Specification Testing in Nonparametric Instrumental Quantile Regression *

Christoph Breunig *

Humboldt-Universität zu Berlin

March 20, 2018

There are many environments in econometrics which require nonseparable modeling of a structural disturbance. In a nonseparable model with endogenous regressors, key conditions are validity of instrumental variables and monotonicity of the model in a scalar unobservable. Under these conditions the nonseparable model is equivalent to an instrumental quantile regression model. A failure of the key conditions, however, makes instrumental quantile regression potentially inconsistent. This paper develops a methodology for testing the hypothesis whether the instrumental quantile regression model is correctly specified. Our test statistic is asymptotically normally distributed under correct specification and consistent against any alternative model. In addition, test statistics to justify the model simplification are established. Finite sample properties are examined in a Monte Carlo study and an empirical illustration.

Keywords: Nonparametric quantile regression, instrumental variable, specification test, local alternative, nonlinear inverse problem.

1. Introduction

Regression models that involve instrumental variables are widely used in economics to overcome endogeneity problems. In these models, assuming the structural disturbances to be additively separable implies that marginal effects do not depend on unobserved characteristics, which may be difficult to justify. This is why their nonseparable extension has received a lot of attention recently. Under certain key conditions, the nonseparable model

*Parts of this paper derive from my doctoral dissertation, completed under the guidance of Enno Mammen. I would like to thank Liangjun Su and three anonymous referees for excellent comments and suggestions that greatly improved the paper. I also thank seminar participants at Boston College, Bristol, Mannheim, Toulouse School of Economics, University College London, WIAS Berlin, and Yale. This work was supported by the DFG-SNF research group FOR916. I am also grateful for support and hospitality of the Cowles Foundation.

*Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany, e-mail: christoph.breunig@hu-berlin.de

is equivalent to an instrumental quantile regression model. These conditions are validity of instruments and monotonicity of the model in a scalar unobservable. If one of these conditions is violated, however, the quantile regression representation is misspecified.
In this paper, we propose a specification test of the instrumental quantile regression model

$$Y = \varphi(Z, q) + U(q) \quad \text{where} \quad \mathbb{P}(U(q) \leqslant 0|W) = q \tag{1.1}$$

for each $0 < q < 1$, where Y is a scalar dependent variable, Z a vector of potentially endogenous regressors, W a vector of instruments, and $U(q)$ an unobservable disturbance.[1] This quantile regression model is equivalent to a nonseparable model (cf. Horowitz and Lee [2007]) given by

$$Y = \varphi(Z, V) \tag{1.2}$$

with

(a.1) the instrumental variable W is independent of V,

(a.2) the function φ is strictly monotonic increasing in the scalar disturbance V, and

(a.3) $V \sim \mathcal{U}(0, 1)$.

Condition *(a.3)* can be assumed without loss of generality if V is continuously distributed with positive density on its support which we assume to hold throughout the paper. The quantile regression model (1.1) for all $0 < q < 1$ is thus misspecified if in its nonseparable version (1.2) the instrument is not valid, that is, W is not independent of V, or the function φ is not monotonic in V.
Specification testing in instrumental variable models is a subject of considerable literature. In the context of nonparametric instrumental mean regression $Y = g(Z) + U$ with $\mathbb{E}[U|W] = 0$, tests for correct specification have been proposed by Gagliardini and Scaillet [2007], Horowitz [2012], and Breunig [2015]. These tests are, however, not robust against potential nonseparability of the structural disturbance. On the other hand, by considering the nonseparable model (1.2) with conditions *(a.1)–(a.3)* not only a failure of the exclusion restriction of the instruments might lead to a misspecified model. Indeed, as argued by Hoderlein and Mammen [2007], in certain applications, such as consumer demand, the monotonicity restriction *(a.2)* might be highly unrealistic. As such, providing a specification test of model (1.2) together with conditions *(a.1)–(a.3)* seems paramount but, as far as we know, has not yet been addressed in the literature.
Research on identification and estimation in nonparametric instrumental quantile regression has been active in the last decade. Chesher [2003] establishes nonparametric identification of derivatives of the unknown functions in a triangular array structure. Chernozhukov and Hansen [2005] and Chernozhukov et al. [2007] give identification conditions and develop a nonparametric minimum distance estimator. Sufficient conditions for local identification are given by Chen et al. [2014]. Horowitz and Lee [2007] propose an estimator based on Tikhonov regularization, Chen and Pouzo [2012] study penalized sieve minimum distance estimation, and Dunker et al. [2014] consider an iteratively regularized Gauß-Newton method. Further, Gagliardini and Scaillet [2012] obtain asymptotic distribution results of a Tikhonov regularized estimator. There is also a large literature on testing quantile regression models with exogenous covariates. In this context particularly relevant is quantile regression testing using an infinite number of quantiles for parametric functions, see Escanciano and Velasco [2010] and, in the nonparametric context, Escanciano and Goh [2014].

In instrumental quantile regression (1.1) for a fixed quantile $0 < q < 1$, Horowitz and Lee [2009] established a test of parametric specification of φ. Chen and Pouzo [2015] consider functionals of semi/nonparametric conditional moment restrictions with possibly nonsmooth generalized residuals. A test of monotonicity in unobservables of φ has been proposed by Hoderlein et al. [2016] but requires conditional exogeneity of Z and hence, is not related to instrumental variables methodology. Recently and independently of this paper, Fève et al. [2016] developed a test of whether Z is independent of the nonseparable disturbance V in the model (1.2).

Our test statistic is based on the L^2–norm of the empirical conditional quantile restriction and involves sieve methodology. The sieve approach makes the statistic easy to implement and further, is convenient to impose additional constraints on the structural function φ. As an example, we discuss a test of additivity of φ with respect to the vector of regressors Z. In addition, we establish a test statistic for testing exogeneity which is robust against nonseparability. More precisely, we establish a test of exogeneity of the regressors Z at some quantile $0 < q < 1$, that is, whether $\mathbb{P}(Y \leqslant \varphi(Z, q)|Z) = q$. This extends the results on nonparametric tests of exogeneity in mean regression suggested by Blundell and Horowitz [2007] and Breunig [2015] to the quantile regression case.

It should also be noted that the test proposed in this paper is a joint test of monotonicity and instrument validity. This is the nature of many nonparametric tests, see, for instance, Chiappori et al. [2015] or Lewbel et al. [2015]. On the other hand, we show in this paper how the sign of $\mathbb{P}(Y \leqslant \varphi(Z, q)|W) - q$ can be exploited to infer on validity of the instrumental variables. As such, in many cases it is possible to detect the cause of a rejection of our test.

We establish the asymptotic distribution of our test statistic under the null hypothesis and its consistency against fixed alternatives. We study the power of our test against a sequence of local alternatives. By Monte Carlo simulations we demonstrate the power properties of our test in finite samples. As an empirical illustration, we study a nonseparable model of the effects of class size on test scores of 4th grade students in Israel. We reject the hypothesis of exogeneity of class size but fail to reject the instrumental variable model.

The remainder of this work is organized as follows. In Section 2, we propose a test statistic and obtain its asymptotic distribution. We further establish consistency of our test. The power of the test is judged by considering a sequence of local alternatives. Section 3 gives several extensions of the previous results. In Section 4 and 5 we study the finite sample properties of our test and give an empirical illustration. All proofs can be found in the appendix.

2. The test statistic and its asymptotic properties

This section begins with the definition of the test statistic and states assumptions required to obtain its asymptotic distribution under the null hypothesis. Moreover, we study power and consistency properties of our test.

2.1. Definition of the test statistic

The quantile regression model (1.1) leads to a nonlinear operator equation, as we see in the following. Let Φ be a Banach space endowed with the norm $\|\phi\|_{Z,p} := (\mathbb{E}\,|\phi(Z)|^p)^{1/p}$ for some integer $p > 0$ and if $p = \infty$ then $\|\phi\|_{Z,\infty} := \sup_z |\phi(z)|$. For simplicity let $\|\phi\|_Z := \|\phi\|_{Z,2}$. Further, let us introduce the Hilbert space $L^2_W := \{\psi : \|\psi\|^2_W := \mathbb{E}\,|\psi(W)|^2 < \infty\}$. We

define a nonlinear operator $\mathcal{T} : \Phi \to L^2_W$ with

$$\mathcal{T}\phi := \mathbb{E}[\mathbb{1}\{Y \leqslant \phi(Z)\}|W] \tag{2.1}$$

for any $\phi \in \Phi$ where $\mathbb{1}$ denotes the indicator function. Thereby, model (1.1) can be rewritten as the operator equation $\mathcal{T}\varphi_q = q$ with $\varphi_q(\cdot) := \varphi(\cdot, q)$ for all $0 < q < 1$.
In many economic applications, for instance when estimating a demand function or Engel curves, the structural function of interest may be assumed to be smooth. This *a priori* knowledge is captured by a set $\mathcal{B} \subset \Phi$ which we introduce below. The set $\mathcal{B}$ may also contain constraints on the function φ_q such as monotonicity, concavity/convexity or additivity (see also Section 3.2) and can also ensure uniqueness of φ_q (see Example 2.1 below). Let us introduce the set $\mathcal{B}^{(0,1)} = \{\phi : \phi(\cdot, q) \in \mathcal{B} \text{ for all } q \in (0,1)\}$. We consider the null hypothesis

$$H_0 : \text{there exists a function } \varphi \in \mathcal{B}^{(0,1)} \text{ such that } \mathcal{T}\varphi_q = q \text{ for all } q \in (0,1). \tag{2.2}$$

The alternative is that there exists no function $\varphi \in \mathcal{B}^{(0,1)}$ solving $\mathcal{T}\varphi_q = q$ for all $q \in (0,1)$. We construct in the following a test statistic based on the L^2–distance. Throughout the paper, we assume that an independent and identically distributed n-sample of (Y, Z, W) is available. Let $\{f_j\}_{j \geqslant 1}$ be a sequence of approximating functions in L^2_W. Then, for any integer $k \geqslant 1$ we denote $f_{\underline{k}}(\cdot) = (f_1(\cdot), \ldots, f_k(\cdot))^t$ and $\mathbf{W}_k = \big(f_{\underline{k}}(W_1), \ldots, f_{\underline{k}}(W_n)\big)^t$ which is a $n \times k$ matrix. A series least square estimator of $\mathbb{E}[\mathbb{1}\{Y \leqslant \phi(Z)\} - q|W = \cdot]$ then writes

$$f_{\underline{l_n}}(\cdot)^t(\mathbf{W}_{l_n}^t \mathbf{W}_{l_n})^- \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q) f_{\underline{l_n}}(W_i)$$

where $(\cdot)^-$ denotes a generalized inverse. Further, we define the sieve least square estimator of φ_q by

$$\widehat{\varphi}_{qn} \in \operatorname*{argmin}_{\phi \in \mathcal{B}_{k_n}} \Big(\sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q) f_{\underline{l_n}}(W_i) \Big)^t (\mathbf{W}_{l_n}^t \mathbf{W}_{l_n})^- \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q) f_{\underline{l_n}}(W_i) \tag{2.3}$$

where $\mathcal{B}_{k_n}$ is a k_n–dimensional sieve space that becomes dense in $\mathcal{B}$ as the sample size n tends to infinity. If $\mathcal{B}$ contains additional constraints then these are imposed in $\mathcal{B}_{k_n}$ on the finite dimensional functions. Here, k_n and l_n grow with sample size n. Clearly, $k_n \leqslant l_n$ for each n is required and in our simulations we choose $l_n = Ck_n$ for some constant $C > 1$ (see also Chen and Christensen [2015] in case of nonparametric instrumental mean regression). The estimator $\widehat{\varphi}_{qn}$ is a simplified version of the penalized sieve minimum distance estimator suggested by Chen and Pouzo [2012].
The test statistic is then given by

$$S_n = \int_0^1 \Big(\sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q) f_{\underline{m_n}}(W_i) \Big)^t (\mathbf{W}_{m_n}^t \mathbf{W}_{m_n})^- \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q) f_{\underline{m_n}}(W_i) dq \tag{2.4}$$

where m_n grows with sample size n. As the test is one sided, we reject the null hypothesis at level α when the standardized version of S_n, namely $3\sqrt{5/m_n}\big(S_n - m_n/6\big)$, is larger than

the $(1-\alpha)$–quantile of $\mathcal{N}(0,1)$. The asymptotic distribution of S_n is derived below under mild restrictions on the dimension parameters k_n, l_n, and m_n. We require that the number of unconditional moment restrictions determined by m_n is asymptotically larger than the dimension of the sieve space $\mathcal{B}_{k_n}$. This corresponds to the test of overidentifying restrictions in parametric models. In contrast to the parametric setting, however, also the number of unconditional moment restrictions used to construct the estimator (determined by l_n) must be asymptotically smaller than the number of moment restrictions used for the test statistic. This ensures that the estimation error in the test statistic becomes asymptotically negligible as we see below.

Our test statistic builds on the nonparametric specification test in instrumental mean regression suggested by Breunig [2015]. Testing in instrumental quantile regression, on the other hand, requires a different methodology. First, the test statistic is a discontinuous function of the unknown structural effect φ_q. Second, instrumental quantile regression leads to a nonlinear inverse problem and hence deriving asymptotic results is more challenging. Third, to verify the conditional moment restrictions for all quantiles we need to integrate over them. In the appendix, we show that the mapping $q \mapsto \varphi_q$ is continuous under mild assumptions. This justifies the use of our L^2–type rather than a sup norm statistic.

2.2. Assumptions and notation.

In order to obtain our asymptotic result we state the following assumptions. Our first assumption gathers conditions which we require for the basis functions $\{f_j\}_{j\geqslant 1}$. In the following, the supports $\mathcal{Z}$ of Z and $\mathcal{W}$ of W are assumed to be bounded (see also Assumption 4). The probability density function (p.d.f.) of W, denoted by p_W, is assumed to be uniformly bounded from above and away from zero.

Assumption 1. *(i) There exists a constant $C > 0$ and a sequence of positive integers $(m_n)_{n\geqslant 1}$ satisfying $\sup_{w\in\mathcal{W}} \|f_{\underline{m_n}}(w)\|^2 \leqslant C m_n$. (ii) The smallest eigenvalue of the matrix $\mathbb{E}[f_{\underline{m}}(W) f_{\underline{m}}(W)^t]$ is bounded away from zero uniformly in m.*

Assumption 1 (i) holds for sufficiently large C if $\{f_j\}_{j\geqslant 1}$ are trigonometric basis functions, B-splines, or wavelets. Assumption 1 (ii) is satisfied if the marginal density of W is uniformly bounded away from zero on $\mathcal{W}$ and $f_{\underline{m_n}}$ forms a vector of orthonormal basis functions. In the following, for any $\phi \in \mathcal{B}^{(0,1)}$ we denote $\phi_q(\cdot) := \phi(\cdot, q)$ for all $0 < q < 1$. In the following, we denote the Fréchet derivative of $\mathcal{T}$ at φ_q by

$$T_q\phi := \mathbb{E}\left[p_{Y|Z,W}\big(\varphi(Z,q), Z, W\big)\phi(Z)\big|W\right]$$

where $p_{Y|Z,W}$ denotes the density of Y conditional on (Z,W). In the following, we denote $|||\phi|||_{Z,p} = \big(\int_0^1 \|\phi(\cdot,q)\|_{Z,p}^p dq\big)^{1/p}$ and $|||\psi|||_W = \big(\int_0^1 \|\psi(\cdot,q)\|_W^2 dq\big)^{1/2}$ for functions $\phi(\cdot,q) \in \Phi$ and $\psi(\cdot,q) \in L^2_W$ for all $q \in (0,1)$.

Assumption 2. *(i) If $|||\mathcal{T}\phi - \mathcal{T}\varphi|||_W^2 = 0$ for some function $\phi \in \mathcal{B}^{(0,1)}$ then it holds $|||\phi - \varphi|||_{Z,p}^2 = 0$. (ii) There exists some constant $0 < \eta < 1$ such that for all $0 < q < 1$ and all functions $\phi \in \{\phi \in \mathcal{B} : \|\phi - \varphi_q\|_{Z,p} \leqslant \varepsilon\}$ for some $\varepsilon > 0$ it holds*

$$\|\mathcal{T}\phi - \mathcal{T}\varphi_q - T_q(\phi - \varphi_q)\|_W \leqslant \eta \|T_q(\phi - \varphi_q)\|_W. \tag{2.5}$$

Assumption 2 (i) ensures identification of φ_q for almost all $0 < q < 1$ on the set $\mathcal{B}$ which we introduce below. Assumption 2 (ii) specifies an upper bound on the Taylor remainder

of $\mathcal{T}$ in a small neighborhood around φ_q. It is also known as the tangential cone condition and frequently used in the analysis of nonlinear operator equations (cf. Hanke et al. [1995] or Dunker et al. [2014] in case of instrumental variable estimation). We provide sufficient conditions for the tangential cone condition in Example 2.1 below and refer to Chen et al. [2014] for further discussions.

Assumption 3. *There exists a sequence $(r_n)_{n\geqslant 1}$ with $r_n = o(1)$ such that for constants $C > 0$ and $\kappa \in (0,1]$ it holds*

$$\max_{1\leqslant j\leqslant m_n} \mathbb{E}\Big[\int_0^1 \sup_{\phi\in\mathcal{B}_n} \big|\, \mathbb{1}\{Y \leqslant \phi(Z,q)\} - \mathbb{1}\{Y \leqslant \varphi(Z,q)\}\big|^2 dq\, f_j^2(W)\Big] \leqslant C r_n^{2\kappa} \tag{2.6}$$

where $\mathcal{B}_n := \{\phi \in \mathcal{B}^{(0,1)} :\, |||\phi - \varphi|||_{Z,p}^2 \leqslant r_n^2\}$.

Assumption 3 states that the function $\varphi_q \mapsto (\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q) f_j(W)$, $1 \leqslant j \leqslant m_n$, is locally uniformly L^2_W continuous for almost all $0 < q < 1$. This condition has also been exploited by Chen et al. [2003] (Theorem 3), Chen [2007] (Lemma 4.2 (i)) or Chen and Pouzo [2012] (Remark c.1). Example 2.2 below gives primitive conditions under which Assumption 3 holds true.

