-g\left(\frac{\theta_0+(k-1)\theta_0}{T}\right)\end{bmatrix}$$

$$=\begin{bmatrix}\bar{m}(1)-\hat{m}(1)\\ \vdots \\ \bar{m}(\theta_0)-\hat{m}(\theta_0)\end{bmatrix}.$$

In addition, both $\bar{m}(t)$ and $\hat{m}(t)$ are θ_0-periodic, and then, $\max_{1\le t\le T}|\bar{m}(t)-\hat{m}(t)|=\max_{1\le t\le\theta_0}|\bar{m}(t)-\hat{m}(t)|$. We can then represent M_2 as averages of the form:

$$\max_{1\le t\le\theta_0}|\bar{m}(t)-\hat{m}(t)|=\max_{1\le t\le\theta_0}\left|\frac{1}{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,t},T}}\hat{g}\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)-g\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)\right|$$

$$\le\max_{1\le t\le\theta_0}\left|\frac{1}{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{i=1}^{T}W_{i,T}\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)\left[g\left(\frac{i}{T}\right)-g\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)\right]\right|$$

$$+\max_{1\le t\le\theta_0}\left|\frac{1}{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{i=1}^{T}W_{i,T}\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)m(i)\right|$$

$$+\max_{1\le t\le\theta_0}\left|\frac{1}{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{i=1}^{T}W_{i,T}\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)\epsilon_{i,T}\right|$$

$$:=M_2^g+M_2^m+M_2^\epsilon.$$

The non-stochastic terms satisfy

$$\sqrt{Th}M_2^g\le C(Th)^{1/2}h^2\le C(Th^5)^{1/2}=o(1), \tag{3.21}$$

$$\sqrt{Th}M_2^m\le C(Th)^{1/2}\frac{1}{T}\le C\left(\frac{h}{T}\right)^{1/2}=o(1). \tag{3.22}$$

On the other hand, for each $t\in\{1,\ldots,T\}$, we have

$$E\left\{\frac{1}{(K^{\theta_0}_{w_{\theta_0,t},T})^2}\sum_{k,k'=1}^{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{i,j=1}^{T}W_{i,T}\left(\frac{w_{\theta_0,t}+(k-1)\theta_0}{T}\right)W_{j,T}\left(\frac{w_{\theta_0,t}+(k'-1)\theta_0}{T}\right)\epsilon_{i,T}\epsilon_{j,T}\right\}$$

$$\le\frac{1}{(K^{\theta_0}_{w_{\theta_0,t},T})^2}\sum_{k,k'=1}^{K^{\theta_0}_{w_{\theta_0,t},T}}\sum_{i,j=1}^{T}(\sup_{x,i}|W_{i,T}(x)|)^2|\mathrm{Cov}(\epsilon_{i,T},\epsilon_{j,T})|\le\frac{C}{Th},$$

and then Chebychev's inequality implies

$$P\left(M_2^{\epsilon} > C\sqrt{\frac{\rho_T}{Th}}\right) \leq \sum_{i=1}^{\theta_0} \frac{C}{Th}\frac{Th}{\rho_T} = o(1).$$

Thus $M_2^{\epsilon} = O_p(\sqrt{\rho_t/(Th)})$. Analogously, we can obtain that

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{K^{\theta_0}_{w_{\theta_0},t,T}} \sum_{k=1}^{K^{\theta_0}_{w_{\theta_0},t,T}} \epsilon_{w_{\theta_0},t+(k-1)\theta_0,T}\right) \\
&\leq \frac{1}{(K^{\theta_0}_{w_{\theta_0},t,T})^{1/2}} \sum_{k,k'=1}^{K^{\theta_0}_{w_{\theta_0},t,T}} |\mathrm{Cov}(\epsilon_{w_{\theta_0},t+(k-1)\theta_0,T}, \epsilon_{w_{\theta_0},t+(k'-1)\theta_0,T})| \leq \frac{C}{K^{\theta_0}_{w_{\theta_0},t,T}} \sum_{k=0}^{\infty} (a^{\theta_0})^k \\
&\leq \frac{C}{T}.
\end{aligned}$$

Therefore,

$$P\left(M_3 > \frac{C}{\sqrt{Th}}\right) \leq \sum_{i=1}^{\theta_0} \frac{C}{T} Th = o(1).$$

Finally, combining these results we obtain

$$\begin{aligned}
\max_{1\leq t\leq T} |\tilde{m}(t) - m(t)| &= o_p((Th)^{-1/2}) + o((Th)^{-1/2}) + O_p((Th)^{-1/2}) + O_p(\rho_T (Th)^{-1/2}) \\
&= O_p((Th)^{-1/2} + \rho_T (Th)^{-1/2}) = O_p(\rho_T (Th)^{-1/2}).
\end{aligned}$$

3.6 References

BALYAEV, Y.; SJÖSTEDT-DE LUNA, S. Weakly approaching sequences of random distributions. *Journal of Applied Probability*, v. 37, n. 3, p. 807-822, 2000. Available in <https://www.jstor.org/stable/3215615?seq=1>. Acessed on 26/05/2020.

BILLINGSLEY, P. *Probability and measure.* John Wiley & Sons, New York, 1995.

BROOMHEAD, D. S.; KING, G. P. Extracting qualitative dynamics from experimental data. *Physica D*, v. 20, n. 2-3, p. 217-236, 1986. Available in <https://www.sciencedirect.com/science/article/abs/pii/016727898690031X>. Acessed on 25/08/2020.

BROOMHEAD, D. S. et al. Singular System Analysis with Application to Dynamical Systems. In: PIKE, E.R.; LUGAITO, L.A. (Ed.). *Chaos, noise and fractals.* Bristol: IOP Publishing, p. 15-27, 1987.

CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, v. 74, n. 368, p. 829-836, 1979. Available in <https://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10481038>. Acessed on 25/08/2020.

CURTISS, J. H. A Note on the Theory of Moment Generating Functions. *The Annals of Mathematical Statistics*, v.13, n. 4, p. 430-433, 1942. Available in <http://www.jstor.org/stable/2235846>. Acessed on 26/05/2020.

EKSTRÖM, M. A general central limit theorem for strong mixing sequences. *Statistics & Probability Letters*, v. 94, p. 236-238, 2014. Available in <https://www.sciencedirect.com/science/article/abs/pii/S0167715214002624>. Acessed on 26/05/2020.

EUBANK, R. L. *Nonparametric regression and spline smoothing.* CRC Press, 1999.

FAN, J. Design-adaptive nonparametric regression. *Journal of the American statistical Association*, v. 87, n. 420, p. 998-1004, 1992. Available in <https://www.jstor.org/stable/2290637?seq=1>. Acessed on 25/08/2020.

GREEN, P. J.; SILVERMAN, B. W. *Nonparametric regression and generalized linear models*: a roughness penalty approach. CRC Press, 1993.

HALL, P.; REIMANN, J.; RICE, J. Nonparametric estimation of a periodic function. *Biometrika*, v. 87, n. 3, p. 545-557, 2000. Available in <https://www.jstor.org/stable/2673629?seq=1>. Acessed on 25/08/2020.

NADARAYA, E. A. On estimating regression. *Theory of Probability & Its Applications*, v. 9, n. 1, p. 141-142, 1964. Available in <https://epubs.siam.org/doi/10.1137/1109020>. Acessed on 25/08/2020.

POLITIS, D. N.; ROMANO, J. P.; WOLF, M. Subsampling for heteroskedastic time series. *Journal of Econometrics*, v. 81, n.2, p. 238-317, 1997. Available in <https://www.sciencedirect.com/science/article/abs/pii/S0304407697865694>. Acessed on 26/05/2020.

RIO, E. . *Asymptotic Theory of Weakly Dependent Random Processes.* Probability Theory and Stochastic Modelling Series (80). Springer Verlag Berlin Heidelberg, Alemanha, 2017.

SUN, Y.; HART, J. D.; GENTON, M. G. Nonparametric inference for periodic sequences. *Technometrics*, v. 54, n. 1, p. 83-96, 2012. Available in <https://www.tandfonline.com/doi/abs/10.1080/00401706.2012.650499>. Acessed on 25/08/2020.

VAART, A. W. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

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://doi.org/10.1093/biomet/ast051>. Acessed on 26/05/2020.

WAHBA, G. *Spline models for observational data*. SIAM, 1990.

WAND, M. P.; JONES, M. C. *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability (60). Chapman & Hall, Boca Raton, Florida, 1994.

WATSON, G. S. Smooth regression analysis. *Sankhyā*: The Indian Journal of Statistics, Series A, v. 26, n. 4, p. 359-372, 1964. Available in <https://www.jstor.org/stable/pdf/25049340.pdf?seq=1>. Acessed on 25/08/2020.

Appendix C - Technical Details

We start by stating some preliminary results. For every $\theta \in \mathbb{N}$, let $\Pi_\theta = X_\theta(X_\theta' X_\theta)^{-1}X_\theta'$ be the matrix of projection onto the column space of X_θ and call $M_\theta = I - \Pi_\theta$ as the annihilator matrix of X_θ, with $X_\theta = [I_\theta \quad I_\theta \quad \dots]'$ being a $T \times \theta$ matrix. To simplify the notations, we will use the shorthands $X \coloneqq X_\theta$, $\Pi \coloneqq \Pi_\theta$ and $M \coloneqq M_\theta$ whenever no risk of confusion exists.

Lemma 3.1. *Let $\theta \in \mathbb{N}$. It holds that:*

(i) Π and M are symmetric and idempotent;

(ii) for a regression model $S = X\beta + \epsilon$, the least residual sum of squares can be written as $RSS(\theta) = S'MS$.

Proof. (i) $\Pi' = [X(X'X)^{-1}X']' = X[(X'X)^{-1}]'X' = X[(X'X)']^{-1}X' = X(X'X)^{-1}X' = \Pi$. Further, $\Pi^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = XI(X'X)^{-1}X' = \Pi$.
The annihilator is also symmetric as $M' = I - \Pi' = I - \Pi = M$. In addition, $M^2 = (I-\Pi)(I-\Pi) = I - 2\Pi + \Pi^2 = I - 2\Pi + \Pi = I - \Pi = M$.
(ii) Since $\Pi S = X(X'X)^{-1}X'S = X\hat{\beta}$, it follows that $S - X\hat{\beta} = S - \Pi S = (I-\Pi)S = MS$. Hence, $RSS(\theta) = (S - X\hat{\beta})'(S - X\hat{\beta}) = S'M'MS = S'M^2S = S'MS$. □

Lemma 3.2. *Let*

$$g_b = (g_b(1/T), \dots, g_b(T/T))' \text{ with } g_b(x) = \sum_{i=1}^{T} W_{i,T}(x)[g(x) - g(i/T)];$$

$$g_m = (g_m(1/T), \dots, g_m(T/T))' \text{ with } g_m(x) = \sum_{i=1}^{T} W_{i,T}(x)m(i); \text{ and}$$

$$g_\epsilon = (g_\epsilon(1/T), \dots, g_\epsilon(T/T))' \text{ with } g_\epsilon(x) = \sum_{i=1}^{T} W_{i,T}(x)\epsilon_{i,T}.$$

Denote $B = \Pi_\theta - \Pi_{\theta_0}$. Then

$$P(\tilde{\theta} \neq \theta_0) \leq \sum_{\substack{1 \leq \theta \leq \Theta_T \\ \theta \neq \theta_0}} P\{\tilde{Q}(\theta, \lambda_T) \leq \tilde{Q}(\theta_0, \lambda_T)\}, \tag{3.23}$$

and, for each $\theta \neq \theta_0$, it holds that

$$\begin{aligned} P\{\tilde{Q}(\theta, \lambda_T) \leq \tilde{Q}(\theta_0, \lambda_T)\} = P\{&V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \leq -B_\theta + W_\theta^{g_b} + W_\theta^{g_m} \\ &- 2S_\theta^{g_b} + 2S_\theta^{g_m} - 2U_\theta^{(g_b,g_m)} - 2U_\theta^{(g_b,g_\epsilon)} - 2U_\theta^{(g_b,\epsilon)} + 2U_\theta^{(g_m,g_\epsilon)} - 2U_\theta^{(g_m,\epsilon)} + S_\theta^{g_\epsilon} \\ &- S_\theta^{\epsilon} + \lambda_T(\theta_0 - \theta)\}, \end{aligned} \tag{3.24}$$

where

$$V_\theta^{(\epsilon,\epsilon)} = -\epsilon' B\epsilon, \qquad V_\theta^{(\epsilon,g_\epsilon)} = -\epsilon' B g_\epsilon, \qquad V_\theta^{(g_\epsilon,g_\epsilon)} = -g_\epsilon' B g_\epsilon, \qquad B_\theta = (X_{\theta_0}\beta)' M_\theta (X_{\theta_0}\beta).$$
$$W_\theta^{g_b} = g_b' B g_b, \qquad W_\theta^{g_m} = g_m' B g_m, \qquad S_\theta^{g_b} = (X_{\theta_0}\beta)' M_\theta g_b, \qquad S_\theta^{g_m} = (X_{\theta_0}\beta)' M_\theta g_m,$$
$$U_\theta^{(g_b,g_m)} = g_b' B g_m, \qquad U_\theta^{(g_b,g_\epsilon)} = g_b' B g_\epsilon, \qquad U_\theta^{(g_b,\epsilon)} = g_b' B\epsilon, \qquad U_\theta^{(g_m,g_\epsilon)} = g_m' B g_\epsilon$$
$$U_\theta^{(g_m,\epsilon)} = g_m' B\epsilon, \qquad S_\theta^{g_\epsilon} = (X_{\theta_0}\beta)' M_\theta g_\epsilon, \qquad S_\theta^{\epsilon} = (X_{\theta_0}\beta)' M_\theta \epsilon.$$

Proof. Since $\tilde\theta \in [\Theta_T]$,

$$w \in \{\tilde\theta \neq \theta_0\} \iff \exists \theta_1 \in [\Theta_T] \setminus \{\theta_0\} : w \in \{\tilde\theta = \theta_1\} \iff w \in \bigcup_{\theta_1 \in [\Theta_T]\setminus\{\theta_0\}} \{\tilde\theta = \theta_1\}.$$

From the subadditivity of the measure, $P(\tilde\theta \neq \theta_0) \leq \sum_{\substack{1\leq\theta_1\leq\Theta_T \\ \theta_1\neq\theta_0}} P(\tilde\theta = \theta_1)$. For each $\theta_1 \in [\Theta_T] \setminus \{\theta_0\}$, the following relations hold

$$\begin{aligned}\tilde\theta = \theta_1 &\iff \arg\min_{1\leq\theta\leq\Theta_T} \tilde Q(\theta,\lambda_T) = \theta_1 \iff \min_{1\leq\theta\leq\Theta_T} \tilde Q(\theta,\lambda_T) = \tilde Q(\theta_1,\lambda_T) \\ &\iff \tilde Q(\theta,\lambda_T) \geq \tilde Q(\theta_1,\lambda_T), \forall\theta \in \{1,\ldots,\Theta_T\}. \end{aligned} \tag{3.25}$$

It implies that, for each $\theta_1 \in [\Theta_T] \setminus \{\theta_0\}$,

$$\{\tilde\theta = \theta_1\} = \bigcap_{1\leq\theta\leq\Theta_T} \{\tilde Q(\theta,\lambda_T) \geq \tilde Q(\theta_1,\lambda_T)\} \subseteq \{\tilde Q(\theta_0,\lambda_T) \geq \tilde Q(\theta_1,\lambda_T)\},$$

and then $P(\tilde\theta = \theta_1) \leq P(\tilde Q(\theta_0,\lambda_T) \geq \tilde Q(\theta_1,\lambda_T))$, by the monotonicity of the measure. Thus $P(\tilde\theta \neq \theta_0) \leq \sum_{\substack{1\leq\theta_1\leq\Theta_T \\ \theta_1\neq\theta_0}} P\{\tilde Q(\theta_1,\lambda_T) \leq \tilde Q(\theta_0,\lambda_T)\}$.

Lemma 3.1 implies that $\overline{\text{RSS}}(\theta) = \tilde S' M_\theta \tilde S$. Therefore, from (3.2),

$$\begin{aligned}\overline{\text{RSS}}(\theta) =& \big(X_{\theta_0}\beta + (g-\hat g) + \epsilon\big)' M_\theta \big(X_{\theta_0}\beta + (g-\hat g) + \epsilon\big) \\ =& \big(X_{\theta_0}\beta + g_b - g_m - g_\epsilon + \epsilon\big)' M_\theta \big(X_{\theta_0}\beta + g_b - g_m - g_\epsilon + \epsilon\big) \\ =& (X_{\theta_0}\beta)' M_\theta (X_{\theta_0}\beta) + g_b' M_\theta g_b + g_m' M_\theta g_m + g_\epsilon' M_\theta g_\epsilon + \epsilon' M_\theta \epsilon \\ &+ 2\big[(X_{\theta_0}\beta)' M_\theta g_b - (X_{\theta_0}\beta)' M_\theta g_m - (X_{\theta_0}\beta)' M_\theta g_\epsilon + (X_{\theta_0}\beta)' M_\theta \epsilon \\ &- g_b' M_\theta g_m - g_b' M_\theta g_\epsilon + g_b' M_\theta \epsilon + g_m' M_\theta g_\epsilon - g_m' M_\theta \epsilon + g_\epsilon' M_\theta \epsilon\big],\end{aligned} \tag{3.26}$$

where $\hat g = (\hat g(1/T),\ldots,\hat g(T/T))'$ with $\hat g(x) = \sum_{i=1}^T W_{i,T}(x) Y_{i,T} = \sum_{i=1}^T W_{i,T}(x)[g(i/T) + m(i) + \epsilon_{i,T}]$.

Since M_{θ_0} annihilates X_{θ_0}, i.e., $M_{\theta_0} X_{\theta_0} = 0$, we immediately see from equation (3.26) that

$$\overline{\text{RSS}}(\theta_0) = g_b' M_{\theta_0} g_b + g_m' M_{\theta_0} g_m + g_\epsilon' M_{\theta_0} g_\epsilon + \epsilon' M_{\theta_0} \epsilon + 2\big(-g_b' M_{\theta_0} g_m$$

$$- g_b' M_{\theta_0} g_\epsilon + g_b' M_{\theta_0} \epsilon + g_m' M_{\theta_0} g_\epsilon - g_m' M_{\theta_0} \epsilon + g_\epsilon' M_{\theta_0} \epsilon\big). \tag{3.27}$$

Hence, from definition (3.10),

$$\begin{aligned}
0 \le \tilde{Q}(\theta_0, \lambda_T) - \tilde{Q}(\theta, \lambda_T) &= \tilde{S}' M_{\theta_0} \tilde{S} - \tilde{S}' M_\theta \tilde{S} + \lambda_T(\theta_0 - \theta) \\
&= -(X_{\theta_0}\beta)' M_\theta (X_{\theta_0}\beta) + g_b'(M_{\theta_0} - M_\theta) g_b + g_m'(M_{\theta_0} - M_\theta) g_m - g_\epsilon'(M_\theta - M_{\theta_0}) g_\epsilon \\
&- \epsilon'(M_\theta - M_{\theta_0})\epsilon + 2\big[-(X_{\theta_0}\beta)' M_\theta g_b + (X_{\theta_0}\beta)' M_\theta g_m + (X_{\theta_0}\beta)' M_\theta g_\epsilon - (X_{\theta_0}\beta)' M_\theta \epsilon \\
&- g_b'(M_{\theta_0} - M_\theta) g_m - g_b'(M_{\theta_0} - M_\theta) g_\epsilon + g_b'(M_{\theta_0} - M_\theta)\epsilon + g_m'(M_{\theta_0} - M_\theta) g_\epsilon \\
&- g_m'(M_{\theta_0} - M_\theta)\epsilon - g_\epsilon'(M_\theta - M_{\theta_0})\epsilon\big]
\end{aligned}$$

which gives the desired result. □

Now we need to investigate the structure of the terms described in equation (3.24).

Given a sample size $T \in \mathbb{N}$, a period $\theta \in \{1, \dots, \Theta_T\}$ and a point $s \in \{1, \dots, \theta\}$, define the subset $A^\theta_{s,T} = \{s + k\theta : k \in \mathbb{N}\} \subseteq [T]$. In addition, denote the cardinality of $A^\theta_{s,T}$ by $K^\theta_{s,T}$. In words, $A^\theta_{s,T}$ is the set of θ-periodic points, starting at s, in $\{1, \dots, T\}$.

Lemma 3.3. *It holds that* $K^\theta_{s,T} = \lfloor \frac{T-s}{\theta} \rfloor + 1$.

Proof. The elements of $A^\theta_{s,T}$ are $s, s+\theta, s+2\theta, \dots, s+(K^\theta_{s,T}-1)\theta$. Clearly, T is an upper bound of $A^\theta_{s,T}$ and $s + K^\theta_{s,T}\theta > T$. Then

$$\begin{aligned}
s + (K^\theta_{s,T} - 1)\theta \le T \text{ and } s + K^\theta_{s,T}\theta > T &\iff K^\theta_{s,T} \le \frac{T-s}{\theta} + 1 \text{ and } K^\theta_{s,T} > \frac{T-s}{\theta} \\
&\iff \frac{T-s}{\theta} - 1 < K^\theta_{s,T} - 1 \le \frac{T-s}{\theta}.
\end{aligned} \tag{3.28}$$

Claim 3. Let $a \in \mathbb{R}$ and $K \in \mathbb{Z}$. Then $a - 1 < K \le a$ implies $K = \lfloor a \rfloor$.

Proof of claim: From $K \le a$, we have $K \le \lfloor a \rfloor$. On the other hand, from $a - 1 < K$, we have $a < K + 1$, which implies $\lfloor a \rfloor < K + 1$ since $\lfloor a \rfloor \le a$. Also, $\lfloor a \rfloor < K + 1$ if, and only if, $\lfloor a \rfloor \le K$. Hence $K = \lfloor a \rfloor$. ■

The application of *Claim* 2 in (3.28) leads to $K^\theta_{s,T} = \left\lfloor \frac{T-s}{\theta} \right\rfloor + 1$. □

We shall see that $K^\theta_{s,T} \overset{a}{\approx} T/\theta$ as a consequence of Lemma 3.3.

Lemma 3.4. $K^\theta_{s,T}$ *is either* $\left\lfloor \frac{T}{\theta} \right\rfloor$ *or* $\left\lfloor \frac{T}{\theta} \right\rfloor + 1$.

Proof. From Lemma 3.3, $K^\theta_{s,T} = \lfloor \frac{T-s}{\theta} \rfloor + 1$.

Claim 4. For any $a, b \in \mathbb{R}$, $\lfloor a \rfloor + \lfloor b \rfloor \le \lfloor a + b \rfloor \le \lfloor a \rfloor + \lfloor b \rfloor + 1$.

Proof of claim: Put $a = \lfloor a \rfloor + c_1$ and $b = \lfloor b \rfloor + c_2$ with $c_1, c_2 \in [0, 1)$. Then $0 \le c_1 + c_2 < 2 \implies 0 \le \lfloor c_1 + c_2 \rfloor \le 1$. Since $\lfloor a + b \rfloor = \lfloor \lfloor a \rfloor + \lfloor b \rfloor + c_1 + c_2 \rfloor = \lfloor a \rfloor + \lfloor b \rfloor + \lfloor c_1 + c_2 \rfloor$, we immediately have the result. ■

By *Claim* 4, we have that $\left\lfloor \frac{T}{\theta} \right\rfloor + \left\lfloor \frac{-s}{\theta} \right\rfloor \le \left\lfloor \frac{T-s}{\theta} \right\rfloor \le \left\lfloor \frac{T}{\theta} \right\rfloor + \left\lfloor \frac{-s}{\theta} \right\rfloor + 1$. Since $\left\lfloor \frac{-s}{\theta} \right\rfloor = -\left\lceil \frac{s}{\theta} \right\rceil$ and $\left\lceil \frac{s}{\theta} \right\rceil = 1$ for $s = 1, \dots, \theta$, it follows that $\left\lfloor \frac{T}{\theta} \right\rfloor - 1 \le \left\lfloor \frac{T-s}{\theta} \right\rfloor \le \left\lfloor \frac{T}{\theta} \right\rfloor$. Thus

$\left\lfloor \frac{T}{\theta} \right\rfloor \leq K^{\theta}_{s,T} \leq \left\lfloor \frac{T}{\theta} \right\rfloor + 1$, and so the integer $K^{\theta}_{s,T}$ can only be $\left\lfloor \frac{T}{\theta} \right\rfloor$ or $\left\lfloor \frac{T}{\theta} \right\rfloor + 1$. $\square$

Let $w^* : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ be a function given by $w^*(x, y) = y - \lfloor \frac{y-1}{x} \rfloor x$ and let the section of w^* at x, denoted by $w^*_x : E_x \to \mathbb{Z}$, be defined as $y \mapsto w^*(x, y)$, where $E_x = \{y : (x, y) \in \mathbb{Z} \times \mathbb{Z}\}$. If $w_T : [\Theta_T] \times [T] \to [T]$ is defined exactly as w^*, i.e., $w_T(\theta, t) = t - \lfloor (t-1)/\theta \rfloor$, then its section satisfies $w_{\theta,T} : [T] \to \{1, \ldots, \theta\}$. For brevity's sake, we will use the shorthand $w_{\theta,T}(t) \stackrel{\text{def}}{=} w_{\theta,t}$. Intuitively, this function is an initial point catcher in the sense that if $t \in A^{\theta}_{s,T}$, then $w_{\theta,t} = s$. Although this function may look fairly technical, it will be conveninent when exploiting the structure of the matrices involved in the estimator (3.10).

Lemma 3.5. *The projection matrix* Π_θ *is given by*

$$\Pi_\theta = X_\theta D X'_\theta,$$

where $D = diag(1/K^{\theta}_{1,T}, \ldots, 1/K^{\theta}_{\theta,T})$, *and corresponds to the first* T *rows and to the first* T *columns of the block matrix* $I_{K^{\theta}_{1,T}} \otimes D$. *That is,*

$$\Pi_\theta = \begin{bmatrix} D & D & \cdots \\ D & D & \\ \vdots & & \ddots \end{bmatrix}_{T \times T} \tag{3.29}$$

Proof. The ith column of $X \stackrel{\text{def}}{=} X_\theta$ is a T-vector with ones in the coordinates $k : k \in A^{\theta}_{i,T}$ and zeros everywhere else. Then we can write $X' = [x_1 \cdots x_\theta]'$ where $x_i = \sum_{k \in A^{\theta}_{i,T}} e_k, \forall i \in \{1, \ldots, \theta\}$, i.e., a summation of canonical vectors of $\mathbb{R}^T$. It follows that

$$(X'X)_{i,j} = x'_i x_j = \begin{cases} K^{\theta}_{i,T} & , \text{ if } i = j \\ 0 & , \text{ if } i \neq j \end{cases} \tag{3.30}$$

Hence $(X'X)^{-1} = D = diag(1/K^{\theta}_{1,T}, \ldots, K^{\theta}_{\theta,T})$.

Now, describe the matrix X in terms of rows as $X = [y_1, \ldots, y_T]'$ with $y_i \in \mathbb{R}^\theta, i \in \{1, \ldots, T\}$. Observe that $\Pi_\theta = XDX'$ is equivalent to

$$\Pi_\theta = \begin{bmatrix} y'_1 D y_1 & \cdots & y'_1 D y_T \\ \vdots & \ddots & \vdots \\ y'_T D y_1 & \cdots & y'_T D y_T \end{bmatrix} \tag{3.31}$$

and that $y_i = e_{w_{\theta,i}}$ with $w_{\theta,i} = i - \lfloor \frac{i-1}{\theta} \rfloor \theta, \forall i \in \{1, \ldots, T\}$. Denote $e_{w_{\theta,i}} = (e_{w_{\theta,i},1}, \ldots, e_{w_{\theta,i},\theta})'$. Then for any $1 \leq i, j \leq T$,

$$y'_i D y_j = e'_{w_{\theta,i}} D e_{w_{\theta,j}} = \sum_{k=1}^{\theta} \sum_{l=1}^{\theta} e_{w_{\theta,i},k} e_{w_{\theta,j},l} (D)_{k,l} = e_{w_{\theta,i},w_{\theta,i}} e_{w_{\theta,j},w_{\theta,j}} (D)_{w_{\theta,i},w_{\theta,j}}$$

$$= \begin{cases} (D)_{w_{\theta,i},w_{\theta,i}} & \text{, if } w_{\theta,i} = w_{\theta,j} \\ 0 & \text{, if } w_{\theta,i} \neq w_{\theta,j} \end{cases}$$

since the only nonzero coordinate of the canonical vector $e_{w_{\theta,i}}$ is the $(w_{\theta,i})$-th coordinate. However, $w_{\theta,i} = w_{\theta,j}$ is equivalent to say that $i,j \in A^{\theta}_{s,T}$ for some $s \in \{1,\ldots,\theta\}$. It determines the desired structure

$$\Pi_\theta = \left[\begin{array}{ccc|ccc|c} D_{1,1} & \cdots & 0 & D_{1,1} & \cdots & 0 & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots \\ 0 & \cdots & D_{\theta,\theta} & 0 & \cdots & D_{\theta,\theta} & \\ \hline D_{1,1} & \cdots & 0 & D_{1,1} & \cdots & 0 & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots \\ 0 & \cdots & D_{\theta,\theta} & 0 & \cdots & D_{\theta,\theta} & \\ \hline & \vdots & & & \vdots & & \ddots \end{array}\right] = \begin{bmatrix} D & D & \cdots \\ D & D & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{3.32}$$

□

Several terms in equation (3.24) of the Lemma 3.2 have in common the product $(I - \Pi_\theta)(X_{\theta_0}\beta)$. The next lemma gives a convenient form, although technical, to deal with this term afterwards.

