Submitted to the Annals of Statistics

TESTING NONPARAMETRIC SHAPE RESTRICTIONS

By Tatiana Komarova and Javier Hidalgo

London School of Economics and Political Science

We describe and examine a test for a general class of shape constraints, such as signs of derivatives, U-shape, quasi-convexity, log-convexity, among others, in a nonparametric framework using partial sums empirical processes. We show that, after a suitable transformation, its asymptotic distribution is a functional of the standard Brownian motion, so that critical values are available. However, due to the possible poor approximation of the asymptotic critical values to the finite sample ones, we also describe a valid bootstrap algorithm.

1. INTRODUCTION. Hypothesis testing is one of the most relevant tasks in empirical work. In this paper, we are interested in a type of testing where neither the null hypothesis nor the alternative have a specific parametric form. This type of hypothesis testing can be denoted as testing for *qualitative* or *shape* restrictions. Examples, widespread in economics and other disciplines, include monotonicity, convexity/concavity, strong convexity, log-convexity, as well as shapes which switch the pattern, being two (related) classical examples the *U-shape* and the *quasi-convexity/concavity*.

The class of shape constraints we are concerned with is quite broad. One example is shape constraints that involve some derivatives of $m(x)$, see (1.1) below, and in particular whether $\partial^r m(x)/\partial x^r \geq 0$ (≤ 0). When $r = 1$ or 2, we have respectively the classical examples of monotonicity or convexity/concavity. A second example involves shapes well examined in the mathematics literature such as log-convexity/concavity. However, the applicability of the methodology proposed below goes beyond these examples and they should be viewed as an illustration of the scope of the approach. More specifically, in Section 2 we give some specific conditions on the type of shape constraints we consider and some examples, whereas some additional examples of shapes of possible interest, such as the quasi-convexity or r- and ρ- convexity/concavity are contained in the supplementary material.

Although there is an ample literature on testing for shape constraints, it mostly focuses on monotonicity or convexity. Examples include [9], [35],

MSC 2010 subject classifications: Primary 05C38, 15A15; secondary 05A15, 15A18

Keywords and phrases: Monotonicity, convexity, concavity, U-shape, quasi-convexity, log-convexity, convexity in means, B-splines, CUSUM transformation, distibution-free-estimation

[31], [42], [63], [15], [59], [22], [25], [1], [6], [20], [3]. It is worth noting the exception in [44], who proposed a consistent test for *U-shape*.

When looking at the regularity conditions in the latter references, some of them, such as [6], [25], [42], focus on regression function with Gaussian white noise model, or on the condition that the explanatory variable are deterministic, see [3], [6], [22] or [35]. However, with random explanatory variables, as in [1], [15] and [31], it is assumed either that they are stochastic independent to the unobserved regression error or that the distribution of the error is symmetric conditional on the explanatory variable. On the other hand, other papers do not provide asymptotic limit theory useful for the purpose of inference or they are tailored to a specific type of shape and their extensions to more general shape properties do not appear straightforward, such as [9], [35] or [44]. Due to these (potential) caveats, one of the purposes of the paper is to examine a testing methodology which is not only applicable for a wide range of shape properties but (**a**) able to perform valid statistical inferences under weak conditions and (**b**) flexible enough to be able to test for more than one shape constraint, for instance testing for monotonicity and log-convexity simultaneously. In addition, our aim is that the proposed methodology would still be easy to implement. In fact, as we shall see it only requires the computation of the *CUSUM* (least squares) of "recursive" residuals.

The methodology we propose in this paper is related to methods used in goodness of fit tests, where the null hypothesis is assumed to belong to a parametric family leaving the alternative nonparametric. To be more precise, consider the nonparametric regression model

$$\begin{aligned} y_i &= m\left(x_i\right) + u_i, \\ E[u_i|x_i] &= 0, \end{aligned} \tag{1.1}$$

where x_i has bounded support $\mathcal{X} =: [\underline{x}, \overline{x}]$ and $m\left(\cdot\right)$ is smooth. More specific conditions on the sequences $\{u_i\}_{i\in\mathbb{N}}$ and $\{x_i\}_{i\in\mathbb{N}}$ will be given in Condition $C1$ in Section 3. Our aim is testing whether the regression function $m\left(x\right)$ possesses the shape properties captured by the null hypothesis

$$H_0: \quad m \in \mathcal{M}_0, \tag{1.2}$$

where the class of interest $\mathcal{M}_0$ is a subset of smooth functions from $\mathcal{X}$ to $\mathbb{R}$, say convexity. Following [61] or [5], we might base the testing procedure on functionals of the partial sums empirical process

$$\mathcal{K}_n\left(x\right) = \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i\mathcal{I}_i\left(x\right), \qquad x \in [\underline{x}, \overline{x}] \tag{1.3}$$

where $\mathcal{I}(\cdot)$ is the indicator function, we abbreviate $\mathcal{I}(x_i < x)$ as $\mathcal{I}_i(x)$, and

$$\widehat{u}_i = y_i - \widehat{m}_{\mathcal{B}}(x_i; L), \quad i = 1, \ldots, n,$$

are the residuals obtained after $m(\cdot)$ has been estimated by some nonparametric estimator $\widehat{m}_{\mathcal{B}}(x_i; L)$, see Section 2 for details.

Unfortunately, after normalization, the limit covariance structure of $\mathcal{K}_n(x)$ depends on $\mathcal{M}_0$, making inferences based on $\mathcal{K}_n(x)$ very difficult to perform, if at all possible. Indeed, when $\mathcal{M}_0$ is the set of some parametric functions, say $m(x) =: m(x;\theta)$, we have that

$$\mathcal{K}_n(x) = \frac{1}{n}\sum_{i=1}^{n} u_i \mathcal{I}_i(x) + \frac{1}{n}\sum_{i=1}^{n}\left(m(x_i;\theta) - m(x_i;\widehat{\theta})\right)\mathcal{I}_i(x)$$

is such that $\left(Eu_i^2\right)^{-1/2} n^{-1/2}\sum_{i=1}^{n} u_i \mathcal{I}_i(x)$ converges to the standard Brownian motion whereas the second term normalized by $n^{1/2}$ converges to a Gaussian random variable which depends on $m(x;\theta)$, and hence on $\mathcal{M}_0$. This was first noticed and shown in [26], and later in a regression model context by [61]. However, in our scenario, we have that

$$\mathcal{K}_n(x) = \frac{1}{n}\sum_{i=1}^{n} u_i \mathcal{I}_i(x) + \frac{1}{n}\sum_{i=1}^{n}\left(m(x_i) - \widehat{m}_{\mathcal{B}}(x_i;L)\right)\mathcal{I}_i(x), \tag{1.4}$$

where the second term is $O_p(n^{-\nu})$, for some $\nu < 1/2$, becoming then the dominant term in the behaviour of $\mathcal{K}_n(x)$. As we describe in Section 3, a consequence is that the asymptotic distribution of $\mathcal{K}_n(x)$ might not be even Gaussian and difficult to characterize, making inferences very cumbersome.

Due to the possible drawbacks of $\mathcal{K}_n(x)$ for the purpose of inference, we shall proceed by considering a transformation of $\mathcal{K}_n(x)$ related to the *CUSUM* of recursive residuals proposed by [10]. More specifically, the asymptotic behaviour of the transformation becomes a standard Brownian motion, and as a consequence, testing can be implemented using standard functionals such as Kolmogorov-Smirnov, Cramér -von-Mises or Anderson-Darling. As a consequence, a nice feature of the transformation is that its asymptotic distribution is pivotal, i.e. it is the same regardless of the shape constraint under consideration.

The remainder of the paper is organized as follows. Section 2 introduces and motivates the *B-splines* to estimate our nonparametric regression function $m(x)$. We then examine how our estimated model captures the shape property of interest, by relating the different shapes to the coefficients of the *B-splines* approximation. In Section 3, we state the regularity conditions, and we motivate and describe a pivotal transformation of $\mathcal{K}_n(x)$ based on the *CUSUM* of recursive residuals. We also describe the local alternatives and show the consistency of the test. Because the Monte-Carlo experiment suggests that the asymptotic critical values do not provide a good approximation to the finite sample ones, Section 4 introduces a valid bootstrap algorithm. Section 5 presents a Monte-Carlo experiment and some empirical examples, whereas Section 6 concludes with a summary and possible extensions of the methodology. All the proofs, which employ a series of lemmas, are confined to the supplementary document. The supplementary document also contains some additional material such as (**i**) additional simulation results for our test including its performance when using the asymptotic critical values, and comparison of its performance to some other tests in the literature in the context of testing for monotonicity, (**ii**) motivation for using *B-splines* instead of some other sieve-type of estimator, and (**iii**) additional examples of shape constraints of interest.

2. NONPARAMETRIC ESTIMATION METHODOLOGY.

A preliminary and key step to provide a test for H_0 in (1.2) is to compute a nonparametric estimator of $m(x)$ subject to the constraints imposed in H_0. When testing for the null hypothesis of either monotonicity or convexity, several nonparametric estimators have been considered in the literature. Early work on isotone/monotone regressions is [11] and [64]. Later approaches include [29], [52] and [48], [36], [21] and [14], with the first two papers incorporating isotonization as part of their methods and the last two papers relying on rearrangement methods. When the null hypothesis is that of convexity, [40] approach is based on estimating $m(\cdot)$ by least squares approach. Asymptotic properties for this estimator are established in [37], [49] and [33], and its global behaviour is examined in [34]. [7] consider an estimator based on first obtaining unconstrained estimate of the derivative of the regression function which is isotonized and then integrated.

However, the previous techniques have some drawbacks either because sometimes are difficult to implement or quite narrow in their scope or they lack asymptotic theory useful for the purpose of inference. Due to this, we shall use a different approach based on *B-splines* and/or penalized *B-splines* known as *P-splines*. There are several reasons why the use of *B-splines* (*P-splines*) is appealing in the context of this paper. One of them is the absence of a dependence between base splines that are separated by a certain distance, as listed in the properties of the *B-spline* basis below. A second motivation is that *B-splines* (*P-splines*) are particularly convenient for testing properties based on the derivatives of the regression function, as discussed later in this section. A further motivation is the ability to write $\mathcal{M}_0$ in (1.2) in terms of restrictions on the coefficients of the *B-splines* approximation to $m(\cdot)$, and as a consequence the implementation of valid asymptotic theory for the test. We should mention that [57], and later extended by [51], introduced monotone regression splines (closely related to *B-splines*) to estimate convex/concave function or functions that are, e.g., both monotone and convex. However one difference of our approach, compared to those in the aforementioned work, is that we let the number of coefficients of the *B-splines* to increase to infinity, and hence the number of constraints. Both [57] and [51] considered the number of constraints fixed. [63] uses quadratic *B-splines* to design a test for monotonicity and cubic *B-splines* to design a test for convexity. It allows for an increasing number of knots but its idea and implementation are different from ours (in particular, it might not be possible to be extended to general shapes). The comparison of the performance of the monotonicity test in [63] and our test is given in the supplementary document.

Let us now describe the *B-splines* and *P-splines* in more detail. *B-splines* or *P-splines* are constructed from polynomial pieces joined at some specific points denoted knots, and whose computation is obtained recursively, see [17], for any degree of the polynomial. In general, the *B-spline* basis of degree q

- takes positive values on the domain spanned by $q + 2$ adjacent knots, and is zero otherwise;
- consists of $q+1$ polynomial pieces each of degree q, and the polynomial pieces join at q inner knots;
- at the joining points, the $(q - 1)$th derivatives are continuous;
- except at the boundaries, it overlaps with $2q$ polynomials pieces of its neighbours;
- at a given x, only $q + 1$ *B-splines* are nonzero.

Suppose that one is interested in approximating the regression function $m(x)$ in the interval $[0, 1]$, where herewith we shall assume, without loss of generality, that $\mathcal{X} = [0, 1]$. Then we split the interval $[0, 1]$ into L' equal length subintervals with $L'+1$ knots[1], where each subinterval will be covered with $q + 1$ *B-splines* of degree q. The total number of knots needed will be $L' + 2q + 1$ (each boundary point 0, 1 is a knot of multiplicity $q + 1$) and the number of *B-splines* is $L = L' + q$. So, denoting the *B-splines* basis of

[1] Although one can, of course, choose nonequidistant subintervals, for simplicity we consider equally spaced knots. One alternative way to locate the knots may based on the quantiles of the x distribution.

degree q by

$$\boldsymbol{P}_L(x) =: (p_{1,L}(x;q), \ldots, p_{L,L}(x;q))', \tag{2.1}$$

we approximate $m(x)$ by a linear combination of $\boldsymbol{P}_L(x)$, that is

$$m_{\mathcal{B}}(x;L) = \sum_{\ell=1}^{L} \beta_\ell p_{\ell,L}(x;q), \tag{2.2}$$

and where henceforth we shall denote the knots as $\{z^k\}$, $k = 1-q, \ldots 0, 1, ..., L'+q+1$, where $0 = z^{1-q} = \ldots = z^1$ and $1 = z^{L'+1} = \ldots = z^{L'+q+1}$.

It is well understood that the choice of the number of knots determines the trade-off between overfitting and underfitting when there are respectively too many or too few knots. The main difference between *B-splines* and *P-splines* is that the latter tend to employ a large number of knots but to avoid oversmoothing they incorporate a penalty function based on the second difference $\triangle^2\beta_\ell$, where $\triangle\beta_\ell = \beta_\ell - \beta_{\ell-1}$.

The methodology and applications of constrained *B-splines* and *P-splines* (that is, those computed under certain constraints on the coefficients) are discussed by many authors, too many to review here. For more detailed discussions, see, among others, the monographs [17] and [24] for *B-splines* and [27], [8] for *P-splines*. Some literature on shape-preserving splines (for standard shapes such as monotonicity, convexity, etc.) includes, among others, [47], [50], [51] and [57].

B-splines possess some properties which turn out to be very useful for the purpose of testing shape constraints. Among them are

$$\begin{aligned}
&(\mathbf{a}) \ \sum_{\ell=1}^{L} p_{\ell,L}(x;q) &&= 1 \quad \text{for all } x \ \text{ and } q. \\
&(\mathbf{b}) \ \frac{\partial m_{\mathcal{B}}(x;L)}{\partial x} &&= m'_{\mathcal{B}}(x;L) \\
& &&= \sum_{\ell=1}^{L-1} \frac{q\triangle\beta_{\ell+1}}{z^{l+1} - z^{l+1-q}} p_{\ell+1,L}(x;q-1).
\end{aligned} \tag{2.3}$$

In particular, (**a**) indicates that *B-splines* are a partition of 1. The property (**b**) states that the derivative of a *B-spline* of degree q becomes a *B-spline* of degree $q-1$. One can derive an expression for the second derivative, and so on. It is worth signaling that other sieve estimators might be used, see the survey in [12], and in particular *Bernstein* polynomials basis as they share some properties similar to those in (2.3). However, because the Bernstein polynomials have an undesirable property of being highly correlated and having a slow bias convergence, they are not useful for the methodology proposed below[2].

