« Powerful nonparametric checks for quantile regression»

Samuel Maistre, Pascal Lavergne and Valentin Patilea

Toulouse
School
of Economics

Powerful nonparametric checks for quantile regression

Samuel Maistre[*], Pascal Lavergne[†] and Valentin Patilea[‡]

June 2014

Abstract

We address the issue of lack-of-fit testing for a parametric quantile regression. We propose a simple test that involves one-dimensional kernel smoothing, so that the rate at which it detects local alternatives is independent of the number of covariates. The test has asymptotically gaussian critical values, and wild bootstrap can be applied to obtain more accurate ones in small samples. Our procedure appears to be competitive with existing ones in simulations. We illustrate the usefulness of our test on birthweight data.

Keywords: Goodness-of-fit test, U-statistics, Smoothing.

[*]CREST (Ensai), France. Email: samuel.maistre@ensai.fr

[†]Toulouse School of Economics, France. Email: pascal.lavergne@ut-capitole.fr

[‡]CREST (Ensai), France. Email: patilea@ensai.fr

1 Introduction

Quantile regression, as introduced by Koenker and Bassett (1978), has emerged as an alternative to mean regression. It allows for a richer data analysis by exploring the effect of covariates at different quantiles of the conditional distribution of the variable of interest. Parametric quantile regression generalizes usual regression are is particularly valuable if variables have asymmetric distributions or heavy tails. Koenker's monograph (2005) and the review of Yu et al. (2003) detail the theory and practice of quantile regression.

As in any statistical modeling exercice, it is crucial to check the fit of a parametric quantile model. There has been a large effort devoted to testing of the fit of parametric mean regressions, however only few lack-of-fit tests of parametric quantile regressions. He and Zhu (2003) extend the approach of Stute (1997) and is based on a vector-weighted cumulative summed process of the residuals. Bierens and Ginther (2002) generalize the integrated conditional moment test of Bierens and Ploberger (1997) to quantile regression. In both cases, the limit distribution of the test statistic is a non-linear functional of a Gaussian process, so that implementation may require rather involved computations to obtain critical values. Zheng (1998) use kernel smoothing over the design space, to obtain an asymptotically pivotal test statistic. Horowitz and Spokoiny (2002) extend such an approach and propose an adaptive procedure to choose the smoothing parameter. As in any multidimensional nonparametric problem, the curse of dimensionality may be detrimental to the performances of the test, see e.g. Lavergne and Patilea (2012) for illustrations.

In this paper, we introduce a new testing methodology that avoids multidimensional smoothing, but still yield an omnibus test. Our test has three specific features. First, it does not require smoothing with respect to all covariates under test. This allows to mitigate the curse of dimensionality that appears with nonparametric smoothing, hence improving the power properties of the test. Second, the test statistic is asymptotically pivotal, while wild bootstrap can be used to obtain small samples critical values of the test. This yields a test whose level is well controlled by bootstrapping, as shown in simulations. Third, our test equally applies whether some of the covariates are discrete.

The paper is organized as follows. In Section 2, we present our testing procedure, we study its asymptotic behavior under the null hypothesis and under a sequence of local alternatives, and we establish the validity of wild bootstrap. In Section 3, we compare the small sample behavior of our test to some existing procedures, and we illustrate its use on birthweight data. Section 3 concludes. Section 4 gathers our technical proofs.

2 Lack-of-Fit Test for Quantile Regression

2.1 Principle and Test

Consider modeling the quantile of a real random variable Y conditional upon covariates $Z \in \mathbb{R}^q$, $q \geq 1$. We assume that $Z = (W, X')'$, where W is continuous and admits a density with respect to the Lebesgue measure, while X may include both continuous and discrete variables. Formally, if $F(\cdot \mid z)$ denotes the conditional distribution of Y given $Z = z$, the τ-th conditional quantile is $Q_\tau(z) = \inf\{y : F(y \mid z) \geq \tau\}$. Assuming $F(\cdot \mid z)$ is absolutely continuous for almost all z, this is equivalent to $F(Q_\tau(z) \mid z) = \tau$. The parametric quantile regression model of interest posits that the conditional τ-th quantile of Y is given by $g(Z; \beta_0)$, where $g(\cdot; \beta)$ is known up to the parameter vector $\beta \in B \subset \mathbb{R}^p$, that is,

$$Y = g(Z; \beta_0) + \varepsilon, \qquad F\left(g(Z; \beta_0) \mid Z\right) = \tau. \tag{2.1}$$

The validity of our parametric quantile regression is thus equivalent to

$$H_0 : \exists \beta_0 \in B : F(g(Z; \beta_0) \mid Z) - \tau = \mathbb{E}\left\{\mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau \mid Z\right\} = 0 \text{ a.s.} \tag{2.2}$$

Hence testing the the correct specification of our parametric quantile regression models reduces to testing a zero conditional mean hypothesis. The alternative hypothesis is then

$$H_1 : \mathbb{P}\left[\mathbb{E}\left\{\mathbb{I}\{Y \leq g(Z; \beta)\} - \tau \mid Z\right\} = 0\right] < 1 \quad \text{for any } \beta \in B.$$

The key element of our testing approach is the following lemma. See also Lavergne et al. (2014) for a related result. First let us introduce some notation. Hereafter, if $g : \mathbb{R}^k \to \mathbb{R}$ is an integrable function, $\mathcal{F}[g]$ denotes its Fourier transform, that is

$$\mathcal{F}[g](t) = \int_{\mathbb{R}^k} \exp(-2\pi i t' u) g(u) du.$$

Lemma 2.1. *Let* (W_1, X_1, U_1) *and* (W_2, X_2, U_2) *be two independent draws of* (W, X, u), *and* $K(\cdot)$ *and* $\psi(\cdot)$ *even functions with (almost everywhere) positive Fourier integrable transforms. Define*

$$I(h) = \mathbb{E}\left[U_1 U_2 h^{-p} K\left((W_1 - W_2)/h\right) \psi(X_1 - X_2)\right].$$

Then for any $h > 0$, $\mathbb{E}[U \mid W, X] = 0$ *a.s.* $\Leftrightarrow I(h) = 0$.

Proof. Let $\langle \cdot, \cdot \rangle$ denote the standard inner product and $\mathcal{F}[K]$ be the Fourier transform of $K(\cdot)$. Using Fourier Inversion Theorem, change of variables, and elementary properties of conditional expectation,

$$
\begin{aligned}
I(h) &= \mathbb{E}\left[U_1 U_2 \int_{\mathbb{R}^p} e^{2\pi i \langle t, W_1 - W_2 \rangle} \mathcal{F}[K](th)\, dt \int_{\mathbb{R}^q} e^{2\pi i \langle s, X_1 - X_2 \rangle} \mathcal{F}[\psi](s)\, ds\right] \\
&= \int_{\mathbb{R}^q} \int_{\mathbb{R}^p} \left|\mathbb{E}\left[\mathbb{E}[U \mid W, X]\, e^{2\pi i \{\langle t, W \rangle + \langle s, X \rangle\}}\right]\right|^2 \mathcal{F}[K](th)\, \mathcal{F}[\psi](s)\, dt ds.
\end{aligned}
$$

Since the Fourier transforms $\mathcal{F}[K]$ and $\mathcal{F}[\psi]$ are strictly positive, $I(h) = 0$ iff

$$\mathbb{E}\left[\mathbb{E}[U \mid W, X] e^{2\pi i\{\langle t, W\rangle + \langle s, X\rangle\}}\right] = 0 \quad \forall t, s \Leftrightarrow \mathbb{E}[U \mid W, X] = 0 \qquad \text{a.s.} \qquad \blacksquare$$

From the above results, it is sufficient to test whether $I(h) = 0$ for any arbitrary h. We chose to consider a sequence of h decreasing to zero when the sample size increases, which is one of the ingredient that allows to obtain a tractable asymptotic distribution for the test statistic. Assume we have at hand a random sample (Y_i, W_i, X_i), $1 \leq i \leq n$, from (Y, W, X). Then we can estimate $I(h)$ by the second-order U-statistic

$$I_n(\beta_0) = I_n(\beta_0; h) = \frac{1}{n(n-1)} \sum_{1 \leq j \neq i \leq n} U_i(\beta_0) U_j(\beta_0) \frac{1}{h} K_h(W_i - W_j) \psi(X_i - X_j)$$

where $U_i(\beta) = \mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau$ and $K_h(\cdot) = K(\cdot/h)$.

For estimating β_0, we follow Koenker and Bassett (1978), who showed that under (2.1) a consistent estimator of β_0 is obtained by minimizing

$$\arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(Y_i - g(Z_i; \beta)), \tag{2.3}$$

where $\rho_\tau(e) = (\tau - \mathbb{I}(e < 0)) e$ is the so-called check function. While this is not a differentiable optimization problem, it is convex and tractable, see e.g. Koenker (2005) for some computational algorithms. Let us define

$$T_n = n h^{1/2} \frac{I_n(\widehat{\beta})}{v_n} \quad \text{where} \quad v_n^2 = \frac{2\tau^2(1-\tau)^2}{n(n-1)} \sum_{j \neq i} h^{-1} K_h^2(W_i - W_j) \psi^2(X_i - X_j). \tag{2.4}$$

An asymptotic α-level test of H_0 is then

Reject H_0 if $T_n \geq z_\alpha$, where z_α is the $(1-\alpha)$−quantile of the standard normal distribution.

Our test statistic is very similar to the one proposed by Zheng (1998), but the latter uses smoothing on all components of Z while we smooth only on the first component W.

The statistic v_n^2 is the variance of $n h^{1/2} I_n(\beta_0)$ conditional on the Z_i under H_0. In general, v_n^2 does not consistently estimate the conditional variance of $n h^{1/2} I_n(\beta)$ under the alternative hypothesis. In some cases v_n^2 overestimates this conditional variance (this is certainly the case for misspecified median regression model because $\tau(1-\tau)$ attains the maximum value at $\tau = 1/2$), so that the test may suffer some power loss. In a mean regression context, Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005) proposed to use a nonparametric estimator of the conditional variance. This might be adapted to quantile regression, but in simulations our test appears to be well-behaved and more powerful than competitors, so we decided in favor of the simplest estimator v_n^2.

4

2.2 Behavior Under the Null Hypothesis

To derive the asymptotic properties of our lack-of-fit test, we introduce our set of assumptions on the data-generating process, the parametric model (2.1), the functions $K(\cdot)$ and $\psi(\cdot)$, and the bandwidth h.

Assumption 2.1. *(a) The random vectors $(\varepsilon_1, Z_1')', \ldots, (\varepsilon_n, Z_n')'$ are independent copies of the random vector $(\varepsilon, Z')' \in \mathbb{R}^{1+q}$. The conditional τth quantile of ε given $Z = (W, X')'$ is equal to zero.*

(b) The variable W admits an absolutely continuous density with the respect of the Lebesgue measure on the real line.

(c) The conditional density $f_\varepsilon(\cdot \mid z)$ of ε given $Z = z$ is uniformly bounded. There exists $a > 0$ such that $f_\varepsilon(\cdot \mid z)$ is differentiable on $(-a, a)$ for any z with $|f_\varepsilon'(0 \mid z)| \leq C\infty$. Moreover, the derivatives $f_\varepsilon'(\cdot \mid z)$ satisfy a uniform Hölder continuity condition, that is there exist positive constants C_2 and c independent of z such that $\forall |u_1|, |u_2| \leq a$, $|f_\varepsilon'(u_1 \mid z) - f_\varepsilon'(u_2 \mid z)| \leq C_2 |u_1 - u_2|^c$.

