Annals of the Institute of Statistical Mathematics (2021) 73:855–889
https://doi.org/10.1007/s10463-020-00771-2

Identifying shifts between two regression curves

Holger Dette[1] · Subhra Sankar Dhar[2] · Weichi Wu[3]

Received: 16 January 2020 / Revised: 18 September 2020
Published online: 6 January 2021

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 two real data examples.

Keywords Comparison of curves · Nonparametric regression · Hypothesis testing

✉ Weichi Wu
wuweichi@mail.tsinghua.edu.cn

Holger Dette
holger.dette@rub.de

Subhra Sankar Dhar
subhra@iitk.ac.in

[1] Fakultät für Mathematik, Ruhr-Universität Bochum, Building IB 2/65, 44780 Bochum, Germany

[2] Department of Mathematics and Statistics, IIT Kanpur, FB 512 A (Faculty Building), Kanpur 208016, India

[3] Department of Industrial Engineering and Center for Statistical Science, Tsinghua University, 114 Weiqing Building, Beijing 100084, China

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

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

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

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 these curves are realizations of one curve but differ in the time and the vertical axes, and consequently, the difference among these 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

$$H_0 : m_1(x) = m_2(x + c) + d \ \text{ for some constants } c, d, \tag{3}$$

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 Sect. 2 that under the null hypothesis (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 (3) 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 a 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 (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 $((m_1')^{-1})' - ((m_2')^{-1})'$ 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) and (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 nonparametric regression functions up to a shift in the covariate and response in Sect. 2.1. The formal testing problem is considered in Sect. 2.2, while we give some theoretical justification for these tools in Sect. 3. A small simulation study is carried out in Sect. 4, illustrating the finite sample properties of the proposed method and two applications are discussed in Sect. 4.3. Finally, all proofs except of the proof of Lemma 1, which justifies our approach, are given in an appendix in Sect. 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

$$Y_{i,s} = m_s\left(\frac{i}{n_s}\right) + e_{i,s},\ i = 1, \dots, n_s,\ s = 1, 2, \tag{4}$$

the error random variables $\{e_{i,1}\}_{i=1,\dots,n_1}$ and $\{e_{i,2}\}_{i=1,\dots,n_2}$ are locally stationary process satisfying some technical conditions that will be described later in Sect. 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 in a hypothesis

$$H_0 : \begin{cases} \text{there exists constants } c \in (0,1) \text{ and } d \in \mathbb{R} \text{ such that} \\ m_1(t) = m_2(t+c) + d, \text{ for all } t \in (0, 1-c). \end{cases} \tag{5}$$

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 horizontal 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 (5) we have

$$((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' = 0, \tag{6}$$

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

Lemma 1 *Assume that the regression functions m_1 and m_2 in (4) 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 (6) 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 $(0, 1-c)$

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

In particular, this yields (subtracting both equations)

$$c = (m_2')^{-1}(u) - (m_1')^{-1}(u) \tag{7}$$

for any $u \in (m_1'(0), m_1'(1-c))$. Taking derivatives on both sides of (7) gives (6) 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 1. □

2.1 Graphical device

According to Lemma 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 horizontal shift c. We now consider a collection of points

$$\mathcal{C}_{n_1,n_2} = \{(t_\ell, \hat{f}_1(t_\ell) - \hat{f}_2(t_\ell)) : t_\ell \in (\hat{a} + \eta, \hat{b} - \eta);\ \ell = 1, \dots, L\}, \tag{8}$$

where $\hat{a} = \widehat{m_1'}(0)$ and $\hat{b} = \widehat{m_1'}(1-\hat{c})$ 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 for $s = 1, 2$ the estimate of the regression function m_s and its derivative m_s' for $t \in [b_{n,s}, 1 - b_{n,s}]$ by

$$(\hat{m}_s(t), b_{n,s}\hat{m}'_s(t))^\top = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n_s} \Big(Y_{i,s} - \beta_0 - \beta_1\Big(\frac{i}{n_s} - t\Big)\Big)^2 K\Big(\frac{i/n_s - t}{b_{n,s}}\Big), \quad (9)$$

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

$$\hat{f}_1(t) = \frac{1}{Nh_{d,1}} \sum_{i=1}^{N} K_d\Big(\frac{\hat{m}'_1(\frac{i}{N}) - t}{h_{d,1}}\Big), \quad (10)$$

$$\hat{f}_2(t) = \frac{1}{Nh_{d,2}} \sum_{i=1}^{N} K_d\Big(\frac{\hat{m}'_2(\frac{i}{N}) - t}{h_{d,2}}\Big) \quad (11)$$

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 (7) we obtain an estimate

$$\hat{c} = \frac{1}{1 - \tilde{c}} \int_0^{(1-\tilde{c})} (\hat{g}_2(\hat{m}'_1(u)) - u)du \quad (12)$$

of the vertical shift c. Here, $\hat{m}'_1$ is the estimate of the derivative of m_1 defined in (9) 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

$$\hat{a} = \hat{m}_1'(0), \ \hat{b} = \hat{m}_1'(1 - \hat{c}) \tag{13}$$

(note that we assume that $c > 0$). We will prove in Theorem 4 below that under the null hypothesis (5) 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 We consider the regression models (4) 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 Sect. 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:

$$m_1(x) = (x - 0.4)^2 \text{ and } m_2(x) = (x - 0.3)^2 - 0.2, \tag{14}$$

$$m_1(x) = (x - 0.4)^2 \text{ and } m_2(x) = x^3, \tag{15}$$

$$m_1(x) = \sin(-\pi x) \text{ and } m_2(x) = \sin(-\pi(x + 0.1)) + \frac{1}{4}, \tag{16}$$

$$m_1(x) = \sin(-\pi x) \text{ and } m_2(x) = -\cos(\pi x). \tag{17}$$

Note that examples (14) and (16) correspond to the null hypothesis, while (15) and (17) represent alternatives. The corresponding plots of the set $\mathcal{C}_{n_1,n_2}$ are shown in Fig. 1, where 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 (5) 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 (5), 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 (5) at a controlled type I error. Recalling the definition of the estimates (10) and (11) of $((m_1')^{-1})'(t)$ and $((m_2')^{-1})'(t)$, we propose to reject the null hypothesis (5) for large values of the statistic

$$T_{n_1,n_2} = \int \left(\hat{f}_1(t) - \hat{f}_2(t)\right)^2 \hat{w}(t)dt, \tag{18}$$

