Identifying shifts between two regression curves

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

Subhra Sankar Dhar
IIT Kanpur
Department of Mathematics & Statistics
Kanpur 208016, India

Weichi Wu
Tsinghua University
Center for Statistics
Department of Industrial Engineering
10084 Beijing China

August 13, 2019

Abstract

This article studies the problem whether two convex (concave) regression functions modelling the relation between a response and covariate in two samples differ by a shift in the horizontal and/or vertical axis. We consider a nonparametric situation assuming only smoothness of the regression functions. A graphical tool based on the derivatives of the regression functions and their inverses is proposed to answer this question and studied in several examples. We also formalize this question in a corresponding hypothesis and develop a statistical test. The asymptotic properties of the corresponding test statistic are investigated under the null hypothesis and local alternatives. In contrast to most of the literature on comparing shape invariant models, which requires independent data the procedure is applicable for dependent and non-stationary data. We also illustrate the finite sample properties of the new test by means of a small simulation study and a real data example.

AMS subject classification: 62G08, 62G10, 62G20

Keywords and phrases: comparison of curves, nonparametric regression, hypothesis testing

1 Introduction

A common problem in statistical analysis is the comparison of two regression models that relate a common response variable to the same covariates for two different groups. If the two regression functions coincide such statistical inference can be performed on the basis of the pooled sample and therefore it is of interest to test hypotheses of this type. More formally, let

$$Y_{i,1} = m_1(t_{i,1}) + e_{i,1} \ , \ \ i = 1, \dots, n_1 \tag{1.1}$$

$$Y_{j,2} = m_2(t_{j,2}) + e_{j,2} \ , \ \ j = 1, \dots, n_2 \tag{1.2}$$

denote two regression models with real valued responses and predictors $t_{\ell,k}$ and random errors $e_{i,1}$ and $e_{j,2}$. Statistical methodology addressing the question, if the two regression functions m_1 and m_2 coincide, has been investigated by many authors, and there exists an enormous amount of literature addressing this important testing problem [see, for example Hall and Hart (1990); Dette and Munk (1998); Dette and Neumeyer (2001); Neumeyer and Dette (2003) for some early and Vilar-Fernández et al. (2007); Neumeyer and Pardo-Fernández (2009); Maity (2012); Degras et al. (2012); Durot et al. (2013); Park et al. (2014) for some more recent references among many others].

Another interesting question in this context is the comparison of the regression curves up to a certain parametric transformation. Such parametric relationship between two regression curves often can be fitted into various real life examples; for instance, as it is mentioned in Härdle and Marron (1990), the growth curves of children may have a simple parametric relationship between them. It may happen that those curves are realizations of one curve but differ in the time and the vertical axes, and consequently, the difference among those set of regression curves can be measured by two unknown quantities, namely, the horizontal shift (i.e., along the covariate axis) and the vertical scale (i.e., along the response axis).

Many authors have worked on this problem. Exemplary we mention the early work by Härdle and Marron (1990); Carroll and Hall (1993); Rønn (2001) and the more recent references Gamboa et al. (2007); Vimond (2010); Collier and Dalalyan (2015) among others. Several authors proposed tests for the hypotheses that the regression curves coincide up to a certain parametric relationship. The proposed methodology is based on the estimation of the parametric form from the given data. In this article we contribute to this literature and propose a simple method to test the hypothesis

$$H_0 : m_1(x) = m_2(x + c) + d \quad \text{for some constants } c, d \ , \tag{1.3}$$

where m_1 and m_2 are convex (or concave) functions. The assumption of a convex or concave regression function is well justified in several applications. For example, production functions are often assumed to be concave [see Varian (1984)], economic theory implies that utility functions are concave [see Matzkin (1991)] or in finance theory restricts call option prices to be convex [see Ait-Sahalia and Duarte (2003)].

We will show in Section 2 that under the null hypothesis (1.3) the functions $((m_1')^{-1})'$ and $((m_2')^{-1})'$ coincide (here and throughout this paper f' denotes the derivative of the function f and f^{-1} its inverse). This fact is utilized to develop a graphical device to check the assumption (2.2) by estimating the difference $((m_1')^{-1})' - ((m_2')^{-1})'$. For this purpose, we use ideas of Dette et al. (2006) who proposed a very simple estimator of the inverse regression function say f based on a kernel density estimation of the random variable $f(U)$, where U is uniformly distributed random variable on the interval $(0, 1)$, and f is either m_1' or m_2'.

The second contribution of this paper is a formal test for the hypothesis (1.3) in the context of dependent and non-stationary data, which is based on a suitable distance between estimates of the functions $((m_1')^{-1})'$ and $((m_2')^{-1}(t))'$. More precisely, we investigate an L^2-norm of a smooth estimator of the difference $((m_1')^{-1})' - ((m_1')^{-1})'$ and derive the asymptotic distribution of the corresponding test statistic under the null hypotheses and local alternatives. The challenges in deriving these results are twofold. First - in contrast to most of the literature - we allow for a very complex dependence structure of the errors in models (1.1) and (1.2). In particular they can be time dependent and non stationary [see, for example Dahlhaus (1997), Mallat et al. (1998), Ombao et al. (2005), Nason et al. (2000), Zhou and Wu (2009), Vogt (2012) for various definitions of non-stationary time series]. A particular difficulty consists in the proof of the asymptotic distribution of the estimated integrated squared difference, which is (after appropriate standardization) normal, but involves higher order derivatives of the regression functions. As these quantities are very difficult to estimate we develop a bootstrap test, which has very good finite sample properties and is based on a Gaussian approximation used in the proof of the weak convergence of the test statistic.

The rest of the article is organized as follows. Section 2 describes the basic methodology adopted in this article. A new graphical device is proposed for comparing two non-parametric regression functions up a to shift in the covariate and response in Section 2.1. The formal testing problem is considered in Section 2.2, while we give some theoretical justification for these tools in Section 3. A small simulation study is carried out in Section 4, illustrating the finite sample properties of the proposed method and an application is discussed in Section 4.3. Finally, all proofs except of the proof of Lemma 2.1, which justifies our approach, are given in an appendix in Section 5.

2 Methodology

Throughout this paper we consider two data sets $\{Y_{i,1}\}_{i=1,..,n_1}$ and $\{Y_{i,2}\}_{i=1,..,n_2}$ that can be modelled as

$$Y_{i,s} = m_s\Big(\frac{i}{n_s}\Big) + e_{i,s}, \quad i = 1, \dots, n_s, \;\; s = 1, 2 \;, \tag{2.1}$$

the error random variables $\{e_{i,1}\}_{i=1,...,n_1}$ and $\{e_{i,2}\}_{i=1,...,n_2}$ are locally stationary process satisfying some technical conditions that will be described later in Section 3.1, and m_1 and m_2, are unknown sufficiently smooth regression functions. We assume that m_1 and m_2 are convex (the case of concave regression functions can be treated in a similar manner) and are interested to investigate in a hypothesis

$$H_0 : \begin{cases} \text{there exists constants } c \in (0,1) \text{ and } d \in \mathbb{R} \text{ such that} \\ m_1(t) = m_2(t+c) + d, \;\text{ for all } t \in (0, 1-c) \end{cases} \tag{2.2}$$

Notice that we assume that information about the sign of a potential vertical shift can be obtained by visible inspection of the data. A corresponding hypothesis with a vertical shift by a negative constant c can be formulated and treated in a similar way, but the details are omitted for the sake of brevity. A key observation is that under the null hypothesis (2.2) we have

$$((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' = 0, \tag{2.3}$$

and this fact motivates us to propose a test statistic and a graphical device based on the the estimate of $((m_1')^{-1}(t))' - ((m_2')^{-1}(t))'$.

Lemma 2.1 *Assume that the regression functions m_1 and m_2 in (2.1) have a strictly increasing first order derivative on the interval $[0, 1]$, then the following statements are equivalent.*

(1) There exists a constant $c \in (0,1)$ *such that* $m_1(t) = m_2(t+c) + d$ *for all* $t \in (0, 1-c)$*.*

(2) Equation (2.3) *holds for all* $u \in (m_1'(0), m_1'(1-c))$*.*

Proof. If condition (1) holds, then

$$m_1'(t) = m_2'(t+c)$$

for all $t \in (0, 1-c)$. Now consider the equation $m_1'(x) = m_2'(x+c) = u$ for some fixed $u \in (m'(0), m'(1-c))$ and note that both derivatives are strictly increasing. Consequently we obtain for a solution in the interval for $(0, 1-c)$

$$x = (m_1')^{-1}(u) \ ; \quad x + c = (m_2')^{-1}(u) \ .$$

In particular, this yields (subtracting both equations)

$$c = (m_2')^{-1}(u) - (m_1')^{-1}(u) \tag{2.4}$$

for any $u \in (m_1'(0), m_1'(1-c))$. Taking derivatives on both sides of (2.4) gives (2.3) and shows that (1) implies (2).

On the other hand, if condition (2) holds, it follows

$$\int_{m_1'(0)}^{s} ((m_1')^{-1})'(u) du = \int_{m_1'(0)}^{s} ((m_2')^{-1})'(u) du,$$

any $s \in (m_1'(0), m_1'(1-c))$, which yields

$$(m_2')^{-1}(s) = (m_1')^{-1}(s) + c$$

for $s \in (m_1'(0), m_1'(1-c))$, where

$$c = (m_2')^{-1}(m_1'(0)).$$

Applying the function m_2' on both sides finally gives

$$m_2'((m_1')^{-1}(s) + c)) = s = m_1'((m_1')^{-1}(s))$$

for $s \in (m_1'(0), m_1'(1-c))$. Using the notation $(m_1')^{-1}(s) = t$ and integrating with respect to t shows that this is equivalent to (1), which completes the proof of Lemma 2.1. $\square$

2.1 Graphical Device

According to Lemma 2.1, under null hypothesis, the points

$$\{(t, f_1(t) - f_2(t)) \mid t \in (m_1'(0), m_1'(1-c))\}$$

lie on the horizontal axis. In order to construct a graphical device, let $\hat{f}_1$ and $\hat{f}_2$ denote suitably chosen uniformly consistent estimates of the functions $f_1 = ((m_1')^{-1})'$ and $f_2 = ((m_2')^{-1})'$, respectively, let $\hat{m}_1'$ denote an estimate of the derivative m_1', and let $\hat{c}$ be an estimate of the vertical shift c. We now consider a collection of points

$$\mathcal{C}_{n_1,n_2} = \{(t_\ell, \hat{f}_1(t_\ell) - \hat{f}_2(t_\ell))\colon t_\ell \in (\hat{a}+\eta, \hat{b}-\eta); \quad \ell = 1, \ldots, L\}, \tag{2.5}$$

where $\hat{a} = \hat{m_1'}(0)$ and $\hat{b} = \hat{m_1'}(1-\hat{c})$ are estimates of $m_1'(0)$ and $m_1'(1-c)$, respectively, η is a small positive constant and L is a positive integer. Under the null hypothesis, the points of $\mathcal{C}_{n_1,n_2}$ should cluster around the horizontal axis.

