The Annals of Statistics
2023, Vol. 51, No. 6, 2299–2317
https://doi.org/10.1214/23-AOS2311

TESTING NONPARAMETRIC SHAPE RESTRICTIONS

BY TATIANA KOMAROVA[1,a] AND JAVIER HIDALGO[2,b]

[1]*Economics Department, University of Manchester,* [a]*tatiana.komarova@manchester.ac.uk*
[2]*Economics Department, London School of Economics and Political Science,* [b]*f.j.hidalgo@lse.ac.uk*

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 a Brownian motion index by the c.d.f. of the regressor. As a result, the test is distribution-free and critical values are readily 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 that 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 $d^r m(x)/dx^r \geq 0$ (≤ 0). When $r = 1$ or 2, we respectively have 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 just an illustration of the scope of the approach. In Section 2, we provide precise conditions for the shape constraints we consider, accompanied by relevant examples. Furthermore, the supplementary material [34] includes additional noteworthy shape examples, such as quasi-convexity, as well as r- and ρ- convexity/concavity.

Although ample literature exists on shape constraint testing, the majority of it focuses on monotonicity and/or convexity. Notable examples include [1, 3, 5, 8, 13, 18, 20, 23, 27, 32, 46, 50]. It is worth mentioning an exception in [35], which proposes a consistent test for *U-shape*.

Regarding the regularity conditions in these references, some of them, such as [5, 20, 32], focus on regression functions in a Gaussian white noise model or in models with deterministic explanatory variables, as seen in [3, 5, 18] or [27]. However, when dealing with random explanatory variables, as in [1, 13] and [23], the assumption is either that they are stochastically independent of the unobserved regression error or that the error's conditional distribution is symmetric given the explanatory variable. On the other hand, other papers either lack asymptotic limit theory useful for inference or are tailored to specific types of shapes, making their extension to more general shape properties nontrivial, such as [8, 27] or [35]. Consequently, one of the paper's objectives is to examine a testing methodology that not only applies to a

Received June 2020; revised July 2023.
MSC2020 subject classifications. Primary 05C38, 15A15; secondary 05A15, 15A18.
Key words and phrases. Monotonicity, convexity, concavity, U-shape, quasi-convexity, log-convexity, convexity in means, B-splines, CUSUM transformation, distribution-free estimation.

wide range of shape properties but also (**a**) enables valid statistical inferences under weak conditions and (**b**) offers the flexibility to test for multiple shape constraints simultaneously such as monotonicity and log-convexity. Furthermore, the proposed methodology is easy to implement, as it only requires the computation of the *CUSUM* (least squares) of "recursive" residuals.

The methodology we suggest in this paper is related to methods used in goodness-of-fit tests that involve a null hypothesis belonging to a parametric family while allowing for a nonparametric alternative. Specifically, we consider the nonparametric regression model

$$y_i = m(x_i) + u_i, \qquad E[u_i|x_i] = 0, \tag{1.1}$$

where x_i are realizations of random variable X with a bounded support $\mathcal{X} =: [\underline{x}, \overline{x}]$ and $m(\cdot)$ is smooth. Section 5 will provide Condition C1, which offers more detailed conditions concerning the sequences $\{u_i\}_{i\in\mathbb{N}}$ and $\{x_i\}_{i\in\mathbb{N}}$. Our aim is testing whether the regression function $m(\cdot)$ possesses the shape properties captured by the null hypothesis

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

where the class of interest $\mathcal{M}_0$ is a subset of smooth functions from $\mathcal{X}$ to $\mathbb{R}$, say convexity. Following [48] or [4], we might base the testing procedure on functionals of the partial sums empirical process

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

Here, $\mathcal{I}(\cdot)$ is the indicator function and $\mathcal{I}_i(x)$ is the abbreviation of $\mathcal{I}(x_i < x)$, and $\widehat{u}_i = y_i - \widehat{m}_{\mathcal{B}}(x_i; L)$, $i = 1, \ldots, n$ are the residuals obtained after $m(\cdot)$ has been estimated by some nonparametric estimator $\widehat{m}_{\mathcal{B}}(\cdot; L)$; 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, suppose for a moment that $\mathcal{M}_0$ is the set of some parametric functions, say $m(x) =: m(x;\theta)$. We then have that

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

is such that $(Eu_i^2)^{-1/2} n^{-1/2} \sum_{i=1}^{n} u_i \mathcal{I}_i(x)$ converges to a Brownian motion indexed by the c.d.f. of the regressor 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 [21], and later in a regression model context by [48]. However, in our scenario, we have that

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

where the second term is $O_p(n^{-\nu})$, for some $\nu < 1/2$, becoming then the dominant term in the behavior of $\mathcal{K}_n(x)$. As we describe in Section 5, 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 [9]. The consequence is that the asymptotic behavior of the transformation follows a Brownian motion indexed by the c.d.f. of the regressor, allowing testing to be implemented using functionals such as Kolmogorov–Smirnov, Cramér–von Mises or Anderson–Darling, among others. As a byproduct, a nice feature of the transformation is that its asymptotic distribution is pivotal, that is, 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 *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 different shapes to the coefficients of the *B-spline* approximation. In Section 3, we present algorithmic steps of our testing procedure with the purpose to provide practitioners with useful guidance. Section 4 gives two sets of examples of shape properties. One example focuses on shapes described by an increasing number of linear inequalities, while the second example shape properties are described by an increasing number of nonlinear inequalities. In Section 5, 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 6 introduces a valid bootstrap algorithm. Section 7 presents Monte Carlo experiments and Section 8 concludes with a summary. All the proofs, which employ a series of lemmas, are confined to the Supplementary Material. The Supplementary Material also contains some extra 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**) empirical applications, (**iii**) motivation for using *B-splines* instead of some other sieve-type estimator and (**iv**) additional examples of shape constraints of interest.

2. Nonparametric estimation methodology. A preliminary and key step in providing a test for H_0 in (1.2) is to compute a nonparametric estimator of $m(\cdot)$ 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 works on isotone/monotone regressions are [10] and [51]. Later approaches include [28, 39, 43] and [17]. When the null hypothesis is convexity, the approach in [31] estimates $m(\cdot)$ using least squares. Statistical properties for this estimator are established in [25, 29, 40] and [26]. An alternative estimator discussed in [6] involves first obtaining an unconstrained derivative estimate of the regression function, which is then isotonized and integrated.

However, the previous techniques have some limitations, such as implementation difficulties, narrow scope or lack of useful asymptotic theory for inference. Hence, we adopt a different approach based on *B-splines* and/or penalized *B-splines*, known as *P-splines*. Using *B-splines* (*P-splines*) offer several advantages for our purposes in this paper. As discussed later in this section, first there is no dependence between base splines separated by a certain distance, as outlined in the properties of the *B-spline* basis. Second, *B-splines* (*P-splines*) are well suited for testing properties based on regression function derivatives. Additionally, we can express $\mathcal{M}_0$ in (1.2) in terms of constraints on the coefficients of the *B-spline* approximation to $m(\cdot)$, enabling the implementation of valid asymptotic theory for the test.

Note that [45], and later extended by [42], introduced monotone regression splines, closely related to *B-splines*, to estimate convex functions or functions that are both monotone and convex. However, our approach differs in that we allow the number of *B-spline* coefficients and constraints to increase to infinity, whereas [45] and [42] considered a fixed number of constraints. Wang and Meyer [50] employed quadratic *B-splines* for a monotonicity test and cubic *B-splines* for a convexity test, allowing for an increasing number of knots. Their approach and implementation differ from ours, and it may not be applicable to general shapes. For a comparison of the performance between the monotonicity test in [50] and our test, please refer to the Supplementary Material.