where the weight function is defined by

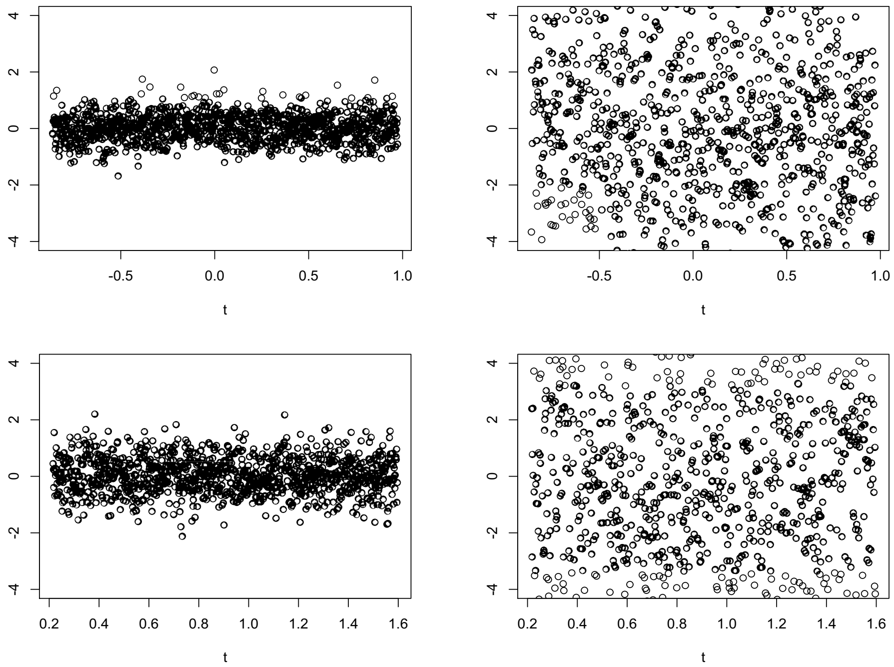

Fig. 1 Plots of the set $\mathcal{C}_{n_1,n_2}$ for different examples. The panels on the left correspond to the models (14) and (16) (null hypothesis) and the panels on the right correspond to the models (15) and (17) (alternative)

$$\hat{w}(t) = \mathbf{1}(\hat{a} + \eta \leq t \leq \hat{b} - \eta), \tag{19}$$

η is a small positive constant and $\hat{a}$ and $\hat{b}$ are defined in (13). In fact, $\hat{w}(t)$ is a consistent estimator of the deterministic weight function

$$w(t) = \mathbf{1}(a + \eta \leq t \leq b - \eta), \tag{20}$$

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

Remark 3 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 Sect. 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 (4). 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 (4). Note that $\{e_{i,1}\}_{i\in\mathbb{N}}$ and $\{e_{i,2}\}_{i\in\mathbb{N}}$ define two triangular arrays although this is not reflected in our notation. In particular, we assume $\{e_{i,s}\}_{i\in\mathbb{N}}$, $s=1,2$ are locally stationary processes in the sense of Zhou and Wu (2009) such that they have the form

$$e_{i,1}=G_1(i/n_1,\mathcal{F}_i),\ 1\le i\le n_1,\ e_{i,2}=G_2(i/n_2,\mathcal{G}_i),\ 1\le i\le n_2. \tag{21}$$

In (21) 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 1

$$\mathbb{E}(G_1(t,\mathcal{F}_0))=0 \quad \text{for } t\in[0,1], \text{ and } \sup_{t\in[0,1]}\|G_1(t,\mathcal{F}_0)\|_4<\infty.$$
$$\mathbb{E}(G_2(t,\mathcal{G}_0))=0 \quad \text{for } t\in[0,1], \text{ and } \sup_{t\in[0,1]}\|G_2(t,\mathcal{G}_0)\|_4<\infty.$$

(a)

$$\sup_{0\le t_1<t_2\le 1}\{\|G_1(t_1,\mathcal{F}_0)-G_1(t_2,\mathcal{F}_0)\|_2/|t_1-t_2|\}<\infty,$$
$$\sup_{0\le t_1<t_2\le 1}\{\|G_2(t_1,\mathcal{G}_0)-G_2(t_2,\mathcal{G}_0)\|_2/|t_1-t_2|\}<\infty.$$

(b)

(c) Let $\{\varepsilon^*_{i,1}\}_{i\in\mathbb{N}}$ denote an independent copy of $\{\varepsilon_{i,1}\}_{i\in\mathbb{N}}$ and define the filtration $\mathcal{F}^*_i=(\varepsilon_{-\infty,1},...,\varepsilon_{-1,1},\varepsilon^*_{0,1},...,\varepsilon_{i,1})$. Similarly, let $\{\varepsilon^*_{i,2}\}_{i\in\mathbb{N}}$ denote an independent copy of $\{\varepsilon_{i,2}\}_{i\in\mathbb{N}}$ and define the filtration $\mathcal{G}^*_i=(\varepsilon_{-\infty,2},...,\varepsilon_{-1,2},\varepsilon^*_{0,2},...,\varepsilon_{i,2})$. There exists a constant $\rho\in(0,1)$ such that for any $k\ge 0$, $\delta_4(k)=O(\rho^k)$, where $\delta_4(k)=\max(\delta_{4,1}(k),\delta_{4,2}(k))$ and

$$\delta_{4,1}(k):=\sup_{t\in[0,1]}\|G_1(t,\mathcal{F}_k)-G_1(t,\mathcal{F}^*_k)\|_4,$$
$$\delta_{4,2}(k):=\sup_{t\in[0,1]}\|G_2(t,\mathcal{G}_k)-G_2(t,\mathcal{G}^*_k)\|_4\,.$$

(d) There exists a constant $v_0>0$ such that the 2×2 matrix $\Sigma^2(t)-v_0I_2$ is strictly positive definite for any $t\in[0,1]$, where I_2 is the 2×2 identity matrix, and $\Sigma^2(t)$ is defined as

$$\Sigma^2(t) = \sum_{s=0}^{\infty} \mathbb{E}\big((G_1(t,\mathcal{F}_0), G_2(t,\mathcal{G}_0))(G_1(t,\mathcal{F}_0), G_2(t,\mathcal{G}_0))^\top\big).$$

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

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 (9), we make the following assumptions.

Assumption 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

(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 4 For $s = 1, 2$, let

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

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

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

where

$$\bar{\omega}_{n,s} = \frac{\log n}{\sqrt{nb_{n,s}}b_{n,s}} + \frac{n^{1/4}\log^2 n}{nb_{n,s}^2} + b_{n,s},\ \ s = 1, 2. \tag{22}$$

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 (8) if it is used with the local linear estimates (9) for the derivatives m_1' and m_2'. It basically gives a theoretical justification for the use of the graphical device proposed in Sect. 2.1. The proof can be found in Sect. 5.2.

Theorem 4 *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 1–4 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 4 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 (5) 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

$$K^\circ(x) = \frac{K(x)x}{\int_{-1}^{1} K(x)x^2 dx}, \tag{23}$$

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

Theorem 5 *Suppose that Assumption 1–4 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* $n_1, n_2 \to \infty$,

$$n_1 b_{n,1}^{9/2} T_{n_1,n_2} - B_n(g) \Rightarrow \mathcal{N}(0, V_T), \tag{24}$$

where the asymptotic bias and variance are given by

$$\begin{aligned} B_n(g) =& \frac{(\int_{-1}^{1} vK_d'(v)dv)^2}{\sqrt{b_{n,1}}} ((K^\circ)' * (K^\circ)'(0)) \\ & \times \sum_{s=1}^{2} c_s r_s^5 \int_{\mathbb{R}} \sigma_s^2(u) w(m_s'(u))(m_s''(u))^{-3} du + \int_0^1 g^2(t) w(t) dt, \\ V_T =& 2\Big(\int_{-1}^{1} vK_d'(v)dv \Big)^4 \\ & \times \sum_{s=1}^{2} c_s^2 r_s^9 \int_{\mathbb{R}} ((K^\circ)' * (K^\circ)'(z))^2 dz \int_{\mathbb{R}} (\sigma_s^2(u) w(m_s'(u))(m_s''(u))^{-3})^2 du \end{aligned}$$

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

Remark 6 Under the null hypothesis, we have $g \equiv 0$ and Theorem 5 can be used to construct a consistent asymptotic level α test for the hypotheses in (5). 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 of $N(0, 1)$, 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 5 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 5 directly leads to the following corollary.