Let $\mathcal{Z} \subset \mathbb{R}^{d_z}$ and for any vector of nonnegative integers $k = (k_1, \ldots, k_{d_z})$ define $|k| = \sum_{j=1}^{d_z} k_j$ and $D^k = \delta^{|k|}/(\delta z_1^{k_1} \ldots \delta z_{d_z}^{k_{d_z}})$. For some integer $p > 0$ we define the norms

$$\|\phi\|_{\alpha,p} = \Big(\sum_{|k|\leqslant\alpha+\alpha_0} \int_{\mathcal{Z}} \big|D^k\phi(z)\big|^p dz\Big)^{1/p}, \quad \|\phi\|_{\alpha,\infty} = \max_{|k|\leqslant\alpha} \sup_{z\in\mathcal{Z}} \big|D^k\phi(z)\big|$$

where α and α_0 are positive integers. We denote the Sobolev spaces associated with the norm $\|\cdot\|_{\alpha,p}$ by

$$W^{\alpha,p} := \{\phi : \mathcal{Z} \to \mathbb{R} :\, \|\phi\|_{\alpha,p} < \infty\}. \tag{2.7}$$

For some constant $\rho > 0$, define $\mathcal{B}$ as the Sobolev ellipsoid of radius ρ given by

$$\mathcal{B} := \mathcal{B}(\alpha) := \{\phi \in W^{\alpha,p} :\, \|\phi\|_{\alpha,p} \leqslant \rho\}. \tag{2.8}$$

While the set of structural functions $\mathcal{B}$ is compact under the norm $\|\cdot\|_{\alpha,\infty}$ (see Lemma A.2 of Santos [2012]) it is not compact under the norm $\|\cdot\|_Z$ (we also refer to Chen and Pouzo [2012] for a further discussion). On the other hand, our sieve space $\mathcal{B}_{k_n}$ used to approximate $\mathcal{B}$ is compact under $\|\cdot\|_Z$ and thus, penalization is not necessary for consistent estimation (see also Chen and Pouzo [2012]). Also additional constraints such as monotonicity can be imposed by $\mathcal{B} = \{\phi \in W^{\alpha,p} :\, \|\phi\|_{\alpha,p} \leqslant \rho,\, \inf_{z\in\mathcal{Z}} \phi'(z) > 0\}$ for scalar z. Such a monotonicty constraint does not necessarily lead to faster rates of convergence, in contrast to an additivity restriction on φ_q. Consequently, we do not treat shape restrictions like monotonicty explicitly but only discuss a test of additivity in Section 3.2. In this context, we also refer to Chetverikov and Wilhelm [2015] for using shape restriction for sieve estimation in instrumental mean regression. The following assumption gathers regularity conditions imposed on the structural functions φ and the supports $\mathcal{Z}$ and $\mathcal{W}$.

Assumption 4. *(i) Let $\alpha_0 > d_z/p$ and $\alpha > d_z/\kappa$. (ii) $\mathcal{Z}$ is bounded, convex and satisfies a uniform cone property. (iii) $\mathcal{W}$ is bounded. (iv) The marginal density of W, denoted by p_W, is bounded from above and uniformly bounded away from zero on $\mathcal{W}$. (v) $p_{Y|Z,W}(\cdot, Z, W)$ is bounded from above.*

Assumption 4 (i) requires α to be large if (2.6) holds only for small $\kappa > 0$ or the dimension d_z is large. Assumption 4 (ii) imposes a weak regularity condition on the shape of $\mathcal{Z}$. For the uniform cone property see, for instance, Paragraph 4.4 in Adams and Fournier [2003]. This property was also used by Santos [2012]. Assumption 4 (v) ensures that $\|T_q\phi\|_W \leqslant C\|\phi\|_Z$ for all $\phi \in L^2_Z$ and some constant $C > 0$.

Example 2.1 (Primitive Conditions for Assumption 2)**.** *Let Φ coincide with the Hilbert space $L^2_Z := \{\phi : \|\phi\|_Z < \infty\}$. If for any $0 < q < 1$ the operator T_q is compact then there exists an orthonormal basis in L^2_Z denoted by $\{e_j\}_{j\geqslant 1}$ satisfying $\|T_q\phi\|^2_W = \sum_{j=1}^{\infty} s^2_{qj}\,\mathbb{E}[\phi(Z)e_j(Z)]^2$ where $(s_{qj})_{j\geqslant 1}$ are the singular values of T_q. If*

$$\mathcal{B} \subset \mathcal{B}_{source,q} := \left\{ \phi \in L^2_Z : \sum_{j=1}^{\infty} s^{-2}_{qj}\,\mathbb{E}[(\phi(Z) - \varphi(Z,q))e_j(Z)]^2 < c_0 \right\}$$

for some constant $c_0 > 0$ then, under mild assumptions on the joint distribution of (Y, Z, W), the function φ_q is identified on $\mathcal{B}$ (cf. Theorem 6 of Chen et al. [2014]). A similar restriction was also imposed by Horowitz and Lee [2007]. If $\mathcal{B} \subset \bigcap_{q\in(0,1)} \mathcal{B}_{source,q}$ then Assumption 2 (i) holds true. Under further assumptions, imposing bounds on the generalized Fourier coefficients is equivalent to imposing smoothness restrictions. To illustrate this relation let Z be a scalar uniformly distributed random variable and assume $s_{qj} = j^{-\zeta}$, $j \geqslant 1$, for some constant $\zeta > 0$. In this case, if $\{e_j\}_{j\geqslant 1}$ are the usual trigonometric basis functions then $\mathcal{B}_{source,q}$ coincides with the Sobolev space of ζ–times differential functions with periodic boundary conditions, while if $s^2_{qj} = \exp(-j^{2\zeta})$, $j \geqslant 1$ and $\zeta > 1$, $\mathcal{B}_{source,q}$ contains only analytic functions (see also Kress [1989]). In this sense, $\mathcal{B}_{source,q}$ links the smoothness of $\phi - \varphi_q$ to the degree of ill-posedness determined by the degree of decay of $(s_{qj})_{j\geqslant 1}$, which is also known as a so-called source condition (cf. Chen and Reiß [2011] or Dunker et al. [2014] for a further discussion).
Under the singular value decomposition of T_q it is also possible to provide primitive conditions for the tangential cone condition (2.5). Assume that the conditional p.d.f. of Y given (Z, W), denoted by $p_{Y|Z,W}$, is continuously differentiable with $|\partial p_{Y|Z,W}(\cdot, Z, W)/\partial y| \leqslant c_1$ and the conditional p.d.f. of Z given W satisfies $p_{Z|W}(\cdot, W) \leqslant c_2 p_Z(\cdot)$, for some constants $c_1, c_2 > 0$. Then by Theorem 6 of Chen et al. [2014] it holds

$$\|\mathcal{T}\phi - \mathcal{T}\varphi_q - T_q(\phi - \varphi_q)\|_W \leqslant c_1\, c_2\, \|\phi - \varphi_q\|^2_Z. \tag{2.9}$$

We further obtain for all $\phi \in \mathcal{B}_{source,q}$ by making use of the Cauchy-Schwarz inequality

$$\begin{aligned}
\|\phi - \varphi_q\|^2_Z &= \sum_{j=1}^{\infty} \frac{s_{qj}}{s_{qj}}\,\mathbb{E}[(\phi(Z) - \varphi(Z,q))e_j(Z)]^2 \\
&\leqslant \Big(\sum_{j=1}^{\infty} s^{-2}_{qj}\,\mathbb{E}[(\phi(Z) - \varphi(Z,q))e_j(Z)]^2 \Big)^{1/2} \Big(\sum_{j=1}^{\infty} s^2_{qj}\,\mathbb{E}[(\phi(Z) - \varphi(Z,q))e_j(Z)]^2 \Big)^{1/2} \\
&\leqslant c_0^{1/2}\, \|T_q(\phi - \varphi_q)\|_W.
\end{aligned}$$

Consequently, the tangential cone condition (2.5) is satisfied if we assume $c_0^{1/2}\, c_1\, c_2 < 1$. We also note that for our test of exogeneity in Section (3.1) only the weaker condition (2.9) is required. □

Example 2.2 (Primitive Conditions for Assumption 3)**.** *Let $F_{Y|ZW}$ denote the cumulative distribution function of Y given (Z, W) and assume that it is Lipschitz continuous with constant $C_L > 0$, that is, $|F_{Y|ZW}(y) - F_{Y|ZW}(y')| \leqslant C_L|y - y'|$ for all (y, y'). Due to Assumption 4 the Sobolev space $W^{\alpha,p}$ can be embedded in $W^{\alpha,\infty}$ (cf. Theorem 6 of Adams and Fournier [2003]). In particular, the supremum norm is bounded on $\mathcal{B}$ and moreover, Assumption 3 holds true. Indeed, $\int_0^1 \|\phi_q - \varphi_q\|_\infty^2 dq \leqslant r_n^2$ implies $\|\phi_q - \varphi_q\|_\infty \leqslant c\, r_n$ for almost all $0 < q < 1$ and some constant $c > 0$. Hence, $\varphi(Z, q) - c\, r_n \leqslant \phi(Z, q) \leqslant \varphi(Z, q) + c\, r_n$ for almost all $0 < q < 1$ and following Chen et al. [2003] (page 1599 – 1600) we observe*

$$
\begin{aligned}
\mathbb{E}\Big[\int_0^1 \max_{\phi\in\mathcal{B}_n} \big(\mathbb{1}\{Y \leqslant \phi(Z,q)\} - \mathbb{1}\{Y \leqslant \varphi(Z,q)\}\big)^2 dq \,\Big| W\Big] \\
&\leqslant \int_0^1 \mathbb{E}\Big[\mathbb{1}\left\{Y \leqslant \varphi(Z,q) + c\, r_n\right\} - \mathbb{1}\left\{Y \leqslant \varphi(Z,q) - c\, r_n\right\}\Big| W\Big] dq \\
&= \int_0^1 \mathbb{E}\Big[F_{Y|ZW}\big(\varphi(Z,q) + c\, r_n\big) - F_{Y|ZW}\big(\varphi(Z,q) - c\, r_n\big)\Big| W\Big] dq \\
&\leqslant C_L\, c\, r_n
\end{aligned}
$$

which implies Assumption 3 with $\kappa = 1/2$. □

Remark 2.1 (Local Overidentification)**.** *The tangential cone condition together with the assumption of injectivity of the Fréchet derivative T_q implies local identification in nonparametric instrumental quantile regression (see Dunker et al. [2014] or Chen et al. [2014]). In this example, we discuss local overidentification restrictions in nonparameteric instrumental quantile regression. As Chen and Santos [2015] point out in their Example 5.2, the range of the Fréchet derivative T_q, is given by*

$$\mathcal{R}_q = \big\{\psi \in L_W^2 : \psi = T_q\phi \text{ for some } \phi \in L_Z^2\big\}.$$

Local overidentification corresponds to the case where the closure of the range $\mathcal{R}_q$ is a strict subset of L_W^2. Under the null hypothesis, the class of structural functions φ is restricted to belong to an ellipsoid $\mathcal{B}$ and thus, consider for each q:

$$\mathcal{R}_q(\mathcal{B}) = \big\{\psi \in L_W^2 : \psi = T_q\phi \text{ for some } \phi \in \mathcal{B}\big\}.$$

Consequently, under mild restrictions on the ellipsoid $\mathcal{B}$ local overidentification can be ensured and the class of functions in the alternative model is not empty. □

The next result formalizes the discussion of the previous remark by providing minimal regularity conditions on the function space $\mathcal{B}$ to ensure overidentification.

Proposition 2.1. *Let Φ coincide with the Hilbert space L_Z^2. For any $0 < q < 1$, assume that the Fréchet derivative T_q is compact with singular values having either a polynomial or exponential decay with eigenfunctions e_j, $j \geqslant 1$. For some $\delta > 0$ assume that $\mathcal{B} \subset \mathcal{B}_\delta := \big\{\phi \in L_Z^2 : \sum_{j=1}^\infty j^{2\delta}\, \mathbb{E}[\phi(Z)e_j(Z)]^2 < C\big\}$ for some constant $C > 0$. Then we have local identification, i.e., the closure of $\mathcal{R}_q(\mathcal{B})$ is a strict subset of L_W^2.*

Note that the restriction on the generalized Fourier coefficients for the functions in $\mathcal{B}_\delta$ is not very restrictive for small $\delta > 0$. In the setting of Example 2.1, where $\{e_j\}_{j\geqslant 1}$ are trigonometric basis functions and $Z \sim \mathcal{U}(0, 1)$, the function class $\mathcal{B}_\delta$ contains δ–times differential functions. In particular, we emphasize that the result in Proposition 2.1 holds for any $\delta > 0$ and thus, $\mathcal{B}$ can be arbitrarily close to the compact set of functions $\big\{\phi \in L_Z^2 : \|\phi\|_Z^2 < C\big\}$, which follows by employing a spectral decomposition of T_q.

Notation For any $\phi \in \mathcal{B}$ we introduce $\Pi_{k_n}\phi \in \mathcal{B}_{k_n}$ satisfying $\|\Pi_{k_n}\phi - \phi\|_{Z,p} = o(1)$. Further, we define

$$\omega_n = \max\Big(n^{-1}l_n, \max_{\phi\in\mathcal{B}_{k_n}} \sum_{j>l_n} \mathbb{E}[(\mathcal{T}\phi(W) - q)f_j(W)]^2, |||T.(\Pi_{k_n}\varphi - \varphi)\,|||_W^2\Big).$$

The rate ω_n captures the variance and bias part for estimating $\mathcal{T}\phi$ for a fixed function ϕ and also contains the bias for approximating the structural function φ in the weak norm induced by the Fréchet derivative of $\mathcal{T}$. Following Chen and Pouzo [2012] we introduce the sieve measure of local ill-posedness by

$$\tau_{k_n} := \max_{\phi\in\mathcal{A}_{k_n}} \Big(\frac{|||\phi - \varphi|||_{Z,p}^2}{|||T.(\phi - \varphi)|||_W^2}\Big)$$

where $\mathcal{A}_{k_n} = \Big\{\phi \in \mathcal{B}_{k_n}^{(0,1)} : |||T.(\phi - \varphi)|||_W^2 > 0\Big\}$. We write $a_n \sim b_n$ when there exist constants $c, c' > 0$ such that $cb_n \leqslant a_n \leqslant c'b_n$ for sufficiently large n.

2.3. Asymptotic distribution under the null hypothesis

The following theorem establishes asymptotic normality of the test statistic S_n after standardization under the null hypothesis H_0.

Theorem 2.2. *Let Assumptions 1–4 be satisfied. Assume that*

$$m_n^{-1} = o(1), \quad m_n = o(n^{1/2}) \tag{2.10}$$

and in addition

$$n\omega_n = o(\sqrt{m_n}) \text{ and } |||\,\Pi_{k_n}\varphi - \varphi\,|||_{Z,p}^2 + \tau_{k_n}\omega_n = o\big(m_n^{-(1+\epsilon)/\kappa}\big) \tag{2.11}$$

for some $\epsilon > 0$. Then we have under H_0

$$3\sqrt{5/m_n}\big(S_n - m_n/6\big) \xrightarrow{d} \mathcal{N}(0,1).$$

To motivate the constants in the sieve mean and variance, respectively, we observe

$$\int_0^1 \mathbb{E}[(\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q)^2|W]dq = \int_0^1 q(1-q)dq = 1/6$$

and

$$\int_0^1 \big(\mathbb{E}[(\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q)(\mathbb{1}\{Y \leqslant \varphi(Z,q')\} - q')|W]\big)^2 d(q,q')$$
$$= 2\int_0^1 (\min(q,q') - qq')^2 d(q,q') = (3\sqrt{5})^{-2},$$

see also the proof of Lemma A.3. The required rate imposed in (2.10) on m_n is milder than the rate requirement $m_n = o(n^{1/3})$ imposed by Breunig [2015] in case of nonparametric instrumental mean regression. This is due to the fact that in the latter case we do not have a lower bound for the sieve standard deviation in general, while in case of quantile regression the sieve standard deviation is $\sqrt{m_n}$ within a positive constant. This can be

exploited to weaken rate restrictions on m_n. Further, note that restriction (2.11) implies $k_n = o(\sqrt{m_n})$ (by using that $l_n \leqslant k_n$). This requirement essentially determines the degree of overidentification required for inference.
Note that the rate restriction $\tau_{k_n}\omega_n = o\big(m_n^{-(1+\epsilon)/\kappa}\big)$ imposed in condition (2.11) implies that the dimension parameter m_n dominates the effect of estimation of the structural function. Consequently, the asymptotic behavior of our test statistic is not affect by the estimation of φ, regardless of the underlying degree of ill-posedness. Note that the rate restriction in condition (2.11) can be ensured by choosing k_n relative to the decay of the sieve measure of local ill-posedness, which is described in more detail in the Example 2.3 below.
In the following, we want to illustrate that condition (2.11) is satisfied under common smoothness restrictions on φ and mapping requirements of the Fréchet derivative T_q.

Remark 2.2. *Consider the Hilbert space case $\Phi = L^2_Z$ and let $\{e_j\}_{j\geqslant 1}$ be an orthonormal basis in L^2_Z. In this case, $\Pi_{k_n}\phi = \sum_{j=1}^{k_n} \mathbb{E}[\phi(Z)e_j(Z)]e_j$. Let us assume the following two conditions.*

(i) Sieve Approximation Error: $\|\Pi_{k_n}\phi - \phi\|_Z = O(k_n^{-\alpha/d_z})$ for all $\phi \in \mathcal{B}$.

(ii) Link condition: $\int_0^1 \|T_q(\Pi_{k_n}\phi - \phi)\|_W^2 dq \leqslant \sum_{j\geqslant 1} \upsilon_j \, \mathbb{E}[(\Pi_{k_n}\phi - \phi)(Z)e_j(Z)]^2$ for all $\phi \in \mathcal{B}$ and some positive nonincreasing sequence $(\upsilon_j)_{j\geqslant 1}$.

If the p.d.f. p_Z of $Z \in [0,1]^{d_z}$ is bounded then it is well known that the sieve approximation error condition holds for splines, wavelets, and Fourier series bases. Due to Assumption 4 (v) the link condition is always satisfied with $\upsilon_j = 1$ for all $j \geqslant 1$. The link condition implies an upper bound for the sieve measure of ill-posedness; that is, $\tau_{k_n} \leqslant C\upsilon_{k_n}$ for some constant $C > 0$ and all $n \geqslant 1$ (cf. Lemma B.2 of Chen and Pouzo [2012]). Consequently, the first part of condition (2.11) simplifies to

$$\max\left(l_n, n\, l_n^{-2\beta/d_w}, n\upsilon_{k_n} k_n^{-2\alpha/d_z}\right) = o(\sqrt{m_n})$$

if $\{\mathcal{T}\phi : \phi \in \mathcal{B}_{k_n}\}$ belongs to a Hölder space with Hölder parameter β. In addition, in the setting of Example 2.2, the second part of condition (2.11) simplifies to

$$m_n^{1+\epsilon} \max\left(n^{-1}l_n,\, l_n^{-2\beta/d_w}, k_n^{-2\alpha/d_z}\right) = o(1)$$

for some $\epsilon > 0$. □

In the following example, we illustrate different mapping properties of the operator T_q which are usually studied in the literature.