We now describe estimators of $m(\cdot)$ under the null hypothesis, and more crucially how we can relate the *B-spline* approximation $m_{\mathcal{B}}(x_i; L)$ to (1.2). In particular, because any $m_{\mathcal{B}}(\cdot; L)$ can be fully characterized by the vector $\beta =: (\beta_1, \ldots, \beta_L) \in \mathbb{R}^L$, a first step will be to examine how we can map the null hypothesis into a set of constraints on β, captured by some subset $S_{q,L} \subset \mathbb{R}^L$, and denoting its associated constraint *B-splines* approximation by

$$\mathcal{M}_{S_{q,L}} =: \{m_{\mathcal{B}}(\cdot; L) \mid \beta = (\beta_1, \ldots, \beta_L) \in S_{q,L}\}. \tag{2.4}$$

We can summarize it in the form of the following condition.

***Condition** $C0$. There is a set $S_{q,L} \subseteq \mathbb{R}^L$ for any $L = L' + q$ that satisfies the following properties:*

(a) $S_{q,L}$ does not depend on the data $\{x_i\}_{i\in\mathbb{Z}}$ and thus it is non-stochastic;
(b) The boundary of $S_{q,L}$ consists of a finite number of surfaces with each surface being explicitly expressed through a continuously differentiable function of one of the components in β in terms of other components of β; that is, each surface forming the boundary can be described by $\beta_\ell = s(\beta_{-\ell})$ for some ℓ with s being continuously differentiable.
(c)
$$\mathcal{H}\left(\mathcal{M}_0, \mathcal{M}_{S_{q,L}}\right) \to 0 \quad \text{as} \quad L \to \infty, \tag{2.5}$$
where $\mathcal{H}$ is the Hausdorff distance in the supremum norm in the space of continuous functions from $\mathcal{X}$ to $\mathbb{R}$.[3]

What Condition $C0$ essentially states is that $\mathcal{M}_0$ can be captured by restrictions on the parameters $\beta = \{\beta_\ell\}_{\ell=1}^L$ which become *both necessary and sufficient* as the system of knots becomes increasingly dense in $\mathcal{X}$. The property of the knot system becoming increasingly dense is implied by the

[2] A further discussion of our motivation to not use other sieves bases can be found in the supplementary document.

[3] That is, if $d(m(x), m_{\mathcal{B}}(x; L)) =: \sup_{x\in\mathcal{X}} |m(x) - m_{\mathcal{B}}(x; L)|$, then the Hausdorff distance is

$$\mathcal{H}\left(\mathcal{M}_0, \mathcal{M}_{S_{q,L}}\right) = \max\{\sup_{m(\cdot)\in\mathcal{M}_0} \inf_{m_{\mathcal{B}}(\cdot;L)\in\mathcal{M}_{S_{q,L}}} d(m(x), m_{\mathcal{B}}(x; L)),\ \sup_{m_{\mathcal{B}}(\cdot;L)\in\mathcal{M}_{S_{q,L}}} \inf_{m(\cdot)\in\mathcal{M}_0} d(m(x), m_{\mathcal{B}}(x; L))\}.$$

requirement (c). In addition, these restrictions on β do not depend on the available data, which adds to the attractiveness of the approach for implementation purposes. Part (b) is required for practical purposes – it ensures that for any finite L the constrained estimation only requires imposing a finite number of inequality constraints.

The idea is then to test the null hypothesis

$$H_0^{\mathcal{B}}: \quad (\beta_1, \ldots, \beta_L) \in S_{q,L} \tag{2.6}$$

with the suitable choice of $S_{q,L}$. Under $C0$, for fixed L, the test in (2.6) can be conceptually regarded as the approximation of the original testing problem in (1.2). However, it is important to stress that as $L \to \infty$, the knot system becomes dense in $\mathcal{X}$ and, thus, the shape property of interest is satisfied on an increasingly dense set of points in $\mathcal{X}$. In addition, for the typical shapes given in Example 1 below, i.e. monotonicity or convexity, those given in (2.6) are equivalent to restrictions in the whole domain $\mathcal{X}$, i.e. $\mathcal{M}_0$ in (1.2).

To obtain an estimator under the null hypothesis, we consider estimation under the constraints in (2.6), that is

$$\widehat{b} = \left(\widehat{b}_1, \ldots, \widehat{b}_L\right) =: \arg \min_{\substack{b_1,\ldots,b_L \\ \text{s.t. } (b_1,b_2,\ldots,b_L) \in S_{q,L}}} \sum_{i=1}^{n} \left(y_i - \sum_{\ell=1}^{L} b_\ell p_{\ell,L}\left(x_i; q\right) \right)^2, \tag{2.7}$$

so that under (1.2)/(2.6), the estimator of $m(x)$ is

$$\widehat{m}_{\mathcal{B}}(x; L) = \widehat{b}' \boldsymbol{P}_L(x). \tag{2.8}$$

As an example, suppose that we are interested in testing for nondecreasing functions. Then, as Example 1 below will indicate, (2.7) becomes

$$\widehat{b} = \left(\widehat{b}_1, \ldots, \widehat{b}_L\right) =: \arg \min_{\substack{b_1,\ldots,b_L \\ \text{s.t. } b_1 \le b_2 \le \ldots \le b_L}} \sum_{i=1}^{n} \left(y_i - \sum_{\ell=1}^{L} b_\ell p_{\ell,L}\left(x_i; q\right) \right)^2, \tag{2.9}$$

which is a quadratic programing problem with linear constraints. When the constraints are nonlinear, such as those in Example 2 below, the constrained estimation may be implemented using global optimization techniques.[4] A further discussion of nonlinear constraints based on Example 2 and their implementation can be found in the supplementary document.

[4]If the unconstrained least squares estimator is in the interior of $S_{q,L}$ then, of course, none of the constraints are binding and the constrained estimation is standard. The computational complications may only happen when some of the constraints are binding.

As an illustration it is worth describing how we can write (2.8) when some constraints are binding, say $\widehat{b}_{\ell_0} = \widehat{b}_{\ell_0+1}$ in (2.9). Denote

$$\widetilde{p}_{\ell,L}(x;q) = \begin{cases} p_{\ell,L}(x;q) & \ell < \ell_0 \\ p_{\ell_0,L}(x;q) + p_{\ell_0+1,L}(x;q) & \ell = \ell_0 \\ p_{\ell+1,L}(x;q) & \ell_0 < \ell \leq L-1 \end{cases}' .$$

Then (2.8) can be written as

$$\widehat{m}_{\mathcal{B}}(x;L) = \sum_{\ell=1}^{\ell_0-1} b_\ell \widetilde{p}_{\ell,L}(x_i;q) + b_{\ell_0}\widetilde{p}_{\ell_0,L}(x_i;q) + \sum_{\ell=\ell_0+1}^{L-1} b_{\ell+1}\widetilde{p}_{\ell,L}(x_i;q)$$

that is, $\{\widetilde{p}_{\ell,L}(x;q)\}_{\ell=1}^{L-1}$ is the set of "effective" polynomials used in the estimated constrained approximation $\widehat{m}_{\mathcal{B}}(x;L)$. Such a system of "effective"polynomials can be defined for any situation of binding set constraints. We will denote this system as $\widetilde{\boldsymbol{P}}_L(x)$ and further denote

$$\widetilde{\boldsymbol{P}}_k =: \widetilde{\boldsymbol{P}}_L(x_k). \tag{2.10}$$

It is worth noting that the unconstraint estimator of $m(x)$ is defined as

$$\begin{aligned} \widetilde{m}_{\mathcal{B}}(x_i;L) &= \widetilde{b}'\boldsymbol{P}_i, \\ \widetilde{b} &= \left(\widetilde{b}_1,\ldots,\widetilde{b}_L\right)' = \left(\frac{1}{n}\sum_{k=1}^n \boldsymbol{P}_k\boldsymbol{P}_k'\right)^+ \frac{1}{n}\sum_{k=1}^n \boldsymbol{P}_k y_k, \end{aligned} \tag{2.11}$$

where B^+ denotes the Moore-Penrose inverse of the matrix B and we abbreviate $\boldsymbol{P}_L(x_k)$ in (2.1) by $\boldsymbol{P}_k$.

We finish this section with two examples of shape constraints. The shapes in Example 1 pertains to the case where the constraints on the coefficients of the *B-splines* approximation are linear, whereas for the set of shapes in Example 2, these constraints are nonlinear (except for some special cases). These examples are meant to illustrate the scope of applicability of our testing methodology rather than to give an exhaustive list of potential applications. Additional examples can be found in the supplementary document.

2.1. ***EXAMPLES***.

EXAMPLE 1 (Constraints on the derivatives of $m(\cdot)$).

$$H_0 : d_r \cdot m^{(r)}(x) \geq c_r, \qquad r \in R, \tag{2.12}$$

where R is a finite subset of $\mathbb{N}^+$, $d_r \in \{-1,1\}$ and c_r are known constants, so that (2.12) allows for inequalities for several derivatives simultaneously. Special cases include testing for (i) ***monotonicity*** *($r=1$ and $c_1=0$), (ii)* ***convexity/concavity*** *($r=2$ and $c_2=0$), (iii)* ***strong λ-convexity*** *($r=2$ and $c_2=\lambda>0$), (iv)* ***monotonicity and concavity simultaneously****, etc.*

The corresponding set $S_{q,L}$ associated to (2.12) is

$$S_{q,L} = \Big\{ (\beta_1,\ldots,\beta_L) \mid \forall\, r \in R, \quad \forall\, z^k, k = 1-q,\ldots,-1,0,1,\ldots,L'+q+1, \\ d_r \cdot m^{(r)}{}_{\mathcal{B}}\left(z^k; L\right) \geq c_r \Big\}.$$

Observe that $S_{q,L}$ imposes only shape constraints at the knots. However, in the leading cases when $c_r=0$, $r\in R$, $S_{q,L}$ has a more familiar structure which guarantees that the shape properties are not only valid at the knots but on the whole domain. E.g., if $R=\{1\}$ and $c_1=0$, then using the property (b) of B-splines we have that

$$S_{q,L} = \left\{ (\beta_1,\ldots,\beta_L) \mid d_1(\beta_{\ell+1}-\beta_\ell) \geq 0, \quad \ell = 1,\ldots,L-1 \right\},$$

which together with the fact that the B-splines are nonnegative, it guarantees that the approximation is monotone on the whole domain.

When $R=\{r\}$, $r>1$, $c_r=0$, then $S_{q,L}$ can be defined in a more convenient way than above. E.g., if we split the interval $[0,1]$ into equidistance subintervals, we can describe $S_{q,L}$ quite easily as

$$S_{q,L} = \left\{ (\beta_1,\ldots,\beta_L) \mid d_r \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \beta_{\ell+k} \geq 0,\ \ell = q,\ldots,L-q+1-r \right\}. \tag{2.13}$$

An even more refined form of $S_{q,L}$, which may be beneficial for small L, would also involve constraints that capture the behaviour of $m_{\mathcal{B}}(x;L)$ around the boundary. These constraints are linear inequalities and are slightly different from those in (2.13) and only involve coefficients corresponding to the B-splines around the boundaries. Just to give an example, for $r=2$, $d_r=1$ and $c_r=0$ (i.e. testing for convexity), the additional inequalities around the boundary are

$$(q-1)\triangle\beta_{q+1} \geq q\triangle\beta_q, \ \ \ldots, \ \ \triangle\beta_3 \geq 2\triangle\beta_2,$$
$$(q-1)\triangle\beta_{L-q+1} \leq q\triangle\beta_{L-q+2}, \ \ \ldots, \ \ \triangle\beta_{L-1} \leq 2\triangle\beta_L.$$

These additional constraints together with those in (2.13) for $r=2$ and $d_r=1$ will ensure the convexity of the approximation on the whole domain.

As $L \to \infty$, the additional constraints becomes less and less important as constraints (2.13) essentially capture the convexity property in the whole domain. However, in finite samples, these constraints around the boundary can be important to increase the power of the test.

Our second example illustrates shape properties well developed in the mathematical literature. For that purpose, it is convenient to give the following definition, see [53].

DEFINITION 1 (mean function). *A function $N : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is called a mean function if*

- *(a) $N(x_1, x_2) = N(x_2, x_1)$, (b) $N(x, x) = x$, (c) $x_1 < N(x_1, x_2) < x_2$ whenever $x_1 < x_2$ and (d) $N(ax_1, ax_2) = aN(x_1, x_2)$ for all $a > 0$.*
- *$N(x, x) = x$*
- *$x_1 < N(x_1, x_2) < x_2$ whenever $x_1 < x_2$*
- *$N(ax_1, ax_2) = aN(x_1, x_2)$ for all $a > 0$*

Examples of mean functions include the *arithmetic mean* (A), the *geometric mean* (G), the *harmonic mean* (H), the *logarithmic mean* and the *identric mean.*

EXAMPLE 2 (MN-convexity). *For any two mean functions M and N, the class of* **MN-*convex*** *is defined as*[5]

$$\mathcal{M}_0 = \left\{ \phi(\cdot) : \ \phi(\cdot) > 0, \quad \forall\, x_1, x_2 \in \mathcal{X} \quad \phi(M(x_1, x_2)) \le N(\phi(x_1), \phi(x_2)) \right\}.$$

When we have different combinations of arithmetic (A), geometric (G) and harmonic (H) means, we end up with the following special cases of MN-convex functions (see e.g. [4]), for instance

1. *m is AG-convex if and only if $\frac{m'(x)}{m(x)}$ is increasing*
2. *m is AH-convex if and only if $\frac{m'(x)}{m^2(x)}$ is increasing*
3. *m is GG-convex if and only if $\frac{xm'(x)}{m(x)}$ is increasing*
4. *m is GH-convex if and only if $\frac{xm'(x)}{m^2(x)}$ is increasing*
5. *m is HG-convex if and only if $\frac{x^2m'(x)}{m(x)}$ is increasing*
6. *m is HH-convex if and only if $\frac{x^2m'(x)}{m^2(x)}$ is increasing*
7. *m is GA-convex if and only if $xm'(x)$ is increasing*

[5] The definition of **MN-concavity** would reverse the inequalities.