Corollary 7 *If the assumptions of Theorem 5 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 (24) holds with*

$$B_n(g) = \frac{(\int vK_d'(v)dv)^2}{\sqrt{b_n}}((K^\circ)' * (K^\circ)')(0)$$
$$\times \sum_{s=1}^{2} \int_{\mathbb{R}} \sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3}du - \int_0^1 g^2(t)w(t)dt$$
$$V_T = 2\Big(\int_{-1}^{1} vK_d'(v)dv\Big)^4$$
$$\times \sum_{s=1}^{2} \int_{\mathbb{R}} ((K^\circ)' * (K^\circ)'(z))^2dz \int_{\mathbb{R}} (\sigma_s^2(u)w(m_s'(u))(m_s''(u))^{-3})^2du.$$

Remark 8 (a) (Kneip and Engel 1995) studied the estimation of nonlinear regression functions under shape invariance. They considered N regression models $Y_{ij} = f_i(t_{ij}) + \varepsilon_{ij}$, $j = 1, \dots, n_i$, $i = 1, \dots, N$ and a shape invariance property of the form $f_i(\theta_{i2}t + \theta_{i3}) = \theta_{i1}\phi(t) + \theta_{i4}$ for $\tilde{N}$ of these models. Utilizing this property, (Kneip and Engel 1995) proposed an estimator of ϕ which is more efficient than the usual nonparametric estimator. The situation considered in the present paper is a special case of the model investigated by these authors, but has its focus on testing. In contrast, (Kneip and Engel 1995) considered the estimation of the function ϕ, for which they derived an improvement by a factor of $\tilde{N}^{-\frac{2k}{2k+1}}$ where k relates to the smoothness of the regression function. However, they did not study related testing problems. Note that in our setting $\tilde{N} \equiv 2$, and therefore, the improvement is limited. It is practically useful to consider also testing for a scaling in addition to a shift, which was also not addressed in Kneip and Engel (1995). We shall leave this problem for future research.

(b) If the two series $\{e_{i,1}\}_{i\in\mathbb{Z}}$ and $\{e_{i,2}\}_{i\in\mathbb{Z}}$ are sequences of independent random variables, then Corollary 7 can be simplified replacing long-run variances $\sigma_1^2(t)$ and $\sigma_2^2(t)$ by the (local) variances $\tilde{\sigma}_1^2(t) = \mathbb{E}(G_1^2(t, \mathcal{F}_0))$ and $\tilde{\sigma}_2^2(t) = \mathbb{E}(G_2^2(t, \mathcal{G}_0))$, respectively. Since Assumption 4 is postulated under short range dependence and refers to the magnitude of the order of bandwidths, it remains unchanged in this case.

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

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

As pointed out by Dette et al. (2006), the choice of $h_{d,s}$ has a negligible impact on the estimators (10) and (11) (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 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 $t \in [m/n_s, 1 - m/n_s]$

$$\hat{\sigma}_s^2(t) = \sum_{j=1}^{n} \frac{m\Delta_{j,s}^2}{2} \omega(t,j), \quad s = 1, 2, \tag{25}$$

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 5 is interesting from a theoretical point of view, it cannot be easily implemented for testing the hypothesis (5). 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 5 is slow (note that the bias in Theorem 5 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 5 in Sect. 5. There we show (see equations (44) and (45)) 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 random variables.

Algorithm 1 (a) Estimate m_1' and m_2' by (9) and estimate the long-run variances σ_1^2 and σ_2^2 by (25).

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

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

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 $W_{(1)} \le W_{(2)} \le \ldots \le W_{(B)}$ be the ordered statistics of $\{W_s, 1 \le s \le B\}$. We reject the null hypothesis (5) at level α, whenever

$$T_{n_1,n_2} > W_{(\lfloor B(1-\alpha) \rfloor)}. \tag{26}$$

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

Remark 9 Algorithm 1 can be extended to comparison of more than two regression curves. For the sake of a simple notation, we consider the case of three samples and

Table 1 The estimated size of the test (26) and the CD test for different sample sizes $n_1 = n_2 = n$ over 1000 runs and $B = 1000$ Bootstrap replications

Model/n	50	100	200	500
(29)	0.054 (0.055)	0.053 (0.053)	0.053 (0.052)	0.050 (0.051)
(30)	0.055 (0.056)	0.054 (0.053)	0.052 (0.053)	0.051 (0.051)
(29)	0.108 (0.110)	0.105 (0.107)	0.104 (0.105)	0.100 (0.101)
(30)	0.111 (0.113)	0.107 (0.108)	0.105 (0.104)	0.102 (0.102)

The level of significance is 5% (upper part) and 10% (lower part). Inside each cell, from the left, the first value corresponds to the test (26), and the second value within parentheses corresponds to the CD test

Table 2 The estimated size of the test (26) and the CD test for unequal sample sizes over 1000 runs and $B = 1000$ Bootstrap replications

(n_1, n_2)	(50, 75)	(100, 150)	(200, 300)	(500, 750)
(29)	0.053 (0.054)	0.052 (0.053)	0.053 (0.052)	0.051 (0.051)
(30)	0.054 (0.053)	0.052 (0.052)	0.052 (0.051)	0.050 (0.051)
(29)	0.109 (0.110)	0.104 (0.106)	0.103 (0.105)	0.101 (0.102)
(30)	0.110 (0.111)	0.105 (0.104)	0.104 (0.103)	0.101 (0.101)

The level of significance is 5% (upper part) and 10% (lower part). Inside each cell, from the left, the first value corresponds to the test (26), and the second value within parentheses corresponds to the CD test

assume that we additionally observe $Y_{i,3} = m_3(\frac{i}{n_3}) + e_{i,3}$, $1 \le i \le n_3$ where $(e_{i,3})_{1\le i\le n_3}$ is a locally stationary process with non-degenerate long-run variance $\sigma_3(\cdot)$ such that $(e_{i,1})_{1\le i\le n_1}$, $(e_{i,2})_{1\le i\le n_2}$ and $(e_{i,3})_{1\le i\le n_3}$ are independent. Consider the hypothesis that

$$H_0 : \begin{cases} \text{there exists constants } c_1, c_2 \in (0,1) \text{ and } d_1, d_2 \in \mathbb{R} \text{ such that} \\ m_1(t) = m_2(t + c_1) + d_1, \text{ for all } t \in (0, 1 - c_1) \\ m_3(t) = m_2(t + c_2) + d_2, \text{ for all } t \in (0, 1 - c_2). \end{cases} \tag{27}$$

Recall the definition of T_{n_1,n_2} in (18). Define T_{n_2,n_3} by replacing $\hat{f}_1(t)$ there with $\hat{f}_3(t)$, where $\hat{f}_3(t)$ is defined via the local linear estimate of $m_3'(t)$ in a way similar to $\hat{f}_1(t)$. We define the test statistic by

$$T_{n_1,n_2,n_3} = \max(T_{n_1,n_2}, T_{n_2,n_3})$$

and propose the following bootstrap test for the null hypothesis (27):