Here the necessary estimates can be constructed in various ways. For example, $\hat{f}_1$ and $\hat{f}_2$ can be obtained using a smooth nonparametric estimate of the derivative of the regression function and calculating the derivative of its inverse. The inversion of the nonparametric estimates of the derivatives m_1 and m_2 might be difficult as these functions are usually not monotone. Possible solutions are to construct isotone (smooth) nonparametric estimates of the derivatives as proposed in Mammen (1991) and Hall and Huang (2001) among others and then calculate the inverse. Here we use a more direct approach related to the work of Dette et al. (2006) who proposed methodology for nonparametric estimation of a monotone regression function based on monotone rearrangements.

To be precise, let K denote a kernel function, $b_{n,1}$, $b_{n,2}$ two bandwidths and define the estimate of the regression function m_s and its derivative m_s' for $t \in [b_{n,s}, 1-b_{n,s}]$ by

$$(\hat{m}_s(t), b_{n,s}\hat{m}_s'(t))^\top = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} \Big(Y_{i,s} - \beta_0 - \beta_1\Big(\frac{i}{n_s} - t\Big)\Big)^2 K\Big(\frac{i/n_s - t}{b_{n,s}}\Big) , \quad s = 1, 2 \tag{2.6}$$

and $\hat{m}_s'(t) = \hat{m}_s'(b_{n,s})$ for $0 \le t \le b_{n,s}$, while $\hat{m}_s'(t) = \hat{m}_s'(1-b_{n,s})$ for $1-b_{n,s} \le t \le 1$. Let K_d be a kernel function, h_d a sufficiently small bandwidth and N a large positive integer (note that this is not the sample size). We define the estimates

$$\hat{f}_1(t) = \frac{1}{Nh_{d,1}} \sum_{i=1}^{N} K_d\Big(\frac{\hat{m}_1'(\frac{i}{N}) - t}{h_{d,1}}\Big), \tag{2.7}$$

$$\hat{f}_2(t) = \frac{1}{Nh_{d,2}} \sum_{i=1}^{N} K_d\Big(\frac{\hat{m}_2'(\frac{i}{N}) - t}{h_{d,2}}\Big). \tag{2.8}$$

for $f_1(t) = ((m_1')^{-1})'(t)$ and $f_2(t) = ((m_2')^{-1})'(t)$, respectively. For the motivation of this definition note that, if the estimates $\hat{m}_s'$ are consistent for m_s' $(s = 1, 2)$, then we can replace for a sufficiently large sample size the estimates by the unknown regression functions, and obtain by a Riemann approximation (if $N \to \infty$, $h_d \to 0$)

$$\begin{aligned}\hat{f}_s(t) &\approx \frac{1}{Nh_d} \sum_{i=1}^{N} K_d\Big(\frac{m_s'(\frac{i}{N}) - t}{h_d}\Big) \approx \frac{1}{h_d} \int_0^1 K_d\Big(\frac{m_s'(x) - t}{h_d}\Big) dx \\ &= \int_{(m_s'(0)-t))/h_d}^{(m_s'(1)-t))/h_d} K_d(u)((m_s')^{-1})'(t + uh_d) du \approx ((m_s')^{-1})'(t) \mathbf{1}\{m_s'(0) < t < m_s'(1)\}.\end{aligned}$$

where $\mathbf{1}(A)$ denotes the indicator functions of the set A and we have used the fact that m_ℓ' is non-decreasing (see Dette et al. (2006) for more details). Finally, the estimate of $(m_2')^{-1}$ can be obtained by integration, that is

$$\hat{g}_2(x) = \int_{m_2'(0)}^{x} \hat{f}_2(t) dt$$

and using (2.4) we obtain an estimate

$$\hat{c} = \frac{1}{1 - \tilde{c}} \int_0^{(1-\tilde{c})} (\hat{g}_2(\hat{m}_1'(u)) - u) du. \tag{2.9}$$

of the vertical shift c. Here $\hat{m}_1'$ is the estimate of the derivative of m_1 defined in (2.6) and

$$\tilde{c} = \hat{g}_2(\hat{m}_1'(0)).$$

is a preliminary consistent estimator of c. The resulting estimates for $a = m_1'(0)$ and $b = m_1(1 - c)$ are then given by

$$\hat{a} \;=\; \hat{m}_1'(0) \ , \quad \hat{b} = \hat{m}_1'(1 - \hat{c})$$

(note that we assume that $c > 0$). We will prove in Theorem 3.1 below that under the null hypothesis (2.2) the points of the set $\mathcal{C}_{n_1,n_2}$ will concentrate around the horizontal axis when the sample sizes are sufficiently large. Therefore we propose a graphical device that plots the points of the set $\mathcal{C}_{n_1,n_2}$.

Example 2.1 We consider the regression models (2.1) with independent standard normal distributed errors and different regression functions where the sample sizes are $n_1 = n_2 = 100$. In this numerical study, $N = 100$, $h_{d,N} = N^{-1/3}$, and bandwidths $b_{n_1,1}$ and $b_{n_2,2}$ are chosen as described in Section 4. The set $\mathcal{C}_{n_1,n_2}$ consists of $L = 1000$ equally spaced points from the interval $(\hat{a} + \eta, \hat{b} - \eta)$, where $\eta = 0.01$. To compute the local linear estimators we use the R package named 'locpol'. The following models are considered in this example:

$$m_1(x) = (x - 0.4)^2 \quad \text{and} \quad m_2(x) = (x - 0.3)^2 - 0.2, \tag{2.11}$$

$$m_1(x) = (x - 0.4)^2 \quad \text{and} \quad m_2(x) = x^3, \tag{2.12}$$

$$m_1(x) = \sin(-\pi x) \quad \text{and} \quad m_2(x) = \sin(-\pi(x + 0.1)) + \frac{1}{4}, \tag{2.13}$$

$$m_1(x) = \sin(-\pi x) \quad \text{and} \quad m_2(x) = -\cos(\pi x). \tag{2.14}$$

Note that examples (2.11) and (2.13) correspond to the null hypothesis, while (2.12) and (2.14) represent alternatives. The corresponding plots of the set $\mathcal{C}_{n_1,n_2}$ are shown in Figure 2.1, where the the left panels clearly support the null hypothesis of a vertical and horizontal shift between the regression functions (the points are clustered around the x-axis). On the other hand, the panels on the right give clear evidence that the null hypothesis (2.2) is not true.

2.2 Investigating shifts in the regression functions by testing

The graphical device discussed in the previous section provides a simple tool of visual examination of the null hypothesis (2.2), but does not give any information about the statistical uncertainty of a decision. In this section we will add to this tool a statistic which can be used to rigorously test the null hypothesis (2.2) at a controlled type I error. Recalling the definition of the estimates (2.7) and (2.8) of $((m_1')^{-1})'(t)$ and $((m_2')^{-1})'(t)$, we propose to reject the null hypothesis (2.2) for large values of the statistic

$$T_{n_1,n_2} = \int \left(\hat{f}_1(t) - \hat{f}_2(t)\right)^2 \hat{w}(t)dt, \tag{2.15}$$

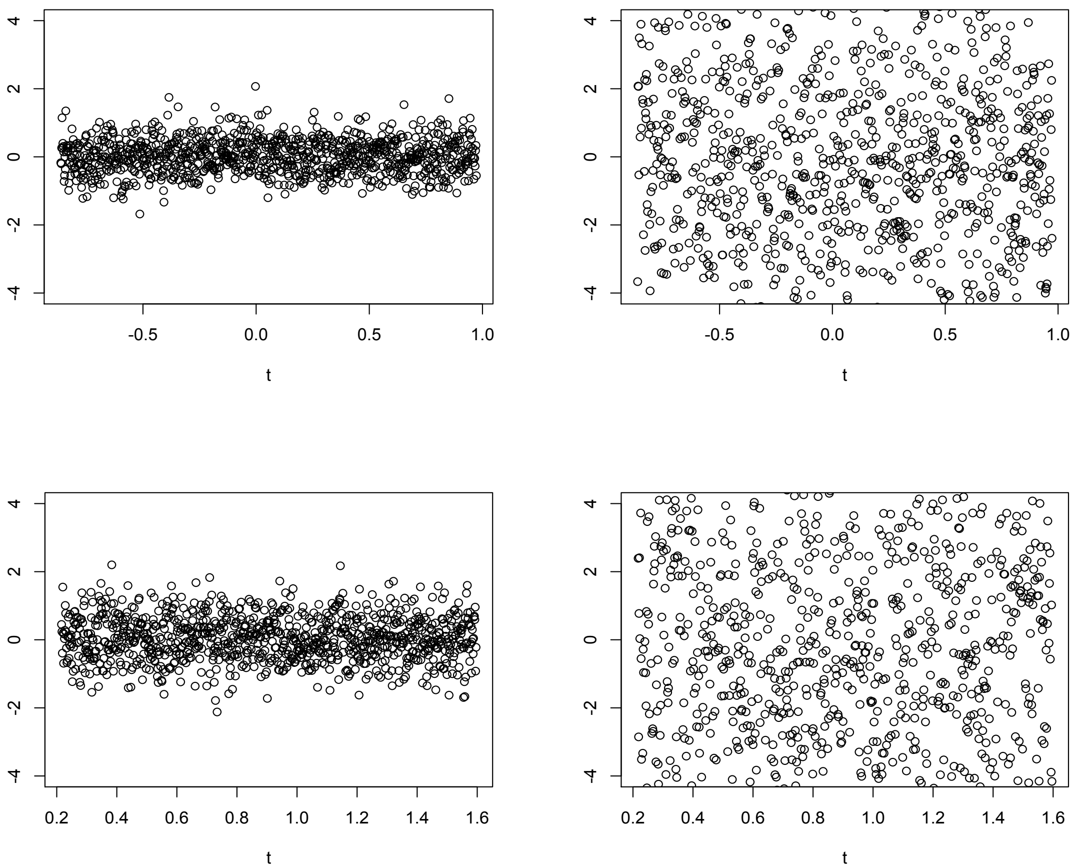

Figure 1: *Plots of the set* $\mathcal{C}_{n_1,n_2}$ *for different examples. The panels on the left correspond to the models* (2.11) *and* (2.13) *(null hypothesis) and the panels on the right correspond to the models* (2.12) *and* (2.14) *(alternative).*

where the weight function is defined by

$$\hat{w}(t) = \mathbf{1}(\hat{a} + \eta \leq t \leq \hat{b} - \eta),$$

η is a small positive constant and $\hat{a}$ and $\hat{b}$ are defined in (2.10). In fact, $\hat{w}(t)$ is a consistent estimator of the deterministic weight function

$$w(t) = \mathbf{1}(a + \eta \leq t \leq b - \eta), \tag{2.16}$$

where $a = m_1'(0), b = m_1'(1-c)$.