Let us delve into the details of *B-splines* (*P-splines*). They are constructed by connecting polynomial pieces at specific points called knots. The computation of these splines is achieved recursively, as described in [15], for any polynomial degree. By construction, the *B-spline* basis of degree q (i) takes positive values on the domain spanned by $q+2$ adjacent knots, and is zero otherwise; (ii) consists of $q+1$ polynomial pieces each of degree q, and the polynomial pieces join at q inner knots; (iii) at the joining points, the $(q-1)$th derivatives are continuous.

Suppose that one is interested in approximating the regression function $m(\cdot)$ on $\mathcal{X}$. Hereafter, 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 $L = L' + q$. Then, denoting the *B-spline* basis of degree q by

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

we approximate $m(x)$ by a linear combination $m_{\mathcal{B}}(x;L) = \sum_{\ell=1}^{L} \beta_\ell p_{\ell,L}(x;q)$ of $\boldsymbol{P}_L(x)$. Henceforth, we shall denote the knots as $\{z_k\}$, $k = 1-q, \dots 0, 1, \dots, L+1$, where $0 = z_{1-q} = \cdots = z_1$ and $1 = z_{L'+1} = \cdots = z_{L+1}$.

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 overfitting 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* (i.e., 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, [15] and [19] for *B-splines* and [7, 22] for *P-splines*. Some literature on shape-preserving splines (for standard shapes such as monotonicity or convexity) includes [38, 41, 42] and [45].

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

$$\begin{aligned} &(\mathbf{a}) \quad \sum_{\ell=1}^{L} p_{\ell,L}(x;q) = 1 \quad \text{for all } x \text{ and } q. \\ &(\mathbf{b}) \quad \frac{dm_{\mathcal{B}}(x;L)}{dx} = \sum_{\ell=1}^{L-1} \frac{q\triangle\beta_{\ell+1}}{z_{\ell+1} - z_{\ell+1-q}} p_{\ell+1,L}(x;q-1). \end{aligned} \tag{2.2}$$

Specifically, (**a**) indicates that *B-splines* are a partition of 1, whereas (**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 forth. Other sieve estimators might be used, (see the survey in [11], and in particular the *Bernstein* polynomials basis as they share some properties similar to those in (2.2). 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. More importantly, we discuss how one can relate the *B-spline* approximation $m_{\mathcal{B}}(\cdot;L)$ to (1.2). Namely, because any

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

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

$m_{\mathcal{B}}(\cdot; L)$ can be fully characterized by $\beta =: (\beta_1, \ldots, \beta_L)' \in \mathbb{R}^L$, 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 $S_{q,L} \subset \mathbb{R}^L$, denoting its associated constraint *B-splines* approximation by

$$\mathcal{M}_{S_{q,L}} =: \{m_{\mathcal{B}}(\cdot; L) | \beta \in S_{q,L}\}. \tag{2.3}$$

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 nonstochastic.

(b) The boundary of $S_{q,L}$ consists of a finite number of surfaces, each of which can be explicitly represented by a continuously differentiable function relating one component of β to the other components. In other words, each such surface can be described as $\beta_\ell = s(\beta_{-\ell})$ for some ℓ with $s(\cdot)$ being a continuously differentiable function.

(c) Let $\mathcal{H}$ represent the Hausdorff distance calculated in the supremum norm within the space of continuous functions from $\mathcal{X}$ to $\mathbb{R}$.[3] Then

$$\mathcal{H}(\mathcal{M}_0, \mathcal{M}_{S_{q,L}}) \to 0 \quad \text{as } L \to \infty.$$

Condition C0 states that $\mathcal{M}_0$ can be captured by constraints on the coefficient vector $\beta \in \mathbb{R}^L$, which become *both necessary and sufficient* as the knot system becomes increasingly dense in $\mathcal{X}$. The requirement for an increasingly dense knot system is implied by part (c). Moreover, these constraints on β are independent of the available data, making the approach appealing for implementation purposes. Part (b) is required for practical reasons, ensuring that constrained estimation only requires a finite number of inequality constraints for any finite L.

The idea is then to test the null hypothesis

$$H_0^{\mathcal{B}}: (\beta_1, \ldots, \beta_L)' \in S_{q,L} \tag{2.4}$$

with a suitable choice of $S_{q,L}$. Under $C0$, for fixed L, the test in (2.4) can be conceptually regarded as the approximation of the original testing problem in (1.2). As $L \to \infty$, however, 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, that is, monotonicity or convexity, the restrictions given in (2.4) are equivalent to the restrictions in the whole domain $\mathcal{X}$.

One can easily obtain the *unconstrained* estimator of $m(\cdot)$ defined as

$$\breve{m}_{\mathcal{B}}(x; L) = \breve{b}' \boldsymbol{P}_L(x), \tag{2.5}$$

where $\breve{b} = (\breve{b}_1, \ldots, \breve{b}_L)' = (\frac{1}{n}\sum_{k=1}^n \boldsymbol{P}_k \boldsymbol{P}_k')^+ \frac{1}{n}\sum_{k=1}^n \boldsymbol{P}_k y_k$, and B^+ denotes the Moore–Penrose inverse of the matrix B. Additionally, $\boldsymbol{P}_k$ is an abbreviation for $\boldsymbol{P}_L(x_k)$ in (2.1). To obtain an estimator under the null hypothesis, we consider estimation under the constraints in (2.4), that is,

$$\widehat{b} =: \arg \min_{(b_1, b_2, \ldots, b_L)' \in S_{q,L}} \sum_{i=1}^n \left(y_i - \sum_{\ell=1}^L b_\ell p_{\ell,L}(x_i; q) \right)^2, \tag{2.6}$$

so that under (1.2)/(2.4) the estimator of $m(\cdot)$ is

$$\widehat{m}_{\mathcal{B}}(x; L) = \widehat{b}' \boldsymbol{P}_L(x). \tag{2.7}$$

[3]If $d_{m,m_{\mathcal{B}}} =: \sup_{x\in\mathcal{X}} |m(x) - m_{\mathcal{B}}(x; L)|$, then the Hausdorff distance is $\mathcal{H}(\mathcal{M}_0, \mathcal{M}_{S_{q,L}}) = \max\{\sup_{m(\cdot)\in\mathcal{M}_0} \inf_{m_{\mathcal{B}}(\cdot;L)\in\mathcal{M}_{S_{q,L}}} d_{m,m_{\mathcal{B}}}, \sup_{m_{\mathcal{B}}(\cdot;L)\in\mathcal{M}_{S_{q,L}}} \inf_{m(\cdot)\in\mathcal{M}_0} d_{m,m_{\mathcal{B}}}\}$.

In the remainder of the paper, we will use notation "$\breve{}$" for unconstrained estimation and notation "$\widehat{}$" for the constrained one.

As an illustration, suppose that we are interested in testing for nondecreasing functions. Then, as Example 1 below will explain, (2.6) becomes

$$\widehat{b} = (\widehat{b}_1, \dots, \widehat{b}_L)' =: \arg \min_{b_1 \le b_2 \le \cdots \le b_L} \sum_{i=1}^{n} \left(y_i - \sum_{\ell=1}^{L} b_\ell p_{\ell,L}(x_i; q) \right)^2, \tag{2.8}$$

which is a quadratic programming problem with linear constraints. When the constraints are nonlinear, such as those in Example 2 appearing later in the paper, the 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 Material.

The next section provides a succinct overview of the algorithm used to perform the transformation of the process $\mathcal{K}(x)$ and construct test statistics. Subsequently, the paper develops the statistical theory that substantiates the aforementioned approach. Prior to delving into that, however, it is important to address a fundamental implementation concern pertaining to the power of our test. Specifically, to ensure that the power of our test is not trivial, it is essential to uphold the binding constraints during the transformation.