Assumption 2.2. *(a) The parameter space B is a compact convex subset of $\mathbb{R}^p$. β_0 is the unique solution of $\min_B \mathbb{E}[\rho_\tau(Y - g(Z, \beta))]$ and is an interior point of B.*

(b) The matrix

$$\mathbb{E}\left[f_\varepsilon(0 \mid Z)\frac{\partial}{\partial \beta}g(Z; \beta_0)\frac{\partial}{\partial \beta'}g'(Z; \beta_0)\right]$$

is finite and nonsingular.

(c) There exists functions $A(\cdot)$, $B(\cdot)$, and $D(\cdot)$, with $\mathbb{E}[A^4(Z)]$, $\mathbb{E}[B^2(Z)] < \infty$, and $\mathbb{E}[D^4(Z)]$, such that

$$\left\|\frac{\partial}{\partial \beta}g(z; \beta)\right\| \leq A(z), \qquad \left\|\frac{\partial}{\partial \beta}g(z; \beta)\frac{\partial}{\partial \beta'}g'(z; \beta)\right\| \leq D(z) \qquad \text{for any } \beta,$$

$$\left\|\frac{\partial}{\partial \beta}g(z; \beta_1) - \frac{\partial}{\partial \beta}g(z; \beta_2)\right\| \leq B(z)\|\beta_1 - \beta_2\| \qquad \text{for any } z, \beta_1, \beta_2.$$

(d) The class of functions $\{g(Z; \beta) : \beta \in B\}$ is a Vapnik-Červonenkis (VC) class.

Assumption 2.3. *(a) The function $K(\cdot)$ is a bounded symmetric univariate density of bounded variation with positive Fourier transform.*

(b) The function $\psi(\cdot)$ is a bounded symmetric multivariate function with positive Fourier transform.

(c) $h \to 0$ and $n^\alpha h^2 \to \infty$ for some $\alpha \in (0, 1)$ as $n \to \infty$.

Our assumptions combine standard assumptions for parametric quantile regression estimation and specific ones for our lack-of-fit test. Among the latter, the conditions on the error term ε impose neither independence of ε and Z, nor a specific form of dependence such as $\varepsilon = s(Z)e$ with e independent of Z as in He and Zhu (2003). Assumption 2.2(d) is a mild technical condition that guarantees suitable uniform rates of convergence for some $U-$processes appearing in the proofs. This condition is satisfied for many parametric models, for instance when $g(Z, \beta) = q(Z'\beta)$ with $q : \mathbb{R} \to \mathbb{R}$ monotone or of bounded variation, see e.g. van der Vaart and Wellner (1996, Section 2.6). Also, if there is $\beta \in B$ such that $g(Z, \beta)$ is squared integrable, then Assumption 2.2(d) follows from 2.2(c). Assumptions on $K(\cdot)$ allows for the use of a triangular, normal, logistic, Student (including Cauchy), or Laplace densities. For $\psi(\cdot)$, one can choose e.g. $\psi(x) = \exp(-\|x\|^2)$, or any multivariate extension of the aforementioned densities. Restrictions on the bandwidth are compatible with optimal choices for regression estimation, see e.g. Härdle and Marron (1985), and for regression checks, see Guerre and Lavergne (2002) and Horowitz and Spokoiny (2002). The following theorem states the asymptotic validity of our test.

Theorem 2.2. *Under the Assumptions 2.1 to 2.3, the test based on T_n has asymptotic level α under H_0.*

2.3 Behavior under Local Alternatives

We now investigate the behavior of our test when H_0 does not hold, and specifically we consider a sequence of local alternatives of the form

$$H_{1n} : \ Y = g(Z; \beta_0) + r_n \delta(Z) + \varepsilon, \qquad F(g(Z; \beta_0) \mid Z) = \tau, \qquad (2.5)$$

where r_n, $n \geq 1$, is a sequence of real numbers tending to zero and $\delta(Z)$ is a real-valued function satisfying

$$\mathbb{E}\left[f_\varepsilon(0 \mid Z) \delta(Z) \frac{\partial}{\partial \beta} g(Z; \beta_0) \right] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(Z)] < \infty. \qquad (2.6)$$

This condition ensures that our sequence of models (2.5) does not belong to the null hypothesis H_0. We do not impose any smoothness restriction on the function $\delta(\cdot)$ as is frequent in this kind of analysis, see e.g. Zheng (1998). As shown in Lemma 4.1 in the Proofs section, $\widehat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2} + r_n^2)$ under H_{1n}. Our next result states that these local alternatives can be detected whenever $r_n^2 n h^{1/2} \to \infty$. Hence our test *does not suffer from the curse of dimensionality* against local alternatives, since its power is unaffected by the number of regressors.

Theorem 2.3. *Under Assumptions 2.1 to 2.3, the test based on T_n is consistent against the sequence of alternatives H_{1n} with $\delta(Z)$ satisfying (2.6) if $r_n^2 n h^{1/2} \to \infty$.*

2.4 Bootstrap Critical Values

The asymptotic approximation of the behavior of T_n may not be satisfactory in small samples as is customary in smoothing-based lack-of-fit tests. This motivates the use of bootstrapping for obtaining critical values. The distribution of T_n depends weakly on the distribution of the error term ε, because $\mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau$ under H_0 is a Bernouilli random variable irrespective of the particular distribution of ε. The same phenomenon is noted by Horowitz and Spokoiny (2002) for their test statistic. Their proposal is thus to naively (or nonparametrically) bootstrap from the empirical distribution of the residuals. This is a valid bootstrap procedure when errors are identically distributed, and it remains asymptotically valid for non identically distributed errors. A first possibility is thus to adopt naive residual bootstrap for our test. Alternatively, He and Zhu (2003) note that one could use any continuous distribution with the τ-th quantile equal to 0. This constitutes a second possibility. While asymptotically valid, these two methods do not account for potential heteroscedastic errors. Thus a third possibility is the wild bootstrap method for quantile regression introduced by Feng et al. (2011). The wild bootstrap procedure for our test works as follows.

1. Let $\widehat{\varepsilon}_i = Y_i - g(Z_i; \widehat{\beta})$, $1 \leq i \leq n$, and $w_1, \cdots w_n$ be bootstrap weights generated independently from a two-point mass distribution with probabilities $1 - \tau$ and τ at $2(1 - \tau)$ and -2τ. Compute $\varepsilon_i^* = w_i|\widehat{\varepsilon}_i|$ and $Y_i^* = g(Z_i; \widehat{\beta}) + \varepsilon_i^*$ for each $i = 1, ..., n$.

2. Use the bootstrap data set $\{Y_i^*, Z_i : i = 1, ..., n\}$ to compute the estimator $\widehat{\beta}^*$, the new $U_i^*(\widehat{\beta}^*) = \mathbb{I}\{Y_i^* \leq g(Z_i; \widehat{\beta}^*)\} - \tau$, and the new test statistic T_n^*.

3. Repeat Steps 1 et 2 many times, and estimate the α-level critical value z_α^* by the $(1 - \alpha)$-th quantile of the empirical distribution of T_n^*.

The bootstrap test then rejects H_0 if $T_n \geq z_\alpha^*$. Alternatively, one could resample residuals in Step 1 by naive bootstrap, or obtain ε_i^* by random draws from e.g. a uniform law on the interval $[-\tau, 1 - \tau]$. The following theorem yields the asymptotic validity of the bootstrap test.

Theorem 2.4. *Under the conditions of Theorem 2.2,*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}\left(T_n^* \leq t \mid Y_1, Z_1, ..., Y_n, Z_n\right) - \Phi(t)| \xrightarrow{p} 0,$$

where $\Phi(\cdot)$ is the standard normal distribution function.

3 Numerical Evidence

3.1 Small Sample Performances

We investigated the performances of our procedure for testing lack-of-fit of a linear median regression for two setups considered by He and Zhu (2003), namely

$$Y = 1 + W + X + \delta \left(W^2 + WX + X^2 \right) + \varepsilon \,, \tag{3.1}$$

$$Y = \delta \log \left(1 + W^2 + X^2 \right) + \varepsilon \,, \tag{3.2}$$

where W follows a standard normal, and X independently follows a binomial of size 5 and probability of success 0.5. For the error term, we considered the three distributions $\mathcal{N}(0, 1)$, $\log \mathcal{N}(0, 1) - 1$ and $\mathcal{N}\left(0, \left(1 + W^2\right)/2\right)$.

For implementation, we chose $\psi(\cdot)$ as the standard normal density and $K(\cdot)$ as triangle density with variance one. We set $\delta = 0$ in Model (3.1) to evaluate the comparative performances of the three possible bootstrapping procedures. Figure 1 reports ou results based on 5000 replications for a sample size of $n = 100$ at nominal level 10%, when the bandwidth is $h = cn^{-1/5}$ with c varying. The three bootstrap methods yield accurate levels for any bandwidth choice when errors are identically distributed, while the use of asymptotic critical values yield large underrejection. In the heteroscedastic case, however, only the wild bootstrap yield an empirical level close to 10%, while the use of naive or uniform bootstrap results in a severely oversized test.

Next, we investigated the power of our test for Models (3.1) and (3.2) with either standard gaussian or heteroscedastic gaussian errors. We compared our test to the one proposed by He and Zhu (2003, hereafter HZ), based on

$$\max_{\|\mathbf{a}\|=1} n^{-1} \sum_{i=1}^{n} \left(\mathbf{a}' \mathbf{R}_n \left(X_i \right) \right)^2 \quad \text{where} \quad \mathbf{R}_n(\mathbf{t}) = n^{-1/2} \sum_{j=1}^{n} \left(\tau - \mathbb{I} \left[U_j(\widehat{\beta}) < 0 \right] \right) Z_j \mathbb{I}\left(Z_j \leq \mathbf{t} \right) \,.$$

We also computed the statistic proposed by Zheng (1998), which in our setup writes

$$\frac{h^{q/2}}{\widehat{\sigma}(n-1)} \sum_{j \neq i} U_i(\widehat{\beta}) U_j(\widehat{\beta}) h^{-q} \tilde{K} \left(\frac{W_i - W_j}{h}, \frac{X_i - X_j}{h} \right)$$

where

$$\widetilde{\sigma}^2 = \frac{2\tau^2 (1-\tau)^2}{n(n-1)} \sum_{j \neq i} h^{-q} \tilde{K}^2 \left(\frac{W_i - W_j}{h}, \frac{X_i - X_j}{h} \right) \,,$$