Remark 2.1 For the construction of the test statistic, other distances between the functions $((\hat{m}_1')^{-1})'(t)$ and $((\hat{m}_2')^{-1})'(t)$ could be considered as well. For the L^2 distance, the derivation of the asymptotic distribution of the statistic T_{n_1,n_2} is already very complicated (see Section 5 for details), but we can make use of a central limit theorem for random quadratic forms [see de Jong (1987)]. Other distances such as the supremum or L^1 distance could be considered as well with additional technical arguments.

3 Asymptotic properties

Before stating the asymptotic distribution of T_{n_1,n_2}, a few concepts and assumptions are stated for model (2.1). For the dependence structure, we use a common concept non-stationarity, which will be described first.

3.1 Locally stationary processes and basic assumptions

Recall the definition of model (2.1) and denote by $\{\mathbf{e}_i\}_{i\in\mathbb{N}} = \{(\varepsilon_{i,1}, \varepsilon_{i,2})^\top\}_{i\in\mathbb{N}}$ the vector of errors. Note that $\{\mathbf{e}_i\}_{i\in\mathbb{N}}$ defines a triangular array although this is not reflected in our notation. In particular we assume $\{\mathbf{e}_i\}_{i\in\mathbb{N}}$ is a locally stationary process in the sense of Zhou and Wu (2009) such that it has the form

$$\mathbf{e}_i = \mathbf{G}(i/n, \boldsymbol{\mathcal{F}}_i) = ((G_1(i/n, \mathcal{F}_i), G_2(i/n, \mathcal{G}_i)^\top), 1 \leq i \leq n, \tag{3.1}$$

where $\boldsymbol{G} : [0,1] \times \mathbb{R}^\infty \to \mathbb{R}^2$ is a measurable nonlinear filter, $\boldsymbol{\mathcal{F}}_i = (..., \boldsymbol{\epsilon}_{i-1}, \boldsymbol{\epsilon}_i)$ is a filtration and $\{\boldsymbol{\epsilon}_i = (\varepsilon_{i,1}, \varepsilon_{i,2})^\top\}_{i\in\mathbb{N}}$ a sequence of independent identically distributed random variables. In (3.1) G_1 and G_2 are the marginal filters and $\mathcal{F}_i = (...., \varepsilon_{i-1,1}, \varepsilon_{i,1})$, $\mathcal{G}_i = (..., \varepsilon_{i-1,2}, \varepsilon_{i,2})$.

Moreover for any p-dimensional vector $\mathbf{v} = (v_1, ..., v_p)^\top$ we define $|\mathbf{v}| = \sqrt{\sum_{i=1}^p v_i^2}$, $\|\mathbf{v}\|_4 = (\mathbb{E}(|\mathbf{v}|^4))^{1/4}$ and make the following basic assumptions.

Assumption 3.1

(a) $\mathbb{E}(\mathbf{G}(t, \boldsymbol{\mathcal{F}}_0)) = 0$ *for* $t \in [0,1]$, *and* $\sup_{t\in[0,1]} \|\mathbf{G}(t, \boldsymbol{\mathcal{F}}_0)\|_4 < \infty$.

(b) $\sup_{0\le t<s\le 1} \|\mathbf{G}(t, \boldsymbol{\mathcal{F}}_0) - \mathbf{G}(s, \boldsymbol{\mathcal{F}}_0)\|_4 < \infty$.

(c) Let $\{\boldsymbol{\epsilon}_i^*\}_{i\in\mathbb{N}}$ *denote an independent copy of* $\{\boldsymbol{\epsilon}_i\}_{i\in\mathbb{N}}$ *and define the filtration* $\boldsymbol{\mathcal{F}}_i^* = (\boldsymbol{\epsilon}_{-\infty}, ..., \boldsymbol{\epsilon}_{-1}, \boldsymbol{\epsilon}_0^*, ..., \boldsymbol{\epsilon}_i)$. *There exists a constant* $\rho \in (0,1)$ *such that for any* $k \ge 0$,

$$\delta_4(k) := \sup_{t\in[0,1]} \|\mathbf{G}(t, \boldsymbol{\mathcal{F}}_k) - \mathbf{G}(t, \boldsymbol{\mathcal{F}}_k^*)\|_4 = O(\rho^k) \ .$$

(d) There exists a constant $\nu_0 > 0$ *such that the* 2×2 *matrix* $\Sigma^2(t) - \nu_0 I_2$ *is strictly positive definite for any* $t \in [0,1]$, *where* I_2 *is the* 2×2 *identity matrix, and* $\Sigma^2(t)$ *is the long run variance of the locally stationary process defined as*

$$\Sigma^2(t) = \sum_{s=0}^{\infty} \mathbb{E}\left(\mathbf{G}(t, \boldsymbol{\mathcal{F}}_0)\mathbf{G}(t, \boldsymbol{\mathcal{F}}_s)^\top\right) .$$

(e) $\Sigma^2(t)$ *is a diagonal matrix with entities* $\sigma_1^2(t)$ *and* $\sigma_2^2(t)$ *(the long-run variances of process* $G_1(\cdot, \mathcal{F}_i)$ *and* $G_2(\cdot, \mathcal{G}_i)$*).*

Note that it follows from the definition of $\delta_4(k)$ that $\delta_4(k) = 0$ for $k \le 0$. Assumptions (d) and (e) ensure that $\sigma_1^2(t)$ and $\sigma_2^2(t)$ are non-degenerate such that $\inf_{t\in[0,1]} \sigma_s^2(t) > 0$ $(s = 1, 2)$.

Recalling the definition of the local linear estimator for the derivatives m_1' and m_2' in (2.6) we make the following assumptions.

Assumption 3.2

(a) The kernel K is a symmetric and twice differentiable function with compact support, say $[-1, 1]$. Furthermore, $\int_{-1}^1 K(x)dx = 1$

(b) The kernel K_d is an even density with compact support, say $[-1, 1]$.

Assumption 3.3

(a) $m_1, m_2 \in \mathcal{C}^{2,1}[0,1]$, where $\mathcal{C}^{2,1}[0,1]$ represents the set of twice continuously differentiable functions, whose second order derivative is Lipschitz continuous on the interval $[0,1]$.

Assumption 3.4 For $s = 1, 2$ let

$$\pi_{n,s} = \frac{\log n}{\sqrt{nb_{n,s}}b_{n,s}} + \frac{n^{1/4}\log^2 n}{nb_{n,s}^2} + b_{n,s}^2 \ , \quad \pi'_{n,s} = \frac{n^{1/4}\log^2 n}{nb_{n,s}^2} + b_{n,s}^2$$

and assume that $\pi_{n,s} = o(h_{d,n})$ $(s = 1, 2)$. Further, assume that

$$nb_{n,s}^2 \to \infty \ , \quad nb_{n,s}^4 \log n \Big(\frac{\pi'_{n,s}}{b_{n,s}} + \frac{\pi_{n,s}^3}{h_d^3} + h_d + \frac{1}{Nh_d} \Big)^2 = o(1),$$
$$\bar{\omega}_n b_{n,s}^{-1/2} \log^2 n = o(1) \ ,$$

where

$$\bar{\omega}_{n,s} = \frac{\log n}{\sqrt{nb_{n,s}}b_{n,s}} + \frac{n^{1/4}\log^2 n}{nb_{n,s}^2} + b_{n,s} \ , \quad s = 1, 2. \tag{3.2}$$

3.2 Asymptotic properties of $\mathcal{C}_{n_1,n_2}$

The following theorem describes the asymptotic properties of the set $\mathcal{C}_{n_1,n_2}$ defined in (2.5) if it is used with the local linear estimates (2.6) for the derivatives m'_1 and m'_2. It basically gives a theoretical justification for the use of the graphical device proposed in Section 2.1. The proof can be found in Section 5.2.

Theorem 3.1 *Define for $\epsilon > 0$ the set*

$$L(\epsilon, g) = \{(x, y) : x \in [m'_1(0) + \eta, m'_1(1 - c) - \eta], |y - g(x)| \leq \epsilon\}.$$

where $g = ((m'_1)^{-1})' - ((m'_2)^{-1})'$. If Assumptions 3.1–3.4 are satisfied, then we have

$$\lim_{n_1, n_2 \to \infty} \mathbb{P}[\mathcal{C}_{n_1,n_2} \subset L(\epsilon, g)] = 1.$$

Under the null hypothesis we have $g \equiv 0$ and

$$L(\epsilon) := L(\epsilon, 0) = \{(x, y) : x \in [m'_1(0) + \eta, m'_1(1 - c) - \eta], |y| \leq \epsilon\}.$$

Theorem 3.1 shows, that for large sample size the points in the set $\mathcal{C}_{n_1,n_2}$ cluster around the horizontal axis if and only if the null hypothesis holds.

3.3 Weak convergence of the test statistic

In this section, we derive the asymptotic distribution of the statistic T_{n_1,n_2}. For this purpose, we define

$$K^\circ(x) = \frac{K(x)x}{\int_{-1}^{1} K(x)x^2 dx}, \tag{3.3}$$

and obtain the following result. The proof is complicated and can be found in Section 5.3.

Theorem 3.2 *Suppose that Assumption 3.1-3.4 hold,* $n_2/n_1 \to c_2$ *for some constant* $c_2 \in (0,\infty)$ *and assume additionally that*

$$\frac{b_{n,1}}{b_{n,2}} \to r_2 \in (0,\infty).$$

Consider local alternatives of the form

$$((m_1')^{-1})'(t) - ((m_2')^{-1})'(t) = \rho_n g(t) + o(\rho_n),$$

where $g \in \mathcal{C}[a,b]$, $\rho_n = (n_1 b_{n,1}^{9/2})^{-1/2}$ *and the order* $o(\rho_n)$ *of the remainder holds uniformly with respect to* t. *Then as* $n_1, n_2 \to \infty$,

$$n_1 b_{n,1}^{9/2} T_{n_1,n_2} - B_n(g) \Rightarrow \mathcal{N}(0, V_T), \tag{3.4}$$

where the asymptotic bias and variance are given by

$$B_n(g) = \frac{(\int_{-1}^{1} vK_d'(v)dv)^2}{\sqrt{b_{n,1}}}((K^\circ)' * (K^\circ)'(0)) \sum_{s=1}^{2} c_s r_s^5 \int_{\mathbb{R}} \sigma_s^2(u) w(m_s'(u))(m_s''(u))^{-3} du + \int_0^1 g^2(t)w(t)dt,$$

$$V_T = 2\Big(\int_{-1}^{1} vK_d'(v)dv\Big)^4 \sum_{s=1}^{2} c_s^2 r_s^9 \int_{\mathbb{R}} ((K^\circ)' * (K^\circ)'(z))^2 dz \int_{\mathbb{R}} (\sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3})^2 du$$

$c_1 = 1$, $r_1 = 1$ *respectively, and* $(K^\circ)' * (K^\circ)'$ *denotes the convolution of the functions* $(K^\circ)'$ *and* $(K^\circ)'$.