Algorithm 2 (a) Generate B copies of standard normally distributed random variables $\{V_{j,1}^{(B)}\}_{j=1}^{n_1}$, $\{V_{j,2}^{(B)}\}_{j=1}^{n_2}$, $\{V_{j,3}^{(B)}\}_{j=1}^{n_3}$ and calculate the statistic

$$\tilde{W}_B = \max(W_B, W_B'),$$

where

$$W_B' = \int_{\mathbb{R}} \Big(\frac{1}{nNb_{n,3}^2 h_{d,3}^2} \Xi_3^{(B)}(t) - \frac{1}{nNb_{n,2}^2 h_{d,2}^2} \Xi_2^{(B)}(t) \Big)^2 \hat{w}_{2,3}(t) dt,$$

and W_B and $\Xi_2^{(B)}(t)$ are defined in Algorithm 1, $\hat{w}_{2,3}(t)$ is defined in analogue to (19) with indices 1, 2 in the formula replaced by 2, 3, and

$$\Xi_3^{(B)}(t) = \sum_{j=1}^{n_3} \sum_{i=1}^{N} K^\circ \Big(\frac{j/n_3 - i/N}{b_{n,3}} \Big) K_d' \Big(\frac{\hat{m}_3'(i/N) - t}{h_{d,3}} \Big) \hat{\sigma}_3 \Big(\frac{j}{n_3} \Big) V_{j,3}^{(B)}.$$

The parameters $b_{n,3}$ and $h_{d,3}$ are used for the nonparametric estimate $\hat{f}_3(t)$ and can be selected in a similar way to $b_{n,v}$ and $h_{d,v}$, $v = 1, 2$.

(b) Let $\tilde{W}_{(1)} \leq \tilde{W}_{(2)} \leq \ldots \leq \tilde{W}_{(B)}$ be the ordered statistics of $\tilde{W}_1, \ldots, \tilde{W}_B$, then the null hypothesis (27) is rejected (at level α), whenever

$$T_{n_1,n_2,n_3} > \tilde{W}_{(\lfloor B(1-\alpha) \rfloor)}.$$

The p value of this test is given by $1 - B^*/B$ where $B^* = \max\{r : \tilde{W}_{(r)} \leq T_{n_1,n_2,n_3}\}$.

4.2 Simulated level and power

In this section, we illustrate the finite sample properties of the test (26) by means of a small simulation study. All presented results are based on 1000 runs and $B = 1000$ bootstrap replications. We consider both scenarios of equal and unequal sample sizes $n_1 = n_2/a$, $a = 1, 1.5$ with $n_1 = 50$, 100, 200 and 500. Throughout this article, the Epanechnikov kernel is considered for all kernels appearing in the test procedure, and we use $N = n$ in (10) and (11). Besides, $h_{d,N} = n^{-1/3}$, and b_{n_1} and b_{n_2} are chosen as described at the beginning of Sect. 4.

For $s = 1$ and 2, we consider model (4) with the error process

$$G_s(t, \mathcal{F}_i) = 0.6(t - 0.3)^2 G_s(t, \mathcal{F}_{i-1,s}) + \eta_{i,s}, \tag{28}$$

where $\mathcal{F}_{i,s} = (\ldots, \eta_{i-1,s}, \eta_{i,s})$. 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. A similar dependence structure has been considered in Dette and Wu (2019), and other locally stationary processes yield similar results. In order to investigate the size of the test (26) and that of the existing tests, we consider the models

$$m_1(x) = (x - 0.4)^2 \text{ and } m_2(x) = (x - 0.3)^2 - 0.2, \tag{29}$$

$$m_1(x) = \sin(-\pi x) \text{ and } m_2(x) = \sin(-\pi(x + 0.1)) + \frac{1}{4}. \tag{30}$$

In Tables 1 and 2, we display the rejection probabilities of the test (26) and the test studied in Collier and Dalalyan (2015) with projection weight (denoted as CD test), where the level of significance is 5% and 10%. The results show a good approximation of the nominal level in all cases under consideration.

In order to study the power of the test (26) and that of the existing tests, we consider the same error processes as in (28) and use the regression functions

$$m_1(x) = (x - 0.4)^2 \text{ and } m_2(x) = x^3, \tag{31}$$

$$m_1(x) = \sin(-\pi x) \text{ and } m_2(x) = -\cos(\pi x). \tag{32}$$

Table 3 The estimated power of the test (26) and the CD test for different sample sizes $n_1 = n_2 = n$ over 1000 runs and $B = 1000$ Bootstrap replications

Model/n	50	100	200	500
(31)	0.611 (0.476)	0.689 (0.541)	0.801 (0.622)	0.888 (0.679)
(32)	0.663 (0.552)	0.712 (0.583)	0.820 (0.657)	0.901 (0.737)
(31)	0.752 (0.593)	0.801 (0.632)	0.896 (0.707)	0.952 (0.754)
(32)	0.790 (0.698)	0.834 (0.736)	0.922 (0.813)	0.994 (0.880)

The level of significance is 5% (upper part) and 10% (lower part). Inside each cell, from the left, the first value corresponds to the test (26), and the second value within parentheses corresponds to the CD test

Table 4 The estimated power of the test (26) and the CD test for unequal sample sizes over 1000 runs and $B = 1000$ Bootstrap replications

Model/(n_1, n_2)	(50, 75)	(100, 150)	(200, 300)	(500, 750)
(31)	0.628 (0.503)	0.703 (0.565)	0.822 (0.663)	0.901 (0.730)
(32)	0.679 (0.567)	0.736 (0.598)	0.842 (0.681)	0.928 (0.759)
(31)	0.783 (0.602)	0.832 (0.657)	0.926 (0.729)	0.977 (0.772)
(32)	0.809 (0.705)	0.850 (0.747)	0.957 (0.838)	0.998 (0.801)

The level of significance is 5% (upper part) and 10% (lower part). Inside each cell, from the left, the first value corresponds to the test (26), and the second value within parentheses corresponds to the CD test

As mentioned in Sect. 1, there are many articles on similar problems, but most of them studied the estimation problem of the transformation. To the best of our knowledge, only (Härdle and Marron 1990) and (Collier and Dalalyan 2015) investigated a similar testing problem as considered in this paper. As the code for the test considered in Collier and Dalalyan (2015) denoted as the CD test is publicly available (see https://code.google.com/p/shifted-curve-testing/), we have compared the estimated size and the estimated power of the test proposed in this paper with that of the CD test. The values reported in Tables 1 and 2 indicate that the test (26) and the CD test approximate the nominal level reasonably well. The results in Tables 3 and 4 show the rejection probabilities under the alternative and demonstrate that in the examples under consideration the test (26) is more powerful than the CD test.

It is also of interest to study the impact of various choices of the bandwidths $b_{n,1}$ and $b_{n,2}$ on the estimated power and size if we do not follow the procedure of choosing $b_{n,1}$ and $b_{n,2}$ as described at the beginning of this section. For this purpose, we investigate the following choices:

(i) $b_{n,1} = n_1^{-1/3}$ and $b_{n,2} = n_2^{-1/3}$,
(ii) $b_{n,1} = n_1^{-1/5}$ and $b_{n,2} = n_2^{-1/5}$,
(iii) $b_{n,1} = n_1^{-1/7}$ and $b_{n,2} = n_2^{-1/7}$.