8. m is HA-convex if and only if $x^2 m'(x)$ is increasing.

AG-convexity is known as log-convexity. To illustrate the form of sets $S_{q,L}$ consider the case of HG-convexity when we can take

$$S_{q,L} = \Big\{(\beta_1, \ldots, \beta_L) \mid \forall\, z^{k_1} < z^{k_2}, \quad k_1, k_2 \in \{1-q, \ldots, L'+q+1\},$$
$$\frac{\left(z^{k_1}\right)^2 m'_{\mathcal{B}}\left(z^{k_1}; L\right)}{m^2_{\mathcal{B}}\left(z^{k_1}; L\right)} \leq \frac{\left(z^{k_2}\right)^2 m'_{\mathcal{B}}\left(z^{k_2}; L\right)}{m^2_{\mathcal{B}}\left(z^{k_2}; L\right)}, \text{ if } \quad \beta_\ell > 0,\ \ell = 1, \ldots, L\Big\}.$$

The sets $S_{q,L}$ for other MN-convex functions are constructed similarly. In some special cases (such as GA-convexity, HA-convexity, or AA-convexity) the constraints on β_ℓs will be linear.

We want to emphasize that the properties of *B-splines* are key for the testing of these hypotheses to be easily implemented.

3. REGULARITY CONDITIONS AND THE TESTING METHODOLOGY.

We start this section by introducing our regularity conditions.

Condition C1 $\{(x_i, u_i)'\}_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed random vectors, where x_i has support on $\mathcal{X} =: [0,1]$ and its probability density function, $f_X(x)$, is bounded away from zero. In addition, $E[u_i|x_i] = 0$, $E[u_i^2|x_i] = \sigma_u^2$, and u_i has finite $4th$ moments.

Condition C2 $m(x)$ is η times continuously differentiable on $[0,1]$, $\eta \geq 1$, and $\partial^\eta m(x)/\partial x^\eta$ is Hölder continuous with exponent $0 < \alpha \leq 1$:

$$|\partial^\eta m(x_1)/\partial x^\eta - \partial^\eta m(x_2)/\partial x^\eta| \leq M_0 |x_1 - x_2|^\alpha,$$

for some finite positive constant M_0.

Condition C3 As $n \to \infty$, L satisfies

$$\left(\frac{L^{1+\eta+\alpha}}{n} + \frac{n}{L^{2(\eta+\alpha)}}\right)\mathcal{I}(\eta+\alpha<2) + \left(\frac{L^3}{n} + \frac{n}{L^4}\right)\mathcal{I}(2 \leq \eta+\alpha) = o(1).$$

As it was done in [61], Condition $C1$ can be weakened to allow for heteroscedasticity, e.g. $E\left[u_i^2 \mid x\right] = \sigma_u^2(x)$. However, the latter condition complicates the technical arguments and for expositional simplicity we omit a detailed analysis of this case. However, in our empirical applications we present examples with heteroscedastic errors and illustrate how to deal with them in practice. Condition $C2$ is a regularity condition on the regression function $m(x)$. In a nutshell, it states that we need slightly more than continuous differentiability of $m(\cdot)$. It guarantees that the approximation error or bias

$$\widetilde{m}(x) =: m_{\mathcal{B}}(x; L) - m(x) \tag{3.1}$$

is $O(L^{-\eta-\alpha})$, see Theorems 3.1 and 4.1 in [2] or [65], see also [13] and references therein. In case of using *P-splines* we also refer to [16] Theorem 2. Condition $C3$ bounds the rate at which L increases to infinity with n.

We now describe the testing methodology in more detail. We shall focus on the null hypothesis (2.6) which is given in terms of the coefficients β, with the alternative hypothesis being the negation of the null. So our testing problem translates into the more familiar testing scenario when the null hypothesis is given as a set of constraints on the parameters of the model. However the main and key difference is that the number of such constraints increases with the sample size.

As we discussed in the introduction, we might employ functionals of (1.3) for the purpose to test for (2.6), that is if

$$\mathcal{K}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i \mathcal{I}_i(x), \qquad x \in [0,1]$$

is significantly different than zero, where $\widehat{u}_i$ are the residuals given by

$$\widehat{u}_i = y_i - \widehat{m}_{\mathcal{B}}(x_i; L), \quad i = 1, \ldots, n. \tag{3.2}$$

Notice that $\mathcal{K}_n(x)$ can be interpreted as a LM type of test. Recall that in a standard regression model the LM test is based on the first order conditions

$$\mathcal{LM}_n(L) = \frac{1}{n}\sum_{i=1}^{n} p_{\ell,L}(x_i; q)\, \widehat{u}_i,$$

so that we test if the residuals and regressors, $p_{\ell,L}(x_i; q)$, satisfy the orthogonality moment condition induced by Condition $C1$.

Using Conditions $C2$ and $C3$, we have that (1.4) is

$$\mathcal{K}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\left(u_i - \sum_{\ell=1}^{L}\left(\widehat{b}_\ell - \beta_\ell\right) p_{\ell,L}(x_i; q)\right)\mathcal{I}_i(x)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\left(m(x_i) - m_{\mathcal{B}}(x_i; L)\right)\mathcal{I}_i(x).$$

Now following Lee and Robinson [46] or Chen and Christensen [13] in a more general context, we obtain that $\sum_{\ell=1}^{L}\left(\widehat{b}_{\ell}-\beta_{\ell}\right)\frac{1}{n}\sum_{i=1}^{n}p_{\ell,L}\left(x_i;q\right)=O_p\left((L/n)^{1/2}\right)$, and denoting $\mathcal{P}_{n,\ell}\left(x;q\right)=:n^{-1}\sum_{i=1}^{n}p_{\ell,L}\left(x_i;q\right)\mathcal{I}_i\left(x\right)$, we conclude that

$$\left(\frac{n}{L}\right)^{1/2}\mathcal{K}_n\left(x\right)=:-\left(\frac{n}{L}\right)^{1/2}\sum_{\ell=1}^{L}\left(\widehat{b}_{\ell}-\beta_{\ell}\right)\mathcal{P}_{n,\ell}\left(x;q\right)\left(1+o_p\left(1\right)\right).$$

The last displayed expression suggests that when β is at the boundary of $S_{q,L}$, the asymptotic distribution is not Gaussian, and so to obtain the asymptotic distribution of $\mathcal{K}_n\left(x\right)$ for inference purposes appears quite difficult, if at all possible.

However, $\mathcal{K}_n\left(x\right)$ can be written as

$$\mathcal{K}_n\left(x\right)=\frac{1}{n}\sum_{i=1}^{n}\widetilde{u}_i\mathcal{I}_i\left(x\right)+o_p\left(n^{-1/2}\right)$$

where

$$\widetilde{u}_i=u_i-\widetilde{\boldsymbol{P}}_i^{\prime}\left(\sum_{k=1}^{n}\widetilde{\boldsymbol{P}}_k\widetilde{\boldsymbol{P}}_k^{\prime}\right)^{+}\sum_{k=1}^{n}\widetilde{\boldsymbol{P}}_k u_k. \tag{3.3}$$

and $\widetilde{\boldsymbol{P}}_k$ is as defined in (2.10) (and, thus, already incorporating all the binding constraints in the constrained estimation).[6]

Now $\widetilde{u}_i$ in (3.3) has the interpretation of being the least squares residuals in an artificial regression model with dependent variable u_i and a vector of “effective” polynomials/explanatory variables $\widetilde{p}_{\ell,L}\left(x_i;q\right)$. So, the latter observation suggests employing the *CUSUM* of recursive residuals for constructing asymptotically pivotal tests, as were proposed by Brown, Durbin and Evans [10], see also Sen [58]. To that end, it is useful to describe how to implement the *CUSUM* of recursive residuals when the restrictions in $S_{q,L}$ are linear first, leaving the more general scenario when some of the constraints are nonlinear for later.

[6] Clearly, $\sum_{\ell=1}^{L}\left(\widehat{b}_{\ell}-\beta_{\ell}\right)p_{\ell,L}\left(x_i;q\right)$ can also be rewritten in terms of polynomials in $\widetilde{\boldsymbol{P}}_k$ only, thus incorporating the binding constraints but for conveying some intuition about its asymptotic behavior it is convenient for us to leave this term as it is.

3.1. *ALL THE CONSTRAINTS ON β_ℓ ARE LINEAR.*

To describe our pivotal transformation, we first recall our notation in (2.1) and (2.10) when testing for monotonicity,

$$\begin{aligned}\boldsymbol{P}_k &= :\boldsymbol{P}_L(x_k)\,,\ \ \widetilde{\boldsymbol{P}}_k =: \widetilde{\boldsymbol{P}}_L(x_k)\,,\quad \text{where}\\ \boldsymbol{P}_L(x) &= :(p_1(x),\ldots,p_L(x))'\,,\\ \widetilde{\boldsymbol{P}}_L(x) &= :\text{set of “effective”polynomials in the constrained } \widehat{m}_{\mathcal{B}}(x;L)\,,\end{aligned}$$

where for notational simplicity we suppress the reference to q and L in $p_{\ell,L}(x;q)$. For example, when the only binding constraint is $\widehat{b}_{\ell_0} = \widehat{b}_{\ell_0+1}$, as described earlier, we have

$$\widetilde{\boldsymbol{P}}_L(x_k) =: (p_1(x),\ldots,p_{\ell_0-1}(x),\widetilde{p}_{\ell_0}(x),p_{\ell_0+2}(x),...,p_L(x))\,.$$

It is obvious that if there were no binding constraints then $\boldsymbol{P}_L(x) \equiv \widetilde{\boldsymbol{P}}_L(x)$. The use of the “correct” $\widetilde{\mathbf{P}}_L(x)$ is crucial for the power of the test. Using $\mathbf{P}_L(x)$ without taking into account the binding constraints will make the test to have only trivial power, see the discussion Section 3.3. However, for the sake of expositional simplicity, in this section we shall consider the case of no binding constraints (and, thus, $\widetilde{\mathbf{P}}_L(x) = \mathbf{P}_L(x)$).[7] With this in mind, for any $x \in \mathcal{X}$, let us define

$$C_n(x) = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{P}_k u_k \mathcal{J}_k(x)\,;\qquad A_n(x) = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{P}_k\boldsymbol{P}'_k\mathcal{J}_k(x) \tag{3.4}$$

where $\mathcal{I}(x \le x_k) =: \mathcal{J}_k(x) = 1 - \mathcal{I}_k(x)$ and using the abbreviation

$$C_{n,i} =: C_n(\widetilde{x}_i)\,;\quad A_{n,i} =: A_n(\widetilde{x}_i)\,, \tag{3.5}$$

where $\widetilde{x}_i = x_i$ if $x_i + n^{-\varsigma} < z^{k(x_i)}$ and $= z^{k(x_i)}$ otherwise, with $z^{k(x)}$ denoting the closest knot z^k, $k = 2,\ldots,L'+1$, bigger than x and $1/2 < \varsigma < 1$. The motivation to make this “trimming” is because when x_i is too close to $z^{k(x_i)}$, the *B-spline* is close but not equal to zero, which induces some technical complications in the proof of our main results. However, in small samples this “trimming” does not appear to be needed, becoming a purely technical argument.

Then the *CUSUM* of (forward) recursive least squares is defined as

$$\mathcal{M}_n(x) =: \frac{1}{n^{1/2}}\sum_{i=1}^{n} v_i\mathcal{I}_i(x) \tag{3.6}$$

[7] When examining the local power of the test in Section 3.3 we shall make explicit the consideration of binding constraints. An additional discussion of the role of the binding constraints is outlined in the supplementary document.

where

$$v_i = u_i - \boldsymbol{P}'_i A^+_{n,i} C_{n,i}. \tag{3.7}$$

Observe that because

$$\sum_{i=1}^{n} g(x_i) = \sum_{i=1}^{n} g\left(x_{(i)}\right), \tag{3.8}$$

by well known arguments, and where $x_{(i)}$ is the i-th order statistic of $\{x_i\}_{i=1}^n$, we might have written (3.6) as

$$\mathcal{M}_n(x) =: \frac{1}{n^{1/2}} \sum_{i=1}^{n} v_{(i)} \mathcal{I}_{(i)}(x),$$

where, with $\boldsymbol{P}_{(i)} =: \mathbf{P}_L\left(x_{(i)}\right)$,

$$\begin{aligned} v_{(i)} &= u_{(i)} - \boldsymbol{P}'_{(i)} A^+_{n,(i)} C_{n,(i)} \\ &= u_{(i)} - \boldsymbol{P}'_{(i)} \left(\frac{1}{n}\sum_{k=i}^{n} \boldsymbol{P}_{(k)} \boldsymbol{P}'_{(k)}\right)^+ \frac{1}{n}\sum_{k=i}^{n} \boldsymbol{P}_{(k)} u_{(k)}. \end{aligned}$$

The latter has the more familiar formulation of *CUSUM* of recursive least squares residuals when the dependent variable is now $u_{(i)}$ and the explanatory variables are $\boldsymbol{P}_{(i)}$, as proposed and formulated by [10].

Now denoting $\mathcal{K}^1_n(x) =: n^{-1/2}\sum_{i=1}^{n} u_i \mathcal{I}_i(x)$, $\mathcal{M}_n(x)$ becomes a linear transformation of $\mathcal{K}^1_n(x)$, i.e.

$$\mathcal{M}_n(x) =: n^{1/2}\left(\mathcal{T}_n \mathcal{K}^1_n\right)(x), \quad x \in (0,1),$$

where, for any function $g(x) \in D[0,1]$,

$$(\mathcal{T}_n g)(x) = g(x) - \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{P}'_i A^+_{n,i} \int_{\widetilde{x}_i}^{1} \boldsymbol{P}_L(z)\, g(dz).$$

$(\mathcal{T}_n \mathcal{K}_n)(x)$, which equals in our case $\left(\mathcal{T}_n \mathcal{K}^1_n\right)(x)$, has the interpretation of being the martingale innovation of $\mathcal{K}_n(x)$ and the transformation $(\mathcal{T}_n g)(x)$ has the limiting version $(\mathcal{T} g)(x)$, defined as

$$(\mathcal{T} g)(x) = g(x) - \int_0^x \boldsymbol{P}'_L(z)\, A^+_L(z) \left(\int_z^1 \boldsymbol{P}_L(w)\, g(dw)\right) f_X(z)\, dz, \quad x < 1.$$

where

$$A_L(x) = \int_x^1 \left(\boldsymbol{P}_L(z)\,\boldsymbol{P}_L'(z)\right) f_X(z)\,dz. \tag{3.9}$$

This type of martingale transformation was proposed by [43] in the standard goodness of fit testing problem, and later used by [62], [45] or [18].