Lemma 3.6. *The expression $(I - \Pi_\theta)(X_{\theta_0}\beta)$ can be written as the vector*

$$(\gamma_{1,T}, \ldots, \gamma_{\theta^x,T}, \gamma_{1,T}, \ldots, \gamma_{\theta^x,T}, \ldots) \in \mathbb{R}^T$$

where

$$\gamma_{s,T} = m(s) - \frac{1}{K^{\theta}_{w_{\theta,s},T}} \sum_{k=1}^{K^{\theta}_{w_{\theta,s},T}} m((k-1)\theta + w_{\theta,s}), \quad \forall s \in \{1,\ldots,\theta^x\} \tag{3.33}$$

with $T > \theta^x$ and θ^x denoting the least common multiple of θ_0 and θ. Moreover, $\gamma_{s,T}$ can be decomposed as

$$\gamma_{s,T} = \xi_s + R_{s,T} \tag{3.34}$$

with $R_{s,T} = R_{1,s,T} + R_{2,s,T}$ and

$$\xi_s = m(s) - \frac{1}{\theta_0} \sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s}) \tag{3.35}$$

$$\mathbb{R}_{1,s,T} = \left(1 - \frac{\theta_0}{K^{\theta}_{w_{\theta,s},T}} \left\lfloor \frac{K^{\theta}_{w_{\theta,s},T}}{\theta_0} \right\rfloor\right) \frac{1}{\theta_0} \sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s}) \tag{3.36}$$

$$\mathbb{R}_{2,s,T} = -\frac{1}{K^{\theta}_{w_{\theta,s},T}} \sum_{k=\theta_0 \left\lfloor K^{\theta}_{w_{\theta,s},T}/\theta_0 \right\rfloor + 1}^{K^{\theta}_{w_{\theta,s},T}} m((k-1)\theta + w_{\theta,s}). \tag{3.37}$$

Proof. From Lemma (3.5), it holds that

$$(I_T - \Pi_\theta)X_{\theta_0}\beta = \begin{bmatrix} I_\theta - D_\theta & -D_\theta & \cdots \\ -D_\theta & I_\theta - D_\theta & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} m(1) \\ \vdots \\ m(\theta_0) \\ m(1) \\ \vdots \end{bmatrix} = \begin{bmatrix} \gamma_{1,T} \\ \gamma_{2,T} \\ \gamma_{3,T} \\ \vdots \end{bmatrix}, \tag{3.38}$$

which immediately gives

$$\begin{aligned} \gamma_1 &= m(1) - \frac{1}{K^{\theta}_{1,T}} \sum_{k=1}^{K^{\theta}_{1,T}} m((k-1)\theta + 1) \\ &\vdots \\ \gamma_\theta &= m(\theta) - \frac{1}{K^{\theta}_{\theta,T}} \sum_{k=1}^{K^{\theta}_{\theta,T}} m((k-1)\theta + 1) \\ \gamma_{\theta+1} &= m(\theta+1) - \frac{1}{K^{\theta}_{1,T}} \sum_{k=1}^{K^{\theta}_{1,T}} m((k-1)\theta + 1 \\ &\vdots \end{aligned}$$

omitting the dependence of the indices of γ on T, for simplicity. Hence, for $i = 1, \ldots, T$, we have that

$$\gamma_i = m(i) + \frac{1}{K^{\theta}_{w_{\theta,i},T}} \sum_{k=1}^{K^{\theta}_{w_{\theta,i},T}} m((k-1)\theta + w_{\theta,i}). \tag{3.39}$$

Given any $i \in \{1, \ldots, \theta^x\}$, observe that

$$\begin{aligned} w_{\theta,i+k\theta^x} &= i + k\theta^x - \left\lfloor \frac{i + k\theta^x - 1}{\theta} \right\rfloor \theta = i + k\theta^x - \left\{ \left\lfloor \frac{i-1}{\theta} \right\rfloor + \frac{k\theta^x}{\theta} \right\} \theta \\ &= i + k\theta^x - \left\lfloor \frac{i-1}{\theta} \right\rfloor \theta - k\theta^x = i - \left\lfloor \frac{i-1}{\theta} \right\rfloor \theta \\ &= w_{\theta,i}, \quad k = 1, \ldots, \left\lfloor \frac{T-i}{\theta^x} \right\rfloor + 1. \end{aligned} \tag{3.40}$$

By (3.39) and (3.40), for all $i \in \{1, \ldots, \theta^x\}$ and all $k \in \{1, \ldots, \lfloor (T-i)/\theta^x \rfloor + 1\}$,

$$\gamma_{i+k\theta^x} = \gamma_i - m(i) + m(i + k\theta^x) = \gamma_i,$$

since θ^x is a multiple of θ_0 and m has period θ_0 by definition. This gives us the desired formula, i.e., $(I-\Pi_\theta)(X_{\theta_0}\beta) = (\gamma_{1,T},\dots,\gamma_{\theta^x,T},\gamma_{1,T},\dots,\gamma_{\theta^x,T},\dots)$.

To decompose $\gamma_{s,t}$, we first consider at the summation

$$\sum_{k=1}^{K^\theta_{w_{\theta,s},T}} m((k-1)\theta + w_{\theta,s}), \quad s\in\{1,\dots,\theta^x\}. \tag{3.41}$$

Note that we are evaluating the θ_0-periodic sequence m at points that are multiples of θ, in (3.41). Therefore, we must have

$$\begin{aligned} m(w_{\theta,s})+\cdots+m(w_{\theta,s}+(\theta_0-1)\theta) &= m(w_{\theta,s}+\theta_0\theta)+\cdots+m(w_{\theta,s}+(2\theta_0-1)\theta)\\ &= m(w_{\theta,s}+2\theta_0\theta)+\cdots+m(w_{\theta,s}+(3\theta_0-1)\theta)\\ &= \cdots\\ &= m(w_{\theta,s}+(k-1)\theta\theta_0)+\cdots+m(w_{\theta,s}+(k\theta_0-1)\theta), \end{aligned} \tag{3.42}$$

for any $k\in\{1,\dots,k_{\max}\}$. Without loss of generality, let $w_{\theta,s}=s\in\{1,\dots,\theta\}$. Define $k_{\max}=\max\{k\in\mathbb{N} : s+(k\theta_0-1)\theta\in A^\theta_{s,T}\}$. Since the greatest element of $A^\theta_{s,T}$ is $s+(K^\theta_{s,T}-1)\theta$, we clearly have that $k_{\max}\theta_0$ is bounded by $K^\theta_{s,T}$, and thus $k_{\max}\le\lfloor K^\theta_{s,T}/\theta_0\rfloor$. On the other hand, $\lfloor K^\theta_{s,T}/\theta_0\rfloor\theta_0\le K^\theta_{s,T}$, which implies $\lfloor K^\theta_{s,T}/\theta_0\rfloor\in\{k\in\mathbb{N} : s+(k\theta_0-1)\theta\in A^\theta_{s,T}\}$. Hence the equality[1], $k_{\max}=\lfloor K^\theta_{s,T}/\theta_0\rfloor$.

From the above observations, we can split (3.41) as a sum of $(\theta_0\theta)$-periodic points, given by (3.42), plus a remainder:

$$\begin{aligned} &\sum_{k=1}^{K^\theta_{w_{\theta,s},T}} m((k-1)\theta+w_{\theta,s})\\ &= \sum_{k=1}^{\lfloor (K^\theta_{w_{\theta,s},T})/\theta_0\rfloor\theta_0} m((k-1)\theta+w_{\theta,s}) + \sum_{k=\lfloor (K^\theta_{w_{\theta,s},T})/\theta_0\rfloor\theta_0+1}^{K^\theta_{w_{\theta,s},T}} m((k-1)\theta+w_{\theta,s})\\ &= \left\lfloor \frac{K^\theta_{w_{\theta,s},T}}{\theta_0}\right\rfloor \sum_{k=1}^{\theta_0} m((k-1)\theta+w_{\theta,s}) + \sum_{k=\lfloor (K^\theta_{w_{\theta,s},T})/\theta_0\rfloor\theta_0+1}^{K^\theta_{w_{\theta,s},T}} m((k-1)\theta+w_{\theta,s}). \end{aligned} \tag{3.43}$$

[1] To gain insight into this result, observe that counting points of the form $s+(k\theta_0-1)\theta$ in $A^\theta_{s,T}$ is equivalent to count points of the form $\theta_0+(k-1)\theta_0$ in the enumeration $E^\theta_{s,T}=\{1,2,\dots,K^\theta_{s,T}\}$. Indeed, the set $\{s+(k\theta_0-1)\theta\}_{k\in\mathbb{N}}\cap A^\theta_{s,T}$ is consituted by the θ_0th, $2\theta_0$th, … points of $A^\theta_{s,T}$. So we rely on the problem of counting the multiples of θ_0 starting at θ_0 in $E^\theta_{s,T}$. Lemma (3.3) tells that the number $k_{\max}$ is exactly $K^{\theta_0}_{\theta_0,K^\theta_{s,T}}=\lfloor (K^\theta_{s,T}-\theta_0)/\theta_0\rfloor+1=\lfloor (K^\theta_{s,T})/\theta_0\rfloor$.

From equations (3.33) and (3.43),

$$\begin{aligned}
\gamma_{s,T} &= m(s) - \frac{1}{K^{\theta}_{w_{\theta,s},T}} \sum_{k=1}^{K^{\theta}_{w_{\theta,s},T}} m((k-1)\theta + w_{\theta,s}) \ (\pm) \ \frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s}) \\
&= m(s) - \frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s}) \\
&+ \frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s}) - \frac{1}{K^{\theta}_{w_{\theta,s},T}} \left\lfloor \frac{K^{\theta}_{w_{\theta,s},T}}{\theta_0} \right\rfloor \sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s}) \\
&- \frac{1}{K^{\theta}_{w_{\theta,s},T}} \sum_{k=\lfloor (K^{\theta}_{w_{\theta,s},T})/\theta_0 \rfloor \theta_0 + 1}^{K^{\theta}_{w_{\theta,s},T}} m((k-1)\theta + w_{\theta,s}) \\
&:= \xi_s + R_{1,s,T} + R_{2,s,T}
\end{aligned}$$

for any $s \in \{1, \ldots, \theta^x\}$, where $(\pm)$ stands for "*plus and minus*". □

From equation (3.33), we immediately see that if $\theta = \theta_0$, then $\gamma_{s,T} = 0$. When $\theta \neq \theta_0$, distinguish between two cases:

(A) $\theta \neq \theta_0$ and θ is not a multiple of θ_0.

(B) $\theta \neq \theta_0$ and θ is a multiple of θ_0.

Lemma 3.7. *The decomposition $\gamma_{s,T} = \xi_s + R_{s,T}$, $s = 1, \ldots, \theta^x$, has the following properties:*

(i) $|R_{s,T}| \leq \frac{C\theta_0}{K^{\theta}_{w_{\theta,s},T}}$, *where C is a positive constant;*

(ii) if case B holds, then $\xi_s = 0, \forall s \in [\theta^x]$;

(iii) if case A holds, then $\exists s \in [\theta^x] : \xi_s \neq 0$; moreover, uniformly on the set $\{s \in [\theta^x] : \xi_s \neq 0\}$, $\exists \eta > 0$ such that $|\xi_s| \geq \eta$.

Proof. (i) Without loss of generality, assume $w_{\theta,s} = s$. Let $s \in \{1, \ldots, \theta\}$. Since for any $x \in \mathbb{R}, n \in \mathbb{Z}$, it holds that $\lfloor x \rfloor = n \iff n \leq x < n+1$, we have

$$\left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} - \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor \right\rfloor = \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor + \left\lfloor -\left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor \right\rfloor = \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor - \left\lceil \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor \right\rceil = 0$$

$$\iff 0 \leq \frac{K^{\theta}_{s,T}}{\theta_0} - \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor < 1 \tag{3.44}$$

$$\iff 0 \leq 1 - \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor \frac{\theta_0}{K^{\theta}_{s,T}} < \frac{\theta_0}{K^{\theta}_{s,T}} \tag{3.45}$$

Then, by Lemma 3.6 and (3.45),

$$|R_{1,s,T}| = \left|\frac{1}{\theta_0}\right| \left| 1 - \frac{\theta_0}{K^{\theta}_{s,T}} \left\lfloor \frac{K^{\theta}_{s,T}}{\theta_0} \right\rfloor \right| \left| \sum_{k=1}^{\theta_0} m((k-1)\theta + s) \right|$$

$$< \frac{1}{\theta_0}\frac{\theta_0}{K^{\theta}_{s,T}}\left|\sum_{k=1}^{\theta_0} m((k-1)\theta+s)\right|$$
$$\leq C\frac{\theta_0}{K^{\theta}_{s,T}} \tag{3.46}$$

with $C=\sup_{t\in\{1,\ldots,\theta_0\}}|m(t)|$. Next, by (3.44), it follows that

$$0 \leq \theta_0\left(\frac{K^{\theta}_{s,T}}{\theta_0}-\left\lfloor\frac{K^{\theta}_{s,T}}{\theta_0}\right\rfloor\right) < \theta_0. \tag{3.47}$$

Thus, by Lemma 3.6 and (3.47),

$$|\mathbb{R}_{2,s,T}| = \left|-\frac{1}{K^{\theta}_{w_{\theta,s},T}}\right|\left|\sum_{k=\theta_0\lfloor K^{\theta}_{w_{\theta,s},T}/\theta_0\rfloor+1}^{K^{\theta}_{w_{\theta,s},T}} m((k-1)\theta+w_{\theta,s})\right|$$
$$\leq \frac{C}{K^{\theta}_{s,T}}\left(K^{\theta}_{s,T}-\theta_0\left\lfloor\frac{K^{\theta}_{s,T}}{\theta_0}\right\rfloor\right)$$
$$< C\frac{\theta_0}{K^{\theta}_{s,T}} \tag{3.48}$$

By combining (3.46) and (3.48), we obtain $|R_{s,T}| \leq C\frac{\theta_0}{K^{\theta}_{s,T}}$.

(ii) Suppose that case B holds, i.e., $\theta = l\theta_0$ for some natural number $l > 1$. Since θ_0 is the period of m, we obtain that

$$\xi_s = m(s) - \frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)l\theta_0+w_{l\theta_0,s}) = m(s) - \frac{\theta_0}{\theta_0}m(w_{l\theta_0,s})$$
$$= m\left(\left(s-\left\lfloor\frac{s-1}{l\theta_0}\right\rfloor l\theta_0\right)+\left\lfloor\frac{s-1}{l\theta_0}\right\rfloor l\theta_0\right) - m\left(s-\left\lfloor\frac{s-1}{l\theta_0}\right\rfloor l\theta_0\right) = 0, \quad \forall s\in\{1,\ldots,\theta^x\}.$$

(iii) Suppose that case A holds and that there is some θ such that $\xi_s = 0$, $\forall s \in \{1,\ldots,\theta^x\}$. But, for any $r\in\mathbb{N}$, it holds that $m(s) = m(s+r\theta^x)$ and

$$w_{\theta,s+r\theta^x} = (s+r\theta^x) - \left\lfloor\frac{(s+r\theta^x)-1}{\theta}\right\rfloor\theta = s - \left\lfloor\frac{s-1}{\theta}\right\rfloor\theta = w_{\theta,s}.$$

Thus, formula (3.35) implies that θ satisfies $\xi_s = 0$, $\forall s\in\mathbb{N}$. Since $s+r\theta\in\mathbb{N}$ and $w_{\theta,s+r\theta} = w_{\theta,s}$ also holds, for all $s,r\in\mathbb{N}$, then $\xi_{s+r\theta}=0$ and

$$\frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta+w_{\theta,s}) = \frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta+w_{\theta,s+r\theta}), \quad s,r\in\mathbb{N}, \tag{3.49}$$

respectively. Hence $m(s) = m(s+r\theta)$, $\forall s,r\in\mathbb{N}$, which implies that m has period θ. As θ_0 is the smallest period of m, we cannot have $\theta < \theta_0$. Then assume $\theta > \theta_0$. Note that

$\theta/\theta_0 - 1 < \lfloor\theta/\theta_0\rfloor < \theta/\theta_0$ and hence $\lfloor\theta/\theta_0\rfloor\theta_0$ is in the interval $(\theta-\theta_0, \theta)$ which contains exactly $\theta_0 - 1$ integers. It means that $\exists k \in \{1,\ldots,\theta_0-1\} : \theta = \lfloor\theta/\theta_0\rfloor\theta_0 + k$, which in turn implies

$$m(s) = m(s+\theta) = m(s+\lfloor\theta/\theta_0\rfloor\theta_0 + k) = m(s+k),\ \forall s \in \mathbb{N},$$

contradicting the fact that $k < \theta_0$ cannot be a period of m. Hence, for every θ under case A, $\exists s \in \{1,\ldots,\theta^x\} : \xi_s \neq 0$.

To prove the next result, let $s \in \{1,\ldots,\theta^x\}$ be such that $\xi_s \neq 0$. Observe that $\theta_0^{-1}\sum_{k=1}^{\theta_0} m((k-1)\theta + w_{\theta,s})$ is the average of θ_0 points of the sequence $\{m(t)\}$. Since the range of $\{m(t)\}$ has at most θ_0 distinct points, the number of possible values that the average can take is at most $\binom{2\theta_0-1}{\theta_0}$ (i.e., the combination of θ_0 values taken θ_0 at a time with repetition). As a consequence, ξ_s can also take only a finite number of values. Denote the finite set of possible values of ξ_s by B_s. Define $B_s^* = B_s^+ \cup B_s^-$, where $B_s^+ = \{x \in B_s : x > 0\}$ and $B_s^- = \{x \in B_s : x < 0\}$. Since B_s^+, B_s^- **are finite sets**, there are $M_1 = \min(B_s^+)$ and $M_2 = \max(B_s^-)$. Set $\eta_s = \min(M_1, -M_2) > 0$. Then $B_s^* \cap \mathcal{B}_{\eta_s}(0) = \emptyset$, where $\mathcal{B}_{\eta_s}(0) = \{x \in \mathbb{R} : |x| < \eta_s\}$ is the open ball centered at zero with radius η_s. It implies that $B_s^* \subseteq [\mathcal{B}_{\eta_s}(0)]^c$. Take $\eta = \min_{s\in\{1,\ldots,\theta^x\}:\xi_s\neq 0}\eta_s$ to obtain $B_s^* \subseteq [\mathcal{B}_{\eta_s}(0)]^c \subseteq [\mathcal{B}_{\eta}(0)]^c$, $\forall s \in \{1,\ldots,\theta^x\} : \xi_s \neq 0$. In words, all possible nonzero values of ξ_s satisfies $|\xi_s| \geq \eta$ uniformly. □

Now we are in position to characterize some asymptotic properties of the terms involved in equation (3.24).

Lemma 3.8. *Let $\{v_t\}$ be an arbitrary divergent sequence of positive numbers. Let $h = O(T^{-1/4})$ and $\Theta_T \leq CT^{2/5-\omega}$ for some small $\omega > 0$. Assume that Conditions 1-4 are fulfilled . Moreover, consider $n \equiv n(\theta) = \#\mathcal{S}$, where $\mathcal{S}$ is the subset of indices $s \in \{1,\ldots,\theta^x\}$ for which $\xi_s \neq 0$. Then there are a sufficiently small constant $c > 0$ and $T_0 \in \mathbb{N}$ such that for all $T > T_0$,*

$$\text{(in case A)}:\quad B_\theta \geq c\frac{nT}{\theta}\quad,\quad P\left(|S_\theta^\epsilon| > v_T\frac{n\sqrt{T}}{\theta}\right) \leq \frac{C}{v_T^2},\quad P\left(|S_\theta^{g_\epsilon}| > v_T\frac{n\sqrt{T/h}}{\theta}\right) \leq \frac{C}{v_T^2}$$

$$|S_\theta^{g_m}| \leq C\frac{n\sqrt{T}}{\theta},\quad |S_\theta^{g_b}| \leq C\frac{n\sqrt{T}}{\theta};$$

$$\text{(in case B)}:\quad B_\theta = 0,\qquad S_\theta^\epsilon = 0,\quad S_\theta^{g_\epsilon} = 0,\quad S_\theta^{g_m} = 0,\quad S_\theta^{g_b} = 0.$$

Moreover, in both cases A and B,

$$P\left(|U_\theta^{(g_m,g_\epsilon)}| > v_T/\sqrt{h}\right) \leq \frac{C}{v_T^2},\qquad P\left(|U_\theta^{(g_b,g_\epsilon)}| > v_T/\sqrt{h}\right) \leq \frac{C}{v_T^2},\quad |W_\theta^{g_m}| \leq C,$$

$$P\left(|U_\theta^{(g_m,\epsilon)}| > v_T/\sqrt{h}\right) \leq \frac{C}{v_T^2},\qquad P\left(|U_\theta^{(g_b,\epsilon)}| > v_T/\sqrt{h}\right) \leq \frac{C}{v_T^2},\quad |W_\theta^{g_b}| \leq C,$$

$$|U_\theta^{(g_b,g_m)}| \le C.$$

Proof. We begin with θ satisfying case A.

Define $S^c = \{1,\dots,\theta^x\} \setminus S$. Lemma 3.7 (i) and Lemma 3.4 imply that $|R_{s,T}| \le C\theta_0/\lfloor T/\theta \rfloor \le C\Theta_T/T$, $\forall s \in \{1,\dots,\theta^x\}$. It follows from Lemmas 3.1, 3.6, 3.7 and the triangle inequality for subtraction that

$$\begin{aligned}
B_\theta &= (X_{\theta_0}\beta)'(I-\Pi_\theta)(X_{\theta_0}\beta) = (X_{\theta_0}\beta)'(I-\Pi_\theta)'(I-\Pi_\theta)(X_{\theta_0}\beta)\\
&= (\gamma_{1,T},\dots,\gamma_{\theta^x,T},\dots)(\gamma_{1,T},\dots,\gamma_{\theta^x,T},\dots)'\\
&= \left\lfloor \frac{T}{\theta^x} \right\rfloor \sum_{k=1}^{\theta^x} \gamma_{k,T}^2 + \sum_{k=\theta^x\lfloor T/\theta^x\rfloor+1}^{T} \gamma_{k,T}^2 \ge \left\lfloor \frac{T}{\theta^x} \right\rfloor \left\{ \sum_{k\in S} \gamma_{k,T}^2 + \sum_{k\in S^c} \gamma_{k,T}^2 \right\}\\
&\ge \left\lfloor \frac{T}{\theta^x} \right\rfloor \sum_{k\in S} |\xi_k + R_{k,T}|^2 = \left\lfloor \frac{T}{\theta^x} \right\rfloor \sum_{k\in S} |\xi_k - (-R_{k,T})|^2 \ge \left\lfloor \frac{T}{\theta^x} \right\rfloor \sum_{k\in S} (|\xi_k| - |R_{k,T}|)^2\\
&= \left\lfloor \frac{T}{\theta^x} \right\rfloor \sum_{k\in S} (|\xi_k|^2 + |R_{k,T}|^2 - 2|\xi_k R_{k,T}|) \ge \left\lfloor \frac{T}{\theta^x} \right\rfloor \sum_{k\in S} |\xi_k|(|\xi_k| - 2|R_{k,T}|)\\
&\ge c_1 \frac{T}{\theta} \sum_{k\in S} [\eta(\eta - 2|R_{k,T}|)] \ge c_1 \frac{T}{\theta} \sum_{k\in S} [\eta c_2]\\
&= c_1 c_2 \eta \frac{nT}{\theta} := c\frac{nT}{\theta}, \quad \forall T \ge T_0,
\end{aligned}$$

for some $T_0 > 0$ and some sufficiently small constants[2] $c_1, c_2 > 0$. The fact that $|R_{k,T}| \to 0$ implies the existence of such positive constant c_2 for all sufficiently large T.

Next, write

$$S_\theta^\epsilon = \sum_{t=1}^{T} \gamma_{t,T}\epsilon_{t,T} = \sum_{t\in I_S} \gamma_{t,T}\epsilon_{t,T} + \sum_{t\in I_{S^c}} \gamma_{t,T}\epsilon_{t,T},$$

where $I_S = \{t : w_{\theta^x,t} \in S\}$ and $I_{S^c} = \{t : w_{\theta^x,t} \in S^c\}$. Using Lemma 3.4, the cardinalities of I_S and I_{S^c} satisfy

$$\#I_S \le n\left(\left\lfloor \frac{T}{\theta^x} \right\rfloor + 1\right) \le n\left(\frac{T}{\theta^x} + 1\right) \le n\left(\frac{T}{\theta} + 1\right) = n\frac{T+\theta}{\theta} \le \frac{2nT}{\theta}; \tag{3.50}$$

$$\#I_{S^c} \le (\theta^x - n)\left(\left\lfloor \frac{T}{\theta^x} \right\rfloor + 1\right) \le \theta^x\left(\frac{T}{\theta^x} + 1\right) \le \theta_0(T+\theta) \le 2\theta_0 T, \tag{3.51}$$

[2] As $\lfloor T/\theta^x \rfloor > T/\theta^x - 1 \ge T/(\theta\theta_0) - 1$ and the convergent sequence

$$\frac{T/(\theta\theta_0) - 1}{T/\theta} = \frac{1}{\theta_0}\left(1 - \frac{\theta}{T}\right) \overset{T\to\infty}{\longrightarrow} \frac{1}{\theta_0}$$

is monotone increasing and strict positive by the assumption $\Theta_T < T$, we can take $c_1 \in (0, 1/\theta_0)$ so that $\lfloor T/\theta^x \rfloor \ge c_1 T/\theta$ holds for all T large enough.

for every T. By Lemma 2.11 of Chapter 2,

$$P\left(|S_\theta^\epsilon| > v_T\sqrt{\frac{nT}{\theta}}\right) \le P\left(\left|\sum_{t\in I_S}\gamma_{t,T}\epsilon_{t,T}\right| > \frac{v_T}{2}\sqrt{\frac{nT}{\theta}}\right) + P\left(\left|\sum_{t\in I_{S^c}}\gamma_{t,T}\epsilon_{t,T}\right| > \frac{v_T}{2}\sqrt{\frac{nT}{\theta}}\right). \tag{3.52}$$

From Lemmas 3.6 and 3.7 (iii), it follows that $|\gamma_{t,T}| \le |\xi_t| + |R_{t,T}| \le \max_{i\in 1,\dots,\theta_0}|m(i)| + C\theta_0/K^\theta_{w_{\theta,t}} \le C$, $\forall t \in \{1,\dots,T\}$. Furthermore, it holds that $E\left[\sum_{t=1}^T \gamma_{t,T}\epsilon_{t,T}\right] = 0$, by assumption. From Chebychev's and Davydov's (Corollary 2.5.1) inequalities, Conditions 1-2 and (3.50),

$$\begin{aligned}
P\left(\left|\sum_{t\in I_S}\gamma_{t,T}\epsilon_{t,T}\right| > \frac{v_T}{2}\sqrt{\frac{nT}{\theta}}\right) &\le \mathrm{Var}\left(\sum_{t\in I_S}\gamma_{t,T}\epsilon_{t,T}\right)\frac{4\theta}{v_T^2 nT} = E\left[\left(\sum_{t\in I_S}\gamma_{t,T}\epsilon_{t,T}\right)^2\right]\frac{4\theta}{v_T^2 nT} \\
&= \frac{4\theta}{v_T^2 nT}E\left[\sum_{t\in I_S}\sum_{l\in I_S}\gamma_{t,T}\ \gamma_{l,T}\ \epsilon_{t,T}\ \epsilon_{l,T}\right] \\
&\le \frac{4\theta}{v_T^2 nT}\sum_{t,l\in I_S}|\gamma_{t,T}|\ |\gamma_{l,T}|\ \left|\mathrm{Cov}(\epsilon_{t,T},\epsilon_{l,T})\right| \\
&\le \frac{C\theta}{v_T^2 nT}\sum_{t,l\in I_S}\left|\mathrm{Cov}(\epsilon_{t,T},\epsilon_{l,T})\right| \\
&\le \frac{C\theta}{v_T^2 nT}\sum_{t,l\in I_S}\alpha(|t-l|)^{(2+\delta)/(4+\delta)}E(|\epsilon_{t,T}|)^{4+\delta}E(|\epsilon_{l,T}|)^{4+\delta} \\
&\le \frac{C\theta}{v_T^2 nT}\sum_{t,l\in I_S}(a^{|t-l|})^{1-2/(4+\delta)} \le \frac{C\theta}{v_T^2 nT}\sum_{t\in I_S}\sum_{l=1}^T a^{|t-l|} \\
&\le \frac{C\theta}{v_T^2 nT}\sum_{t\in I_S}\underbrace{\sum_{w=0}^\infty 2a^w}_{\le C} \le \frac{C\theta}{v_T^2 nT}\frac{2nT}{\theta} \le \frac{C}{v_T^2}.
\end{aligned} \tag{3.53}$$

On the other hand, Lemma 3.7(i) implies that $|\gamma_{i,T}| = |R_{i,T}| \le C\Theta_T/T, \forall i \in S^c$, which in turn gives

$$\begin{aligned}
\mathrm{Var}\left(\sum_{t\in I_{S^c}}\gamma_{t,T}\epsilon_{t,T}\right) = E\left[\left(\sum_{t\in I_{S^c}}\gamma_{t,T}\epsilon_{t,T}\right)^2\right] &\le \sum_{t\in I_{S^c}}\sum_{l\in I_{S^c}}|\gamma_{t,T}||\gamma_{l,T}||\mathrm{Cov}(\epsilon_{t,T},\epsilon_{l,T})| \\
&\le C\left(\frac{\Theta_T}{T}\right)^2 2\theta_0 T\sum_{w=0}^\infty 2a^w \le C\frac{\Theta_T^2}{T},
\end{aligned}$$

using (3.51). Then, by Chebychev's inequality,

$$P\left(\left|\sum_{t\in I_{S^c}}\gamma_{t,T}\epsilon_{t,T}\right| > \frac{v_T}{2}\sqrt{\frac{nT}{\theta}}\right) \le C\frac{\Theta_T^2}{T}\frac{\theta}{v_T^2 nT} \le C\frac{\Theta_T^3}{v_T^2T^2} \le C\frac{T^{3(2/5-\omega)-2}}{v_T^2}$$
$$\le C\frac{T^{-4/5}}{v_T^2} \le \frac{C}{v_T^2}. \tag{3.54}$$

By combining the inequalities (3.52)-(3.54), we have that

$$P\left(\left|S_\theta^\epsilon\right| > \frac{v_T}{2}\sqrt{\frac{nT}{\theta}}\right) \le \frac{C}{v_T^2}, \tag{3.55}$$

for any T.