Remark 3.1 Under the null hypothesis, we have $g \equiv 0$ and Theorem 3.2 can be used to construct a consistent asymptotic level α test for the hypotheses in (2.2). More precisely, the null hypothesis is rejected whenever

$$T_{n_1,n_2} > \frac{\hat{B}_n(0) + z_{1-\alpha}\hat{V}_T^{\frac{1}{2}}}{n_1 b_{n,1}^{\frac{9}{2}}} ,$$

where $z_{1-\alpha}$ is the corresponding $(1-\alpha)$-th quantile, and $\hat{B}_n(0)$ and $\hat{V}_T$ are appropriate estimates of the asymptotic bias (for $g(t) \equiv 0$) and variance, respectively. Moreover, Theorem 3.2 also shows that this test is able to detect alternatives converging to the null hypothesis at a rate $\rho_n = (n_1 b_{n,1}^{9/2})^{1/2}$. In this case, the asymptotic power of the test is approximately given by

$$\Phi\Big(\frac{\int g^2(t)w(t)dt}{V_T^{1/2}} - z_{1-\alpha}\Big),$$

where Φ is the cumulative distribution function of the standard normal distribution,

In the case where the sample sizes n_1 and n_2 are equal Theorem 3.2 directly leads to the following corollary.

Corollary 3.1 *If the assumptions of Theorem 3.2 are satisfied, the sample sizes and bandwidths are equal (i.e. $n_1 = n_2$ $b_{n,1} = b_{n,2} = b_n$), the weak convergence in* (3.4) *holds with*

$$B_n(g) = \frac{(\int vK_d'(v)dv)^2}{\sqrt{b_n}}((K^\circ)' * (K^\circ)')(0) \sum_{s=1}^{2} \int_{\mathbb{R}} \sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3}du - \int_0^1 g^2(t)w(t)dt$$

$$V_T = 2\Big(\int_{-1}^{1} vK_d'(v)dv\Big)^4 \sum_{s=1}^{2} \int_{\mathbb{R}} ((K^\circ)' * (K^\circ)'(z))^2 dz \int_{\mathbb{R}} (\sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3})^2 du.$$

4 Implementation and simulation study

We begin with some details regarding the implementation of the test. The calculation of the test statistic requires the specification of the bandwidths and we use the general Cross Validation (GCV) method proposed in Zhou and Wu (2010). Specifically, let $\hat{m}_s(\cdot, b)$ denote the estimate of the regression function m_s with bandwidth b, then we consider

$$\hat{b}_{n_s,s} = \text{argmin}_b \frac{n_s^{-1} \sum_{i=1}^{n_s} (Y_{i,s} - \hat{m}_s(i/n_s, b))^2}{(1 - K(0)(n_s b)^{-1})^2}.$$

As pointed out by Dette et al. (2006), the choice of $h_{d,s}$ has a negligible impact on the the estimators (2.7) and (2.8) (and the corresponding test) as long as it is chosen sufficiently small. As a rule of thumb, we choose $h_{d,s}$ as $n_s^{-1/3}$.

For the estimation of the the long-variance we define for $s = 1, 2$ the partial sum $S_{k,r,s} = \sum_{i=k}^{r} Y_{i,s}$, for some $m \geq 2$

$$\Delta_{j,s} = \frac{S_{j-m+1,j,s} - S_{j+1,j+m,s}}{m},$$

and for $t \in [m/n, 1 - m/n]$

$$\hat{\sigma}_s^2(t) = \sum_{j=1}^{n} \frac{m\Delta_{j,s}^2}{2} \omega(t, j), \quad s = 1, 2, \tag{4.1}$$

where for some bandwidth $\tau_{n,s} \in (0, 1)$,

$$\omega(t, i) = H\Big(\frac{i/n_s - t}{\tau_{n,s}}\Big) / \sum_{i=1}^{n} H\Big(\frac{i/n_s - t}{\tau_{n,s}}\Big).$$

Here H is a symmetric kernel function with compact support $[-1, 1]$ and $\int H(x)dx = 1$. For $t \in [0, m/n_s)$ and $t \in (1 - m/n_s, 1]$ we define $\hat{\sigma}_s^2(t) = \hat{\sigma}_s^2(m/n_s)$ and $\hat{\sigma}^2(t) = \hat{\sigma}^2(1 - m/n_s)$, respectively. The consistency of these estimators has been shown in Theorem 4.4 of Dette and Wu (2019).

4.1 Bootstrap

Although Theorem 3.2 is interesting from a theoretical point of view, it cannot be easily implemented for testing the hypothesis (2.2). The asymptotic bias and variance depend on the long run variances σ_1^2, σ_2^2 and the first and second derivative of the regression functions $m_1(\cdot)$ and $m_2(\cdot)$. In general, these quantities are difficult to estimate. Furthermore, it is well known, that - even in the case of independence - the convergence rate of statistics as considered in Theorem 3.2 is slow (note that the bias in Theorem 3.2 is of order $1/\sqrt{b_{n,1}}$). As an alternative we therefore propose a bootstrap test which does not require the estimation of the derivatives and addresses the problem of slow convergence rate.

The bootstrap procedure is motivated by technical arguments used in the proof of Theorem 3.2 in Section 5. There we show (see equations (5.12) and (5.13)) that under the null hypothesis, the statistic T_{n_1,n_2} can be approximated by the statistic

$$\int_{\mathbb{R}} U_n^2(t)w(t)dt,$$

where

$$U_n(t) = \frac{1}{nNb_{n,1}^2 h_{d,1}^2} \sum_{j=1}^{n_1} \sum_{i=1}^{N} K^\circ\Big(\frac{j/n_1 - i/N}{b_{n,1}}\Big) K_d'\Big(\frac{m_1'(i/N) - t}{h_{d,1}}\Big) \sigma_1\Big(\frac{j}{n_1}\Big) V_{j,1}$$
$$- \frac{1}{nNb_{n,2}^2 h_{d,2}^2} \sum_{j=1}^{n_2} \sum_{i=1}^{N} K^\circ\Big(\frac{j/n_2 - i/N}{b_{n,2}}\Big) K_d'\Big(\frac{m_2'(i/N) - t}{h_{d,2}}\Big) \sigma_2\Big(\frac{j}{n_2}\Big) V_{j,2}$$

and $\{V_{j,1}, j \in \mathbb{Z}\}, \{V_{j,2}, j \in \mathbb{Z}\}$, are sequences of independent standard normal distributed random variables.

Algorithm 4.1

(a) Estimate m_1' and m_2' by (2.6) and estimate the long run variances σ_1^2 and σ_2^2 by (4.1).

(b) Generate B copies of standard normal distributed random variables $\{V_{j,1}^{(B)}\}_{j=1}^{n_1}$, $\{V_{j,2}^{(B)}\}_{j=1}^{n_2}$ and calculate the statistic

$$W_B = \int_{\mathbb{R}} \Big(\frac{1}{nNb_{n,1}^2 h_{d,1}^2} \Xi_1^{(B)}(t) - \frac{1}{nNb_{n,2}^2 h_{d,2}^2} \Xi_2^{(B)}(t)\Big)^2 w(t)dt,$$

where

$$\Xi_1^{(B)}(t) = \sum_{j=1}^{n_1} \sum_{i=1}^{N} K^\circ\Big(\frac{j/n_1 - i/N}{b_{n,1}}\Big) K_d'\Big(\frac{\hat{m}_1'(i/N) - t}{h_{d,1}}\Big) \hat{\sigma}_1\Big(\frac{j}{n_1}\Big) V_{j,1}^{(B)},$$
$$\Xi_2^{(B)}(t) = \sum_{j=1}^{n_2} \sum_{i=1}^{N} K^\circ\Big(\frac{j/n_2 - i/N}{b_{n,2}}\Big) K_d'\Big(\frac{\hat{m}_2'(i/N) - t}{h_{d,2}}\Big) \hat{\sigma}_2\Big(\frac{j}{n_2}\Big) V_{j,2}^{(B)}.$$

(c) Let $W_{(1)} \leq W_{(2)} \leq \ldots \leq W_{(B)}$ be the ordered statistics of $\{W_s, 1 \leq s \leq B\}$. We reject the null hypothesis (2.2) at level α, whenever

$$T_{n_1,n_2} > W_{(\lfloor B(1-\alpha) \rfloor)}. \tag{4.2}$$

The p-value of this test is given by $1 - B^*/B$. where $B^* = \max\{r : W_{(r)} \leq T_{n_1,n_2}\}$.

4.2 Simulated level and power

In this section we illustrate the finite sample properties of the test (4.2) by means of a small simulation study. All presented results are based on 1000 runs and $B = 500$ bootstrap replications. We consider equal sample sizes $n_1 = n_2 = n = 100$, 200 and 500. Throughout this article, the Epanechnikov kernel (e.g., see Silverman (1998)) is considered for all kernels appearing in the test procedure, and we use $N = n$ in (2.7) and (2.8). Besides, $h_{d,N} = n^{-1/3}$, and b_{n_1} and b_{n_2} are chosen as described at the beginning of Section 4.