and $\tilde{K}$ is a triangle kernel applied to the norm of its argument. We apply the wild bootstrap procedure to compute the critical values of all tests. Figure 2 reports power curves of the
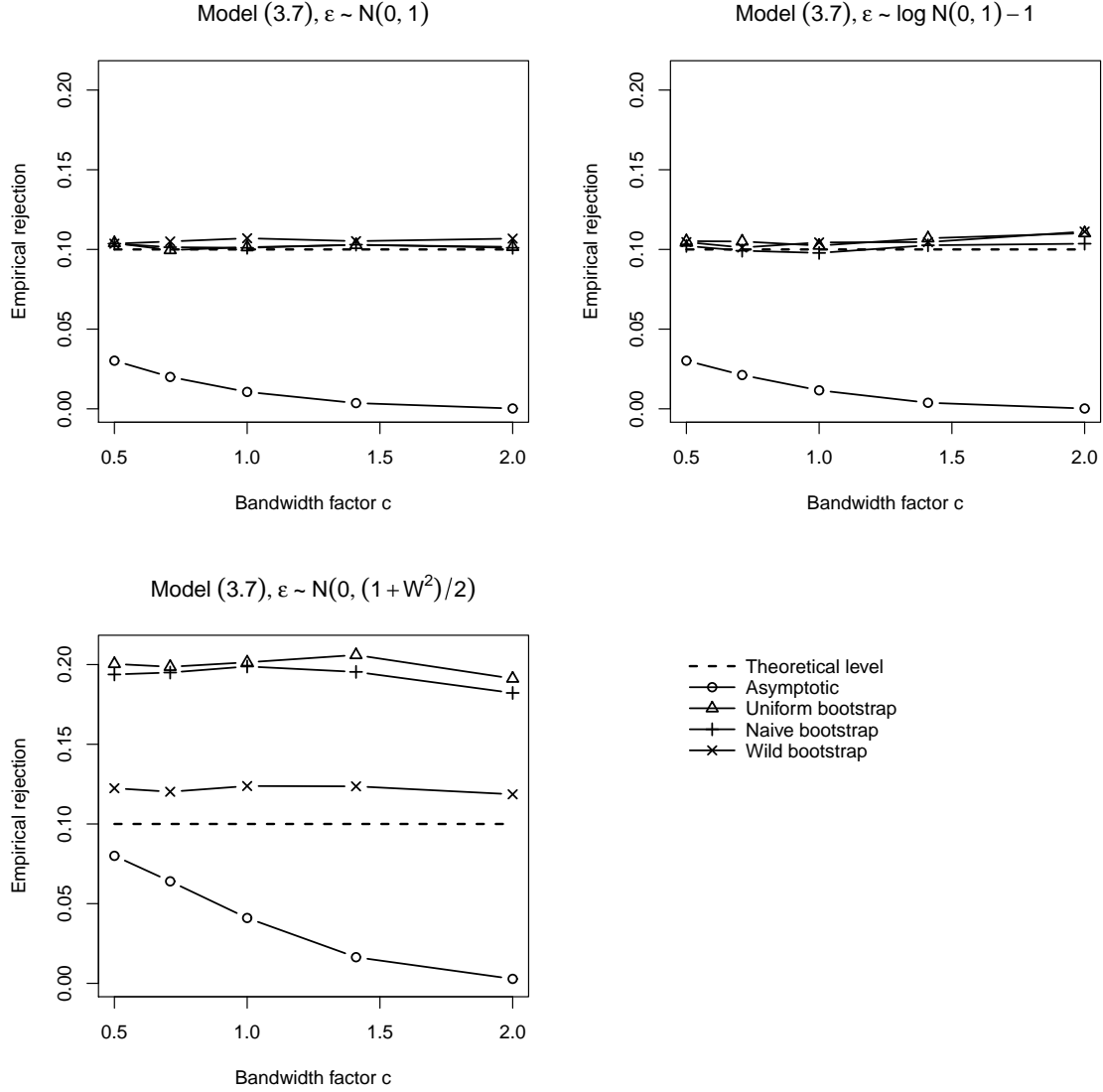
Figure 1: Empirical rejections under H_0 as a function of the bandwidth, $n = 100$

different tests as a function of δ based on 2500 replications. For the linear Model (3.1), all tests perform almost similarly. Our test is a bit more powerful, especially for a larger bandwidth, which was expected given our theoretical analysis. For the nonlinear Model (3.2), the power advantage of our test is more pronounced. Its power can be as large as twice the power of the test by He and Zhu (2003).

3.2 Empirical Illustration

We studied some parametric quantile models for children birthweight using data analyzed by Abrevaya (2001) and Koenker and Hallock (2001), who gave a detailed data description. We focused on median regression and the 10th percentile quantile regression. Models are
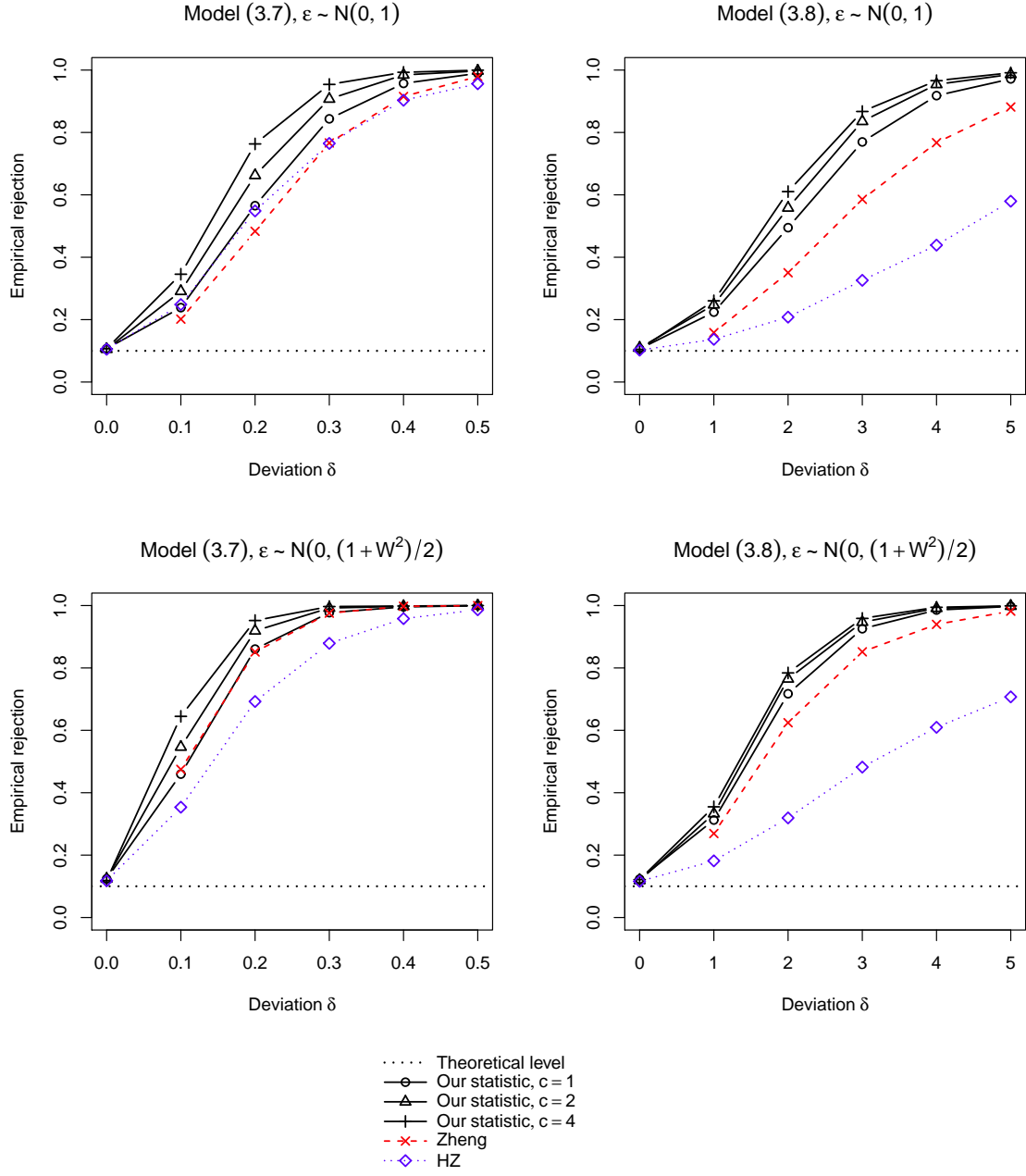
Figure 2: Power curves for models (3.1) and (3.2), $n = 100$.

estimated and tested on a subsample of 1168 smoking college graduate mothers. We first analyzed the simple model considered by He and Zhu (2003), which is linear in weight gain during pregnancy (WTGAIN), average number of cigarettes per day (CIGAR), and age (AGE). When implementing our test, we chose age as the W variable, and we standardize all explanatory variables. Other details are identical to what was done in simulations. For both quantiles, HZ test does not reject this specification. Our test does not reject the linear median regression at 10% level, but detects misspecification for the lower decile regression when $c = 2$.

Since the more detailed analysis of Abrevaya (2001) and Koenker and Hallock (2001) suggests that birthweight is quadratic in age, we then considered this variation. None of the tests detects a misspecified model. Finally, we considered a more complete model similar to Abrevaya (2001), where we added the explanatory binary variables BOY (1 if child is male), BLACK (1 if mother is black), MARRIED (1 if married), and NOVISIT (1 if no prenatal visit during the pregnancy). HZ test does not reject the model at either quantiles. Our test however indicates a misspecified median regression model at 10% level, while it does not reject the model for the lower decile. Our limited empirical exercice suggests that our new test, beside existing procedures such as the test by He and Zhu (2003), is a valuable addition to the practitioner toolbox.

Table 1: Application: estimation results and tests p-values

	$\tau = 0.5$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.1$
CIGAR	-5.35	-7.53	-5.05	-8.36	-5.07	-8.07
	(2.28)	(4)	(2.3)	(3.53)	(2.36)	(3.25)
WTGAIN	8.09	14.73	7.69	14.96	8.31	15.91
	(1.33)	(0.75)	(1.32)	(1.2)	(1.31)	(1.4)
AGE	-9.34	-5.13	43.6	133.67	78.59	117.62
	(3.82)	(4.47)	(50.59)	(30.11)	(45.85)	(48.42)
AGESQ			-0.84	-2.23	-1.38	-1.94
			(0.81)	(0.5)	(0.72)	(0.82)
BOY					137.22	-5.22
					(34.35)	(47.33)
BLACK					-177.78	-124.18
					(75.09)	(69.17)
MARRIED					21.62	41.75
					(48.39)	(54.66)
NOVISIT					-211.62	-275.15
					(406.72)	(112.5)
HZ	0.347	0.227	0.266	0.356	0.272	0.135
Our test c=1	0.791	0.165	0.738	0.942	0.068	0.972
Our test c=2	0.704	0.044	0.741	0.968	0.078	0.796

4 Proofs

We first recall some definitions. For the definition of a VC-class, we refer to Section 2.6.2 of van der Vaart and Wellner (1996). Next, let $\mathcal{G}$ be a class of real-valued functions on a set $\mathcal{S}$. We call $\mathcal{G}$ an *Euclidean(c,d)* family of functions, or simply *Euclidean*, for the envelope G if there exists positive constants c and d with the following properties: if $0 < \epsilon \leq 1$ and λ is a measure for which $\int G^2 d\lambda < \infty$, then there are functions $g_1, \ldots, g_j$ in $\mathcal{G}$ such that (i) $j \leq c\epsilon^{-d}$; and (ii) for each g in $\mathcal{G}$ there is an g_i with $\int |g - g_i|^2 d\lambda \leq \epsilon^2 \int G^2 d\lambda$. The constants c and d must not depend on λ. See e.g. Nolan and Pollard (1987) or Sherman (1994). Recall that if $\mathcal{F}$ is a VC-class of functions then the class $\{\mathbb{I}\{f \geq 0\} : f \in \mathcal{F}\}$ is Euclidean for the envelope $F \equiv 1$, see van der Vaart and Wellner (1996) Lemma 2.6.18(iii) and Theorem 2.6.7 or Pakes and Pollard (1989). Bellow, we shall use this property with the VC-classes of functions of $\{\varepsilon + g(Z, \beta_0) - g(Z, \beta) : \beta \in B\}$ and $\{\varepsilon + g(Z, \beta_0) + r_n\delta(Z) - g(Z, \beta) : \beta \in B\}$.

In the following, $F_\varepsilon(\cdot \mid x)$ is the conditional distribution function of ε given $Z = z$; that means $F_\varepsilon(0 \mid \cdot) \equiv \tau$. Below C, C_1, C_2,... denote constants, not necessarily the same as before and possibly changing from line to line.