Exemplary we display in Table 5, the simulated power for these bandwidths in the scenario considered in Table 3 for a 5% level of significance. We observe that the estimated size (these results are not displayed for the sake of brevity) and the

Table 5 The estimated power of the test (26) for different sample sizes $n_1 = n_2 = n$ over 1000 runs and $B = 1000$ Bootstrap replications

Model	b_n	$n = 50$	$n = 100$	$n = 200$	$n = 500$
(31)	$n^{-1/3}$	0.622	0.677	0.814	0.878
(31)	$n^{-1/5}$	0.634	0.685	0.826	0.899
(31)	$n^{-1/7}$	0.627	0.681	0.801	0.891
(32)	$n^{-1/3}$	0.685	0.733	0.826	0.895
(32)	$n^{-1/5}$	0.692	0.755	0.840	0.919
(32)	$n^{-1/7}$	0.676	0.739	0.818	0.902

The level of significance is 5%. The upper part of the table shows the estimated power for model (31), and the lower part of the table shows the estimated power for model (32)

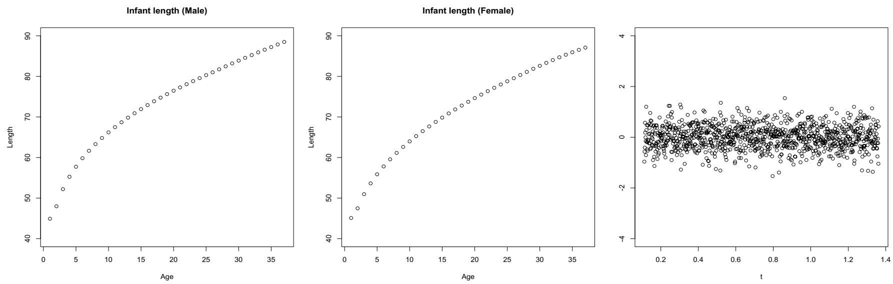

Fig. 2 Plots of the length of the male (middle part) and female (left part) infants and plot of the set $\mathcal{C}_{n_1,n_2}$ (right part)

estimated power is not varying more than 3%. In all other examples, we observed a similar behaviour indicating some robustness of the test with respect to the choice of the bandwidths.

4.3 Real data analysis

Growth Data of Male and Female Infants We use the test (26) and the graphical device to investigate the validity of assertion (5) 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 3 years (here $n_1 = n_2 = 37$). The data are depicted in Fig. 2 and indicate that the unknown regression functions associated with male and female may be differ by a shift in the horizontal and/or vertical axis. Therefore, we model the negative values of this data by two regression models of the form (4) 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 (12) and $\hat{d} = \hat{m}_1(0) - \hat{m}_2(\hat{c}) = 0.087$ as estimate of the vertical shift d.

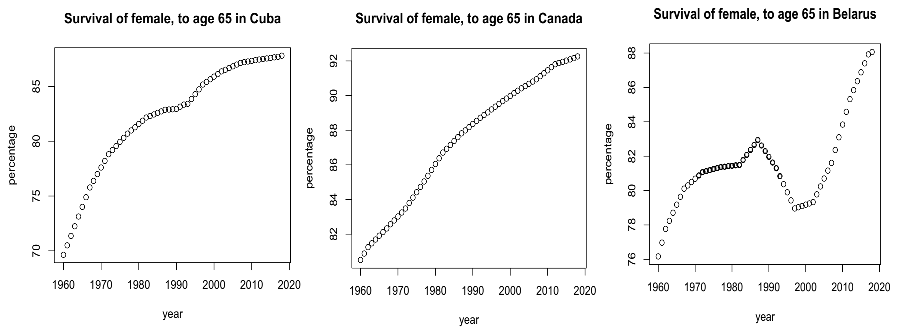

Fig. 3 Survival percentage of female to age 65 in Cuba, Canada and Belarus

In the right part of Fig. 2, we plot the points of the set $\mathcal{C}_{n_1,n_2}$ defined in (8) 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 Sect. 4). The figure clearly indicates the existence of a vertical and horizontal shift between the regression functions as formulated in the null hypothesis (5).

Finally, we also investigate the performance of the test (26) for this data set, where all parameters required for the bootstrap test are chosen as described in Sect. 4. For $B = 1000$ bootstrap replications, we obtain the p value 0.799, which gives no indication to reject the null hypothesis and is consistent with the conclusion made by graphical inspection.

Survival to Age 65, Female We here use the test (26) and the graphical device to investigate the validity of assertion (5) for the data related to survival of female to age sixty-five in three countries, namely Cuba, Canada and Belarus. This data set is available from https://data.worldbank.org/indicator/SP.DYN.TO65.FE.ZS?locations=CU-CA-BY and consists of the survival of female to age sixty-five in Cuba, Canada and Belarus from 1960 to 2018, i.e. we have $n_1 = n_2 = 59$ to test (5) for any two countries. The data for each country are displayed in Fig. 3, and the diagrams indicate that the regression functions associated with data of Cuba and Canada may differ by a shift in the horizontal and/or vertical axis, but such relation may not hold with the regression function associated with the data of Belarus. In this study, we implement the test (26) for three cases, namely the comparison between Cuba and Canada, the comparison between Cuba and Belarus, and the comparison between Canada and Belarus. For the data of Cuba and Canada, we obtain $\hat{c} = 0.243$ as estimate for the horizontal shift using the statistic (12) and $\hat{d} = \hat{m}_1(0) - \hat{m}_2(\hat{c}) = 2.226$ as estimate of the vertical shift. In order to validate (5) for the data of Cuba and Canada, we display in the first diagram from the left in Fig. 4 the points of the set $\mathcal{C}_{n_1,n_2}$ defined in (8) using $L = 1000$ equally spaced points in the interval $(\hat{a} + \eta, \hat{b} - \eta)$,

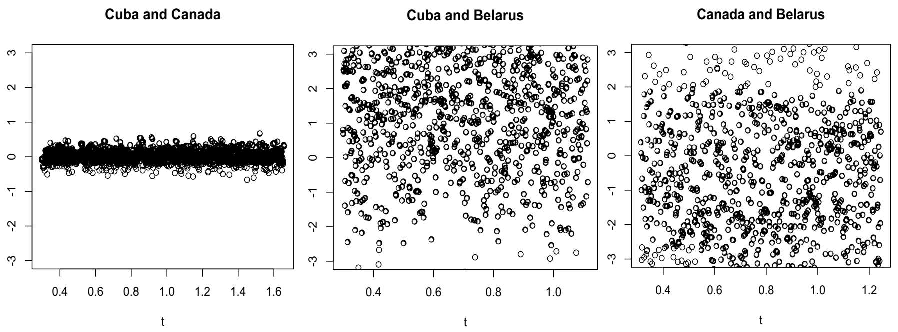

Fig. 4 Plots of $\mathcal{C}_{n_1,n_2}$ for the data related to Survival to Age 65, Female described in Sect. 4.3.

where $\hat{a} = \hat{m}_1'(0) = 0.298$, $\hat{b} = \hat{m}_1'(1 - \hat{c}) = 1.657$, and $\eta = 0.001$ (the smoothing parameters are chosen as described in Sect. 4). The figure clearly indicates the existence of a vertical and horizontal shift between the regression functions as formulated in the null hypothesis (5). Next, we also investigate the performance of the test (26) for this Cuba and Canada data set, where all parameters required for the bootstrap test are chosen as described in Sect. 4. For $B = 1000$ bootstrap replications, we obtain the p value 0.683, which favours the assertion stated in the null hypothesis and is consistent with the conclusion made by graphical inspection. For the data of Cuba and Belarus, we obtain $\hat{c} = 0.462$ as estimate for the horizontal shift using the statistic (12) and $\hat{d} = \hat{m}_1(0) - \hat{m}_2(\hat{c}) = 1.176$ as estimate of the vertical shift. The results for the graphical device described in Sect. 2.1 are displayed in the middle panel of Fig. 4, where we plot the points of the set $\mathcal{C}_{n_1,n_2}$ using $L = 1000$ equally spaced points in the interval $(\hat{a} + \eta, \hat{b} - \eta)$ with $\hat{a} = \hat{m}_1'(0) = 0.298$, $\hat{b} = \hat{m}_1'(1 - \hat{c}) = 1.114$, and $\eta = 0.001$ (the smoothing parameters are chosen as described in Sect. 4). The plot clearly indicates that there does not exist a vertical or horizontal shift between the regression functions as formulated in the null hypothesis (5). The test (26) for this data set yields the p value 0.061 and a rejection of the null hypothesis at the level 10% (all parameters required for the bootstrap test are chosen as described in Sect. 4 and $B = 1000$ bootstrap replications are used).