Similarly,

$$S_\theta^{g_\epsilon} = \sum_{t\in I_S}\gamma_{t,T}\sum_{i=1}^T W_{i,T}(t/T)\epsilon_{i,T} + \sum_{t\in I_{S^c}}\gamma_{t,T}\sum_{i=1}^T W_{i,T}(t/T)\epsilon_{i,T} := A_1^{g_\epsilon} + A_2^{g_\epsilon}.$$

Using the definition of J_x, given by (2.33) and (2.34), Lemma 2.8 of Chapter 2 and the assumption that K has compact support,

$$\begin{aligned}\mathrm{Var}(A_1^{g_\epsilon}) &= E\left\{\sum_{t,l\in I_S}\sum_{i,j=1}^T \gamma_{t,T}\gamma_{l,T}W_{i,T}(t/T)\epsilon_{i,T}W_j(l/T)\epsilon_{j,T}I(i\in J_{t/T})I(j\in J_{l/T})\right\}\\
&\le \sum_{t,l\in I_S}\sum_{(i,j)\in J_{t/T}\times J_{l/T}} |\gamma_{t,T}\gamma_{l,T}||\mathrm{Cov}(\epsilon_{i,T},\epsilon_{j,T})| \sup_{i\in[T]}\sup_{x\in[0,1]}|W_{i,T}(x)| \sup_{j\in[T]}\sup_{x\in[0,1]}|W_{j,T}(x)|\\
&\le \frac{C}{(Th)^2}\sum_{t,l\in I_S}\sum_{(i,j)\in J_{t/T}\times J_{l/T}}|\mathrm{Cov}(\epsilon_{i,T},\epsilon_{j,T})| \le \frac{C}{(Th)^2}\sum_{t,l\in I_S}\sum_{i\in J_{t/T}} 2\sum_{j=0}^\infty a^j\\
&\le \frac{C}{(Th)^2}\left(\frac{2nT}{\theta}\right)^2 k_T \le C\left(\frac{n}{\theta}\right)^2\frac{T}{h},\end{aligned} \tag{3.56}$$

for all T sufficiently large, where $k_T = \#J_x = O(Th)$ by Lemma 2.2 of Chapter 2. Since $|\gamma_{i,T}| \le C\Theta_T/T = o(1)$, $\forall i\in S^c$, $\mathrm{Var}(A_2^{g_\epsilon})$ is dominated by $\mathrm{Var}(A_1^{g_\epsilon})$. Therefore, for T sufficiently large,

$$P\left(|A_i^{g_\epsilon}| > \frac{v_T}{2}\frac{n\sqrt{T/h}}{\theta}\right) \le C\frac{T}{h}\left(\frac{n}{\theta}\right)^2\left(\frac{2\theta}{v_Tn\sqrt{T/h}}\right)^2 = \frac{4C}{v_T^2}, \quad i\in\{1,2\}$$

and thus,

$$P\left(\left|S_\theta^{g_\epsilon}\right| > v_T\frac{n\sqrt{T/h}}{\theta}\right) \le \frac{C}{v_T^2}. \tag{3.57}$$

From the proofs of Theorems 3.1 and 2.4, $|\sum_{i=1}^T W_{i,T}(x)m(i)| \le C/T$ and $|\sum_{i=1}^T W_{i,T}(x)[g(x)-$

$g(i/T)]| \leq Ch^2$ hold uniformly over $x \in [0,1]$ for T large enough, respectively. Then

$$\begin{aligned}
|S_\theta^{g_b}| &\leq \sum_{t=1}^T |\gamma_{t,T}| \left| \sum_{i=1}^T W_{i,T}(t/T)[g(t/T) - g(i/T)] \right| \leq Ch^2 \sum_{t=1}^T |\gamma_{t,T}| \\
&\leq Ch^2 \Bigg(\sum_{t \in I_S} |\gamma_{t,T}| + \sum_{t \in I_{S^c}} |\gamma_{t,T}| \Bigg) \\
&\leq Ch^2 \left(\frac{2nT}{\theta} + 2\theta_0 T \frac{\Theta_T}{T} \right) = 2Ch^2 \Bigg(\frac{nT}{\theta} + \underbrace{\theta_0 \Theta_T}_{=o(T)} \Bigg) \\
&\leq C \frac{n}{\theta} T h^2 \leq C \frac{n\sqrt{T}}{\theta} \qquad (3.58)
\end{aligned}$$

for all sufficiently large T, using the hypothesis that $h = O(T^{-1/4})$. By Condition 5, $T^{-1} = o(h^2)$, implying that the term $|S_\theta^{g_m}|$ is dominated by $|S_\theta^{g_b}|$. Hence $|S_\theta^{g_m}| \leq Cn\sqrt{T}/\theta$, also holds for T large enough.

Now, let θ satisfy case B. Using similar arguments as for Lemma 3.7(ii)'s proof ,

$$\begin{aligned}
\gamma_{s,T} &= m(s) - \frac{1}{K^\theta_{w_{\theta,s},T}} \sum_{k=1}^{K^\theta_{w_{\theta,s},T}} m((k-1)\theta + w_{\theta,s}) \\
&= m(s) - \frac{1}{K^\theta_{w_{l\theta_0,s},T}} \sum_{k=1}^{K^{l\theta_0}_{w_{l\theta_0,s},T}} m((k-1)l\theta_0 + w_{l\theta_0,s}) \\
&= m(s) + m(w_{l\theta_0,s}) \\
&= m\left(w_{l\theta_0,s} + \left\lfloor \frac{s-1}{l\theta_0} \right\rfloor l\theta_0 \right) + m(w_{l\theta_0,s}) = 0, \quad \forall s \in \{1, \ldots, \theta^x\},
\end{aligned}$$

for some $1 < l \in \mathbb{N}$. Then, by Lemma 3.6, $(I - \Pi_\theta) X_{\theta_0} \beta$ is the zero T-vector. We thus have $B_\theta = S_\theta^\epsilon = S_\theta^{g_\epsilon} = S_\theta^{g_m} = S_\theta^{g_b} = 0$.

It remains to bound the terms $W_\theta^{g_b}, W_\theta^{g_m}, U_\theta^{(g_b,g_m)}, U_\theta^{(g_b,g_\epsilon)}, U_\theta^{(g_b,\epsilon)}, U_\theta^{(g_m,g_\epsilon)}$ and $U_\theta^{(g_m,\epsilon)}$, which do not have $(I - \Pi_\theta) X_{\theta_0} \beta$ in their formulas. We start with the non-stochastic terms $W_\theta^{g_b}, W_\theta^{g_m}, U_\theta^{(g_b,g_m)}$. From Lemma 3.5,

$$\begin{aligned}
(\Pi_\theta - \Pi_{\theta_0}) g_b &= \left(\begin{bmatrix} D_\theta & D_\theta & \cdots \\ D_\theta & D_\theta & \\ \vdots & & \ddots \end{bmatrix} - \begin{bmatrix} D_{\theta_0} & D_{\theta_0} & \cdots \\ D_{\theta_0} & D_{\theta_0} & \\ \vdots & & \ddots \end{bmatrix} \right) g_b \\
&= \begin{bmatrix} \frac{1}{K^\theta_{1,T}} \sum_{k=1}^{K^\theta_{1,T}} g_b\left(\frac{1+(k-1)\theta}{T} \right) \\ \vdots \\ \frac{1}{K^\theta_{\theta,T}} \sum_{k=1}^{K^\theta_{\theta,T}} g_b\left(\frac{\theta+(k-1)\theta}{T} \right) \\ \vdots \end{bmatrix} - \begin{bmatrix} \frac{1}{K^{\theta_0}_{1,T}} \sum_{k=1}^{K^{\theta_0}_{1,T}} g_b\left(\frac{1+(k-1)\theta_0}{T} \right) \\ \vdots \\ \frac{1}{K^{\theta_0}_{\theta_0,T}} \sum_{k=1}^{K^{\theta_0}_{\theta_0,T}} g_b\left(\frac{\theta_0+(k-1)\theta_0}{T} \right) \\ \vdots \end{bmatrix}
\end{aligned}$$

where $D_\theta = diag(1/K^\theta_{1,T}, \ldots, 1/K^\theta_{\theta,T})$ and $D_{\theta_0} = diag(1/K^{\theta_0}_{1,T}, \ldots, 1/K^{\theta_0}_{\theta_0,T})$. Hence,

$$\begin{aligned}|g_b'(\Pi_\theta - \Pi_{\theta_0})g_b| &\leq \left|\sum_{l=1}^{T} g_b\Big(\frac{l}{T}\Big)\Bigg[\frac{1}{K^\theta_{w_{\theta,l},T}}\sum_{k=1}^{K^\theta_{w_{\theta,l},T}} g_b\Big(\frac{w_{\theta,l}+(k-1)\theta}{T}\Big)\Bigg]\right| \\ &+ \left|\sum_{l=1}^{T} g_b\Big(\frac{l}{T}\Big)\Bigg[\frac{1}{K^{\theta_0}_{w_{\theta_0,l},T}}\sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,l},T}} g_b\Big(\frac{w_{\theta_0,l}+(k-1)\theta_0}{T}\Big)\Bigg]\right| \\ &:= D_1^{g_b} + D_2^{g_b}. \end{aligned} \tag{3.59}$$

Since $\sup_{x\in[0,1]}|g_b(x)| = O(h^2)$ and $h = O(T^{-1/4})$, it holds that

$$D_1^{g_b} \leq C\sum_{l=1}^{T} h^2\Bigg[\frac{1}{K^\theta_{w_{\theta,l},T}}\sum_{k=1}^{K^\theta_{w_{\theta,l},T}} h^2\Bigg] \leq CTh^4 = O(1).$$

Clearly, we also have $D_2^{g_b} = O(1)$. Therefore, for T sufficiently large

$$|W_\theta^{g_b}| \leq C. \tag{3.60}$$

Since $\sup_{x\in[0,1]}|g_m(x)| = O(1/T)$ and so, dominated by $\sup_{x\in[0,1]}|g_b(x)|$, we have that the terms $|W_\theta^{g_m}|$ and $|U_\theta^{(g_b,g_m)}|$ are also dominated by $|W_\theta^{g_b}|$. Then, for T sufficiently large

$$|W_\theta^{g_m}| \leq C, \tag{3.61}$$

$$|U_\theta^{(g_b,g_m)}| \leq C. \tag{3.62}$$

We finally turn to the stochastic terms $U_\theta^{(g_b,g_\epsilon)}, U_\theta^{(g_b,\epsilon)}, U_\theta^{(g_m,g_\epsilon)}$ and $U_\theta^{(g_m,\epsilon)}$. Note that for arbitrary $x, y \in \mathbb{R}^T$, $x'By = y'Bx$. Then write

$$\begin{aligned}U_\theta^{(g_b,g_\epsilon)} &= \sum_{l=1}^{T} g_\epsilon(l/T)\Bigg[\frac{1}{K^\theta_{w_{\theta,l},T}}\sum_{k=1}^{K^\theta_{w_{\theta,l},T}} g_b\Big(\frac{w_{\theta,l}+(k-1)\theta}{T}\Big)\Bigg] \\ &- \sum_{l=1}^{T} g_\epsilon(l/T)\Bigg[\frac{1}{K^{\theta_0}_{w_{\theta_0,l},T}}\sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,l},T}} g_b\Big(\frac{w_{\theta_0,l}+(k-1)\theta_0}{T}\Big)\Bigg] \\ &:= G_1^{(g_b,g_\epsilon)} - G_2^{(g_b,g_\epsilon)}. \end{aligned} \tag{3.63}$$

Then, from Davydov's inequality,

$$\mathrm{Var}\big(G_1^{(g_b,g_\epsilon)}\big) = E\Bigg\{\sum_{l,t=1}^{T} g_\epsilon(l/T)g_\epsilon(t/T)\frac{1}{K^\theta_{w_{\theta,l},T}K^\theta_{w_{\theta,t},T}}$$

$$\times \Bigg[\sum_{k=1}^{K^\theta_{w_{\theta,l},T}} g_b\bigg(\frac{w_{\theta,l}+(k-1)\theta}{T}\bigg)\Bigg]\Bigg[\sum_{k'=1}^{K^\theta_{w_{\theta,t},T}} g_b\bigg(\frac{w_{\theta,t}+(k'-1)\theta}{T}\bigg)\Bigg]\Bigg\}$$
$$\leq C\sum_{l,t=1}^{T} \sum_{(i,j)\in J_{t/T}\times J_{l/T}} |\mathrm{Cov}(\epsilon_{i,T}\epsilon_{j,T})| \sup_{i\in[T]}\sup_{x\in[0,1]}|W_{i,T}(x)| \sup_{j\in[T]}\sup_{x\in[0,1]}|W_{j,T}(x)|h^4$$
$$\leq \frac{Ch^4}{(Th)^2}\sum_{l,t=1}^{T} \sum_{(i,j)\in J_{t/T}\times J_{l/T}} |\mathrm{Cov}(\epsilon_{i,T}\epsilon_{j,T})| \leq C\frac{h^2}{T^2}T^2 k_T \leq CTh^3,$$

for T large enough, where J_x and k_T are defined as in (3.56). Also, $\mathrm{Var}\big(G_2^{(g_b,g_\epsilon)}\big) \leq CTh^3$. By Chebychev's inequality,

$$P\bigg(|U_\theta^{(g_b,g_\epsilon)}| > v_T/\sqrt{h}\bigg) \leq P\bigg(|G_1^{(g_b,g_\epsilon)}| + |G_2^{(g_b,g_\epsilon)}| > v_T/\sqrt{h}\bigg)$$
$$\leq P\bigg(|G_1^{(g_b,g_\epsilon)}| > \frac{v_T}{2\sqrt{h}}\bigg) + P\bigg(|G_2^{(g_b,g_\epsilon)}| > \frac{v_T}{2\sqrt{h}}\bigg)$$
$$\leq C\frac{Th^4}{v_T^2} \leq \frac{C}{v_T^2}, \tag{3.64}$$

for T large enough. Analogously as in (3.63), we decompose

$$U_\theta^{(g_b,g_\epsilon)} = G_1^{(g_m,g_\epsilon)} - G_2^{(g_b,g_\epsilon)},$$
$$U_\theta^{(g_m,\epsilon)} = G_1^{(g_m,\epsilon)} - G_2^{(g_m,\epsilon)},$$
$$U_\theta^{(g_b,\epsilon)} = G_1^{(g_b,\epsilon)} - G_2^{(g_b,\epsilon)}.$$

It can be easily seen that the sequence $\mathrm{Var}\big(G_i^{(g_b,g_\epsilon)}\big)$ dominates $\mathrm{Var}\big(G_i^{(g_m,g_\epsilon)}\big)$, $\mathrm{Var}\big(G_i^{(g_m,\epsilon)}\big)$ and $\mathrm{Var}\big(G_i^{(g_b,\epsilon)}\big), i \in \{1,2\}$. Hence,

$$P\bigg(|U_\theta^{(g_m,g_\epsilon)}| > v_T/\sqrt{h}\bigg) \leq \frac{C}{v_T^2}, \tag{3.65}$$
$$P\bigg(|U_\theta^{(g_m,\epsilon)}| > v_T/\sqrt{h}\bigg) \leq \frac{C}{v_T^2}, \tag{3.66}$$
$$P\bigg(|U_\theta^{(g_b,\epsilon)}| > v_T/\sqrt{h}\bigg) \leq \frac{C}{v_T^2}, \tag{3.67}$$

for T large enough.

□

Say that a real sequence a_T is $\Theta(b_T)$ if there are constants $m, M > 0$ such that $b_T m \leq a_T \leq M b_T$ for all sufficiently large T.

Lemma 3.9. *Suppose the conditions of Lemma 3.8 hold. Assume that* $h = \Theta(T^{-1/4})$,

$\lambda_T = o(T)$ and $T^{1/4}\Theta_T\rho_T^{1/2} = o(\lambda_T)$, where ρ_T is a positive sequence slowly diverging to infinity. Then for all $\theta \neq \theta_0 : 1 \leq \theta \leq \Theta_T$ and all T sufficiently large,

$$\Pr\{Q(\theta, \lambda_T) \leq Q(\theta_0, \lambda_T)\} \leq \frac{C}{\Theta_T\rho_T}.$$

Proof. In general, given a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}$-measurable sets $A, A_1, \ldots, A_k$, it holds that

$$A = A \cap \left[\Big(\bigcap_{i\in[k]} A_i\Big) \cup \Big(\bigcap_{i\in[k]} A_i\Big)^c\right] = \Big[A \cap \Big(\bigcap_{i\in[k]} A_i\Big)\Big] \cup \Big[A \cap \Big(\bigcup_{i\in[k]} A_i^c\Big)\Big]$$
$$\subseteq \Big[A \cap \Big(\bigcap_{i\in[k]} A_i\Big)\Big] \cup \Big(\bigcup_{i\in[k]} A_i^c\Big),$$

and then, by the monotonicity and subadditivity of the measure, $P(A) \leq P\Big[A\cap\Big(\bigcap_{i\in[k]} A_i\Big)\Big] + \sum_{i=1}^k P(A_i^c)$.

Set $v_T = \sqrt{\Theta_T\rho_T}$ and $\rho_T = \ln\ln T$. Then $\rho_T = o(T^a)$, for any $a > 0$. Suppose that θ satisfies the case A. By the above result together with Lemma 3.8, we obtain

$$\begin{aligned}
P\{\tilde{Q}(\theta,\lambda_T) \leq \tilde{Q}(\theta_0, \lambda_T)\} \leq{}& P\bigg\{V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \leq -B_\theta + W_\theta^{g_b} + W_\theta^{g_m} \\
&- 2S_\theta^{g_b} + 2S_\theta^{g_m} - 2U_\theta^{(g_b,g_m)} - 2U_\theta^{(g_b,g_\epsilon)} - 2U_\theta^{(g_b,\epsilon)} + 2U_\theta^{(g_m,g_\epsilon)} - 2U_\theta^{(g_m,\epsilon)} + S_\theta^{g_\epsilon} \\
&- S_\theta^\epsilon + \lambda_T(\theta_0 - \theta),\ \big|S_\theta^\epsilon\big| \leq v_T\frac{n\sqrt{T}}{\theta},\ \big|S_\theta^{g_\epsilon}\big| \leq v_T\frac{n\sqrt{T/h}}{\theta},\ |U_\theta^{(g_m,g_\epsilon)}| \leq v_T/\sqrt{h}, \\
&|U_\theta^{(g_b,g_\epsilon)}| \leq v_T/\sqrt{h},\ |U_\theta^{(g_b,\epsilon)}| \leq v_T/\sqrt{h},\ |U_\theta^{(g_m,\epsilon)}| \leq v_T/\sqrt{h}\bigg\} \\
&+ P\bigg(\big|S_\theta^\epsilon\big| > v_T\frac{n\sqrt{T}}{\theta}\bigg) + P\bigg(\big|S_\theta^{g_\epsilon}\big| > v_T\frac{n\sqrt{T/h}}{\theta}\bigg) + P\bigg(|U_\theta^{(g_m,g_\epsilon)}| > v_T/\sqrt{h}\bigg) \\
&+ P\bigg(|U_\theta^{(g_b,g_\epsilon)}| > v_T/\sqrt{h}\bigg) + P\bigg(|U_\theta^{(g_b,\epsilon)}| > v_T/\sqrt{h}\bigg) + P\bigg(|U_\theta^{(g_m,\epsilon)}| > v_T/\sqrt{h}\bigg) \\
\leq{}& P\bigg\{V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \leq -B_\theta + Cv_T\frac{n\sqrt{T/h}}{\theta} + \lambda_T(\theta_0 - \theta)\bigg\} + \frac{C}{v_T^2},
\end{aligned}$$

for all T sufficiently large.

If $\theta \geq \theta_0$, then

$$\bigg(Cv_T\frac{n\sqrt{T/h}}{\theta} + \lambda_T(\theta_0 - \theta)\bigg)\frac{\theta}{nT} \leq \frac{Cv_T}{\sqrt{Th}} \leq C(T^{-7/20-\omega}\ln\ln T)^{1/2} = o(1),$$

for T large enough. If $\theta < \theta_0$, then

$$\bigg(Cv_T\frac{n\sqrt{T/h}}{\theta} + \lambda_T(\theta_0 - \theta)\bigg)\frac{\theta}{nT} \leq \frac{Cv_T}{\sqrt{Th}} + \frac{\theta_0^2\lambda_T}{T} = o(1),$$

by the hypothesis $\lambda_T = o(T)$. Therefore, regardless of whether $\theta > \theta_0$ or $\theta < \theta_0$, it holds that $\forall \delta_2 > 0 : \exists T_1 \in \mathbb{N} : \forall T \geq T_1 : |Cv_T n\sqrt{T/h}/\theta + \lambda_T(\theta_0 - \theta)| \leq \delta_2 nT/\theta$. Hence, for every $\delta_2 : c > \delta_2 > 0$, Lemma 3.8 implies that there exists $T_2 \geq T_1 > 0$ such that for all $T > T_2$

$$-B_\theta + Cv_T n/\theta\sqrt{T/h} + \lambda_T(\theta_0 - \theta) \leq -(c - \delta_2)nT/\theta = -C_1 nT/\theta, \tag{3.68}$$

for some positive constant C_1. Applying Lemma 2.11 two times, we have

$$\begin{aligned} P\{\tilde{Q}(\theta, \lambda_T) \leq \tilde{Q}(\theta_0, \lambda_T)\} &\leq P\left\{V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \leq -C_1 nT/\theta\right\} + \frac{C}{v_T^2} \\ &\leq P\left\{|V_\theta^{(\epsilon,\epsilon)}| + |V_\theta^{(\epsilon,g_\epsilon)}| + |V_\theta^{(g_\epsilon,g_\epsilon)}| \geq C_1 nT/\theta\right\} + \frac{C}{v_T^2} \\ &\leq P\left\{|V_\theta^{(\epsilon,\epsilon)}| \geq CnT/\theta\right\} + P\left\{|V_\theta^{(\epsilon,g_\epsilon)}| \geq CnT/\theta\right\} \\ &+ P\left\{|V_\theta^{(g_\epsilon,g_\epsilon)}| \geq CnT/\theta\right\} + \frac{C}{v_T^2} \\ &:= P_1 + P_2 + P_3 + \frac{C}{v_T^2}, \end{aligned} \tag{3.69}$$

for T sufficiently large. Now we need to bound each of the probabilities P_1, P_2 and P_3.

We start with P_3. Analogously to the decomposition as for (3.59) in the Lemma 3.8's proof, we can write

$$\begin{aligned} V_\theta^{(g_\epsilon,g_\epsilon)} &= g_\epsilon'(\Pi_{\theta_0} - \Pi_\theta)g_\epsilon \\ &= \sum_{l=1}^{T} g_\epsilon(l/T)\left[\frac{1}{K^{\theta_0}_{w_{\theta_0,l},T}} \sum_{k=1}^{K^{\theta_0}_{w_{\theta_0,l},T}} g_\epsilon\left(\frac{w_{\theta_0,l} + (k-1)\theta_0}{T}\right)\right] \\ &- \sum_{l=1}^{T} g_\epsilon(l/T)\left[\frac{1}{K^{\theta}_{w_{\theta,l},T}} \sum_{k=1}^{K^{\theta}_{w_{\theta,l},T}} g_\epsilon\left(\frac{w_{\theta,l} + (k-1)\theta}{T}\right)\right] \\ &:= V_{\theta,1}^{(g_\epsilon,g_\epsilon)} - V_{\theta,2}^{(g_\epsilon,g_\epsilon)}. \end{aligned} \tag{3.70}$$

From Lemma 2.11 we have that

$$\begin{aligned} P_3 &\leq P\left(|V_{\theta,1}^{(g_\epsilon,g_\epsilon)} - V_{\theta,2}^{(g_\epsilon,g_\epsilon)}| \geq C\frac{nT}{\Theta_T}\right) \leq P\left(|V_{\theta,1}^{(g_\epsilon,g_\epsilon)}| + |V_{\theta,2}^{(g_\epsilon,g_\epsilon)}| \geq C\frac{nT}{\Theta_T}\right) \\ &\leq P\left(|V_{\theta,1}^{(g_\epsilon,g_\epsilon)}| \geq \frac{C}{2}\frac{nT}{\Theta_T}\right) + P\left(|V_{\theta,2}^{(g_\epsilon,g_\epsilon)}| \geq \frac{C}{2}\frac{nT}{\Theta_T}\right) \\ &:= P_{3,a} + P_{3,b}. \end{aligned} \tag{3.71}$$

Denote $J_{l,t} = J_{t/T} \times J_{l/T} \times J_{[w_{\theta,l}+(k-1)\theta]/T} \times J_{[w_{\theta,t}+(k-1)\theta]/T}$, for any $l, t \in [T]$, where $J_x, x \in [0,1]$, is defined as in (3.56). The application of Theorem 2.1 of Rio (2017) with the hypothesis that $E(\epsilon_{t,T}^4(\ln(1+\epsilon_{t,T})^3)) < \infty$ gives

$$\begin{aligned}
\mathrm{Var}\big(V_{\theta,2}^{(g_\epsilon,g_\epsilon)}\big) &\le E\bigg\{\bigg[\sum_{l=1}^{T} g_\epsilon(l/T)\bigg[\frac{1}{K^\theta_{w_{\theta,l},T}}\sum_{k=1}^{K^\theta_{w_{\theta,l},T}} g_\epsilon\Big(\frac{w_{\theta,l}+(k-1)\theta}{T}\Big)\bigg]\bigg]^2\bigg\}\\
&\le \frac{C}{(Th)^4}\sum_{l,t=1}^{T}\sum_{k=1}^{K^\theta_{w_{\theta,l},T}}\sum_{k=1}^{K^\theta_{w_{\theta,t},T}}\sum_{i,i',j,j'=1}^{T}|E(\epsilon_{i,T}\epsilon_{i',T}\epsilon_{j,T}\epsilon_{j',T})|\\
&\times \frac{1}{K^\theta_{w_{\theta,l},T}K^\theta_{w_{\theta,t},T}} I((i,i',j,j')\in J_{l,t})\\
&\le \frac{C}{(Th)^4}\sum_{l,t=1}^{T}\bigg\{3\Big(\sum_{i,i'\in J_{t/T}\times J_{l/T}}|E(\epsilon_{p,T}\epsilon_{q,T})|\Big)^2\\
&+48\sum_{k=1}^{T}\int_0^1[\min(\alpha^{-1}(u),n)]^3Q_k^4(u)du\bigg\}\\
&\le \frac{C}{T^2h^4}\Big\{\underbrace{k_T^2+T}_{=O((Th)^2)}\Big\}\le \frac{C}{h^2},
\end{aligned} \tag{3.72}$$

for T large enough, where $k_T = \#J_x = O(Th)$, $(p,q,r,s) = (i,i',j,j') - \min J_{l/T} + 1$, $\alpha^{-1}(u) = \inf\{k \in \mathbb{N} : \alpha(k) \le u\}$, $Q_k(u) = \inf\{t > 0 : P(|\epsilon_k| > t) \le u\}$.

The same bound holds for $\mathrm{Var}\big(V_{\theta,1}^{(g_\epsilon,g_\epsilon)}\big)$. By Chebychev's inequality

$$P_3 \le C\Big(\frac{\Theta_T}{Th}\Big)^2 \le CT^{-7/10-2\omega} = CT^{-2/5+\omega}T^{-3/10-3\omega} \le \frac{C}{\Theta_T\rho_T}. \tag{3.73}$$

The remaining stochastic terms can be decomposed analogously as

$$\begin{aligned}
V_\theta^{(\epsilon,g_\epsilon)} &= V_{\theta,1}^{(\epsilon,g_\epsilon)} - V_{\theta,2}^{(\epsilon,g_\epsilon)},\\
V_\theta^{(\epsilon,\epsilon)} &= V_{\theta,1}^{(\epsilon,\epsilon)} - V_{\theta,2}^{(\epsilon,\epsilon)}.
\end{aligned}$$

Using Theorem 2.1 of Rio (2017) again, we have

$$\mathrm{Var}\big(V_{\theta,1}^{(\epsilon,g_\epsilon)}\big), \mathrm{Var}\big(V_{\theta,2}^{(\epsilon,g_\epsilon)}\big) \le C\frac{1}{(Th)^2}\sum_{i,i',j,j'=1}^{T}|E(\epsilon_{i,T}\epsilon_{i',T}\epsilon_{j,T}\epsilon_{j',T})| \le \frac{C}{h^2};$$

$$\begin{aligned}
\mathrm{Var}\big(V_{\theta,1}^{(\epsilon,\epsilon)}\big), \mathrm{Var}\big(V_{\theta,2}^{(\epsilon,\epsilon)}\big) &\le \frac{C\Theta_T^2}{T^2}\sum_{i,i'=1}^{T}\sum_{k=1}^{K^\theta_{w_{\theta,i},T}}\sum_{k'=1}^{K^\theta_{w_{\theta,i'},T}}|E(\epsilon_{i,T}\epsilon_{i',T}\epsilon_{w_{\theta,i+(k-1)\theta},T}\epsilon_{w_{\theta,i'+(k'-1)\theta},T})|\\
&\le \frac{C\Theta_T^2}{T^2}\sum_{i,i',j,j'=1}^{T}|E(\epsilon_{i,T}\epsilon_{i',T}\epsilon_{j,T}\epsilon_{j',T})| \le C\Theta_T^2.
\end{aligned}$$

for T large enough. Then $P_2 \le C/(\Theta_T \rho_T)$. In addition, from Chebychev's inequality,

$$P_1 \le CT^{-2/5-4\omega} = CT^{-2/5+\omega}T^{-5\omega} \le \frac{C}{\Theta_T \rho_T}, \tag{3.74}$$

for all sufficiently large T.