Example 2.3. *Consider the Hilbert space setting of Remark 2.2 with conditions (i) and (ii). In addition assume that the reverse link condition $\int_0^1 \|T_q\phi\|_W^2 dq \geqslant c\sum_{j\geqslant 1} \upsilon_j \, \mathbb{E}[\phi(Z)e_j(Z)]^2$ for $\phi \in \mathcal{B}$ and some constant $c > 0$ is satisfied. In the setting of Example 2.1, we have $\int_0^1 s_{qj}^2 dq > \upsilon_j$ for all $j \geqslant 1$ implying that T_q is nonsingular for almost all $0 < q < 1$ (since any countable union of null sets is null). For simplicity, let Z and W be scalars. Further, let $\max\left(n^{-1}l_n, l_n^{-2\beta}\right) \sim n^{-1}k_n$ and $k_n \sim n^{\chi}$ for some constant $\chi > 0$ which is specified in the following two cases.*

(i) Mildly ill-posed case: If $\upsilon_{k_n} \sim k_n^{-2\zeta}$ for some $\zeta \geqslant 0$ then in order for (2.11) to hold we require $m_n \sim n^{\iota}$ with $0 < \iota < 1/3$ and

$$(1 - \iota/2)/(2\alpha + 2\zeta) < \chi < \iota/2.$$

Further, $\int_0^1 \|\Pi_{k_n}\varphi_q - \varphi_q\|_Z^2 dq + \tau_{k_n}\omega_n = O(k_n^{-2\alpha} + k_n^{2\zeta+1}n^{-1})$ *which is* $o(m_n^{-2/\kappa})$ *if* $\iota/(\alpha\kappa) < \chi < (1-2\iota/\kappa)/(2\zeta+1)$*. Thus, condition* (2.11) *is satisfied if*

$$\max\Big((1-\iota/2)/(2\alpha+2\zeta), \iota/(\alpha\kappa)\Big) < \chi < \min\Big(\iota/2, (1-2\iota/\kappa)/(2\zeta+1)\Big).$$

(ii) Severely ill-posed case: If $\upsilon_{k_n} \sim \exp\big(-k_n^{2\zeta}\big)$ *for some* $\zeta > 0$ *then* $\int_0^1 \|\Pi_{k_n}\varphi_q - \varphi_q\|_Z^2 dq + \tau_{k_n}\omega_n = O(k_n^{-2\alpha} + \exp(k_n^{2\zeta})k_n n^{-1})$*. Thereby, condition* (2.11) *is satisfied if, for example,* $m_n = o\big((\log n)^{\alpha\kappa/\zeta}\big)$ *and* $k_n \sim (\log n)^{1/\zeta}$*.*

In both situations we conclude that the dimension parameter m_n *is required to be larger than the dimension* k_n *of the sieve space for* n *sufficiently large. Roughly speaking we require more moment restrictions implied by the instrument than parameters we want to estimate. This corresponds to the test of overidentification in the parametric framework.* □

In contrast to a test integrated over all quantiles, one might be interested to check model (1.1) for one specific quantile. In this case, we consider the test statistic

$$S_n(q) = \Big(\sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q) f_{\underline{m_n}}(W_i)\Big)^t (\mathbf{W}_{m_n}^t \mathbf{W}_{m_n})^- \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q) f_{\underline{m_n}}(W_i) \tag{2.12}$$

If $S_n(q)$ becomes too large then we reject the null hypothesis H_0. The derivation of the asymptotic behavior of $S_n(q)$ is similar as in Theorem 2.2. Indeed, only the Lebesgue measure over $(0,1)$ has to be replaced by the Dirac measure which has its mass at the quantile of interest.

Corollary 2.3. *Let Assumptions 1 and 4 be satisfied. For a fixed quantile* $q \in (0,1)$*, let Assumptions 2, 3, and conditions* (2.10) *and* (2.11) *hold. If there exists a function* $\varphi_q \in \mathcal{B}$ *with* $\mathcal{T}\varphi_q = q$ *then*

$$(2m_n)^{-1/2}\Big(\frac{1}{q(1-q)} S_n(q) - m_n\Big) \xrightarrow{d} \mathcal{N}(0,1).$$

In addition, one might be interested in certain regions of quantile functions. Let μ denote any measure on $(0,1)$. Again, the next result is a direct implication of Theorem 2.2 and hence we omit its proof.

Corollary 2.4. *Let Assumptions 1 and 4 be satisfied. For all* q *in the support of* μ*, let Assumptions 2, 3, and conditions* (2.10) *and* (2.11) *hold. If there exists a function* $\varphi \in \mathcal{B}$ *with* $\int |\mathcal{T}\varphi_q - q| d\mu(q) = 0$ *then*

$$\Big(2m_n \int_0^1 (\min(q,q') - qq')^2 d\mu(q) d\mu(q')\Big)^{-1/2} \Big(\int_0^1 S_n(q) d\mu(q) - m_n \int_0^1 q(1-q) d\mu(q)\Big) \xrightarrow{d} \mathcal{N}(0,1).$$

As mentioned in the introduction, our test is a joint test of instrument validity and monotonicity of φ in its second entry. The following remark illustrates how the test statistic $S_n(q)$ integrated over a subset of $(0,1)$ can be useful to detect which kind of deviation exists.

Remark 2.3 (Detecting the Kind of Deviation)**.** *Suppose that the structural function is strictly monotonically increasing in its second entry for values $q \in (0, q')$ given some $q' \in (0,1)$ (can be checked using Corollary 2.4). Further, let $q \mapsto \varphi(\cdot, q)$ be either nonincreasing or decreasing on (q', q''). This can be assured by letting q'' close to q' and assuming that φ does not oscillate for $q \geqslant q'$. If W is a valid instrument, employing model equation (1.2) and $V \sim \mathcal{U}(0,1)$ yields*

$$\begin{aligned}\mathbb{P}(Y \leqslant \varphi(Z,q)|W) &= \mathbb{P}(\varphi(Z,V) \leqslant \varphi(Z,q)|W) \\ &\leqslant \mathbb{P}(V \leqslant q|W) \\ &= q\end{aligned}$$

for all $q \leqslant q''$ and q'' sufficiently close to q'. The last inequality holds regardless whether the function $q \mapsto \varphi(\cdot, q)$ is strictly monotone or not. Consequently, if $\inf_{w \in \mathcal{W}} \mathbb{P}(Y \leqslant \varphi(Z,q)|W = w) > q$ for some $q \in (q', q'')$ we may conclude that W is not a valid instrument. The analysis of a one sided test based on this inequality is beyond the scope of this paper. On the other hand, we can check the kind of deviation by using the estimator $\inf_{w \in \mathcal{W}} f_{\underline{m_n}}(w)^t \big[n^{-1} \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q) f_{\underline{m_n}}(W_i)\big]$. Further, confidence statements can be achieved by using resampling methods. □

Remark 2.4 (Implementation of the Test Statistic)**.** *This remark provides some details on the implementation of our test statistic. First, replace the $(0,1)$–integral by a sum of uniformly distributed random variables. Second, for any integer $m_n \leqslant n^{1/2}$ estimate the structural effect φ_q given in (2.3) for each grid point q, each paramter k_n with $k_n^2 \leqslant m_n$ and $l_n = 2k_n$. Third, compute the standardized test statistic S_n such that it is maximized w.r.t. m_n and minimized w.r.t. k_n. That is, we choose k_n to provide a good model fit and m_n to increase the power of the test. The choice of the dimension parameters capture essential rate requirements imposed to achieve asymptotic normality and is also motivated by simulation results. This leads to a so-called minimum-maximum principle, see also Subsection 4.1 for more details.* □

2.4. Consistency against a fixed alternative

Let us first establish consistency when H_0 does not hold, that is, there exists no function φ belonging to $\mathcal{B}^{(0,1)}$ which solves $\mathcal{T}\varphi_q = q$ for all $0 < q < 1$. The following proposition shows that our test has the ability to reject a false null hypothesis with probability 1 as the sample size grows to infinity. In the following analysis of the asymptotic power of our testing procedure we let $\varphi_q = \operatorname{argmin}_{\phi \in \mathcal{B}} \|\mathcal{T}\phi - q\|_W$. So if H_0 is false then $\int_0^1 \|\mathcal{T}\varphi_q - q\|_W^2 dq > 0$ since p_W is uniformly bounded from below.

Proposition 2.5. *Assume that H_0 does not hold. Let Assumptions 1–4 be satisfied. Consider a sequence $(\gamma_n)_{n \geqslant 1}$ satisfying $\gamma_n = o(n/\sqrt{m_n})$. If conditions (2.10) and (2.11) hold we have*

$$\mathbb{P}\Big(3\sqrt{5/m_n}\big(S_n - m_n/6\big) > \gamma_n\Big) = 1 + o(1).$$

2.5. Limiting behavior under local alternatives

In the following, we study the power of the test, that is, the probability to reject a false hypothesis against a sequence of linear local alternatives that tends to zero as the sample

size tends to infinity. We proceed similarly as Ait-Sahalia et al. [2001] (Section 3.3). More precisely, let $(\varphi_{qn})_{n\geqslant 1}$ be a sequence of (nonstochastic) functions satisfying $n\int_0^1 \|\mathcal{T}\varphi_{qn} - \mathcal{T}\varphi_q\|_W^2 dq = o(\sqrt{m_n})$ where $\varphi_q = \operatorname{argmin}_{\phi\in\mathcal{B}} \|\mathcal{T}\phi - q\|_W$. Then we consider alternative models defined by φ_{qn} with

$$\int_0^1 \big\|\mathcal{T}\varphi_{qn} - q - \delta_n\xi_q\big\|_W^2 dq = o(\delta_n^2) \quad \text{where} \quad \delta_n^2 = \sqrt{m_n}/(3\sqrt{5}\,n). \tag{2.13}$$

Here, $\xi_q \in L_W^2$ is a function satisfying $\int_0^1 \|\xi_q\|_W^2 dq > 0$. The next result establishes asymptotic normality for the standardized test statistic S_n.

Proposition 2.6. *Let Assumptions 1–4 be satisfied. Assume that $(\varphi_{qn})_{n\geqslant 1}$ satisfies (2.13) and $n\int_0^1 \|\mathcal{T}\varphi_{qn} - \mathcal{T}\varphi_q\|_W^2 dq = o(\sqrt{m_n})$. If conditions (2.10) and (2.11) hold we have*

$$3\sqrt{5/m_n}\big(S_n - m_n/6\big) \xrightarrow{d} \mathcal{N}\Big(\sum_{j=1}^{\infty}\int_0^1 \mathbb{E}[\xi_q(W)f_j(W)]^2 dq, 1\Big).$$

From Proposition 2.6 we see that our test can detect local linear alternatives at the rate δ_n. If $\{f_j\}_{j\geqslant 1}$ forms an orthonormal basis in L_W^2 then δ_n coincides with $m_n^{1/4}n^{-1/2}$ within a constant. Hence, our test has the same power against local linear alternatives as the test of Hong and White [1995] who consider parametric specification testing.

2.6. Inference based on bootstrap

Nonparametric tests that rely on the asymptotic normal approximation may perform poorly in finite samples. An alternative approach is to use bootstrap approximation. It is known that bootstrap based procedures could approximate finite sample distributions more accurately. In the following, we propose a bootstrap version of our test statistic S_n.
The bootstrap procedure is based on a sequence of independent and identically distributed random variables ε_i, $1 \leqslant i \leqslant n$, drawn independently of the original data (Y_i, X_i, W_i), $1 \leqslant i \leqslant n$. Following Chen and Pouzo [2015] we then consider the bootstrap residual function

$$\varepsilon_i\big(\mathbb{1}\{Y_i \leqslant \varphi_q(Z_i)\} - q\big).$$

Let $\widehat{\varphi}_{qn}^*$ be the bootstrap version of the sieve least squares estimator (2.3), which is computed in the same way but where only $\big(\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q\big)$ is replaced by $\varepsilon_i\big(\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q\big)$. The bootstrap version S_n^* of our test statistic S_n given in (2.4) builds on $\widehat{\varphi}_{qn}^*$. More precisely, S_n^* is computed as the test statistic S_n but where only $\big(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q\big)$ is replaced by $\varepsilon_i\big(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}^*(Z_i)\} - q\big)$.

Assumption 5. *Let $(\varepsilon_i)_{i\geqslant 1}$ be an independent and identically distributed sequence of random variables drawn independently of (Y, Z, W) such that $\mathbb{E}[\varepsilon] = 1$, $\mathbb{V}ar(\varepsilon) =: \sigma_\varepsilon^2 \in (0, \infty)$ and $\mathbb{E}[|\varepsilon - 1|^4] < \infty$*

Assumption 5 corresponds to Assumption Boot.1 of Chen and Pouzo [2015]. We slightly strengthen their assumption by imposing a fourth moment restriction, which we require to derive asymptotic validity of the bootstrap procedure. Due to the bootstrap innovations

ε_i the constants in the sieve mean and sieve standard deviation change. For the bootstrap test S_n^* we obtain the sieve mean constant

$$\int_0^1 \mathbb{E}[\varepsilon^2(\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q)^2|W]dq = (\sigma_\varepsilon^2 + 1)/6$$

and the sieve standard deviation constant

$$\Big(\int_0^1 \big(\mathbb{E}[\varepsilon^2(\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q)(\mathbb{1}\{Y \leqslant \varphi(Z,q')\} - q')|W]\big)^2 d(q,q')\Big)^{1/2} = (\sigma_\varepsilon^2 + 1)/(3\sqrt{5}).$$

Chen and Pouzo [2015] show that the bootstrap version of the sieve estimator $\widehat{\varphi}_{qn}^*$ converges at the same rate as $\widehat{\varphi}_{qn}$. Thus, following line by line the proof of Theorem 2.2 and using the imposed restrictions on the weights ε_i we obtain the following result.

Corollary 2.7. *Let the assumptions of Theorem 2.7 be satisfied. Under Assumption 5 and null hypothesis H_0 we have*

$$3\sqrt{5/(m_n(\sigma_\varepsilon^2 + 1))}\big(S_n^* - m_n(\sigma_\varepsilon^2 + 1)/6\big) \xrightarrow{d} \mathcal{N}(0,1).$$

It should be emphasized the asymptotic validity of the bootstrap procedure is, in particular, due to the rate condition (2.11), which ensures that the asymptotic distribution of S_n^* is not affected by the estimation of the structural function. The next result establishes consitency of the bootstrap test against fixed alternatives.

Corollary 2.8. *Assume that H_0 does not hold and that the assumptions of Proposition 2.5 are satisfied. Under Assumption 5 we have*

$$\mathbb{P}\Big(3\sqrt{5/(m_n(\sigma_\varepsilon^2 + 1))}\big(S_n^* - m_n(\sigma_\varepsilon^2 + 1)/6\big) > \gamma_n\Big) = 1 + o(1).$$

3. Extensions

As we see in this section, our testing procedure can potentially be applied to a much wider range of situations. We now discuss corollaries that generalize the previous results in different ways. For the following analysis we focus on a fixed quantile $q \in (0,1)$.

3.1. Testing exogeneity

Falsely assuming exogeneity of the regressors leads to inconsistent estimators while on the other hand treating exogenous regressors as if they were endogenous can lower rate of convergence dramatically. In this subsection, we develop a nonparametric test of exogeneity that is robust against possible nonseparability of unobservables. The test statistic is similar to the statistic $S_n(q)$ given in (2.12) but where $\widehat{\varphi}_{qn}$ is replaced by an estimator of the conditional quantile function.

In contrast to the previous section, we assume here that there exists a unique function φ_q satisfying $Y = \varphi_q(Z) + U_q$ with $\mathbb{P}(U_q \leqslant 0|W) = q$ and for some $q \in (0,1)$. The relation between Z and W is thus restricted through this maintained hypothesis. Under the maintained hypothesis, we propose a test whether the vector of regressors Z is exogenous at a quantile $q \in (0,1)$, that is,

$$H_0^e : \mathbb{P}(U_q \leqslant 0|Z) = q.$$

In the following, we denote the conditional quantile function by φ_q^e which satisfies $\mathbb{P}(Y \leqslant \varphi_q^e(Z)|Z) = q$. The null hypothesis H_0^e is satisfied if and only if the structural function φ_q coincides with the conditional quantile function φ_q^e. Further, under nonsingularity of the operator $\mathcal{T}$, hypothesis H_0^e is equivalent to

$$\mathcal{T}\varphi_q^e = q. \tag{3.1}$$

Our test of exogeneity, which we propose below, is based on this equation or equivalently on $\mathbb{P}(Y \leqslant \varphi_q^e(Z)|W) = q$. More precisely, to test exogeneity we replace in the statistic $S_n(q)$ given in (2.12) the estimator of φ_q by an estimator of φ_q^e.
In the following, $\widehat{\varphi}_{qn}^e$ denotes an estimator for the conditional quantile function φ_q^e. For instance, an estimator of φ_q^e is given by

$$\widehat{\varphi}_{qn}^e = \operatorname*{argmin}_{\phi \in \mathcal{B}_{k_n}} \sum_{i=1}^n \varrho_q\big(Y_i - \phi(Z_i)\big) \tag{3.2}$$

where $\varrho_q(u) = |u| - (2q-1)u$ is the check function and here, $\mathcal{B}_{k_n} = \big\{\phi \in \mathcal{B} : \phi(\cdot) = \sum_{j=1}^{k_n} \beta_j e_j(\cdot)\big\}$. For B-spline basis functions and an additional penalty this estimator was proposed by Koenker et al. [1994]. In the following, let p_Z and $p_{Z|W}$ denote the marginal density of Z and the conditional density of Z given W, respectively.

Assumption 6. *(i) There exists a function $\varphi_q \in \mathcal{B}$ such that $\mathcal{T}\varphi_q = q$. (ii) $p_{Y|Z,W}(\cdot, Z, W)$ is continuously differentiable, $|\partial p_{Y|Z,W}(\cdot, Z, W)/\partial y| \leqslant C$ and $p_{Z|W}(\cdot, W) \leqslant C p_Z(\cdot)$ for some constant $C > 0$. (iii) There exists a sequence $(R_n^e)_{n\geqslant 1}$ with $R_n^e = o(1)$ such that $\|\widehat{\varphi}_{qn}^e - \varphi_q^e\|_Z^2 = O_p(R_n^e)$.*

Assumption 6 (i) formalizes the maintained hypothesis of a correctly specified nonparametric instrumental quantile moment equation. Section 2 provides a test for it. Due to Assumption 6 (ii) we do not require Assumption 2 (ii) but can rather rely on an upper bound of the Taylor reminder of $\mathcal{T}$ obtained by Chen et al. [2014]. In this sense, the test of exogeneity presented below requires weaker restrictions on the local curvature of $\mathcal{T}$ than in case of specification testing. Assumption 6 specifies a rate requirement for the L_Z^2 distance of the estimator $\widehat{\varphi}_{qn}^e$. For instance, under H_0^e, Assumption 6 (iii) is satisfied with $R_n^e = k_n/n + k_n^{-2r}$ when $\widehat{\varphi}_{qn}^e$ is given by the estimator (3.2) with the B-splines basis functions $\{e_j\}_{j\geqslant 1}$ and Z is scalar, see He and Shi [1994]. The same rate is obtained by Horowitz and Lee [2005] in the case of multivariate Z in an additive quantile regression model.
For a test of the null hypothesis H_0^e we replace in the definition of $S_n(q)$ given in (2.12) the estimator $\widehat{\varphi}_{qn}$ by $\widehat{\varphi}_{qn}^e$. That is,

$$S_n^e(q) = \Big(\sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}^e(Z_i)\} - q) f_{\underline{m_n}}(W_i)\Big)^t (\mathbf{W}_{m_n}^t \mathbf{W}_{m_n})^- \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}^e(Z_i)\} - q) f_{\underline{m_n}}(W_i)$$

We reject the hypothesis H_0^e if $S_n^e(q)$ becomes too large. The next result establishes asymptotic normality of our test statistic $S_n^e(q)$ under the null hypothesis.