For the data of Canada and Belarus, we obtain $\hat{c} = 0.379$ and $\hat{d} = \hat{m}_1(0) - \hat{m}_2(\hat{c}) = 1.115$ as estimate for the horizontal and vertical shift, respectively. The points of the set $\mathcal{C}_{n_1,n_2}$ using $L = 1000$ equally spaced points in the interval $(\hat{a} + \eta, \hat{b} - \eta)$ are displayed in the right panel of Fig. 4, where $\hat{a} = \hat{m}_1'(0) = 0.314$, $\hat{b} = \hat{m}_1'(1 - \hat{c}) = 1.245$, and $\eta = 0.001$ (the smoothing parameters are chosen as described in Sect. 4). Again the result clearly indicates that there is no vertical or horizontal shift between the regression functions. The test (26) for the Canada and Belarus data set yields the p value 0.053 and a rejection of the null hypothesis at the level 10% (all parameters required for the bootstrap test are chosen as described in Sect. 4 and $B = 1000$ bootstrap replications are used).

We conclude this section with the implementation of the multiple test for the hypothesis (27) described in Remark 9, i.e. the data for Cuba, Canada and Belarus together. For $B = 1000$ bootstrap replications, we obtain the p value 0.034 yielding a rejection of the hypothesis in (27) at level 5%.

Acknowledgements The first author has been supported by the Collaborative Research Center "Statistical modelling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1, C1) of the German Research Foundation (DFG). The second author is supported by MATRICS (MTR/2019/000039), a research grant from the SERB, Government of India and in part by SFB 823. The third author is supported by NSFC Young program (No.11901337) (a research grant of China) and in part by SFB 823. The authors are grateful to the referees and the Associate Editor for their constructive comments on an earlier version of this paper and to Martina Stein, who typed parts of this paper with considerable technical expertise.

Appendix: Proofs

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 10 *Let*

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

and assume that the Assumption 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*

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

where $\Sigma(t)$ *is the square root of the long-run variance matrix* $\Sigma^2(t)$ *defined in Assumption 1.*

Proposition 11 *Let Assumption 1 and 2 be satisfied.*

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

$$\sup_{t\in[b_{n,s},1-b_{n,s}]}\Big|\hat{m}_s'(t)-m_s'(t)-\frac{1}{n_s b_{n,s}^2}\sum_{i=1}^{n_s}K^\circ\Big(\frac{i/n_s-t}{b_{n,s}}\Big)e_{i,s}\Big| = O_P\Big(\frac{1}{n_s b_{n,s}^2}+b_{n,s}^2\Big) \tag{33}$$

where the kernel K° *is defined in* (23).

(ii) *For* $s=1,2$

$$\sup_{t\in[b_{n,s},1-b_{n,s}]}\Big|\frac{1}{n_s b_{n_s}^2}\sum_{i=1}^{n_s}K^\circ\Big(\frac{i/n_s-t}{b_{n,s}}\Big)\Big(e_{i,s}-\sigma_s(i/n)V_{i,s}\Big)\Big| = o_p\Big(\frac{\log^2 n_s}{n_s^{3/4}b_{n,s}^2}\Big), \tag{34}$$

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

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

$$\sup_{t\in[b_{n,s},1-b_{n,s}]}|\hat{m}_s'(t)-m_s'(t)| = O_p\Big(\frac{\log n_s}{\sqrt{n_s b_{n,s}}b_{n,s}}+\frac{\log^2 n_s}{n_s^{3/4}b_{n,s}^2}+b_{n,s}^2\Big). \tag{35}$$

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

$$\sup_{t\in[0,b_{n,s}]\cup[1-b_{n,s},1]}|\hat{m}_s'(t)-m_s'(t)| = O_p\Big(\frac{\log n_s}{\sqrt{nb_{n,s}}b_{n,s}}+\frac{\log^2 n_s}{n_s^{3/4}b_{n,s}^2}+b_{n,s}\Big). \tag{36}$$

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

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

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 (34) follows from (33), Proposition 10, the summation by parts formula and similar arguments to derive equation (44) in Zhou (2010).

(iii) + (iv): Following Lemma C.3 of supplement of Dette and Wu (2019), we have

$$\sup_{t\in[b_{n,s},1-b_{n,s}]} \Big|\frac{1}{n_s b_{n,s}} \sum_{i=1}^{n_s} K^\circ\Big(\frac{i/n_s - t}{n_s b_{n,s}}\Big)\Big(\sigma_s(\frac{i}{n_s})V_{i,s}\Big)\Big| = O_p\Big(\frac{\log n_s}{\sqrt{n_s b_{n,s}}}\Big). \quad (37)$$

Finally, (35) follows from (33) (34) and (37) and (36) is obtained by similar arguments using Lemma C.3 in the supplement of Dette and Wu (2019). This completes the proof of Proposition 11. □

Proof of Theorem 4

We only prove the result in the case $g \equiv 0$. The general case follows by the same arguments. Under Assumptions 1 and 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 $\hat f_1(t)$ and $\hat f_2(t)$ are defined in (10) and (11), respectively. Next, since under the null hypothesis (5), $((m_1')^{-1}(t))' - ((m_2')^{-1})'(t) = 0$ for all $t \in (a + \eta, b - \eta)$, (See Lemma 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 $\mathbb{P}[\mathcal{C}_{n_1,n_2} \subset L(\epsilon)] = 1$. □

Proof of Theorem 5

To simplify the notation, we prove Theorem 5 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

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

which is obtained from T_n by replacing the weight function $\hat{w}$ in (18) by its deterministic analogue (20). We shall show Theorem 5 in two steps proving the assertions

$$nb_n^{9/2}\tilde{T}_n - B_n(g) \Rightarrow \mathcal{N}(0, V_T) \tag{38}$$

$$nb_n^{9/2}(T_n - \tilde{T}_n) = o_p(1). \tag{39}$$

For (39), the difference $\tilde{T}_n - T_n$ is contributed by $\hat{w}(t) - w(t)$. The arguments of proving (38) are useful for the proof of (39) and are mathematically involved. We shall discuss the proof of (38) in detail in the next subsection.