Finally, it is worth mentioning that in (3.4) we might have employed $\mathcal{J}_k(x) = \mathcal{I}(x < x_k)$ instead of our definition $\mathcal{J}_k(x) = \mathcal{I}(x \le x_k)$. However, because by definition of *B-splines* the matrix $A_{n,i}$, and hence $A_L(x_i)$, might be singular, if we employed $\mathcal{J}_k(x) = \mathcal{I}(x < x_k)$, then it would not be guaranteed that

$$\boldsymbol{P}_i' - \boldsymbol{P}_i' A_{n,i}^+ A_{n,i} = 0.$$

On the other hand, Theorem 12.3.4 in [39] yields that the last displayed equation holds true when $\mathcal{J}_k(x) = \mathcal{I}(x \le x_k)$.

Denote $\mathcal{U}(x) =: \sigma_u \mathcal{B}(F_X(x))$, where $\mathcal{B}(z)$ is the standard Brownian motion and $F_X(x)$ the distribution function of x_i. Then,

THEOREM 1. *Under Conditions $C1 - C3$, we have that*

$$\mathcal{M}_n(x) \overset{weakly}{\Rightarrow} \mathcal{U}(x); \quad x \in [0,1].$$

Unfortunately, we do not observe u_i, so that to implement the pivotal transformation $(\mathcal{T}_n g)(x)$, we replace v_i by $\widehat{v}_i$, where $\widehat{v}_i$ is defined as v_i in (3.7) but where we replace u_i by $\widehat{u}_i$ as defined in (3.2), yielding the statistic

$$\widetilde{\mathcal{M}}_n(x) =: \frac{1}{n^{1/2}} \sum_{i=1}^n \widehat{v}_i \mathcal{I}_i(x). \tag{3.10}$$

THEOREM 2. *Assuming that H_0 holds true, under Conditions $C1 - C3$, we have that*

$$\widetilde{\mathcal{M}}_n(x) \overset{weakly}{\Rightarrow} \mathcal{U}(x); \quad x \in [0,1].$$

Denote the estimator of the variance of u_i, σ_u^2, by

$$\widehat{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^n \widehat{u}_i^2.$$

PROPOSITION 1. *Under Conditions $C1 - C3$, we have that $\widehat{\sigma}_u^2 \overset{P}{\to} \sigma_u^2$.*

We then have the following corollary.

Corollary 1. *Under H_0 and assuming Conditions $C1-C3$, for any continuous functional $g:\mathbb{R}\to\mathbb{R}^+$,*

$$g\left(\widetilde{\mathcal{M}}_n\left(x\right)/\widehat{\sigma}_u\right)\overset{d}{\to}g\left(\mathcal{U}\left(x\right)/\sigma_u\right).$$

Proof. The proof is standard using Theorem 2, Proposition 1 and the continuous mapping theorem, so it is omitted. ∎

Denoting $\tilde{n}=:n-L$ and $\widetilde{\mathcal{M}}_n\left(x^q\right)=\widetilde{\mathcal{M}}_{n,q}$, where $x^q=q/n$, standard functionals are the Kolmogorov-Smirnov, Cramér-von-Mises and Anderson-Darling tests defined respectively as

$$\begin{aligned}\mathcal{KS}_n &= \sup_{q=1,\ldots,\tilde{n}}\left|\frac{\widetilde{\mathcal{M}}_{n,q}}{\widehat{\sigma}_u}\right|\overset{d}{\to}\sup_{x\in(0,1)}\left|\mathcal{B}\left(F_X\left(x\right)\right)\right| \\ \mathcal{C}v\mathcal{M}_n &= \frac{1}{\tilde{n}}\sum_{q=1}^{\tilde{n}}\frac{\widetilde{\mathcal{M}}_{n,q}^2}{\widehat{\sigma}_u^2}\overset{d}{\to}\int_0^1\mathcal{B}^2\left(F_X\left(x\right)\right)dx,\end{aligned}\tag{3.11}$$

$$\mathcal{AD}_n=\frac{1}{\tilde{n}}\sum_{q=1}^{\tilde{n}}\frac{\widetilde{\mathcal{M}}_{n,q}^2}{\widehat{\sigma}_u^2x^q\left(1-x^q\right)}\overset{d}{\to}\int_0^1\frac{\mathcal{B}^2\left(F_X\left(x\right)\right)}{F_X\left(x\right)\left(1-F_X\left(x\right)\right)}dx.$$

3.2. ***NONLINEAR CONSTRAINTS ON*** β_ℓ.

We turn now our attention to describing the *CUSUM* of recursive residuals when some constraints describing $S_{q,L}$ may be non-linear. First, if the constrained were no binding, the pivotal transformation would be conducted in the same way as in the previous section. Thus, it suffices to discuss the case when some of the constraints are binding, i.e. some of the elements in $\widehat{b}$ are at the boundary of $S_{q,L}$.

To that end, we first describe $\widetilde{\boldsymbol{P}}_L\left(x\right)$. The main difference with the linear scenario is that the constraints described by the boundary of $S_{q,L}$ are given by implicit functions. In particular, for the type of shapes in Example 2, we have that the boundary is given by implicit functions in the form $H\left(\beta_{\ell_0-2},\beta_{\ell_0-1},\beta_{\ell_0}\right)=0$ whose explicit solutions $\beta_{\ell_0}=h\left(\beta_{\ell_0-2},\beta_{\ell_0-1}\right)$ are obtained either analytically or numerically[8]. Then, if, for instance, we have only one binding constraint, for the purpose of conducting our (asymptotic) pivotal transformation, instead of approximating $m(\cdot)$ by the linear function

[8]Please see the supplementary document for more details on the form of constraints in Example 2.

$\sum_{k=1}^{L} \beta_k p_k(x_i)$, we consider the approximation given by

$$g\left(x_i; \beta_{-\ell_0}\right) =: \sum_{k=1}^{\ell_0 - 1} \beta_k p_k(x_i) + h\left(\beta_{-\ell_0}\right) p_{\ell_0}(x_i) + \sum_{k=\ell_0+1}^{L} \beta_k p_k(x_i), \tag{3.12}$$

where $\beta_{-\ell_0} = \left(\beta_{\ell_0 - 2}, \beta_{\ell_0 - 1}\right)$.

Then $\widetilde{\boldsymbol{P}}_L(x)$ will be given by the vector of first derivatives of $g\left(x; \beta_{-\ell_0}\right)$ with respect to the parameters. That is

$$\widetilde{\boldsymbol{P}}_L(x) =: \widetilde{\boldsymbol{P}}_L\left(x; \beta_{-\ell_0}\right) = \frac{\partial}{\partial \beta_{-\ell_0}} g\left(x; \beta_{-\ell_0}\right) =: \left\{\widetilde{p}_\ell\left(x; \beta_{-\ell_0}\right)\right\}_{\ell=1; \neq \ell_0}^{L}.$$

It is easy to see that $\widetilde{p}_\ell\left(x; \beta_{-\ell_0}\right) =: p_\ell(x) + \frac{\partial h\left(\beta_{-\ell_0}\right)}{\partial \beta_\ell} p_{\ell_0}(x)$, for $\ell \neq \ell_0$. Then, the *CUSUM* of recursive residuals becomes

$$\mathcal{M}_n(x) =: \frac{1}{n^{1/2}} \sum_{i=1}^{n} \widehat{v}_i \mathcal{I}_i(x),$$

where, with $\widehat{u}_i = y_i - g\left(x_i; \widehat{\beta}_{-\ell_0}\right)$,

$$\widehat{v}_i = \widehat{u}_i - \widetilde{\mathbf{P}}_i'\left(\widehat{b}_{-\ell_0}\right) \mathcal{D}_n^{+}\left(i; \widehat{b}_{-\ell_0}\right) \sum_{k=1}^{n} \widetilde{P}_k\left(\widehat{b}_{-\ell_0}\right) \widehat{u}_k \mathcal{J}_k(\widetilde{x}_i),$$

$\widetilde{\boldsymbol{P}}_i\left(\beta_{-\ell_0}\right) =: \widetilde{\boldsymbol{P}}_L\left(x_i; \beta_{-\ell_0}\right)$, and $\mathcal{D}_n\left(x; \beta_{-\ell_0}\right) = \sum_{k=1}^{n} \widetilde{\mathbf{P}}_k\left(\beta_{-\ell_0}\right) \widetilde{\mathbf{P}}_k'\left(\beta_{-\ell_0}\right) \mathcal{J}_k(x)$, and $\mathcal{D}_n\left(i; \beta_{-\ell_0}\right) = \mathcal{D}_n\left(\widetilde{x}_i; \beta_{-\ell_0}\right)$ with $\widetilde{x}_i$ defined in the same way as in Section 3.1. Note that by employing $\widetilde{p}_\ell\left(x_i; \beta_{-\ell_0}\right)$ instead of $p_\ell(x_i)$, we have automatically incorporated our binding restriction in our pivotal transformation. As when the constraints were linear, we have the following result.

THEOREM 3. *Assuming that H_0 holds true, under Conditions $C1 - C3$, we have that*

$$\widetilde{\mathcal{M}}_n(x) \overset{weakly}{\Rightarrow} \mathcal{U}(x); \quad x \in [0, 1].$$

3.3. *POWER AND LOCAL ALTERNATIVES.*

We now discuss the power and Pitman's alternative of our tests. For that purpose, consider the alternative hypothesis

$$H_1 \equiv \quad E[y \mid x] = m(x); \quad m(\cdot) \notin \mathcal{M}_0 \tag{3.13}$$

in a set $\mathcal{X}_1 =: [a_1, a_2] \subseteq \mathcal{X}$, which is assumed to be an interval for notational simplicity. Let's denote by $\breve{m}(x)$ the best approximation in $\mathcal{M}_0$ to $m(x)$ in the $\mathcal{L}_2$-norm, and denote by $\mathcal{X}_2$ the set where $\breve{m}(x)$ belongs to the "*boundary*" of the null hypothesis. What we mean by "*boundary*" can be better described using a couple of examples. When $\mathcal{M}_0$ is the set of nondecreasing functions, the "*boundary*" function is a constant. On the other hand, if we were interested in testing for convexity, the "*boundary*" function is a straight line. It is worth mentioning that the set $\mathcal{X}_2 =: [\breve{a}_1, \breve{a}_2]$, where $\breve{m}(x)$ belongs to the "*boundary*", does not need to coincide with the set $\mathcal{X}_1$. However, $\mathcal{X}_2$ satisfies that $\mathcal{X}_2 \supseteq \mathcal{X}_1$. For instance, if $\mathcal{M}_0$ is the set of nondecreasing functions and

$$m(x) = x\mathcal{I}(x < 1/4) + (1/2 - x)\mathcal{I}(1/4 \leq x < 3/4) + (x-1)\mathcal{I}(3/4 \leq x < 1),$$

it is quite obvious that $\breve{m}(x) = 0$ in $\mathcal{X} = [0,1]$. However $\mathcal{X}_1 = (1/4, 3/4)$ but $\mathcal{X}_2 = [0,1]$.

Now, it is worth observing that we can rewrite (3.13) as

$$E[y \mid x] = m(x) =: \breve{m}(x) + \breve{m}_1(x),$$

where by construction we can take $\breve{m}_1(x) = 0$ if $x \notin \mathcal{X}_2$. In addition, to simplify some notation for any L' it is convenient to employ the approximation $\mathcal{X}_2 = [\overline{\ell}/L', \overline{L}/L']$, where indices $\overline{\ell}$ and $\overline{L}$ are chosen to guarantee the ratios $\overline{\ell}/L'$ and $\overline{L}/L'$ to be closest to $\breve{a}_1$ and $\breve{a}_2$, respectively. This will yield that $\overline{\ell}/L' \to \breve{a}_1$ and $\overline{L}/L' \to \breve{a}_2$.

When the null hypothesis is written in terms of the r-th derivative of $m(x)$, as in Example 1 with $c_r = 0$, the "*boundary*" function satisfies that $\partial^r E[y \mid x]/\partial x^r = 0$ in $\mathcal{X}_2$, i.e. "*boundary*" function becomes a polynomial of order less than or equal to $r-1$. Regarding the scenarios described in Example 2, we have that a "*boundary*" function is a solution to the first order differential equation given for some constant c by

$$\begin{aligned} &(\mathbf{i})\ \frac{x^\gamma m'(x)}{m(x)} = c \quad \text{or}\ (\mathbf{ii})\ \frac{x^\gamma m'(x)}{m^2(x)} = c, \quad \gamma = 0,1,2 \\ &\text{or}\ (\mathbf{iii})\ x^\gamma m'(x) = c, \quad \gamma = 1,2, \end{aligned} \tag{3.14}$$

depending on the exact shape property under consideration. From here, standard arguments yield that

$$\frac{m'(x)}{m(x)} = \frac{c}{x^\gamma} \Rightarrow \log m(x) = c\left\{x\mathbf{1}(\gamma=0) + \log x\mathbf{1}(\gamma=1) - \frac{1}{x}\mathbf{1}(\gamma=2)\right\} + b \tag{3.15}$$

for $(\mathbf{i})$ and some constant b, whereas for those in $(\mathbf{ii})$ we obtain that

$$\frac{m'(x)}{m^2(x)} = \frac{c}{x^\gamma} \Rightarrow \frac{1}{m(x)} = -c\left\{x\mathbf{1}(\gamma=0) + \log x\mathbf{1}(\gamma=1) - \frac{1}{x}\mathbf{1}(\gamma=2)\right\} + b \tag{3.16}$$

and for (**iii**) $m(x) = c\left\{\log x\mathbf{1}(\gamma=1) - x^{-1}\mathbf{1}(\gamma=2)\right\} + b$. Those examples illustrate how to describe the "*boundary*" functions in a general context.