For $s = 1$ and 2, we consider model (2.1) with the error process

$$G_s(t, \mathcal{F}_i) = 0.6(t - 0.3)^2 G(t, \mathcal{F}_{i-1,s}) + \eta_{i,s}, \tag{4.3}$$

where $\mathcal{F}_{i,s} = (..., \eta_{i-1,s}, \eta_{i,s})$. We assume that $\eta_{i,1}$ are i.i.d standard normal random variables, and $\eta_{i,2}$ are i.i.d. copies of the random variable $t_5/\sqrt{5/3}$, where t_5 denotes the t-distribution with 5 degrees of freedom. For the regression functions we consider the models

$$m_1(x) = (x - 0.4)^2 \quad \text{and} \quad m_2(x) = (x - 0.3)^2 - 0.2, \tag{4.4}$$

$$m_1(x) = \sin(-\pi x) \quad \text{and} \quad m_2(x) = \sin(-\pi(x + 0.1)) + \frac{1}{4}. \tag{4.5}$$

In Table 1 we display the rejection probabilities of the test (4.2), where the level of significance is 5% and 10%. The results show a good approximation of the nominal level in all cases under consideration.

model	$n = 100$	$n = 200$	$n = 500$
(4.4)	0.057	0.054	0.051
(4.5)	0.059	0.057	0.054
(4.4)	0.111	0.108	0.104
(4.5)	0.116	0.112	0.103

Table 1: *The estimated size of the test* (4.2) *for different sample sizes* $n_1 = n_2 = n$. *The level of significance is* 5% *(upper part) and* 10% *(lower part).*

In order to study the power of the test (4.2) we consider the same error processes as in (4.3) and used the regression functions

$$m_1(x) = (x - 0.4)^2 \quad \text{and} \quad m_2(x) = x^3, \tag{4.6}$$

$$m_1(x) = \sin(-\pi x) \quad \text{and} \quad m_2(x) = -\cos(\pi x). \tag{4.7}$$

The simulated power is displayed in Table 2 and the results indicate that the test detects the alternatives reasonably well.

model	$n = 100$	$n = 200$	$n = 500$
(4.6)	0.563	0.647	0.778
(4.7)	0.617	0.744	0.822
(4.6)	0.722	0.847	0.899
(4.7)	0.777	0.868	0.971

Table 2: *The estimated power of the test* (4.2) *for different sample sizes* $n_1 = n_2 = n$. *The level of significance is* 5% *(upper part) and* 10% *(lower part).*

4.3 Real data analysis

In this section, we use the test (4.2) and the graphical device described in Section 2.1 to investigate the validity of assertion (2.2) for growth data of male and female infants. This data set is available from `https://www.cdc.gov/growthcharts/html_charts/lenageinf.htm#males` and consists of the monthly growth of length of male and female infants in the first three years (here $n_1 = n_2 = 37$). The data is depicted in Figure 2 and indicates that the relation between length and age in both groups might be concave. Therefore we model the negative values of this data by two regression models of the form (2.1) with convex regression functions, where group 1 represents the male and group 2 the female infants. For this data, we obtain $\hat{c} = 0.046$ as estimate for the horizontal shift using the statistic (2.9) and $\hat{d} = \hat{m}_1(0) - \hat{m}_2(\hat{c}) = 0.087$ as estimate of the vertical shift d.

We begin illustrating the application of the graphical device described in Section 2.1. In Figure 3 we plot the points of the set $\mathcal{C}_{n_1,n_2}$ in (2.5) using $L = 1000$ equally spaced points in the interval $(\hat{a} + \eta, \hat{b} - \eta)$, where $\hat{a} = \hat{m}_1'(0) = 0.112$, $\hat{b} = \hat{m}_1'(1 - \hat{c}) = 1.362$, and $\eta = 0.001$ is chosen (the smoothing parameters are chosen as described in Section 4). The figure clearly indicates the existence of a vertical and horizontal shift between the regression functions as

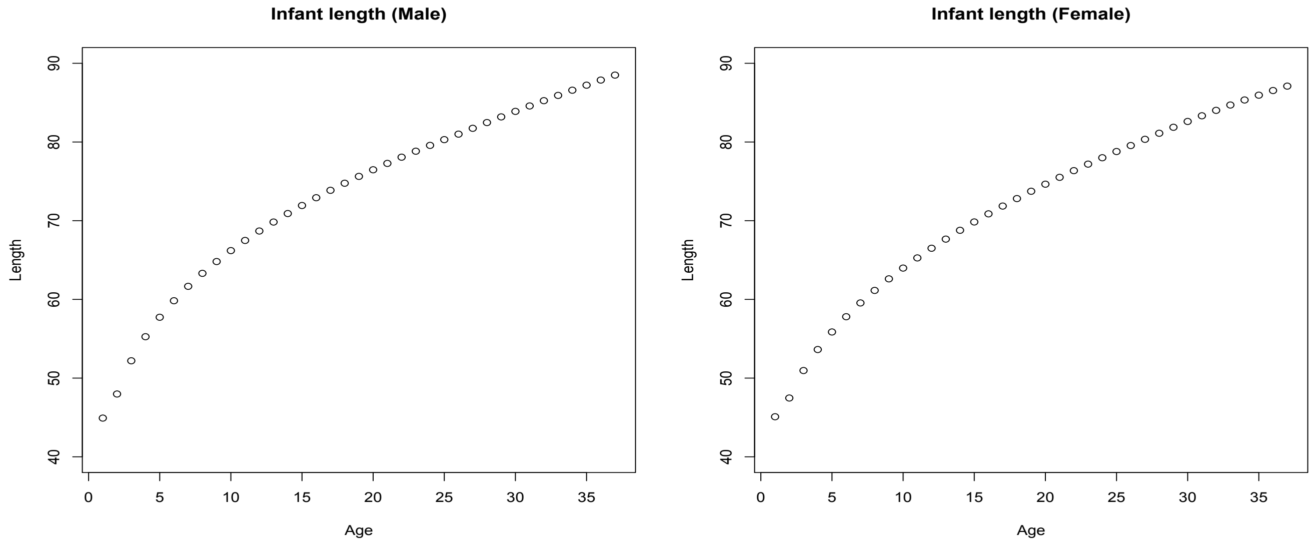

Figure 2: *Plots of the length of the male (right part) and female (left parts) infants for different age.*

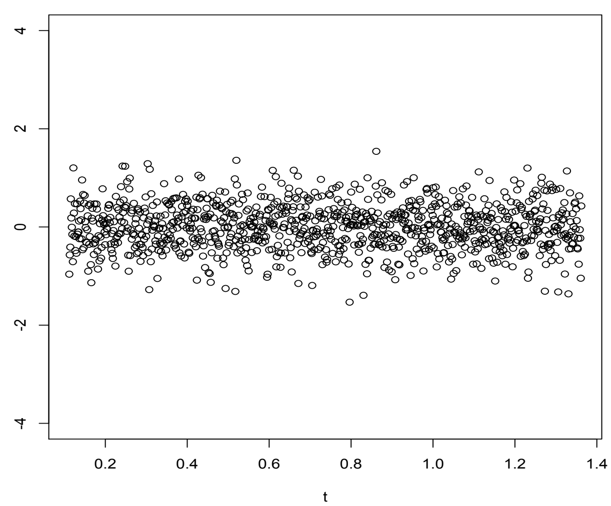

Figure 3: *Plots of* $\mathcal{C}_{n_1,n_2}$ *for the real data described in Section 4.3.*

formulated in the null hypothesis (2.2).

Finally, we also investigate the performance of the test (4.2) for this data set, where all parameters required for the bootstrap test are chosen as described in Section 4. For $B = 500$ bootstrap replications, we obtain the p-value 0.781, which gives no indication to reject the null hypothesis and is consistent with the conclusion made by graphical inspection.

5 Appendix : Proofs

5.1 Preliminaries

In this section, we state a few auxiliary results, which will be used later in the proof. We begin with Gaussian approximation. A proof of this result can be found in Wu and Zhou (2011).

Proposition 5.1 *Let*

$$\boldsymbol{S}_i = \sum_{s=1}^{i} \mathbf{e}_i,$$

and assume that the Assumption 3.1 is satisfied. Then on a possibly richer probability space, there exists a process $\{\mathbf{S}_i^\dagger\}_{i\in\mathbb{Z}}$ *such that*

$$\{\boldsymbol{S}_i^\dagger\}_{i=0}^n \overset{\mathcal{D}}{=} \{\boldsymbol{S}_i\}_{i=0}^n$$

(equality in distribution), and a sequence of independent 2*-dimensional standard normal distributed random variables* $\{\mathbf{V}_i\}_{i\in\mathbb{Z}}$*, such that*

$$\max_{1\le j\le n} \Big| \sum_{i=1}^{j} \mathbf{S}_i^\dagger - \sum_{i=1}^{j} \Sigma(i/n)\mathbf{V}_i \Big| = o_p(n^{1/4}\log^2 n),$$

where $\Sigma(t)$ *is the square root of the long-run variance matrix* $\Sigma^2(t)$ *defined in Assumption 3.1.*

Proposition 5.2 *Let Assumption 3.1 and 3.2 be satisfied.*

(i) For $s = 1, 2$ *we have*

$$\sup_{t\in[b_{n,s},1-b_{n,s}]} \Big| \hat{m}_s'(t) - m_s'(t) - \frac{1}{n_s b_{n,s}^2} \sum_{i=1}^{n_s} K^\circ\Big(\frac{i/n_s - t}{b_{n,s}}\Big) e_{i,s} \Big| = O_P\Big(\frac{1}{n_s b_{n,s}^2} + b_{n,s}^2\Big) \quad (5.1)$$

where the kernel K° is defined in (3.3).

(ii) For $s = 1, 2$

$$\sup_{t\in[b_{n,s},1-b_{n,s}]} \Big| \frac{1}{n_s b_{n_s}^2} \sum_{i=1}^{n_s} K^\circ\Big(\frac{i/n_s - t}{b_{n,s}}\Big)\Big(e_{i,s} - \sigma_s(i/n)V_{i,s}\Big)\Big| = o_p\Big(\frac{\log^2 n_s}{n_s^{3/4} b_{n,s}^2}\Big), \quad (5.2)$$

where $\{V_{i,s},\ i = 1, \ldots, n_s, s = 1, 2\}$ denotes a sequence of independent standard normal distributed random random variables.

(iii) For $s = 1, 2$ we have

$$\sup_{t\in[b_{n,s},1-b_{n,s}]} |\hat{m}_s'(t) - m_s'(t)| = O_p\Big(\frac{\log n_s}{\sqrt{n_s b_{n,s}} b_{n,s}} + \frac{\log^2 n_s}{n_s^{3/4} b_{n,s}^2} + b_{n,s}^2\Big). \quad (5.3)$$

(iv) For $s = 1, 2$ we have

$$\sup_{t\in[0,b_{n,s}]\cup[1-b_{n,s},1]} |\hat{m}_s'(t) - m_s'(t)| = O_p\Big(\frac{\log n_s}{\sqrt{n b_{n,s}} b_{n,s}} + \frac{\log^2 n_s}{n_s^{3/4} b_{n,s}^2} + b_{n,s}\Big). \quad (5.4)$$

Proof:

(i): Define for $s = 1, 2$ and $l = 0, 1, 2$

$$R_{n,s,l}(t) = \frac{1}{n_s b_{n,s}} \sum_{i=1}^{n_s} Y_{i,s} K\Big(\frac{i/n - t}{b_{n,s}}\Big)\Big(\frac{i/n_s - t}{b_{n,s}}\Big)^l,$$
$$S_{n,s,l}(t) = \frac{1}{n_s b_{n,s}} \sum_{i=1}^{n_s} K\Big(\frac{i/n_s - t}{b_{n,s}}\Big)\Big(\frac{i/n_s - t}{b_{n,s}}\Big)^l$$

Straightforward calculations show that

$$(\hat{m}_s(t), b_{n,s}\hat{m}_s'(t))^\top = S_{n,s}^{-1}(t) R_{n,s}(t) \qquad (s = 1, 2),$$

where

$$R_{n,s}(t) = \begin{pmatrix} R_{n,s,0}(t) \\ R_{n,s,1}(t) \end{pmatrix} \ , \quad S_{n,s}(t) = \begin{pmatrix} S_{n,s,0} & S_{n,s,1} \\ S_{n,s,1} & S_{n,s,2}. \end{pmatrix}$$

Note that Assumption 3.2 gives

$$S_{n,s,0}(t) = 1 + O\Big(\frac{1}{n_s b_s}\Big), S_{n,s,1}(t) = O\Big(\frac{1}{n_s b_{n,s}}\Big), S_{n,s,2}(t) = \int_{-1}^{1} K(x)x^2 dx + O\Big(\frac{1}{n_s b_{n,s}}\Big)$$

uniformly with respect to $t \in [b_{n,s}, 1 - b_{n,s}]$. The first part of the proposition now follows by a Taylor expansion of $R_{n,s,l}(t)$.