4.1 Proof of Theorem 2.2

Proof. First, we prove that if $\mathcal{H}_0$ holds

$$n\sqrt{h}\left\{W_n(\widehat{\beta}) - W_n(\beta_0)\right\} = o_{\mathbb{P}}(1). \tag{4.1}$$

Let us introduce some simplifying notation:

$$G_i(\beta, \beta_0) = g(Z_i; \beta) - g(Z_i; \beta_0), \quad \psi_{ij} = \psi(X_i - X_j), \quad K_{h,ij} = K_h(W_i - W_j). \tag{4.2}$$

Under $\mathcal{H}_0$

$$
\begin{aligned}
W_n(\beta) &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau] [\mathbb{I}\{Y_j \leq g(Z_j; \beta)\} - \tau] K_{h,ij}\psi_{ij} \\
&= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(0 \mid Z_i)] \\
&\qquad\qquad \times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0)\} - F_\varepsilon(0 \mid Z_j)] K_{h,ij}\psi_{ij}.
\end{aligned}
$$

By a Taylor expansion, decompose

$$F_\varepsilon(0 \mid Z_i) = F_\varepsilon(G_i(\beta, \beta_0) \mid Z_i) - f_\varepsilon(0 \mid Z_i)\dot{g}'(Z_i; \beta_0)(\beta - \beta_0) + O_{\mathbb{P}}\left(\|\beta - \beta_0\|^2\right).$$

We can write $W_n(\beta) - W_n(\beta_0) = \{W_{1n}^0(\beta) - W_{1n}^0(\beta_0)\} + 2W_{2n}^0(\beta) + W_{3n}^0(\beta) + R_n^0$ where

$$W_{1n}^0(\beta) = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(G_i(\beta, \beta_0) \mid Z_i)]$$
$$\times [\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0)\} - F_\varepsilon(G_j(\beta, \beta_0) \mid Z_j)] K_{h,ij} \psi_{ij}$$

$W_{2n}^0(\beta) = (\beta - \beta_0)' \widetilde{W}_{2n}^0(\beta)$ with

$$\widetilde{W}_{2n}^0(\beta) = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(G_i(\beta, \beta_0) \mid Z_i)]$$
$$\times f_\varepsilon(0 \mid Z_j) \dot{g}(Z_j; \beta_0) K_{h,ij} \psi_{ij},$$

$W_{3n}^0(\beta) = (\beta - \beta_0)' \widetilde{W}_{3n}^0 (\beta - \beta_0)$ with

$$\widetilde{W}_{3n}^0 = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon(0 \mid Z_i) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_j; \beta_0) f_\varepsilon(0 \mid Z_j) K_{h,ij} \psi_{ij} = O_\mathbb{P}(1).$$

The rate of $\widetilde{W}_{3n}^0$ follows simply by computing its mean and variance. By Assumption 2.1(c) and Assumption 2.2(c) it is easy to check that $|R_n^0| \leq \|\beta - \beta_0\|^2 O_\mathbb{P}(1)$. For deriving the order of $\widetilde{W}_{2n}^0$, apply Hoeffding decomposition and write $h\widetilde{W}_{2n}^0(\beta) = V_n^2(\beta) + V_n^1(\beta)$ with V_n^1, V_n^2 degenerate $U-$processes or order 1 and 2, respectively. In view of Assumptions 2.2(d) and 2.3(a), apply Corollary 4 of Sherman (1994) and deduce that $V_n^2(\beta) = O_\mathbb{P}(n^{-1})$ uniformly in β (and h). Next, if $\dot{g}^{(l)}$ denotes the lth component of the vector of first-order derivatives $\dot{g}$, $1 \leq l \leq p$, and

$$\pi^{(l)}(Z_i) = \mathbb{E}\left[f_\varepsilon(0 \mid Z_j) \dot{g}^{(l)}(Z_j; \beta_0) h^{-3/4} K_{h,ij} \psi_{ij} \mid Z_i\right]$$

we can rewrite the lth component of the vector $V_n^1(\beta)$ as

$$\frac{h^{3/4}}{n} \sum_{i=1}^n [\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0)\} - F_\varepsilon(G_i(\beta, \beta_0) \mid Z_i)] \pi^{(l)}(Z_i).$$

By Hölder inequality, Assumption 2.1(c), Assumption 2.2(c) and a change of variables,

$$\left|\pi^{(l)}(X_i)\right| \leq \mathbb{E}\left[f_\varepsilon(0 \mid Z_j) \left|\dot{g}^{(l)}(Z_j; \beta_0)\right| h^{-3/4} K_{h,ij} \left|\psi_{ij}\right| \mid Z_i\right]$$
$$\leq C_1 \mathbb{E}^{1/4}\left[A^4(Z_j)\right] \mathbb{E}^{3/4}\left[h^{-1} K_{h,ij}^{4/3} \mid Z_i\right]$$
$$\leq C_2,$$

for any $1 \leq l \leq p$. Now, by Corollary 4 of Sherman (1994), $h^{-3/4} V_n^1(\beta) = O_\mathbb{P}(n^{-1/2})$ uniformly in β. Deduce that

$$\sup_\beta |W_{2n}^0(\beta)| \leq \|\beta - \beta_0\| O_\mathbb{P}\left(h^{-1}n^{-1} + h^{-1/4}n^{-1/2}\right).$$

14

Finally, by Lemma 1 of Zheng (1998), for any $\alpha \in (0, 1)$

$$\sup_{\beta} |W_{1n}^0(\beta) - W_{1n}^0(\beta_0)| = O_{\mathbb{P}}\left(h^{-1}n^{-1-\alpha/4}\right)$$

uniformly over $O_{\mathbb{P}}\left(n^{-1/2}\right)$ neighborhoods of β_0. Gathering the results and using Lemma 4.1 with $\delta(\cdot) \equiv 0$ we obtain (4.1). Now, it remains to check that $nh^{1/2}W_n(\beta_0)/v_n$ converges in law to a standard normal distribution. This result easily follows as a particular case of Lemma 4.1 below. ∎

4.2 Proof of Theorem 2.3

First, we state the behavior of $\widehat{\beta}$, the estimator of β_0 under the sequence of local alternatives H_{1n}. The proof is provided in the Supplementary Materials.

Lemma 4.1. *Suppose that Assumptions 2.1, 2.2 hold, let $\delta(\cdot)$ be a function such that Condition (2.6) holds, and let r_n, $n \geq 1$ be a sequence of real numbers such that $r_n \to 0$. If $\widehat{\beta} = \arg\min_{\beta \in B} \Gamma_n(\beta)$ with $\Gamma_n(\beta) = \sum_{i=1}^n \rho_\tau(Y_i - g(Z_i; \beta))$, then under $\mathcal{H}_0$, $\widehat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$ and under H_{1n} defined in (2.5), $\widehat{\beta} - \beta_n = O_{\mathbb{P}}(n^{-1/2})$ where*

$$\beta_n = \beta_0 - r_n^2 \left[\mathbb{E}\left[f_\varepsilon(0 \mid Z)\dot{g}(Z; \beta_0)\dot{g}'(Z; \beta_0)\right]\right]^{-1} \mathbb{E}\left[f_\varepsilon'(0 \mid Z)\delta^2(Z)\dot{g}(Z; \beta_0)\right].$$

Lemma 4.1 shows in particular that under $\mathcal{H}_{1n}$, $\widehat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2} + r_n^2)$. To our best knowledge, this result on the behavior of $\widehat{\beta}$ under the local alternatives is new. He and Zhu (2003) only considered the case $r_n = n^{-1/2}$ while Zheng (1998) assumed $\widehat{\beta} - \beta^* = O_{\mathbb{P}}(n^{-1/2})$ under $\mathcal{H}_{1n}$, for some fixed β^*. Our Lemma 4.1 indicates that such $\sqrt{n}-$convergence assumptions on the local alternatives may be too restrictive. Below, we improve the point (C) in the Theorem of Zheng (1998) also because we can take into account the rates of convergence of $\widehat{\beta}$ under the alternatives slower than $O_{\mathbb{P}}(n^{-1/2})$.

In the case of a fixed deviation from the null hypothesis, that is $r_n \equiv 1$, the tools used for proving Theorem 2.3 could be easily adapted to show the $\sqrt{n}-$convergence of $\widehat{\beta}$ to β^* that minimizes the map $\beta \mapsto \mathbb{E}[\rho_\tau(Y - g(Z, \beta))] = \mathbb{E}[\rho_\tau(g(Z, \beta_0) + \delta(Z) + \varepsilon - g(Z, \beta))]$. The consistency of the test is then a consequence of the fact that $nh^{1/2}I_n(\beta^*)$ tends to infinity.

Let $\delta_i = \delta(Z_i)$ and let $G_i(\beta, \beta_0)$ and $K_{h,ij}$ be defined as in equation (4.2). Under $\mathcal{H}_{1n}$

$$
\begin{aligned}
W_n(\beta) &= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} \left[\mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau\right] \left[\mathbb{I}\{Y_j \leq g(Z_j; \beta)\} - \tau\right] K_{h,ij}\psi_{ij} \\
&= \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} \left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n\delta_i\} - F_\varepsilon(0 \mid Z_i)\right] \\
&\qquad\qquad \times \left[\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0) - r_n\delta_j\} - F_\varepsilon(0 \mid Z_j)\right] K_{h,ij}\psi_{ij}.
\end{aligned}
$$

Let us decompose

$$
\begin{aligned}
F_\varepsilon \left(0 \mid Z_i \right) = {} & F_\varepsilon \left(G_i(\beta, \beta_0) - r_n \delta_i \mid Z_i \right) - f_\varepsilon \left(0 \mid Z_i \right) \left\{ \dot{g}'(Z_i; \beta_0) \left(\beta - \beta_0 \right) - r_n \delta_i \right\} \\
& - 2^{-1} r_n^2 f_\varepsilon' \left(0 \mid Z_i \right) \delta_i^2 + O_\mathbb{P} \left(\| \beta - \beta_0 \|^2 + r_n \| \beta - \beta_0 \| \right) + o_\mathbb{P} \left(r_n^2 \right).
\end{aligned}
$$

We can write

$$
W_n(\beta) = W_{1n}(\beta) + 2[W_{2n}(\beta) + W_{3n}(\beta) + W_{4n}(\beta)] + W_{5n}(\beta) + 2W_{6n}(\beta) + W_{7n} + R_n
$$

where

$$
\begin{aligned}
W_{1n}(\beta) = {} & \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} \left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon \left(G_i(\beta, \beta_0) - r_n \delta_i \mid Z_i \right) \right] \\
& \times \left[\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0) - r_n \delta_j\} - F_\varepsilon \left(G_j(\beta, \beta_0) - r_n \delta_j \mid Z_j \right) \right] K_{h,ij} \psi_{ij}
\end{aligned}
$$

$W_{2n}(\beta) = (\beta - \beta_0)' \widetilde{W}_{2n}(\beta)$ with

$$
\begin{aligned}
\widetilde{W}_{2n}(\beta) = {} & \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} \left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon \left(G_i(\beta, \beta_0) - r_n \delta_i \mid Z_i \right) \right] \\
& \times f_\varepsilon \left(0 \mid Z_j \right) \dot{g}(Z_j; \beta_0) K_{h,ij} \psi_{ij},
\end{aligned}
$$

$$
\begin{aligned}
W_{3n}(\beta) = {} & \frac{r_n h^{-1}}{n(n-1)} \sum_{j \neq i} \left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon \left(G_i(\beta, \beta_0) - r_n \delta_i \mid Z_i \right) \right] \\
& \times f_\varepsilon \left(0 \mid Z_j \right) \delta_j K_{h,ij} \psi_{ij},
\end{aligned}
$$

$$
\begin{aligned}
W_{4n}(\beta) = {} & \frac{r_n^2 h^{-1}}{2n(n-1)} \sum_{j \neq i} \left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n \delta_i\} - F_\varepsilon \left(G_i(\beta, \beta_0) - r_n \delta_i \mid Z_i \right) \right] \\
& \times f_\varepsilon' \left(0 \mid Z_j \right) \delta_j^2 K_{h,ij} \psi_{ij},
\end{aligned}
$$

$W_{5n}(\beta) = (\beta - \beta_0)' \widetilde{W}_{5n} (\beta - \beta_0)$ with

$$
\widetilde{W}_{5n} = \frac{h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon \left(0 \mid Z_i \right) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_j; \beta_0) f_\varepsilon \left(0 \mid Z_j \right) K_{h,ij} \psi_{ij} = O_\mathbb{P}(1),
$$

$W_{6n}(\beta) = (\beta - \beta_0)' \widetilde{W}_{6n}$ with

$$
\widetilde{W}_{6n} = \frac{r_n h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon \left(0 \mid Z_i \right) \delta_i f_\varepsilon \left(0 \mid Z_j \right) \dot{g}(X_j; \beta_0) K_{h,ij} \psi_{ij} = O_\mathbb{P}(r_n),
$$

$$
W_{7n} = \frac{r_n^2 h^{-1}}{n(n-1)} \sum_{j \neq i} f_\varepsilon \left(0 \mid Z_i \right) \delta(X_i) f_\varepsilon \left(0 \mid Z_j \right) \delta(Z_j) K_{h,ij} \psi_{ij} = C_1 r_n^2 + o_\mathbb{P}(r_n^2)
$$

with $C_1 > 0$ and R_n a reminder term that is negligible because of the properties of f'_ε and $\dot{g}$. Note that the $U-$statistics $\widetilde{W}_{5n}$, $\widetilde{W}_{6n}$ and W_{7n} depend only on the X_i. Their orders are obtained from elementary calculations of mean and variance.