Let us illustrate how we can write (2.7) and transform the linear space of the *B-splines* basis when some constraints are binding in the constrained monotonicity estimation. Suppose we have $\widehat{b}_{\ell_0} = \widehat{b}_{\ell_0+1}$ in (2.8). Denote

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

Then (2.7) can be written as

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

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

$$\widetilde{\boldsymbol{P}}_k =: \widetilde{\boldsymbol{P}}_L(x_k). \tag{2.9}$$

It is $\widetilde{\boldsymbol{P}}_L(x)$, and not the original system $\boldsymbol{P}_L(x)$, that has to be used in the transformation. Further examples on how the binding constraints can be enforced in the *CUSUM*/Khmaladze transformation are in Section 5.2.

3. Algorithm.

Below are algorithmic steps of our testing procedure giving a quick guide to the practitioners.

STEP 1 Order the sample $\{(x_i, y_i)\}_{i=1}^{n}$ in the ascending order of x. Without a loss of generality, we will assume that the original sample is already ordered in this way.

[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.

STEP 2 Find a constrained estimator (2.7), compute the residuals $\widehat{u}_i = y_i - \widehat{m}_{\mathcal{B}}(x_i; L)$, $i = 1, \dots, n$ and define the system $\widetilde{\mathbf{P}}_L(x)$ of "effective polynomials" using all the binding constraints participating in $\widehat{m}_{\mathcal{B}}(\cdot; L)$. Let L^e denote the number of polynomials in this system.

STEP 3 For each $j = 1, \dots, \tilde{n}$, where $\tilde{n} =: n - L^e - 1$, compute

$$\widehat{v}_j = \widehat{u}_j - \widetilde{\boldsymbol{P}}_j' \left(\sum_{k=1}^{n} \widetilde{\boldsymbol{P}}_k \widetilde{\boldsymbol{P}}_k' \mathcal{I}_j(x_k) \right)^+ \sum_{k=1}^{n} \widetilde{\boldsymbol{P}}_k \mathcal{I}_j(x_k) \widehat{u}_k. \tag{3.1}$$

STEP 4 Compute the estimate of the variance of u_i, σ_u^2, as $\breve{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^{n} \breve{u}_i^2$, where $\breve{u}_i$ are unconstrained residuals $\breve{u}_i = y_i - \breve{m}_{\mathcal{B}}(x_i; L)$.

STEP 5 Compute $\widetilde{M}_{\tilde{n}}(x_i) = \frac{1}{\sqrt{\tilde{n}}} \sum_{j=1}^{\tilde{n}} \hat{v}_j \mathcal{I}_j(x_i)$ and calculate the values of standard functionals such as the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling defined respectively as

$$\mathcal{KS}_{\tilde{n}} = \sup_{i=1,\dots,\tilde{n}} \left| \frac{\widetilde{M}_{\tilde{n}}(x_i)}{\breve{\sigma}_u} \right|, \qquad \mathcal{C}v\mathcal{M}_{\tilde{n}} = \sum_{i=1}^{\tilde{n}} \frac{\widetilde{M}_{\tilde{n}}(x_i)^2}{\tilde{n} \breve{\sigma}_u^2}, \qquad \mathcal{AD}_{\tilde{n}} = \sum_{i=1}^{\tilde{n}} \frac{\widetilde{M}_{\tilde{n}}(x_i)^2 / \tilde{n}}{\breve{\sigma}_u^2 \widehat{F}_X(x_i)}, \tag{3.2}$$

where $\widehat{F}_X$ denotes the empirical c.d.f. of X.[5] Compare them to the critical values $\mathcal{KS}_{\tilde{n}}^*(\alpha_0)$, $\mathcal{C}v\mathcal{M}_{\tilde{n}}^*(\alpha_0)$, $\mathcal{AD}_{\tilde{n}}^*(\alpha_0)$, respectively, for a chosen significance level α_0. If, for example, $\mathcal{KS}_{\tilde{n}} > \mathcal{KS}_{\tilde{n}}^*(\alpha_0)$, reject the null by Kolmogorov–Smirnov at the significance level α_0.

4. Examples. We give two sets of examples of shape constraints. In Example 1, the shapes in $S_{q,L}$ are described by linear inequalities, whereas in Example 2, the constraints are given by nonlinear inequalities (except for some special cases). These examples are meant to illustrate the scope of applicability of the proposed testing methodology rather than to give an exhaustive list of potential applications. Additional examples can be found in the Supplementary Material.

EXAMPLE 1. Our first set of examples is mainly described by derivatives on $m(\cdot)$. More specifically,

$$H_0 : a_r \cdot d^r m(x)/dx^r \geq c_r \quad \forall x \in \mathcal{X}, r \in R, \tag{4.1}$$

where R is a finite subset of $\mathbb{N}^+$, $a_r \in \{-1, 1\}$ and c_r are known constants, so that (4.1) allows for inequalities on 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 (4.1) is

$$S_{q,L} = \{\beta \in \mathbb{R}^L | \forall r \in R, \forall z_k, a_r \cdot d^r m_{\mathcal{B}}(z_k; L)/dx^r \geq c_r\}.$$

Note 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. For instance, let $R = \{1\}$ and $c_1 = 0$. Then, using the property (b) of *B-splines*, we conclude that

$$S_{q,L} = \{\beta \in \mathbb{R}^L | a_1(\beta_{\ell+1} - \beta_\ell) \geq 0, \ell = 1, \dots, L-1\},$$

[5] One could, of course, center the process $\widetilde{M}_{\tilde{n}}(x)$ to ensure that it converges to a Brownian bridge indexed by the empirical c.d.f. of X. Then $\mathcal{AD}_{\tilde{n}}$ would be defined in a standard manner as follows: $\mathcal{AD}_{\tilde{n}} = \sum_{i=1}^{\tilde{n}} \frac{\widetilde{M}_{\tilde{n}}(x_i)^2/\tilde{n}}{\breve{\sigma}_u^2 \widehat{F}_X(x_i)(1-\widehat{F}_X(x_i))}$.

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

More generally, for $R = \{r\}$, $r > 1$, $c_r = 0$, splitting $[0, 1]$ into equidistant subintervals leads to the following form of $S_{q,L}$:

$$S_{q,L} = \left\{\beta \in \mathbb{R}^L | a_r \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \beta_{\ell+k} \geq 0, \ell = q, \dots, L-q+1-r\right\}.$$

An even more refined form of $S_{q,L}$, which may be beneficial for small L, would also involve constraints that capture the behavior of $m_{\mathcal{B}}(\cdot; L)$ around the boundary. This is further discussed in the Supplementary Material.

Our second set of examples illustrates shape properties well developed in the mathematical literature. It builds on a general notion of a *mean function* given in [44] and the Supplementary Material to this paper. Examples of mean functions include the *arithmetic* (A), the *geometric* (G), the *harmonic* (H), the *logarithmic* and the *identric* means.

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

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

Using different combinations of arithmetic (A), geometric (G) and harmonic (H) means, we end up with many cases of MN-convex functions. Among them are: (a) m is AG-convex if and only if $m'(x)/m(x)$ is increasing; (b) m is GG-convex if and only if $xm'(x)/m(x)$ is increasing; (c) m is HG-convex if and only if $x^2 m'(x)/m(x)$ is increasing; (d) m is HH-convex if and only if $x^2 m'(x)/m^2(x)$ is increasing.