Hence, in case A,

$$P\{\tilde{Q}(\theta, \lambda_T) \le \tilde{Q}(\theta_0, \lambda_T)\} \le \frac{C}{\Theta_T \rho_T}, \tag{3.75}$$

for all sufficiently large T.

Next, let us focus in case B. By Lemma 3.8,

$$\begin{aligned}
&P\{\tilde{Q}(\theta, \lambda_T) \le \tilde{Q}(\theta_0, \lambda_T)\} \le P\bigg\{V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \le W_\theta^{g_b} + W_\theta^{g_m} - 2U_\theta^{(g_b,g_m)} \\
&- 2U_\theta^{(g_b,g_\epsilon)} - 2U_\theta^{(g_b,\epsilon)} + 2U_\theta^{(g_m,g_\epsilon)} - 2U_\theta^{(g_m,\epsilon)} + \lambda_T(\theta_0 - \theta),\ |U_\theta^{(g_m,g_\epsilon)}| \le \upsilon_T/\sqrt{h}, \\
&\quad |U_\theta^{(g_b,g_\epsilon)}| \le \upsilon_T/\sqrt{h},\ |U_\theta^{(g_b,\epsilon)}| \le \upsilon_T/\sqrt{h},\ |U_\theta^{(g_m,\epsilon)}| \le \upsilon_T/\sqrt{h}\bigg\} + \frac{C}{\upsilon_T^2} \\
&\le P\bigg\{V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \le C\upsilon_T/h + \lambda_T(\theta_0 - \theta)\bigg\} + \frac{C}{\upsilon_T^2},
\end{aligned}$$

for T large enough. Since $\theta_0 - \theta < 0$ and $T^{1/4}\rho_T^{1/2}\Theta_T = o(\lambda_T)$, it follows that, for all T sufficiently large, there is $C_4 > 0$ satisfying

$$\frac{\upsilon_T}{h} + \lambda_T(\theta_0 - \theta) \le CT^{1/4}(\Theta_T\rho_T)^{1/2} + \lambda_T(\theta_0 - \theta) \le \lambda_T(C\delta_4 + (\theta_0 - \theta)) \le -C_4\lambda_T,$$

for any $\delta_4 > 0$ small enough so that $C\delta_4 + (\theta_0 - \theta) < 0$. We thus have that

$$\begin{aligned}
P\{\tilde{Q}(\theta, \lambda_T) \le \tilde{Q}(\theta_0, \lambda_T)\} &\le P\bigg\{V_\theta^{(\epsilon,\epsilon)} + V_\theta^{(\epsilon,g_\epsilon)} + V_\theta^{(g_\epsilon,g_\epsilon)} \le -C\lambda_T\bigg\} + \frac{C}{\upsilon_T^2} \\
&\le P\bigg\{|V_\theta^{(\epsilon,\epsilon)}| \ge C\lambda_T\bigg\} + P\bigg\{|V_\theta^{(\epsilon,g_\epsilon)}| \ge C\lambda_T\bigg\} \\
&+ P\bigg\{|V_\theta^{(g_\epsilon,g_\epsilon)}| \ge C\lambda_T\bigg\} + \frac{C}{\upsilon_T^2} \\
&:= Q_1 + Q_2 + Q_3 + \frac{C}{\upsilon_T^2}.
\end{aligned} \tag{3.76}$$

Along the same lines as for the case A, we have that

$$\begin{aligned}
Q_2, Q_3 &\le \frac{C}{(\lambda_T h)^2} \le \frac{C}{\Theta_T^2 \rho_T} \le \frac{C}{\Theta_T \rho_T}; \\
Q_1 &\le C\frac{\Theta^2}{\lambda_T^2} \le C\frac{\Theta_T^{5/4}\Theta_T^{3/4}}{T^{1/2}\Theta_T^2\rho_T} \le \frac{C}{T^{5\omega/4}\Theta_T^{5/4}\rho_T} \le \frac{C}{\Theta_T\rho_T}.
\end{aligned}$$

Therefore, in case B,

$$P\{\tilde{Q}(\theta,\lambda_T) \le \tilde{Q}(\theta_0,\lambda_T)\} \le \frac{C}{\Theta_T \rho_T},$$

for sufficiently large T. □

Lemma 3.10. *Let $x \in [0,1]$, $i \in \{1,\dots,\theta_0\}$ and $\theta_0 \in \{1,\dots,\Theta_T\}$, with $1 \le \Theta_T < T$, be given with T large enough so that the set*

$$J_{x,i} = \{k \in \{1,\dots,K_{i,T}^{\theta_0}\} : (i+(k-1)\theta_0)/T \in C_x\}$$

where

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

is well-defined and nonempty. Then the cardinality of $J_{x,i}$ is $O(K_{i,T}^{\theta_0} h)$. Denote $\gamma_{i,x,T,k}^{\theta_0} = (i+(k-1)\theta_0)/T$, $\forall T, k \in \mathbb{N}$. Under Condition 4, for all sufficiently large T and $j \in \mathbb{N}$,

$$\left| \frac{1}{K_{i,T}^{\theta_0}} \sum_{k=1}^{K_{i,T}^{\theta_0}} K_h\big(\gamma_{i,x,T,k}^{\theta_0} - x\big)\left(\frac{\gamma_{i,x,T,k}^{\theta_0} - x}{h}\right)^j - \int_0^1 K_h(u-x)\left(\frac{u-x}{h}\right)^j du \right| \le \frac{C}{Th}.$$

Proof. Define $k_* = \min J_{x,i}$, $k^* = \max J_{x,i}$, $\bar{C}_x = \sup C_x$ and $\underset{\sim}{C}_x = \inf C_x$. For brevity's sake, let $\gamma_k := \gamma_{i,x,T,k}^{\theta_0}$ and $J_{x,i}^* = J_{x,i} \setminus \{k_*\}$. Along the same lines of Lemma 2.2's proof in the previous chapter, we can find that $\#J_{x,i} = O(Th/\theta_0) = O(K_{i,T}^{\theta_0} h)$, by Lemma 3.4, and that $0 \le \bar{C}_x - \gamma_{k^*} \le \theta_0/T$ and $0 \le \gamma_{k_*} - \underset{\sim}{C}_x \le \theta_0/T$ hold. Furthermore, we have

$$\begin{aligned}
&\left| \frac{1}{K_{i,T}^{\theta_0}} \sum_{k \in J_{x,i}} K_h(\gamma_k - x)\left(\frac{\gamma_k - x}{h}\right)^j - \int_{C_x} K_h(u-x)\left(\frac{u-x}{h}\right)^j du \right| \\
&\le \left| \frac{1}{K_{i,T}^{\theta_0}} \sum_{k \in J_{x,i}^*} K_h(\gamma_k - x)\left(\frac{\gamma_k - x}{h}\right)^j - \sum_{k \in J_{x,i}^*} \int_{\gamma_k}^{\gamma_k + \theta_0/T} K_h(u-x)\left(\frac{u-x}{h}\right)^j du \right| \\
&+ \frac{1}{K_{i,T}^{\theta_0}} K_h(\gamma_{k_*} - x)\left|\frac{\gamma_{k_*} - x}{h}\right|^j + \int_{\underset{\sim}{C}_x}^{\gamma_{k_*}} K_h(u-x)\left|\frac{u-x}{h}\right|^j du \\
&+ \int_{i^*/T}^{\bar{C}_x} K_h(u-x)\left|\frac{u-x}{h}\right|^j du \\
&\le \sum_{k \in J_{x,i}^*} \left| \frac{1}{K_{i,T}^{\theta_0}} K_h(\gamma_k - x)\left(\frac{\gamma_k - x}{h}\right)^j - \frac{\theta_0}{T} K_h\Big(\xi_k - x\Big)\left(\frac{\xi_k - x}{h}\right)^j \right| + \frac{C}{Th} \\
&\le \frac{1}{K_{i,T}^{\theta_0}} \sum_{k \in J_{x,i}^*} \left| K_h(\gamma_k - x)\left(\frac{\gamma_k - x}{h}\right)^j - K_h\Big(\xi_k - x\Big)\left(\frac{\xi_k - x}{h}\right)^j \right|
\end{aligned}$$

$$
\begin{aligned}
&+ \underbrace{\left|\frac{1}{K_{i,T}^{\theta_0}} - \frac{\theta_0}{T}\right|}_{\le C/T^2} \sum_{k \in J_{x,i}^*} \left| K_h\Big(\xi_k - x\Big)\Big(\frac{\xi_k - x}{h}\Big)^j \right| + \frac{C}{Th} \\
&\le \frac{1}{K_{i,T}^{\theta_0}} \sum_{k \in J_{x,i}^*} \left| K_h(\gamma_k - x)\Big(\frac{\gamma_k - x}{h}\Big)^j - K_h\Big(\xi_k - x\Big)\Big(\frac{\xi_k - x}{h}\Big)^j \right| + \frac{C}{T} + \frac{C}{Th} \\
&\le \frac{1}{K_{i,T}^{\theta_0}} \sum_{k \in J_{x,i}^*} \Bigg\{ K_h(\gamma_k - x) \left| \Big(\frac{\gamma_k - x}{h}\Big)^j - \Big(\frac{\xi_i - x}{h}\Big)^j \right| \\
&+ \left|\frac{\xi_i - x}{h}\right|^j \Big| K_h(\gamma_k - x) - K_h(\xi_k - x) \Big| \Bigg\} + \frac{C}{Th} \le \frac{C}{Th},
\end{aligned}
$$

where $\xi_k \in (\gamma_k, \gamma_k + \theta_0/T)$ for each $k \in J_{x,i}^*$. To see that $|1/K_{i,T}^{\theta_0} - \theta_0/T| \le C/T^2$ holds, note that the facts

$$
\frac{T}{\theta_0} - 1 < \left\lfloor \frac{T}{\theta_0} \right\rfloor \le \frac{T}{\theta_0} \iff a_T := \frac{1}{T/\theta_0 - 1} - \frac{1}{T/\theta_0} > \frac{1}{\lfloor T/\theta_0 \rfloor} - \frac{1}{T/\theta_0} \ge 0
$$

and

$$
\frac{T}{\theta_0} < \left\lfloor \frac{T}{\theta_0} \right\rfloor + 1 \le \frac{T}{\theta_0} + 1 \iff 0 > \frac{1}{\lfloor T/\theta_0 \rfloor + 1} - \frac{1}{T/\theta_0} \ge \frac{1}{T/\theta_0 + 1} - \frac{1}{T/\theta_0} := -b_T,
$$

imply

$$
\begin{aligned}
\lim_{T\to\infty} T^2 a_T &= \lim_{T\to\infty} \frac{T^2\theta_0^2}{T^2 - \theta_0 T} = \theta_0^2 \\
\lim_{T\to\infty} T^2 b_T &= \lim_{T\to\infty} \frac{T^2\theta_0^2}{T^2 + \theta_0 T} = \theta_0^2,
\end{aligned}
$$

and thus $T^2 a_T, T^2 b_T$ are convergent nonnegative sequences, which in turn imply that there is $C > 0$ such that both a_T and b_T are bounded by C/T^2. By Lemma 3.4, $K_{i,T}^{\theta_0}$ is either $\lfloor T/\theta_0 \rfloor$ or $\lfloor T/\theta_0 \rfloor + 1$. Therefore, $|1/K_{i,T}^{\theta_0} - \theta_0/T| \le \max(a_T, b_T) \le C/T^2$. □

Lemma 3.11. *Let $T \in \mathbb{N}$ be given. Let $\{\epsilon_{t,T} : 1 \le t \le T, T \ge 1\}$ be a strong mixing triangular array on $(\Omega, \mathcal{F}, P)$ with mixing sequence α_T and $\{a_{t,T}(x) : 1 \le t \le T, T \ge 1\}$ be a triangular array of finite real numbers. Finally, let $J \subseteq [T]$ be a set and k_T its cardinality with k_T being a sequence diverging to infinity. Then the sub-array $\{a_{t,T}\epsilon_{t,T} I(t \in J) : 1 \le t \le T, T \ge 1\}$ is also strongly mixing with mixing coefficients $\alpha'_T(j)$ bounded by $\alpha_T(j)$, for any $0 \le j < T$.*

Proof. By definition,

$$
\alpha'_T(j) = \sup_{1 \le k \le T-j} \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{1,T}^k, B \in \mathcal{F}_{k+j,T}^T\}, \quad 0 \le j < T
$$

where $\mathcal{F}_{i,T}^k = \sigma(a_{l,T}\epsilon_{l,T}^* : i \leq l \leq k)$ and $\epsilon_{t,T}^* = \epsilon_{t,T}I(t \in J)$. For any $1 \leq i \leq k \leq T$, we have

$$\begin{aligned}
\mathcal{F}_{i,T}^k = \sigma(a_{l,T}\epsilon_{l,T}^* : i \leq l \leq k) &= \sigma(\cup_{l=i}^k \sigma(a_{l,T}\epsilon_{l,T}^*)) \\
&\subseteq \sigma(\cup_{l=i}^k \sigma(a_{l,T}, \epsilon_{l,T}^*)) \\
&= \sigma\{\cup_{l=i}^k [(\epsilon_{l,T}^*)^{-1}(\mathbb{B}_{\mathbb{R}}) \cup \{\emptyset, \Omega\}]\} \\
&= \sigma(\cup_{l=i}^k (\epsilon_{l,T}^*)^{-1}(\mathbb{B}_{\mathbb{R}})) = \sigma(\epsilon_{l,T}^* : i \leq l \leq k), \qquad (3.77)
\end{aligned}$$

since the sigma-algebra generated by a constant is the trivial sigma-algebra. To justify the inclusion in (3.77), consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = xy$.

Claim 5. The function $f : (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measurable.

Proof of claim: The Borel sigma-algebra on $\mathbb{R}^2$ is defined as the sigma-algebra generated by the set of open sets in $\mathbb{R}^2$. That is, it is the smallest sigma-algebra containing all open sets in $\mathbb{R}^2$. Furthermore, it is well known that $f : (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measurable if, and only if, $\forall a \in \mathbb{R}, \{(x, y) : f(x, y) < a\} \in \mathcal{B}_{\mathbb{R}^2}$. Since f is continuous, $\{f^{-1}(-\infty, a)\}$ is open, and hence must be in $\mathcal{B}_{\mathbb{R}^2}$, for any $a \in \mathbb{R}$. ■

Define the random vector $Z = (X, Y)$. As $XY = f(Z)$ and f is Borel,

$$\sigma(XY) = \sigma(f(Z)) = \{(Z^{-1} \circ f^{-1})(A) : A \in \mathcal{B}_{\mathbb{R}}\} \subseteq \{Z^{-1}(B) : B \in \mathcal{B}_{\mathbb{R}}\} = \sigma(X, Y).$$

From (3.77), for any $0 \leq j < T$ and any $1 \leq k \leq T - j$, it holds that

$$\begin{aligned}
\mathcal{F}_{1,T}^k \subseteq \sigma(\epsilon_{l,T}^* : 1 \leq l \leq k) &= \sigma\{(\cup_{l \in [k] \cap J} \sigma(\epsilon_{l,T})) \cup (\cup_{l \in [k] \setminus J} \sigma(0))\} \\
&= \sigma\{(\cup_{l \in [k] \cap [d_T]} \sigma(\epsilon_{l,T})) \cup \{\emptyset, \Omega\}\} \\
&= \sigma\{\cup_{l \in [k] \cap [d_T]} \sigma(\epsilon_{l,T})\} \subseteq \sigma\{(\cup_{l=1}^k \sigma(\epsilon_{l,T}))\} \qquad (3.78)
\end{aligned}$$

and similarly

$$\mathcal{F}_{k+j,T}^T \subseteq \sigma(\epsilon_{l,T}^* : k + j \leq l \leq T) \subseteq \sigma\{(\cup_{l=k+j}^T \sigma(\epsilon_{l,T}))\}. \qquad (3.79)$$

The inclusions (3.78) and (3.79) imply the result. □

Appendix D - General Central Limit Theorems for mixing arrays

Politis et al. (1997) obtained a Central Limit Theorem (CLT) for strong mixing sequences without the strict stationarity assumption. CLTs for strong mixing sequences are traditionally proved using Bernstein's method. The main idea of this method is to split a sum $X_1 + \ldots + X_n$ into a sum of nearly independent random variables (the big blocks) and a sum of other terms (small blocks) which is asymptotically negligible if properly normalized. In order to derived this result, we need the following lemmas.

Lemma 3.12 (Ibragimov's Bound)**.** *Let $\{X_t\}$ be a sequence of random vectors defined on a probability space and let $\mathcal{F}_a^b = \sigma(X_t : a \leq t \leq b)$. Also, denote the mixing coefficient corresponding to $\{X_t\}$ by α_X. Let Y_1, Y_2 be random variables measurable with respect to $\mathcal{F}_{-\infty}^n, \mathcal{F}_{n+m}^\infty$, respectively. In addition, let Y_1', Y_2' be independent random variables having the same probability distribution as Y_1, Y_2, respectively. Denote the characteristic functions of $Y_1 + Y_2$ and $Y_1' + Y_2'$ by φ and φ', respectively. Then $\sup_t|\varphi(t) - \varphi'(t)| \leq 16\alpha_X(m)$.*

Proof. By Euler's formula and Billingsley's inequality, we have

$$\begin{aligned}
|\varphi(t) - \varphi'(t)| &= \big|E\big(e^{it(Y_1+Y_2)}\big) - E\big(e^{it(Y_1'+Y_2')}\big)\big| = \big|E\big(e^{itY_1}e^{itY_2}\big) - E\big(e^{itY_1}\big)E\big(e^{itY_2}\big)\big| \\
&= \big|\mathrm{Cov}(\cos tY_1 + i\sin tY_1, \cos tY_2 + i\sin tY_2)\big| \\
&\leq |\mathrm{Cov}(\cos tY_1, \cos tY_2)| + |\mathrm{Cov}(\sin tY_1, \sin tY_2)| \\
&+ |\mathrm{Cov}(\cos tY_1, \sin tY_2)| + |\mathrm{Cov}(\sin tY_1, \cos tY_2)| \\
&\leq 16\alpha_X(m),
\end{aligned}$$

since $\operatorname{ess\,sup}|\cos(tY_i)| \leq 1$ and $\operatorname{ess\,sup}|\sin(tY_i)| \leq 1$ (and thus are in L^∞), for any t and any $i = 1, 2$. □

Lemma 3.13 (Doukhan's Moment Bound)**.** *Let $\{X_i\}$ be a sequence of mean zero random variables and denote the corresponding mixing coefficient by α_X. Define, for $\tau \geq 2$ and $\delta > 0$*

$$\begin{aligned}
C(\tau, \delta) &= \sum_{k=0}^{\infty} (k+1)^{\tau-2} \alpha_X^{\delta/(\tau-\delta)}(k), \\
L(\tau, \delta, d) &= \sum_{i=1}^{d} \|X_i\|_{\tau+\delta}^{\tau}, \\
D(\tau, \delta, d) &= \max(L(\tau, \delta, d), [L(2, \delta, d)]^{\tau/2}).
\end{aligned}$$

Then $E\big|\sum_{i=1}^d X_i\big|^\tau \leq BD(\tau, \delta, d)$, where B is a constant depending only on τ, δ and α_X.

In particular, if τ is an even integer, then

$$E\Big|\sum_{i=1}^{d} X_i\Big|^{\tau} \le B(\tau,\delta)D(\tau,\delta,d)$$

where $B(\tau,\delta)$ can be computed recursively, e.g., for τ up to 4,

$$\begin{aligned}
&B(1,\delta) \le 1;\\
&B(2,\delta) \le 18\max\{1, C(2,\delta)\};\\
&B(3,\delta) \le 102\max\{1, C(3,\delta)\};\\
&B(4,\delta) \le 3024\max\{1, C^2(4,\delta)\}.
\end{aligned}$$

If we additionally assume that $\|X_i\|_{2+2\delta} \le \Delta,\ \forall i$, then

$$E\Big|\sum_{i=1}^{d} X_i\Big|^{2+2\delta} \le \Gamma d^{1+\delta/2}$$

where $\Gamma = \{3024\max[1, C^2(4,\delta)]\}2^{4[1+4(2-\delta)/\delta]}\Delta^{(2+\delta)(1+\delta/2)}$.

The proof of this lemma is in the Appendix A of Politis et al. (1997).

Theorem 3.4 (Lyapunov's CLT)**.** *Suppose that $\{X_i\}$ is a sequence of independent random variables such that, for each i, $E(X_i) = \mu_i < \infty$ and $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$. Define $s_n^2 = \sum_{k=1}^{n}\sigma_k^2$. If there exists $\delta > 0$ so that $|X_i|^{2+\delta}$ are integrable and the Lyaponov's condition holds, i.e.,*

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{k=1}^{n} E\{|X_k-\mu_k|^{2+\delta}\} = 0,$$

then

$$\frac{1}{s_n}\sum_{k=1}^{n}(X_k-\mu_k) \xrightarrow{d} N(0,1).$$

For a proof of the Lyapunov's CLT, see Theorem 27.3 of Billingsley (1995).

Theorem 3.5 (Politis' CLT)**.** *Let $\{X_{n,i} : 1 \le i \le d_n\}$ be a triangular array of mean zero random variables. Denote the strong mixing coefficient corresponding to the nth row by α_n. Define*

$$S_{n,k,a} = \sum_{i=a}^{a+k-1} X_{n,i}; \quad T_{n,k,a} = k^{-1/2}\sum_{i=a}^{a+k-1} X_{n,i}; \quad \text{and} \ \ \sigma_{n,k,a}^2 = \mathrm{Var}(T_{n,k,a}).$$

Assume the conditions: for some $\delta > 0$,

(A.1) $\|X_{n,i}\|_{2+2\delta} \le \Delta,\ \forall n, i$;

(A.2) $\sigma_{n,k,a}^2 \to \sigma^2$ *uniformly in a, i.e., for any sequence $k_n := k$ that tends to infinity, $\sup_a|\sigma_{n,k,a}^2 - \sigma^2| \to 0$ as $n\to\infty$;*

(A.3) $\sum_{k=0}^{\infty}(k+1)^2\alpha_n^{\delta/(\delta+4)}(k) \le K,\ \forall n,$
where Δ, K *are finite constants, independent of* n, k *or* a. *Then*

$$T_{n,d_n,1} \xrightarrow{d} N(0,\sigma^2),\ \ i.e.,\ d_n^{-1/2}\sum_{i=1}^{dn} X_{n,i} \to N(0,\sigma^2).$$

Proof. For each row n, and given lengths b_n, l_n, define

$$\begin{aligned}
U_{n,i} &= \sum_{k=1}^{b_n} X_{n,(i-1)(b_n+l_n)+k}, \quad 1 \le i \le r_n \\
V_{n,i} &= \sum_{k=b_n+1}^{b_n+l_n} X_{n,(i-1)(b_n+l_n)+k}, \quad 1 \le i \le r_n - 1; \\
V_{n,r_n} &= X_{n,(r_n-1)(b_n+l_n)+1} + \cdots + X_{n,d_n},
\end{aligned}$$

where r_n is the greatest integer i so that $(i-1)(b_n+l_n)+b_n < d_n$. Then $S_{n,d_n,1} = \sum_{i=1}^{r_n} U_{n,i} + \sum_{i=1}^{r_n} V_{n,i}$. Note that representing $S_{n,d_n,1}$ in this way, the indices of the sum is splitted into alternating blocks of lengths b_n and l_n. We want to choose l_n small enough so that $d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i} \xrightarrow{p} 0$ but big enough so that $d_n^{-1/2}\sum_{i=1}^{r_n} U_{n,i}$ can be approximated by a sum of independent random variables also normalized by $d_n^{-1/2}$.

Choose $b_n = \lfloor d_n^{3/4} \rfloor$ and $l_n = \lfloor d_n^{3/4} \rfloor$. Observing that r_n is $\lfloor (d_n - b_n)/(b_n + l_n) \rfloor$ or $\lfloor (d_n - b_n)/(b_n + l_n) \rfloor + 1$, we have the asymptotic equivalences[3]: $b_n \sim d_n^{3/4}, l_n \sim d_n^{1/4}$ and $r_n \sim d_n^{1/4}$.

Firstly, we show that $d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i} \xrightarrow{p} 0$ as $n \to \infty$. Since its expectation is zero, it is sufficient to prove that the variance vanishes. Under assumptions *A.1* and *A.3*, note that Lemma 3.13 implies[4]

$$\begin{aligned}
E|V_{n,i}|^2 = E\left|\sum_{k=1}^{l_n} X_{n,(i-1)(b_n+l_n)+b_n+k}\right|^2 &\le B(2,\delta)D(2,\delta,l_n) \\
&\le 18\max\{1, C(2,\delta)\}L(2,\delta,l_n) \le 18Kl_n\Delta^2 \coloneqq Cl_n.
\end{aligned}$$

Then the application of Minkowski's inequality $r_n - 1$ times gives

$$\left[\mathrm{Var}\left(d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i}\right)\right]^{1/2} = \left[\int\left(\sum_{i=1}^{r_n} d_n^{-1/2}V_{n,i}\right)^2 dP\right]^{1/2} \le \sum_{i=1}^{r_n}\left[\int\left(d_n^{-1/2}V_{n,i}\right)^2 dP\right]^{1/2}$$

[3]Since $d_n - b_n \sim d_n$ and $b_n + l_n \sim d_n^{3/4}$, we have $\lim_{n\to\infty} \frac{d_n^{3/4}(d_n-b_n)}{d_n(b_n+l_n)} = 1$.

[4]For every $k \ge 0$, it holds that $1 \le (k+1)^2$ and $\alpha_n^{\delta/(2+\delta)}(k) \le \alpha_n^{\delta/(4+\delta)}(k)$, with the observation that $0 \le \alpha_n \le 1/4, \forall n$. Thus $C(2,\delta) \le C(4,\delta) \le K$. Further, note that when $i = r_n$, V_{n,r_n} is the sum of at most $l_n + b_n$ terms. In this case, $E|V_{n,r_n}|^2 \le C(b_n + l_n)$, for some $C > 0$.

$$
\begin{aligned}
&= d_n^{-1/2}\left\{\sum_{i=1}^{r_n-1}[E(V_{n,i}^2)]^{1/2} + [E(V_{r_n,i}^2)]^{1/2}\right\}\\
&\le d_n^{-1/2}\left\{r_n(Cl_n)^{1/2} + [C(b_n+l_n)]^{1/2}\right\} = d_n^{-1/2}O(d_n^{3/8})\\
&= O(d_n^{-1/8}) = O(o(1)) = o(1)
\end{aligned}
$$

From Chebychev's inequality, $P(|d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i}| > \omega) \le \omega^{-2}\operatorname{Var}(d_n^{-1/2}\sum_{i=1}^{r_n} V_{n,i})$ for any $\omega > 0$. Taking the limit over n gives the convergence in probability.

Let $U'_{n,i}, 1 \le i \le r_n$, be independent random variables so that $U'_{n,i}$ has the same distribution as $U_{n,i}$, for each $1 \le i \le r_n$. Define the sums $F_{n,k} = d_n^{-1/2}\sum_{j=1}^{k} U_{n,i}$ and $F'_{n,k} = d_n^{-1/2}\sum_{j=1}^{k} U'_{n,i}$, and their characteristic functions $\varphi_{F_{n,k}}, \varphi_{F'_{n,k}}$, respectively. Then, for any t and n,

$$
\begin{aligned}
\left|\varphi_{F_{n,r_n}}(t) - \varphi_{F'_{n,r_n}}(t)\right| &= \left|Ee^{itF_{n,r_n}} - Ee^{itF'_{n,r_n}}\right|\\
&= \Big|Ee^{itF_{n,r_n}} - Ee^{itF'_{n,r_n}} \pm Ee^{itF_{n,r_n-1}}Ee^{it\frac{U'_{n,r_n}}{\sqrt{d_n}}}\\
&\pm Ee^{itF_{n,r_n-2}}\Pi_{j=r_n-1}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}} \pm \cdots \pm Ee^{itF_{n,2}}\Pi_{j=3}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}}\Big|\\
&= \Big|Ee^{itF_{n,r_n}} - Ee^{itF'_{n,r_n}} + Ee^{itF_{n,r_n-1}}Ee^{it\frac{U'_{n,r_n}}{\sqrt{d_n}}} - Ee^{itF_{n,r_n-1}}Ee^{it\frac{U_{n,r_n}}{\sqrt{d_n}}}\\
&+ Ee^{itF_{n,r_n-2}}\Pi_{j=r_n-1}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}} - Ee^{itF_{n,r_n-2}}Ee^{it\frac{U_{n,r_n}}{\sqrt{d_n}}}\Pi_{j=r_n}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}}\\
&\pm \cdots\\
&+ Ee^{itF_{n,2}}\Pi_{j=3}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}} - Ee^{itF_{n,2}}Ee^{it\frac{U_{n,3}}{\sqrt{d_n}}}\Pi_{j=4}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}}\Big|\\
&= \Big|\sum_{k=1}^{r_n-2}\big(Ee^{itF_{n,k+1}} - Ee^{itF_{n,k}}Ee^{it\frac{U_{n,k+1}}{\sqrt{d_n}}}\big)\Pi_{j=k+2}^{r_n}Ee^{it\frac{U'_{n,j}}{\sqrt{d_n}}}\\
&+ Ee^{itF_{n,r_n}} - Ee^{itF_{n,r_n-1}}Ee^{it\frac{U_{n,r_n}}{\sqrt{d_n}}}\Big|\\
&\le \left|\sum_{k=1}^{r_n-1}\big(Ee^{itF_{n,k+1}} - Ee^{itF_{n,k}}Ee^{it\frac{U_{n,k+1}}{\sqrt{d_n}}}\big)\right| \le \sum_{k=2}^{r_n}\left|\varphi_{F_{n,k}} - \varphi_{F'_{n,k}}\right|\\
&\le r_n(16\alpha_n(l_n)),
\end{aligned}
$$

using Lemma 3.12 (Ibragimov's Bound) and the fact that $|\varphi_{U_{n,k}}| \le 1,\ \forall k = 1, \ldots, r_n$. Purposely choose $\alpha_n(k) \le K/k^2$, which allows the mixing coefficient to decrease slow enough so that condition $A.3$ is violated. Then

$$16r_n\alpha_n(l_n) \le 16r_nK/l_n^2 = O(d_n^{-1/4}) = o(1).$$

Since $\alpha_n(k)$ has to be strictly less that K/k^2 (i.e., decrease at a faster rate) to satisfy condition $A.3$, then $\sup_t\left|\varphi_{F_{n,r_n}}(t) - \varphi_{F'_{n,r_n}}(t)\right| \to 0$ as $n \to \infty$. Hence, $F_{n,r_n} \xrightarrow{d} F'_{n,r_n}, n \to$

∞.