Next, we can write $W_{1n}(\beta) = \{W_{1n}(\beta) - W_{1n}(\beta_0)\} + W_{1n}(\beta_0)$. As $W_{1n}(\beta_0)$ is centered, its order in probability is given by the variance. We have

$$
\begin{aligned}
\mathrm{Var}(W_{1n}(\beta_0) \mid Z_1, ..., Z_n) &= \frac{1}{n^2(n-1)^2} \sum_{i \neq j} F_\varepsilon\left(-r_n\delta_i \mid Z_i\right)\left[1 - F_\varepsilon\left(-r_n\delta_i \mid Z_i\right)\right] \\
&\qquad \times F_\varepsilon\left(-r_n\delta_j \mid Z_j\right)\left[1 - F_\varepsilon\left(-r_n\delta_j \mid Z_j\right)\right]h^{-2}K^2_{h,ij}\psi_{ij}\left(\mu\right) \\
&\leq \frac{h^{-1}}{16n(n-1)}\left[\frac{1}{n(n-1)}\sum_{i \neq j} h^{-1}K^2_{h,ij}\psi_{ij}\right]
\end{aligned}
$$

The expectation of the last $U-$statistic in the display converges to a constant while the variance tends to zero. As $W_{1n}(\beta_0)$ is of zero conditional mean given the Z_i, deduce that the variance of $W_{1n}(\beta_0)$ is bounded by $Cn^{-2}h^{-1}$. By Chebyshev's inequality, $W_{1n}(\beta_0) = o_\mathbb{P}\left(r_n^2\right)$, provided that $r_n^2 nh^{1/2} \to \infty$. Next, let

$$
\begin{aligned}
H_{1n}(Z_i, Z_j, \beta) &= \left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n\delta_i\} - F_\varepsilon\left(G_i(\beta, \beta_0) - r_n\delta_i \mid Z_i\right)\right] \\
&\quad \times \left[\mathbb{I}\{\varepsilon_j \leq G_j(\beta, \beta_0) - r_n\delta_j\} - F_\varepsilon\left(G_j(\beta, \beta_0) - r_n\delta_j \mid Z_j\right)\right]K_{h,ij}\psi_{ij}, \qquad \beta \in B.
\end{aligned}
$$

By the arguments used for Lemma 4.1 above, the class of functions $\{H_{1n}(\cdot, \cdot, \beta) : \beta \in B\}$ is *Euclidean(c,d)* for an envelope with a finite fourth moment, with c and d independent of n. Now, we can use equation (A.11) of Zheng (1998) and his Lemma 1 with the condition (ii) replaced by $\mathbb{E}[H_{1n}(\cdot, \beta) - H_{1n}(\cdot, \beta_0)]^2 \leq \Lambda \|\beta - \beta_0\|$. By a close inspection of the proof of Zheng's Lemma 1, see his equations (A.2) to (A.5), it is obvious to adapt his conclusion and to deduce that in our setup for any $0 < \alpha < 1$

$$
W_{1n}(\beta) - W_{1n}(\beta_0) = O_\mathbb{P}\left(n^{-1}h^{-1}\|\beta - \beta_0\|^{\alpha/2}\right) = O_\mathbb{P}\left(n^{-1}h^{-1}\left\{r_n + n^{-1/4}\right\}^\alpha\right)
$$

uniformly over $O_\mathbb{P}(r_n^2 + n^{-1/2})$ neighborhoods of β_0. Thus, when $n^{1/2}r_n^2 \to \infty$, we have

$$
W_{1n}(\widehat{\beta}) - W_{1n}(\beta_0) = O_\mathbb{P}\left(n^{-1}h^{-1}r_n^\alpha\right) = O_\mathbb{P}\left(n^{-1/2}\right) = o_\mathbb{P}\left(r_n^2\right),
$$

whereas in the case where $n^{1/2}r_n^2$ is bounded, use $nh^{1/2}r_n^2 \to \infty$ and take α sufficiently close to one to obtain

$$
W_{1n}(\widehat{\beta}) - W_{1n}(\beta_0) = O_\mathbb{P}\left(n^{-1-\alpha/4}h^{-1}\right) = o_\mathbb{P}\left(r_n^2\right).
$$

The remaining terms W_{2n}, W_{3n} and W_{4n} can be treated in the following way. By Hoeffding's decomposition

$$
r_n^{-1}hW_{3n}(\beta) = U_n^2(\beta) + U_n^1(\beta)
$$

17

with U_n^1, U_n^2 degenerate $U-$processes or order 1 and 2, respectively. In view of Assumption 2.2(d) and the fact that $K\left(\cdot\right)$ is bounded, apply Corollary 4 of Sherman (1994) to deduce that $U_n^2(\beta) = O_\mathbb{P}\left(n^{-1}\right)$ uniformly in β. If $K_{h,ij}\left(\theta\right) = K_h((X_i - X_j)'\theta)$ and

$$\xi\left(Z_i\right) = \mathbb{E}\left[\mathbb{E}\left\{f_\varepsilon\left(0 \mid Z_j\right)\delta\left(Z_j\right) \mid Z_j'\theta\right\}h^{-3/4}K_{h,ij}\psi_{ij} \mid Z_i\right]$$

we can write

$$U_n^1(\beta) = \frac{h^{3/4}}{n}\sum_i\left[\mathbb{I}\{\varepsilon_i \leq G_i(\beta, \beta_0) - r_n\delta_i\} - F_\varepsilon\left(G_i(\beta, \beta_0) - r_n\delta_i \mid Z_i\right)\right]\xi\left(Z_i\right).$$

By Hölder inequality, Assumption 2.1(c) and a change of variables,

$$\left|\xi\left(Z_i\right)\right| \leq \mathbb{E}^{1/4}\left[\delta^4(Z_j)\right]\mathbb{E}^{3/4}\left[h^{-1}K_{h,ij}^{4/3} \mid Z_i\right] \leq C,$$

for some $C > 0$. Now, by Corollary 4 of Sherman (1994), $h^{-3/4}U_n^1(\beta) = O_\mathbb{P}\left(n^{-1/2}\right)$ uniformly in β. As $nh^{1/2}r_n^2 \to \infty$, deduce that

$$\sup_\beta |W_{3n}(\beta)| = O_\mathbb{P}\left(r_nh^{-1}n^{-1} + r_nh^{-1/4}n^{-1/2}\right) = o_\mathbb{P}(r_n^2).$$

By similar arguments, $\sup_\beta |W_{4n}(\beta)| = o_\mathbb{P}(r_n^2)$ (here apply Hölder inequality with $p = q = 2$) and W_{3n}, $\sup_\beta |\widetilde{W}_{2n}(\beta)| = O_\mathbb{P}\left(h^{-1}n^{-1} + h^{-1/4}n^{-1/2}\right)$, and thus

$$\sup_\beta |W_{2n}(\beta)| = O_\mathbb{P}(r_n^2 + n^{-1/2})O_\mathbb{P}\left(h^{-1}n^{-1} + h^{-1/4}n^{-1/2}\right) = o_\mathbb{P}(r_n^2).$$

Collecting results, under $\mathcal{H}_{1n}$, $T_n \geq Cnh^{1/2}r_n^2\{1 + o_\mathbb{P}(1)\}$ or some constants $C > 0$. Now, the proof is complete.

4.3   Proof of Theorem 2.4

Let $W_n^*(\beta)$ be the statistic obtained after replacing $U_i\left(\beta\right)$ with $U_i^*\left(\beta\right) = \mathbb{I}\{Y_i^* \leq g(Z_i; \beta)\} - \tau$ in the formula of $W_n(\beta)$. The proof of the bootstrap procedure consistency follows the steps of the proof of Theorem 2.2, but requires several specific ingredients: (a) the convergence in law of $nh^{1/2}W_n^*(\widehat{\beta})/v_n$ *conditionally* upon the original sample; and (b) the $O_\mathbb{P}\left(n^{-1/2}\right)$ rate for $\widehat{\beta}^* - \widehat{\beta}$, and the negligibility of $W_n^*(\widehat{\beta}^*) - W_n^*(\widehat{\beta})$ given the original sample. If S_{1n}^* and S_{2n}^* denote bootstrapped statistics, S_{1n}^* is bounded in probability given the sample if

$$\lim_{M\to\infty}\mathbb{P}[|S_{1n}^*| > M \mid Y_1, Z_1, \cdots, Y_n, Z_n] = o_p(1).$$

while S_{2n}^* is asymptotically negligible given the sample if

$$\forall\epsilon > 0, \qquad \mathbb{P}[|S_{2n}^*| > \epsilon \mid Y_1, Z_1, \cdots, Y_n, Z_n] = o_p(1).$$

The asymptotic normality of $nh^{1/2}W_n^*(\widehat{\beta})/v_n$ given the sample is obtained below from a martingale central limit theorem as stated in Hall and Heyde (1980).