Corollary 3.1. *Let Assumptions 1, 2 (i), 3, 4 and 6 hold. Let m_n satisfy condition (2.10). Consider the estimator $\widehat{\varphi}_{qn}^e$ given in (3.2) where k_n satisfies*

$$nR_n^e = o(\sqrt{m_n}) \quad \text{and} \quad R_n^e = o\big(m_n^{-(1+\epsilon)/\kappa}\big) \tag{3.3}$$

for some $\epsilon > 0$. Then we have under H_0^e

$$(2m_n)^{-1/2}\Big(\frac{1}{q(1-q)}\,S_n^e(q) - m_n\Big) \xrightarrow{d} \mathcal{N}(0,1).$$

Example 3.1. *Let us illustrate when condition (3.3) holds true. Let $m_n \sim n^\iota$ with $0 < \iota < 1/3$. Then for (3.3) to hold let $k_n \sim n^\chi$ where $\chi > 0$ satisfies*

$$\max\Big(\frac{1-\iota/2}{2r}, \frac{\iota}{r\kappa}\Big) < \chi < \min\Big(\frac{\iota}{2}, 1 - \frac{2\iota}{\kappa}\Big).$$

Hence, we require $r > 2/\kappa$ which is a slightly stronger restriction than Assumption 4 (i). □

In the following, we study the power of the test, that is, the probability to reject a false hypothesis against a sequence of linear local alternatives that tends to zero as the sample size tends to infinity. More precisely, let $(\varphi_{qn}^e)_{n\geqslant 1}$ be a sequence of (nonstochastic) functions satisfying

$$\big\|\mathcal{T}\varphi_{qn}^e - q - \delta_n \xi_q^e\big\|_W^2 = o(\delta_n^2) \quad \text{where} \quad \delta_n^2 = \sqrt{2m_n}. \tag{3.4}$$

Here, $\xi_q^e \in L_W^2$ is a function satisfying $\|\xi_q^e\|_W^2 > 0$. The next result establishes asymptotic normality for the standardized test statistic $S_n^e(q)$.

Corollary 3.2. *Let Assumptions 1, 2 (i), 3, 4, and 6 be satisfied. Assume that $(\varphi_{qn}^e)_{n\geqslant 1}$ satisfies (3.4). If condition (3.3) holds true we have*

$$(2m_n)^{-1/2}\Big(\frac{1}{q(1-q)}\,S_n^e(q) - m_n\Big) \xrightarrow{d} \mathcal{N}\Big(\sum_{j=1}^{\infty}\int_0^1 \mathbb{E}[\xi_q^e(W) f_j(W)]^2 dq, 1\Big).$$

3.2. Testing additivity

The test statistic given in (2.4) is also convenient to check additional restrictions on the structural effect φ_q for $0 < q < 1$. These additional restrictions can be easily imposed by constraints on the functions of the sieve space $\mathcal{B}_{k_n}$. For instance, one may impose an additive structure of the quantile structural effects.

By assuming an additive structure of φ_q one might reduce the effect of dimensionality of the regressors on the convergence rate of an estimator (cf. Chen and Pouzo [2012] in case of instrumental quantile regression). Applying this structure leads, however, to inconsistent estimators in general if the function φ_q does not obey an additive form. Our aim in the following is to test whether

$$H_0^{add}: \text{there exist functions } \varphi_q^1, \varphi_q^2 \in \mathcal{B} \text{ such that } \mathbb{P}(Y \leqslant \varphi_q^1(Z') + \varphi_q^2(Z'')|W) = q.$$

Similarly as above we obtain the test statistic

$$S_n^{add}(q) = \Big(\sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}^{add}(Z_i)\} - q) f_{\underline{m_n}}(W_i)\Big)^t (\mathbf{W}_{m_n}^t \mathbf{W}_{m_n})^- \sum_{i=1}^n (\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}^{add}(Z_i)\} - q) f_{\underline{m_n}}(W_i)$$

Here the estimator $\widehat{\varphi}_{qn}^{add} = (\widehat{\varphi}_{qn}^1, \widehat{\varphi}_{qn}^2)$ of $\varphi_q = (\varphi_q^1, \varphi_q^2)$ is given by (2.3) where the sieve basis is a tensor product of basis functions that depend either on Z' or Z''. For a more detailed discussion we refer to Section 6 of Chen and Pouzo [2012]. The next asymptotic normality result is a direct consequence of Corollary 2.3 and hence its proof is omitted.

Corollary 3.3. *Given the conditions of Corollary 2.3 we have under H_0^{add}*

$$(2m_n)^{-1/2}\Big(\frac{1}{q(1-q)}\,S_n^{add}(q) - m_n\Big) \xrightarrow{d} \mathcal{N}(0,1).$$

4. Monte Carlo simulation

In this section, we study the finite-sample performance of our test by presenting the results of a Monte Carlo investigation. There are 1000 Monte Carlo replications in each experiment. Results are presented for the nominal levels 0.05. Let Φ denote the cumulative standard normal distribution function. Throughout this simulation study, realizations (Z, W) were generated by $Z = \Phi\big(\zeta\omega + \sqrt{1-\zeta^2}\,\varepsilon\big)$ and $W = \Phi(\omega)$ where ω is independent of ε and $\omega, \varepsilon \sim \mathcal{N}(0,1)$. Here, the constant $\zeta > 0$ determines the degree of correlation between Z and W and is varied in the experiments.

4.1. Testing a Nonparametric Specification

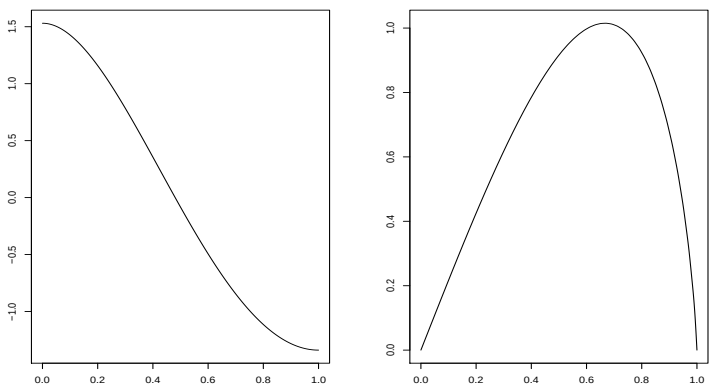
We begin with the finite-sample analysis of our test statistics in case of nonparametric specification testing. To analyze the finite sample power we distinguish in the following between a failure of the null hypothesis caused either by a lack of instrument validity or by non-monotonicity of the structural function in unobservables.

Failure of instrument validity. We first generate realizations of Y under the null hypothesis H_0. Recall that under H_0 there exists a function $\varphi \in \mathcal{B}^{(0,1)}$ such that $\mathbb{P}(Y \leqslant \varphi(Z,q)|W) = q$ for all $q \in (0,1)$. In the following finite sample analysis, we restrict $\mathcal{B}^{(0,1)}$ to contain continuously differentiable functions only. Under H_0 we generate realizations of Y from the nonseparable model

$$Y = \phi(Z)(1 + V/6) + V/2 \tag{4.1}$$

where $V = \vartheta\,\varepsilon + \sqrt{1-\vartheta^2}\,\epsilon$ with $\epsilon \sim \mathcal{N}(0,1)$ independent of (ω, ε) and $\vartheta = 0.7$. We consider the function $\phi(z) = \sum_{j=1}^{\infty} j^{-4}\cos(j\pi z)$. For computational reasons we truncate the infinite sum at 100. The resulting function is displayed in Figure 1. Since ϕ is continuously differentiable the null hypothesis H_0 is satisfied with $\varphi(z,q) = \phi(z)\big(1 + F_V^{-1}(q)/6\big) + F_V^{-1}(q)/2$, where F_V^{-1} denotes the quantile function of V.

Figure 1: Graphs of ϕ and φ^e



When H_0 is false we generate realizations of Y from

$$Y = (\phi(Z) + \rho_j(Z))(1 + V/6) + V/2 \tag{4.2}$$

Table 1: Empirical rejection probabilities for the standardized test statistic $3\sqrt{5/m_n}\big(S_n - m_n/6\big)$ and its bootstrap version $3\sqrt{5/(m_n(\sigma_\varepsilon^2+1))}\big(S_n^* - m_n(\sigma_\varepsilon^2+1)/6\big)$ with varying dimension parameters k_n and m_n with $l_n = 2k_n$.

Sample Size	*Model*		*Emp. Rejection prob. using S_n*			*Emp. Rejection prob. using S_n^**		
		k_n \ m_n	20	25	30	20	25	30
500	H_0 true	4	**0.085**	0.083	0.082	**0.064**	0.052	0.050
	ρ_1		**0.317**	0.289	0.259	**0.252**	0.224	0.196
	ρ_2		**0.337**	0.302	0.289	**0.298**	0.248	0.215
	ρ_3		**0.393**	0.354	0.341	**0.356**	0.308	0.301
	ρ_3		**0.739**	0.701	0.670	**0.748**	0.680	0.658
	H_0 true	5	0.076	0.076	0.080	0.044	0.032	0.048
	ρ_1		0.195	0.179	0.169	0.106	0.106	0.047
	ρ_2		0.200	0.194	0.174	0.130	0.116	0.064
	ρ_3		0.171	0.153	0.152	0.082	0.082	0.064
	ρ_3		0.270	0.257	0.228	0.168	0.140	0.095
1000	H_0 true	4	0.077	**0.082**	0.081	0.060	0.074	**0.076**
	ρ_1		0.630	0.587	0.553	0.576	0.540	0.502
	ρ_2		0.636	0.582	0.549	0.576	0.544	0.492
	ρ_3		0.738	0.697	0.670	0.710	0.662	0.638
	ρ_3		0.905	0.882	0.864	0.938	0.924	0.896
	H_0 true	5	0.203	0.192	0.178	0.098	0.104	0.094
	ρ_1		0.554	**0.518**	0.495	0.420	**0.396**	0.380
	ρ_2		0.629	**0.596**	0.549	0.532	**0.478**	0.460
	ρ_3		0.423	**0.410**	0.385	0.338	**0.314**	0.272
	ρ_3		0.622	**0.593**	0.574	0.576	**0.550**	0.520

where $\rho_j(z) = 10\,j\,(z\,\mathbb{1}\{z \leqslant 0.25\} + (z-1)\,\mathbb{1}\{z > 0.25\})$ for $j = 1,2$ and $\rho_j(z) = (z/2c_j)\,\mathbb{1}\{0.5 - c_j \leqslant z < 0.5 + c_j\}$ for $j = 3,4$, with $c_3 = 0.1$ and $c_4 = 0.05$. Here, the variable V is generated as in (4.1). Under (4.2), the structural function φ satisfying the quantile restriction $\mathbb{P}(Y \leqslant \varphi(Z,q)|W) = q$ is given by $\varphi(z,q) = (\phi(z) + \rho_j(z))(1 + F_V^{-1}(q)/6) + F_V^{-1}(q)/2$. So $\varphi(\cdot,q)$ is not continuously differentiable and thus, H_0 is false. Due to the ill-posed inverse problem estimation of $\varphi(\cdot,q)$ we cannot choose k_n sufficiently large to capture such irregularities which implies finite sample power of our test against those alternatives. This corresponds to the analysis of Horowitz [2011] in the instrumental mean regression case.

For each quantile $0 < q < 1$, we estimate the structural function using the estimator $\widehat{\varphi}_{qn}$ given in (2.3) with B-splines as approximation basis functions. More precisely, for the sieve space $\mathcal{B}_{k_n}$ we use B-splines of order 2 with 1 knot or 2 knots (hence $k_n = 4$ or $k_n = 5$) and for the criterion function we use B-splines of order 2 with 5 knots or 7 knots (hence $l_n = 2k_n$), respectively. We thus follow Chen and Christensen [2015] and choose l_n to be

a constant multiple of k_n. Also for the vector of basis functions $f_{\underline{m_n}}$, used to construct the test statistic, we use B-spline basis of order 2 with knots varying between 17, 22 or 27 (hence $m_n = 20$, $m_n = 25$ or $m_n = 30$).
The empirical rejection probabilities of our standardized test statistic $3\sqrt{5/m_n}\big(S_n - m_n/6\big)$ at nominal level 0.05 are shown in Table 1. We approximate the integral over the quantiles on $(0,1)$ by the mean of a random sample from the uniform $(0,1)$ distribution. As we see from Table 1, our test is less sensitive with respect to the choice of m_n than to the choice of k_n, which is not surprising and well known from nonparametric instrumental variable estimation problems, see also Chen and Pouzo [2015]. Table 1 shows the empirical rejection probabilities for the sample sizes 500 and 1000. We see that as the sample size increases the finite sample rejection probabilities become larger in the alternative models. For $k_n = 4$ we see that the finite sample coverage improves slightly as the sample size increases. This is not the case for $k_n = 5$ which appears to be an inappropriate choice implying a large variance.
In Table 1 we also compare our testing procedure to a bootstrap version of it. We consider the generalized residual bootstrap as proposed in Subsection 2.6. We generate the bootstrap weights by $\varepsilon \sim \mathcal{N}(1, \sigma_\varepsilon^2)$, independently of (Y, X, W), where $\sigma_\varepsilon = 0.5$. We run 200 bootstrap evaluations per Monte Carlo replication. In particular, from Table 1 we see that bootstrap leads to an improvement in the finite sample coverage in the true model. In this sense, the bootstrap test statistic is less sensitive to the choice of k_n under the true model. Similar to Chen and Pouzo [2015] (see p. 1059), we see only a minor improvement of the bootstrap test in the alternative models but we expect that it improves further as the number of bootstrap runs in increased.
As we fix the dimension parameter $l_n = 2k_n$, two dimension parameters remain to be chosen by the econometrician, namely, k_n and m_n. While proposing an adaptive testing procedure is beyond the scope of this paper, we want to provide an heuristic argument for the parameter choice. Intuitively, we want to choose k_n such that we have a good model fit, i.e., a small value of test statistic, and m_n to have good power properties, i.e., a large value of the test statistics. Moreover, the choice should reflect the rate requirement from our theory, that is, $k_n \leqslant l_n = o(m_n^{1/2})$ and $m_n = o(n^{1/2})$. We implement such a heuristic parameter choice criterion via the following minimum-maximum principle. That is, if $\{s(k_n, m_n)\}$ denotes the standardized value of our test S_n with dimension parameters k_n and m_n then we choose these parameters such that

$$\min_{k_n < n^{1/4}} \max_{k_n^2 \leqslant m_n < n^{1/2}} \{s(k_n, m_n)\}.$$

The values of this minimum-maximum principle (over the range $m_n \in \{20, 25, 30\}$ and $k_n \in \{4, 5\}$) are shown in bold in Table 1. Note that the requirement $k_n < n^{1/4}$ implies $k_n \leqslant 4$ when $n = 500$ and $k_n \leqslant 5$ when $n = 1000$. Further, $m_n < n^{1/2}$ implies $m_n \leqslant 22$ for $n = 500$ and $m_n \leqslant 31$ for $n = 1000$. We see that this criterion helps to avoid choosing the dimension parameter k_n too large which would yield inaccurate coverage. Such a rule, however, does not account for ill-posedness of the estimation problem and hence, k_n might still be chosen too large. We thus could calculate the sieve measure of ill-posedness by estimating the first k_n minimal eigenvalues of T_q (see also Chen and Pouzo [2015]).

Failure of monotonicity in unobservables. We study the finite sample power of our test when φ is not strictly monotonic in the structural disturbance V. Realizations of Y were

generated from

$$Y = \Phi(Z + V)V^2 \tag{4.3}$$

where $V = \Phi\big((\vartheta\,\varepsilon + \sqrt{1-\vartheta^2}\,\epsilon)/4\big)$ with $\epsilon \sim \mathcal{N}(0,1)$ and where $\vartheta = 0.8$. When H_0 is false

Table 2: Empirical rejection probabilities for the standardized test statistic $3\sqrt{5/m_n}\big(S_n - m_n/6\big)$ and $3\sqrt{5/(m_n(\sigma_\varepsilon^2+1))}\big(S_n^* - m_n(\sigma_\varepsilon^2+1)/6\big)$ using varying dimension parameters k_n and m_n with $l_n = 2k_n$.

Sample Size	*Model*		*Emp. Rejection prob. using S_n*			*Emp. Rejection prob. using S_n^**		
		k_n \ m_n	20	25	30	20	25	30
500	(4.3)	4	**0.043**	0.066	0.079	**0.022**	0.044	0.058
	(4.4) with j=1		**0.390**	0.433	0.393	**0.338**	0.324	0.298
	(4.4) with j=2		**0.966**	0.967	0.959	**0.984**	0.970	0.964
	(4.5) with j=1		**0.441**	0.492	0.435	**0.376**	0.372	0.342
	(4.5) with j=2		**0.976**	0.979	0.968	**0.994**	0.982	0.978
	(4.3)	5	0.048	0.063	0.083	0.024	0.030	0.036
	(4.4) with j=1		0.183	0.247	0.215	0.132	0.126	0.110
	(4.4) with j=2		0.671	0.710	0.649	0.722	0.662	0.602
	(4.5) with j=1		0.219	0.278	0.259	0.154	0.144	0.112
	(4.5) with j=2		0.721	0.746	0.672	0.766	0.704	0.650
1000	(4.3)	4	0.042	0.080	0.082	0.032	0.037	**0.038**
	(4.4) with j=1		0.717	0.712	0.681	0.696	0.677	0.636
	(4.4) with j=2		1.000	1.000	0.999	1.000	1.000	1.000
	(4.5) with j=1		0.751	0.768	0.737	0.752	0.733	0.694
	(4.5) with j=2		1.000	0.999	0.999	1.000	1.000	1.000
	(4.3)	5	0.044	0.055	**0.057**	0.030	0.030	0.042
	(4.4) with j=1		0.452	**0.435**	0.394	0.414	**0.368**	0.332
	(4.4) with j=2		0.966	**0.953**	0.932	0.982	**0.974**	0.968
	(4.5) with j=1		0.515	**0.490**	0.441	0.490	**0.442**	0.400
	(4.5) with j=2		0.971	**0.961**	0.950	0.984	**0.982**	0.982

we generate

$$Y = \Phi(Z + V)(V - 0.5)^{2j} \tag{4.4}$$

or

$$Y = \Phi(Z + V)\Phi^{-2j}(V) \tag{4.5}$$

for $j = 1, 2$. In the alternative models, the structural disturbance enters the model in a nonmonotonic way. We construct the statistic S_n and its bootstrap counterpart S_n^* as described in the previous paragraph.

Table 2 depicts the empirical rejection probabilities of our test against the alternative models (4.4) and (4.5). Again we observe that our test is not very sensitive to the choice of the dimension parameter m_n. Our test becomes somewhat less powerful for large k_n. But in contrast to the alternatives involving discontinuous functions in the previous paragraph, the choice of k_n is not as sensitive. For each choice of parameter k_n, our test becomes more powerful as the sample size increases from 500 to 1000. For $n = 1000$ we see that the parameter choice $k_n = 5$ leads to a more accurate finite sample coverage. This is captured by the minimum-maximum principle as introduced above. Again, the resulting values of the test statistic using this criterion over the range $m_n \in \{20, 25, 30\}$ and $k_n \in \{4, 5\}$ are shown in bold. Again we observe that the boostrap version of the test statistic behaves similarly as the statistic S_n.