(ii): The fact asserted in (5.2) follows from (5.1), Proposition 5.1, the summation by parts formula and similar arguments to derive equation (44) in Zhou (2010).

(iii) + (iv): Following Lemma 10.3 of Dette and Wu (2019), we have

$$\sup_{t\in[b_{n,s},1-b_{n,s}]} \Big|\frac{1}{n_s b_{n,s}} \sum_{i=1}^{n_s} K^\circ\Big(\frac{i/n_s - t}{n_s b_{n,s}}\Big)\Big(\sigma_s(\frac{i}{n_s})V_{i,s}\Big)\Big| = O_p\Big(\frac{\log n_s}{\sqrt{n_s b_{n,s}}}\Big). \tag{5.5}$$

Finally, (5.3) and (5.4) follow from (5.1) (5.2) and (5.5), which completes the proof of Proposition 5.2. □

5.2 Proof of Theorem 3.1

We only prove the result in the case $g \equiv 0$. The general case follows by the same arguments. Under Assumptions 3.1 and 3.2, it follows from the proof of Theorem 4.1 in Dette and Wu (2019) that

$$\sup_{t\in(a+\eta,b-\eta)} \big[\big(\hat{f}_1(t) - \hat{f}_2(t)\big) - \big(((m_1')^{-1}(t))' - ((m_2')^{-1})'(t)\big)\big] \to 0$$

in probability, where $\hat{f}_1^{-1}(t)$ and $\hat{f}_2(t)$ are defined in (2.7) and (2.8), respectively. Next, since under the null hypothesis (2.2), $((m_1')^{-1}(t))' - ((m_2')^{-1})'(t) = 0$ for all $t \in (a+\eta, b-\eta)$, (See Lemma 2.1) we have under the null hypothesis,

$$\sup_{t\in(a+\eta,b-\eta)} \big[\hat{f}_1(t) - \hat{f}_2(t)\big] \to 0$$

in probability. In other words, under H_0, for any $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\Big[\sup_{t\in(a+\eta,b-\eta)} \big|\hat{f}_1(t) - \hat{f}_2(t)\big| < \epsilon\Big] = 1,$$

and hence, under the null hypothesis $g \equiv 0$, we have $\mathbb{P}[\mathcal{C}_{n_1,n_2} \subset L(\epsilon)] = 1$. □

5.3 Proof of Theorem 3.2

To simplify the notation, we prove Theorem 3.2 in the case of equal sample sizes and equal bandwidths. The general case follows by the same arguments with an additional amount of notation. In this case $c_2 = r_2 = 1$ and we omit the subscript in bandwidths if no confusion arises, for example we write $n_1 = n_2 = n$, $b_{n,1} = b_{n,2} = b_n$ and use a similar notation for other symbols depending on the sample size. In particular, we write T_n for T_{n_1,n_2} if $n = n_1 = n_2$.

Define the statistic

$$\tilde{T}_n = \int \Big(\hat{f}_1(t) - \hat{f}_2(t)\Big)^2 w(t)dt$$

which is obtained from T_n by replacing the weight function $\hat{w}$ in (2.15) by its deterministic analogue (2.16). We shall show Theorem 3.2 in two steps proving the assertions

$$nb_n^{9/2}\tilde{T}_n - B_n(g) \Rightarrow \mathcal{N}(0, V_T) \tag{5.6}$$

$$nb_n^{9/2}(T_n - \tilde{T}_n) = o_p(1). \tag{5.7}$$

5.3.1 Proof of (5.6)

By simple algebra, we obtain the decomposition

$$\tilde{T}_n = \int (I_1(t) - I_2(t) + II(t))^2 w(t)dt,$$

where for $s = 1, 2$

$$I_s(t) = \frac{1}{Nh_d}\sum_{i=1}^{N}\Big(K_d\Big(\frac{\hat{m}_s'(i/N) - t}{h_d}\Big) - K_d\Big(\frac{m_s'(i/N) - t}{h_d}\Big)\Big), \tag{5.8}$$

$$II(t) = \frac{1}{Nh_d}\sum_{i=1}^{N}\Big(K_d\Big(\frac{m_1'(i/N) - t}{h_d}\Big) - K_d\Big(\frac{m_2'(i/N) - t}{h_d}\Big)\Big). \tag{5.9}$$

Observing the estimate on page 471 of Dette et al. (2006) it follows

$$\frac{1}{Nh_d}\sum_{i=1}^{N} K_d\Big(\frac{m_s'(i/N) - t}{h_d}\Big) = \Big(((m_s')^{-1}(t))' + O\Big(h_d + \frac{1}{Nh_d}\Big)\Big) \ ,$$

$(s = 1, 2)$ which yields the estimate

$$II(t) = ((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' + O\Big(h_d + \frac{1}{Nh_d}\Big) \tag{5.10}$$

uniformly with respect to $t \in [a+\eta, b-\eta]$. For the two other terms we use a Taylor expansion and obtain the decomposition

$$I_s(t) = I_{s,1}(t) + I_{s,2}(t) \quad (s = 1, 2),$$

where

$$I_{s,1}(t) = \frac{1}{Nh_d^2} \sum_{i=1}^{N} K_d'\Big(\frac{m_s'(i/N) - t}{h_d}\Big)(\hat{m}_s'(i/N) - m_s'(i/N)),$$
$$I_{s,2}(t) = \frac{1}{2Nh_d^3} \sum_{i=1}^{N} K_d''\Big(\frac{m_s'(i/N) - t + \theta_s(\hat{m}_s'(i/N) - m_s(i/N))}{h_d}\Big)(\hat{m}_s'(i/N) - m_s'(i/N))^2$$

for some $\theta_s \in [-1, 1]$ $(s = 1, 2)$. By part (iii) and (iv) of Proposition 5.2 and the same arguments that were used in the online supplement of Dette and Wu (2019), to obtain the bound for the term $\Delta_{2,N}$ in the proof of their Theorem 4.1 it follows that

$$I_{s,2}(t) = O_p\Big(\frac{\pi_n^2}{h_d^3}(h_d + \pi_n)\Big) = O_p\Big(\frac{\pi_n^3}{h_d^3}\Big) \ (s = 1, 2), \tag{5.11}$$

uniformly with respect to $t \in [a+\eta, b-\eta]$. Here we used the fact that the number of non-zero summands in $I_{s,2}(t)$ is of order $O(h_d + \pi_n)$.

Next, for the investigation of the difference $I_{1,1}(t) - I_{2,1}(t)$, we define $\mathbf{m}' = (m_1, m_2)$ and consider the vector

$$K_d'\Big(\frac{\mathbf{m}'(i/N) - t}{h_d}\Big) = \Big(K_d'\Big(\frac{m_1'(i/N) - t}{h_d}\Big), -K_d'\Big(\frac{m_2'(i/N) - t}{h_d}\Big)\Big)^\top.$$

By part (i) and (ii) of Proposition 5.2, it follows that there exists independent 2-dimensional standard normal distributed random vectors $\mathbf{V}_i$ such that

$$I_{1,1}(t) - I_{2,1}(t) = \frac{1}{nNb_n^2h_d^2} \sum_{j=1}^{n} \sum_{i=1}^{N} K^\circ\Big(\frac{j/n - i/N}{b_n}\Big)(K_d')^T\Big(\frac{\mathbf{m}'(i/N) - t}{h_d}\Big)\Sigma(j/n)\mathbf{V}_j$$
$$+ O_p(\pi_n' h_d^{-1}).$$

uniformly with respect to $t \in [a+\eta, b-\eta]$. Combining this estimate with equations (5.10) and (5.11), it follows

$$T_n = \int \big(U_n(t) + ((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' + R_n^\dagger(t)\big)^2 w(t)dt, \tag{5.12}$$

where

$$U_n(t) = \frac{1}{nNb_n^2h_d^2}\sum_{j=1}^n\sum_{i=1}^N K^\circ\Big(\frac{j/n - i/N}{b_n}\Big)(K_d')^\top\Big(\frac{\mathbf{m}'(i/N)-t}{h_d}\Big)\Sigma(j/n)\mathbf{V}_j, \tag{5.13}$$

and the remainder $R_n^\dagger(t)$ can be estimated as follows

$$\sup_{t\in[a+\eta,b-\eta]}|R_n^\dagger(t)| = O_p\Big(\frac{\pi_n'}{h_d} + \frac{\pi_n^3}{h_d^3} + h_d + \frac{1}{Nh_d}\Big). \tag{5.14}$$

We now study the asymptotic properties of to the quantities

$$nb_n^{9/2}\int (U_n(t))^2 w(t)dt, \tag{5.15}$$

$$nb_n^{9/2}\int U_n(t)((m_1^{-1}(t))' - (m_2^{-1}(t))')w(t)dt, \tag{5.16}$$

$$nb_n^{9/2}\int U_n(t)R_n^\dagger(t)w(t)dt, \tag{5.17}$$

which determine the asymptotic distribution of T_n since the bandwidth conditions yield under local alternatives in the case $(m_1^{-1}(t))' - (m_2^{-1}(t))' = \rho_n g(t)$,

$$nb_n^{9/2}\int \rho_n^2(t)w(t) = \int g^2(t)w(t)dt, \tag{5.18}$$

and the other parts of the expansion are negligible, i.e.,

$$nb_n^{9/2}\int (R_n^\dagger(t))^2 w(t)dt = o(1), \tag{5.19}$$

$$nb_n^{9/2}\int \rho_n g(t)R_n^\dagger(t)w(t)dt = o(1). \tag{5.20}$$

Asymptotic properties of (5.15)**:** To address the expressions related to $U_n(t)$ in (5.15) - (5.17) note that

$$U_n(t) = U_{n,1}(t) - U_{n,2}(t),$$

where

$$U_{n,s}(t) = \frac{1}{nNb_n^2h_d^2}\sum_{j=1}^{n}\sum_{i=1}^{N} K^\circ\Big(\frac{j/n - i/N}{b_n}\Big)K_d'\Big(m_s'(i/N) - th_d\Big)\sigma_s(j/n)V_{j,s}$$

for $s = 1, 2$, and $\{V_{j,s}\}$ are independent standard normal distributed random variables. In order to simplify the notation, we define the quantities

$$U_{n,s}(t) = \sum_{j=1}^{n} G(m_s'(\cdot), j, t)V_{j,s} \qquad (s = 1, 2),$$

where

$$G(m_s'(\cdot), j, t) = \frac{1}{nNb_n^2h_d^2}\sum_{i=1}^{N} K^\circ\Big(\frac{j/n - i/N}{b_n}\Big)K_d'\Big(\frac{m_s'(i/N) - t}{h_d}\Big)\sigma_s(j/n).$$