To fix ideas, we shall explicitly consider the case when $\mathcal{M}_0$ is the set of nondecreasing functions, discussing more general scenarios in Remark 2 below. Suppose that our optimization problem given in (2.9) ended up with $\widehat{b}_{\ell_0} = ... = \widehat{b}_{L_0}$, so that

$$\begin{aligned} \widehat{m}(x_i) &= \left(\widehat{b}_1, \ldots, \widehat{b}_{\ell_0 - 1}, \widehat{b}_{\ell_0}, \widehat{b}_{L_0+1}, ..., \widehat{b}_L\right)\widetilde{\mathbf{P}}_i, \\ &= \sum_{k=1}^{\ell_0-1}\widehat{b}_k p_k(x) + \widehat{b}_{\ell_0}\sum_{k=\ell_0}^{L_0} p_k(x) + \sum_{k=L_0+1}^{L}\widehat{b}_k p_k(x), \end{aligned} \tag{3.17}$$

where $\widehat{u}_i = y_i - \widehat{m}(x_i)$ with $\widetilde{\mathbf{P}}_L(w)$ in (2.10) being

$$\widetilde{\mathbf{P}}_L(w) = \left(p_1(w), ..., p_{\ell_0-1}(w), \widetilde{p}_{\ell_0}(w), p_{L_0+1}(w), ..., p_L(w)\right). \tag{3.18}$$

Observe that due to the properties of the *B-splines*, we have that $\widetilde{p}_{\ell_0}(x) = \sum_{k=\ell_0}^{L_0} p_k(x)$ is equal to 1 when $x \in \left[\frac{\ell_0}{L'}, \frac{L_0-q}{L'}\right]$.

It worth observing that, as mentioned above, in this case we have that $m_{\mathcal{B}}(x;L)$ in (2.4) becomes

$$\breve{m}_{\mathcal{B}}(x;L) = \sum_{k=1}^{\overline{\ell}-1}\beta_k p_k(x) + \beta_{\overline{\ell}}\sum_{k=\overline{\ell}}^{\overline{L}} p_k(x) + \sum_{k=\overline{L}+1}^{L}\beta_k p_k(x) \tag{3.19}$$

so that

$$\dot{\mathbf{P}}_L(x) = \left(p_1(x), ..., p_{\overline{\ell}-1}(x), \dot{p}_{\overline{\ell}}(x), p_{\overline{L}+1}(x), ..., p_L(x)\right), \tag{3.20}$$

and where similarly as above, $\dot{p}_{\overline{\ell}}(x) = \sum_{k=\overline{\ell}}^{\overline{L}} p_k(x)$ is equal to 1 when $x \in \mathcal{X}_2'$ with $\mathcal{X}_2' = \left[\frac{\overline{\ell}}{L'}, \frac{\overline{L}-q}{L'}\right]$. The latter implies that we can consider ℓ_0/L' and L_0/L' as estimators of $\overline{\ell}/L'$ and $\overline{L}/L'$ respectively, which we will show in the proof of Proposition 2 below to be consistent. That is,

$$\left|\frac{\ell_0 - \overline{\ell}}{L'}\right| + \left|\frac{L_0 - \overline{L}}{L'}\right| = o_p(1). \tag{3.21}$$

Define

$$\mathcal{L}_L(x) = \int_{[0;x]\cap\mathcal{X}_2'}\left\{\breve{m}_1(v) - \dot{\mathbf{P}}_L'(v)\widetilde{A}_L^{+}(v)\int_{[v;1]\cap\mathcal{X}_2'}\dot{\mathbf{P}}_L(w)\breve{m}_1(w)f_X(w)\,dw\right\}f_X(v)\,dv,$$

which is different from zero in $\mathcal{X}_2'$. Indeed, because $f_X(x) > 0$, we have that $\mathcal{L}_L(x) =: 0$ a.e. on $\mathcal{X}_2'$ iff for v a.e. on $\mathcal{X}_2'$

$$\breve{m}_1(v) - \dot{\mathbf{P}}_L'(v)\widetilde{A}_L^+(v)\int_{[v;1]\cap\mathcal{X}_2'}\dot{\mathbf{P}}_L(w)\breve{m}_1(w)f_X(w)\,dw =: 0. \tag{3.22}$$

But the latter means that $\breve{m}_1(x)$ belongs to the space span by $\dot{\mathbf{P}}_L(w)$. However, the latter is ruled out since $\breve{m}_1(x) \notin \mathcal{M}_0$ in $\mathcal{X}_2'$ and any linear combination of $\dot{\mathbf{P}}_L(w)$ is a constant function in $\mathcal{X}_2'$ and hence belonging to $\mathcal{M}_0$. We shall remark that $\dot{\mathbf{P}}_L(w)$ depends on $\mathcal{M}_0$, via the boundary component of $\breve{m}(x)$.

PROPOSITION 2. *Assuming Conditions C1 − C3, under H_1 in (3.13), we have that*

$$\widetilde{\mathcal{M}}_n(x) - n^{1/2}\mathcal{L}_L(x) \overset{weakly}{\Rightarrow} \mathcal{U}(x) + \mathcal{V}(x), \quad x \in [0,1], \tag{3.23}$$

where $\mathcal{V}(x)$ is a non-degenerate random variable.

The first consequence of Proposition 2 is that our tests would reject H_0 with probability 1 as n increases to infinity. Indeed, this is the case as $\mathcal{L}(x)$ is a nonzero function in $\mathcal{X}_2'$, so that for any continuous functional $g : \mathcal{X} \to \mathbb{R}^+$, we have that by standard arguments,

$$1/g\left(\widetilde{\mathcal{M}}_n(x)\right) \overset{P}{\to} 0.$$

Next, we examine the Pitman's alternatives for which the test has non-trivial power. For that purpose, consider the Pitman's alternatives

$$H_a \equiv \quad E[y_i \mid x_i] =: \breve{m}(x) + n_1^{-1/2}\breve{m}_1(x),$$

where $\breve{m}(x)$ and $\breve{m}_1(x)$ satisfy respectively the same conditions as above. Then, Proposition 2 yields that

$$\widetilde{\mathcal{M}}_n(x) - \mathcal{L}_L(x) \overset{weakly}{\Rightarrow} \mathcal{U}(x) + \mathcal{V}(x), \quad x \in [0,1].$$

REMARK 1. *It is important to remark that if $\widetilde{\mathbf{P}}_L(w) = \mathbf{P}_L(w)$, then (3.22) would indeed be $o_p(1)$ regardless whether $m(x) \in \mathcal{M}_0$ or not. The consequence would be that a test based on $\widetilde{\mathcal{M}}_n(x)$ given in (3.10) would have no power. So for the test to have power it is crucial that we employ $\widetilde{\mathbf{P}}_L(w)$ when performing the pivotal transformation in (3.6) or (3.10).*[9]

[9]See some additional discussion in the supplement document.

REMARK 2. (**i**) *We have seen that when testing for monotonicity, the* "boundary" *function is a constant and the associated set of "effective" polynomials* $\dot{\mathbf{P}}_L(w)$ *was given in* (3.20)*, whereas the corresponding approximation* $\mathcal{M}_{S_{q,L}}$ *becomes* $m_{\mathcal{B}}(x;L)$ *given in* (3.19)*. Observe that when* $x \in \mathcal{X}_2' = \left[\frac{\overline{\ell}}{L'}, \frac{\overline{L}-q}{L'}\right]$ *our approximation function is a constant.*

Suppose now that $\mathcal{M}_0$ *denotes the set of convex functions. As we mentioned above, the* "boundary" *function becomes a straight line or in terms of* $\mathcal{S}_{q,L}$*, we have that the coefficients* β_ℓ*,* $\ell = \overline{\ell}, ..., \overline{L}$*, satisfy* $\beta_\ell - 2\beta_{\ell-1} + \beta_{\ell-2} = 0$ *or* $\beta_\ell = \beta_{\overline{\ell}} + (\ell - \overline{\ell})(\beta_{\overline{\ell}+1} - \beta_{\overline{\ell}})$*. Thus, after standard algebra, the restrictions on the coefficients* β_ℓ *yield that the associated set of "effective" polynomials* $\dot{\mathbf{P}}_L(w)$ *now becomes*

$$\dot{\mathbf{P}}_L(w) = \left(p_1(w), ..., p_{\overline{\ell}-1}(w), \dot{p}_{\overline{\ell}}(w), \dot{p}_{\overline{\ell}+1}(w), p_{\overline{L}+1}(w), ..., p_L(w)\right)$$

with $\dot{p}_{\overline{\ell}}(w) =: p_{\overline{\ell}}(w) - \sum_{\ell=\overline{\ell}+2}^{\overline{L}} (\ell - \overline{\ell} - 1) p_\ell(w)$ *and* $\dot{p}_{\overline{\ell}+1}(w) =: p_{\overline{\ell}+1}(w) + \sum_{\ell=\overline{\ell}+2}^{\overline{L}} (\ell - \overline{\ell}) p_\ell(w)$*. The last displayed expression yields that* $\mathcal{M}_{S_{q,L}}$*, i.e.* (3.19)*, becomes now*

$$\breve{m}_{\mathcal{B}}(x;L) = \sum_{k=1}^{\overline{\ell}-1} \beta_k p_k(x) + \beta_{\overline{\ell}} \dot{p}_{\overline{\ell}}(x) + \beta_{\overline{\ell}+1} \dot{p}_{\overline{\ell}+1}(x) + \sum_{k=\overline{L}+1}^{L} \beta_k p_k(x).$$

Notice that in this case when $x \in \mathcal{X}_2'$ *our approximation function is a straight line. The latter comes from the observation that the properties given in* (2.3) *implies that with equidistant knots when* $x \in \mathcal{X}_2'$*, we have that*

$$\begin{aligned} \frac{\partial \breve{m}_{\mathcal{B}}(x;L)}{\partial x} &= \sum_{\ell=1}^{L-1} \frac{q \triangle \beta_{\ell+1}}{z^{l+1} - z^{l+1-q}} p_{\ell+1,L}(x;q-1) \\ &= L' (\triangle \beta_{\ell+1}) \sum_{\ell=1}^{L-1} p_{\ell+1,L}(x;q-1) \end{aligned}$$

because $\triangle \beta_{\ell+1}$ *is a constant when* $\overline{\ell} \le \ell \le \overline{L}$*. From here we conclude because* $\sum_{\ell=1}^{L-1} p_{\ell+1,L}(x;q-1) = 1$ *when* $x \in \mathcal{X}_2'$*. Of course, as neither* $\overline{\ell}$ *nor* $\overline{L}$ *are known, the values would be replaced by* ℓ_0 *and* L_0 *satisfying* (3.21)*.*

General cases in Example 1 *can be handled similarly, that is testing of the sign of the* r*-th derivative. This follows from the observation that when we impose the null hypothesis, the approximation* $m_{\mathcal{B}}(x;L)$ *of* $m(x)$*, that is*

$$\mathcal{M}_{s_{q,L}} =: \{m_{\mathcal{B}}(\cdot;L) \mid \beta = (\beta_1, \ldots, \beta_L) \in S_{q,L}\}, \tag{3.24}$$

becomes a polynomial of degree less than or equal to $r-1$ *in* $\mathcal{X}_2$*, which is the boundary function and where the coefficients* β_ℓ*,* $\ell = \overline{\ell}+1, ..., \overline{L}+1$*, satisfy* $\triangle^r \beta_{\ell+r} = 0$*.*

(ii) *When the null hypothesis is nonlinear in parameters, as those cases given in Example 2, we make use of the fact that a "*boundary*" function takes the form of one of the solutions given in* (3.15) *or* (3.16) *say. For instance, when testing for* log-convexity*, the "*boundary*" function becomes* $\Delta(w) = \exp(b+aw)$*, and our approximation* $m_{\mathcal{B}}(x;L)$ *in* (3.24) *becomes*

$$\breve{m}_{\mathcal{B}}(x;L) = \sum_{k=1}^{\overline{\ell}-1} \beta_k p_k(x) + \exp(b+ax)\,\mathcal{I}\left(x \in \mathcal{X}_2'\right) + \sum_{k=\overline{L}-q}^{L} \beta_k p_k(x),$$

whereas now the associated set of "effective polynomials" $\dot{\mathbf{P}}_L(w)$ *would be*

$$\begin{aligned} \dot{\mathbf{P}}_L(x) &= \Big(p_1(x), ..., p_{\overline{\ell}-1}(x), \dot{p}_{\overline{\ell}}(x), p_{\overline{L}-q}(x), ..., p_L(x)\Big), \\ \dot{p}_{\overline{\ell}}(x) &= :\exp(b+ax)(1,x)\,\mathcal{I}\left(x \in \mathcal{X}_2'\right) \end{aligned}$$

since $\dot{p}_{\overline{\ell}}(x) = \partial \exp(b+ax)/\partial(b,a)$*. Of course, neither* $\overline{\ell}$ *nor* $\overline{L}$ *are known, so as we argued in the proof of Proposition 2, they will be replaced by preliminary "estimates"* ℓ_0 *and* L_0*. We can see that this approach is a natural extension when the interest was, say, to test for monotonicity. Indeed, in this case we have that* (3.19) *can be written as*

$$\breve{m}_{\mathcal{B}}(x;L) = \sum_{k=1}^{\overline{\ell}-1} \beta_k p_k(x) + \beta_{\overline{\ell}}\mathcal{I}\left(x \in \mathcal{X}_2'\right) + \sum_{k=\overline{L}-q}^{L} \beta_k p_k(x)$$

with $\beta_k = \beta_{\overline{\ell}}$ *for* $\ell = \overline{L}-q, ..., \overline{L}$ *and because* $\sum_{k=\overline{\ell}}^{\overline{L}} p_k(x)$ *is equal to* 1 *when* $x \in \mathcal{X}_2' = \left[\frac{\overline{\ell}}{L'}, \frac{\overline{L}-q}{L'}\right]$*.*

4. BOOTSTRAP ALGORITHM.

One of our motivations to introduce a bootstrap algorithm for our test(s) is that although it is pivotal, our Monte Carlo experiment suggests that they suffer from small sample biases. When the asymptotic distribution does not provide a good approximation to the finite sample one, a standard approach to improve its performance is to employ bootstrap algorithms, as they provide small sample refinements. In fact, our Monte Carlo simulation does suggest that the bootstrap, to be described below, does indeed give a better finite sample approximation. The notation for the bootstrap is as usual and we shall implement the fast algorithm of WARP by [30] in the Monte Carlo experiment.

The bootstrap is based on the following *3 STEPS*.