4.2. Testing exogeneity

Realizations Y were generated by

$$Y = \varphi^e(Z) + V/2$$

where V is generated as described in model (4.1), that is, $V = \vartheta\,\varepsilon + \sqrt{1-\vartheta^2}\,\epsilon$ with $\epsilon \sim \mathcal{N}(0,1)$ independent of (ω, ε). The function φ^e is given by $\varphi^e(z) = \sum_{j=1}^{\infty}(-1)^{j+1}\, j^{-2}\sin(j\pi z)$. Again, for computational reasons we truncate the infinite sum at 100. The resulting function is displayed in Figure 1. Note that ϑ determines the degree of endogeneity of Z and is varied among the experiments. The null hypothesis $H_0 : \mathbb{P}(Y \leqslant \varphi^e(Z)|Z) = q$ holds true if $\vartheta = 0$ and is false otherwise. In the following, we perform a test at the median $q = 0.5$. As our test relies on the equation $\mathbb{P}(Y \leqslant \varphi^e(Z)|W) = q$ we expect our test to have more power as the correlation between W and Z increases.

The test statistic is implemented as described in Section 4.2. To estimate the structural effect we make use of the estimator $\widehat{\varphi}^e_{qn}$ of He and Shi [1994] given in (3.2). Here, we use B-splines of order 2 with 1 knot (hence $k_n = 4$) or 2 knots (hence $k_n = 5$). In contrast to the previous section, the choice of the dimension parameter k_n is not affected by the ill-posedness of the underlying inverse problem. As above, the vector of basis functions $f_{\underline{m_n}}$ is also constructed with B-spline basis of order 2 with knots varying between 17, 22 or 27 (hence $m_n = 20$, $m_n = 25$ or $m_n = 30$).

Table 3 depicts the empirical rejection probabilities with varying number of basis functions. As we see from Table 3, our test becomes more powerful for larger ζ; that is, for instruments with a stronger correlation to the covariates Z. From Table 3 we see that the test of exogeneity becomes somewhat less powerful for larger values of m_n. On the other hand, the test seems not to be too sensitive with respect to the choice of the dimension parameters k_n and m_n. We also see from Table 3 that the finite sample coverage and power properties of the test improve as the sample size increases from 500 to 1000.

Similarly as above, a guideline for smoothing parameter choice in practice is given by the following minimum-maximum principle. That is, if $\{s^e_q(k_n, m_n)\}$ denotes the standardized value of our test $S^e_n(q)$ with dimension parameters k_n and m_n then choose these parameters such that

$$\min_{k_n < n^{1/4}} \max_{k_n^2 \leqslant m_n < n^{1/2}} \{s^e_q(k_n, m_n)\}.$$

Again this criterion takes the rate condition for the asymptotic theory into account. In Table 3 the resulting values of the test statistic using this criterion over the range $m_n \in \{20, 25, 30\}$ and $k_n \in \{4, 5\}$ are shown in bold.

Table 3: Empirical rejection probabilities for the standardized test statistic $(2m_n)^{-1/2}\big(4\,S_n^e(0.5) - m_n\big)$ with varying dimension parameters k_n and m_n.

ζ	ϑ		*Emp. Rejection prob. using $S_n^e(0.5)$ with $n = 500$*			*Emp. Rejection prob. using $S_n^e(0.5)$ with $n = 1000$*		
		k_n \ m_n	20	25	30	20	25	30
0.4	0.00	4	**0.064**	0.064	0.064	0.064	0.062	0.056
	0.30		**0.172**	0.161	0.139	0.350	0.290	0.264
	0.35		**0.231**	0.204	0.176	0.497	0.436	0.392
	0.40		**0.319**	0.275	0.256	0.659	0.605	0.546
	0.45		**0.425**	0.389	0.334	0.821	0.775	0.717
0.7	0.00		**0.067**	0.067	0.057	0.054	0.059	0.049
	0.30		**0.273**	0.246	0.219	0.664	0.584	0.542
	0.35		**0.393**	0.363	0.321	0.859	0.800	0.755
	0.40		**0.571**	0.515	0.465	0.970	0.947	0.908
	0.45		**0.746**	0.680	0.619	0.997	0.990	0.982
0.4	0.00	5	0.065	0.067	0.063	0.059	**0.057**	0.055
	0.30		0.170	0.154	0.148	0.335	**0.287**	0.264
	0.35		0.227	0.202	0.179	0.501	**0.428**	0.388
	0.40		0.315	0.278	0.256	0.667	**0.598**	0.553
	0.45		0.429	0.386	0.355	0.824	**0.775**	0.715
0.7	0.00		0.061	0.057	0.055	0.049	0.041	**0.045**
	0.30		0.247	0.221	0.201	0.647	**0.581**	0.525
	0.35		0.393	0.353	0.318	0.858	**0.797**	0.727
	0.40		0.571	0.495	0.438	0.966	**0.940**	0.905
	0.45		0.725	0.658	0.598	0.997	**0.990**	0.983

5. An empirical illustration

To illustrate our testing procedure, we present an empirical application concerning estimation of the effects of class size on students' performance on standardized tests. Angrist and Lavy [1999] studied the effects of class size on test scores of 4th and 5th grade students in Israel. In this empirical illustration, we focus on 4th grade reading comprehension which was also considered by Horowitz [2011].
In this empirical example we study the model

$$Y_{sc} = \varphi(Z_{sc}, V_{sc}) + D_{sc}\,\beta(V_{sc}) \tag{5.1}$$

where Y_{sc} is the average reading comprehension test score of 4th grade students in class c of school s, Z_{sc} is the number of students in class c of school s, D_{sc} is the fraction of disadvantaged students in class c of school s with unknown scalar function β, $V_{sc} = U_s + \varepsilon_{sc}$ where U_s is an unobserved school-specific random effect, and ε_{sc} is an unobserved, independently over classes and schools distributed random variable.

The class size Z_{sc} may be endogenous, for instance, due to the socioeconomic background of the students. To identify the causal effect of class size on scholar achievement Angrist and Lavy [1999] use *Maimonides' rule* as instruments. According to this administrative rule, maximum class size is given by 40 pupils and will be split if the number of enrolled students exceeds this number. More precisely, assuming that cohorts are divided into classes of equal size, Maimonides rule is described by

$$W_{sc} = E_s / \lceil 1 + (E_s - 1)/40 \rceil$$

where E_s denotes enrollment in school s and $\lceil x \rceil$ denotes the largest integer less or equal to x. Note that Horowitz [2011] could show that a linear relation between class size and scholar achievement as used by Angrist and Lavy [1999] is misspecified. To apply our tests, we consider a subsample where only one representative class per school is considered. By doing so, we avoid that rejection of a hypothesis may be caused by within class correlation. Moreover, only schools with at least two classes are considered which leads to a sample size of 707.

In the following, we want to test nonparametrically whether class size is endogenous at the 0.5–quantile. The null hypothesis is that $\mathbb{P}(Y_{sc} \leqslant \varphi(Z_{sc}, q) + D_{sc}\,\beta(q)|Z_{sc}) = q$ where $q = 0.5$. The value of our test statistic $S_n^e(0.5) = (2m_n)^{-1/2}\big(4\,S_n^e(0.5) - m_n\big)$ is given by 1.885. For the choice of smoothing parameters k_n and m_n we applied the minimum-maximum principle as described in Section 4.2. The resulting dimension parameters are $k_n = 4$ and $m_n = 23$.[2] We thus, reject the hypothesis of exogeneity at the 0.05 nominal level. In particular, in model (5.1) under conditions *(a.1)– (a.3)* we conclude that Z_{sc} is not independent of V_{sc}.

We now test whether the model (5.1) with conditions *(a.1)– (a.3)* is correctly specified. We construct our test statistic using B-splines as described in Section 4.1. For the choice of smoothing parameters k_n and m_n we applied the minimum-maximum principle as described in Section 4.2. As in the Monte Carlo section we choose $l_n = 2k_n$. Our test statistic attains the value 1.4152 and thus fails to reject the nonseparable model (5.1) with conditions *(a.1)– (a.3)* at the 0.05 nominal level. This value of the test statistic is obtained when $k_n = 4$ and $m_n = 26$. For the fixed quantile $q = 0.5$, we also performed a test of $\mathbb{P}(Y_{sc} \leqslant \varphi(Z_{sc}, q) + D_{sc}\,\beta(q)|W_{sc}) = q$. In this case, our test statistic attains the value 0.981 and again fails to reject the hypothesis.[3]

For the full sample, Figure 2 depicts estimators of the structural effect φ_q for the quantiles $q \in \{0.25, 0.5, 0.75\}$ where the number of disadvantaged students is smaller than 15% (in this case $n = 688$). The solid black lines are the estimators and the blue lines are the 90% pointwise bootstrap confidence intervals (we account for within school correlation by using schools as the bootstrap sampling units, see also Horowitz [2011]). We can see that the confidence intervals are tight enough to reject the hypothesis that the quantile structural effects are overall upward sloping. In particular, we see that the effect of class size variation on test scores is more severe for lower performing classes.

6. Conclusion

In this paper, we develop a nonparametric specification test for the quantile regression model (1.1). The test statistic is easy to implement and a natural extension of specification testing in a parametric framework. We establish the asymptotic distribution of our test under the null hypothesis. Our test is consistent against a fixed alternative and we study its power

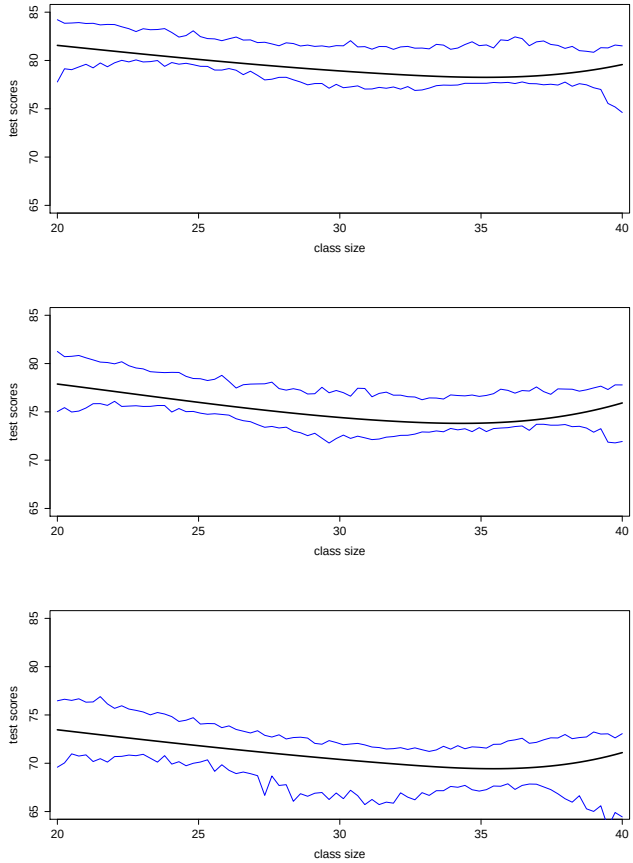


Figure 2: Estimated structural effects (solid black lines) for $q \in \{0.25, 0.5, 0.75\}$ and 90% confidence intervals (blue lines)

properties by considering a sequence of local alternatives. We also provide extensions of our test theory concerning model simplification. We demonstrate via Monte Carlo simulations that our testing procedure performs well in finite samples. The usefulness of our testing procedure is illustrated by an empirical example.

Notes

[1]Since conditional expectations are defined only up to equality almost surely, all (in)equalities with conditional expectations and/or random variables are understood as (in)equalities almost surely.

[2]The value of the test for other choices of k_n is 2.254 for $k_n = 3$ and 2.182 for $k_n = 5$ where $l_n = 2k_n$ and m_n is maximized over the range k_n^2 to 26 (being the largest integer smaller than $\sqrt{707}$).

[3]This is not the case if k_n is chosen too small or too large. For instance if $k_n = 4$ or $k_n = 9$, respectively, then the value of the test statistic is 2.064 or 3.420 (as above maximized of m_n and $l_n = 2k_n$).

References

R. A. Adams and J. J. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier/Academic Press, Amsterdam, 2003.

Y. Ait-Sahalia, P. J. Bickel, and T. M. Stoker. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105(2):363–412, 2001.

J. D. Angrist and V. C. Lavy. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575, 1999.

A. M. Awad. Conditional central limit theorems for martingales and reversed martingales. *The Indian Journal of Statistics, Series A*, 43:10–106, 1981.

R. Blundell and J. Horowitz. A nonparametric test of exogeneity. *Review of Economic Studies*, 74(4):1035–1058, 2007.

C. Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015.

X. Chen. Large sample sieve estimation of semi-nonparametric models. Handbook of Econometrics, pages 5549 – 5632. Elsevier, 2007.

X. Chen and T. M. Christensen. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 188(2):447–465, 2015.

X. Chen and D. Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 2012.

X. Chen and D. Pouzo. Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica*, 83(3):1013–1079, 2015.

X. Chen and M. Reiß. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27(03):497–521, 2011.

X. Chen and A. Santos. Overidentification in regular models. Cowles foundation discussion papers, Cowles Foundation for Research in Economics, Yale University, 2015.

X. Chen, O. Linton, and I. Van Keilegom. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71:1591–1608, 2003.

X. Chen, V. Chernozhukov, S. Lee, and W. K. Newey. Local identification of nonparametric and semiparametric models. *Econometrica*, 82(2):785–809, 2014.

V. Chernozhukov and C. Hansen. An iv model of quantile treatment effects. *Econometrica*, 73:245–261, 2005.

V. Chernozhukov, G. Imbens, and W. K. Newey. Instrumental variable estimation of nonseparable models. *Journal of Econometrics*, 139(1):4–14, 2007.

A. Chesher. Identification in nonseparable models. *Econometrica*, 71(5):1405–1441, 2003.

D. Chetverikov and D. Wilhelm. Nonparametric instrumental variable estimation under monotonicity. *arXiv preprint arXiv:1507.05270*, 2015.

P.-A. Chiappori, I. Komunjer, and D. Kristensen. Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, 188(1):22–39, 2015.

F. Dunker, J.-P. Florens, T. Hohage, J. Johannes, and E. Mammen. Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression. *Journal of Econometrics*, 178:444–455, 2014.

J. C. Escanciano and S.-C. Goh. Specification analysis of linear quantile models. *Journal of Econometrics*, 178:495–507, 2014.

J. C. Escanciano and C. Velasco. Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, 159(1):209–221, 2010.

F. Fève, J.-P. Florens, and I. Van Keilegom. Estimation of conditional ranks and tests of exogeneity in nonparametric nonseparable models. Technical Report just-accepted, 2016.

P. Gagliardini and O. Scaillet. A specification test for nonparametric instrumental variable regression. Swiss Finance Institute Research Paper No. 07-13, 2007.

P. Gagliardini and O. Scaillet. Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica*, 80(4):1533–1562, 2012.

M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72(1):21–37, Nov. 1995.

X. He and P. Shi. Convergence rate of b-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, 3(3-4):299–308, 1994.

S. Hoderlein and E. Mammen. Identification of marginal effects in nonseparable models without monotonicity. *Econometrica*, 75(5):1513–1518, 2007.

S. Hoderlein, L. Su, H. White, and T. T. Yang. Testing for monotonicity in unobservables under unconfoundedness. *Journal of Econometrics*, 193(1):183–202, 2016.

Y. Hong and H. White. Consistent specification testing via nonparametric series regression. *Econometrica*, 63:1133–1159, 1995.

J. L. Horowitz. Specification testing in nonparametric instrumental variables estimation. *Journal of Econometrics*, 167:383–396, 2012.

J. L. Horowitz. Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2):347–394, 2011.

J. L. Horowitz and S. Lee. Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75:1191–1208, 2007.

J. L. Horowitz and S. Lee. Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 152(2):141–152, 2009.

J. L. Horowitz and S. Lee. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100(472):1238–1249, 2005.

R. Koenker, P. Ng, and S. Portnoy. Quantile smoothing splines. *Biometrika*, 81(4):673–680, 1994.

R. Kress. *Linear integral equations*, volume 82 of *Applied Mathematical Sciences.* Springer, New York, NY, 2 edition, 1989.

A. Lewbel, X. Lu, and L. Su. Specification testing for transformation models with an application to generalized accelerated failure-time models. *Journal of Econometrics*, 184 (1):81–96, 2015.

W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168, 1997.

A. Santos. Inference in nonparametric instrumental variables with partial identification. *Econometrica*, 80(1):213–275, 2012.

A. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)*. Springer, 2000.

A. Appendix

A.1. Proofs of Section 3.

In the appendix, $f_{\underline{m_n}}$ denotes an m_n dimensional vector with entries f_j for $1 \leqslant j \leqslant m_n$. Moreover, $\|\cdot\|$ is the usual Euclidean norm. For ease of notation, let $\mathbf{X}_i = (Y_i, Z_i, W_i)$ for $1 \leqslant i \leqslant n$ with realizations $\mathbf{x} = (y, z, w) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{W}$. Let $\mathcal{H}$ be a class of measurable functions with a measurable envelope function H. Then $N(\varepsilon, \mathcal{H}, L_X^2)$ and $N_{[]}(\varepsilon, \mathcal{H}, L_X^2)$, respectively, denote the covering and bracketing numbers for the set $\mathcal{H}$. In addition, let $J_{[]}(1, \mathcal{H}, L_X^2)$ denote a bracketing integral of $\mathcal{H}$, that is,

$$J_{[]}(1, \mathcal{H}, L_X^2) = \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon \|H\|_X, \mathcal{H}, L_X^2)} d\varepsilon.$$

Throughout the proofs, we will use $C > 0$ to denote a generic finite constant that may be different in different uses. Further, for ease of notation we write $\int$ for $\int_0^1$, $\sum_i$ for $\sum_{i=1}^n$, and $\sum_{i'<i}$ for $\sum_{i=1}^n \sum_{i'=1}^{i-1}$. For any $\phi, \psi \in L_W^2$, the inner product in L_W^2 is denoted by $\langle \phi, \psi \rangle_W = \mathbb{E}[\phi(W)\psi(W)]$ and let $F_{m_n}\phi = \sum_{j=1}^{m_n} \langle \phi, f_j \rangle_W f_j$. In the following, we denote $\widehat{Q}_n = n^{-1} \sum_i f_{\underline{m_n}}(W_i) f_{\underline{m_n}}(W_i)^t$. By Assumption 1, the eigenvalues of $\mathbb{E}[f_{\underline{m_n}}(W) f_{\underline{m_n}}(W)^t]$ are bounded away from zero and hence, it may be assumed that $\mathbb{E}[f_{\underline{m_n}}(W) f_{\underline{m_n}}(W)^t] = I_{m_n}$ where I_{m_n} denotes the m_n dimensional identity matrix (cf. Newey [1997], p. 161).
In the following result, we establish continuity of the mapping $q \mapsto \varphi(\cdot, q)$ under the tangential cone condition and a mild assumption on the sieve approximation error for φ_q.