Lemma 4.1. *Under the assumptions of Theorem 2.4,*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(nh^{1/2}W_n^*(\widehat{\beta})/v_n \le t \mid Y_1, Z_1, ..., Y_n, Z_n\right) - \Phi(t)\right| \to 0, \qquad in\ probability.$$

Proof. The proof is based on the Central limit Theorem (CLT) for martingale arrays, see Corollary 3.1 of Hall and Heyde (1980). Recall that $U_i^*(\widehat{\beta}) = \mathbb{I}\{Y_i^* \le g(Z_i; \widehat{\beta}^*)\} - \tau$. Define the martingale array $\left\{S_{n,m}^*, \mathcal{F}_{n,m}^*, 1 \le m \le n, n \ge 1\right\}$ where $S_{n,1}^* = 0$ and $S_{n,m}^* = \sum_{i=2}^{m} G_{n,i}^*$ with

$$G_{n,i}^* = \frac{2h^{-1/2}}{n-1} U_i^*(\widehat{\beta}) \sum_{j=1}^{i-1} U_j^*(\widehat{\beta}) K_{h,ij}\psi_{ij},$$

and $\mathcal{F}_{n,m}^*$ is the σ-field generated by $\{\overline{Z}, \eta_1, \ldots, \eta_m\}$ where $\overline{Z} = \{Y_1, \ldots, Y_n, Z_1, \ldots, Z_n\}$. Thus $nh^{1/2}W_n^*(\widehat{\beta}) = S_{n,n}^*$. Next define

$$
\begin{aligned}
V_n^{2*} &= \sum_{i=2}^{n} \mathbb{E}\left[G_{n,i}^{2*} \mid \mathcal{F}_{n,i-1}^*\right] \\
&= \frac{4h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} U_j^*(\widehat{\beta})U_k^*(\widehat{\beta})K_{h,ij}K_{h,ik}\psi_{ij}\psi_{ik} \\
&= \frac{4h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=2}^{n}\sum_{j=1}^{i-1} U_j^{*2}(\widehat{\beta})K_{h,ij}^2\psi_{ij}^2 \\
&\quad + \frac{8h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=3}^{n}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1} U_j^*(\widehat{\beta})U_k^*(\widehat{\beta})K_{h,ij}K_{h,ik}\psi_{ij}\psi_{ik} \\
&= A_n^* + B_n^*.
\end{aligned}
$$

Recall that

$$v_n^2 = \frac{2h^{-1}\tau^2(1-\tau)^2}{n(n-1)} \sum_{j\ne i} K_{h,ij}^2\psi_{ij}^2$$

and by standard calculations of the means and variance it could be shown to tend to a positive constant. Next, note that

$$\mathbb{E}\left[A_n^* \mid \overline{Z}\right] = \frac{4h^{-1}\tau(1-\tau)}{(n-1)^2} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \mathbb{E}\left[U_j^{*2}(\widehat{\beta}) \mid \overline{Z}\right] K_{h,ij}^2\psi_{ij}^2 = \frac{n}{n-1}\,v_n^2.$$

19

Moreover,

$$
\begin{aligned}
\mathbb{E}\left[\operatorname{Var}\left(A_n^* \mid \overline{Z}\right)\right] &= \frac{16\tau^2(1-\tau)^2}{h^2(n-1)^4} \\
&\quad \times \sum_{i=2}^{n}\sum_{i'=2}^{n}\sum_{j=1}^{i\wedge i'-1} \mathbb{E}\left[\mathbb{E}\left[U_j^{*4}(\widehat{\beta}) - \tau^2(1-\tau)^2|\overline{Z}\right]K_{h,ij}^2 K_{h,i'j}^2 \psi_{ij}^2 \psi_{i'j}^2\right] \\
&= \frac{16\tau^4(1-\tau)^4\{\tau(1-\tau)(1-3\tau(1-\tau))-1\}}{h^2(n-1)^4} \\
&\quad \times \sum_{i=2}^{n}\sum_{i'=2}^{n}\sum_{j=1}^{i\wedge i'-1} \mathbb{E}\left[K_{h,ij}^2 K_{h,i'j}^2 \psi_{ij}^2 \psi_{i'j}^2\right] \\
&= \frac{32\tau^4(1-\tau)^4(\tau(1-\tau)(1-3\tau(1-\tau))-1)}{h^2(n-1)^4} \\
&\quad \times \sum_{i=3}^{n}\sum_{i'=2}^{i-1}\sum_{j=1}^{i'-1} \mathbb{E}\left[K_{h,ij}^2 K_{h,i'j}^2 \psi_{ij}^2 \psi_{i'j}^2\right] \\
&\quad + \frac{16\tau^4(1-\tau)^4(\tau(1-\tau)(1-3\tau(1-\tau))-1)}{h^2(n-1)^4} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \mathbb{E}\left[K_{h,ij}^4 \psi_{ij}^4\right] \\
&= O(n^{-1}) + O(n^{-2}h^{-1})
\end{aligned}
$$

because ψ_{ij}, $\mathbb{E}\left[h^{-1}K_{h,ij}^4\right]$ and $\mathbb{E}\left[h^{-2}K_{h,ij}^2 K_{h,i'j}^2\right]$ are bounded for all pairwise distinct indexes i, i' and j. Deduce that $A_n^*/v_n^2 \to 1$ in probability. On the other hand,

$$
\mathbb{E}\left[B_n^{*2}\right] = \frac{8\tau^4(1-\tau)^4}{h^2(n-1)^4} \sum_{i=3}^{n}\sum_{j=2}^{i-1}\sum_{k=1}^{j-1} \mathbb{E}\left[K_{h,ij}^2 K_{h,ik}^2 \psi_{ij}^2 \psi_{ik}^2\right] = O(n^{-1})
$$

so that $V_n^{2*}/v_n^2 \to 1$ in probability. To use the CLT it remains to check the Lindeberg condition. For any $\epsilon > 0$,

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=2}^{n}\mathbb{E}\left[G_{n,i}^{*2}\mathbb{I}(G_{n,i}^{*2} > \epsilon) \mid \mathcal{F}_{n,i-1}^*\right]\right] &\leq \epsilon^{-4}\mathbb{E}\left[\sum_{i=2}^{n}\mathbb{E}\left[G_{n,i}^{*4} \mid \mathcal{F}_{n,i-1}^*\right]\right] \\
&\leq \frac{16\tau^3(1-\tau)^3\{1-3\tau(1-\tau)\}}{\epsilon^4 h^2(n-1)^4} \sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{i-1} \mathbb{E}\left[K_{h,ij}^2 K_{h,ik}^2 \psi_{ij}^2 \psi_{ik}^2\right] \\
&\leq \frac{32\tau^3(1-\tau)^3\{1-3\tau(1-\tau)\}}{\epsilon^4 h^2(n-1)^4} \sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1} \mathbb{E}\left[K_{h,ij}^2 K_{h,ik}^2 \psi_{ij}^2 \psi_{ik}^2\right] \\
&\quad + \frac{16\tau^3(1-\tau)^3\{1-3\tau(1-\tau)\}}{\epsilon^4 h^2(n-1)^4} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \mathbb{E}\left[K_{h,ij}^4 \psi_{ij}^4\right] \\
&= O(n^{-1}) + O(n^{-2}h).
\end{aligned}
$$

Eventually, applying the CLT for martingale arrays along the subsequences of V_n^{2*} that converge almost surely to the limit of v_n^2 and subsequences for which the Lindeberg condition is satisfied almost surely, the result follows. ∎

To obtain the $O_{\mathbb{P}}\left(n^{-1/2}\right)$ rate for $\widehat{\beta}^* - \widehat{\beta}$, and the negligibility of $W_n^*(\widehat{\beta}^*) - W_n^*(\widehat{\beta})$ given the original sample, we use a conditional version of the moment inequality for $U-$processes proved by Sherman (1994). Before stating this new result that has its own interest let us introduce some more notation: for k a positive integer let $(n)_k = n(n-1)...(n-k+1)$ and let $\mathbf{i_k^n} = (i_1, ..., i_k)$ be a $k-$tuple of distinct integers from the set $\{1, ..., n\}$. Similarly, $\mathbf{i_k^{2n}} = (i_1, ..., i_k)$ denotes a $k-$tuples of distinct integers from $\{1, ..., 2n\}$. Moreover, a function g on $\mathcal{S}^k$ is called degenerate if for each $i = 1, ..., k$, and all $s_1, ..., s_{i-1}, s_{i+1}, ..., s_k \in \mathcal{S}$, $\mathbb{E}[g(s_1, ..., s_{i-1}, S, s_{i+1}, ..., s_k)] = 0$.

We state the following Lemma. The proof is provided in the Supplementary Materials.

Lemma 4.2. *Let k be a positive integer and $\mathcal{G}$ a degenerate class of real-valued functions on $\mathbb{R}^{1+q} \times ... \times \mathbb{R}^{1+q}$. Suppose $\mathcal{G}$ is* Euclidean(c,d) *for a squared integrable envelope and some $c, d > 0$. Fix $z_1, ..., z_n \in \mathbb{R}^q$ and let $u_1, ..., u_n, u_{n+1}, ..., u_{2n}$ be independent copies of the random variable u. For $i = 1, ..., n$, let $v_i = (u_i, z_i)$ and $v_{n+i} = (u_{n+i}, z_i)$. Define $g_{\mathbf{i_k^n}}(u_{i_1}, ..., u_{i_k}) = g(v_{i_1}, ..., v_{i_k})$ and define $g_{\mathbf{i_k^{2n}}}$ similarly. Suppose that for any $k-$tuple $\mathbf{i_k^n}$, the function $g_{\mathbf{i_k^n}}$ is degenerate as a function of u_i variables (necessarily the same property holds also for any $k-$tuple $\mathbf{i_k^{2n}}$). Let*

$$U_{n,z_1,...,z_n}^k(g) = (n)_k^{-1} \sum_{\mathbf{i_k^n}} g_{\mathbf{i_k^n}}(u_{i_1}, ..., u_{i_k}), \quad U_{2n,z_1,...,z_n}^k(g) = (2n)_k^{-1} \sum_{\mathbf{i_k^{2n}}} g_{\mathbf{i_k^{2n}}}(u_{i_1}, ..., u_{i_k}).$$

Then for any $\alpha \in (0,1)$, there exists a constant Λ depending only on α and k (and independent of n and the sequence $z_1, ..., z_n$) such that

$$\mathbb{E}\left[\sup_{\mathcal{G}} |n^{k/2} U_{n,z_1,...,z_n}^k(g)|\right] \leq \Lambda \mathbb{E}^{1/2}\left[\sup_{\mathcal{G}}\{U_{2n,z_1,...,z_n}^k(g^2)\}^{\alpha}\right].$$

To establish the rate of $\widehat{\beta}^* - \widehat{\beta}$ given the sample, it suffices to consider a simplified version of our Lemma 4.1. By Lemma 4.2, $\sup_\beta \left|n^{-1}\Gamma_n^*(\beta) - \mathbb{E}\left[\rho_\tau(Y - g(Z;\beta)) \mid \overline{Z}\right]\right|$ is asymptotically negligible given the sample $\overline{Z} = \{Y_1, ..., Y_n, Z_1, ..., Z_n\}$. Reconsidering the arguments for the consistency of argmax estimators along almost surely convergent subsequences depending on $\overline{Z}$, deduce that $\widehat{\beta}^* - \widehat{\beta}$ is a asymptotically negligible given the sample $\overline{Z}$. Next, define the empirical process

$$\nu_n^*(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{\psi_\tau\left(Y_i^* - g(Z_i;\beta)\right) - \mathbb{E}[\psi_\tau\left(Y_i - g(Z_i;\beta)\right) \mid \overline{Z}]\right\} \dot{g}(Z_i;\beta)$$

21

indexed by β. Lemma 4.2 guarantees that $\sup_\beta |\nu_n^*(\beta)|$, and in particular $\nu_n^*(\widehat{\beta}^*) - \nu_n^*(\widehat{\beta})$, are bounded in probability given the sample. Proceeding like in equation (4.6) of the Supplementary Materials, that is using the directional derivative of $\Gamma_n^*(\beta)$ at $\widehat{\beta}^*$ along any direction γ, deduce

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau \left(Y_i^* - g(Z_i; \widehat{\beta}^*) \right) \dot{g}(Z_i; \widehat{\beta}^*)$$

is bounded in probability given the sample (conditional negligibility could be also derived but boundedness given the sample suffices for the present purpose). Since for all i,

$$\mathbb{E} \left[\psi_\tau \left(Y_i^* - g(Z_i; \widehat{\beta}^*) \right) \mid \overline{Z} \right] = F_{\varepsilon^*} \left(g(Z_i; \widehat{\beta}^*) - g(Z_i; \widehat{\beta}) \mid \overline{Z} \right) - \tau,$$

and for any sample $\overline{Z}$, the distribution function $F_{\varepsilon^*}(\cdot \mid \overline{Z})$ is that of the uniform law on $[-\tau, 1-\tau]$, the boundedness of $\sqrt{n}(\widehat{\beta}^* - \widehat{\beta})$ follows by a Taylor expansion of $F_{\varepsilon^*}(\cdot \mid \overline{Z})$ around the origin, exactly like in the proof of Lemma 4.1 in the case $r_n = 0$. The case of the wild bootstrap and linear quantile regression follows as a consequence of Theorem 1 of Feng et al. (2011). The arguments of Theorem 1 of Feng et al. (2011) could be adapted to nonlinear models using a linearization like in the proof of Lemma 4.1. The details are omitted.