STEP 1 Compute the unconstrained residuals

$$\widetilde{u}_i = y_i - \widetilde{m}_{\mathcal{B}}\left(x_i; L\right), \quad i = 1, ..., n$$

with $\widetilde{m}_{\mathcal{B}}\left(x_i; L\right)$ as defined in (2.11).

STEP 2 Obtain a random sample of size n from the empirical distribution of $\left\{\widetilde{u}_i - \frac{1}{n}\sum_{i=1}^n \widetilde{u}_i\right\}_{i=1}^n$. Denote such a sample as $\{u_i^*\}_{i=1}^n$ and compute the bootstrap analogue of the regression model using $\widehat{m}_{\mathcal{B}}\left(x_i; L\right)$, that is

$$y_i^* = \widehat{m}_{\mathcal{B}}\left(x_i; L\right) + u_i^*, \ \ i = 1, ..., n. \tag{4.1}$$

STEP 3 Compute the bootstrap analogue of $\widetilde{\mathcal{M}}_n\left(x\right)$ as

$$\widetilde{\mathcal{M}}_n^*\left(x\right) =: \frac{1}{n^{1/2}}\sum_{i=1}^n \widehat{v}_i^* \mathcal{I}_i\left(x\right)$$

where

$$\widehat{v}_i^* = \widehat{u}_i^* - \boldsymbol{P}'_i A_{n,i}^+ C_{n,i}^*; \quad C_{n,i}^* =: C_{n,i}^*\left(\widetilde{x}_i\right) = \frac{1}{n}\sum_{k=1}^n \boldsymbol{P}_k \widehat{u}_k^* \mathcal{J}_k\left(\widetilde{x}_i\right)$$

with $\widehat{u}_i^* = y_i^* - \boldsymbol{P}'_i A_n^+\left(0\right) C_n^*\left(0\right)$, $i = 1, ..., n$.

THEOREM 4. *Under Conditions $C1 - C3$, we have that for any continuous function $g : \mathbb{R} \to \mathbb{R}^+$, (in probability),*

$$g\left(\widetilde{\mathcal{M}}_n^*\left(x\right)\right) \overset{d}{\Rightarrow} g\left(\mathcal{U}\left(x\right)\right).$$

Finally, we can replace $\widehat{u}_i^*$ by y_i^* in the computation of $\widetilde{\mathcal{M}}_n^*\left(x\right)$. That is,

COROLLARY 2. *Under Conditions $C1 - C3$, we have that*

$$\widetilde{\mathcal{M}}_n^*\left(x\right) - \widetilde{\widetilde{\mathcal{M}}}_n^*\left(x\right) = 0,$$

where

$$\widetilde{\widetilde{\mathcal{M}}}_n^*\left(x\right) =: \frac{1}{n^{1/2}}\sum_{i=1}^n \left(y_i^* - \boldsymbol{P}'_i A_{n,i}^+ \frac{1}{n}\sum_{k=1}^n \boldsymbol{P}_k y_k^* \mathcal{J}_k\left(x_i\right)\right)\mathcal{I}_i\left(x\right).$$

The proof of Corollary 2 is immediate by Lemma 1 in the supplementary document and therefore is omitted.

5. MONTE CARLO EXPERIMENTS AND EMPIRICAL EXAMPLES.

5.1. *MONTE CARLO EXPERIMENTS*.

In this section we present the results of several computational experiments. All the results in this section are given for cubic splines with different number of knots. We present the results for *B-splines* as well as for *P-splines* with penalties on the second differences of coefficients. The penalty parameter is chosen by cross-validation in the unconstrained estimation described in [27]. In the tables, "*KS*" refers to the Kolmogorov-Smirnov test statistic, "*CvM*" refers to the Cramér-von Mises test statistic and "*AD*" to the Anderson-Darling integral test statistic. All three test statistics are based on a Brownian bridge. $L'+1$ denotes the number of equidistant knots (including the boundary points) on the interval of interest. For example, when $L'=6$ and the interval is $[0,1]$, we consider knots $0, 1/6, 1/3, 1/2, 2/3, 5/6, 1$. In the implementation of *P-splines* in simulations, every simulation draw will give a different cross-validation parameter. In our simulation results for each L' we use a modal value of these cross-validation parameters.

In all the scenarios below

$$X \sim \mathcal{U}[0,1], \quad U \sim \mathcal{N}(0,\sigma^2), \quad U \perp X.$$

In Scenarios 1, 3-5 the interval of interest is $[0,1]$ whereas in Scenario 2 of U-shape we consider individually intervals $[0, s_0]$ and $[s_0, 1]$ with s_0 being the switch point.

In the *WARP* bootstrap implementation, the demeaned residuals and x are drawn independently. Rejection rates are for 2000 simulations. In all the tables with Monte Carlo testing results N denotes the number of observations in each simulation and σ stands for the standard deviation in the error distribution.

Even though *B-splines* and *P-splines* deliver asymptotically equivalent results, the evidence from Monte Carlo experiments in Scenarios 3 and 4 suggests that in a finite sample the use of *P-splines* gives a better power of the test and also leads to a more stable power across different L (equivalently, L'). For this reason, we are inclined to recommend using *P-splines* in practice.

The supplementary document contains additional results. In particular, it shows the performance of our test using asymptotic critical values. The results support our proposal to use bootstrap critical values in practice. The supplement also gives testing results for sample sizes $N=100$ and $N=200$. In addition, in the supplement we compare our test to those in [35], [31]

Setting	Method	B-splines		P-splines	
		10%	5%	10%	5%
$L' = 6$	KS bootstrap	0.113	0.054	0.101	0.05
$N = 1000$	CvM bootstrap	0.1035	0.0475	0.1005	0.044
$\sigma = 0.25$	AD bootstrap	0.1065	0.054	0.104	0.052
$L' = 9$	KS bootstrap	0.102	0.044	0.119	0.052
$N = 1000$	CvM bootstrap	0.101	0.048	0.11	0.0455
$\sigma = 0.25$	AD bootstrap	0.0985	0.042	0.1045	0.048
$L' = 14$	KS bootstrap	0.0945	0.043	0.105	0.0555
$N = 1000$	CvM bootstrap	0.098	0.0425	0.0955	0.045
$\sigma = 0.25$	AD bootstrap	0.093	0.043	0.096	0.049
$L' = 19$	KS bootstrap	0.089	0.0485	0.101	0.058
$N = 1000$	CvM bootstrap	0.105	0.0555	0.1025	0.0495
$\sigma = 0.25$	AD bootstrap	0.1085	0.0545	0.1065	0.049

TABLE 1
Tests for monotonically increasing regression function in Scenario 1a.

and [63] when testing for monotonicity. Regarding the power of the test in Scenarios 3 and 4, we find that when using *P-splines*, our test has a superior performance to them for small noise to signal ratios and performs at least as well as these alternative tests for small noise to signal ratios (when power is very close to 1). In particular, this further supports our recommendation of using *P-splines* in practice.

Scenario 1 (test for monotonicity). We consider the following regression functions defined on $[0, 1]$:

$$m\left(x\right) = x^{\frac{13}{4}}, \quad \text{(Scenario 1a)}$$
$$m\left(x\right) = -(x-0.5)^2 \cdot 1(x < 0.5) + (x-0.5)^2 \cdot 1(x \geq 0.5), \quad \text{(Scenario 1b)}$$

Functions in Scenarios 1a and 1b have different degrees of smoothness. In Scenario 1a, the function is twice continuously differentiable and its second derivative is Hölder continuous with the exponent $\frac{1}{4}$ whereas in Scenario 1b the function is smooth and its first derivative is Lipschitz. The results are summarized in Tables 1 and 2.

Since it may be of interest to explore the cases of various regularities of $m(\dot{})$, in the supplement we consider two additional Scenarios 1c and 1d. In Scenario 1c, $m(\cdot)$ is smooth but its derivative is not Hölder continuous. In Scenario 1d, $m(\cdot)$ is infinitely differentiable.

Scenario 2 (test for U-shape). The regression function is defined as

$$m(x) = 10\left(\log(1+x) - 0.33\right)^2.$$

Setting	Method	B-splines		P-splines	
		10%	5%	10%	5%
$L' = 6$	KS bootstrap	0.097	0.0525	0.1065	0.053
$N = 1000$	CvM bootstrap	0.1005	0.0515	0.0975	0.0445
$\sigma = 0.25$	AD bootstrap	0.0975	0.058	0.096	0.045
$L' = 9$	KS bootstrap	0.1055	0.0515	0.096	0.0425
$N = 1000$	CvM bootstrap	0.099	0.0505	0.0995	0.041
$\sigma = 0.25$	AD bootstrap	0.094	0.0445	0.099	0.0465
$L' = 14$	KS bootstrap	0.0925	0.048	0.092	0.044
$N = 1000$	CvM bootstrap	0.0885	0.046	0.0925	0.049
$\sigma = 0.25$	AD bootstrap	0.0875	0.0465	0.0965	0.0425
$L' = 19$	KS bootstrap	0.093	0.04	0.098	0.042
$N = 1000$	CvM bootstrap	0.0915	0.049	0.098	0.0475
$\sigma = 0.25$	AD bootstrap	0.0885	0.043	0.1005	0.0465

TABLE 2
Tests for monotonically increasing regression function in Scenario 1b.

		Continuously joined				Smoothly joined			
Setting	Method	B-splines		P-splines		B-splines		P-splines	
		10%	5%	10%	5%	10%	5%	10%	5%
$L' = 4$	KS	0.0935	0.048	0.113	0.051	0.098	0.062	0.1105	0.055
$N = 1000$	CvM	0.106	0.046	0.1	0.0515	0.101	0.0575	0.1005	0.05
$\sigma = 0.25$	AD	0.107	0.048	0.098	0.055	0.1015	0.0615	0.094	0.0505
$L' = 6$	KS	0.1105	0.0555	0.112	0.051	0.107	0.0575	0.0935	0.0495
$N = 1000$	CvM	0.101	0.0585	0.099	0.0525	0.1	0.0575	0.0955	0.048
$\sigma = 0.25$	AD	0.1015	0.0565	0.098	0.047	0.1055	0.055	0.0965	0.0485

TABLE 3
Tests for U-shape with the switch at $s_0 = e^{0.33} - 1$ in Scenario 2. $L' + 1$ denotes the number of equidistant knots on each subinterval $[0, s_0]$ and $[s_0, 1]$.

The graph of this function is U-shaped with the switch point at $s_0 = e^{0.33} - 1$. In simulations s_0 is taken to be known.

The results are summarized in Table 3. We use two different *B-splines* – one on $[0, s_0]$ and the other on $[s_0, 1]$. We analyze the properties of the testing procedure in two approaches. In the first approach additional restrictions are imposed for the two *B-splines* to be joined continuously at s_0, and in the second approach these two *B-splines* are joined smoothly at s_0 (see details in Example 2).

Scenario 3 (analysis of power of the test). Take the regression function

$$\begin{aligned} m(x) &= 10(x - 0.5)^3 - \exp(-100(x - 0.25)^2)) \cdot \mathcal{I}(x < 0.5) \\ &\quad + (0.1(x - 0.5) - \exp(-100(x - 0.25)^2)) \cdot \mathcal{I}(x >= 0.5). \end{aligned}$$

and depicted in Figure 1. As expected, the power of the test depends on the variance of the error. The results are summarized in Table 4.

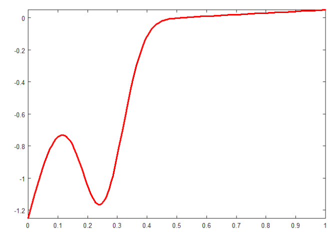

FIG 1. *Plot of the regression function in Scenario 3.*

The power of monotonicity tests based on this regression function is considered in [31] and a similar regression function is considered in [35]. Note that [31] considers smaller sample sizes and also smaller standard deviation of noise with $\sigma = 0.1$.

Scenario 4 (analysis of power of the test). The regression function

$$m(x) = x + 0.415 \exp(-ax^2), \quad a > 0.$$

and depicted in Figure 2. The left-hand side graph in Figure 2 is for the case $a = 50$ and the right-hand side graph in Figure 2. In the latter case the non-monotonicity dip is smaller. These situations are considered to be challenging for monotonicity tests as these functions are somewhat close to the set of monotone functions (in any conventional metric). As expected, the power of the test depends on the value of parameter a and also depends on the variance of the error. The results are summarized in Table 5.

The power of monotonicity tests based on this regression function is examined in [31] and a similar regression function was considered in [9]. Note that [31] uses smaller sample sizes and also only $a = 50$ and $\sigma = 0.1$ to analyze power implications.

Scenario 5 (test for log-convexity). We take the following regression function:

$$m(x) = \exp(x^2), \quad x \in [0, 1].$$

The results are summarized in Table 6. In this case, the results for *P-splines* are the same as for *B-splines* as the cross-validation criterion indicated 0 as the optimal penalty parameter in the overwhelming majority of simulations.

Setting	Method	B-splines		P-splines	
		10%	5%	10%	5%
$L' = 6$	KS bootstrap	1	0.998	1	0.9885
$N = 1000$	CvM bootstrap	0.9155	0.7335	0.9975	0.9895
$\sigma = 0.5$	AD bootstrap	0.985	0.9375	0.9995	0.998
$L' = 9$	KS bootstrap	0.93	0.823	0.999	0.998
$N = 1000$	CvM bootstrap	0.862	0.766	0.998	0.9935
$\sigma = 0.5$	AD bootstrap	0.9195	0.8085	0.9985	0.9935
$L' = 12$	KS bootstrap	0.864	0.8175	0.996	0.9885
$N = 1000$	CvM bootstrap	0.8505	0.7895	0.989	0.9725
$\sigma = 0.5$	AD bootstrap	0.8655	0.799	0.9885	0.974
$L' = 19$	KS bootstrap	0.639	0.5375	0.9815	0.9515
$N = 1000$	CvM bootstrap	0.5295	0.395	0.9435	0.8795
$\sigma = 0.5$	AD bootstrap	0.571	0.428	0.9445	0.892
$L' = 6$	KS bootstrap	1	1	1	1
$N = 1000$	CvM bootstrap	1	1	1	1
$\sigma = 0.25$	AD bootstrap	1	1	1	1
$L' = 9$	KS bootstrap	1	1	1	1
$N = 1000$	CvM bootstrap	1	1	1	1
$\sigma = 0.25$	AD bootstrap	1	1	1	1
$L' = 12$	KS bootstrap	0.9995	0.9995	1	1
$N = 1000$	CvM bootstrap	0.996	0.9945	1	1
$\sigma = 0.25$	AD bootstrap	1	0.996	1	1
$L' = 19$	KS bootstrap	0.996	0.986	1	1
$N = 1000$	CvM bootstrap	0.9155	0.836	1	1
$\sigma = 0.25$	AD bootstrap	0.955	0.891	1	1

TABLE 4
Tests for monotonicity in Scenario 3.