We now show that $F'_{n,r_n} \xrightarrow{d} N(0,\sigma^2)$. Write

$$\begin{aligned}\frac{1}{r_nb_n}\operatorname{Var}\left(\sum_{i=1}^{r_n} U'_{n,i}\right) &= \frac{1}{r_nb_n}\sum_{i=1}^{r_n}\operatorname{Var}(U'_{n,i}) = \frac{1}{r_nb_n}\sum_{i=1}^{r_n}E({U'_{n,i}}^2)\\ &= \frac{1}{r_n}\sum_{i=1}^{r_n}E[(b_n^{-1/2}U_{n,i})^2] = \frac{1}{r_n}\sum_{i=1}^{r_n}E\left[\left(\frac{1}{b_n^{1/2}}\sum_{k=1}^{b_n}X_{n,(i-1)(b_n+l_n)+k}\right)^2\right].\end{aligned}$$

Condition A.2 implies that $\operatorname{Var}(b_n^{-1/2}U_{n,i}) \to \sigma^2$ uniformly in i. Then

$$\begin{aligned}\left|\frac{1}{r_n}\sum_{i=1}^{r_n}\operatorname{Var}(b_n^{-1/2}U_{n,i}) - \sigma^2\right| &= \left|\frac{1}{r_n}\sum_{i=1}^{r_n}\left[\operatorname{Var}(b_n^{-1/2}U_{n,i}) - \sigma^2\right]\right|\\ &\le \frac{1}{r_n}\sum_{i=1}^{r_n}\left|\operatorname{Var}(b_n^{-1/2}U_{n,i}) - \sigma^2\right|\\ &\le \sup_i\left|\operatorname{Var}(b_n^{-1/2}U_{n,i}) - \sigma^2\right| \to 0 \qquad (3.80)\end{aligned}$$

Assume $K \ge 1$ without loss of generality. From Lemma 3.13,

$$\frac{1}{b_n^{(2+\delta)/2}}E|U'_{n,i}|^{2+\delta} \le \{3024\max[1,K^2]\}2^{4[1+4(2-\delta)/\delta]}\Delta^{(2+\delta)(1+\delta/2)} := C \qquad (3.81)$$

Combining (3.80) and (3.81), we have

$$\begin{aligned}\left[\operatorname{Var}\left(\sum_{k=1}^{r_n}U'_{n,i}\right)\right]^{-(2+\delta)/2}\sum_{i=1}^{r_n}E|U'_{n,i}|^{2+\delta} &= r_n^{-\frac{2+\delta}{2}}\left[\frac{1}{r_nb_n}\operatorname{Var}\left(\sum_{k=1}^{r_n}U'_{n,i}\right)\right]^{-\frac{2+\delta}{2}}\sum_{i=1}^{r_n}E\left|\frac{U'_{n,i}}{b_n^{1/2}}\right|^{2+\delta}\\ &\le r_n^{-\frac{2+\delta}{2}}O(1)O(r_n) = O(r_n^{-\delta/2}) = o(1).\end{aligned}$$

Since the Lyapunov's condition is satisfied, we use Theorem 3.4 to obtain that

$$\frac{\sum_{i=1}^{r_n}U'_{n,i}}{Var(\sum_{i=1}^{r_n}U'_{n,i})^{1/2}} \xrightarrow{d} N(0,1). \qquad (3.82)$$

As showed in (3.80), $(r_nb_n)^{-1}\operatorname{Var}(\sum_{k=1}^{r_n}U'_{n,i}) \to \sigma^2$, but $r_nb_n \sim d_n$. Therefore

$$\lim_{n\to\infty}\frac{1}{d_n}\operatorname{Var}(\sum_{k=1}^{r_n}U'_{n,i}) = \lim_{n\to\infty}\frac{r_nb_n}{d_n}\lim_{n\to\infty}\frac{1}{r_nb_n}\operatorname{Var}(\sum_{k=1}^{r_n}U'_{n,i}) = \sigma^2,$$

and hence,

$$\sqrt{\frac{\operatorname{Var}(\sum_{k=1}^{r_n}U'_{n,i})}{d_n}} \to \sigma. \qquad (3.83)$$

For simplicity, denote $Y_n = \sum_{i=1}^{r_n}U_{n,i}/\sqrt{s_n}$ and $b_n = \sqrt{s_n/d_n}$, where $s_n = \operatorname{Var}(\sum_{k=1}^{r_n}U'_{n,i})$, and consider $Y \sim N(0,1)$. Assume $Y_n \xrightarrow{d} Y$ and $b_n \to \sigma$ to represent expressions (3.82) and (3.83), respectively. It follows that $\forall x \in \mathbb{R} : \forall \epsilon_1, \epsilon_2 > 0 : \exists N_1, N_2 \in \mathbb{N} : n \geq N_1 \implies |P(Y_n \leq x/b_n) - P(Y \leq x/b_n)| < \epsilon_1/2$ and $n \geq N_2 \implies |P(Y \leq x/b_n) - P(Y \leq x/\sigma)| < \epsilon_2/2$, since the distribution function of Y is continuous. In particular, for $\epsilon_1 = \epsilon_2$,

$$\begin{aligned} |P(b_nY_n \leq x) - P(bY \leq x)| &\leq |P(Y_n \leq x/b_n) - P(Y \leq x/b_n)| \\ &+ |P(Y \leq x/b_n) - P(Y \leq x/\sigma)| \\ &\leq \epsilon_1 \end{aligned}$$

Thus $b_nY_n \xrightarrow{d} \sigma Y$. That is,

$$\sqrt{\frac{\mathrm{Var}(\sum_{k=1}^{r_n} U'_{n,i})}{d_n}} \frac{\sum_{i=1}^{r_n} U'_{n,i}}{Var(\sum_{i=1}^{r_n} U'_{n,i})^{1/2}} = \frac{\sum_{i=1}^{r_n} U'_{n,i}}{d_n^{1/2}} \xrightarrow{d} \sigma N(0,1). \tag{3.84}$$

Since $d_n^{-1/2} \sum_{i=1}^{r_n} V_{n,i} \xrightarrow{p} 0$ and (3.84) hold, the application of Slutsky's theorem gives

$$\frac{S_{n,d_n,1}}{d_n^{1/2}} = \frac{\sum_{i=1}^{r_n} U_{n,i} + \sum_{i=1}^{r_n} V_{n,i}}{d_n^{1/2}} \xrightarrow{d} N(0,\sigma^2).$$

□

The next theorem is due to Ekström (2014) who provided a more general CLT without imposing the condition A.2 of Politi's CLT. Belyaev and Sjöstedt-de Luna (2000) introduced the notion of *weakly approaching sequences of distributions*, generalizing the concept of weak convergence of distributions without the need to have a limiting distribution. Two sequences of distribution laws $\{\mathcal{L}(Y_n)\}$ and $\{\mathcal{L}(X_n)\}$ of random variables $\{Y_n\}$ and $\{X_n\}$, respectively, are said to *weakly approach each other* if for any bounded continuous function f, we have $E(f(Y_n)) - E(f(X_n)) \to 0$ as $n \to \infty$, and we write $\mathcal{L}(Y_n) \overset{w.a.}{\longleftrightarrow} \mathcal{L}(X_n),\ n \to \infty$.

Theorem 3.6 (Politis-Ekstrom's CLT)**.** *Let $\{X_{n,i} : 1 \leq i \leq d_n\}$ be a triangular array of mean zero random variables and consider the notations of Theorem 3.5. If conditions A.1 and A.3 of Theorem 3.5 hold true, then*

$$\mathcal{L}(T_{n,d_n,1}) \overset{w.a.}{\longleftrightarrow} N(0, \sigma^2_{n,d_n,1}).$$

Corollary 3.6.1. *Under the assumptions of Theorem 3.6, if we additionaly have $\sigma^2_{n,d_n,1} \to \sigma^2$ as $n \to \infty$, then $T_{n,d_n,1} \xrightarrow{d} N(0,\sigma^2)$.*

Proof. Let $\{X_n\} : X_n \sim N(0, \sigma^2_{n,d_n,1})$ and $Y_n = T_{n,d_n,1}$ for every n. Denote $M_n(t)$ as the moment generating function associated to X_n for each n. Then $\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty} e^{\sigma^2_{n,d_n,1} t^2/2} = e^{\sigma^2 t^2/2}$, for all $t \in \mathbb{R}$, by hypothesis. Therefore $X_n \xrightarrow{d} X$ with $X \sim N(0,\sigma^2)$, from Theorem 3 of Curtiss (1942). Using portmanteau's Lemma (VAART, 1998, p. 6) we have that for all bounded continuous function f,

$$\lim_{n\to\infty} E(f(Y_n)) - E(f(X)) = \lim_{n\to\infty} E(f(Y_n)) - E(f(X_n)) + \lim_{n\to\infty} E(f(X_n)) - E(f(X)) = 0$$

implying $Y_n \overset{d}{\to} X$, as desired. $\square$

To illustrate the applicability of the Politis-Ekstrom's CLT, consider the triangular array $\{X_{t,T}\} = \{X_{t,T} I(t \in J_{x,T})\}$ in (3.18) where the set $J_{x,T} = \{t \in [T] : t/T \in (x - h, x + h)\}$, for some $x \in (h, 1-h)$, has cardinality k_T. For each T, the smallest element of $J_{x,T} \subseteq [T]$ does not need to be 1. It does not mean, however, that we cannot use the CLT. If $\{\epsilon_{t,T}\}$ satisfies the conditions of Theorem 3.6, so does $\{X_{t,T}\}$ by Lemma 3.11. Since, for each T, $J_{x,T}$ is a finite set, then there is a bijection $f_T : J_{x,T} \to \{1, \dots, \lfloor k_T \rfloor\}$ which enables us to treat the array $\{X_{t,T} I(t \in J_{x,T})\} = \{X_{t,T}, t \in J_{x,T}, T \geq 1\}$ equivalently as $\{Z_{i,T}, 1 \leq i \leq \lfloor k_T \rfloor, T \geq 1\}$ where $Z_{i,T} \coloneqq X_{f_T(t),T}$.

Appendix E - A note on the proof of Vogt and Linton

In the proof of Lemma A4 in the supplementary material of Vogt and Linton (2014), a similar bound problem to that of inequality (3.72), in this study, appeared. An upper bound for a sum of products of four random variables was required in order to prove the lemma and, ultimately, to prove the consistency of the period estimator. For this, they introduced the concept of *separated* indices. In their words, "we say that an index i_1 is separated from the indices $i_2, \ldots, i_d$ if $|i_1 - i_k| > C_2 \log T$ for a sufficiently large constant C_2 and all $k = 2, \ldots, d$". This concept allowed the authors to split the following summation as

$$\sum_{l,l'=1}^{T} \sum_{k=1}^{K_{l_\theta,T}^{\lfloor\theta\rfloor}} \sum_{k'=1}^{K_{l_\theta,T}^{'\lfloor\theta\rfloor}} E[\epsilon_{(k-1)\theta+l_\theta}\epsilon_l\epsilon_{(k'-1)\theta+l'_\theta}\epsilon_{l'}] = \sum_{(l,l',k,k')\in\Gamma} E[\epsilon_{(k-1)\theta+l_\theta}\epsilon_l\epsilon_{(k'-1)\theta+l'_\theta}\epsilon_{l'}] + \sum_{(l,l',k,k')\in\Gamma^c} E[\epsilon_{(k-1)\theta+l_\theta}\epsilon_l\epsilon_{(k'-1)\theta+l'_\theta}\epsilon_{l'}],$$

where "Γ is the set of tuples (l, l', k, k') such that none of the indices $l, l', (k-1)\theta, (k'-1)\theta + l'_\theta$ is separated from the others and Γ^c is its complement" and $i_\theta := w_{\theta,i}$ in our notation. After bounding the sum over Γ, the sum over its complement Γ^c was bounded using the argument that "for any tuple $(l, l', k, k') \in \Gamma^c$, there exists an index, say l, which is separated from the others". However, such set of indices is only a proper subset of Γ^c.

To make our argument clearer, let us give an explicit definition of Γ based on Vogt and Linton (2014):

$$\begin{aligned}\Gamma &= \{(l, l', k, k') \in [T]^4 : \text{none of } l, l', (k-1)\theta + l_\theta, (k'-1)\theta + l'_\theta \\ &\qquad \text{is separated from the others}\} \\ &= \{(l, l', k, k') \in [T]^4 : \text{every index } l, l', (k-1)\theta + l_\theta, (k'-1)\theta + l'_\theta \\ &\qquad \text{is not separated from the others}\} \\ &= \{(i_1, i_2, i_3, i_4) \in [T]^4 : \forall j \in [4] : \forall k \in [4] \setminus \{j\} : |f(i_j) - f(i_k)| \leq C_2 \log T\}\end{aligned}$$

where

$$f(i_j) = \begin{cases} i_j & \text{,if } j \in \{1, 2\} \\ (i_j - 1)\theta + i_1 - \lfloor (i_1 - 1)/\theta \rfloor & \text{,if } j = 3 \\ (i_j - 1)\theta + i_2 - \lfloor (i_2 - 1)/\theta \rfloor & \text{,if } j = 4 \end{cases}.$$

Therefore, its complement is given by

$$\Gamma^c = \{(i_1, i_2, i_3, i_4) \in [T]^4 : \exists j \in [4] : \exists k \in [4] \setminus \{j\} : |f(i_j) - f(i_k)| > C_2 \log T\}.$$

Note that Vogt and Linton considered the set

$$\{(i_1, i_2, i_3, i_4) \in [T]^4 : \exists j \in [4] : \forall k \in [4] \setminus \{j\} : |f(i_j) - f(i_k)| > C_2 \log T\},$$

which is a proper subset of Γ^c, and is not sufficient for the proof.

As an example, assume $\{\epsilon_i\}$ i.i.d. (and thus, strongly mixing) with mean zero and finite variance. The tuple (l, l, k, k) with $l = 1$ and $k = C_2 \log T/\theta + 2$ is in Γ^c for T large enough. But $\mathrm{Cov}(\epsilon_l, \epsilon_l \epsilon_{1+\theta+C_2 \log T} \epsilon_{1+\theta+C_2 \log T}) = E(\epsilon_l^2) E(\epsilon_{1+\theta+C_2 \log T}^2) = [E(\epsilon_l^2)]^2 = C < \infty$, by the hypotheses. Hence, the argument of Vogt and Linton that for any $(l, l', k, k') \in \Gamma^c$, $\mathrm{Cov}(\epsilon_l, \epsilon_{l'} \epsilon_{(k-1)\theta+l_\theta} \epsilon_{(k'-1)\theta+l'_\theta}) < CT^{-C_3}$ for arbitrarily large $C_3 > 0$ does not hold.

4 NONPARAMETRIC ESTIMATION OF A SMOOTH TREND IN THE PRESENCE OF A PERIODIC SEQUENCE: FINITE SAMPLE BEHAVIOR AND APPLICATIONS

Abstract. We investigate the finite sample behavior of the estimators obtained by reversing the procedure of Vogt and Linton (2014). We suggest a plug-in type bandwidth for the trend estimator. Our simulations showed a good performance for the suggested bandwidth selector and a fairly robust behavior of the period estimator over different bandwidths. We complement the study with two applications: one in global temperature data and the other in the estimation of the non-accelerating inflation rate of unemployment.

Keywords: Nonparametric regression. Asymptotic analysis. Monte Carlo Simulation.

JEL Codes. C14; C15; C22.

4.1 Introduction

Vogt and Linton (2014) proposed a three-step procedure to estimate a trend function in the presence of a periodic sequence. In the previous section, we showed some asymptotic properties of the estimators derived by reversing their original estimation procedure. Desirable large sample properties, such as (uniform) consistency and asymptotic normality was proved. In practice, we may be interested in assessing the finite sample behaviour of such estimators. As these estimators depend on a bandwidth choice, we must have a suitable bandwidth selection criteria.

In the present section, we investigate the finite sample behaviour of the estimators involved in the reversed estimation procedure of Vogt and Linton (2014). A plug-in type bandwidth is proposed in order to estimate the trend function, in the first step. Our simulation exercise showed a good performance for the proposed bandwidth. Although we do not provide an optimal bandwidth selection for the period estimator, we employ a simulation exercise to evaluate the sensitivity of the estimator for different bandwidth choices having the plug-in bandwidth, used in the first step, 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 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 estudied so far.

4.2 Bandwidth selection for the trend estimator

The local polynomial regression requires the choice of a bandwidth parameter. The bandwidth selection is usually done by a *cross-validation* algorithm or a *plug-in* method (see Wand and Jones, 1994; Fan and Gijbels, 1996). In this section we focus on a plug-in type bandwidth based on minimizing the Mean Integrated Squared Error (MISE) for the trend estimator $\hat{g}$, defined by (3.5) in Chapter 3.

Consider the model (3.1) of Chapter 3: for any $T \in \mathbb{N}$, $\{Y_{t,T} : t = 1, \dots, T\}$ follows

$$Y_{t,T} = g(t/T) + m(t) + \epsilon_{t,T}, \quad t \in \{1, \dots, T\}, \tag{4.1}$$

where $g : [0, 1] \to \mathbb{R}$ is a deterministic trend function, m is a θ_0-periodic real sequence and $\{\epsilon_{t,T}\}_{t=1}^{T}$ is a zero mean random sequence. Assume again that g and m are normalized so that $\sum_{t=1}^{\theta_0} m(t) = 0$. In matrix form, $Y_T = g_T + m_T + \epsilon_T$ where $Y_T = (Y_{1,T}, \dots, Y_{T,T})'$, $g_T = (g(1/T), \dots, g(T/T))'$, $m_T = (m(1), \dots, m(T))'$ and $\epsilon_T = (\epsilon_{1,T}, \dots, \epsilon_{T,T})'$.

Define the naive local linear estimator for $B_T := (g(x), g'(x))'$ as

$$\hat{B}_T(x) := \begin{bmatrix} \hat{g}(x) \\ \hat{g}'(x) \end{bmatrix} = (A_T' W_T A_T)^{-1}(A_T' W_T Y_T) := S_T^{-1} D_T,$$

where

$$A_T = \begin{bmatrix} 1 & (x_1 - x) \\ \vdots & \vdots \\ 1 & (x_T - x) \end{bmatrix}, \; S_T = \begin{bmatrix} s_0 & s_1 \\ s_1 & s_2 \end{bmatrix}, \; W_T = \frac{1}{T}\operatorname{diag}(K_h(x_1 - x), \dots, K_h(x_T - x)),$$
$$D_T = (d_0, d_1)',$$

with

$$s_k = \frac{1}{T}\sum_{t=1}^{T}(x_t - x)^k K_h(x_t - x), \; k \in \mathbb{N},$$
$$d_k = \frac{1}{T}\sum_{t=1}^{T}(x_t - x)^k K_h(x_t - x) Y_{t,T}, \; k \in \mathbb{N},$$

$K_h(u) = K(u/h)/h$ and $x_i = i/T$. For simplicity, the dependence of the matrices S_T and D_T on $x \in [0,1]$ and the dependence of the design points on T were omitted.

Denote the covariance matrix of the errors ϵ_T by Γ_T. The exact Mean Squared Error (MSE) of the naive trend estimator in (3.5) is given by

$$MSE(x,h) = b_T^2(x) + V_T(x), \tag{4.2}$$

where

$$b_T(x) = e_1' S_T^{-1} A_T' W_T (g_T - A_T' B_T(x) + m_T)$$

and

$$V_T(x) = e_1' S_T^{-1} A_T' W_T \Gamma_T W_T A_T S_T^{-1} e_1.$$

Following the ideas of Fan et al. (1996) and Fernández and Fernández (2001), we approximate the bias by a 2nd order Taylor expansion,

$$b_T(x) \approx b_T^*(x) = e_1' S_T^{-1} \begin{bmatrix} s_2 \\ s_3 \end{bmatrix} g_T''(x)/2 + e_1' S_T^{-1} A_T' W_T m_T. \tag{4.3}$$

Given appropriate estimators for g'', θ_0, m and Γ_T, an estimate $\widehat{MSE}(x,h)$ of (4.2) is obtined. Define the estimator for the Mean Integrated Squared Errors (MISE) by means of right Riemann sums approximation:

$$\widehat{MISE}(h) = \int \widehat{MSE}(x,h)dx \approx \frac{1}{T}\sum_{i=1}^{T} \widehat{MSE}(i/T,h). \tag{4.4}$$

Then the plug-in bandwidth is selected from

$$h_{\text{opt}} = \arg\min_{h} \frac{1}{T}\sum_{i=1}^{T} \widehat{MSE}(i/T,h). \tag{4.5}$$

Since the first $\tilde{\theta}$ points of the periodic sequence estimator $\tilde{\beta}$ defined in (3.12) do not necessarily sum zero, we heuristically propose the selector h^*_{opt} which is the particular case of (4.5) that uses $\tilde{\beta} - 1'_{\tilde{\theta}}\tilde{\beta}$ as the estimator of m_T.

In the next sections, the estimation procedures will be carried out with the Epanechnikov kernel.

4.2.1 Simulation: plug-in bandwidth performance

In this section we analyze the finite sample performance of the proposed bandwidth selector via a Monte Carlo experiment. The data generating process is the same as that of Section 6 of Vogt and Linton (2014). Model (4.1) is simulated with

$$m(t) = \sin\Big(\frac{2\pi}{\theta_0}t + \frac{3\pi}{2}\Big); \quad g(u) = 2u^2; \quad \epsilon_t = 0.45\epsilon_{t-1} + \eta_t,$$

$\theta_0 = 60$ and $\eta_t \overset{i.i.d.}{\sim} N(0, \sigma^2_\eta)$. To achieve strict stationarity, assume $\epsilon_0 \sim N(0, \sigma^2_\eta/(1 - 0.45^2))$. In a first step, we approximate the MISE between the trend function g and the estimator $\hat{g}$,

$$MISE(h) = E\int (\hat{g}(x;h) - g(x))^2 dx, \tag{4.6}$$

on a grid of equally spaced bandwidth values consisted of 300 points from 0.1 to 1, by means of Riemann sums and through 500 Monte Carlo simulations. Obviously, the function g in (4.6) is assumed to be known in order to make the computation feasible. By minimizing the approximated MISE in h, we obtain a numerical approximation of $h_{\min} = \arg\min_h MISE(h)$. This is done for the sample sizes $T \in \{160, 250, 500\}$ and for the error variances $\sigma^2_\eta \in \{0.2, 0.4, 0.6\}$.

In the second step, another 500 random samples are generated and the selector h_{opt} is computed for every sample. We perform Monte Carlo approximations once more to calculate the expected value and the standard deviation of h_{opt} as well as the MSE between $MISE(h_{\text{opt}})$ and $MISE(h_{\min})$, denoted as $\Delta M(h_{\text{opt}})$, which will serve as an efficiency measurement. That is, $\Delta M(h_{\text{opt}}) = E(MISE(h_{\text{opt}}) - MISE(h_{\min}))^2$. This exercise is done for every choice of sample sizes and error variances mentioned in the first step and is replicated to the selector h^*_{opt}.

Now, we describe the computation of the bandwidth h_{opt}. With the help of the pilot bandwidth, $h_{\text{pilot}} = 0.5$, we estimate g, θ_0 and m using the estimators $\hat{g}, \tilde{\theta}$ and $\tilde{m}$ proposed in Chapter 3. The exact procedure to estimate the period will be detailed in the next section. With $\tilde{m}$ and $\hat{g}$ in hand, we can calculate the residuals $\hat{\epsilon}_t = Y_t - \tilde{m}(t) - \hat{g}(t/T), t \in \{1, \ldots, T\}$, which will be used to estimate the covariance matrix Γ_T of the first order autoregressive errors. The natural estimator for Γ_T has the structure

$$\hat{\Gamma}_T = \hat{\sigma}^2_\epsilon \begin{bmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \cdots & \hat{\rho}^{T-1} \\ \hat{\rho} & 1 & \hat{\rho} & \cdots & \hat{\rho}^{T-2} \\ \hat{\rho}^2 & \hat{\rho} & 1 & \cdots & \hat{\rho}^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}^{T-1} & \hat{\rho}^{T-2} & \hat{\rho}^{T-3} & \cdots & 1 \end{bmatrix}. \tag{4.7}$$

with

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^{T} \hat{\epsilon}_t} \quad \text{and} \quad \hat{\sigma}^2_\epsilon = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}^2_t. \tag{4.8}$$

In (4.7), $\hat{\sigma}^2_\epsilon$ and $\hat{\rho}^j$ are estimators for the variance and the j-order autocorrelation of $\{\epsilon_t\}$, respectively. The integral of the functional derivative g'' is estimated by fitting a second order polynomial to g, globally, through the parametric fit $Y_t - \tilde{m}(t) = \hat{\alpha}_1(t/T) + \hat{\alpha}_2(t/T)^2$ with $\hat{\alpha}$ being the generalized least squares estimate associated to the matrix (4.7). The resulting estimator is defined by $\hat{g}'' = 2\hat{\alpha}$. Since $\hat{g}''(t/T) = 2\hat{\alpha}$ is constant for $t \in \{1, \ldots, T\}$, we have $\widehat{\int g''^2} = (2\hat{\alpha})^2$ by means of Riemann sum approximation. This simple procedure (known as "Rule-of-thumb") is somewhat crude but requires little programming.

The heuristic selector h^*_{opt} is computed analogously except that periodic sequence m is estimated via $\tilde{\beta} - 1'_{\tilde{\theta}}\tilde{\beta}$.

Table 1 shows that h_{opt} and h^*_{opt} performed well, specially for the sample sizes 160 and 500. We can also see that their efficiencies worsen as the error's variance gets bigger.

4.3 Sensitivity of the period estimator over bandwidths

In this section, we analyze the finite sample behavior of the period estimator along a set of different bandwidth values using Monte Carlo experiments.

We follow the heuristic procedure proposed by Vogt and Linton (2014), for selecting the regularization parameter of the period estimator. Consider the simple model without trend

$$S_{t,T} = m(t) + \epsilon_{t,T}, \tag{4.9}$$

where $\{\epsilon_{t,T}\}_{t=1}^T$ have the same joint distribution as $\{u_t\}_{t=1}^T$ with $\{u_t : t \in \mathbb{Z}\}$ being a sequence of independent and identically distributed zero mean random variables which

Table 1: Plug-in bandwidths

	$h_{\min}$	Mean h^*_{opt}	Mean h_{opt}	St. Dev. h^*_{opt}	St. Dev. h_{opt}	$\Delta M \times 10^6$ h^*_{opt}	$\Delta M \times 10^6$ h_{opt}
$T = 160$							
$\sigma^2_\epsilon = 0.25$	0.90	0.91	0.81	0.08	0.09	2.41	9.35
$\sigma^2_\epsilon = 0.5$	0.91	0.88	0.80	0.12	0.12	9.23	18.39
$\sigma^2_\epsilon = 0.75$	0.92	0.86	0.79	0.14	0.13	33.62	37.23
$T = 250$							
$\sigma^2_\epsilon = 0.25$	0.58	0.64	0.64	0.16	0.15	1.68	1.69
$\sigma^2_\epsilon = 0.5$	0.58	0.68	0.69	0.16	0.16	2.77	2.48
$\sigma^2_\epsilon = 0.75$	0.59	0.69	0.70	0.18	0.17	5.62	5.27
$T = 500$							
$\sigma^2_\epsilon = 0.25$	0.32	0.34	0.39	0.05	0.05	0.09	0.14
$\sigma^2_\epsilon = 0.5$	0.42	0.40	0.44	0.09	0.07	0.51	0.53
$\sigma^2_\epsilon = 0.75$	0.43	0.47	0.50	0.14	0.12	1.89	1.76
$T = 800$							
$\sigma^2_\epsilon = 0.25$	0.28	0.31	0.36	0.03	0.03	0.00	0.08
$\sigma^2_\epsilon = 0.5$	0.34	0.35	0.40	0.04	0.04	0.11	0.30
$\sigma^2_\epsilon = 0.75$	0.35	0.38	0.43	0.07	0.06	0.46	0.62

* The table presents the expectation, standard deviation and the efficiency measurement associated with each bandwidth selector. Here, $\Delta M(\hat{h}) = E(MISE(\hat{h}) - MISE(h_{\min}))^2$.

also has finite variance. As showed by Vogt and Linton (2014), when $\theta = r\theta_0$ for some integer r, it holds that $E\{RSS(r\theta_0)\} + \sigma^2 r\theta_0 = E\{RSS(\theta_0)\} + \sigma^2\theta_0$. This suggests to choose the penalization parameter λ_T larger than σ^2 in order to avoid choosing multiples of θ_0. However, the penalization should not be too large otherwise the criterion function Q at θ_0, defined in (3.8), becomes larger than the criterion function at $\theta = 1$. From this heuristics, they proposed to choose the regularization parameter as

$$\lambda_T = \sigma^2 k_T, \tag{4.10}$$

where k_T is a slowly divergent sequence and $\sigma^2 = E(u^2) < \infty$. To meet the conditions of Theorem 3.2, k_T should grow slightly faster than $T^{1/4}$.