A straightforward calculation (using the change of variable $v = (m_s'(u) - t)/h_d$) shows that

$$\begin{aligned} G(m_s'(\cdot), j, t) &= \frac{1}{nb_n^2h_d^2}\int_0^1 K^\circ\left(\frac{j/n - u}{b_n}\right) K_d'\left(\frac{m_s'(u) - t}{h_d}\right)\sigma_s(j/n)du + O(\delta_n) \\ &= \frac{1}{nb_n^2h_d}\sigma_s(j/n)\int_{\mathcal{A}_s(t)} K_d'(v)((m_s')^{-1}(t + h_dv))'K^\circ\left(\frac{j/n - (m_s')^{-1}(t + h_dv)}{b_n}\right)dv \\ &+ O(\delta_n), \end{aligned}$$

where the interval $\mathcal{A}_s(t)$ is defined by

$$\mathcal{A}_s(t) = \Big(\frac{m_s'(0) - t}{h_d}, \frac{m_s'(1) - t}{h_d}\Big),$$

the remainder is given by

$$\delta_n = O\Big(\Big(\frac{1}{nb_n^2h_d^2N}\Big)\mathbf{1}\Big(\Big|\frac{j/n - (m_s')^{-1}(t)}{b_n + Mh_d}\Big| \le 1\Big)\Big),$$

and $\mathbf{1}(A)$ denote the indicator function of the set A. As the kernel $K_d'(\cdot)$ has a compact support and is symmetric, it follows by a Taylor expansion for any t with $w(t) \neq 0$

$$\int_{\mathcal{A}_s(t)} K_d'(v)((m_s')^{-1}(t+h_d v))' K^\circ\Big(\frac{j/n-(m_s')^{-1}(t+h_d v)}{b_n}\Big)dv$$
$$= -\frac{h_d}{b_n}(((m_s')^{-1}(t))')^2 (K^\circ)'\Big(\frac{j/n-(m_s')^{-1}(t)}{b_n}\Big)\int K_d'(v)vdv\Big(1+O\Big(b_n+\frac{h_d^2}{b_n^2}\Big)\Big)$$

With the notation

$$\tilde{G}(m_s'(\cdot),j,t) = \frac{-1}{nb_n^3}(K^\circ)'\Big(\frac{j/n-(m_s')^{-1}(t)}{b_n}\Big)\sigma_s(j/n)(((m_s')^{-1})'(t))^2\int vK_d'(v)dv$$

$(s=1,2)$ we thus obtain the approximation

$$\begin{aligned}
\int U_n^2(t)w(t) &= \sum_{s=1}^{2}\sum_{j=1}^{n} V_{j,s}^2 \int G^2(m_s'(\cdot),j,t)^2 w(t)dt \qquad (5.21)\\
&+\sum_{s=1}^{2}\sum_{1\le i\neq j\le n} V_{i,s}V_{j,s}\int G(m_s'(\cdot),i,t)G(m_s'(\cdot),j,t)w(t)dt\\
&-2\sum_{1\le i\le n} V_{i,1}V_{i,2}\int G(m_1'(\cdot),i,t)G(m_2'(\cdot),i,t)w(t)dt\\
&=\sum_{s=1}^{2}\sum_{j=1}^{n} V_{j,s}^2\Big(\int \tilde{G}^2(m_s'(\cdot),j,t)^2 w(t)dt(1+r_{i,s})\Big)\\
&+\sum_{s=1}^{2}\sum_{1\le i\neq j\le n} V_{i,s}V_{j,s}\Big(\int \tilde{G}(m_s'(\cdot),i,t)\tilde{G}(m_s'(\cdot),j,t)w(t)dt(1+r_{i,j,s})\Big)\\
&-2\sum_{1\le i\le n} V_{i,1}V_{i,2}\Big(\int \tilde{G}(m_1'(\cdot),i,t)\tilde{G}(m_2'(\cdot),i,t)w(t)dt(1+r_{i,s}')\Big),
\end{aligned}$$

where the remainder satisfy

$$\max\Big(\max_{i,j,s=1,2}(|r_{i,j,s}|),\max_{i,s=1,2}(|r_{i,s}|),\max_{i,s=1,2}(|r_{i,s}'|)\Big)=o(1).$$

Let us now consider the statistcis $\tilde{U}_{n,s}(t)=\sum_{j=1}^{n}\tilde{G}(m_s'(\cdot),j,t)V_{j,s}$ $(s=1,2)$, and

$$\tilde{U}_n(t)=\tilde{U}_{n,1}(t)-\tilde{U}_{n,2}(t), \qquad (5.22)$$

then, by the previous calculations, it follows that

$$nb_n^{9/2}\Big(\int U_n^2(t)w(t)dt - \int \tilde{U}_n^2(t)w(t)dt\Big) = o_P(1), \tag{5.23}$$

and therefore, we investigate the weak convergence of the statistic $nb_n^{9/2}\int \tilde{U}_n^2(t)w(t)dt$ in the following. For this purpose we use a similar decomposition as in (5.21) and obtain

$$\begin{aligned}
\int \tilde{U}_n^2(t)w(t)dt &= \sum_{s=1}^{2}\int (\tilde{U}_{n,s}(t))^2 w(t)dt - 2\int (\tilde{U}_{n,1}(t)\tilde{U}_{n,2}(t))w(t)dt \\
&= \sum_{s=1}^{2}\sum_{j=1}^{n} V_{j,s}^2 \int \tilde{G}^2(m_s'(\cdot),j,t)^2 w(t)dt \\
&+ \sum_{s=1}^{2}\sum_{1\le i\ne j\le n} V_{i,s}V_{j,s}\int \tilde{G}(m_s'(\cdot),i,t)\tilde{G}(m_s'(\cdot),j,t)w(t)dt \\
&- 2\sum_{1\le i\le n} V_{i,1}V_{i,2}\int \tilde{G}(m_1'(\cdot),i,t)\tilde{G}(m_2'(\cdot),i,t)w(t)dt \\
&:= D_1 + D_2 + D_3,
\end{aligned} \tag{5.24}$$

where the last equation defines D_1, D_2 and D_3 in an obvious manner. Elementary calculations (using a Taylor expansion and the fact that the kernels have compact support) show that

$$\begin{aligned}
\mathbb{E}(D_1) &= \sum_{s=1}^{2}\sum_{j=1}^{n}\int \Big(\frac{-1}{nb_n^3}(K^\circ)'\Big(\frac{j/n-(m_s')^{-1}(t)}{b_n}\Big)\sigma_s(j/n)(((m_s')^{-1})'(t))^2\int vK_d'(v)dv\Big)^2 w(t)dt \\
&= \sum_{s=1}^{2}\sum_{j=1}^{n}\int \Big(\frac{1}{nb_n^3}(K^\circ)'\Big(\frac{j/n-(m_s')^{-1}(t)}{b_n}\Big)\sigma_s((m_s')^{-1}(t))(((m_s')^{-1})'(t))^2\int vK_d'(v)dv\Big)^2 \\
&\times w(t)dt(1+O(b_n)).
\end{aligned} \tag{5.25}$$

Using the estimate

$$\frac{1}{nb_n}\sum_{j=1}^{n}\Big((K^\circ)'\Big(\frac{j/n-(m_s')^{-1}(t)}{b_n}\Big)\Big)^2 = \int ((K^\circ)'(x))^2dx\Big(1+O\Big(\frac{1}{nb_n}\Big)\Big), \tag{5.26}$$

(uniformly with respect to $t \in [a+\eta, b-\eta]$), (5.25) and (5.26) gives

$$\mathbb{E}(D_1) = \frac{1}{nb_n^5} \sum_{s=1}^{2} \int ((K^\circ)'(x))^2 dx \int \Big(\sigma_s((m_s')^{-1}(t))(((m_s')^{-1}(t))')^2 \int vK_d'(v)dv \Big)^2 w(t)dt$$
$$\times \Big(1 + O\Big(b_n + \frac{1}{nb_n}\Big)\Big),$$

which implies

$$\mathbb{E}(nb_n^{9/2} D_1) = B_n(0) + O\Big(\sqrt{b_n} + \frac{1}{nb_n^{3/2}}\Big), \tag{5.27}$$

where $B_n(g)$ is defined in Theorem 3.2 (and we use the notation with the function $g \equiv 0$). Here we used the change of variable $(m_s')^{-1}(t) = u$, and afterwards, $((m_s')^{-1})'(t) = \frac{1}{m_s''((m_s')^{-1}(t))}$. Similar arguments establish that

$$\mathrm{Var}(D_1) = O\Big(\sum_{s=1}^{2}\sum_{j=1}^{n} (\int \tilde{G}^2(m_s'(\cdot), j, t)^2 w(t)dt)^2\Big) = O\Big(\frac{nb_n^2}{n^4 b_n^{12}}\Big) = O\Big(\frac{1}{n^3 b_n^{10}}\Big),$$

where the first estimate is obtained from the fact that $\int G^2(m_s'(\cdot), j, t)w(t)dt = O(b_n/(nb_n^3))$. This leads to the estimate

$$\mathrm{Var}(nb_n^{9/2} D_1) = O\Big(\frac{1}{nb_n}\Big). \tag{5.28}$$

For the term D_3 in the decomposition (5.24) it follows that

$$\mathbb{E}(D_3^2) = 4 \sum_{1 \le i \le n} \Big(\int \tilde{G}(m_1'(\cdot), i, t)\tilde{G}(m_2'(\cdot), i, t)w(t)dt \Big)^2$$
$$= \frac{4(\int vK_d'(v)dv)^4}{n^4 b^{12}} \sum_i \Big(\int (((m_1')^{-1})')^2(t)(((m_2')^{-1})'(t)(K^\circ)'\Big(\frac{i/n - (m_1')^{-1}(t)}{b_n}\Big)$$
$$(K^\circ)'\Big(\frac{i/n - (m_2')^{-1}(t)}{b_n}\Big)w(t)dt\Big)^2 \sigma_1^2(i/n)\sigma_2^2(i/n) = O((n^3 b_n^{11})^{-1})$$