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

$$S_{q,L} = \left\{\beta \in \mathbb{R}^L | \forall z_{k_1} < z_{k_2}, k_1, k_2 \in \{1-q, \dots, L+1\}, \right.$$
$$\left. \frac{(z_{k_1})^2 m'_{\mathcal{B}}(z_{k_1}; L)}{m^2_{\mathcal{B}}(z_{k_1}; L)} \leq \frac{(z_{k_2})^2 m'_{\mathcal{B}}(z_{k_2}; L)}{m^2_{\mathcal{B}}(z_{k_2}; L)}, \text{ and } \beta_\ell > 0, \ell = 1, \dots, L\right\}.$$

Sets $S_{q,L}$ for other MN-convex functions are constructed similarly. This is further discussed in the Supplementary Material, along with a longer list of specific MN-convex functions.

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

5. Regularity conditions and the testing methodology.

We shall now present statistical foundations of our testing methodology. For that purpose, we introduce 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, $f_X(\cdot)$, 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. $m(\cdot)$ is η times continuously differentiable on $[0, 1]$, $\eta \geq 1$ and $d^\eta m(x)/dx^\eta$ is Hölder continuous with exponent $0 < \alpha \leq 1$:

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

for some finite positive constant M_0.

CONDITION C3. As $n \to \infty$, L satisfies

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

As in [48], Condition C1 can be weakened to allow for heteroscedastic errors, that is $E[u_i^2|x] = \sigma_u^2(x)$. Heteroscedasticity complicates technical arguments, and thus, for expositional simplicity we omit a detailed analysis of this case. In our empirical applications in the Supplementary Material, however, we present examples of models with heteroscedastic errors and illustrate how to deal with them in practice. Condition C2 is a regularity condition on the regression function $m(\cdot)$. Essentially, it asserts that, at the minimum, we need slightly more than continuous differentiability of $m(\cdot)$. It guarantees that the approximation error or the bias $m^\dagger(x) =: m_{\mathcal{B}}(x; L) - m(x)$ is $O(L^{-\eta-\alpha})$; see Theorems 3.1 and 4.1 in [2] or [52], and also see [12] and the references therein. In case of using *P-splines*, we also refer to [14] Theorem 2. Condition C3 bounds the rate at which L increases to infinity with n.

We now give details of the testing methodology. Naturally, we shall focus on the null hypothesis (2.4), which is given in terms of the coefficients β, with the alternative hypothesis being the negation of the null. Then 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 here is that the number of such constraints increases with the sample size.

With the objective of conducting tests for (2.4), we employ functionals of (1.3) with the purpose of detecting if they are significantly different than zero. In (1.3), $\widehat{u}_i$ are the constrained residuals given by

$$\widehat{u}_i = y_i - \widehat{m}_{\mathcal{B}}(x_i; L), \quad i = 1, \dots, n, \tag{5.1}$$

so that $\mathcal{K}_n(x)$ can be interpreted as a *LM* type of test. Recall that in a standard regression model, the *LM* test would be based on the first-order conditions $\mathcal{LM}_n(L) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{P}_L(x_i)\widehat{u}_i$, so that one tests if the residuals and regressors $\boldsymbol{P}_L(x_i)$ satisfy the orthogonality moment condition induced by Condition C1.

We have that (1.4) is

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

Using Conditions C2 and C3 and following Lee and Robinson [37] or Chen and Christensen [12] in a more general context, we obtain that

$$\sum_{\ell=1}^L (\widehat{b}_\ell - \beta_\ell)\frac{1}{n}\sum_{i=1}^n p_{\ell,L}(x_i; q) = O_p\big((L/n)^{1/2}\big).$$

Denoting $\mathcal{P}_{n,\ell}(x; q) =: n^{-1}\sum_{i=1}^n p_{\ell,L}(x_i; q)\mathcal{I}_i(x)$, we conclude that

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

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

However, as shown in the proof of Theorem 2, $\mathcal{K}_n(x)$ can be expressed as $\mathcal{K}_n(x) = \frac{1}{n}\sum_{i=1}^n \widetilde{u}_i\mathcal{I}_i(x) + o_p(n^{-1/2})$, where

$$\widetilde{u}_i = u_i - \widetilde{\boldsymbol{P}}_i'\left(\sum_{k=1}^n \widetilde{\boldsymbol{P}}_k\widetilde{\boldsymbol{P}}_k'\right)^+ \sum_{k=1}^n \widetilde{\boldsymbol{P}}_k u_k, \tag{5.2}$$

and $\widetilde{\boldsymbol{P}}_k$ is as defined in (2.9) (and thus, already incorporating all the binding constraints in the estimation).[6]

Now $\widetilde{u}_i$ in (5.2) represents the least squares residuals in a regression model with the dependent variable u_i and the vector of "effective" polynomials $\widetilde{\boldsymbol{P}}_L(x_i)$ as the set of explanatory variables. This observation suggests the use of the *CUSUM* of "recursive" residuals to construct asymptotically pivotal tests, as proposed by Brown, Durbin and Evans [9]; see also Sen [47]. First, we describe the implementation of the *CUSUM* of "recursive" residuals when the restrictions in $S_{q,L}$ are linear. More general scenarios, where some constraints are nonlinear, are addressed later.

5.1. *All the constraints on β_ℓ are linear.* To describe our pivotal transformation, we first recall our notation in (2.1) and (2.9) when testing for monotonicity,

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

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

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

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" $\widetilde{\boldsymbol{P}}_L(x)$ is crucial for the power of the test. Using $\boldsymbol{P}_L(x)$ without taking into account the binding constraints will make the test to have only trivial power; see the discussion in Section 5.3 and the Supplementary Material. However, for the sake of expositional simplicity, in this section we shall consider the case of no binding constraints (thus, $\widetilde{\boldsymbol{P}}_L(x) = \boldsymbol{P}_L(x)$).[7]

With this in mind, for any $x \in \mathcal{X}$, let us define

$$C_n(x) = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{P}_k u_k \mathcal{J}_k(x), \qquad A_n(x) = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{P}_k \boldsymbol{P}_k' \mathcal{J}_k(x), \tag{5.3}$$

where $\mathcal{I}(x \le x_k) =: \mathcal{J}_k(x) = 1 - \mathcal{I}_k(x)$. We will use the abbreviations

$$C_{n,i} =: C_n(\widetilde{x}_i), \qquad A_{n,i} =: A_n(\widetilde{x}_i),$$

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, $k = 2, \dots, L' + 1$, bigger than x and $1/2 < \varsigma < 1$.[8] Then the *CUSUM* of (forward) "recursive" least squares is defined as

$$\mathcal{M}_n(x) =: \frac{1}{n^{1/2}} \sum_{i=1}^{n} v_i \mathcal{I}_i(x), \tag{5.4}$$

[6] $\sum_{\ell=1}^{L}(\widehat{b}_\ell - \beta_\ell) p_{\ell,L}(x_i; q)$ can be rewritten in terms of $\widetilde{\boldsymbol{P}}_i$ only. However, to convey some intuition about its asymptotic behavior, it is convenient to leave this term as it is.

[7] When examining the local power of the test in Section 5.3 we shall make explicit the consideration of binding constraints. An additional discussion of the role of the binding constraints is outlined in the Supplementary Material.

[8] We make this "trimming" 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.

where $v_i = u_i - \boldsymbol{P}_i' A_{n,i}^+ C_{n,i}$. Note that because $\sum_{i=1}^n \phi(x_i) = \sum_{i=1}^n \phi(x_{(i)})$, where $x_{(i)}$ is the ith order statistic of $\{x_i\}_{i=1}^n$, we could write (5.4) as

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

$$v_{(i)} = u_{(i)} - \boldsymbol{P}_{(i)}' \left(\frac{1}{n} \sum_{k=i}^n \boldsymbol{P}_{(k)} \boldsymbol{P}_{(k)}' \right)^+ \frac{1}{n} \sum_{k=i}^n \boldsymbol{P}_{(k)} u_{(k)} \quad \text{with } \boldsymbol{P}_{(i)} =: \mathbf{P}_L(x_{(i)}).$$

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 [9].