Given a bandwidth choice, let $\tilde{S} = (\tilde{S}_{1,T}, \ldots, \tilde{S}_{T,T})' = Y - \hat{g}$ be the estimted data obtained from the naive trend estimates and let $\breve{\beta}_\theta = (X'_\theta X_\theta)^{-1} X'_\theta \tilde{S}$. Since σ^2 is unknown in (4.10), it is replaced by the standard estimator

$$\breve{\sigma}^2 = \frac{\sum_{i=1}^T \breve{\epsilon}^2_{t,T}}{T} \tag{4.11}$$

where $\breve{\epsilon} = \tilde{S} - \breve{m}$ with $\breve{m} = \breve{\beta}_{\breve{\theta}}$ and $\breve{\theta} = \arg\min_{1 \le \theta \le \Theta_T} \|\tilde{S} - X_\theta \breve{\beta}_\theta\|$. As noted by the authors, although $\breve{\theta}$ is an inconsistent estimator of θ_0, it equals $r\theta_0$ for some $r \in \mathbb{N}$ with probability approaching to one . Since multiples of θ_0 are also periods of m, we could use $\check{m}$ as a preliminary estimator of m in order to calculate the residuals.

The model used in the simulation exercise is exactly the same as that in Section 1.2. To be comparable with the results of Vogt and Linton (2014), we perform 1000 simulations for the sample sizes $T = \{160, 250, 500\}$ and for $\sigma_\eta^2 \in \{0.2, 0.4, 0.6\}$. We choose three different bandwidths values based on h_{opt} defined in the previous section: $h_{\text{opt}}^{1.5}$, $h_{\text{opt}}^{5/4}$, $h_{\text{opt}}^{0.5}$ and h_{opt} itself. The plug-in bandwidth h_{opt}, for each case, is considered to be its expected value, obtained in the simulation of Section 1.2 (which is presented in the third column of Table 1). Note that $h_{\text{opt}}^{5/4}$ is a choice satisfying Theorem 3.2 since the bandwidth is assumed to be $\Theta(T^{-1/4})$ and h_{opt} is of order $\Theta(T^{-1/5})$, according to Appendix G.

The selection rule above for the penalization parameter does not take into account the dependence structure of the autoregressive errors. As long as the correlation is not too strong, σ^2 should dominate the long-run variance justifying the use of the rule under several dependent cases. Based on Appendix F, we use the rule

$$\lambda_T = \sigma^2 \left(\frac{1 + \rho^\theta}{1 - \rho^\theta} \right) T^{1/4}, \tag{4.12}$$

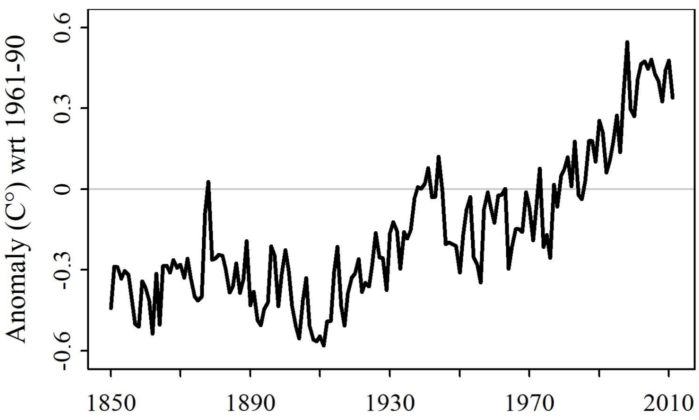
with the autocorrelation parameter being estimated by $\check{\rho} = \sum_{t=2}^T \check{\epsilon}_t \check{\epsilon}_{t-1} / \sum_{t=1}^T \hat{\epsilon}_t$ and the variance σ^2, by (4.11).

Table 2: Empirical probabilities that $\tilde{\theta} = 60$ and that $55 \leq \tilde{\theta} \leq 65$.

	$P(\tilde{\theta} = 60)$				$P(55 \leq \tilde{\theta} \leq 65)$			
	T=160	T=250	T=500	T=800	T=160	T=250	T=500	T=800
Chosen bandwidth: $h_{\text{opt}}^{0.5}$								
σ_ϵ^2 =0.25	0.20	0.43	0.96	1.00	1.00	1.00	1.00	1.00
σ_ϵ^2 =0.5	0.16	0.27	0.85	1.00	0.97	0.98	1.00	1.00
σ_ϵ^2 =0.75	0.14	0.24	0.68	0.99	0.90	0.99	1.00	1.00
Chosen bandwidth: h_{opt}								
σ_ϵ^2 =0.25	0.21	0.43	0.96	1.00	1.00	0.99	1.00	1.00
σ_ϵ^2 =0.5	0.15	0.27	0.85	1.00	0.97	0.98	1.00	1.00
σ_ϵ^2 =0.75	0.13	0.25	0.68	0.99	0.90	0.99	1.00	1.00
Chosen bandwidth: $h_{\text{opt}}^{5/4}$								
σ_ϵ^2 =0.25	0.21	0.43	0.96	1.00	1.00	1.00	1.00	1.00
σ_ϵ^2 =0.5	0.15	0.25	0.85	1.00	0.96	0.98	1.00	1.00
σ_ϵ^2 =0.75	0.13	0.24	0.68	0.99	0.89	0.99	1.00	1.00
Chosen bandwidth: $h_{\text{opt}}^{1.5}$								
σ_ϵ^2 =0.25	0.19	0.41	0.95	1.00	1.00	0.99	1.00	1.00
σ_ϵ^2 =0.5	0.15	0.26	0.85	1.00	0.96	0.98	1.00	1.00
σ_ϵ^2 =0.75	0.14	0.23	0.67	0.99	0.88	0.99	1.00	1.00

Table 2 presents the empirical probabilities $P(\tilde{\theta} = 60)$ and $P(55 \leq \tilde{\theta} \leq 65)$ for different sample sizes, error variances and bandwidth choices. Overall, the period estimator $\tilde{\theta}$ per-

Figure 4.1: Yearly temperature anomalies.



formed fairly robust over different bandwidths and with a good accuracy when compared with the results of Vogt and Linton (2014).

Table 4 and Figures 4.9 - 4.10 in Appendix I present additional results for h^*_{opt} and for sample sizes which are in the 60-periodic orbit or $120, 140$ and 160. In all cases, the period estimator showed to be robust as well.

In the absence of a bandwidth selection rule for the period estimator $\tilde{\theta}$, the robustness over bandwidths is highly desirable. In this exercise, if one chooses the same bandwidth h_{opt} selected in the first step of our estimation procedure, his period estimator would be almost as accurate as that obtained using other bandwidth choices considered above.

4.4 Applications

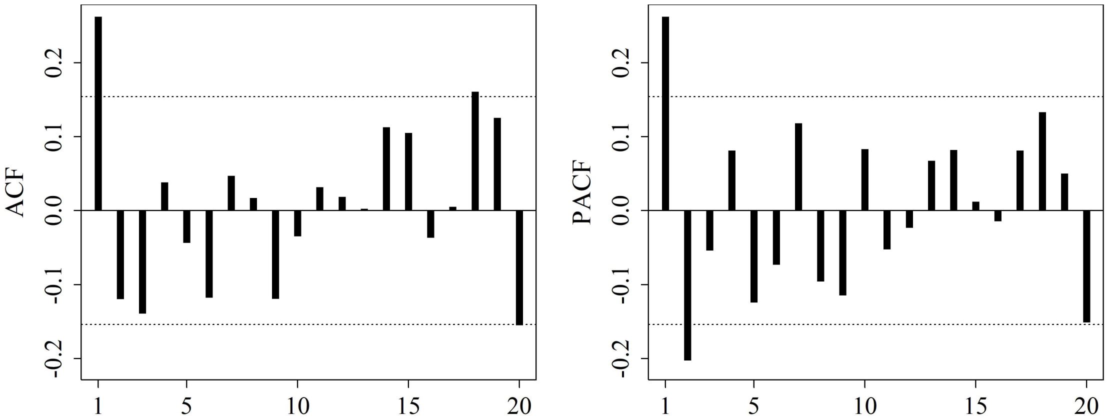
4.4.1 Global temperature anomalies

We illustrate the applicability of the proposed reversed three-step procedure to the HadCrut3[1] data used in Vogt and Linton (2014). The data refers to the yearly global mean temperature anomalies from 1850 to 2011. More specifically, these are temperature deviations from the average of 1961-1990 measured in degree Celsius. As pointed by the authors, the global mean temperature records suggest that there has been a significant upward trend in the temperatures (BLOOMFIELD, 1992; HANSEN et al., 2002) and some existing researches indicate that the global temperature system possesses an oscillation with period in the region between 60 and 70 years (SCHLESINGER; RAMANKUTTY, 1994; MAZZARELLA, 2007). Figure 4.1 depicts the data.

We fit the model (3.1) to the temperature data and estimate the trend function g, the unknown period θ_0 and the periodic sequence m, in this order. Since the period estimator $\tilde{\theta}$ shown to be robust to different bandwidths in Section 4.3, we choose the

[1] The dataset have been developed by the Climatic Research Unit in conjunction with the Hadley Centre. It can be accessed by the link: https://crudata.uea.ac.uk/cru/data/crutem3/HadCRUT3-gl.dat

Figure 4.2: Autocorrelation and partial autocorrelation functions of the pilot residuals.



same bandwidth, h^*_{opt}, in all steps of the estimation.

We employed the pilot bandwidth $h_{\text{p}} = 0.5$ for the computation of the selector h^*_{opt}. As described in section 4.2, we need preliminary estimates $g^{(p)}, \theta_0^{(p)}, m^{(p)}$ and $\Gamma_T^{(p)}$ of g, θ_0, m and Γ_T, respectively, to approximate the $\widehat{MISE}$, and it is accomplished based on h_{p}. Then our reversed estimation procedure is used to obtain $g^{(p)}, \theta_0^{(p)}$ and the centered $m^{(p)}$ with the bandwidth h_p, producing the residuals $\epsilon^{(p)} = Y - g^{(p)} - m^{(p)}$.

Figure 4.2 depicts the autocorrelation and partial autocorrelation functions of $\epsilon^{(p)}$ from where we can conjecture that we are dealing with a first order moving average error proccess. Inspecting various ARMA models we found that the lowest Bayesian information criterion (BIC) is associated with the MA(1) model. Therefore, we estimate the covariance matrix Γ_T by

$$\hat{\Gamma}_T = \hat{\sigma}^2_\eta \begin{bmatrix} 1+\hat{\rho}^2 & \hat{\rho} & 0 & \cdots & 0 \\ \hat{\rho} & 1+\hat{\rho}^2 & \hat{\rho} & \cdots & 0 \\ 0 & \hat{\rho} & 1+\hat{\rho}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\hat{\rho}^2 \end{bmatrix}. \tag{4.13}$$

with $\hat{\rho}$ and $\hat{\sigma}^2_\eta$ being the maximum likelihood estimates of the moving average coefficient and the variance of the innovations, respectively. Having all preliminary estimates in hand, we approximate $\widehat{MISE}(h)$ and then perform a numerical minimization over h. We obtained the minimum point $h^*_{\text{opt}} \approx 0.43$. By selecting this bandwidth, we perform our estimation procedure again to obtain the final estimates $\hat{g}, \tilde{\theta}$ and $\tilde{\beta}$. In assuming MA(1) errors we highlight that the penalization rule becomes just $\lambda_T = \sigma^2 T^{1/4}$. This selection rule for the penalization parameter was used to estimate both $\theta_0^{(p)}$ and $\tilde{\theta}$.

Figure 4.3 (a) shows the values of the estimated trend $\hat{g}$. In particular, $\hat{g}$ becomes monotone increasing after the year of 1874, indicting a predominant upward trend within the time interval under analysis. We found an oscillation with period 63 which is depicted

Figure 4.3: Estimated values for the trend function, the period and the periodic sequence.

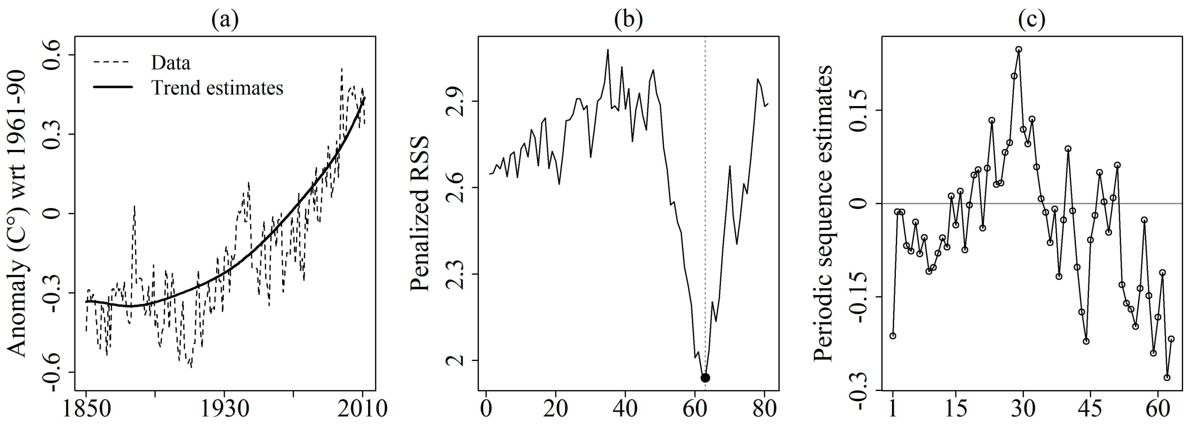
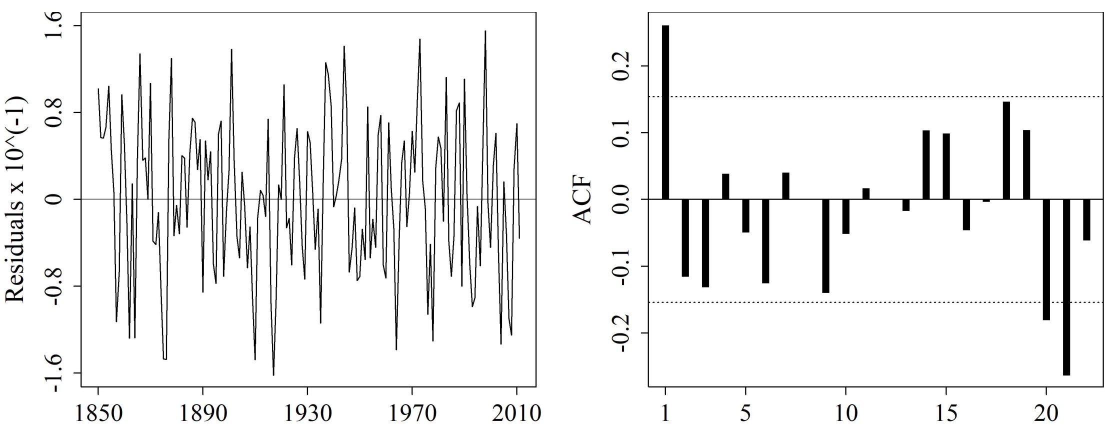


Figure 4.4: Residuals.



in Figure 4.3 (b) where we can see an evident downward spike in the criterion function. The estimated 63-periodic sequence is illustrated in Figure 4.3 (c). Therefore, the estimated results are consistent with the evidences found in the climate change literature.

The estimated residuals $\tilde{\epsilon} = Y - \hat{g} - \tilde{\beta} X_{\tilde{\theta}}$ are reported at Figure 4.4 as well as its autocorrelation function. The residuals do not appear to have a strong trend or periodic behavior. In addition, the autocorrelation function of the residuals do not appear to show a strong dependence over time.

4.4.2 Australian non-accelerating inflation rate of unemployment

The tradeoff between inflation and unemployment has been investigated by many economists, giving rise to some ideas that are now central in mainstream macroeconomics. One of the most widely known economic concept is the Phillip's Curve (PHILLIPS, 1958) which stablishes an inverse relationship between inflation and unemployment. Phillips (1958) found a relatively stable negative correlation between the rate of change in nominal wages and the unemployment rate in United Kingdom. Later, Samuelson and Solow (1960) showed a similar relationship in United States, but focusing in inflation rates rather than in rates of change in nominal wages. They also championed that it could be used as a policy tool (HALL; HART, 2012). Ideally, by determining the fiscal and monetary policy to change the aggregate demand, poliymakers would be able to choose any pair of unemployment and inflation rates on the Phillip's Curve.

Seminal works of Friedman (1968) and Phelps (1967, 1968) introduced the idea that monetary attempts to keep the unemployment low at the cost of higher inflation would be just temporarily successful. When the inflation expectations be adjusted to the new rate of monetary growth, the unemployment rate comes back to its *natural rate*. Many authors do consider the natural rate of unemployment and the non-accelerating inflation rate of unemployment (NAIRU) as synomyms (GORDON, 1997; STAIGER et al., 1997; STIGLITZ, 1997; MANKIW, 1985; BALL; MANKIW, 2002), i.e., as the unemployment rate consistent with stable (or non-accelerating) inflation. In this section, both concepts will be treated as equivalent.

In the Phillip's Curve literature, the Friedman-Phelps framework can be expressed as (BALL; MANKIW, 2002; BALL; MAZUMDER, 2019; FUHRER et al., 2009)

$$\pi_t = \pi_{t-1} + \alpha(u_t^* - u_t) + v_t, \quad \alpha > 0, \quad t \in \{1, \ldots, T\} \tag{4.14}$$

where π_t is the inflation rate, u_t is the unemployment rate, u_t^* is the NAIRU and v_t is an error term. Equation 4.14 is commonly called the *accelerationist* Phillip's Curve. It differs from the basic Phillip's Curve mainly because it includes the (time-varying) NAIRU and the lagged inflation rate which is implicity assumed to be the expected inflation rate at the current time, $E_t(\pi_t) = \pi_{t-1}$.

Equation 4.14 is equivalent to

$$\frac{\Delta\pi_t}{\alpha} + u_t = u_t^* + \frac{v_t}{\alpha} \tag{4.15}$$

where $\Delta\pi_t = \pi_t - \pi_{t-1}$. Once α is known and observations of $\Delta\pi_t$ and u_t are given, Ball and Mankiw (2002) suggested that u_t^* could be estimated from (4.15) using standard trend extraction tools. At a first step, they assumed u_t^* constant to obtain an ordinary least squares (OLS) estimate $\hat{\alpha}$ for the parameter α from model (4.14), and then use $\hat{\alpha}$ in (4.15) to estimate u_t^* as the trend of the Hodrick-Prescott (HP) filter.

We will extend the approach of Ball and Mankiw (2002) in order to illustrate our estimation procedure using australian data. Our aim is to provide central estimates for the time-varying NAIRU[2].

[2]The confidence intervals for the local linear trend estimator has nothing to do with the standard deviations of stochastic NAIRUs that often appear in the literature. While the former relates to estimation errors, the latter relates to the variance assumed in the NAIRU's dynamics.

Given $T \in \mathbb{N}$, assume the observations $\{(\Delta\pi_{t,T}, u_{t,T})\}_{t=1}^{T}$ follow the model

$$\Delta\pi_{t,T} = y_{t,T} + m_1(t) \tag{4.16}$$

$$u_{t,T} = x_{t,T} + m_2(t) \tag{4.17}$$

$$y_{t,T} = \alpha(f(t/T) - x_{t,T}) + v_{t,T}, \tag{4.18}$$

for any $t \in \{1, \ldots, T\}$, where m_1, m_2 are two unknown deterministic periodic sequences with fundamental periods θ_1, θ_2 respectively, f is an unknown deterministic smooth function interpreted as the NAIRU and $\{v_{t,T}\}$ is a strictly stationary and strongly mixing stochastic process[3]. Additionally, assume that $\sum_{i=1}^{\theta_0} m_1(i)/\alpha + m_2(i) = 0$ and denote the fundamental period of $m_1/\alpha + m_2$ by θ_0. In particular, if both m_1 and m_2 have period one (aperiodic), we rely on a model similar to (4.14). By our model, equation (4.18) is equivalent to

$$\frac{\Delta\pi_{t,T}}{\alpha} + u_{t,T} = f(t/T) + m_\alpha(t) + \frac{v_t}{\alpha}, \tag{4.19}$$

where $m_\alpha = m_1/\alpha + m_2$. Given an initial estimate $a^{(0)}$ of α, we can obtain estimates $f^{(0)}, \theta^{(0)}, m^{(0)}$ of f, θ_0, m_α using our proposed method.

To gain finite sample insights, suppose that $f^{(0)}, \theta^{(0)}$ are given and the following regression model is used to re-estimate α in (4.18),

$$\Delta\pi_{t,T} = \beta(f^{(0)}(t/T) - u_{t,T}) + v_{t,T} \tag{4.20}$$

for $t \in \{1, \ldots, T\}$. Then we would be ignoring the seasonal term of $\Delta\pi$ resulting in a biased least squares estimate (see Appendix H). If the seasonal term m_2 is not orthogonal to $\Delta\pi$, then it also has to be taken into account in order to separate the partial effects of $x_{t,T}$ and of m_2. Therefore, one can suggest to use the model

$$\Delta\pi_{t,T} = \beta(f^{(0)}(t/T) - u_{t,T}) + \sum_{i=1}^{\theta^{(0)}} \beta_i D_{i,t,T} + v_{t,T} \tag{4.21}$$

where $D_{i,t,T} = I(t \in \{1, \ldots, T\} : t = i + k\theta^{(0)}$ for some $k \in \mathbb{N})$, which is simply a periodic dummy variable. If, say, $\theta_0 = LCM(\theta_1, \theta_2)$[4], then $\Delta\pi$ and u are θ_0-periodic, even though it is not necessarily their least periods. It can be shown that the least squares estimate of

[3]Theorem 6 in Section 28.5 of Fristedt and Gray (1996) implies that a stationary proccess is strongly mixing if and only if it is ergodic. Technically, we also need to ensure that data generating proccess of $(y_{t,T}, x_{t,T})$ satisfies $E(x_{t,T}v_{t,T}) = 0$ and $E(x_{t,T}^2) < \infty$, $\forall 1 \leq i \leq T, \forall T \in \mathbb{N}$, and is jointly stationary ergodic to obtain consistent ordinary least squares estimates (see Proposition 2.1(a) of Hayashi, 2000).

[4]Although it holds in most cases, there are situations where it is false. For example, take $\alpha = 1, m_1(t) = \{(-1)^t\}$ and $m_2 = -m_1$ both periodic with least period equal to 2. Then $m_1/\alpha + m_2 = \{0, 0, \ldots\}$ which has least period 1. On the other hand, take $\alpha = 1, m_1 = \{1, 2, 3, 4, 1, 2, 3, 4, \ldots\}$ and $m_2 = (0, 0, -2, -2, 0, 0, -2, -2, \ldots)$ both periodic sequences with least period 4. But $m_1/\alpha + m_2 = \{1, 2, 1, 2, \ldots\}$ which has period 2. These examples show that if our assumption fails, θ_0 may not be a (multiple) period of m_1 or m_2.

β from (4.21) is the same to that of obtained using (4.20) but with $\Delta\pi$ and $f^{(0)} - u$ being priorly $\theta^{(0)}$-deseasonalized (see Appendix H). From these facts, we can conclude that there may be an excessive number of seasonal dummies in the model (4.21), if $\theta_0 = LCM(\theta_1, \theta_2)$. This implies a possible loss of efficiency (see Appendix H).

We briefly describe the NAIRU estimation from our model as follows:

(a) Calculate the OLS estimate $a^{(0)}$ of α from (4.18) assuming f constant;
(b) Given $a^{(0)}$, estimate f, θ_0 using our proposed estimators $f^{(0)}, \theta^{(0)}$ for the trend and period, respectively, from (4.19);
(c) Given $f^{(0)}, \theta^{(0)}$, estimate α by the OLS estimate $\hat{\beta}$ obtained from (4.21);
(d) Given $\hat{\beta}$, estimate f again using our proposed naive trend estimator $f^{(\text{final})}$ from (4.19).

Based on Ball and Mankiw (2002), we also estimate the time-varying NAIRU using the HP filter as follows

(a) Calculate the OLS estimate $a^{(0)}$ of α from (4.14) assuming u_t^* constant;
(b) Given $a^{(0)}$, use the HP filter estimate $u_t^{(0)}$ of u_t^* from (4.15);
(c) Given $u_t^{(0)}$, estimate α by the OLS estimate $a^{(1)}$ obtained from (4.14);
(d) Given $a^{(1)}$, estimate u_t^* again using the HP filter estimate $u_t^{(\text{HP})}$ from (4.15).

According to the estimates of Reserve Bank of Australia (RBA), the NAIRU was around 7 per cent in early 1980 and declined to around 6 per cent in 1985. It reached a peak in the mid-1990s at around 7 per cent and, subsequently, declined more or less steadily since then to around 5 per cent in early 2017 (CUSBERT et al., 2017).

We used annual data[5] from 1968 to 2019 to provide australian NAIRU estimates for the period 1980-2017. The estimation for the trend function and for the periodic sequence are done in the same way as in section 4.4.1, with a pilot bandwith equal to 0.3. We report that the bandwidth selection of h^*_{opt} considered a MA(2) error proccess, and the value $h^*_{\text{opt}} \cong 0.24$ was obtained.

A periodic sequence of period 13 was captured in the estimated time series $\Delta\Pi_{t,T}/a^{(0)} + u_{t,T}$. By observing the criterion function in Figure 4.6(a), the heuristicaly selected penalization parameter λ_T, defined in (4.10), should perhaps be slightly increased. Nevertheless, the downward spike at period 13 is evident, producing the periodic sequence illustrated in Figure 4.6(b).

The final estimates for the time-varying NAIRU are presented in Figure 4.7. It shows our proposed method estimates and the estimates of the method which uses the HP filter. We consider the Hodrick-Prescott's penalization parameters $\lambda \in \{10, 100, 400\}$, which are usual for annual data (RAVN; UHLIG, 2002). One can see that our proposed method

[5]The inflation data is the growth rate of OECD's CPI (total all items for Australia), code CPALTT01AUA657N, retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPALTT01AUA657N. Unemployment rate data (aged 15 and over, all persons for Australia), code LRUNTTTTAUA156S, can be obtained in https://fred.stlouisfed.org/series/LRUNTTTTAUA156S.

produced fairly different NAIRU estimates from that obtained through HP's estimates. Furthermore, our estimates seem to be in line with those described by RBA (see Figure 4.12 in Appendix I), except for the beginning of the sample where we obtained values around 6 per cent instead of 7 per cent. However our aim is not to hit the exact values estimated from RBA since their model treat the NAIRU as stochastic and many other variables are considered as well as its lagged values. As mentioned by Setterfield et al. (1992) for the case of Canada, the NAIRU estimates are sensitive to model specification and the definition of variables.

Figure 4.5: Unemployment rates and first differences of the inflation rates .

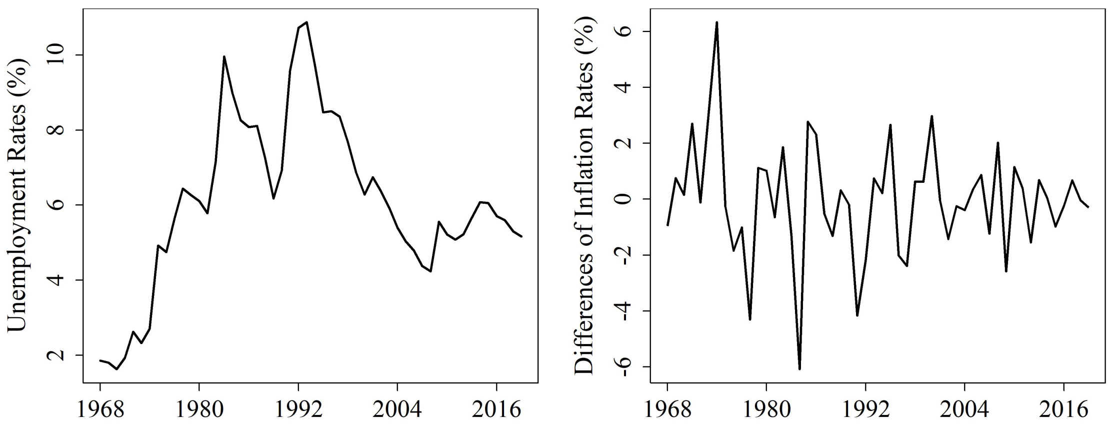


Figure 4.6: Estimated period and associated periodic sequence.