Lemma A.1. *Let Assumption 2 be satisfied. Assume for almost all $q \in (0,1)$ there exists a function φ_q with $\mathcal{T}\varphi_q = q$, let T_q be compact, and $\|\varphi_q - \Pi_k \varphi_q\|_Z = o(1)$ as $k \to \infty$. Then the mapping $q \mapsto \varphi(\cdot, q)$ is continuous.*

Proof. For some $q \in (0,1)$, since the linear operator T_q is compact there exists singular value decomposition of it denoted by $\{s_{qj}, e_j, f_j\}_{j \geqslant 1}$. For any $\varepsilon > 0$ and k sufficiently large, let us define $\delta = (1 - \eta)\, \varepsilon\, s_{qk}/3$. We consider $q' \in (0,1)$ such that $|q - q'| < \delta$. Since q, q' satisfy the quantile restriction we have $\|\mathcal{T}\varphi_q - \mathcal{T}\varphi_{q'}\|_W < \delta$. Let us further denote

$r_k(q) = \|\Pi_k\varphi_q - \varphi_q\|_W$. We have $r_k(q) \leqslant \varepsilon/6$ by assumption for all q. By Assumption 2 (ii) and the triangular inequality it holds

$$\begin{aligned}
\|\mathcal{T}\varphi_q - \mathcal{T}\varphi_{q'}\|_W &\geqslant (1-\eta)\|T_q(\varphi_q - \varphi_{q'})\|_W \\
&= (1-\eta)\|T_q\Pi_k(\varphi_q - \varphi_{q'}) - T_q(\Pi_k\varphi_q - \varphi_q) + T_q(\Pi_k\varphi_{q'} - \varphi_{q'})\|_W \\
&\geqslant (1-\eta)\Big(\|T_q\Pi_k(\varphi_q - \varphi_{q'})\|_W - \|T_q(\Pi_k\varphi_q - \varphi_q)\|_W - \|T_q(\Pi_k\varphi_{q'} - \varphi_{q'})\|_W\Big) \\
&\geqslant (1-\eta)\, s_{qk}\Big(\|\Pi_k(\varphi_q - \varphi_{q'})\|_Z - r_k(q) - r_k(q')\Big) \\
&\geqslant (1-\eta)\, s_{qk}\Big(\|\varphi_q - \varphi_{q'}\|_Z - 2r_k(q) - 2r_k(q')\Big),
\end{aligned}$$

using that $(s_{qj})_{j\geqslant 1}$ is a nonincreasing sequence. This implies

$$\begin{aligned}
\|\varphi_q - \varphi_{q'}\|_Z &\leqslant (1-\eta)^{-1}\, s_{qk}^{-1}\,\delta + 2r_k(q) + 2r_k(q') \\
&\leqslant (1-\eta)^{-1}\, s_{qk}^{-1}\,\delta + 2\,\varepsilon/3 \\
&\leqslant \varepsilon,
\end{aligned}$$

which proves the result. □

Proof of Proposition 2.1. For any $q \in (0,1)$, since the linear operator T_q is compact there exists singular value decomposition of it denoted by $\{s_{qj}, e_j, f_j\}_{j\geqslant 1}$. Let us denote $\langle \phi, e_j\rangle_Z := \mathbb{E}[\phi(Z)e_j(Z)]$. Consider first the polynomial case, i.e., $s_{qj} = j^{-\zeta}$ for some $\zeta > 0$. Consider the set of functions $\Psi = \Big\{\psi \in L^2_W : \langle \psi, f_j\rangle_W \sim j^{-2(\zeta+\delta')-1} \text{ for some } \delta' < \delta\Big\}$. This set is obviously compact with $\Psi \in L^2_W \subset \Big\{\psi \in L^2_W : \sum_{j=1}^\infty j^{2\zeta}\langle \psi, f_j\rangle_W^2 < C\Big\} \subset L^2_W$. For any $\psi \in \Psi$ there exists $\phi \in L^2_Z$ such that $\|T_q\phi - \psi\|_W = 0$ or equivalently

$$\sum_{j=1}^\infty \Big(j^{-\zeta}\langle \phi, e_j\rangle_Z - \langle \psi, f_j\rangle_W\Big)^2 = 0.$$

Thus for each $j \geqslant 1$ we have $\langle \phi, e_j\rangle_Z = j^{\zeta}\langle \psi, f_j\rangle_W$ and thus

$$\langle \phi, e_j\rangle_Z^2 \sim j^{-1-2\delta'}$$

and since $\delta' < \delta$ the function ϕ does not belong to $\mathcal{B}$. Consequently, for each $\psi \in \Psi$ the associated function $\phi \in L^2_Z$ does not belong to $\mathcal{B}$, i.e., Ψ does not belong to the closure of $\mathcal{R}_q(\mathcal{B})$. In the exponential case, i.e., $s_{qj}^2 = \exp(-j^{2\zeta})$, the result follows immediately from the above, which completes the proof. □

Proof of Theorem 2.2. Since we have $\|\widehat{Q}_n - I_{m_n}\|^2 = o_p(m_n^2/n)$ it is sufficient to prove that $3\sqrt{5/m_n}\big(\sum_{j=1}^{m_n}\int |n^{-1/2}\sum_i(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - q)f_j(W_i)|^2 dq - m_n/6\big) \xrightarrow{d} \mathcal{N}(0,1)$. The

proof is based on the decomposition

$$\begin{aligned}
&\sum_{j=1}^{m_n}\int \big|n^{-1}\sum_i(\mathbb{1}\{Y_i\leqslant\widehat{\varphi}_{qn}(Z_i)\}-q)f_j(W_i)\big|^2dq\\
&=\sum_{j=1}^{m_n}\int \big|n^{-1}\sum_i(\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}-q)f_j(W_i)\big|^2dq\\
&-\frac{2}{n^2}\sum_{j=1}^{m_n}\int\Big(\sum_i(\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}-q)f_j(W_i)\Big)\\
&\qquad\qquad\times\Big(\sum_i\big(\mathbb{1}\{Y_i\leqslant\widehat{\varphi}_{qn}(Z_i)\}-\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}\big)f_j(W_i)\Big)dq\\
&+\sum_{j=1}^{m_n}\int\big|n^{-1}\sum_i\big(\mathbb{1}\{Y_i\leqslant\widehat{\varphi}_{qn}(Z_i)\}-\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}\big)f_j(W_i)\big|^2dq=I_n-2II_n+III_n.
\end{aligned}\tag{A.1}$$

Consider I_n. We calculate further

$$\begin{aligned}
m_n^{-1/2}\big(nI_n-m_n/6\big)&=\frac{1}{\sqrt{m_n}n}\sum_i\sum_{j=1}^{m_n}\Big(\int|(\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}-q)f_j(W_i)|^2dq-1/6\Big)\\
&+\frac{1}{\sqrt{m_n}n}\sum_{i\neq i'}\sum_{j=1}^{m_n}\int\big(\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}-q\big)\big(\mathbb{1}\{Y_{i'}\leqslant\varphi(Z_{i'},q)\}-q\big)f_j(W_i)f_j(W_{i'})dq
\end{aligned}$$

where the first summand tends in probability to zero as $n\to\infty$. Indeed,we have

$$\mathbb{E}\int|(\mathbb{1}\{Y\leqslant\varphi(Z,q)\}-q)f_j(W)|^2dq=\mathbb{E}[f_j^2(W)]\int q(1-q)dq=1/6$$

for all $j\geqslant 1$ and hence,

$$\begin{aligned}
&\mathbb{E}\Big|\frac{1}{\sqrt{m_n}n}\sum_i\sum_{j=1}^{m_n}\Big(\int|(\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}-q)f_j(W_i)|^2dq-1/6\Big)\Big|^2\\
&\leqslant\frac{1}{m_nn}\int\mathbb{E}\Big|\sum_{j=1}^{m_n}|(\mathbb{1}\{Y\leqslant\varphi(Z,q)\}-q)f_j(W)|^2-\mathbb{E}\,|(\mathbb{1}\{Y\leqslant\varphi(Z,q)\}-q)f_j(W)|^2\Big|^2dq\\
&\qquad\leqslant\frac{1}{m_nn}\sup_{w\in\mathcal{W}}\|f_{\underline{m_n}}(w)\|^4\int\mathbb{E}\,|\,\mathbb{1}\{Y\leqslant\varphi(Z,q)\}-q|^4dq\\
&\qquad\leqslant O(m_n/n)=o(1)
\end{aligned}$$

by using $\sup_{w\in\mathcal{W}}\|f_{\underline{m_n}}(w)\|^2\leqslant Cm_n$. Therefore, to establish $3\sqrt{5/m_n}(nI_n-m_n/6)\xrightarrow{d}\mathcal{N}(0,1)$ it is sufficient to show

$$\frac{3\sqrt{5}}{\sqrt{m_n}n}\sum_{i\neq i'}\sum_{j=1}^{m_n}\int\big(\mathbb{1}\{Y_i\leqslant\varphi(Z_i,q)\}-q\big)\big(\mathbb{1}\{Y_{i'}\leqslant\varphi(Z_{i'},q)\}-q\big)f_j(W_i)f_j(W_{i'})dq\xrightarrow{d}\mathcal{N}(0,1).$$

This follows from Lemma A.3. Consider III_n. Let us denote $\mathcal{B}_n:=\{\phi\in\mathcal{B}^{(0,1)}:|||\phi-\varphi|||_{Z,p}^2\leqslant m_n^{-(1+c)/\kappa}\}$ for some constant $c>0$ and $\mathcal{B}_{qn}:=\{\phi_q:\phi\in\mathcal{B}_n\}\subset\mathcal{B}$. Further, we

denote for $1 \leqslant j \leqslant m_n$ and $1 \leqslant i \leqslant n$

$$h_{qj}(\mathbf{X}_i, \phi_q) = \big(\mathbb{1}\{Y_i \leqslant \phi(Z_i, q)\} - \mathbb{1}\{Y_i \leqslant \varphi(Z_i, q)\}\big) f_j(W_i)$$

and the classes $\mathcal{H}_{qjn} = \{h_{qj}(\cdot, \phi_q) : \phi_q \in \mathcal{B}_{qn}\}$ and $\mathcal{H}_{qj} = \{h_{qj}(\cdot, \phi_q) : \phi_q \in \mathcal{B}\}$. We observe

$$\begin{aligned} III_n &= \sum_{j=1}^{m_n} \int \big| n^{-1} \sum_i h_{qj}(\mathbf{X}_i, \widehat{\varphi}_{qn}) \big|^2 dq \\ &\leqslant 2\eta_p \,|||\, \mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi \,|||_W^2 + 2 \sum_{j=1}^{m_n} \int \big| n^{-1} \sum_i h_{qj}(\mathbf{X}_i, \widehat{\varphi}_{qn}) - \big\langle \mathcal{T}\widehat{\varphi}_{qn} - \mathcal{T}\varphi_q, f_j \big\rangle_W \big|^2 dq. \end{aligned}$$

From (A.4) in Lemma A.2 together with condition $n\tau_n = o(\sqrt{m_n})$ we deduce $n \,|||\, \mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi|||_W^2 = o_p(\sqrt{m_n})$. Further, we observe for every $\phi_q \in \mathcal{B}_{qn}$ that

$$\big| h_{qj}(\mathbf{X}_i, \phi_q) \big|^2 \leqslant \max_{\phi_q \in \mathcal{B}_{qn}} \big| \big(\mathbb{1}\{Y_i \leqslant \phi(Z_i, q)\} - \mathbb{1}\{Y_i \leqslant \varphi(Z_i, q)\}\big) f_j(W_i) \big|^2 =: H_{qj}^2(\mathbf{X}_i)$$

and hence, H_{qj} is an envelope function of the class $\mathcal{H}_{qjn}$ and due to Assumption 3 we have $\mathbb{E}[\int H_{qj}^2(\mathbf{X}) dq] \leqslant C m_n^{-(1+c)}$. Moreover, (A.5) in Lemma A.2 together with condition (2.11) implies $|||\widehat{\varphi}_{\cdot n} - \varphi|||_{Z,p}^2 = o_p\big(m_n^{-(1+c)/\kappa}\big)$ and thereby

$$\begin{aligned} &\mathbb{P}\Big(\sum_{j=1}^{m_n} \int \big| n^{-1} \sum_i h_{qj}(\mathbf{X}_i, \widehat{\varphi}_{qn}) - \big\langle \mathcal{T}\widehat{\varphi}_{qn} - \mathcal{T}\varphi_q, f_j \big\rangle_W \big|^2 dq > \varepsilon \Big) \\ &\qquad \leqslant \sum_{j=1}^{m_n} \varepsilon^{-1} \,\mathbb{E} \sup_{\phi \in \mathcal{B}_n} \int \Big| n^{-1/2} \sum_i h_{qj}(\mathbf{X}_i, \phi_q) - \mathbb{E}\, h_{qj}(\mathbf{X}, \phi_q) \Big|^2 dq + o(1) \\ &\qquad \leqslant \sum_{j=1}^{m_n} \varepsilon^{-1} \int \mathbb{E} \max_{\phi_q \in \mathcal{B}_{qn}} \Big| n^{-1/2} \sum_i h_{qj}(\mathbf{X}_i, \phi_q) - \mathbb{E}\, h_{qj}(\mathbf{X}, \phi_q) \Big|^2 dq + o(1) \\ &\leqslant \sum_{j=1}^{m_n} \varepsilon^{-1} \int \Big(\mathbb{E} \max_{\phi_q \in \mathcal{B}_{qn}} \Big| n^{-1/2} \sum_i h_{qj}(\mathbf{X}_i, \phi_q) - \mathbb{E}\, h_{qj}(\mathbf{X}, \phi_q) \Big| + \big(\mathbb{E}\, |H_{qj}(\mathbf{X})|^2 \big)^{1/2} \Big)^2 dq + o(1) \end{aligned}$$

where the last inequality is due to Theorem 2.14.5 of van der Vaart and Wellner [2000]. We further conclude by applying the last display of Theorem 2.14.2 of van der Vaart and Wellner [2000]

$$\mathbb{E} \max_{\phi_q \in \mathcal{B}_{qn}} \Big| n^{-1/2} \sum_i h_{qj}(\mathbf{X}_i, \phi_q) - \mathbb{E}\, h_{qj}(\mathbf{X}, \phi_q) \Big| \leqslant C J_{[]}(1, \mathcal{H}_{qjn}, L_{\mathbf{X}}^2) \big(\mathbb{E}\, |H_{qj}(\mathbf{X})|^2 \big)^{1/2}$$

for all $0 < q < 1$. Now since $\max_{1 \leqslant j \leqslant m_n} \mathbb{E} \int |H_{qj}(\mathbf{X})|^2 dq \leqslant C m_n^{-(1+c)}$ for n sufficiently large it is sufficient to show that $\max_{1 \leqslant j \leqslant m_n} J_{[]}(1, \mathcal{H}_{qjn}, L_{\mathbf{X}}^2) < C$ for all $0 < q < 1$. From Lemma 4.2 (i) of Chen [2007] we deduce

$$\begin{aligned} N_{[]}(\varepsilon \big(\mathbb{E}\, |H_{qj}(\mathbf{X})|^2 \big)^{1/2}, \mathcal{H}_{qjn}, L_{\mathbf{X}}^2) &\leqslant N_{[]}\Big(\varepsilon, \big(\mathbb{E}\, |H_{qj}(\mathbf{X})|^2 \big)^{-1/2} \mathcal{H}_{qjn}, L_{\mathbf{X}}^2 \Big) \\ &\leqslant N_{[]}\Big(\varepsilon, \mathcal{H}_{qj}, L_{\mathbf{X}}^2 \Big) \\ &\leqslant N\Big(\Big(\frac{\varepsilon}{2C} \Big)^{2/\kappa}, \mathcal{B}, \|\cdot\|_{Z,p} \Big) \\ &\leqslant N\Big(\Big(\frac{\varepsilon}{2C} \Big)^{2/\kappa}, \mathcal{B}, \|\cdot\|_{\infty} \Big). \end{aligned}$$

Employing condition $\alpha_0 > d_z/p$ and Theorem 6.2 Part II of Adams and Fournier [2003] yields that $W^{\alpha,p}$ is compactly embedded in $W^{\alpha,\infty}$. Thereby, $\mathcal{B} \subset W^{\alpha,p}$ is totally bounded in $W^{\alpha,\infty}$ which implies $\|\phi\|_{\alpha,\infty} \leqslant C$ for all $\phi \in \mathcal{B}$. Let $W_C^{\alpha,\infty} := \{W^{\alpha,\infty} : \|\phi_q\|_{\alpha,\infty} \leqslant C\}$. Now Theorem 2.7.1 of van der Vaart and Wellner [2000] gives

$$\log N\big(\varepsilon^{2/\kappa}, \mathcal{B}, \|\cdot\|_\infty\big) \leqslant \log N\big(\varepsilon^{2/\kappa}, W_C^{\alpha,\infty}, \|\cdot\|_\infty\big) \leqslant C\varepsilon^{-2d_z/(\alpha\kappa)}$$

where C depends on the diameter of $\mathcal{Z}$. Now due to Assumption 4 (i) it is straightforward to see that $\max_{1\leqslant j\leqslant m_n} J_{[]}(1, \mathcal{H}_{qjn}, L^2_{\mathbf{X}}) < C$ and hence, $nIII_n = o_p(\sqrt{m_n})$.
Consider II_n. We observe

$$\begin{aligned}
nII_n &= \sum_{j=1}^{m_n} \int \Big(\sum_i (\mathbb{1}\{Y_i \leqslant \varphi(Z_i,q)\} - q) f_j(W_i)\Big)\Big(n^{-1}\sum_i h_{qj}(\mathbf{X}_i, \widehat{\varphi}_{qn})\Big) dq \\
&= \sum_{j=1}^{m_n} \int \Big(\sum_i (\mathbb{1}\{Y_i \leqslant \varphi(Z_i,q)\} - q) f_j(W_i)\Big)\Big(n^{-1}\sum_i h_{qj}(\mathbf{X}_i, \widehat{\varphi}_{qn}) - \big\langle \mathcal{T}\widehat{\varphi}_{qn} - \mathcal{T}\varphi_q, f_j\big\rangle_W\Big) dq \\
&\qquad + \sum_{j=1}^{m_n} \int \Big(\sum_i (\mathbb{1}\{Y_i \leqslant \varphi(Z_i,q)\} - q) f_j(W_i)\Big)\big\langle \mathcal{T}\widehat{\varphi}_{qn} - \mathcal{T}\varphi_q, f_j\big\rangle_W dq \\
&= C_{n1} + C_{n2}.
\end{aligned}$$