Finally, using Lemma 4.2, derive conditional versions of Lemma 1 of Zheng (1998) and of Corollary 4 of Sherman in the case of constant envelopes. Combine these results with the fact that $\sqrt{n}(\widehat{\beta}^* - \widehat{\beta})$ is bounded in probability given the sample and follow the lines of the proof of Theorem 2.2 above to deduce that for any $\varepsilon > 0$

$$\mathbb{P} \left(nh^{1/2} \left| W_n^*(\widehat{\beta}^*) - W_n^*(\widehat{\beta}) \right| > \varepsilon \mid Y_1, Z_1, ..., Y_n, Z_n \right) \to 0, \quad \text{in probability.}$$

References

ABREVAYA, J. (2001): "The effects of demographics and maternal behavior on the distribution of birth outcomes," *Empirical Economics*, 26, 247–257.

BIERENS, H. J. AND D. K. GINTHER (2002): "Integrated Conditional Moment testing of quantile regression models," in *Economic Applications of Quantile Regression*, ed. by B. Fitzenberger, R. Koenker, and J. A. Machado, Physica-Verlag HD, Stud. Empir. Econom., 307–324.

BIERENS, H. J. AND W. PLOBERGER (1997): "Asymptotic theory of integrated conditional moment tests," *Econometrica*, 65, 1129–1151.

DE LA PEÑA, V. H. AND E. GINÉ (1999): *Decoupling*, Probab. Appl. (N. Y.), Springer-Verlag, New York, from dependence to independence, Randomly stopped processes. *U*-statistics and processes. Martingales and beyond.

FENG, X., X. HE, AND J. HU (2011): "Wild bootstrap for quantile regression," *Biometrika*, 98, 995–999.

GUERRE, E. AND P. LAVERGNE (2002): "Optimal Minimax Rates for Nonparametric Specification Testing in Regression Models," *Econometric Theory*, 18, 1139–1171.

——— (2005): "Data-Driven Rate-Optimal Specification Testing in Regression Models," *Ann. Statist.*, 33, pp. 840–870.

HALL, P. AND C. C. HEYDE (1980): *Martingale limit theory and its application*, New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers], probab. Math. Statist.

HÄRDLE, W. AND J. S. MARRON (1985): "Optimal bandwidth selection in nonparametric regression function estimation," *Ann. Statist.*, 13, 1465–1481.

HE, X. AND L.-X. ZHU (2003): "A lack-of-fit test for quantile regression," *J. Amer. Statist. Assoc.*, 98, 1013–1022.

HOROWITZ, J. L. AND V. G. SPOKOINY (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model against a Nonparametric Alternative," *Econometrica*, 69, 599–631.

——— (2002): "An adaptive, rate-optimal test of linearity for median regression models," *J. Amer. Statist. Assoc.*, 97, 822–835.

KOENKER, R. (2005): *Quantile regression*, vol. 38 of *Econom. Soc. Monogr.*, Cambridge University Press, Cambridge.

KOENKER, R. AND G. BASSETT, JR. (1978): "Regression quantiles," *Econometrica*, 46, 33–50.

KOENKER, R. AND K. F. HALLOCK (2001): "Quantile Regression," *J. Econ. Perspect.*, 15, 143–156.

LAVERGNE, P., S. MAISTRE, AND V. PATILEA (2014): "A significance test for covariates in nonparametric regression." arXiv:1403.7063 [math.ST].

LAVERGNE, P. AND V. PATILEA (2012): "One for all and all for one: regression checks with many regressors," *J. Bus. Econom. Statist.*, 30, 41–52.

NOLAN, D. AND D. POLLARD (1987): "*U*-processes: rates of convergence," *Ann. Statist.*, 15, 780–799.

PAKES, A. AND D. POLLARD (1989): "Simulation and the asymptotics of optimization estimators," *Econometrica*, 57, 1027–1057.

SHERMAN, R. P. (1994): "Maximal inequalities for degenerate U-processes with applications to optimization estimators," *Ann. Statist.*, 22, 439–459.

STUTE, W. (1997): "Nonparametric model checks for regression," *Ann. Statist.*, 25, 613–641.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak convergence and empirical processes*, Springer Ser. Statist., New York: Springer-Verlag, with applications to statistics.

YU, K., Z. LU, AND J. STANDER (2003): "Quantile regression: applications and current research areas," *The Statistician*, 52, 331–350.

ZHENG, J. X. (1998): "A consistent nonparametric test of parametric regression models under conditional quantile restrictions," *Econometric Theory*, 14, 123–138.

Supplementary Material to "Powerful nonparametric checks for quantile regression"

Lemma 4.1. *Suppose that Assumptions 2.1, 2.2 hold, let $\delta(\cdot)$ be a function such that Condition (2.6) holds, and let r_n, $n \geq 1$ be a sequence of real numbers such that $r_n \to 0$. If $\widehat{\beta} = \arg\min_{\beta \in B} \Gamma_n(\beta)$ with $\Gamma_n(\beta) = \sum_{i=1}^n \rho_\tau(Y_i - g(Z_i; \beta))$, then under $\mathcal{H}_0$, $\widehat{\beta} - \beta_0 = O_\mathbb{P}(n^{-1/2})$ and under H_{1n} defined in (2.5), $\widehat{\beta} - \beta_n = O_\mathbb{P}(n^{-1/2})$ where*

$$\beta_n = \beta_0 - r_n^2 \left[\mathbb{E}\left[f_\varepsilon(0 \mid Z) \dot{g}(Z; \beta_0) \dot{g}'(Z; \beta_0) \right] \right]^{-1} \mathbb{E}\left[f_\varepsilon'(0 \mid Z) \delta^2(Z) \dot{g}(Z; \beta_0) \right].$$

Proof. It is easy to check that

$$|\rho_\tau(a - b) - \rho_\tau(a)| \leq |b| \max(\tau, 1 - \tau) \leq |b|. \tag{4.3}$$

Combine this with the Mean Value Theorem and Assumption 2.2(c) to check the conditions of Lemma 2.13 of Pakes and Pollard (1989) and to derive the Euclidean property for an integrable envelope for the family of functions $\{(y, z) \mapsto \rho_\tau(y - g(z; \beta)) : \beta \in B\}$.

Next, we study the consistency of $\widehat{\beta}$ under H_0. By the uniform law of large numbers, $\sup_\beta |n^{-1} \Gamma_n(\beta) - \mathbb{E}[\rho_\tau(Y - g(Z; \beta))]| \to 0$, in probability (use for instance Lemma 2.8 of Pakes and Pollard 1989). This uniform convergence, the identification condition in Assumption 2.2(a), the continuity of $g(z; \cdot)$ for any z, and usual arguments used for proving consistency of argmax estimators, allow to deduce $\widehat{\beta} - \beta_0 = o_\mathbb{P}(1)$. To obtain the consistency under the local alternatives approaching H_0, it suffices to prove $\sup_{\beta \in B} |\Delta_n(\beta)| \to 0$ in probability, where

$$\Delta_n(\beta) = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau(l(\varepsilon_i, Z_i; \beta) + r_n \delta(Z_i)) - \rho_\tau(l(\varepsilon_i, Z_i; \beta)) \right\}$$

and $l(u, z; \beta) = u + g(z; \beta_0) - g(z; \beta)$. By inequality (4.3),

$$|\Delta_n(\beta)| \leq \frac{|r_n|}{n} \sum_{i=1}^n |\delta(Z_i)|.$$

Consequently, $\Delta_n(\beta) = o_\mathbb{P}(1)$ uniformly over $\beta \in B$, and thus the consistency follows.

Define $\psi_\tau(e) = \tau - \mathbb{I}(e < 0)$ as the derivative of ρ_τ. To obtain the rate of convergence of $\widehat{\beta}$ under $\mathcal{H}_{1n}$ (in particular under H_0 by taking $r_n \equiv 0$) consider the empirical process

$$\nu_n(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \psi_\tau(Y_i - g(Z_i; \beta)) - \mathbb{E}[\psi_\tau(Y_i - g(Z_i; \beta)) \mid Z_i] \right\} \dot{g}(Z_i; \beta)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \psi_\tau(l(\varepsilon_i, Z_i; \beta) + r_n \delta(Z_i)) - \mathbb{E}[\psi_\tau(l(\varepsilon_i, Z_i; \beta) + r_n \delta(Z_i)) \mid Z_i] \right\} \dot{g}(Z_i; \beta)$$

indexed by β. First, let us notice that

$$\nu_n(\beta) - \nu_n(\beta_0) = o_{\mathbb{P}}(1) \tag{4.4}$$

uniformly over $o_{\mathbb{P}}(1)$ neighborhoods of β_0, as a consequence of Corollary 8 of Sherman (1994). Indeed, by Lemma 2.13 of Pakes and Pollard (1989), the class of functions $\{\dot{g}(\cdot; \beta) : \beta \in B\}$ is Euclidean for a squared integrable envelope. Next, by the VC-class property of the regression functions $\{g(\cdot; \beta), \beta \in B\}$, the class of functions $\{(u, z) \mapsto \psi_\tau(l(u, z; \beta) + r_n \delta(z)) : \beta \in B\}$ is *Euclidean(c,d)* for a constant envelope. See Lemma 2.12 of Pakes and Pollard (1989). Moreover, the constants c and d can be taken independent of n, see, for instance, the proof of Lemma 2.6.18(v) of van der Vaart and Wellner (1996). Finally, by repeated applications of the Mean Value Theorem and Assumptions 2.1(c) and 2.2(c), for any z, β_1, β_2 we have

$$
\begin{aligned}
&| \mathbb{E}\left[\psi_\tau(l(\varepsilon, z; \beta_1) + r_n \delta(z))\right] - \mathbb{E}\left[\psi_\tau(l(\varepsilon, z; \beta_2) + r_n \delta(z))\right] | \\
&\leq \quad |F_\varepsilon(g(z; \beta_1) - g(z; \beta_0) - r_n \delta(z) \mid z) - F_\varepsilon(g(z; \beta_2) - g(z; \beta_0) - r_n \delta(z) \mid z)| \\
&\leq \quad f_\varepsilon(v_n \mid z) |g(z; \beta_1) - g(z; \beta_2)| \\
&\leq \quad C A(z) \|\beta_1 - \beta_2\|
\end{aligned} \tag{4.5}
$$

for some v_n between $g(z; \beta_1) - g(z; \beta_0) - r_n \delta(z)$ and $g(z; \beta_2) - g(z; \beta_0) - r_n \delta(z)$. By Pakes and Pollard (1989, Lemma 2.13), the class of functions $\{z \mapsto \mathbb{E}\left[\psi_\tau(l(\varepsilon, z; \beta) + r_n \delta(z))\right] : \beta \in B\}$ is *Euclidean(c,d)* for an envelope with a finite fourth moment, with c and d independent of n. Deduce that the empirical process $\nu_n(\beta)$, $\beta \in B$, is indexed by a class of functions that is Euclidean for a squared integrable envelope. Finally, condition (ii) of Corollary 8 of Sherman (1994), can be checked from inequalities like in (4.5) and conditions on $|\dot{g}(z; \beta) - \dot{g}(z; \beta_0)|$.