Hence,

$$nb_n^{9/2} D_3 = O_p\Big(\Big(\frac{1}{nb_n^2}\Big)^{1/2}\Big). \tag{5.29}$$

Finally we investigate the term D_2 using a central limit theorem for quadratic forms [see de Jong (1987)]. For this purpose define the terms (note that $(K^\circ)'(\cdot)$ is symmetric and has bounded support)

$$
\begin{aligned}
V_{s,n} &= \sum_{1\le i\neq j\le n} \Big((K^\circ)'\Big(\frac{i/n-(m_s')^{-1}(t)}{b_n}\Big)(K^\circ)'\Big(\frac{j/n-(m_s')^{-1}(t)}{b_n}\Big)\sigma_s(\frac{i}{n})\sigma_s(\frac{j}{n})(((m_s')^{-1})'(t))^4 w(t)dt\Big)^2 \\
&= n^2\int_0^1\int_0^1 \Big(\int_{\mathbb{R}}(K^\circ)'\Big(\frac{u-(m_s')^{-1}(t)}{b_n}\Big)(K^\circ)'\Big(\frac{v-(m_s')^{-1}(t)}{b_n}\Big)\sigma_s(u)\sigma_s(v)(((m_s')^{-1})'(t))^4 w(t)dt\Big)^2 \\
&\times dudv(1+o(1)) \\
&= n^2 b_n^2\int_0^1\int_0^1 \Big(\int_{\mathbb{R}}(K^\circ)'(y)(K^\circ)'(\frac{v-u}{b_n}+y)\sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3}dy\Big)^2 dudv(1+o(1)) \\
&= n^2 b_n^3\int((K^\circ)'*(K^\circ)'(z))^2dz\int(\sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3})^2du(1+o(1)) \ ,
\end{aligned}
$$

then $\lim_{n\to\infty} V_{s,n}/(n^2b_n^3)$ exists ($s=1,2$) and

$$
\lim_{n\to\infty}\frac{2(\int vK_d'(v)dv)^4 n^2 b_n^9}{(nb_n^3)^4}(V_{1,n}+V_{2,n}) = V_T,
$$

where the asymptotic variance V_T is defined in Theorem 3.2. Now similar arguments as used in the proof of Lemma 4 in Zhou (2010) show that

$$
nb_n^{9/2}D_2 \Rightarrow N(0,V_T),
$$

Combining this statement with (5.23), (5.24), (5.27), (5.28), and (5.29) finally gives

$$
nb_n^{9/2}\int U_n^2(t)w(t)dt - B_n(0) \Rightarrow N(0,V_T). \tag{5.30}
$$

Asymptotic properties of (5.16)**:** Note that

$$
\int U_n(t)((m_1^{-1})'-(m_2^{-1})')w(t)dt = \int(U_{n,1}(t)-U_{n,2}(t))((m_1^{-1})'-(m_2^{-1})')w(t)dt,
$$

where

$$
\begin{aligned}
\int(U_{n,s}(t)((m_1^{-1})'-(m_2^{-1})')w(t)dt &= \sum_{j=1}^n V_{j,s}\int G(m_s'(\cdot),j,t)(\rho_n g(t)+o(\rho_n))w(t)dt \\
&= O_p\Big(\Big(\frac{nb_n^2\rho_n^2}{n^2b_n^6}\Big)^{1/2}\Big) = O_p\Big(\frac{\rho_n}{(nb_n^4)^{1/2}}\Big).
\end{aligned}
$$

Observing that

$$\int G(m_s'(\cdot),j,t)\rho_n g(t)w(t)dt = O(\rho_n b_n/(nb_n^3)),$$

the bandwidth conditions and the definition of ρ_n give for $s=1,2$,

$$nb_n^{9/2}\int (U_{n,s}(t)((m_1^{-1})' - (m_2^{-1})')w(t)dt = O_p(b_n^{1/4}). \tag{5.31}$$

Asymptotic properties of (5.17)**:** Note that it follows for the term (5.17)

$$\Big|\int U_{n,s}(t)R_n^\dagger(t)w(t)dt\Big| \le \sup_t |R_n^\dagger(t)| \int \sup_t \Big|\sum_{j=1}^n V_{j,s}G(m_s'(\cdot),j,t)\Big| w(t)dt.$$

Observing that $\sum_j G^2(m_s'(\cdot),j,t) = O(nb_n/(nb_n^3)^2)$ we have

$$\sup_t \Big|\sum_{j=1}^n V_{j,s}G(m_s'(\cdot),j,t)\Big| = O_p\Big(\frac{\log^{1/2} n}{n^{1/2}b_n^{5/2}}\Big),$$

and the conditions on the bandwidths and (5.14) yield

$$\begin{aligned} nb_n^{9/2}&\Big|\int (U_{n,s}(t)(R_n^\dagger(t))w(t)dt\Big| \\ &= O_p\Big(\frac{\log^{1/2} n}{n^{1/2}b_n^{5/2}}\Big(\frac{\pi_n'}{h_d} + \frac{\pi_n^3}{h_d^2} + h_d + \frac{1}{Nh_d}\Big)nb_n^{9/2}\Big) = o_p(1). \end{aligned} \tag{5.32}$$

The proof of assertion (5.6) is now completed using the decomposition (5.12) and the results (5.18), (5.19), (5.20), (5.30), (5.31) and (5.32).

5.3.2 Proof of (5.7)

From the proof of (5.6) we have the decomposition

$$\begin{aligned} T_n - \tilde T_n &= \int (I_1(t) - I_2(t) + II(t))^2(\hat w(t) - w(t))dt \\ &= \int \big(U_n(t) + ((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' + R_n^\dagger(t)\big)^2 (\hat w(t) - w(t))dt, \end{aligned}$$

where quantities I_s, II, $U_n(t)$ and $R_n^\dagger(t)$ are defined in (5.8), (5.9), and (5.13). By the proof of (5.6), it then suffices to show that

$$nb_n^{9/2}\int(U_n(t))^2(\hat{w}(t)-w(t))dt = o_p(1).$$

Using the same arguments as given in the proof of (5.6), this assertion follows from

$$nb_n^{9/2}\int(\tilde{U}_n(t))^2(\hat{w}(t)-w(t))dt = o_p(1).$$

where $\tilde{U}_n(t)$ is defined in (5.22). Recalling the definition of a, b in (2.16) it then follows (using similar arguments as given for the derivation of (5.5)) that

$$\sup_{t\in[a,b]}|\tilde{U}_n(t)| = O_p\left(\frac{\log n}{\sqrt{nb_n}b_n^2}\right).$$

Furthermore, together with part (iii) of Proposition 5.2 it follows that

$$\int(\tilde{U}_n(t))^2(\hat{w}(t)-w(t))dt \le \sup_{t\in[a,b]}|\tilde{U}_n(t)|)^2\int|\hat{w}(t)-w(t)|dt = O_p\left(\frac{\bar{\omega}_n\log^2 n}{nb_n^5}\right),$$

where $\bar{\omega}_n$ is defined in (3.2). Thus by our choices of bandwidth $nb_n^{9/2}\frac{\bar{\omega}_n\log^2 n}{nb_n^5} = o(1)$, from which result (ii) follows.

Finally, the assertion of the Theorem 3.2 follows from (5.6) and (5.7). □

Acknowledgments. This work has been supported in part by the Collaborative Research Center "Statistical modelling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The authors are grateful to Martina Stein, who typed parts of this paper with considerable technical expertise.

References

Ait-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116:9–47.

Carroll, R. J. and Hall, P. (1993). Semiparametric comparison of regression curves via normal likelihoods. *Australian Journal of Statistics*, 34(3):471–487.

Collier, O. and Dalalyan, A. S. (2015). Curve registration by nonparametric goodness-of-fit testing. *Journal of Statistical Planning and Inference*, 162:20–42.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25(1):1–37.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75(2):261–277.

Degras, D., Xu, Z., Zhang, T., and Wu, W. B. (2012). Testing for parallelism among trends in multiple time series. *IEEE Transactions on Signal Processing*, 60(3):1087–1097.

Dette, H. and Munk, A. (1998). Nonparamatric comparison of several regression curves: exact and asymptotic theory. *Annals of Statistics*, 26(6):2339–2368.

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Annals of Statistics*, 29(5):1361–1400.

Dette, H., Neumeyer, N., and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12:469–490.

Dette, H. and Wu, W. (2019+). Detecting relevant changes in the mean of non-stationary processes - a mass excess approach. *The Annals of Statistics, to appear.*

Durot, C., Groeneboom, P., and Lopuhaä, H. P. (2013). Testing equality of functions under monotonicity constraints. *Journal of Nonparametric Statistics*, 25(4):939–970.

Gamboa, F., Loubes, J., and Maza, E. (2007). Semi-parametric estimation of shifts. *Electronic Journal of Statistics*, 1:616–640.

Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, 85(412):1039–1049.

Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, 29(3):624–647.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, 18(1):63–89.

Maity, A. (2012). A powerful test for comparing multiple regression functions. *Journal of Nonparametric Statistics*, 24(3):563–576.

Mallat, S., Papanicolaou, G., and Zhang, Z. (1998). Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics*, 26(1):1–47.

Mammen, E. (1991). Estimating a smooth monotone regression function. *Annals of Statistics*, 19(2):724–740.

Matzkin, R. L. (1991). Semiparametric estimation of monotone and concave utility functionsfor polychotomous choice models. *Econometrica*, 59(5):1315–1327.

Nason, G. P., Von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):271–292.

Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves - an empirical process approach. *Annals of Statistics*, 31(3):880–920.

Neumeyer, N. and Pardo-Fernández, J. C. (2009). A simple test for comparing regression curves versus one-sided alternatives. *Journal of Statistical Planning and Inference*, 139(12):4006–4016.

Ombao, H., Von Sachs, R., and Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531.

Park, C., Hannig, J., and Kang, K.-H. (2014). Nonparametric comparison of multiple regression curves in scale-space. *Journal of Computational and Graphical Statistics*, 23(3):657–677.

Rønn, B. (2001). Nonparametric maximum likelihood estimation for shifted curves. *Journal of the Royal Statistical Society, Ser. B*, 63(2):243–259.

Silverman, B. W. (1998). CHAPMAN & HALL, London.

Varian, H. R. (1984). The nonparametric approach to production analysis. *Econometrica*, 52(3):579–597.

Vilar-Fernández, J. M., Vilar-Fernández, J. A., and González-Manteiga, W. (2007). Bootstrap tests for nonparametric comparison of regression curves with dependent errors. *TEST*, 16(1):123–144.

Vimond, M. (2010). Efficient estimation for a subclass of shape invariant models. *Annals of Statistics*, 38(3):1885–1912.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633.

Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica*, 21(3):1397–1413.

Zhou, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics*, 38(4):2187–2217.

Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics*, 37(5B):2696–2729.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):513–531.