Proof of (38)

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 $s = 1, 2$

$$I_s(t) = \frac{1}{Nh_d} \sum_{i=1}^{N} \left(K_d\left(\frac{\hat{m}'_s(i/N) - t}{h_d}\right) - K_d\left(\frac{m'_s(i/N) - t}{h_d}\right)\right), \tag{40}$$

$$II(t) = \frac{1}{Nh_d} \sum_{i=1}^{N} \Big(K_d\Big(\frac{m_1'(i/N) - t}{h_d}\Big) - K_d\Big(\frac{m_2'(i/N) - t}{h_d}\Big)\Big). \quad (41)$$

We shall study $I_s(t)$, $s = 1, 2$ and $II(t)$ via borrowing the idea of proof of Theorem 4.1 of Dette and Wu (2019). In fact, the functions m_s' and $\hat{m}_s'$ for $s = 1, 2$ here play a similar role as the functions μ and $\tilde{\mu}_{b_n}$ in Dette and Wu (2019). 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

$$II(t) = ((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' + O\Big(h_d + \frac{1}{Nh_d}\Big) \quad (42)$$

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

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

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

for some $\theta_s \in [-1, 1]$ ($s = 1, 2$). In the following, we shall prove (38) in the following steps.

(a) Using arguments of Dette and Wu (2019) we show that the leading term of $I_s(t)$ is $I_{s,1}(t)$, $s = 1, 2$.
(b) Using Proposition 11, we approximate the leading term of $I_{1,1}(t) - I_{2,1}(t)$ via a Gaussian process. Therefore, T_n has the form

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

where $U_n(t)$ is a Gaussian process and $R_n^{\dagger}(t)$ is a negligible remaining term.
(c) Under the considered alternative hypothesis, the asymptotic distribution is determined by $\int U_n(t)^2 w(t)dt$ and $\int (((m_1')^{-1}(t))' - ((m_2')^{-1}(t))')^2 w(t)dt$. The latter

accounts for a part of the bias. The former produces another part of the bias and determines the asymptotic stochastic behaviour. All terms in the expansion of T_n that involves $R_n^\dagger(t)$ are negligible.

(c_1) Further calculations show $U_n(t) = U_{n,1}(t) - U_{n,2}(t)$, where

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

where $G(m_s'(\cdot), j, t)$ depends on the kernels, curves, long-run variances and bandwidths. Using a Riemann sum approximation, we can further simplify the leading term of $G(m_s'(\cdot), j, t)$, and hence, the leading term of $\int (U_n(t))^2(t)w(t)dt$ is a quadratic Gaussian.

(c_2) The asymptotic normality is then guaranteed by Theorem 2.1 of de Jong (1987). The mean and variance are obtained via straightforward but tedious calculations and certain arguments from Zhou (2010).

(d) Finally, we show that $\int U_n(t)(((m_1')^{-1}(t))' - ((m_2')^{-1}(t))')w(t)dt$ and $\int U_n(t)R_n^\dagger(t)w(t)dt$ are negligible.

Step (a): By part (iii) and (iv) of Proposition 11 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

$$I_{s,2}(t) = O_p\Big(\frac{\pi_n^2}{h_d^3}(h_d + \pi_n)\Big) = O_p\Big(\frac{\pi_n^3}{h_d^3}\Big)\,(s = 1, 2), \tag{43}$$

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

Step (b): 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 11, it follows that there exists independent 2-dimensional standard normally distributed random vectors $\mathbf{V}_i$ such that

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

uniformly with respect to $t \in [a + \eta, b - \eta]$. Combining this estimate with equations (42) and (43), it follows

$$T_n = \int \big(U_n(t) + ((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' + R_n^{\dagger}(t)\big)^2 w(t)dt, \tag{44}$$

where

$$U_n(t) = \frac{1}{nNb_n^2 h_d^2} \sum_{j=1}^{n} \sum_{i=1}^{N} K^{\circ}\Big(\frac{\frac{j}{n} - \frac{i}{N}}{b_n}\Big)(K_d')^{\top}\Big(\frac{\mathbf{m}'(\frac{i}{N}) - t}{h_d}\Big)\Sigma(\tfrac{j}{n})\mathbf{V}_j, \tag{45}$$

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

$$\sup_{t\in[a+\eta,b-\eta]} |R_n^{\dagger}(t)| = O_p\Big(\frac{\pi_n'}{h_d} + \frac{\pi_n^3}{h_d^3} + h_d + \frac{1}{Nh_d}\Big). \tag{46}$$

Step (c): We now study the asymptotic properties of to the quantities

$$nb_n^{9/2} \int (U_n(t))^2 w(t)dt, \tag{47}$$

$$nb_n^{9/2} \int U_n(t)((m_1^{-1}(t))' - (m_2^{-1}(t))')w(t)dt, \tag{48}$$

$$nb_n^{9/2} \int U_n(t)R_n^{\dagger}(t)w(t)dt, \tag{49}$$

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

$$nb_n^{9/2} \int \rho_n^2(t)g^2(t)w(t) = \int g^2(t)w(t)dt, \tag{50}$$

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

$$nb_n^{9/2} \int (R_n^{\dagger}(t))^2 w(t)dt = o(1), \tag{51}$$

$$nb_n^{9/2} \int \rho_n g(t)R_n^{\dagger}(t)w(t)dt = o(1). \tag{52}$$

Step (c_1): *Asymptotic properties of* (47): To address the expressions related to $U_n(t)$ in (47)–(49) note that

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

where

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

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

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

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

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

where $\mathcal{A}_s(t)\mathcal{A}_s(t) = \big(\frac{m_s'(0)-t}{h_d}, \frac{m_s'(1)-t}{h_d}\big)$, the remainder is given by

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

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 $w(t) \neq 0$

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

With the notation

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

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

$$
\begin{aligned}
\int U_n^2(t)w(t) &= \sum_{s=1}^{2}\sum_{j=1}^{n} V_{j,s}^2 \int G^2(m_s'(\cdot),j,t)^2 w(t)dt \\
&\quad + \sum_{s=1}^{2}\sum_{1\le i\ne j\le n} V_{i,s}V_{j,s}\int G(m_s'(\cdot),i,t)G(m_s'(\cdot),j,t)w(t)dt \\
&\quad - 2\sum_{1\le i\le n} V_{i,1}V_{i,2}\int G(m_1'(\cdot),i,t)G(m_2'(\cdot),i,t)w(t)dt \\
&= \sum_{s=1}^{2}\sum_{j=1}^{n} V_{j,s}^2\left(\int \tilde{G}^2(m_s'(\cdot),j,t)^2 w(t)dt(1+r_{i,s})\right) \\
&\quad + \sum_{s=1}^{2}\sum_{1\le i\ne j\le n} V_{i,s}V_{j,s}\left(\int \tilde{G}(m_s'(\cdot),i,t)\tilde{G}(m_s'(\cdot),j,t)w(t)dt(1+r_{i,j,s})\right) \\
&\quad - 2\sum_{1\le i\le n} V_{i,1}V_{i,2}\left(\int \tilde{G}(m_1'(\cdot),i,t)\tilde{G}(m_2'(\cdot),i,t)w(t)dt(1+r_{i,s}')\right),
\end{aligned}
\tag{53}
$$

where the remainder satisfy

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

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

$$
\tilde{U}_n(t) = \tilde{U}_{n,1}(t) - \tilde{U}_{n,2}(t), \tag{54}
$$

then, by the previous calculations, it follows that

$$
nb_n^{9/2}\left(\int U_n^2(t)w(t)dt - \int \tilde{U}_n^2(t)w(t)dt\right) = o_P(1), \tag{55}
$$