		$a = 50$				$a = 20$			
Setting	Method	B-splines		P-splines		B-splines		P-splines	
		10%	5%	10%	5%	10%	5%	10%	5%
$L' = 6$	KS	0.382	0.2425	0.575	0.423	0.25	0.139	0.338	0.216
$N = 1000$	CvM	0.3385	0.2245	0.5165	0.384	0.2465	0.1335	0.311	0.212
AD $\sigma = 0.5$	AD	0.434	0.228	0.6015	0.457	0.2475	0.1395	0.3355	0.221
$L' = 9$	KS	0.381	0.2545	0.572	0.4655	0.2	0.124	0.3405	0.221
$N = 1000$	CvM	0.363	0.2355	0.536	0.3915	0.2055	0.132	0.3385	0.206
$\sigma = 0.5$	AD	0.4735	0.3165	0.6095	0.4555	0.228	0.145	0.348	0.206
$L' = 12$	KS	0.3975	0.27	0.5995	0.4745	0.22	0.1375	0.343	0.2185
$N = 1000$	CvM	0.3905	0.2565	0.5545	0.4075	0.2335	0.152	0.3215	0.2065
$\sigma = 0.5$	AD	0.486	0.3505	0.614	0.5035	0.2525	0.1625	0.3415	0.2195
$L' = 19$	KS	0.4045	0.284	0.5905	0.476	0.2185	0.1235	0.3455	0.2265
$N = 1000$	CvM	0.418	0.308	0.5525	0.414	0.232	0.145	0.318	0.1995
$\sigma = 0.5$	AD	0.497	0.37	0.6125	0.4915	0.2485	0.155	0.341	0.2205
$L' = 6$	KS	0.9	0.8405	0.986	0.9625	0.5795	0.4605	0.756	0.6415
$N = 1000$	CvM	0.8295	0.7025	0.9615	0.913	0.5895	0.4195	0.741	0.626
$\sigma = 0.25$	AD	0.939	0.854	0.9835	0.9665	0.608	0.4505	0.7275	0.6295
$L' = 9$	KS	0.919	0.8385	0.986	0.9685	0.484	0.3355	0.6805	0.5645
$N = 1000$	CvM	0.8355	0.7275	0.966	0.911	0.46	0.347	0.6495	0.5065
$\sigma = 0.25$	AD	0.937	0.8695	0.99	0.9615	0.4995	0.374	0.6655	0.512
$L' = 12$	KS	0.9235	0.842	0.9865	0.9705	0.461	0.334	0.6785	0.5585
$N = 1000$	CvM	0.863	0.756	0.97	0.9325	0.436	0.318	0.6525	0.4965
$\sigma = 0.25$	AD	0.951	0.8895	0.9865	0.9705	0.492	0.3575	0.658	0.517
$L' = 19$	KS	0.925	0.8505	0.9865	0.974	0.4835	0.3505	0.698	0.585
$N = 1000$	CvM	0.8835	0.802	0.9745	0.9355	0.4595	0.345	0.6625	0.495
$\sigma = 0.25$	AD	0.941	0.8985	0.987	0.9695	0.486	0.3665	0.666	0.5105
$L' = 6$	KS	1	1	1	1	1	0.998	1	1
$N = 1000$	CvM	1	1	1	1	0.9985	0.9965	1	1
$\sigma = 0.1$	AD	1	1	1	1	0.9995	0.9975	1	1
$L' = 9$	KS	1	1	1	1	0.9935	0.9875	1	1
$N = 1000$	CvM	1	1	1	1	0.9915	0.9825	1	0.9955
$\sigma = 0.1$	AD	1	1	1	1	0.9915	0.9865	1	0.997
$L' = 12$	KS	1	1	1	1	0.968	0.9535	0.9995	0.998
$N = 1000$	CvM	1	1	1	1	0.9539	0.932	0.9975	0.995
$\sigma = 0.1$	AD	1	1	1	1	0.969	0.951	0.998	0.997
$L' = 19$	KS	1	1	1	1	0.973	0.9515	0.9995	0.9995
$N = 1000$	CvM	1	1	1	1	0.9525	0.9285	0.9985	0.9955
$\sigma = 0.1$	AD	1	1	1	1	0.966	0..948	0.999	0.997

TABLE 5

Tests for monotonicity in Scenario 4.

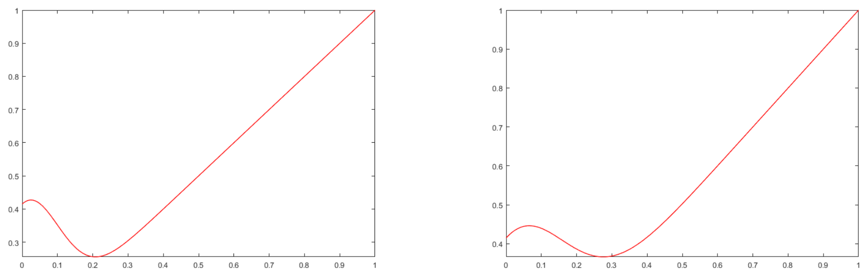

FIG 2. *Plot of the regression function in Scenario 4. The left-hand side graph is for $a = 50$ and the right-hand side graph is for $a = 20$.*

Setting	Method	B-splines	
		10%	5%
$L' = 6$	KS bootstrap	0.1045	0.0535
$N = 1000$	CvM bootstrap	0.1015	0.054
$\sigma = 0.25$	AD bootstrap	0.1015	0.0495
$L' = 9$	KS bootstrap	0.098	0.0435
$N = 1000$	CvM bootstrap	0.1	0.0445
$\sigma = 0.25$	AD bootstrap	0.1	0.048
$L' = 12$	KS bootstrap	0.1135	0.053
$N = 1000$	CvM bootstrap	0.0935	0.0465
$\sigma = 0.25$	AD bootstrap	0.095	0.0495

TABLE 6
Tests for log-convexity in Scenario 5.

5.2. ***APPLICATIONS.***

1. **US presidential elections** Here we use data on the US 2016 presidential elections across different counties. We consider counties that have both urban and rural populations and analyze the effect of rural population on votes received by Donald J. Trump.

Figure 3 is a scatter plot of the percentage of the rural population and the share of votes received by Donald J. Trump with the fitted curve obtained using cubic *B-splines* with $L'+1 = 13$ uniform knots in the range of values of the percentage of the rural population (the minimum value of the percentage is $3.13 \cdot 10^{-4}$ and the maximum value is). The fitted curve is obtained under the monotonicity restriction.

We conduct the tests for a) monotonicity and b) monotonicity and concavity simultaneously. In order to correct for heteroscedasticity of the errors, we estimate the scedastic function $\widehat{\sigma}^2(x)$ using residuals obtained in the unconstrained estimation using cubic *B-splines* with the same set of knots.

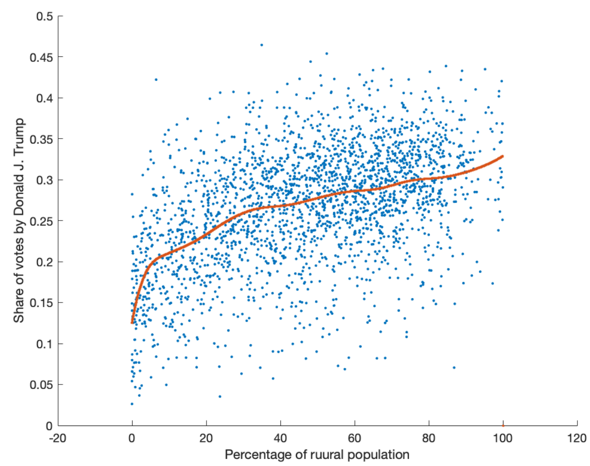

Fig 3. *US presidential election data for $N = 2,397$ counties that have both urban and rural population. Plot of the percentage of the rural population and share of votes received by Donald J. Trump and the constrained (under monotonicity) fit by cubic* B-splines *with $L' + 1 = 13$ uniform knots in the domain of the percentage of the rural population.*

The scedastic function $\widehat{\sigma}^2(x)$ is estimated by regressing the logarithm of the squared unconstrained residuals on a linear combination of first-order *B-splines* with $L' + 1$ uniform knots in the domain of the percentage of the rural population.

We then consider the constrained residuals divided by $\widehat{\sigma}(x)$ when calculating *KS*, *CvM* and *AD* test statistics and unconstrained residuals divided by $\widehat{\sigma}(x)$ when drawing bootstrap samples. After a bootstrap sample of residuals is drawn, we multiply each residual by the corresponding $\widehat{\sigma}(x)$ when generating a bootstrap sample of observations of the dependent variable.

We implement the testing procedure by conducting the pivotal *CUSUM* transformation from the left end of the support (in the theoretical description throughout the paper we implemented it from the right end of the support) as based on the visual analysis the violations of monotonicity or concavity are more likely to happen at the right end. In the case of *P-splines*, we use the same *B-spline* basis, take the second-order penalty and choose the penalization constant using the ordinary cross-validation criterion as in Eilers and Marx (1996). The penalty enters unconstrained optimization problems as well as constrained ones.

Tables 7-8 present results of our testing by showing the test statistics

		B-splines			P-splines		
Setting	Method	Test statistic	Bootstrap c.v.		Test statistic	Bootstrap c.v.	
			10%	5%		10%	5%
$L' = 12$	KS	0.7350	1.1484	1.2673	0.8854	0.7991	0.8878
	CvM	0.0456	0.2848	0.3782	0.1596	0.1153	0.1515
	AD	0.3170	1.4965	2.0232	0.9466	0.7556	0.9932
$L' = 18$	KS	0.7970	1.0856	1.2018	0.8072	0.7710	0.8529
	CvM	0.0615	0.2586	0.3479	0.1285	0.1060	0.1405
	AD	0.3919	1.4242	1.7918	0.7605	0.7060	0.9298
$L' = 24$	KS	0.8210	1.1213	1.2619	0.7177	0.7551	0.8413
	CvM	0.1474	0.2847	0.3858	0.1102	0.1012	0.1331
	AD	0.7839	1.5422	2.0622	0.6630	0.683	0.9107

TABLE 7
US presidential elections data. Test statistics and bootstrap critical values under the null hypothesis of monotonicity of the regression function. Bootstrap critical values are from 1000 bootstrap replications.

and also bootstrap critical values using both *B-splines* and *P-splines* for several L'. As we can see from Table 7, we do not reject monotonicity at the 5% level even though testing using *P-splines* supports monotonicity less confidently. As for the test for monotonicity and concavity together, even though the approach with *B-splines* does not reject it at the 5% level the approach with *P-splines* does. Even though asymptotically *B-splines* and *P-splines* deliver equivalent results, as discussed previously, in a finite sample *P-splines* deliver a better power of the test (as well as lead to the power that is more stable with regard to the choice of L). We therefore rely on the conclusion delivered by *P-splines* and, thus, reject that the regression function is both monotone and concave at the 5% level.

2. **Energy consumption in the Southern region of Russia.**

The data are on daily energy consumption (in MWh) and average daily temperature (in Celsius) in the Southern region of Russia in the period from February 1, 2016 till January 31, 2018. The data have been downloaded from the official website of System Operator of the Unified Energy System of Russia.[10]

We provide tests for U-shape with a switch at 17.6° and convexity using the approaches outlined in the previous section. In order to correct for heteroscedasticity of the errors, we estimate the scedastic function $\widehat{\sigma}^2(x)$ using residuals obtained in the unconstrained estimation using *B-splines* (or *P-splines*, respectively). The scedastic function is estimated using cubic

[10]http://so-ups.ru/

Setting	Method	B-splines Test statistic	B-splines Bootstrap c.v. 10%	B-splines Bootstrap c.v. 5%	P-splines Test statistic	P-splines Bootstrap c.v. 10%	P-splines Bootstrap c.v. 5%
$L' = 12$	KS	0.7496	1.0721	1.1815	1.2258	0.9341	1.02
	CvM	0.0609	0.2385	0.3094	0.3853	0.1676	0.2167
	AD	0.7695	1.3467	1.6553	2.1236	1.0527	1.3136
$L' = 18$	KS	0.8901	1.1475	1.2652	1.089	0.9194	0.9846
	CvM	0.1630	0.2161	0.2915	0.2789	0.1597	0.1967
	AD	1.3998	1.2201	1.5434	1.661	1.0078	1.2151
$L' = 24$	KS	0.9658	1.0483	1.1524	1.0376	0.9217	0.9869
	CvM	0.1687	0.2245	0.2982	0.2911	0.1521	0.1935
	AD	1.2835	1.3051	1.6117	1.7665	0.9532	1.1633

TABLE 8
US presidential elections data. Test statistics and bootstrap critical values under the null hypothesis of monotonicity and concavity of the regression function. Bootstrap critical values are from 1000 bootstrap replications.

B-splines with 6 uniform knots and in the form of

$$\sigma^2(x) = \left(\sum_{k=1}^{8} c_k p_k\left(x; 8\right) \right)^2 .$$

Figure 4 gives scatter plots of the data together with fitted curves obtained under the *U-shape* constraint with the switch at $s_0 = 17.6°$. This constraint fit is obtained in accordance with the technique in the previous section. Namely, we consider individual *B-spline* fits on intervals $[\underline{x}, s_0]$ and $[s_0, \overline{x}]$, where $\underline{x}$ and $\overline{x}$ are respectively lowest and highest values of the temperature in the sample. On each interval we use $L' + 1 = 5$ uniform knots. The left-hand side figure only imposes the continuity of the fitted curve at the switch point, whereas the right-hand side figure imposes continuous differentiability.

Tables 9-11 present results of our testing. Namely, Table 9 shows test statistics for the null hypothesis of *U-shaped* regression function and also bootstrap critical values using both *B-splines* and *P-splines* in case when two *B-spline* curves are joined at the switch point in a continuous way. Table 10 presents analogous results for the null hypothesis of *U-shaped* regression function when two *B-spline* curves are joined at the switch point in a continuously differentiable way. Table 11 gives results for the null hypothesis of convexity. In all the cases our pivotal transformation is conducted from the right end of the support. The bootstrap critical values are obtained on the basis of 400 bootstrap replications. As we can see, the null hypothesis of a *U-shaped* relationship with the switch point at 17.6° is not rejected at the 5% level by any type of the test, whereas convexity is rejected.