The Cauchy Schwarz inequality implies for all $\varepsilon > 0$

$$\begin{aligned}
\mathbb{P}(|C_{n1}| > \varepsilon\sqrt{m_n}) &\leqslant (\varepsilon\sqrt{m_n})^{-1}\Big(\int q(1-q)dq\Big)^{1/2} \\
&\quad \times \sum_{j=1}^{m_n}\Big(\int \mathbb{E}\max_{\phi_q\in\mathcal{B}_{qn}} \big|n^{-1/2}\sum_i h_{qj}(\mathbf{X}_i,\phi_q) - \mathbb{E}\, h_{qj}(\mathbf{X},\phi_q)\big|^2 dq\Big)^{1/2} + o(1) \\
&= o(1)
\end{aligned}$$

where the last equality follows similarly to the proof of $nIII_n = o_p(\sqrt{m_n})$. Consider C_{n2}. Let us introduce the function for $1 \leqslant j \leqslant m_n$ and $1 \leqslant i \leqslant n$

$$t_{qn}(\mathbf{X}_i, \phi_q) := \big(\mathbb{1}\{Y_i \leqslant \varphi(Z_i,q)\} - q\big)\big(F_{m_n}\mathcal{T}\phi_q - F_{m_n}\mathcal{T}\varphi_q\big)(W_i)$$

and the sets $\mathcal{D}_n := \big\{\phi \in \mathcal{B}^{(0,1)} : n\,|||\,\mathcal{T}\phi - \mathcal{T}\varphi|||^2_W \leqslant \sqrt{m_n}\big\}$, $\mathcal{D}_{qn} := \{\phi_q : \phi \in \mathcal{D}_n\} \subset \mathcal{B}$, $\mathcal{G}_q := \{t_{qn} : \phi \in \mathcal{B}\}$, and $\mathcal{G}_{qn} := \{t_{qn} : \phi \in \mathcal{D}_{qn}\}$. We calculate

$$\mathbb{P}\big(|C_{n2}| > \varepsilon\sqrt{m_n}\big) \leqslant \sqrt{n}(\varepsilon\sqrt{m_n})^{-1}\,\mathbb{E}\int \max_{\phi_q\in\mathcal{D}_{qn}}\Big|\frac{1}{\sqrt{n}}\sum_i t_{qn}(\mathbf{X}_i,\phi_q)\Big| dq + o(1).$$

Since p_W is uniformly bounded away from zero, $n\,|||\,\mathcal{T}\phi - \mathcal{T}\varphi|||^2_W \leqslant \sqrt{m_n}$, and $\|F_{m_n}(\mathcal{T}\phi_q - \mathcal{T}\varphi_q)\|_W \leqslant C\|\mathcal{T}\phi_q - \mathcal{T}\varphi_q\|_W$ for all $\phi \in \mathcal{D}_n$ we have $|F_{m_n}(\mathcal{T}\phi_q - \mathcal{T}\varphi_q)(w)| \leqslant C\, m_n^{1/4} n^{-1/2}$ for almost all $0 < q < 1$ and p_W–almost all w. Consequently, $t_{qn}(\mathbf{x}, \phi_q) \leqslant C\, m_n^{1/4} n^{-1/2}$ p_W–almost surely. We conclude by again applying the last display of Theorem 2.14.2 of van der Vaart and Wellner [2000]

$$\mathbb{E}\max_{\phi_q\in\mathcal{D}_{qn}}\Big|\frac{1}{\sqrt{n}}\sum_i t_{qn}(\mathbf{X}_i,\phi_q)\Big| \leqslant C J_{[]}(1, \mathcal{G}_{qn}, L^2_{\mathbf{X}})\, m_n^{1/4} n^{-1/2}.$$

As above it can be seen that $J_{[]}(1, \mathcal{G}_{qn}, L^2_{\mathbf{X}}) < C$ for all $0 < q < 1$. Indeed, from Assumption 2 (ii) we conclude $\|\mathcal{T}\phi - \mathcal{T}\varphi_q\|_W \leqslant (1+\eta)\|T_q(\phi - \varphi_q)\|_W$ and further, Assumption 4 (v) yields $\|F_{m_n}(\mathcal{T}\phi - \mathcal{T}\varphi_q)\|_W \leqslant C(1+\eta)\eta_p\|\phi - \varphi_q\|_Z$. Hence, the mapping $\phi \mapsto F_{m_n}\mathcal{T}\phi$ is Lipschitz continuous at φ_q and we may apply Theorem 2.7.11 of van der Vaart and Wellner [2000] which yields

$$\begin{aligned} N_{[]}(\varepsilon\big(n^{-1}\sqrt{m_n}\big)^{1/2}, \mathcal{G}_n, L^2_{\mathbf{X}}) &\leqslant N_{[]}(\varepsilon, \mathcal{G}_q, L^2_{\mathbf{X}}) \\ &\leqslant N_{[]}\big(\varepsilon, \{F_{m_n}\mathcal{T}\phi - F_{m_n}\mathcal{T}\varphi_q : \phi \in \mathcal{B}\}, L^2_W\big) \\ &\leqslant N\Big(\frac{\varepsilon}{2C}, \mathcal{B}, \|\cdot\|_\infty\Big). \end{aligned}$$

Thereby, $C_{n2} = o_p(\sqrt{m_n})$, which completes the proof. □

In the following we make use of the notation $g_{qj}(\mathbf{X}_i, \phi) := (\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q)f_j(W_i)$, $1 \leqslant j \leqslant m_n$, $1 \leqslant i \leqslant n$, for any $\phi \in \mathcal{B}$.

Proof of Proposition 2.5. For the proof it is sufficient to show $n^{-1}S_n \geqslant \int \|\mathcal{T}\varphi_q - q\|^2_W dq/2 + o_p(1)$. Since $\int \|n^{-1}\sum_i(\mathbb{1}\{Y_i \leqslant \widehat{\varphi}_{qn}(Z_i)\} - \mathbb{1}\{Y_i \leqslant \varphi_q(Z_i)\})f_{\underline{m_n}}(W_i)\|^2 dq = o_p(1)$ (cf. proof of Theorem 2.2 together with Lemma A.2) we obtain

$$\begin{aligned} \int \big\|n^{-1}\sum_i(\mathbb{1}\{Y_i \leqslant \varphi_q(Z_i)\} - q)f_{\underline{m_n}}(W_i)\big\|^2 dq \\ &= \int \big\|\,\mathbb{E}[((\mathcal{T}\varphi_q)(W) - q)f_{\underline{m_n}}(W)]\big\|^2 dq + o_p(1) \\ &\geqslant \int \|\mathcal{T}\varphi_q - q\|^2_W dq/2 + o_p(1), \end{aligned}$$

which proves the result. □

Proof of Proposition 2.6. Since $\varphi_q = \operatorname{argmin}_{\phi\in\mathcal{B}} \|\mathcal{T}\phi - q\|_W$ we obtain as in the proof of Theorem 2.2 by employing the results of Lemma A.2 that

$$S_n = \sum_{j=1}^{m_n} \int \Big|n^{-1/2}\sum_i g_{qj}(\mathbf{X}_i, \varphi_q)\Big|^2 dq + o_p(\sqrt{m_n}).$$

Further, we calculate

$$\begin{aligned} \sum_{j=1}^{m_n} \int \Big|n^{-1/2}\sum_i g_{qj}(\mathbf{X}_i, \varphi_q)\Big|^2 dq &= \sum_{j=1}^{m_n} \int \Big|n^{-1/2}\sum_i \big(g_{qj}(\mathbf{X}_i, \varphi_q) - \mathbb{E}\, g_{qj}(\mathbf{X}_i, \phi)\big)\Big|^2 dq \\ &+ 2\sum_{j=1}^{m_n} \int \Big(n^{-1/2}\sum_i \big(g_{qj}(\mathbf{X}_i, \varphi_q) - \mathbb{E}\, g_{qj}(\mathbf{X}, \varphi_q)\big)\Big)\sqrt{n}\,\mathbb{E}\, g_{qj}(\mathbf{X}, \varphi_q) dq \\ &+ n\sum_{j=1}^{m_n} \int \Big|\,\mathbb{E}\, g_{qj}(\mathbf{X}, \varphi_q)\Big|^2 dq \\ &= I_n + 2II_n + III_n. \end{aligned}$$

We have $3\sqrt{5/m_n}\big(I_n - m_n/6\big) \xrightarrow{d} \mathcal{N}(0,1)$. Further, since $\mathbb{E}[(\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q)^2|W] \leqslant 1$ we obtain

$$\mathbb{E}\,|II_n|^2 \leqslant n\,\mathbb{E}\int \big|(\mathbb{1}\{Y \leqslant \varphi(Z,q)\} - q)\sum_{j=1}^{m_n}\mathbb{E}\,g_{qj}(\mathbf{X},\varphi_q)\big|^2 dq$$
$$\leqslant n\int \Big|\sum_{j=1}^{m_n}\mathbb{E}\,g_{qj}(\mathbf{X},\varphi_q)\Big|^2 dq \leqslant n\int \|\mathcal{T}\varphi_q - q\|_W^2 dq \quad \text{(A.2)}$$

and hence $II_n = O_p((n\int \|\mathcal{T}\varphi_q - q\|_W^2 dq)^{1/2})$. Moreover, since $n\int \|\mathcal{T}\varphi_{qn} - \mathcal{T}\varphi_q\|_W^2 dq = o(\sqrt{m_n})$ and by employing relation (2.13) it is easily seen that

$$\frac{3\sqrt{5}}{\sqrt{m_n}} III_n = \frac{3\sqrt{5}\,n}{\sqrt{m_n}}\int \|\mathcal{T}\varphi_{qn} - q - \delta_n\xi_q\|_W^2 dq + \sum_{j=1}^{\infty}\int \mathbb{E}[\xi_q(W)f_j(W)]^2 dq + o(1),$$

which proves the result. □

Proof of Corollary 2.8. For the proof it is sufficient to show $n^{-1}S_n^* \geqslant \int \|\mathcal{T}\varphi_q - q\|_W^2 dq/2 + o_{p^*}(1)$ with probability approaching one. Chen and Pouzo [2015] show that the bootstrap version of the sieve estimator $\widehat{\varphi}_{qn}^*$ converges at the same rate as $\widehat{\varphi}_{qn}$. In light of the proof of Proposition 2.5, it is sufficient to show

$$\int \big\|n^{-1}\sum_i \varepsilon_i(\mathbb{1}\{Y_i \leqslant \varphi_q(Z_i)\} - q)f_{\underline{m_n}}(W_i)\big\|^2 dq$$
$$= \int \big\|\,\mathbb{E}[((\mathcal{T}\varphi_q)(W) - q)f_{\underline{m_n}}(W)]\big\|^2 dq + o_p(1)$$
$$\geqslant \int \|\mathcal{T}\varphi_q - q\|_W^2 dq/2 + o_p(1),$$

using that ε is independent of W and $\mathbb{E}[\varepsilon] = 1$ as well as $\mathbb{V}ar(\varepsilon) < \infty$, which proves the result. □

Proof of Corollary 3.1. In light of the proof of Theorem 2.2 is sufficient to prove $n\|\mathcal{T}\widehat{\varphi}_{qn}^e - \mathcal{T}\varphi_q\|_W^2 = o_p(\sqrt{m_n})$. Due to Assumption 6 (ii) we obtain as in the proof of Theorem 6 of Chen et al. [2014] that

$$\|\mathcal{T}\widehat{\varphi}_{qn}^e - \mathcal{T}\varphi_q - T_q(\widehat{\varphi}_{qn}^e - \varphi_q)\|_W \leqslant C\|\widehat{\varphi}_{qn}^e - \varphi_q\|_Z^2$$

and consequently,

$$\|\mathcal{T}\widehat{\varphi}_{qn}^e - \mathcal{T}\varphi_q\|_W \leqslant C\big(\|T_q(\widehat{\varphi}_{qn}^e - \varphi_q)\|_W + \|\widehat{\varphi}_{qn}^e - \varphi_q\|_Z^2\big).$$

Moreover, by applying $\sup_y p_{Y|Z,W}(y,Z,W) \leqslant C$ and Jensen's inequality we have

$$\begin{aligned}\|T_q(\widehat{\varphi}_{qn}^e - \varphi_q)\|_W^2 &= \int_{\mathcal{W}} |\int_{\mathcal{Z}} p_{Y|Z,W}(\varphi(z,q),z,w)(\widehat{\varphi}_{qn}^e - \varphi_q)(z)p_{Z|W}(z,w)dz|^2 p_W(w)dw\\ &\leqslant C\|\widehat{\varphi}_{qn}^e - \varphi_q\|_Z^2\\ &= O_p(R_n^e)\\ &= o_p(\sqrt{m_n}/n),\end{aligned}$$

by employing the rate conditions (3.3) and Assumption 6 (iii). □

Proof of Proposition 3.2. Due to the rate restriction (3.3) we may follow the proof of Theorem 2.2 and Corollary 3.1 and hence obtain

$$S_n^{\mathrm{e}}(q) = \sum_{j=1}^{m_n} \Big| n^{-1/2} \sum_i (\mathbb{1}\{Y_i \leqslant \varphi_q^{\mathrm{e}}(Z_i)\} - q) f_j(W_i) \Big|^2 + o_p(\sqrt{m_n}).$$

Thus, by following line by line the Proposition 2.6, we obtain the result. □

A.2. Technical assertions.

We can not apply the consistency and rate of convergence results of Chen and Pouzo [2012] when the null hypothesis H_0 fails. The following Lemma extends their results to possibly misspecified instrumental quantile regression. Recall that under misspecification $\varphi_q = \operatorname{argmin}_{\phi\in\mathcal{B}} \|\mathcal{T}\phi - q\|_W$ does not satisfy $\mathcal{T}\varphi_q = q$.

Lemma A.2. *Let Assumptions 1–4 hold true. Then*

$$|||\widehat{\varphi}_{\cdot n} - \varphi|||_{Z,p}^2 = o_p(1), \tag{A.3}$$

$$|||\mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi|||_W^2 = O_p\Big(\omega_n + \int \|\mathcal{T}\varphi_q - q\|_W^2 dq\Big), \tag{A.4}$$

$$|||\widehat{\varphi}_{\cdot n} - \varphi|||_{Z,p}^2 = O_p\Big(\,|||\, \Pi_{k_n}\varphi - \varphi \,|||_{Z,p}^2 + \tau_{k_n}\big(\omega_n + \int \|\mathcal{T}\varphi_q - q\|_W^2 dq\big)\Big). \tag{A.5}$$

Proof. Proof of (A.3). We define $\mathcal{R}_n := \max\big(n^{-1}l_n, \max_{\phi\in\mathcal{B}_{k_n}} \sum_{j>l_n} \mathbb{E}[(\mathcal{T}\phi(W)-q)f_j(W)]^2\big)$. From the proof of Proposition 2.5 we have that

$$\sum_{j=1}^{l_n} \mathbb{E} \max_{\phi\in\mathcal{B}_{k_n}} \Big| n^{-1} \sum_i \mathbb{1}\{Y_i \leqslant \phi(Z_i)\} f_j(W_i) - \mathbb{E}[\mathbb{1}\{Y \leqslant \phi(Z)\} f_j(W)] \Big|^2 = O(n^{-1}l_n). \tag{A.6}$$

Consequently, we observe

$$\int \big\| n^{-1} \sum_i (\mathbb{1}\{Y_i \leqslant \Pi_{k_n}\varphi_q(Z_i)\} - q) f_{\underline{l_n}}(W_i) \big\|^2 dq \leqslant 2 \int \|\mathcal{T}\Pi_{k_n}\varphi_q - q\|_W^2 dq + O_p(\mathcal{R}_n).$$

Further, using the elementary inequality $(a-b)^2 \geqslant a^2/2 - b^2$ and again applying relation (A.6) gives

$$\begin{aligned}\int \big\| n^{-1} \sum_i (\mathbb{1}\{Y_i \leqslant \phi_q(Z_i)\} - q) f_{\underline{l_n}}(W_i) \big\|^2 dq &\geqslant \int \|F_{l_n}(\mathcal{T}\phi_q - q)\|_W^2 dq/2 \\ &\quad - \sum_{j=1}^{l_n} \max_{\phi\in\mathcal{B}_{k_n}} \Big| n^{-1} \sum_i \mathbb{1}\{Y_i \leqslant \phi(Z_i)\} f_j(W_i) - \mathbb{E}\,\mathbb{1}\{Y \leqslant \phi(Z)\} f_j(W) \Big|^2 \\ &\geqslant C \int \|\mathcal{T}\phi_q - q\|_W^2 dq - O_p(\mathcal{R}_n).\end{aligned}$$

Let us denote $\mathcal{A}_{k_n} = \{\phi \in \mathcal{B}_{k_n}^{(0,1)} : |||\phi - \varphi|||_{Z,p}^2 \geqslant \varepsilon\}$ for some $\varepsilon > 0$. Since $\mathcal{T}$ is continuous and $\varphi_q = \operatorname{argmin}_{\phi\in\mathcal{B}} \|\mathcal{T}\phi - q\|_W$ is unique we have that $\min_{\phi\in\mathcal{A}_{k_n}} \int \|\mathcal{T}\phi_q - q\|_W^2 dq$ is strictly

positive for all $n \geqslant 1$. Therefore, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\,|||\, \widehat{\varphi}_{\cdot n} - \varphi |||_{Z,p}^2 \geqslant \varepsilon \Big) \\
&\leqslant \mathbb{P}\Big(\min_{\phi \in \mathcal{A}_{k_n}} \int \Big\| \sum_i (\mathbb{1}\{Y_i \leqslant \phi(Z_i, q)\} - q) f_{\underline{l_n}}(W_i) \Big\|^2 dq \\
&\qquad\qquad \leqslant \int \Big\| \sum_i (\mathbb{1}\{Y_i \leqslant \Pi_{k_n} \varphi(Z_i, q)\} - q) f_{\underline{l_n}}(W_i) \Big\|^2 dq \Big) \\
&\leqslant \mathbb{P}\Big(\min_{\phi \in \mathcal{A}_{k_n}} \int \|\mathcal{T}\phi_q - q\|_W^2 dq \leqslant \int \|\mathcal{T}\Pi_{k_n}\varphi_q - q\|_W^2 dq + O_p(\mathcal{R}_n) \Big) = o(1)
\end{aligned}$$

since $\int \|\mathcal{T}\Pi_{k_n}\varphi_q - q\|_W^2 dq = \int \|\mathcal{T}\varphi_q - q\|_W^2 dq + o(1)$, $\mathcal{R}_n = o(1)$, and making use of $\min_{\phi \in \mathcal{A}_{k_n}} \int \|\mathcal{T}\phi_q - q\|_W^2 dq > \int \|\mathcal{T}\varphi_q - q\|_W^2 dq + o(1)$. Proof of (A.4). For some $\varepsilon > 0$ let us denote $\mathcal{D}_{k_n} = \{\phi \in \mathcal{B}_{k_n}^{(0,1)} : |||\mathcal{T}\phi - \mathcal{T}\varphi|||_W^2 \geqslant \varepsilon\omega_n\}$. Therefore, we obtain as above

$$\begin{aligned}
&\mathbb{P}\Big(\,|||\, \mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi |||_W^2 \geqslant \varepsilon\omega_n \Big) \\
&\qquad\qquad \leqslant \mathbb{P}\Big(\min_{\phi \in \mathcal{D}_{k_n}} \int \|\mathcal{T}\phi_q - q\|_W^2 dq \leqslant \int \|\mathcal{T}\Pi_{k_n}\varphi_q - q\|_W^2 dq + O_p(\mathcal{R}_n) \Big).
\end{aligned}$$