Now denoting $\mathcal{K}_n^1(x) =: n^{-1} \sum_{i=1}^n u_i \mathcal{I}_i(x)$, the process $\mathcal{M}_n(x)$ becomes a linear transformation of $\mathcal{K}_n^1(x)$, that is, $\mathcal{M}_n(x) =: n^{1/2}(\mathcal{T}_n \mathcal{K}_n^1)(x)$, $x \in (0, 1)$, where, for any real-valued function $g \in D[0, 1]$,

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

Thus, because the process $(\mathcal{T}_n \mathcal{K}_n)(x) = (\mathcal{T}_n \mathcal{K}_n^1)(x) + o_p(n^{-1/2})$, we can interpret $(\mathcal{T}_n \mathcal{K}_n)(x)$ as being the martingale innovation of $\mathcal{K}_n(x)$ and where the transformation $(\mathcal{T}_n g)(x)$ has the limiting version $(\mathcal{T} g)(x)$, defined as

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

$$\text{with } A_L(x) = \int_x^1 \left(\boldsymbol{P}_L(w) \boldsymbol{P}_L'(w) \right) f_X(w)\, dw.$$

This type of martingale transformation was proposed by [33] in the standard goodness-of-fit testing problem, and later used by [36, 49] or [16].

Finally, it is worth mentioning that in (5.3) we could 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 [30] 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 $F_X(\cdot)$ is the distribution function of X. Then we have the following.

THEOREM 1. *Assuming that H_0 holds true, under Conditions C1–C3, we have that* $\mathcal{M}_n(x) \overset{\text{weakly}}{\Rightarrow} \mathcal{U}(x)$, $x \in [0, 1]$.

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 analogously to v_i but where u_i is replaced with $\widehat{u}_i$ as defined in (5.1), yielding the statistic

$$\widetilde{\mathcal{M}}_n(x) =: \frac{1}{n^{1/2}} \sum_{i=1}^n \widehat{v}_i \mathcal{I}_i(x). \tag{5.5}$$

THEOREM 2. *Assuming that H_0 holds true, under Conditions C1–C3, we have that* $\widetilde{\mathcal{M}}_n(x) \overset{\text{weakly}}{\Rightarrow} \mathcal{U}(x)$, $x \in [0, 1]$.

We compute the estimate of σ_u^2 as $\breve{\sigma}_u^2 = 1/n \sum_{i=1}^n \breve{u}_i^2$, where $\breve{u}_i$ are the unconstrained residuals $\breve{u}_i = y_i - \breve{m}_{\mathcal{B}}(x_i; L)$. We then have the following.

PROPOSITION 1. *Under Conditions C1–C3*, $\breve{\sigma}_u^2 \xrightarrow{P} \sigma_u^2$.

COROLLARY 1. *Under H_0 and assuming Conditions C1–C3, for any continuous functional* $g : \mathbb{R} \to \mathbb{R}^+$, $g(\widetilde{\mathcal{M}}_n(x)/\breve{\sigma}_u) \xrightarrow{d} g(\mathcal{U}(x)/\sigma_u)$.

Corollary 1 forms a basis for testing for H_0 using functionals of $\widetilde{\mathcal{M}}_n(x)/\breve{\sigma}_u$, such as the ones given in (3.2).

5.2. *Nonlinear constraints on β_ℓ.* Now let us turn our attention to the *CUSUM* of "recursive" residuals with nonlinear constraints defining $S_{q,L}$. If none of the constraints are binding, the pivotal transformation proceeds in a manner akin to the previous section. Therefore, our focus will mainly be on scenarios where some constraints are binding, indicating that certain elements in $\widehat{b}$ are at the boundary of the set $S_{q,L}$.

To that end, we first describe $\widetilde{\boldsymbol{P}}_L(x)$. The main difference with the linear scenario is that the constraints described by the boundary of $S_{q,L}$ are now given by implicit functions. In particular, for the type of shapes in Example 2, we have that the boundary is given by implicit functions of the form $H(\beta_{\ell_0-2}, \beta_{\ell_0-1}, \beta_{\ell_0}) = 0$ whose explicit solutions $\beta_{\ell_0} = h(\beta_{\ell_0-2}, \beta_{\ell_0-1})$ are obtained either analytically or numerically.[9] 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 $\sum_{k=1}^L \beta_k p_k(x)$, we would consider the approximation given by

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

where $\beta_{-\ell_0} = (\beta_{\ell_0-2}, \beta_{\ell_0-1})$. Then $\widetilde{\boldsymbol{P}}_L(x)$ will be given by the vector of first derivatives of $g(x; \beta_{-\ell_0})$ with respect to the parameters. That is,

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

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

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

where $\widehat{v}_i = \widehat{u}_i - \widetilde{\boldsymbol{P}}_i'(\widehat{b}_{-\ell_0}) \mathcal{D}_n^+(i; \widehat{b}_{-\ell_0}) \sum_{k=1}^n \widetilde{\boldsymbol{P}}_k(\widehat{b}_{-\ell_0}) \widehat{u}_k \mathcal{J}_k(\widetilde{x}_i)$, and we set $\widetilde{\boldsymbol{P}}_i(\beta_{-\ell_0}) =: \widetilde{\boldsymbol{P}}_L(x_i; \beta_{-\ell_0})$, $\mathcal{D}_n(x; \beta_{-\ell_0}) =: \sum_{k=1}^n \widetilde{\boldsymbol{P}}_k(\beta_{-\ell_0}) \widetilde{\boldsymbol{P}}_k'(\beta_{-\ell_0}) \mathcal{J}_k(x)$ and $\mathcal{D}_n(i; \beta_{-\ell_0}) =: \mathcal{D}_n(\widetilde{x}_i; \beta_{-\ell_0})$ with $\widetilde{x}_i$ determined in the same way as in Section 5.1, and $\widehat{u}_i = y_i - g(x_i; \widehat{\beta}_{-\ell_0})$. By employing $\widetilde{p}_\ell(x_i; \beta_{-\ell_0})$ instead of $p_\ell(x_i)$, we have incorporated the 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* $\widetilde{\mathcal{M}}_n(x) \overset{weakly}{\Rightarrow} \mathcal{U}(x)$, $x \in [0, 1]$.

[9] Please see the Supplementary Material for more details on the form of constraints in Example 2.

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

$$H_1 : E[y|x] = m(x); \quad m(\cdot) \notin \mathcal{M}_0 \tag{5.6}$$

in a set $\mathcal{X}_1 =: [a_1, a_2] \subseteq \mathcal{X}$, which is assumed to be an interval for notational simplicity. Let $m^{\diamond}(\cdot)$ represent the best approximation in $\mathcal{M}_0$ to $m(\cdot)$ based on the $\mathcal{L}_2$-norm. Furthermore, let $\mathcal{X}_2$ denote the set where $m^{\diamond}(\cdot)$ resides on the "*boundary*"of the null hypothesis. When $\mathcal{M}_0$ is the set of nondecreasing functions, the "*boundary*" function is a constant. If we are interested in testing for convexity, the "*boundary*" function is a straight line. Analogously, the notion of "*boundary*" can be extended to any shape property. Note that $\mathcal{X}_2$ does not need to coincide with the set $\mathcal{X}_1$ even though it holds that $\mathcal{X}_2 \supseteq \mathcal{X}_1$. For instance, if $\mathcal{M}_0$ is the set of nondecreasing functions and

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

then $m^{\diamond}(x) = 0$ in $\mathcal{X}$ =[0, 1]. However, $\mathcal{X}_1 = (1/4, 3/4)$ whereas $\mathcal{X}_2 = [0, 1]$.