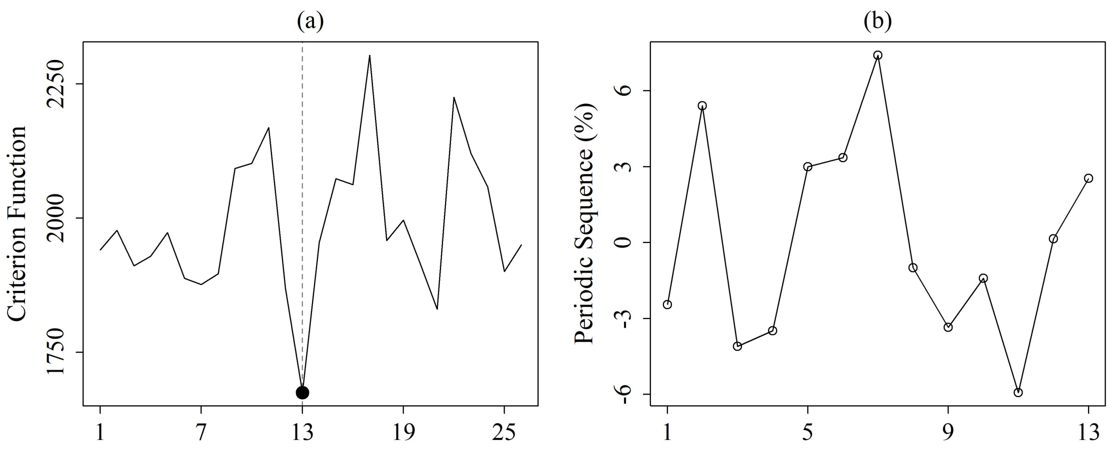
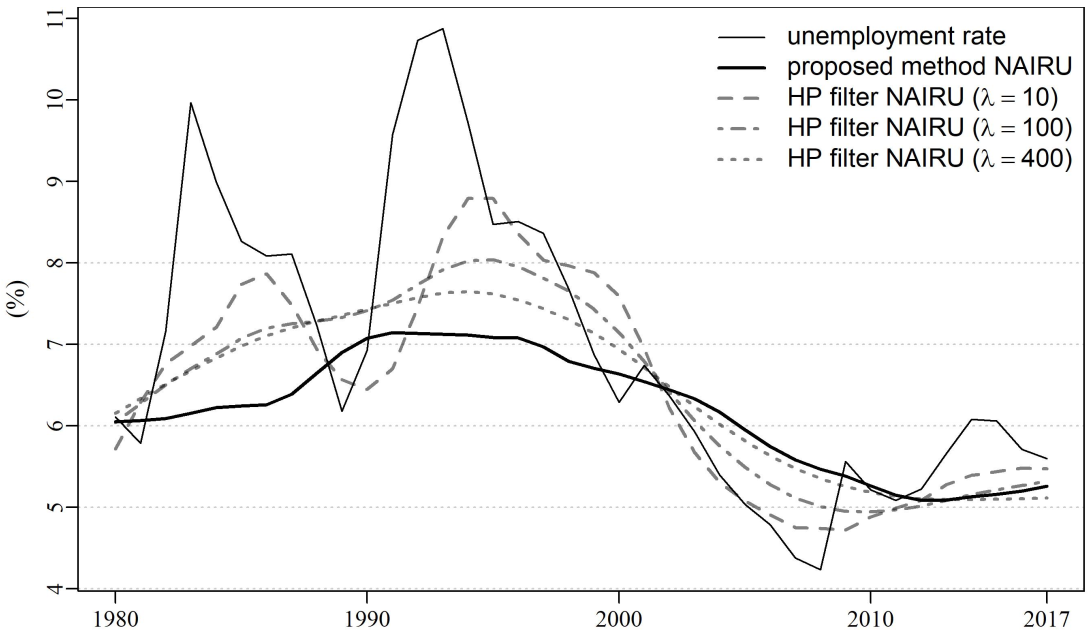


Figure 4.7: NAIRU estimates.



4.5 References

BALL, L.; MAZUMDER, S. A Phillips Curve with Anchored Expectations and Short-Term Unemployment. *Journal of Money, Credit and Banking*, v. 51, n. 1, p. 111-137, 2019. Available in <https://ideas.repec.org/a/wly/jmoncb/v51y2019i1p111-137.html>. Acessed on 25/08/2020.

BALL, L.; MANKIW, N.G. The NAIRU in theory and practice. *Journal of Economic Perspectives*, v. 16, n. 4, p. 115-136, 2002. Available in <https://ideas.repec.org/p/nbr/nberwo/8940.html>. Acessed on 25/08/2020.

BLOOMFIELD, P. Trends in global temperature. *Climate change*, v. 21, n. 1, p. 1-16, 1992.

CUSBERT, T. Estimating the NAIRU and the Unemployment Gap. *Reserve Bank of Australia Bulletin*, p. 13-22, june, 2017. Available in <https://www.rba.gov.au/publications/bulletin/2017/jun/pdf/bu-0617-2-estimating-the-nairu-and-the-unemployment-gap.pdf>. Acessed on 25/08/2020.

FAN, J.; GIJBELS, I. *Local polynomial modelling and its applications*: monographs on statistics and applied probability. CRC Press, 1996.

FAN, J. et al. A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica*, v. 6, n. 1, p. 113-127, 1996. Available in <https://www.jstor.org/stable/24306002?seq=1>. Acessed on 25/08/2020.

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.

FRIEDMAN, M. The Role of Monetary Policy. *American Association Review*, v. 58, n. 1, p. 1-17, 1968. Available in <https://www.jstor.org/stable/1831652?seq=1>. Acessed on 25/08/2020.

FRISTEDT, B.; GRAY, L. *A Modern Approach to Probability Theory*. New York: Springer, 1996.

FUHRER, J. The Phillips curve in historical context. In: FUHRER, J. et al. *Understanding Inflation and the Implications for Monetary Policy: A Phillips Curve Retrospective*. London: MIT Press, 2009, p. 3-68.

GORDON, R.J. The time-varying NAIRU and its implications for economic policy. *Journal of Economic Perspectives*, v. 11, n. 1, p. 11-32, 1997. Available in <https://www.jstor.org/stable/2138249?seq=1>. Acessed on 25/08/2020.

HALL, T.E.; HART, W.R. The Samuelson-Solow Phillips curve and the great inflation. *History of Economics Review*, v. 55, n. 1, p. 62-72, 2012. Available in <https://www.tandfonline.com/doi/abs/10.1080/18386318.2012.11682193>. Acessed on 25/08/2020.

HANSEN, J. et al. Global warming continues. *Science*, v. 295, n. 5553, p. 275-276, 2002. Available in <https://science.sciencemag.org/content/295/5553/275.3>. Acessed on 25/08/2020.

HART, J.D. Kernel Regression Estimation With Time Series Errors. *Journal of the Royal Statistical Society*, Series B, v. 53, n. 1, p. 173-187, 1991. Available in <https://www.jstor.org/stable/2345733?seq=1>. Acessed on 25/08/2020.

HAYASHI, F. *Econometrics*. Princeton University Press, 2000.

MANKIW, N.G. Small menu costs and large business cycles: A macroeconomic model of monopoly. *The Quarterly Journal of Economics*, v. 100, n. 2, p. 529-537, 1985. Available in <https://ideas.repec.org/a/oup/qjecon/v100y1985i2p529-538..html>. Acessed on 25/08/2020.

MAZZARELLA, A. The 60-year solar modulation of global air temperature: the Earth's rotation and atmospheric circulation connection. *Theoretical and Applied Climatology*, v. 88, n. 3-4, p. 193-199, 2007. Available in <https://link.springer.com/article/10.1007/s00704-005-0219-z>. Acessed on 25/08/2020.

PHELPS, E.S. Phillips curves, expectations of inflation and optimal unemployment over time. *Economica*, v. 34, n. 13, p. 254-281, 1967. Available in <https://www.jstor.org/stable/2552025?seq=1>. Acessed on 25/08/2020.

PHELPS, E.S. Money-wage dynamics and labor-market equilibrium. *Journal of political economy*, v. 76, n. 4(2), p. 678-711, 1968. Available in <https://www.jstor.org/stable/1830370?seq=1>. Acessed on 25/08/2020.

PHILLIPS, A.W. The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957. *Economica*, v. 25, n. 100, p. 283-299, 1958. Available in <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0335.1958.tb00003.x>. Acessed on 25/08/2020.

RAVN, M.O.; UHLIG, H. On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of economics and statistics*, v. 84, n. 2, p. 371-376, 2002. Available in <https://www.jstor.org/stable/3211784?seq=1>. Acessed on 25/08/2020.

SAMUELSON, P.A.; SOLOW, R.M. Analytical aspects of anti-inflation policy. *The American Economic Review*, v. 50, n. 2, p. 177-194, 1960. Available in <https://www.jstor.org/stable/1815021?seq=1>. Acessed on 25/08/2020.

SCHLENSINGER, M.E.; RAMANKUTTY, N. An oscillation in the global climate system of period 65-70 years. *Nature*, v. 367, n. 6465, p. 723-726, 1994. Available in <https://www.nature.com/articles/367723a0>. Acessed on 25/08/2020.

SETTERFIELD, M.A.; GORDON, D.V.; OSBERG, L. Searching for a Will o'the Wisp: An Empirical Study of the NAIRU in Canada. *European Economic Review*, v. 36, n. 1, p. 119-136, 1992. Available in <https://www.sciencedirect.com/science/article/abs/pii/001429219290020W>. Acessed on 25/08/2020.

STAIGER, D.O.; STOCK, J.H.; WATSON, M.W. How precise are estimates of the natural rate of unemployment? In: ROMER, C.D.; ROMER, D.H. *Reducing Inflation: Motivation and Strategy.* NBER books, University of Chicago Press, p. 195-246, 1997.

STIGLITZ, J. Reflections on the natural rate hypothesis. *Journal of Economic Perspectives*, v. 11, n. 1, p. 3-10, 1997. Available in <https://ideas.repec.org/a/aea/jecper/v11y1997i1p3-10.html>. Acessed on 25/08/2020.

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.

Appendix F - Penalization parameter selection

In this section, we clarify the penalization parameter selection employed in our simulations and derive the asymptotic plug-in bandwidth for the naive trend estimator $\hat{g}$.

To motivate formula (4.12) for selecting the penalization parameter, consider the model (4.9) with $\{u_t\}$ being a weakly stationary autoregressive error proccess of order 1. By denoting $\text{RSS}(\theta)$ as the residual sum of squares associated with the least squares estimator based on the period θ, Vogt and Linton (2014) in page 8 of their supplementary material, showed that

$$\frac{\text{RSS}(\theta)}{T} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_{t,T}^2 - \sum_{s=1}^{\theta}\frac{1}{T}\left(\frac{1}{K_{s,T}^{\theta}}\sum_{t,t'=1}^{T} I_s(t)I_s(t')\epsilon_{t,T}\epsilon_{t',T}\right),$$

where $I_s(t) = I(t = k\theta + s \text{ for some } k \in \mathbb{N})$ with I being the indicator function. Therefore

$$E\left[\frac{\text{RSS}(\theta)}{T}\right] = \sigma_u^2 - \sum_{s=1}^{\theta}\frac{1}{T}\left(\frac{1}{K_{s,T}^{\theta}}\sum_{k,k'=1}^{K_{s,T}^{\theta}} \text{Cov}(u_{s+(k-1)\theta}, u_{s+(k'-1)\theta})\right),$$

where σ_u^2 is the variance of the proccess. Now, let $c : \mathbb{Z} \to \mathbb{R}$ be the autocorrelation function, and observe that

$$\begin{aligned}
\frac{1}{K_{s,T}^{\theta}}\sum_{k,k'=1}^{K_{s,T}^{\theta}} \text{Cov}(u_{s+(k-1)\theta}, u_{s+(k'-1)\theta}) &= \frac{\sigma_u^2}{K_{s,T}^{\theta}}\sum_{k,k'=1}^{K_{s,T}^{\theta}} c(\theta|k-k'|) \\
&= \sigma_u^2\left[c(0) + \frac{2}{K_{s,T}^{\theta}}\sum_{k=1}^{K_{s,T}^{\theta}}\sum_{k'=k+1}^{K_{s,T}^{\theta}} c(\theta|k-k'|)\right] \\
&= \sigma_u^2\left[c(0) + \frac{2}{K_{s,T}^{\theta}}\sum_{d=1}^{K_{s,T}^{\theta}-1} (K_{s,T}^{\theta} - d)c(d\theta)\right] \\
&= \sigma_u^2\left[c(0) + 2\sum_{d=1}^{K_{s,T}^{\theta}-1} c(d\theta) - \frac{2}{K_{s,T}^{\theta}}\sum_{d=1}^{K_{s,T}^{\theta}-1} dc(d\theta)\right].
\end{aligned}$$

Also, note that $\sum_{d=1}^{\infty}|c(d\theta)| \leq \infty$, by the stationarity assumption[6]. Let $\epsilon > 0$ be arbitrary. Then there exists $T_\epsilon \in \mathbb{N}$ such that for every $K_{s,T}^{\theta} \geq K_{s,T_\epsilon}^{\theta}$ we have $\sum_{K_{s,T}^{\theta} \leq d}|c(d\theta)| < \epsilon$. Therefore, for any $T > T_\epsilon$,

$$\frac{1}{K_{s,T}^{\theta}}\sum_{d=1}^{K_{s,T}^{\theta}-1}|dc(d\theta)| \leq \frac{1}{K_{s,T}^{\theta}}\left[\sum_{d=1}^{K_{s,T_\epsilon}^{\theta}-1}|dc(d\theta)| + \sum_{K_{s,T_\epsilon}^{\theta} \leq d \leq K_{s,T}^{\theta}}|dc(d\theta)|\right]$$

[6]Precisely, $\sum_{d=1}^{\infty}|c(d\theta)| = |\phi^{\theta}|/(1-|\phi^{\theta}|)$ where ϕ is the autoregressive coefficient of $\{u_t\}$.

$$\leq \frac{1}{K^{\theta}_{s,T}}\left[\sum_{d=1}^{K^{\theta}_{s,T_\epsilon}-1} |dc(d\theta)| + \sum_{K^{\theta}_{s,T_\epsilon}\leq d \leq K^{\theta}_{s,T}} K^{\theta}_{s,T}|c(d\theta)|\right]$$

$$\leq \frac{1}{K^{\theta}_{s,T}}\sum_{d=1}^{K^{\theta}_{s,T_\epsilon}-1} |dc(d\theta)| + \sum_{K^{\theta}_{s,T_\epsilon}\leq d} |c(d\theta)|$$

$$< \frac{1}{K^{\theta}_{s,T}}\sum_{d=1}^{K^{\theta}_{s,T_\epsilon}-1} |dc(d\theta)| + \epsilon.$$

By taking limits on both sides, we have $1/K^{\theta}_{s,T}\sum_{d=1}^{K^{\theta}_{s,T}-1}|dc(d\theta)| \to 0$, since $\epsilon > 0$ is arbitrary. Thus

$$\frac{1}{K^{\theta}_{s,T}}\sum_{k,k'=1}^{K^{\theta}_{s,T}} \mathrm{Cov}(u_{s+(k-1)\theta}, u_{s+(k'-1)\theta}) \overset{a}{\approx} \sigma_u^2\left[c(0) + 2\sum_{d=1}^{\infty} c(d\theta)\right] = \sigma_u^2\left[1 + 2\sum_{d=1}^{\infty}\phi^{d\theta}\right]$$
$$= \sigma_u^2\frac{1+\phi^{\theta}}{1-\phi^{\theta}}.$$

With these observations, we obtain that

$$E\left[\frac{\mathrm{RSS}(\theta)}{T}\right] = \sigma_u^2 - \sum_{s=1}^{\theta}\frac{1}{T}\left(\sigma_u^2\frac{1+\phi^{\theta}}{1-\phi^{\theta}} + o(1)\right) = \sigma_u^2 - \sigma_u^2\frac{\theta}{T}\frac{1+\phi^{\theta}}{1-\phi^{\theta}} + o\left(\frac{1}{T}\right)$$

Hence,

$$E[\mathrm{RSS}(\theta_0)] + \theta_0\sigma_u^2\frac{1+\phi^{\theta_0}}{1-\phi^{\theta_0}} \overset{a}{\approx} E[\mathrm{RSS}(r\theta_0)] + r\theta_0\sigma_u^2\frac{1+\phi^{r\theta_0}}{1-\phi^{r\theta_0}}$$

where $\sigma_u^2 = \sigma^2/(1-\phi^2)$ with σ^2 being the variance of the error of the autoregressive proccess. Thus,

$$E[\mathrm{RSS}(\theta_0)] + \theta_0\lambda_T\frac{1+\phi^{\theta_0}}{1-\phi^{\theta_0}} \overset{a}{\approx} E[\mathrm{RSS}(r\theta_0)] + \sigma_u^2\left(r\theta_0\frac{1+\phi^{r\theta_0}}{1-\phi^{r\theta_0}} - \theta_0\frac{1+\phi^{\theta_0}}{1-\phi^{\theta_0}}\right) + \theta_0\lambda_T\frac{1+\phi^{\theta_0}}{1-\phi^{\theta_0}}$$
$$\leq E[\mathrm{RSS}(r\theta_0)] + r\theta_0\lambda_T\frac{1+\phi^{r\theta_0}}{1-\phi^{r\theta_0}}$$

if $\lambda_T \geq \sigma_u^2$. This reasoning justifies the use of (4.12), when T is large enough. Clearly, similar arguments can be used to justify this type of selection for general stationary ARMA(p,q) proccesses.

Appendix G - Asymptotic plug-in bandwidth

Next, we derive an asymptotic plug-in method to select the bandwidth for the trend estimator $\hat{g}$. We strengthen the assumptions on the model (4.1) by requiring that the error process is strictly stationary: 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

(**B1**) The covariance structure of the process $\{\epsilon_{t,T}\}$ satisfies $\mathrm{Cov}(\epsilon_{i,T}, \epsilon_{i+k,T}) = \sigma_\epsilon^2 c(k),\ |k| = 0, 1, \dots,\ \forall T \in \mathbb{N}$, and $\sum_{k=1}^{\infty} k|c(k)| < \infty$;

(**B2**) The bandwidth sequence $h_n := h$ satisfies $h > 0,\ h \to 0$ and $Th^2 \to \infty$;

(**B3**) g is second continuously differentiable on $[0, 1]$;

(**B4**) The kernel function K is symmetric around zero, Lipschitz continuous and differentiable in its compact support.

Note that Conditions 1 and 2 imply B1[7]. Without loss of generality, we assume $\operatorname{supp} K = [-1, 1]$ and $\int K(u)du = 1$.

Define the term

$$d_k^* = \frac{1}{T}\sum_{t=1}^{T}(x_t - x)^k K_h(x_t - x)\epsilon_{t,T},\ k \in \{0, 1\}. \tag{4.22}$$

From Proposition 1 and 2 of Fernández and Fernández (2001) or Theorem 1 of Hart (1991), we have the following results.

Proposition 4.1. *Let $x \in (h, 1-h)$. Under B2 and B4, we have*

$$\lim_{T\to\infty} h^{-j} s_j = \mu_j, \quad \forall j \in \{0, 1, 2, 3\}, \tag{4.23}$$

where $\mu_j = \int u^j K(u)du$. In particular, $\lim_{T\to\infty} H^{-1}S_T H^{-1} = S$, where $H = \mathrm{diag}(1, h)$ and the 2×2 matrices S_T and S are given by $(S_T)_{i,j} = s_{i+j-2}$ and $(S)_{i,j} = \mu_{i+j-2}$, respectively. Furthermore, if $B1, B2$ and $B4$ hold, then

$$\lim_{T\to\infty} Th\,\mathrm{Cov}(h^{-i}d_i^*, h^{-j}d_j^*) = v_{j+i}c(\epsilon),\ \forall i, j \in \{0, 1\}, \tag{4.24}$$

where $v_l = \int u^l K^2(u)du$ and $c(\epsilon) = \sigma_\epsilon^2[c(0) + 2\sum_{l=1}^{\infty} c(l)]$. Equivalently,

$$\lim_{T\to\infty} ThE(H^{-1}D_T^* D_T^{*\prime} H^{-1}) = \tilde{D}c(\epsilon)$$

in matrix form, where $\tilde{D} = (d_0^*, d_1^*)'$ and $\tilde{D} = (v_0, v_1)'$.

[7] Using the ratio test, we have $(k+1)a^{k+1}/(ka^k) \to a < 1$ as $k \to \infty$, if $0 < a < 1$. Then, from Davydov's inequality, $\sum_{k=1}^{\infty} k|\sigma_\epsilon^2 c(k)| \le C\sum_{k=1}^{\infty} ka^k < \infty$, if Conditions 1 and 2 holds. Since the variance is finite, $\sum_{k=1}^{\infty} k|c(k)| \le \infty$.

Consider the notations of section 4.2 and Proposition 4.1 and define $\mu = (\mu_2, \mu_3)'$.

Theorem 4.1. *Under B1-B4, for any $x \in (h, 1-h)$, the asymptotic expressions for the bias and the variance of $\hat{B}_T(x)$ are, respectively,*

$$Bias(\hat{B}_T(x)) \overset{a}{\approx} \frac{h^2 g''(x)}{2} S^{-1}\mu,$$
$$\mathrm{Var}(\hat{B}_T(x)) \overset{a}{\approx} \frac{1}{Th} c(\epsilon) S^{-1}\tilde{D}S^{-1}.$$

Proof. We start with the derivation of the bias. Write

$$E[\hat{B}_T(x)] = S_T^{-1}A_T'W_T(g+m) := G + M,$$

where $g = (g(x_1), \dots, g(x_T))'$ and $m = (m(1), \dots, m(T))'$, with $x_i = i/T$, omitting the dependence of both G and M on x and T . The 2nd-order Taylor expansion of g about x is given by

$$\begin{aligned} g &= \begin{bmatrix} g(x) + (x_1 - x)g'(x) + (x_1 - x)^2 g''(x)/2 + o((x_1 - x)^2) \\ \vdots \\ g(x) + (x_T - x)g'(x) + (x_T - x)^2 g''(x)/2 + o((x_T - x)^2) \end{bmatrix} \\ &= A_T B_T(x) + g''(x)/2 \begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_T - x)^2 \end{bmatrix} + o\left(\begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_T - x)^2 \end{bmatrix} \right). \end{aligned}$$

Then, recalling that $S_T = A_T'W_TA_T$,

$$\begin{aligned} G &= B(x) + \frac{g''(x)}{2} S_T^{-1}A_T'W_T \begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_T - x)^2 \end{bmatrix} + o(1)S_T^{-1}A_T'W_T \begin{bmatrix} (x_1 - x)^2 \\ \vdots \\ (x_T - x)^2 \end{bmatrix} \\ &= B(x) + S_T^{-1}\left(\frac{g''(x)}{2} \begin{bmatrix} s_2 \\ s_3 \end{bmatrix} + o\begin{pmatrix} h^2 \\ h^3 \end{pmatrix} \right). \end{aligned} \tag{4.25}$$

Define $E = \{i \in \{1, \dots, \theta_0\} : K_{i,T}^{\theta_0} = \lfloor T/\theta_0 \rfloor\}$. Turning to the term M, we have

$$\begin{aligned} e_1'A_T'W_Tm &= \frac{1}{T}\sum_{t=1}^{T} K_h(x_t - x)m(t) = \frac{1}{T}\sum_{t=1}^{\theta_0} m(t) \sum_{k=1}^{K_{t,T}^{\theta_0}} K_h(x_{t+(k-1)\theta_0} - x) \\ &= \frac{1}{T}\sum_{t=1}^{\theta_0} m(t)K_{t,T}^{\theta_0}\left\{ \int_0^1 K_h(u-x)du + O(1/(Th)) \right\} \\ &= \frac{1}{T}\left\{ \int_{-1}^1 K(w)dw + O(1/(Th)) \right\} \left[\sum_{t\in E} m(t)K_{t,T}^{\theta_0} + \sum_{t\in E^c} m(t)K_{t,T}^{\theta_0} \right] \end{aligned}$$

$$= \frac{1}{T}\Big\{\underbrace{\int_{-1}^{1} K(w)dw}_{=1} + O(1/(Th))\Big\} \underbrace{\sum_{t\in E^c} m(t)}_{=O(1)}$$
$$= \frac{1}{T}\sum_{t\in E^c} m(t) + O(1/(T^2h)) := M_1 + O(1/(T^2h)),$$

using condition B4 and Lemmas 3.4 and 3.10 of Chapter 3. Similarly,

$$e_2' A_T' W_T m = \frac{1}{T}\sum_{t=1}^{T} K_h(x_t - x)(x_t - x)m(t)$$
$$= \frac{h}{T}\Big\{\underbrace{\int_{-1}^{1} K(w)wdw}_{=0} + O(1/(Th))\Big\} \underbrace{\sum_{t\in E^c} m(t)}_{=O(1)} = O(1/T^2).$$

Therefore,

$$M = S_T^{-1}\left(\begin{bmatrix} M_1 \\ 0 \end{bmatrix} + O\begin{pmatrix} (T^2h)^{-1} \\ (T^2h)^{-1} \end{pmatrix}\right). \tag{4.26}$$

Hence, from equations (4.25) and (4.26),

$$E[\hat{B}_T(x)] = B(x) + S_T^{-1}\left(\frac{g''(x)}{2}\begin{bmatrix} s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} M_1 \\ 0 \end{bmatrix} + O\begin{pmatrix} (T^2h)^{-1} \\ (T^2h)^{-1} \end{pmatrix} + o\begin{pmatrix} h^2 \\ h^3 \end{pmatrix}\right)$$

On the other hand, we have

$$S_T^{-1} D_T^* = \hat{B}(x) - E[\hat{B}(x)].$$

Then,

$$\hat{B}_T(x) - B(x) = S_T^{-1} D_T^* + S_T^{-1}\left(\frac{g''(x)}{2}\begin{bmatrix} s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} M_1 \\ 0 \end{bmatrix} + O\begin{pmatrix} (T^2h)^{-1} \\ (T^2h)^{-1} \end{pmatrix} + o\begin{pmatrix} h^2 \\ h^3 \end{pmatrix}\right).$$

This equation is convenient since it decomposes $\hat{B}_T(x) - B(x)$ into a bias part and a stochastic part. From Proposition 4.1,

$$HE[\hat{B}_T(x) - B(x)] = HS_T^{-1}\left(\frac{g''(x)}{2}\begin{bmatrix} h^2\mu_2 + o(h^2) \\ o(h^3) \end{bmatrix} + \begin{bmatrix} M_1 \\ 0 \end{bmatrix}\right.$$
$$\left. + O\begin{pmatrix} (T^2h)^{-1} \\ (T^2h)^{-1} \end{pmatrix} + o\begin{pmatrix} h^2 \\ h^3 \end{pmatrix}\right)$$
$$= \frac{h^2 g''(x)}{2} HS_T^{-1}H\begin{bmatrix} \mu_2 \\ 0 \end{bmatrix} + HS_T^{-1}HH^{-1}\begin{bmatrix} M_1 \\ 0 \end{bmatrix}$$

$$
\begin{aligned}
&+ O((T^2h)^{-1}) H S_T^{-1} H H^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + o(h^2) H S_T^{-1} H \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&\overset{a}{\approx} \frac{h^2 g''(x)}{2} S^{-1} \begin{bmatrix} \mu_2 \\ 0 \end{bmatrix} + S^{-1} H^{-1} \begin{bmatrix} M_1 \\ 0 \end{bmatrix} \\
&+ S^{-1} \left(o \begin{pmatrix} h^2 \\ h^2 \end{pmatrix} + O \begin{pmatrix} (T^2h)^{-1} \\ (T^2h)^{-1} \end{pmatrix} \right) \\
&= \frac{h^2 g''(x)}{2} S^{-1} \mu + \begin{bmatrix} M_1 \\ 0 \end{bmatrix} + S^{-1} o \begin{pmatrix} h^2 \\ h^2 \end{pmatrix} \\
&\overset{a}{\approx} \frac{h^2 g''(x)}{2} S^{-1} \mu,
\end{aligned}
$$

since $M_1 = O(1/T) = o(h^2)$ by B2. On the other hand,

$$
\begin{aligned}
\mathrm{Var}(H\hat{B}_T(x)) &= E(H S_T^{-1} D_T^* D_T^{*\prime} S_T^{-1\prime} H) = E(H S_T^{-1} H H^{-1} D_T^* D_T^{*\prime} H^{-1} H S_T^{-1\prime} H) \\
&\overset{a}{\approx} \frac{1}{Th} c(\epsilon) S^{-1} \tilde{D} S^{-1}.
\end{aligned}
$$

□

Corollary 4.1.1. *Under B1-B4, for any $x \in (h, 1-h)$, the asymptotic Mean Squared Error (MSE) of the trend estimator $\hat{g}(x)$ satisfies*

$$
MSE(\hat{g}(x), h) := MSE(x, h) \overset{a}{\approx} \frac{h^4 g''^2(x) \mu_2^2}{4} + \frac{c(\epsilon) v_0}{Th}. \tag{4.27}
$$

Theorem 4.1 implies that if h converges to zero slow enough, i.e., $1/T = o(h^2)$, then the asymptotic bias is the same as that for the model assuming the periodic component is known. When h^2 is allowed to converge to zero faster than $1/T$, i.e., $Th^2 = o(1)$, then the local asymptotic bias is dominated by the periodic component $M_1 = O(1/T)$. In this case, the bias-variance trade-off disappears, and the smaller h is chosen, the greater the asymptotic MSE will be. It corroborates with the intuition that h should not be chosen too small when estimating the trend of model (4.1) in the presence of the periodic sequence.

Hence, the Asymptotic Mean Integrated Squared Error (AMISE) is

$$
AMISE(h) = \frac{h^4 \int g''^2 \mu_2^2}{4} + \frac{c(\epsilon) v_0}{Th}. \tag{4.28}
$$

Given good estimators for the integral of the functional g''^2 and for $c(\epsilon)$, say $\int \hat{g}''^2$ and $\hat{c}(\epsilon)$, it makes sense to select h as the minimizer of formula (4.28) which is

$$
h_{\text{as}} = \left(\frac{v_0}{\mu_2^2} \frac{\hat{c}(\epsilon)}{\int \hat{g}''^2 T} \right)^{1/5}. \tag{4.29}
$$

Appendix H - Some effects of periodic sequences on least squares estimates

To make the arguments as simple as possible, suppose that the random sample (y, x) has x as deterministic. Although similar conclusions are obtained with stochastic x by imposing another set of classical least squares conditions (see Hayashi, 2000). Here we clarify the known consequences of omitting relevant variables or including superfluous variables as well as the equivalence between the least squares estimates obtained by a periodic adjustment within the regression model and that obtained using priorly periodic adjusted variables. Assume that the regression model is

$$y = x\alpha + D\beta + \epsilon \tag{4.30}$$

where y is a T-vector of dependent variables, x is a $T \times d$ matrix of "fixed" regressors, $D := D_\theta = (I_\theta, I_\theta, \dots)'$ is a $T \times \theta$ matrix of periodic dummies with I_θ being the $\theta \times \theta$ identity matrix and ϵ is a T-vector of errors. In this appendix, we always assume that the true model fulfills the Gauss-Markov conditions, thus resulting in a best linear unbiased estimator (BLUE). However, the estimated model is

$$u = x\alpha + \epsilon. \tag{4.31}$$

Since the least squares (LS) estimate is $\hat{\alpha} = (x'x)^{-1}x'y$, equation (4.30) implies

$$\hat{\alpha} = \alpha + (x'x)^{-1}x'(D\beta + \epsilon).$$

We immediately see that $\hat{\alpha}$ is biased and the bias term is given by $(x'x)^{-1}x'D\beta$.