On the other hand, because $\widehat{\beta}$ minimizes $\Gamma_n(\beta)$ defined in (2.3) over β, the directional derivative of $\Gamma_n(\beta)$ at $\widehat{\beta}$ along any direction γ (with $\|\gamma\| = 1$) is nonnegative. That is

$$
\begin{aligned}
0 \quad &\leq \quad \lim_{t \to 0} t^{-1}\left[\Gamma_n(\widehat{\beta} + t\gamma) - \Gamma_n(\widehat{\beta})\right] \\
&= \quad - \sum_{\{Y_i \neq g(Z_i; \widehat{\beta})\}} \psi_\tau\left(Y_i - g(Z_i; \widehat{\beta})\right) \gamma' \dot{g}(Z_i; \widehat{\beta}) \\
&\quad + \lim_{t \to 0} \sum_{\{Y_i = g(Z_i; \widehat{\beta})\}} t^{-1}\rho_\tau\left(g(Z_i; \widehat{\beta}) - g(Z_i; \widehat{\beta} + t\gamma)\right) \\
&= \quad - \sum_{\{Y_i \neq g(Z_i; \widehat{\beta})\}} \psi_\tau\left(Y_i - g(Z_i; \widehat{\beta})\right) \gamma' \dot{g}(Z_i; \widehat{\beta}) \\
&\quad - \sum_{\{Y_i = g(Z_i; \widehat{\beta})\}} \psi_\tau\left(-\gamma' \dot{g}(Z_i; \widehat{\beta})\right) \gamma' \dot{g}(Z_i; \widehat{\beta}) \\
&= \quad -D_{1n}(\widehat{\beta}) - D_{2n}(\widehat{\beta}).
\end{aligned} \tag{4.6}
$$

By Assumption 2.2, $|D_{2n}(\widehat{\beta})|$ is bounded by $\sum_{\{Y_i = g(Z_i; \widehat{\beta})\}} A(Z_i)$. As, for any x, the error term u has a continuous law given $Z = z$, the number of observations with $Y_i = g(Z_i; \widehat{\beta})$ is bounded in probability as the sample size tends to infinity. On the other hand, the moment condition on $A(\cdot)$ implies that $\max_{1 \leq i \leq n} A(Z_i) = o_{\mathbb{P}}\left(n^{1/2}\right)$. As γ is an arbitrary direction, it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau \left(Y_i - g(Z_i; \widehat{\beta})\right) \dot{g}(Z_i; \widehat{\beta}) = o_{\mathbb{P}}(1). \tag{4.7}$$

Finally, since $\widehat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$ and $\tau = F_\varepsilon(0 \mid Z_i)$, deduce that

$$
\begin{aligned}
\nu_n(\beta_0) &= \nu_n(\widehat{\beta}) + o_{\mathbb{P}}(1) && \text{[by (4.4)]} \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\left[\psi_\tau \left(Y_i - g(Z_i; \widehat{\beta})\right) \mid Z_i\right] \dot{g}(Z_i; \widehat{\beta}) + o_{\mathbb{P}}(1) && \text{[by (4.7)]} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[F_\varepsilon \left(g(Z_i; \widehat{\beta}) - g(Z_i; \beta_0) - r_n \delta(Z_i) \mid Z_i\right) - \tau\right] \dot{g}(Z_i; \widehat{\beta}) + o_{\mathbb{P}}(1) \\
&= \left\{\frac{1}{n} \sum_{i=1}^{n} f_\varepsilon(0 \mid Z_i) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_i; \beta_0)\right\} \sqrt{n}\left(\widehat{\beta} - \beta_0\right) \\
&\quad -r_n \left\{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_\varepsilon(0 \mid Z_i) \delta(Z_i) \dot{g}(Z_i; \beta_0)\right\} \\
&\quad +r_n^2 \sqrt{n} \left\{\frac{1}{n} \sum_{i=1}^{n} f'_\varepsilon(0 \mid Z_i) \delta^2(Z_i) \dot{g}(Z_i; \beta_0)\right\} \\
&\quad +o_{\mathbb{P}}\left(\sqrt{n}\|\widehat{\beta} - \beta_0\|\right) + o_{\mathbb{P}}\left(r_n^2 \sqrt{n}\right),
\end{aligned}
$$

where the last equality is based on a local expansions of $F_\varepsilon(\cdot \mid z)$ and $g(z; \cdot)$. By the law of large numbers, the central limit theorem and the fact that $\nu_n(\beta_0) = O_{\mathbb{P}}(1)$ and the random vector $f_u(0 \mid Z) \delta(Z) \dot{g}(Z; \beta_0)$ has zero mean, we obtain

$$\mathbb{E}[f_\varepsilon(0 \mid Z) \dot{g}(Z; \beta_0) \dot{g}'(Z; \beta_0)] \sqrt{n}\left(\widehat{\beta} - \beta_0\right) + r_n^2 \sqrt{n} \mathbb{E}[f'_\varepsilon(0 \mid Z) \delta^2(Z) \dot{g}(Z; \beta_0)] = O_{\mathbb{P}}(1)$$

from which the result follows. ∎

Lemma 4.2. *Let k be a positive integer and $\mathcal{G}$ a degenerate class of real-valued functions on $\mathbb{R}^{1+q} \times \ldots \times \mathbb{R}^{1+q}$. Suppose $\mathcal{G}$ is* Euclidean(c,d) *for a squared integrable envelope and some $c, d > 0$. Fix $z_1, \ldots, z_n \in \mathbb{R}^q$ and let $u_1, \ldots, u_n, u_{n+1}, \ldots, u_{2n}$ be independent copies of the random variable u. For $i = 1, \ldots, n$, let $v_i = (u_i, z_i)$ and $v_{n+i} = (u_{n+i}, z_i)$. Define $g_{\mathbf{i_k^n}}(u_{i_1}, \ldots, u_{i_k}) = g(v_{i_1}, \ldots, v_{i_k})$ and define $g_{\mathbf{i_k^{2n}}}$ similarly. Suppose that for any $k-$tuple $\mathbf{i_k^n}$, the function $g_{\mathbf{i_k^n}}$ is degenerate as a function of u_i variables (necessarily the same property holds also for any $k-$tuple $\mathbf{i_k^{2n}}$). Let*

$$U_{n,z_1,\ldots,z_n}^k(g) = (n)_k^{-1} \sum_{\mathbf{i_k^n}} g_{\mathbf{i_k^n}}(u_{i_1}, \ldots, u_{i_k}), \quad U_{2n,z_1,\ldots,z_n}^k(g) = (2n)_k^{-1} \sum_{\mathbf{i_k^{2n}}} g_{\mathbf{i_k^{2n}}}(u_{i_1}, \ldots, u_{i_k}).$$

Then for any $\alpha \in (0, 1)$, there exists a constant Λ depending only on α and k (and independent of n and the sequence $z_1, ..., z_n$) such that

$$\mathbb{E}\left[\sup_{\mathcal{G}} |n^{k/2} U_{n, z_1, ..., z_n}^k(g)|\right] \leq \Lambda \mathbb{E}^{1/2}\left[\sup_{\mathcal{G}}\{U_{2n, z_1, ..., z_n}^k(g^2)\}^\alpha\right].$$

Proof. We sketch the steps of the proof that follows the lines of the proof of the Main Corollary in Sherman (1994). For the sake of simplicity, we only consider the case of Euclidean families for a constant envelope. Fix n and $z_1, ..., z_n$ arbitrarily.

i) Symmetrization inequality. For each $g \in \mathcal{G}$ define $\widetilde{g}(\mathbf{i_k^n})$ as a sum of 2^k terms, each having the form

$$(-1)^r g_{\mathbf{i_k^n}}(u_{i_1}^*, \ldots, u_{i_k}^*)$$

with $u_{i_j}^*$ equal to either u_{i_j} or u_{n+i_j} where i_j ranges over the set $\{1, ..., n\}$, and r is the number of elements $u_{i_1}^*, ..., u_{i_k}^*$ belonging to $\{u_{n+1}, ..., u_{2n}\}$. Independently, take a sample $\sigma_1, ..., \sigma_n$ of Rademacher random variables, that is symmetric variables on the two points set $\{-1, 1\}$. Let Φ be a convex function on $[0, \infty)$. Then

$$\mathbb{E}\Phi\left(\sup_{\mathcal{G}}\left|\sum_{\mathbf{i_k^n}} g_{\mathbf{i_k^n}}(u_{i_1}, \ldots, u_{i_k})\right|\right) \leq \mathbb{E}\Phi\left(\sup_{\mathcal{G}}\left|\sum_{\mathbf{i_k^n}} \sigma_{i_1} \ldots \sigma_{i_k} \widetilde{g}(\mathbf{i_k^n})\right|\right). \qquad (4.8)$$

The proof of this inequality is omitted as it can be derived with only formal changes from the proof of Sherman (1994)'s symmetrization inequality. It can be also be derived from the lines of de la Peña and Giné (1999), Theorem 3.5.3 (see also Remark 3.5.4 of de la Peña and Giné).

ii) Maximal inequality. The following arguments are similar to those in Sherman (1994), section 5. Define the stochastic process

$$Z(g) = n^{k/2} \sum_{\mathbf{i_k^n}} \sigma_{i_1} \ldots \sigma_{i_k} \widetilde{g}(\mathbf{i_k^n}), \quad g \in \mathcal{G}$$

and the pseudo-metric $d_{U_{2n}^k}(g_1, g_2) = [U_{2n, z_1, ..., z_n}^k(|g_1 - g_2|^2)]^{1/2}$. Finally, let us remark that for each g, by Cauchy-Schwarz inequality and the definitions of $\widetilde{g}(\mathbf{i_k^n})$ and $g_{\mathbf{i_k^{2n}}}$ we have

$$\sum_{\mathbf{i_k^n}} \widetilde{g}(\mathbf{i_k^n})^2 \leq 2^k \sum_{\mathbf{i_k^{2n}}} g_{\mathbf{i_k^{2n}}}^2(u_{i_1}, ..., u_{i_k}) = 2^k (2n)_k U_{2n, z_1, ..., z_n}^k(g^2)$$

which is the counterpart of inequality (5) of Sherman (1994). Now, we have all the ingredients to continue exactly as in the proof of Sherman's maximal inequality and to deduce that for any positive integer m

$$\mathbb{E}\left[\sup_{\mathcal{G}} |n^{k/2} U_{n, z_1, ..., z_n}^k(g)|\right] \leq \Gamma \mathbb{E}\left[\int_0^{\delta_n^k} [D(x, d_{U_{2n}^k}, \mathcal{G})]^{1/2m} dx\right]$$

where $D(\epsilon, d_{U_{2n}^k}, \mathcal{G})$ are the *packing numbers* of the set $\mathcal{G}$ with respect to the pseudometric $d_{U_{2n}^k}$, $\delta_n^k = \sup_{\mathcal{G}} \sqrt{U_{2n,z_1,...,z_n}^k(g^2)}$ and Γ is a constant depending only on m and k.

iii) Moment inequality for Euclidean families. If $\mathcal{G}$ is *Euclidean(c,d)* for a constant envelope equal to one, then the packing number $D(\epsilon, d_{U_{2n}^k}, \mathcal{G})$ is bounded by $c\epsilon^{-d}$. To check this, apply the definition of an Euclidean family for $\mathcal{G}$ with μ the measure that places mass $(2n)_k^{-1}$ at each of the $(2n)_k$ pairs (v_i, v_j), $1 \le i \ne j \le 2n$. Finally, our result follows using the arguments of the Main Corollary of Sherman (1994). ∎