We can rewrite (5.6) as

$$E[y|x] = m(x) =: m^{\diamond}(x) + m_1^{\diamond}(x),$$

where by construction we can take $m_1^{\diamond}(x) = 0$ if $x \notin \mathcal{X}_2$.

To fix ideas, we shall explicitly consider the case when $\mathcal{M}_0$ is the set of nondecreasing functions, discussing more general scenarios in the Supplementary Material. Suppose that our optimization problem given in (2.8) ended up with $\widehat{b}_{\ell_0} = \cdots = \widehat{b}_{L_0}$, so that

$$\begin{aligned}\widehat{m}_{\mathcal{B}}(x_i; L) &= (\widehat{b}_1, \ldots, \widehat{b}_{\ell_0-1}, \widehat{b}_{\ell_0}, \widehat{b}_{L_0+1}, \ldots, \widehat{b}_L)\widetilde{\boldsymbol{P}}_i, \\ &= \sum_{k=1}^{\ell_0-1} \widehat{b}_k p_k(x_i) + \widehat{b}_{\ell_0} \sum_{k=\ell_0}^{L_0} p_k(x_i) + \sum_{k=L_0+1}^{L} \widehat{b}_k p_k(x_i),\end{aligned}$$

and where $\widetilde{\boldsymbol{P}}_L(x)$ in (2.9) is $\widetilde{\boldsymbol{P}}_L(x) = (p_1(x), \ldots, p_{\ell_0-1}(x), \widetilde{p}_{\ell_0}(x), p_{L_0+1}(x), \ldots, p_L(x))$. Due to the properties of the *B-splines*, $\widetilde{p}_{\ell_0}(x) = \sum_{k=\ell_0}^{L_0} p_k(x)$ is equal to 1 when $x \in [\frac{\ell_0}{L'}, \frac{L_0-q}{L'}]$. In this case, $m_{\mathcal{B}}(\cdot; L)$ in (2.3) becomes

$$\begin{aligned}m_{\mathcal{B}}^{\diamond}(x; L) &= \sum_{k=1}^{\overline{\ell}-1} \beta_k p_k(x) + \beta_{\overline{\ell}} \sum_{k=\overline{\ell}}^{\overline{L}} p_k(x) + \sum_{k=\overline{L}+1}^{L} \beta_k p_k(x) \\ &= (\beta_1, \ldots, \beta_{\bar{\ell}-1}, \beta_{\bar{\ell}}, \beta_{\bar{L}+1}, \ldots, \beta_L)\dot{\boldsymbol{P}}_L(x)\end{aligned}$$

with $\dot{\boldsymbol{P}}_L(x) = (p_1(x), \ldots, p_{\overline{\ell}-1}(x), \dot{p}_{\overline{\ell}}(x), p_{\overline{L}+1}(x), \ldots, p_L(x))$ and $\dot{p}_{\overline{\ell}}(x) = \sum_{k=\overline{\ell}}^{\overline{L}} p_k(x)$. Similar to above, $\dot{p}_{\overline{\ell}}(x)$ equals 1 when $x \in \mathcal{X}_2 = [\overline{\ell}/L', \overline{L} - q/L']$. 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 in the sense that $|\ell_0 - \overline{\ell}|/L' + |L_0 - \overline{L}|/L' = o_p(1)$.

Define

$$\mathcal{L}_L(x) = \int_{[0;x]\cap\mathcal{X}_2} \{m_1^{\diamond}(v) - \dot{\boldsymbol{P}}'_L(v)\widetilde{A}_L^{+}(v) \int_{[v;1]\cap\mathcal{X}_2} \dot{\boldsymbol{P}}_L(w) m_1^{\diamond}(w) f_X(w)\, dw\} f_X(v)\, dv,$$

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

$$m_1^{\diamond}(v) - \dot{\boldsymbol{P}}'_L(v)\widetilde{A}_L^{+}(v) \int_{[v;1]\cap\mathcal{X}'_2} \dot{\boldsymbol{P}}_L(w) m_1^{\diamond}(w) f_X(w)\, dw = 0. \tag{5.7}$$

The latter means that $m_1^{\diamond}(\cdot)$ belongs to the space spanned by $\dot{\boldsymbol{P}}_L(\cdot)$. This, however, is ruled out since $m_1^{\diamond}(x) \notin \mathcal{M}_0$ in $\mathcal{X}_2$ and any linear combination of $\dot{\boldsymbol{P}}_L(w)$ is a constant function in $\mathcal{X}_2$, and hence belonging to $\mathcal{M}_0$. We shall remark that $\dot{\boldsymbol{P}}_L(w)$ depends on $\mathcal{M}_0$, via the boundary component of $m^{\diamond}(\cdot)$.

PROPOSITION 2. *Assuming Conditions* C1–C3, *under* H_1 *in* (5.6),

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

where $\mathcal{V}(x)$ *is a nondegenerate random variable.*

The first consequence of Proposition 2 is that our tests would reject H_0 with a probability of 1 as $n \to \infty$. This is because $\mathcal{L}(x)$ is a nonzero function in $\mathcal{X}_2$, so that for any continuous functional $g : \mathbb{R} \to \mathbb{R}^+$ such that $1/g(x) \to 0$ as $|x| \to \infty$, standard arguments establish that $1/g(\widetilde{\mathcal{M}}_n(x)) \overset{p}{\to} 0$.

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

$$H_a \equiv \; E[y_i|x_i] =: m^{\diamond}(x_i) + n^{-1/2} m_1^{\diamond}(x_i),$$

where $m^{\diamond}(\cdot)$ and $m_1^{\diamond}(\cdot)$ satisfy respectively the same conditions as above. Then Proposition 2 yields that

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

REMARK 1. If $\widetilde{\boldsymbol{P}}_L(w) = \boldsymbol{P}_L(w)$, then (5.7) would be $o_p(1)$ regardless of whether $m(\cdot) \in \mathcal{M}_0$ or not. Consequently, a test based on $\widetilde{\mathcal{M}}_n(x)$ in (5.5) would lack power. Therefore, to ensure the test has power, it is crucial to use $\widetilde{\boldsymbol{P}}_L(w)$ when performing the pivotal transformation in (5.4) or (5.5).[10]

The Supplementary Material considers general cases when the null hypothesis is written in terms of the rth derivative of $m(\cdot)$, as in Example 1, and also the scenarios described in Example 2.

6. Bootstrap algorithm. We introduce a bootstrap algorithm for our test to address the small sample biases observed in our Monte Carlo experiments, despite the pivotal nature of our test. When the asymptotic distribution fails to adequately approximate the finite sample distribution, employing bootstrap algorithms is a standard approach to improve performance and provide small sample refinements. Our Monte Carlo simulations confirm that the bootstrap algorithm, to be described below, yields a superior finite sample approximation. We will utilize the fast WARP algorithm developed by [24] in our Monte Carlo experiments.

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

STEP 1 Compute the unconstrained residuals $\breve{u}_i = y_i - \breve{m}_{\mathcal{B}}(x_i; L)$, $i = 1, \ldots, n$, with $\breve{m}_{\mathcal{B}}(x_i; L)$ as defined in (2.5).

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

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

[10] See some additional discussion in the Supplementary Material.