Method	B-splines Test statistic	Bootstrap critical value 10%	5%	1%	P-splines Test statistic	Bootstrap critical value 10%	5%	1%
KS	0.6204	1.0997	1.2139	1.4053	0.2482	0.5898	0.6199	0.7207
CvM	0.0698	0.3098	0.4048	0.5885	0.0038	0.0434	0.0498	0.0621
AD	0.5119	1.6541	2.1098	2.925	0.0647	0.397	0.4312	0.5324

TABLE 9

Energy consumption data. Test statistics and bootstrap critical values under the null hypothesis of U-shaped regression function with the switch as 17.6°. *Two B-spline curves are joined continuously at the switch point. Bootstrap critical values are from 400 bootstrap replications.*

Method	B-splines Test statistic	Bootstrap critical value 10%	5%	1%	P-splines Test statistic	Bootstrap critical value 10%	5%	1%
KS	0.8481	1.023	1.1749	1.407	0.505	0.6143	0.6524	0.7747
CvM	0.1472	0.2211	0.3247	0.5553	0.0469	0.0476	0.0542	0.0752
AD	0.9114	1.3676	1.8002	2.9199	0.2878	0.3713	0.4151	0.5547

TABLE 10

Energy consumption data. Test statistics and bootstrap critical values under the null hypothesis of U-shaped regression function with the switch as 17.6°. *Two B-spline curves are joined smoothly at the switch point. Bootstrap critical values are from 400 bootstrap replications.*

Method	B-splines Test statistic	Bootstrap critical value 10%	5%	1%	P-splines Test statistic	Bootstrap critical value 10%	5%	1%
KS	2.9812	1.1537	1.2652	1.5579	3.4626	0.5478	0.6891	1.0121
CvM	2.7713	0.3327	0.4389	0.71	3.2622	0.0402	0.0716	0.2076
AD	14.361	1.823	2.3809	3.856	17.23	0.3332	0.484	1.2158

TABLE 11

Energy consumption data. Test statistics and bootstrap critical values under the null hypothesis of convexity of the regression function. Bootstrap critical values are from 400 bootstrap replications.

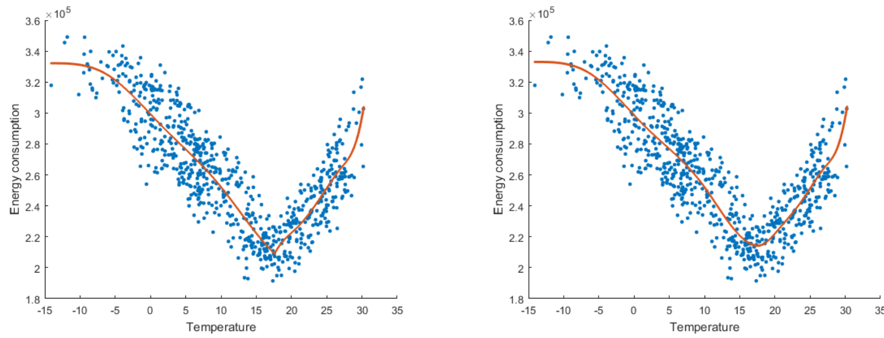

Fig 4. *Energy consumption data. Plot of temperature and energy consumption and the constrained fit (under U-shape with the switch at* $17.6°$*) using cubic B-spline with 5 uniform knots on each subinterval of temperature values. On the left-hand side the fitted curve is continuous at the switch point. On the right-hand side the fitted curve is continuously differentiable at the switch point.*

6. CONCLUSION.

This paper proposes a methodology for testing a wide range of shape properties of a regression function. The methodology relies on applying a pivotal transformation to the partial sums empirical process in a nonparametric setting where *B-splines* or *P-splines* have been used to approximate the functional space under the null hypothesis. We establish that the proposed pivotal transformation eliminates the effect of nonparametric estimation and results in asymptotically pivotal testing. To the best of our knowledge, this paper is the first implementation of the pivotal transformation in a nonparametric setting.

In our main examples we considered shape constraints that can be written as inequality constraints on the coefficients of the approximating regression splines. The generality of our procedure allows to test several shape properties simultaneously. The implementation is especially easy when the inequality constraints are linear, which is the case for shape properties expressed as linear inequality constraints on the derivatives.

References.

[1] Abrevaya, J. and Jiang, W. (2005). A Nonparametric Approach to Measuring and Testing Curvature, *Journal of Business & Economic Statistics*, **23**, 1-19.

[2] Agarwal, G.G. and Studden, W.J. (1980). Asymptotic Integrated Mean Square Error Using Least Squares and Bias Minimizing Splines, *The Annals of Statistics*, **6**, 1307-1325.

[3] Akakpo, N., Balabdaoui F. and C. Durot (2014). Testing monotonicity via local least concave majorants, *Bernoulli*, **20**, 514-544.

[4] Anderson, G.D., Vamanamurthy, M.K. and Vuorinen, M. (2007). Generalized convexity and inequalities, *Journal of Mathematical Analysis and Applications*, **335**, 1294-1308.

[5] Andrews, D.W.K. (1997). A Conditional Kolmogorov Test, *Econometrica*, **65**, 1097-1128.

[6] Baraud, Y., Huet, S. and Laurent, B. (2005). Testing Convex Hypotheses on the Mean of a Gaussian Vector. Application to Testing Qualitative Hypotheses on a Regression Function, *The Annals of Statistics*, **33**, 214-257.

[7] Birke, M. and Dette, H. (2007). Estimating a Convex Function in Nonparametric Regression, *Scandinavian Journal of Statistics*, **34**, 384-404.

[8] Bollaerts, K., Eilers, P.H. C. and van Mechelen, I. (2006). Simple and multiple P-splines regression with shape constraints, *British Journal of Mathematical and Statistical Psychology*, **59**, 451-469.

[9] Bowman, A.W., Jones, M.C. and Gijbels, I. (1998). Testing monotonicity of regression, *Journal of Computation and Graphic Statistics*, **7**, 489-500.

[10] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion), *Journal of the Royal Statistical Society B*, **37**, 149-192.

[11] Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters, *Annals of Mathematical Statistics*, **26**, 607-616.

[12] Chen, X.(2007). Large Sample Sieve Estimation of Semi-Nonparametric Models, chapter 76 in *Handbook of Econometrics, 6B*, Elsevier.

[13] Chen, X., and Christensen, J.D. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions, *Journal of Econometrics*, **188**, 447-465.

[14] Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement, *Biometrika*, **96**, 559-575.

[15] Chetverikov, D. (2019). Testing regression monotonicity in econometric models, *Econometric Theory*, **35 (4)**, 729-776.

[16] Claeskens, G., Krivobokova, T. and Opsomer, J.D. (2009). Asymptotic properties of penalized spline estimators, *Biometrika*, **96**, 529-544.

[17] De Boor, C. (1978). *A Practical Guide to Splines*. Springer, Berlin.

[18] Delgado, M.A., Hidalgo, J. and Velasco, C. (2005). Distribution free goodness-of-fit tests for linear processes, *Annals of Statistics*, **33,** 2568-2609.

[19] Demko, S., Moss, W. and Smith, P. (1984): “Decay rates for inverses of band matrices,” *Mathematics and Computation*, **43**, 491-499.

[20] Dette, H., Hoderlein, S. and Neumayer, N. (2016). Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutzky Negative Semidefiniteness, *Journal of Econometrics*, **191 (1)**, 129-144.

[21] Dette, H., Neumeyer, N. and Pilz, K.F. (2006). A simple nonparametric estimator of a strictly monotone regression function, *Bernoulli*, **12 (3)**, 469-490.

[22] Diack, C. and Thomas-Agnan, C. (1998). A nonparametric test of the non-Convexity of regression, *Journal of Nonparametric Statistics*, **9**, 335-362.

[23] Dierckx, P. (1980). Algorithm 42. An algorithm for cubic spline fitting with convexity constraints, *Computing*, **24**, 349-371.

[24] Dierckx, P. (1993). *Curve and surface fitting with splines*. Oxford: Clarendon.

[25] Dümbgen, L. and Spokoiny, V.G. (2001). Multiscale Testing of Qualitative Hypotheses, *The Annals of Statistics*, **29**, 124-152.

[26] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution*

Function. Society for Industrial and Applied Mathematic, Philadelphia.
[27] Eilers, P.H. C. and Marx, B.D. (1996). Flexible Smoothing with B-splines and Penalties, *Statistical Science*, **11**, 89-102.
[28] Fonseca, C.M. Petronilho, J. (2001): "Explicit inverses of some tridiagonal matrices, " *Linear Algebra and its Applications*, *325*, 7–21.
[29] Friedman, J.H. and Tibshirani, R.J. (1984). The monotone smoothing of scatterplots, *Technometrics*, **26**, 243-250.
[30] Giacomini, R., Politis, D.N. and White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators, *Econometric Theory*, **29**, 567-589.
[31] Ghosal, S., Sen, A. and van der Vaart, A. (2000). Testing monotonicity of regression, *The Annals of Statistics*, **28**, 1054-1082.
[32] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, Almqvist and Wiksell, Stockholm.
[33] Groeneboom, P., Jongbloed, G. and Wellner, J.A. (2001). Estimation of a convex function: Characterizations and asymptotic theory, *The Annals of Statistics*, **29**, 1653-1698.
[34] Guntuboyina, A. and Sen, B. (2015). Global risk bounds and adaptation in univariate convex regression, *Probability Theory and Related Fields*, **163**, 379-411.
[35] Hall, P. and Heckman, N. E. (2000). Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions, *The Annals of Statistics*, **28**, 20-39.
[36] Hall, P. and Huang, L-S. (2001). Nonparametric Kernel Regression Subject to Monotonicity Constraints, *The Annals of Statistics*, **29**, 624-647.
[37] Hanson, P.L. and Pledger, G. (1976). Consistency in concave regression, *The Annals of Statistics*, **4**, 1038-1050.
[38] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits, *The Annals of Statistics*, **21**, 1926-1947.
[39] Harville, D.A. (2008). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag.
[40] Hildreth, C. (1954). Point estimates of ordinates of concave functions, *Journal of American Statistical Association*, **49**, 598-619.
[41] Hong, Y. and White, H. (1995). Consistent Specification Testing via Nonparametric Series Regression, *Econometrica*, **63**, 1133-1160.
[42] Juditsky, A. and Nemirovski, A. (2002). On nonparametric tests of positivity/monotonicity/convexity, *The Annals of Statistics*, **30**, 498-527.
[43] Khmaladze, E.V. (1981). Martingale approach to the theory of goodness of fit tests, *Theory Probability and its Applications*, **26**, 240-257.
[44] Kostyshak, S. (2017). Non-Parametric Testing of U-Shaped Relationships. Working paper, University of Florida.
[45] Koul, H.L. and Stute, W. (1999). Nonparametric model checks for time series, *The Annals of Statistics*, **27**, 204-237.
[46] Lee, Y., and Robinson, P.M. (2016). Series estimation under cross-sectional dependence, *Journal of Econometrics*, **190**, 1-17.
[47] Li, W., Naik, D. and Swetits, J. (1996). A data smoothing technique for piecewise convex/concave curves, *SIAM Journal on Scientific Computing*, **17**, 517-537.
[48] Mammen, E. (1991a). Estimating a Smooth Monotone Regression Function, *The Annals of Statistics*, **19**, 724-740.
[49] Mammen, E. (1991b). Nonparametric regression under qualitative smoothness assumptions, *The Annals of Statistics*, **19**, 741-759.
[50] Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restric-

tions, *Scandinavian Journal of Statistics*, **26**, 239-252.

[51] MEYER, M.C. (2008). Inference using shape-restricted regression splines, *The Annals of Applied Statistics*, **2**, 1013-1033.

[52] MUKERJEE, H. (1988). Monotone nonparametric regression, *The Annals of Statistics*, **16**, 741-750.

[53] NICULESCU, C.P. (2003). Convexity according to means, *Mathematical Inequalities and Applications*, **6**, 571-579.

[54] NIKABADZE, A. AND STUTE, W. (1997). Model checks under random censorship, *Statistics & Probability Letters*, **32**, 249-259.

[55] O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems (with Discussion), *Statistical Science*, **1**, 505-527.

[56] O'SULLIVAN, F. (1988). Nonparametric estimation of relative risk using splines and crossvalidation, *SIAM Journal on Scientific and Statistical Computing*, **9(3)**, 531-542

[57] RAMSAY, J.O. (1988). Monotone regression splines in action, *Statistical Science*, **3(4)**, 425-461.

[58] SEN, P.K. (1982). Invariance principles for recursive residuals, *The Annals of Statistics*, **10**, 307-312.

[59] SCHLEE, W. (1982). Nonparametric Tests of the Monotony and Convexity of Regression, in *Nonparametric Statistical Inference*. Amsterdam: North-Holland.

[60] STEPHENS, M.A. (1992). Introduction to Kolmogorov (1933). On the empirical determination of a distribution function. *Breakthroughs in Statistics, Volume II. Methodology and Distribution* (Samuel Kotz and Norman L. Johnson., eds.), Springer-Verlag, New York, 93-105.

[61] STUTE, W. (1997). Nonparametric model checks for regression, *The Annals of Statistics*, **25**, 613-641.

[62] STUTE, W., THIES, S. AND ZHU, L.-X. (1998b). Model checks for regression: An innovation process approach, *The Annals of Statistics*, **26**, 1916-1934.

[63] WANG, J.C. AND MEYER, M.C. (2011). Testing the monotonicity or convexity of a function using regression splines, *The Canadian Journal of Statistics*, **39 (1)**, 89-107.

[64] WRIGHT, F.T. (1981). The Asymptotic Behaviour of Monotone Regression Estimates, *The Annals of Statistics*, **2**, 443-448.

[65] ZHOU, S., SHEN, X. AND D.A. WOLFE (1998). Local Asymptotics for Regression Splines and Confidence Regions, *The Annals of Statistics*, **26**, 1760-1782.

ECONOMICS DEPARTMENT
LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
HOUGHTON STREET
LONDON WC2A 2AE
U.K. E-MAIL: t.komarova@lse.ac.uk

ECONOMICS DEPARTMENT
LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
HOUGHTON STREET
LONDON WC2A 2AE
U.K. E-MAIL: f.j.hidalgo@lse.ac.uk