and therefore, we investigate the weak convergence of $nb_n^{9/2}\int \tilde{U}_n^2(t)w(t)dt$ in the following. For this purpose, we use a similar decomposition as in (53) and obtain

$$
\begin{aligned}
\int \tilde{U}_n^2(t)w(t)dt &= \sum_{s=1}^{2}\int (\tilde{U}_{n,s}(t))^2 w(t)dt - 2\int (\tilde{U}_{n,1}(t)\tilde{U}_{n,2}(t))w(t)dt \\
&= \sum_{s=1}^{2}\sum_{j=1}^{n} V_{j,s}^2 \int \tilde{G}^2(m_s'(\cdot),j,t)^2 w(t)dt \\
&\quad + \sum_{s=1}^{2}\sum_{1\le i\ne j\le n} V_{i,s}V_{j,s}\int \tilde{G}(m_s'(\cdot),i,t)\tilde{G}(m_s'(\cdot),j,t)w(t)dt \\
&\quad - 2\sum_{1\le i\le n} V_{i,1}V_{i,2}\int \tilde{G}(m_1'(\cdot),i,t)\tilde{G}(m_2'(\cdot),i,t)w(t)dt \\
&:= D_1 + D_2 + D_3,
\end{aligned}
\tag{56}
$$

where the last equation defines D_1, D_2 and D_3 in an obvious manner.

Step (c_2): Elementary calculations (using a Taylor expansion and the fact that the kernels have compact support) show that

$$
\begin{aligned}
\mathbb{E}(D_1) &= \sum_{s=1}^{2}\sum_{j=1}^{n}\int \Big(\frac{\sigma_s(\frac{j}{n})}{nb_n^3}(K^\circ)'\Big(\frac{\frac{j}{n}-(m_s')^{-1}(t)}{b_n}\Big)(((m_s')^{-1})'(t))^2\kappa_K\Big)^2 w(t)dt \\
&= \sum_{s=1}^{2}\sum_{j=1}^{n}\int \Big(\frac{\sigma_s((m_s')^{-1}(t))}{nb_n^3}(K^\circ)'\Big(\frac{\frac{j}{n}-(m_s')^{-1}(t)}{b_n}\Big)(((m_s')^{-1})'(t))^2\kappa_K\Big)^2 \\
&\quad \times w(t)dt(1+O(b_n)),
\end{aligned}
\tag{57}
$$

where $\kappa_K = \int vK_d'(v)dv$. Using the estimate

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

(uniformly with respect to $t \in [a+\eta, b-\eta]$) and (57) gives

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

which implies

$$
\mathbb{E}(nb_n^{9/2}D_1) = B_n(0) + O\Big(\sqrt{b_n}+\frac{1}{nb_n^{3/2}}\Big),
\tag{58}
$$

where $B_n(g)$ is defined in Theorem 5 (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

$$\mathrm{Var}(nb_n^{9/2} D_1) = O\Big(\frac{1}{nb_n} \Big). \tag{59}$$

For the term D_3 in the decomposition (56), it follows that

$$\begin{aligned}
\mathbb{E}(D_3^2) &= 4 \sum_{1\le i\le n} \Big(\int \tilde{G}(m_1'(\cdot), i, t)\tilde{G}(m_2'(\cdot), i, t)w(t)dt \Big)^2 \\
&= \frac{4\kappa_K^4}{n^4 b^{12}} \sum_i \Big(\int (((m_1')^{-1})')^2(t)(((m_2')^{-1})'(t)(K^\circ)'\Big(\frac{i/n - (m_1')^{-1}(t)}{b_n} \Big) \\
&\quad (K^\circ)'\Big(\frac{i/n - (m_2')^{-1}(t)}{b_n} \Big) w(t)dt \Big)^2 \sigma_1^2(i/n)\sigma_2^2(i/n) = O((n^3 b_n^{10})^{-1}).
\end{aligned}$$

Hence,

$$nb_n^{9/2} D_3 = O_p\Big(\Big(\frac{1}{nb_n} \Big)^{1/2} \Big). \tag{60}$$

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

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

then $\lim_{n\to\infty} V_{s,n}/(n^2 b_n^3)$ 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 5. Now similar arguments as 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 (55), (56), (58), (59), and (60) finally gives

$$nb_n^{9/2} \int U_n^2(t)w(t)dt - B_n(0) \Rightarrow N(0, V_T). \tag{61}$$

Step (d):

Asymptotic properties of (48): Define $d(t) = ((m_1^{-1})' - (m_2^{-1})')$ and note that

$$\int U_n(t)d(t)w(t)dt = \int (U_{n,1}(t) - U_{n,2}(t))d(t)w(t)dt,$$

where

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

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 $s = 1, 2$,

$$nb_n^{9/2} \int (U_{n,s}(t)((m_1^{-1})' - (m_2^{-1})')w(t)dt = O_p(b_n^{1/4}). \tag{62}$$

Asymptotic properties of (49): Note that it follows for the term (49)

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

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 (46) yield

$$\begin{aligned} &nb_n^{9/2} \Big| \int (U_{n,s}(t)(R_n^{\dagger}(t))w(t)dt \Big| \\ &= O_p\Big(\frac{\log^{1/2} n}{n^{1/2} b_n^{5/2}}\Big(\frac{\pi_n'}{h_d} + \frac{\pi_n^3}{h_d^2} + h_d + \frac{1}{Nh_d}\Big)nb_n^{9/2}\Big) = o_p(1).\end{aligned} \tag{63}$$

The proof of assertion (38) is now completed using the decomposition (44) and the results (50), (51), (52), (61), (62) and (63). □

Proof of (39)

From the proof of (38), we have the decomposition

$$T_n - \tilde{T}_n = \int \left(U_n(t) + ((m_1')^{-1}(t))' - ((m_2')^{-1}(t))' + R_n^{\dagger}(t)\right)^2 (\hat{w}(t) - w(t))dt,$$

where quantities I_s, II, $U_n(t)$ and $R_n^{\dagger}(t)$ are defined in (40), (41), and (45). By the proof of (38), 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 (38), 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 (54). Recalling the definition of a, b in (20) it then follows (using similar arguments as given for the derivation of (37)) 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 11 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 (22). 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 Theorem 5 follows from (38) and (39). □

References

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

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

Collier, O., 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.

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

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

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

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

Dette, H., Wu, W. (2019). Detecting relevant changes in the mean of non-stationary processes: A mass excess approach. *The Annals of Statistics*, *47*(6), 3578–3608.

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

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

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

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

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

Kneip, A., Engel, J. (1995). Model estimation in nonlinear regression under shape invariance. *The Annals of Statistics*, *46*, 551–570.

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

Mallat, S., Papanicolaou, G., 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 functions for polychotomous choice models. *Econometrica*, *59*(5), 1315–1327.

Nason, G. P., Von Sachs, R., 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., Dette, H. (2003). Nonparametric comparison of regression curves: An empirical process approach. *Annals of Statistics*, *31*(3), 880–920.

Neumeyer, N., 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., Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, *100*(470), 519–531.

Park, C., Hannig, J., 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, Series B*, *63*(2), 243–259.

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

Vilar-Fernández, J. M., Vilar-Fernández, J. A., 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., 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., Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics*, *37*(5B), 2696–2729.

Zhou, Z., 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.