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

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

$$\widehat{v}_i^* = \widehat{u}_i^* - \widetilde{\boldsymbol{P}}_i' \left(\sum_{k=1}^{n} \widetilde{\boldsymbol{P}}_k \widetilde{\boldsymbol{P}}_k' \mathcal{J}_k(x_i) \right)^+ \sum_{k=1}^{n} \widetilde{\boldsymbol{P}}_k \mathcal{J}_k(x_i) \widehat{u}_k^*,$$

and $\widehat{u}_i^* = y_i^* - \widetilde{\boldsymbol{P}}_i' (\sum_{k=1}^{n} \widetilde{\boldsymbol{P}}_k \widetilde{\boldsymbol{P}}_k')^+ \sum_{k=1}^{n} \widetilde{\boldsymbol{P}}_k y_k^*$, $i = 1, \ldots, n$.

THEOREM 4. *Under Conditions* C1–C3, *we have that* $g(\widetilde{\mathcal{M}}_n^*(x)) \xrightarrow{d} g(\mathcal{U}(x))$ *for any continuous* (*with probability* 1) *function* $g : \mathbb{R} \to \mathbb{R}^+$.

Corollary 2 shows that $\widehat{u}_i^*$ can be substituted with y_i^* in the computation of $\widetilde{\mathcal{M}}_n^*(x)$.

COROLLARY 2. *Under* $C1$–$C3$, *we have* $\widetilde{\mathcal{M}}_n^*(x) - \widetilde{\widetilde{\mathcal{M}}}_n^*(x) = 0$, *where*

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

7. Monte Carlo experiments. In our computational experiments, all the results are for cubic splines with varying numbers of knots. We include results for both *B-splines* and *P-splines*, with penalties on the second differences of coefficients determined through cross-validation as described in [22]. We use the modal value of these cross-validation parameters obtained across simulation draws. The tables provide the results of Kolmogorov–Smirnov ("*KS*"), Cramér–von Mises ("*CvM*") and Anderson–Darling ("*AD*") test statistics.[11] The number of equidistant knots (including boundary points) on the interval of interest is denoted as $L' + 1$. In all the scenarios, $X \sim \mathcal{U}[0, 1]$, $U \sim \mathcal{N}(0, \sigma^2)$ and $U \perp X$.

The rejection rates are based on the bootstrap critical values derived from the *WARP* bootstrap implementation where the demeaned residuals and x are drawn independently. The rejection rates are determined based on 2000 simulations. n represents the number of observations in each simulation.

Even though *B-splines* and *P-splines* deliver asymptotically equivalent results, the results in Scenario 3 suggest that in finite samples, *P-splines* yield better test power and stability across different L (or L') compared to *B-splines*. Therefore, we recommend using *P-splines* in practice.

The Supplementary Material contains additional results. Specifically, Scenario 4 gives additional illustrations of the power of the test, and Scenario 5 shows the performance of the test when nonlinear constraints on β are involved (log-convex regression function). It also shows the performance of our test for Scenarios 1–3 using asymptotic critical values. The results support our proposal to use bootstrap critical values in practice. The Supplementary Material also gives testing results for sample sizes $n = 100$ and $n = 200$. In addition, in the Supplementary Material we compare our test to those in [23, 27] and [50] when testing for monotonicity. Regarding the power of the test in Scenario 3, we find that when using *P-splines*, our test has a superior performance to those for large 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.

[11] We first center the empirical process in a way that it converges to a Brownian bridge indexed by the c.d.f. of X. Our additional simulations confirm that results with such a centering are very similar to results based on the original process without the centering.

TABLE 1
Tests for monotonically increasing regression function in Scenarios 1a and 1b

		Scenario 1a				Scenario 1b			
		B-splines		P-splines		B-splines		P-splines	
Setting	Method	10%	5%	10%	5%	10%	5%	10%	5%
$L' = 6$	KS	0.113	0.054	0.101	0.05	0.097	0.0525	0.1065	0.053
$n = 1000$	CvM	0.1035	0.0475	0.1005	0.044	0.1005	0.0515	0.0975	0.0445
$\sigma = 0.25$	AD	0.1065	0.054	0.104	0.052	0.0975	0.058	0.096	0.045
$L' = 14$	KS	0.0945	0.043	0.105	0.0555	0.0925	0.048	0.092	0.044
$n = 1000$	CvM	0.098	0.0425	0.0955	0.045	0.0885	0.046	0.0925	0.049
$\sigma = 0.25$	AD	0.093	0.043	0.096	0.049	0.0875	0.0465	0.0965	0.0425
$L' = 19$	KS	0.089	0.0485	0.101	0.058	0.093	0.04	0.098	0.042
$n = 1000$	CvM	0.105	0.0555	0.1025	0.0495	0.0915	0.049	0.098	0.0475
$\sigma = 0.25$	AD	0.1085	0.0545	0.1065	0.049	0.0885	0.043	0.1005	0.0465

SCENARIO 1 (Test for monotonicity). Regression functions on [0, 1] are

$$m(x) = x^{13/4}, \quad \text{(Scenario 1a)}$$

$$m(x) = -(x-0.5)^2 \cdot \mathcal{I}(x < 0.5) + (x-0.5)^2 \cdot \mathcal{I}(x \geq 0.5), \quad \text{(Scenario 1b)}.$$

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 Table 1.

Since it might be of interest to explore a broader spectrum of cases involving different levels of smoothness for $m(\cdot)$, in the Supplementary Material we consider two additional Scenarios 1c and 1d. In Scenario 1c, $m(\cdot)$ is smooth but its derivative is not Hölder continuous. In 1d, $m(\cdot)$ is infinitely differentiable.

SCENARIO 2 (Test for *U-shape*). The regression function is defined as

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

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 2. We use two different *B-splines*—one on $[0, s_0]$ and the other on $[s_0, 1]$. We join these *B-splines* continuously at s_0, and in another approach join them smoothly at s_0.

SCENARIO 3 (Analysis of power of the test). The regression function is

$$m(x) = x + 0.415e^{-ax^2}, \quad a > 0,$$

(its graph can be found in the Supplementary Material). We consider $a = 50$ and $a = 20$. In the latter case, the nonmonotonicity dip is smaller. These situations are deemed challenging for monotonicity tests as these functions are close to the set of monotone functions (in any conventional metric). As expected, the power of the test depends on the value of a and also depends on the variance of the error. The results for $\sigma = 0.25$ and $\sigma = 0.1$ are summarized in Table 3. The Supplementary Material contains additional results for $\sigma = 0.5$.

The power of monotonicity tests for such $m(\cdot)$ is examined in [23]. A similar regression function is studied in [8]. Note that [23] uses smaller sample sizes and also only $a = 50$ and $\sigma = 0.1$ to analyze power implications.

TABLE 2
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* $[s_0, 1]$

		Continuously joined				Smoothly joined			
		B-splines		P-splines		B-splines		P-splines	
Setting	Method	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

8. Conclusion. This paper proposes a methodology to test various shape properties of a regression function. The methodology involves applying a transformation to the empirical process of partial sums in a nonparametric setting, where *B-splines* or *P-splines* are used to approximate the functional space under the null. We prove that the proposed transformation eliminates the impact of nonparametric estimation and yields asymptotically pivotal testing. To the best of our knowledge, this paper is the first to implement this transformation in a nonparametric context.

In our main examples, we examine shape constraints expressed as inequality constraints on the coefficients of the approximating regression splines. The flexibility of our approach enables the simultaneous testing of multiple shape properties. When the inequality constraints are linear, the implementation becomes particularly straightforward, as is the case for shape properties expressed as linear inequality constraints on the derivatives.