Now assume that the model is given by

$$y = x\alpha + \epsilon,$$

but the estimated model is

$$y = x\alpha + D\beta + \epsilon.$$

Let $M = [x \quad D]$ be the $n \times (d + k)$ augmented matrix. The solution for the least squares problem of the estimated model is

$$(\hat{\alpha}, \hat{\beta})' = (M'M)^{-1}M'y = \left[\begin{pmatrix} x' \\ D' \end{pmatrix} \begin{pmatrix} x & D \end{pmatrix} \right]^{-1} \begin{bmatrix} x'y \\ D'y \end{bmatrix}$$
$$= \begin{bmatrix} x'x & x'D \\ D'x & D'D \end{bmatrix}^{-1} \begin{bmatrix} x'y \\ D'y \end{bmatrix} := \begin{bmatrix} A & B \\ C & E \end{bmatrix}^{-1} \begin{bmatrix} x'y \\ D'y \end{bmatrix},$$

and then

$$\begin{aligned}\hat{\alpha} &= \left[\begin{array}{cc}(A - BE^{-1}C)^{-1} & -(A - BE^{-1}C)^{-1}BE^{-1}\end{array}\right]\left[\begin{array}{c}x'y \\ D'y\end{array}\right] \\ &= (A - BE^{-1}C)^{-1}(x'y - BE^{-1}D'y) \\ &= (x'x - x'D(D'D)^{-1}D'x)^{-1}(x'y - x'D(D'D)^{-1}D'y).\end{aligned}$$

The projection matrix $\Pi = D(D'D)^{-1}D'$ was already studied in Lemma 3.5, where we found that

$$\Pi = \left[\begin{array}{ccc}K & K & \cdots \\ K & K & \\ \vdots & & \ddots\end{array}\right]_{T\times T}$$

with $K = \text{diag}(1/K^{\theta}_{1,T}, \ldots, 1/K^{\theta}_{\theta,T})$ and $K^{\theta}_{i,T} = \lfloor (T-i)/\theta \rfloor + 1$. The annihilator-like matrix $M = I_T - \Pi$ acts as a periodic adjustment matrix since it subtracts from any T-vector its θ-periodic means. Using the technical notation as that of Chapter 2, we can explicitly obtain that the i-th coordinate of My is given by $y_i - 1/K^{\theta}_{w_{\theta,i},T}\sum_{t=1}^{K^{\theta}_{w_{\theta,i},T}} y_t$ with $w_{\theta,i} = i - \lfloor (i-1)/\theta \rfloor$, where one can see the interpretation of M as a periodic (or seasonal) adjustment matrix. Hence

$$\hat{\alpha} = (x'Mx)^{-1}x'My. \tag{4.32}$$

Since the true model is $y = x\alpha + \epsilon$, we have from (4.32) that

$$\hat{\alpha} = \alpha + (x'Mx)^{-1}x'M\epsilon, \tag{4.33}$$

revealing an unbiased estimator. It is well known that the covariance matrix of the BLUE estimator a for the true model is $\sigma^2_{\epsilon}(x'x)^{-1}$. By Lemma 3.1($i$), M is symmetric and idempotent. Then the covariance matrix of $\hat{\alpha}$ is given by

$$E[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)'] = E[(x'Mx)^{-1}(x'M\epsilon\epsilon'Mx)(xMx')^{-1}] = \sigma^2_{\epsilon}(x'Mx)^{-1}.$$

A general relative efficiency analysis can be made by introducing the following partial order relation: we say that two Hermitian matrices A and B with equal dimensions satisfy $A \succeq B$ if $A - B$ is positive semi-definite.

Claim 6. Let A and B be two real, symmetric and positive definite $T \times T$ matrices. The matrix $A - B$ is positive semi-definite if, and only if, $B^{-1} - A^{-1}$ is positive semi-definite.
Proof of claim: By the positive definiteness hypothesis, there exist unique square root matrices $A^{1/2}$ and $B^{1/2}$ with inverse matrices $A^{-1/2}$ and $B^{-1/2}$, respectively. It holds that $A \succeq B \iff B^{-1/2}AB^{-1/2} \succeq I \iff \lambda_{inf} \geq 1$, where λ_{inf} is the infimum of the spectrum of $B^{-1/2}AB^{-1/2}$. To see this, let $M = B^{-1/2}AB^{-1/2} - I$ and let v be an

eigenvector corresponding to a eigenvalue λ, and observe that $\langle Mv, v\rangle = \langle v, v\rangle(\lambda - 1) \geq 0$. Further, since $B^{-1/2}AB^{-1/2} = (B^{-1/2}A^{1/2})(A^{1/2}B^{-1/2})$ and commuting matrices share the same eigenvalue spectrum, we must have $A^{1/2}B^{-1}A^{1/2} \succeq I \iff B^{-1} \succeq A^{-1}$.

■

By setting $A = (x'Mx)^{-1}$ and $B = (x'x)^{-1}$, we see that A and B are positive definite since $x'x$ and $(Mx)'(Mx)$ are positive semi-definite and invertible. Also $B^{-1} - A^{-1} = x'\Pi x = (\Pi x)'(\Pi x)$ is positive semi-definite. By the above claim, a is more efficient that $\hat{\alpha}$ in the sense than $\sigma_\epsilon^2(x'Mx)^{-1} - \sigma_\epsilon^2(x'x)^{-1}$ is positive semi-definite, unless $\Pi x = 0$ (that is, $x'D = 0$). Using the relation $\succeq$, we arrived with the conclusion that the only case where both estimators are equally efficient is when x is uncorrelated with D.

One final observation is that equation (4.32) implies the equivalence between the least squares estimate of model

$$y = x\alpha + D\beta + \epsilon$$

and the least squares estimate obtained from a model that uses seasonal adjusted variables

$$My = Mx\alpha + \epsilon,$$

by the symmetry and idempotency of M.

Appendix I - Additional reports

Table 3 shows the performance of the asymptotic plug-in h_{as} for the simulation exercise of Section 4.2.1. The asymptotic selector performs poorly when compared to h_{opt} and h^*_{opt}, specially for the smaller sample sizes $T \in \{160, 250\}$. This suggests that the bias part of the $MISE(h)$ of $\hat{g}$ is highly affected by the periodic component m, discouranging the use of asymptotic plug-in rules for small samples.

Table 3: Asymptotic plug-in bandwidth performance

		Mean	St. Dev.	$\Delta M \times 10^5$
	h_{min}	h_{as}	h_{as}	h_{as}
$T = 160$				
$\sigma^2 = 0.25$	0.90	0.33	0.08	586.47
$\sigma^2 = 0.5$	0.91	0.41	0.35	383.14
$\sigma^2 = 0.75$	0.92	0.43	0.19	364.77
$T = 250$				
$\sigma^2 = 0.25$	0.58	0.30	0.04	16.87
$\sigma^2 = 0.5$	0.58	0.35	0.11	7.33
$\sigma^2 = 0.75$	0.59	0.38	0.11	7.84
$T = 500$				
$\sigma^2 = 0.25$	0.32	0.26	0.02	0.19
$\sigma^2 = 0.5$	0.42	0.30	0.03	0.14
$\sigma^2 = 0.75$	0.43	0.33	0.05	0.14
$T = 800$				
$\sigma^2 = 0.25$	0.28	0.24	0.01	0.02
$\sigma^2 = 0.5$	0.34	0.28	0.02	0.01
$\sigma^2 = 0.75$	0.35	0.30	0.03	0.02

* The table presents the expectation, standard deviation and the efficiency measurement associated with each bandwidth selector. Here, $\Delta M(\hat{h}) = E(MISE(\hat{h}) - MISE(h_{\text{min}}))^2$.

Table 4 shows the sensitiveness of $\tilde{\theta}$ over different bandwidth values, which are powers of h^*_{opt} and h_{as}. Eventhough the behavior of h_{as} is distinct from the other selectors, the accuracy of the period estimator $\tilde{\theta}$ remained roughly unchanged for all selected bandwidths and for each pair (σ^2_ϵ, T). Tables 4 and 2 offer a strong evidence that the estimator $\tilde{\theta}$ is robust with respect to bandwidth choices, for the considered model.

Figure 4.8 presents the results of the simulation exercise of Section 4.2.1 for the sample sizes $120, 140, 160, 240, 260, 280, 420, 440$ and 460. The expected values of the bandwidth selector h_{as} are flatter than that of h_{opt} and h^*_{opt}. This is due to the absence of the periodic component in the asymptotic MISE defined in (4.28). On the other hand, the exact MSE, defined in (4.2), depends directly on m. It implies that the integrated bias carries the 60-periodic behavior of m, which in turn produces the oscillatory behavior on the means of h_{opt} and h^*_{opt} depicted in Figure 4.8. As can be seen from the plots of the

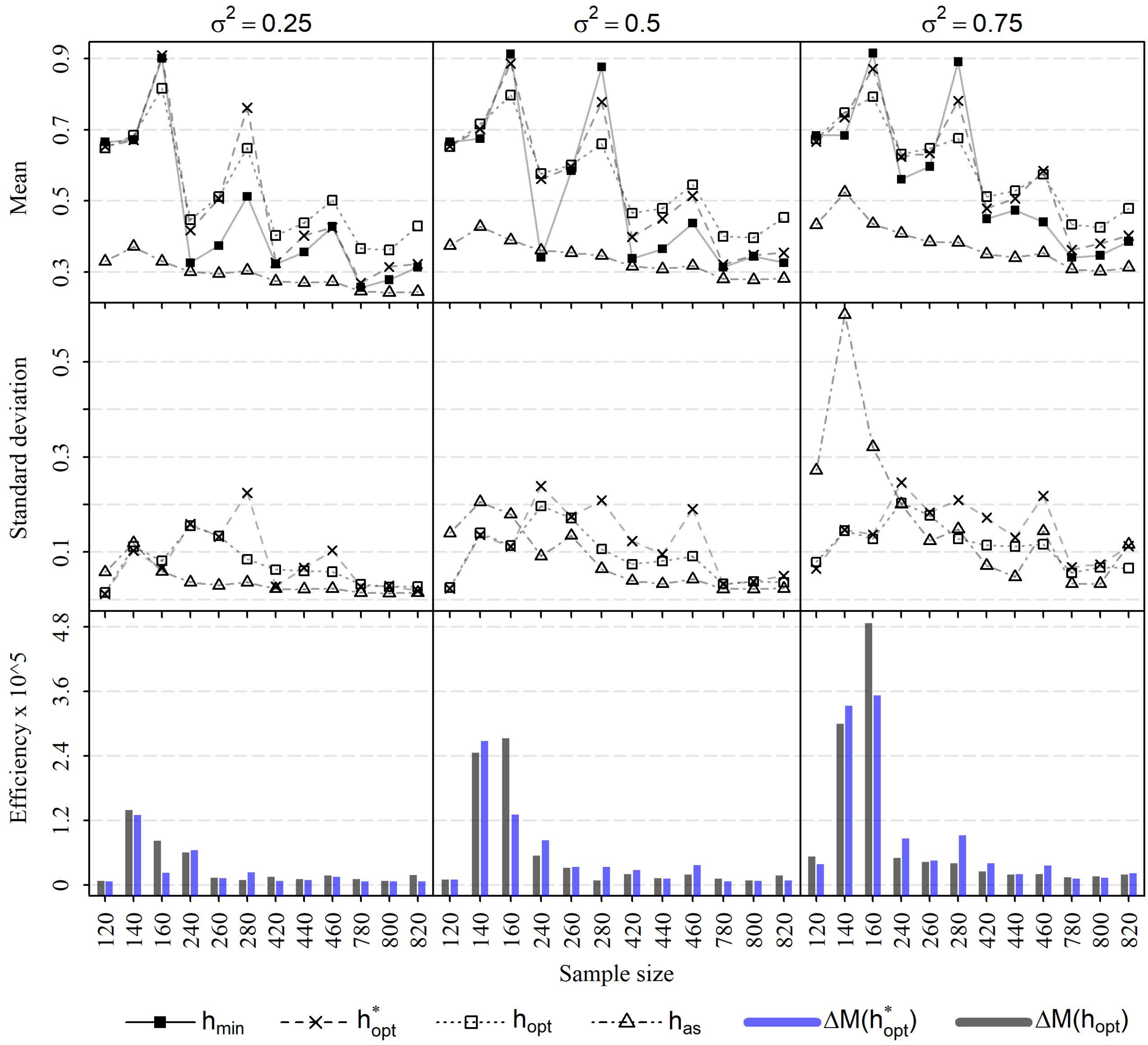
Table 4: Sensitivity of $\tilde{\theta}$ based on h^*_{opt} and h_{as}.

	$P(\tilde{\theta}=60)$				$P(55 \leq \tilde{\theta} \leq 65)$			
	T=160	T=250	T=500	T=800	T=160	T=250	T=500	T=800
Chosen bandwidth: $h^{*0.5}_{\text{opt}}$								
σ^2_ϵ =0.25	0.20	0.43	0.96	1.00	1.00	1.00	1.00	1.00
σ^2_ϵ =0.5	0.16	0.27	0.85	1.00	0.97	0.98	1.00	1.00
σ^2_ϵ =0.75	0.14	0.24	0.67	0.99	0.90	0.99	1.00	1.00
Chosen bandwidth: h^*_{opt}								
σ^2_ϵ =0.25	0.20	0.43	0.96	1.00	1.00	0.99	1.00	1.00
σ^2_ϵ =0.5	0.16	0.26	0.85	1.00	0.97	0.98	1.00	1.00
σ^2_ϵ =0.75	0.14	0.25	0.68	0.99	0.90	0.99	1.00	1.00
Chosen bandwidth: $h^{*5/4}_{\text{opt}}$								
σ^2_ϵ =0.25	0.20	0.43	0.96	1.00	1.00	1.00	1.00	1.00
σ^2_ϵ =0.5	0.16	0.26	0.85	1.00	0.97	0.98	1.00	1.00
σ^2_ϵ =0.75	0.14	0.24	0.68	0.99	0.90	0.99	1.00	1.00
Chosen bandwidth: $h^{*1.5}_{\text{opt}}$								
σ^2_ϵ =0.25	0.20	0.41	0.95	1.00	1.00	0.99	1.00	1.00
σ^2_ϵ =0.5	0.15	0.26	0.85	1.00	0.96	0.98	1.00	1.00
σ^2_ϵ =0.75	0.13	0.23	0.67	0.99	0.89	0.99	1.00	1.00
Chosen bandwidth: $h^{0.5}_{\text{as}}$								
σ^2_ϵ =0.25	0.19	0.44	0.96	1.00	1.00	0.99	1.00	1.00
σ^2_ϵ =0.5	0.15	0.27	0.85	1.00	0.96	0.98	1.00	1.00
σ^2_ϵ =0.75	0.13	0.25	0.68	0.99	0.89	0.99	1.00	1.00
Chosen bandwidth: h_{as}								
σ^2_ϵ =0.25	0.12	0.40	0.95	1.00	0.95	0.99	1.00	1.00
σ^2_ϵ =0.5	0.12	0.25	0.86	1.00	0.90	0.98	1.00	1.00
σ^2_ϵ =0.75	0.12	0.24	0.68	0.99	0.83	0.99	1.00	1.00
Chosen bandwidth: $h^{5/4}_{\text{as}}$								
σ^2_ϵ =0.25	0.09	0.39	0.95	1.00	0.87	0.99	1.00	1.00
σ^2_ϵ =0.5	0.11	0.23	0.84	1.00	0.82	0.98	1.00	1.00
σ^2_ϵ =0.75	0.08	0.23	0.67	0.99	0.73	0.99	1.00	1.00
Chosen bandwidth: $h^{1.5}_{\text{as}}$								
σ^2_ϵ =0.25	0.08	0.34	0.94	1.00	0.77	0.99	1.00	1.00
σ^2_ϵ =0.5	0.08	0.20	0.85	1.00	0.63	0.98	1.00	1.00
σ^2_ϵ =0.75	0.07	0.23	0.67	0.99	0.55	0.99	1.00	1.00

approximated MISE's minimum point, $h_{\min}$, this periodic behavior should be captured by any reasonable bandwidth selector for $\hat{g}$ when the sample is relatively small.

The simulation exercise of Section 4.3 is extended to the same samples sizes as in Figure 4.8, and the results are presented in Figures 4.9, 4.10 and 4.11 for bandwidth variations with respect to $h^*_{\text{opt}}, h_{\text{opt}}$ and h_{as}, respectively. Note that Figures 4.9 and 4.10 are almost the same. More importantly, regardless of whether the chosen base is h^*_{opt} or h_{opt}, the accuracy of $\tilde{\theta}$ does not change too much along the expoents $0.5, 1, 1.25$ and 1.5, for each pair (σ^2_ϵ, T). This property does not hold when the sample size is small $(120, 140$

Figure 4.8: Bandwidth selection performance for the trend estimator $\hat{g}$



and 160) and the chosen base is h_{as}, as can be seen in Figure 4.11.

Figure 4.12 reproduces the NAIRU estimates of Cusbert et al. (2017).

Table 5 presents the OLS outputs of step (c) of the estimation schemes described in Section 4.4.2. Model 1 is given by (4.21) where the regressor *gap* consists in the difference between the NAIRU and the unemployment rate, and X_i is the i-th column of the 52×13 dummy matrix $X = [I_{13} \quad I_{13} \quad \ldots]'$ with I_{13} being the 13×13 identity matrix. Models 2, 3 and 4 relate to estimates of equation (4.14) when u_t^* is previously estimated by HP filter using penalization parameters $10, 100$ and 400, respectively.

Figure 4.9: Bandwidth sensitiveness of $\tilde{\theta}$ based on h^*_{opt}

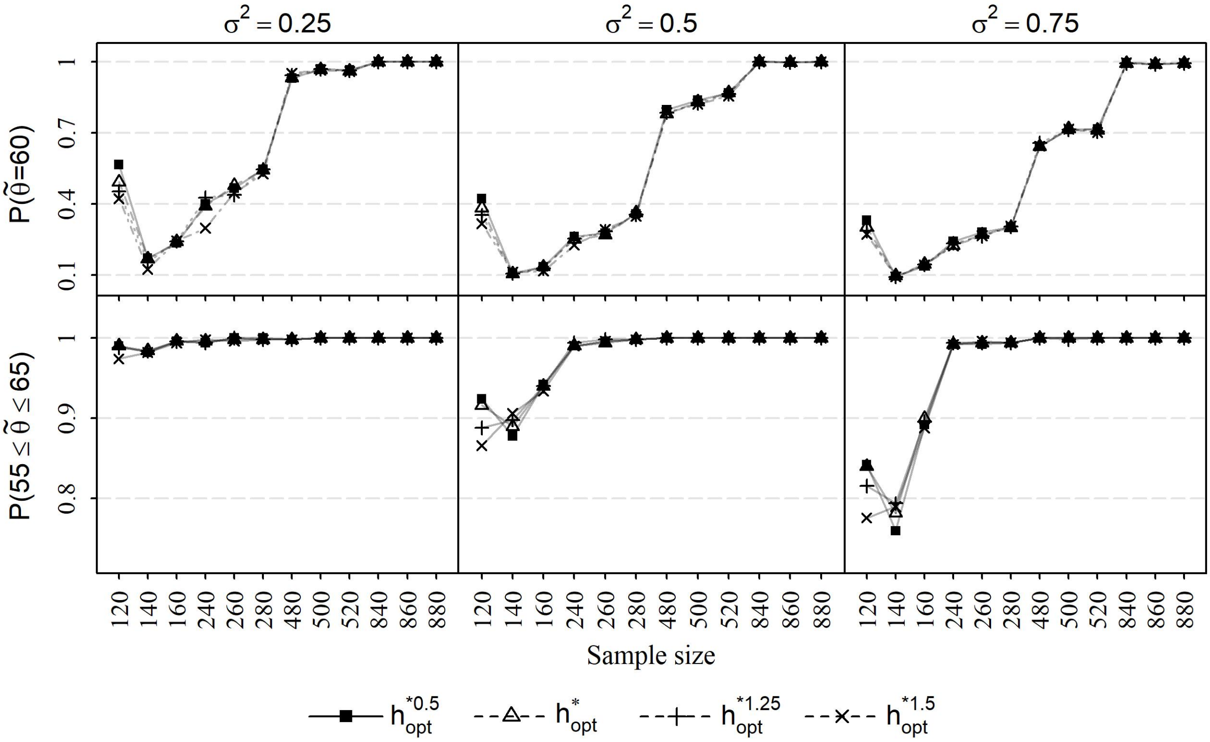


Figure 4.10: Bandwidth sensitiveness of $\tilde{\theta}$ based on h_{opt}

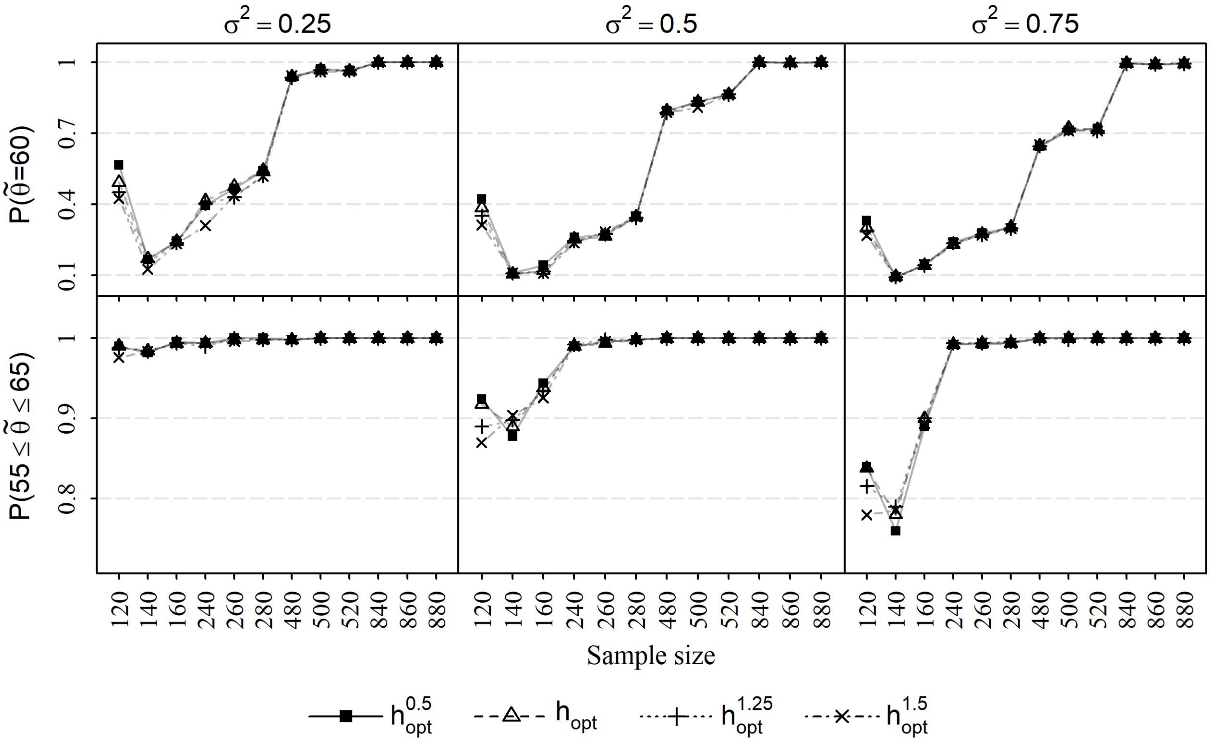


Figure 4.11: Bandwidth sensitiveness of $\tilde{\theta}$ based on h_{as}

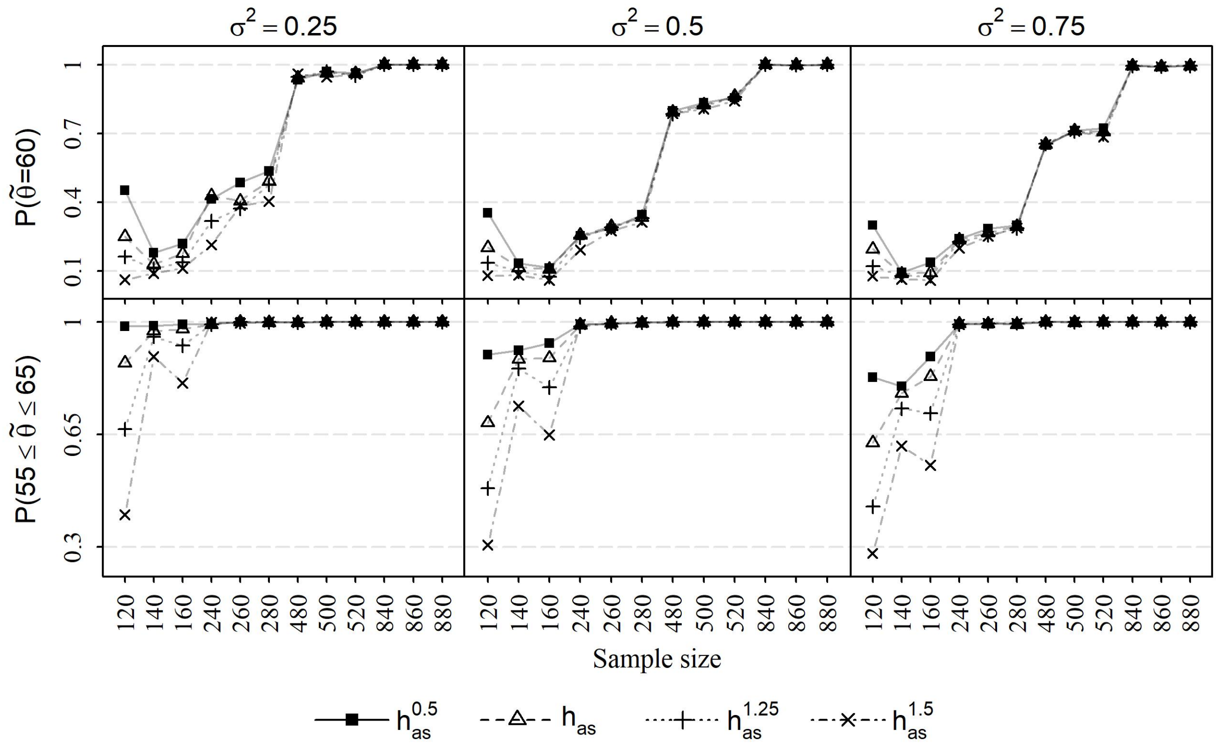
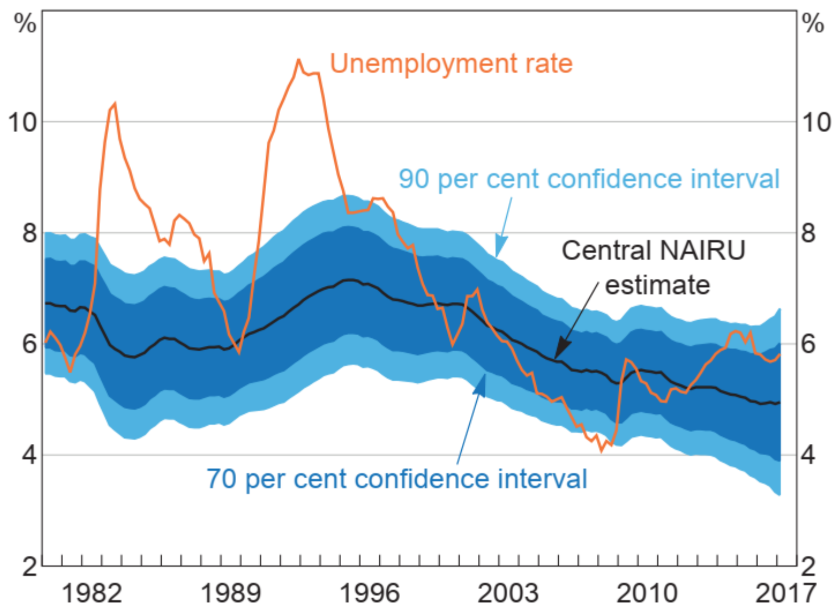


Figure 4.12: Quarterly NAIRU estimates from RBA.



Note: reprinted from Cusbert et al. (2017).

Table 5: Least squares regression outputs.

	Model 1	Model 2	Model 3	Model 4
gap	0.36^{*}	0.51^{***}	0.51^{***}	0.47^{**}
	(0.13)	(0.10)	(0.14)	(0.15)
X_1	-0.78			
	(0.85)			
X_2	1.66			
	(0.85)			
X_3	-1.26			
	(0.85)			
X_4	-1.08			
	(0.85)			
X_5	0.93			
	(0.84)			
X_6	1.04			
	(0.85)			
X_7	2.30^{**}			
	(0.85)			
X_8	-0.29			
	(0.85)			
X_9	-1.05			
	(0.85)			
X_{10}	-0.44			
	(0.84)			
X_{11}	-1.82^{*}			
	(0.85)			
X_{12}	0.08			
	(0.85)			
X_{13}	0.83			
	(0.85)			
R^2	0.49	0.34	0.20	0.16
Adj. R^2	0.30	0.33	0.19	0.15
Num. obs.	52	52	52	52

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

5 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.