Further, it holds $\int \|\mathcal{T}\Pi_{k_n}\varphi_q - q\|_W^2 dq \leqslant 2\,|||\,\mathcal{T}\Pi_{k_n}\varphi - \mathcal{T}\varphi\,|||_W + 2\int \|\mathcal{T}\varphi_q - q\|_W^2 dq$. We thus obtain

$$\begin{aligned}
&\mathbb{P}\Big(\,|||\, \mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi |||_W^2 \geqslant \varepsilon\,\omega_n \Big) \\
&\leqslant \mathbb{P}\Big(\min_{\phi \in \mathcal{D}_{k_n}} \int \|\mathcal{T}\phi_q - q\|_W^2 dq \leqslant 2|||\mathcal{T}\Pi_{k_n}\varphi - \mathcal{T}\varphi|||_W^2 + 2\int \|\mathcal{T}\varphi_q - q\|_W^2 dq + O_p(\mathcal{R}_n) \Big).
\end{aligned}$$

For all $\phi \in \mathcal{D}_{k_n}$ and $0 < q < 1$ we have

$$\|\mathcal{T}\phi_q - q\|_W^2 \geqslant \|\mathcal{T}\varphi_q - q\|_W^2 \geqslant \|\mathcal{T}\phi_q - \mathcal{T}\varphi_q\|_W^2/2 - \|\mathcal{T}\phi_q - q\|_W^2$$

and hence, $\|\mathcal{T}\phi_q - q\|_W^2 \geqslant \|\mathcal{T}\phi_q - \mathcal{T}\varphi_q\|_W^2/4$. Thereby, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\,|||\, \mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi |||_W^2 \geqslant \varepsilon\,\omega_n \Big) \\
&\leqslant \mathbb{P}\Big(\frac{1}{4} \min_{\phi \in \mathcal{D}_{k_n}} |||\mathcal{T}\phi - \mathcal{T}\varphi|||_W^2 \leqslant 2\,|||\,\mathcal{T}\Pi_{k_n}\varphi - \mathcal{T}\varphi\,|||_W^2 + 2\int \|\mathcal{T}\varphi_q - q\|_W^2 dq + O_p(\mathcal{R}_n) \Big) \\
&\qquad \leqslant \mathbb{P}\Big(\frac{\varepsilon}{4}\omega_n \leqslant 2\eta \int \|T_q(\Pi_{k_n}\varphi_q - \varphi_q)\|_W^2 dq + 2\int \|\mathcal{T}\varphi_q - q\|_W^2 dq + O_p(\mathcal{R}_n) \Big)
\end{aligned}$$

which goes to zero for all $n \geqslant 1$ as $\varepsilon \to \infty$. Proof of (A.5). Note that $\|T_q(\phi - \varphi_q)\|_W \leqslant (1-\eta)^{-1}\|\mathcal{T}\phi - \mathcal{T}\varphi_q\|_W$ for all ϕ in a sufficiently small neighborhood around φ_q. Thereby, due to (A.3) we obtain

$$|||\widehat{\varphi}_{\cdot n} - \varphi|||_{Z,p}^2 = O_p\Big(\,|||\, \Pi_{k_n}\varphi - \varphi \,|||_{Z,p}^2 + \tau_{k_n} \,|||\, \mathcal{T}\widehat{\varphi}_{\cdot n} - \mathcal{T}\varphi \,|||_W^2 \Big).$$

Hence, the result follows by applying (A.4). □

The following lemma is similar to Lemma A.2 of Breunig [2015]. In the following, however, we provide the proof for the sake of completeness. For all $\phi \in \mathcal{B}$ recall the definition $g_j(\mathbf{X}_i, \phi) = (\mathbb{1}\{Y_i \leqslant \phi(Z_i)\} - q) f_j(W_i)$ for all $1 \leqslant j \leqslant m_n$ and $1 \leqslant i \leqslant n$. Let us introduce $\mathcal{X}_{ii'} := 6\sqrt{5}/(\sqrt{m_n} n) \sum_{j=1}^{m_n} \int g_j(\mathbf{X}_i, \varphi_q) g_j(\mathbf{X}_{i'}, \varphi_q) dq$ and

$$Q_{ni} := \begin{cases} \sum_{l=1}^{i-1} \mathcal{X}_{li}, & \text{for } i = 2, \ldots, n, \\ 0, & \text{for } i = 1 \text{ and } i > n. \end{cases} \tag{A.7}$$

Then clearly

$$\begin{aligned} &3\sqrt{5}/(\sqrt{m_n} n) \sum_{i \neq i'} \sum_{j=1}^{m_n} \int g_j(\mathbf{X}_i, \varphi_q) g_j(\mathbf{X}_{i'}, \varphi_q) dq \\ &\qquad = 6\sqrt{5}/(\sqrt{m_n} n) \sum_{i < i'} \sum_{j=1}^{m_n} \int g_j(\mathbf{X}_i, \varphi_q) g_j(\mathbf{X}_{i'}, \varphi_q) dq = \sum_{i < i'} \mathcal{X}_{ii'} = \sum_{i=1}^{n} Q_{ni}. \end{aligned}$$

Let $\mathcal{B}_{ni} := \mathcal{B}((Z_1, Y_1, W_1), \ldots, (Z_i, Y_i, W_i))$, $1 \leqslant i \leqslant n$, $n \geqslant 1$, be the σ-algebra generated by $(Z_1, Y_1, W_1), \ldots, (Z_i, Y_i, W_i)$. Since $g_j(\mathbf{X}_i, \varphi_q)$, $1 \leqslant i \leqslant n$, are centered random variables it follows that $\{(\sum_{i'=1}^{i} Q_{ni'}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale for each $n \geqslant 1$ and hence $\{(Q_{ni}, \mathcal{B}_{ni}),\ i \geqslant 1\}$ is a Martingale difference array for each $n \geqslant 1$.

Lemma A.3. *Let Q_{ni} be defined as in (A.7). Let Assumption 1 and condition (2.10) be satisfied. Then, we have $\sum_{i=1}^{\infty} Q_{ni} \xrightarrow{d} \mathcal{N}(0, 1)$.*

Proof. For the proof we have to show that the Martingale difference array $\{(Q_{ni}, \mathcal{B}_{ni}),\ i \geqslant 1\}$, $n \geqslant 1$, satisfies the conditions

$$\sum_{i=1}^{\infty} \mathbb{E}\,|Q_{ni}|^2 \leqslant 1 \quad \text{for all } n \geqslant 1, \tag{A.8}$$

$$\sum_{i=1}^{\infty} Q_{ni}^2 = 1 + o_p(1), \tag{A.9}$$

$$\sup_{i \geqslant 1} |Q_{ni}| = o_p(1). \tag{A.10}$$

Then the result follows by Awad [1981]. Proof of (A.8). Since $\mathbb{E}[(\mathbb{1}\{Y \leqslant \varphi(Z, q)\} - q)(\mathbb{1}\{Y \leqslant \varphi(Z, q')\} - q') | W] = \min(q, q') - qq'$ we have

$$\int \Big(\mathbb{E}\left[g_j(\mathbf{X}, \varphi_q) g_{j'}(\mathbf{X}, \varphi_{q'}) \right] \Big)^2 d(q, q') = \int (\min(q, q') - qq')^2 d(q, q')\, \mathbb{1}_{\{j = j'\}} = \mathbb{1}_{\{j = j'\}}/90,$$

where we used that $\mathbb{E}[f_j(W) f_{j'}(W)]^2 = \mathbb{1}_{\{j = j'\}}$ and

$$\begin{aligned} \int (\min(q, q') - qq')^2 d(q, q') &= \int \Big(\int_0^q (q' - qq')^2 dq' + \int_q^1 (q - qq')^2 dq' \Big) dq \\ &= \frac{2}{3} \int q^3 (1 - q)^2 dq \\ &= 1/90. \end{aligned}$$

Observe that $\mathbb{E}[\mathcal{X}_{1i}\mathcal{X}_{1i'}] = 0$ for $i \neq i'$ and thus, for $i = 2, \dots, n$ we have

$$\begin{aligned}
\mathbb{E}\,|Q_{ni}|^2 &= \mathbb{E}\,|\mathcal{X}_{1i} + \cdots + \mathcal{X}_{i-1,i}|^2 \\
&= (i-1)\,\mathbb{E}\,|\mathcal{X}_{12}|^2 \\
&= \frac{(6\sqrt{5})^2(i-1)}{n^2 m_n}\,\mathbb{E}\,\Big|\sum_{j=1}^{m_n}\int g_j(\mathbf{X}_1,\varphi_q)g_j(\mathbf{X}_2,\varphi_q)dq\Big|^2 \\
&= \frac{180\,(i-1)}{n^2 m_n}\sum_{j,j'=1}^{m_n}\int\big(\mathbb{E}[g_j(\mathbf{X},\varphi_q)g_{j'}(\mathbf{X},\varphi_{q'})]\big)^2 d(q,q') \\
&= \frac{2(i-1)}{n^2}.
\end{aligned}$$

Thereby, we conclude

$$\sum_{i=1}^{n}\mathbb{E}\,|Q_{ni}|^2 = \frac{2}{n^2}\sum_{i=1}^{n-1} i = \frac{n(n-1)}{n^2} = 1 - \frac{1}{n} \tag{A.11}$$

which proves (A.8).
Proof of (A.9). Using relation (A.11) we observe

$$\mathbb{E}\,\Big|\sum_{i=1}^{n} Q_{ni}^2 - 1\Big|^2 = \sum_{i=1}^{n}\mathbb{E}\,Q_{ni}^4 + 2\sum_{i<i'}\mathbb{E}\,Q_{ni}^2 Q_{ni'}^2 - 1 + o(1) =: I_n + II_n - 1 + o(1).$$

Consider I_n. Observe that

$$\begin{aligned}
\mathbb{E}\,|Q_{ni}|^4 = \mathbb{E}\,\Big|\sum_{i'=1}^{i-1}\mathcal{X}_{i'i}\Big|^4 &\leqslant \int \mathbb{E}\,\Big|\frac{6\sqrt{5}}{n\sqrt{m_n}}\sum_{j=1}^{m_n} g_j(\mathbf{X}_i,\varphi_q)\sum_{i'=1}^{i-1} g_j(\mathbf{X}_{i'},\varphi_q)\Big|^4 dq \\
&\leqslant \frac{C}{n^4 m_n^2}\sup_{w\in\mathcal{W}}\|f_{\underline{m_n}}(w)\|^4\Big((i-1)\,\mathbb{E}\,\|f_{\underline{m_n}}(W)\|^4 + 3(i-1)(i-2)(\mathbb{E}\,\|f_{\underline{m_n}}(W)\|^2)^2\Big)
\end{aligned}$$

where we used that $\mathbb{E}[g_j(\mathbf{X},\varphi_q)] = 0$ for $0 < q < 1$. Since $\sum_{i=1}^{n} 3(i-1)(i-2) = n(n-1)(n-2)$ we conclude

$$I_n \leqslant C\Big(\frac{n(n-1)}{2n^4}\,\mathbb{E}\,\|f_{\underline{m_n}}(W)\|^4 + \frac{n(n-1)(n-2)}{n^4}(\mathbb{E}\,\|f_{\underline{m_n}}(W)\|^2)^2\Big) = o(1)$$

since $(\mathbb{E}\,\|f_{\underline{m_n}}(W)\|^2)^2 \leqslant \mathbb{E}\,\|f_{\underline{m_n}}(W)\|^4 \leqslant C m_n^2$. We calculate for $i < i'$

$$\begin{aligned}
Q_{ni}^2 Q_{ni'}^2 = \Big(\sum_{k=1}^{i-1}\mathcal{X}_{ki}^2\Big)\Big(\sum_{k=1}^{i'-1}\mathcal{X}_{ki'}^2\Big) &+ \Big(\sum_{k=1}^{i-1}\mathcal{X}_{ki}^2\Big)\Big(\sum_{k\neq k'}^{i'-1}\mathcal{X}_{ki'}\mathcal{X}_{k'i'}\Big) \\
&+ \Big(\sum_{k\neq k'}^{i-1}\mathcal{X}_{ki}\mathcal{X}_{k'i}\Big)\Big(\sum_{k=1}^{i'-1}\mathcal{X}_{ki'}^2\Big) + \Big(\sum_{k\neq k'}^{i-1}\mathcal{X}_{ki}\mathcal{X}_{k'i}\Big)\Big(\sum_{k\neq k'}^{i'-1}\mathcal{X}_{ki'}\mathcal{X}_{k'i'}\Big) \\
&=: A_{ii'} + B_{ii'} + C_{ii'} + D_{ii'}.
\end{aligned}$$

Consider $A_{ii'}$. Exploiting relation (A.11) and using $\sum_{i<i'}(i-1) = \sum_{i'=1}^{n}(i'-1)(i'-2)/2 = n(n-1)(n-2)/6$ and further $\sum_{i<i'}(i-1)(i'-3) = \sum_{i'=1}^{n}(i'-3)(i'-2)(i'-1)/2 =$

$n(n-1)(n-2)(n-3)/8$ we obtain

$$
\begin{aligned}
2\sum_{i<i'} \mathbb{E}\, A_{ii'} &= 4\,\mathbb{E}\,\mathcal{X}_{12}^2\mathcal{X}_{23}^2 \sum_{i<i'}(i-1) + 2(\mathbb{E}\,\mathcal{X}_{12}^2)^2 \sum_{i<i'}(i-1)(i'-3) + o(1) \\
&\leqslant C\frac{n(n-1)(n-2)}{n^4 m_n^2} \sum_{j,l=1}^{m_n} \int \mathbb{E}\left[g_j^2(\mathbf{X},\varphi_q) g_l^2(\mathbf{X},\varphi_{q'})\right] d(q,q') \\
&\qquad\qquad + \frac{n(n-1)(n-2)(n-3)}{n^4} + o(1)
\end{aligned}
$$

since $\int \mathbb{E}[g_j(\mathbf{X},\varphi_q) g_{j'}(\mathbf{X},\varphi_{q'})] d(q,q') = \mathbb{1}\{j=j'\}/90$. Moreover, applying Cauchy Schwarz's inequality twice gives

$$
\sum_{j,l=1}^{m_n} \int \mathbb{E}\left[g_j^2(\mathbf{X},\varphi_q) g_l^2(\mathbf{X},\varphi_{q'})\right] d(q,q') \leqslant \sup_{w\in\mathcal{W}} \|f_{\underline{m_n}}(w)\|^4 \leqslant C m_n^2.
$$

Thereby, it holds $2\sum_{i<i'} \mathbb{E}\, A_{ii'} = 1 + o(1)$. Now consider $B_{ii'}$. Since $\{f_l\}_{l\geqslant 1}$ forms an orthonormal basis on the support of W we obtain

$$
\begin{aligned}
&\mathbb{E}\Big(\sum_{k=1}^{i-1} \mathcal{X}_{ki}^2\Big)\Big(\sum_{k\neq k'}^{i'-1} \mathcal{X}_{ki'}\mathcal{X}_{k'i'}\Big) = 2\sum_{k=1}^{i-1} \mathbb{E}\,\mathcal{X}_{ki}^2 \mathcal{X}_{ki'} \mathcal{X}_{ii'} \\
&\leqslant \frac{C(i-1)}{n^4 m_n^2} \sum_{j,j'=1}^{m_n} \int \mathbb{E}\Big| g_j(\mathbf{X}_1,\varphi_q) g_{j'}(\mathbf{X}_1,\varphi_q) g_j(\mathbf{X}_2,\varphi_q) g_{j'}(\mathbf{X}_2,\varphi_{q'}) \\
&\qquad\qquad \times q(1-q) \sum_{l=1}^{m_n} g_l^2(\mathbf{X}_1,\varphi_q)\Big| d(q,q',q'') \\
&\qquad\qquad \leqslant \frac{C(i-1)}{n^4 m_n}\Big(\sum_{j,j'=1}^{m_n} \int \mathbb{E}\,|g_j(\mathbf{X},\varphi_q) g_{j'}(\mathbf{X},\varphi_q)|^2 d(q,q')\Big) \leqslant \frac{C(i-1)m_n}{n^4}.
\end{aligned}
$$

This, together with relation (A.11), yields $\sum_{i<i'} \mathbb{E}\, B_{ii'} = o(1)$. Further, it is easily seen that $\sum_{i<i'} \mathbb{E}\, C_{ii'} = o(1)$. Consider $D_{ii'}$. Using twice the law of iterated expectation gives

$$
\begin{aligned}
&\mathbb{E}\, D_{ii'} = \mathbb{E}\Big(\sum_{k\neq k'}^{i-1} \mathcal{X}_{ki}\mathcal{X}_{k'i}\Big)\Big(\sum_{k\neq k'}^{i'-1} \mathcal{X}_{ki'}\mathcal{X}_{k'i'}\Big) = 4\sum_{k<k'}^{i-1} \mathbb{E}\,\mathcal{X}_{ki}\mathcal{X}_{k'i}\mathcal{X}_{ki'}\mathcal{X}_{k'i'} \\
&= 4\sum_{k<k'}^{i-1} \mathbb{E}\left[\mathcal{X}_{ki}\mathcal{X}_{k'i}\,\mathbb{E}[\mathcal{X}_{ki'}\mathcal{X}_{k'i'}|(Y_k,Z_k,W_k),(Y_{k'},Z_{k'},W_{k'}),(Y_i,Z_i,W_i)]\right] \\
&\leqslant \frac{C}{n^2 m_n} \sum_{k<k'}^{i-1} \mathbb{E}\Big[\mathbb{E}[\mathcal{X}_{ki}\mathcal{X}_{k'i}|(Y_k,Z_k,W_k),(Y_{k'},Z_{k'},W_{k'})] \\
&\qquad\qquad \times \sum_{j,j'=1}^{m_n} \int \mathbb{E}[g_j(\mathbf{X},\varphi_q) g_{j'}(\mathbf{X},\varphi_{q'})] g_j(\mathbf{X}_k,\varphi_q) g_{j'}(\mathbf{X}_{k'},\varphi_{q'}) d(q,q')\Big] \\
&\leqslant \frac{C}{n^4 m_n^2} \int \mathbb{E}\Big|\sum_{j,j'=1}^{m_n} \mathbb{E}[g_j(\mathbf{X},\varphi_q) g_{j'}(\mathbf{X},\varphi_{q'})] g_j(\mathbf{X}_1,\varphi_q) g_{j'}(\mathbf{X}_2,\varphi_{q'})\Big|^2 d(q,q')(i-1)(i-2) \\
&\qquad\qquad \leqslant \frac{C}{n^4 m_n}(i-1)(i-2).
\end{aligned}
$$

again using that $\mathbb{E}[g_j(\mathbf{X}, \varphi_q) g_{j'}(\mathbf{X}, \varphi_{q'})]$ is only different from zero whenever $j = j'$. Consequently, we obtain

$$\sum_{i<i'} \mathbb{E}\, D_{ii'} \leqslant \frac{C}{n^4 m_n} \sum_{i<i'} (i-1)(i-2) = \frac{C\, n(n-1)(n-2)(n-3)}{m_n n^4} = o(1)$$

and hence $2 \sum_{i<i'} \mathbb{E}\, Q_{ni}^2 Q_{ni'}^2 = 1 + o(1)$.
Proof of (A.10). Note that $\mathbb{P}\big(\sup_{i \geqslant 1} |Q_{ni}| > \varepsilon\big) \leqslant \sum_{i=1}^n \mathbb{P}\big(Q_{ni}^2 > \varepsilon^2\big)$ and, hence the assertion follows from the Markov inequality. □