TABLE 3
Tests for monotonicity in Scenario 3

		$a = 50$				$a = 20$			
		B-splines		P-splines		B-splines		P-splines	
Setting	Method	10%	5%	10%	5%	10%	5%	10%	5%
$L' = 6$	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
$L' = 12$	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
$L' = 12$	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

Funding. We appreciate financial support from STICERD.

SUPPLEMENTARY MATERIAL

Supplement to "Testing nonparametric shape restrictions" (DOI: 10.1214/23-AOS2311SUPP; .pdf). Supplementary information.

REFERENCES

[1] ABREVAYA, J. and JIANG, W. (2005). A nonparametric approach to measuring and testing curvature. *J. Bus. Econom. Statist.* **23** 1–19. MR2108688 https://doi.org/10.1198/073500104000000316

[2] AGARWAL, G. G. and STUDDEN, W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.* **8** 1307–1325. MR0594647

[3] AKAKPO, N., BALABDAOUI, F. and DUROT, C. (2014). Testing monotonicity via local least concave majorants. *Bernoulli* **20** 514–544. MR3178508 https://doi.org/10.3150/12-BEJ496

[4] ANDREWS, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica* **65** 1097–1128. MR1475076 https://doi.org/10.2307/2171880

[5] 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. *Ann. Statist.* **33** 214–257. MR2157802 https://doi.org/10.1214/009053604000000896

[6] BIRKE, M. and DETTE, H. (2007). Estimating a convex function in nonparametric regression. *Scand. J. Stat.* **34** 384–404. MR2346646 https://doi.org/10.1111/j.1467-9469.2006.00534.x

[7] BOLLAERTS, K., EILERS, P. H. C. and VAN MECHELEN, I. (2006). Simple and multiple P-splines regression with shape constraints. *Br. J. Math. Stat. Psychol.* **59** 451–469. MR2282224 https://doi.org/10.1348/000711005X84293

[8] BOWMAN, A. W., JONES, M. C. and GIJBELS, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.* **7** 489–500.

[9] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. MR0378310

[10] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894 https://doi.org/10.1214/aoms/1177728420

[11] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics* **6B**. Elsevier, Amsterdam.

[12] CHEN, X. and CHRISTENSEN, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *J. Econometrics* **188** 447–465. MR3383220 https://doi.org/10.1016/j.jeconom.2015.03.010

[13] CHETVERIKOV, D. (2019). Testing regression monotonicity in econometric models. *Econometric Theory* **35** 729–776. MR3974977 https://doi.org/10.1017/s0266466618000282

[14] CLAESKENS, G., KRIVOBOKOVA, T. and OPSOMER, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96** 529–544. MR2538755 https://doi.org/10.1093/biomet/asp035

[15] DE BOOR, C. (1978). *A Practical Guide to Splines*. *Applied Mathematical Sciences* **27**. Springer, New York-Berlin. MR0507062

[16] DELGADO, M. A., HIDALGO, J. and VELASCO, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Ann. Statist.* **33** 2568–2609. MR2253096 https://doi.org/10.1214/009053605000000606

[17] DETTE, H., NEUMEYER, N. and PILZ, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12** 469–490. MR2232727 https://doi.org/10.3150/bj/1151525131

[18] DIACK, C. A. T. and THOMAS-AGNAN, C. (1998). A nonparametric test of the non-convexity of regression. *J. Nonparametr. Stat.* **9** 335–362. MR1646904 https://doi.org/10.1080/10485259808832749

[19] DIERCKX, P. (1993). *Curve and Surface Fitting with Splines*. *Monographs on Numerical Analysis*. The Clarendon Press, Oxford University Press, New York. MR1218172

[20] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961 https://doi.org/10.1214/aos/996986504

[21] DURBIN, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. *Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics*, *No.* 9. SIAM, Philadelphia, PA. MR0305507

[22] EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11** 89–121. MR1435485 https://doi.org/10.1214/ss/1038425655

[23] GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082. MR1810919 https://doi.org/10.1214/aos/1015956707

[24] 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. MR3064050 https://doi.org/10.1017/S0266466612000655

[25] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742 https://doi.org/10.1214/aos/1015345958

[26] GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621 https://doi.org/10.1007/s00440-014-0595-3

[27] HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39. MR1762902 https://doi.org/10.1214/aos/1016120363

[28] HALL, P. and HUANG, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **29** 624–647. MR1865334 https://doi.org/10.1214/aos/1009210683

[29] HANSON, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann. Statist.* **4** 1038–1050. MR0426273

[30] HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York. MR1467237 https://doi.org/10.1007/b98818

[31] HILDRETH, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619. MR0065093

[32] JUDITSKY, A. and NEMIROVSKI, A. (2002). On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.* **30** 498–527. MR1902897 https://doi.org/10.1214/aos/1021379863

[33] KHMALADZE, È. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.* **26** 240–257.

[34] KOMAROVA, T. and HIDALGO, J. (2023). Supplement to "Testing nonparametric shape restrictions." https://doi.org/10.1214/23-AOS2311SUPP

[35] KOSTYSHAK, S. (2017). Non-parametric testing of U-shaped relationships. Working paper, University of Florida.

[36] KOUL, H. L. and STUTE, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27** 204–236. MR1701108 https://doi.org/10.1214/aos/1018031108

[37] LEE, J. and ROBINSON, P. M. (2016). Series estimation under cross-sectional dependence. *J. Econometrics* **190** 1–17. MR3425269 https://doi.org/10.1016/j.jeconom.2015.08.001

[38] LI, W., NAIK, D. and SWETITS, J. (1996). A data smoothing technique for piecewise convex/concave curves. *SIAM J. Sci. Comput.* **17** 517–537. MR1374294 https://doi.org/10.1137/S1064827592239822

[39] MAMMEN, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19** 724–740. MR1105841 https://doi.org/10.1214/aos/1176348117

[40] MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759. MR1105842 https://doi.org/10.1214/aos/1176348118

[41] MAMMEN, E. and THOMAS-AGNAN, C. (1999). Smoothing splines and shape restrictions. *Scand. J. Stat.* **26** 239–252. MR1707587 https://doi.org/10.1111/1467-9469.00147

[42] MEYER, M. C. (2008). Inference using shape-restricted regression splines. *Ann. Appl. Stat.* **2** 1013–1033. MR2516802 https://doi.org/10.1214/08-AOAS167

[43] MUKERJEE, H. (1988). Monotone nonparameteric regression. *Ann. Statist.* **16** 741–750. MR0947574 https://doi.org/10.1214/aos/1176350832

[44] NICULESCU, C. P. (2003). Convexity according to means. *Math. Inequal. Appl.* **6** 571–579. MR2013520 https://doi.org/10.7153/mia-06-53

[45] RAMSAY, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3** 425–461.

[46] SCHLEE, W. (1982). Nonparametric tests of the monotony and convexity of regression. In *Nonparametric Statistical Inference*, *Vol. I*, *II* (*Budapest*, 1980). *Colloquia Mathematica Societatis János Bolyai* **32** 823–836. North-Holland, Amsterdam. MR0719743

[47] SEN, P. K. (1982). Invariance principles for recursive residuals. *Ann. Statist.* **10** 307–312. MR0642743

[48] STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316 https://doi.org/10.1214/aos/1031833666

[49] STUTE, W., THIES, S. and ZHU, L.-X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26** 1916–1934. MR1673284 https://doi.org/10.1214/aos/1024691363

[50] WANG, J. C. and MEYER, M. C. (2011). Testing the monotonicity or convexity of a function using regression splines. *Canad. J. Statist.* **39** 89–107. MR2815340 https://doi.org/10.1002/cjs.10094

[51] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

[52] ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277 https://doi.org/10.1214/aos/1024691356