

Powerful nonparametric checks for quantile regression

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

Abstract

We propose a new and simple lack-of-fit test for a parametric quantile regression. It 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 and in an empirical application with several covariates.

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

*CREST (Ensaï), France ; Université de Lyon, France and Université catholique de Louvain, Belgium. Email: samuel.maistre@uclouvain.be

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

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

1 Introduction

Parametric quantile regression, as introduced by Koenker and Bassett (1978), has emerged as a powerful 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. It is then particularly valuable if variables have asymmetric distributions or heavy tails. The monograph of Koenker (2005) and the review of Yu et al. (2003) detail the theory and practice of quantile regression.

As in any statistical modeling exercise, it is crucial to check the fit of a parametric quantile model. Sound inference hinges on the correct functional specification of the regression function, but the possibility of misspecification in a parametric framework cannot be ignored, especially as applied researchers tend to choose functional forms on the basis of parsimony and tractability. While there has been a large effort devoted to testing the fit of parametric mean regressions, only a few lack-of-fit tests of parametric quantile regressions have been proposed. Two competing approaches have been investigated. The first essentially compares the parametric fit with a nonparametric one. In this vein, Zheng (1998) builds a test on kernel smoothing over the design. Horowitz and Spokoiny (2002) extend this approach to propose an adaptive procedure that allow for a data-driven choice of the smoothing parameter. But, 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. A second approach is based on a weighted cumulative sum (cumsum) process of the residuals. Along these lines, He and Zhu (2003) extend the approach developed by Stute (1997) for mean regression checks, Bierens and Ginther (2001) generalize the integrated conditional moment test of Bierens and Ploberger (1997) to quantile regression, and Escanciano and Goh (2014) consider testing the correct specification of a linear quantile function for a continuum of quantile levels. Conde-Amboage et al. (2015) develop a cumsum-type test suitable for high-dimensional covariates. The smoothing approach and the cumsum one are related. As shown by Fan and Li (2000) in the context of goodness-of-fit tests of a parametric regression, the test statistic proposed by Bierens and Ploberger (1997) is essentially the same as the kernel-smoothing test of Härdle and Mammen (1993) but with a smoothing parameter that is held fixed independently of the sample size. This however has huge consequences on the properties of the test. The kernel-smoothing statistic has an asymptotically pivotal distribution, while the limit distribution of its competitor is a non-linear functional of a Gaussian process. The power properties of the two tests are also pretty different depending on the kind of alternatives considered, see e.g. Fan and Li (2000) and Guerre and Lavergne (2002). Practically, one observes in small samples different performances of the two types of test, whether in terms of size control or power, in the case where many covariates are present, see e.g. Lavergne and Patilea (2010).

In this work, we introduce a new testing methodology that mixes some features of the two main approaches investigated up to date. Our approach has three specific features. First, it does not require smoothing with respect to all covariates, but only on one continuous covariate. The test is omnibus, but mitigates the curse of dimensionality that appears with nonparametric smoothing, hence weakening the influence of the smoothing parameter and improving the power properties of the test with respect

to a standard smooth test. Our empirical application shows that indeed our test can be less sensitive to bandwidth and more powerful than such a competitor when several covariates are present. Moreover, in practice, when there are several continuous covariates, the covariate on which we smooth does not influence dramatically the test's outcome. Second, because there is still some smoothing, the test statistic is asymptotically pivotal, while bootstrap can be used to obtain accurate critical values of the test in small samples. We find indeed in our simulations that our test's level is well controlled by wild bootstrapping. Third, our test easily applies in a model with continuous as well as discrete covariates, but avoids sample splitting with respect to the modalities of the discrete variables. This feature also contributes to a better detection of an incorrect parametric model as illustrated in our empirical application.

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 procedure to compute critical values. In Section 3, we compare the small sample behavior of our test to some existing procedures, and we illustrate its use and advantages in analyzing data on children birthweight. 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 $Z \in \mathbb{R}^q$, $q \geq 2$. We assume that $Z = (W, X)'$, where W is a one-dimensional continuous random variable that admits a density with respect to the Lebesgue measure, while X may include both continuous and discrete variables.

We assume that the conditional distribution of Y given $Z = z$ is absolutely continuous for almost all z . The τ -th conditional quantile of Y given Z is then defined as $Q_\tau(z) = \inf\{y : \mathbb{P}(Y \leq y \mid Z = z) \geq \tau\}$. 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,

$$Y = g(Z; \beta_0) + \varepsilon, \quad \mathbb{P}(Y \leq g(Z; \beta_0) \mid Z) = \tau. \quad (2.1)$$

The validity of the parametric quantile regression is thus equivalent to

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

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

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

The key element of our testing approach is the following lemma, that for our purpose will be applied to the random variable $U = \mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau$. Hereafter, if $g : \mathbb{R}^k \rightarrow \mathbb{R}$ 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) . Let $K(\cdot)$ and $\psi(\cdot)$ be bounded, even, integrable functions with (almost everywhere) positive, integrable Fourier transforms, and assume $\int_{\mathbb{R}} K(v) dv > 0$. Assume $\mathbb{E}(|U|^2) < \infty$, and define*

$$I(h) = \mathbb{E} [U_1 U_2 h^{-1} K((W_1 - W_2)/h) \psi(X_1 - X_2)].$$

Then for any $h > 0$,

$$\mathbb{E}[U | W, X] = 0 \text{ a.s.} \Leftrightarrow I(h) = 0.$$

Moreover, if $\mathbb{P}(\mathbb{E}[U | W, X] = 0) < 1$, then $\inf_{h \in (0,1]} I(h) > 0$.

The result could be proved following the lines of Lavergne et al. (2015, Lemma 1). The details are provided in the Supplementary Material. From the above result, the null hypothesis $H_0 : \mathbb{E}[U | W, X] = 0$ a.s., with $U = \mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau$, reduces to $I(h) = 0$ for an arbitrary h . We consider a sequence of h decreasing to zero when the sample size increases, which is one of the ingredients to obtain an asymptotically pivotal 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) . If we knew β_0 , we could estimate $I(h)$ by the second-order U-statistic

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

where $U_i(\beta) = \mathbb{I}\{Y_i \leq g(Z_i; \beta)\} - \tau$ and $K_h(\cdot) = K(\cdot/h)$. To estimate β_0 , we follow Koenker and Bassett (1978), who showed that under (2.1) a consistent estimator of β_0 is obtained as

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

where $\rho_{\tau}(e) = e(\tau - \mathbb{I}(e < 0))$ 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

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

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 similar to the one proposed by Zheng (1998), with the fundamental difference that the latter uses a multidimensional smoothing kernel over $Z = (W, X)'$, that is $h^{-q} \tilde{K}((W_i - W_j)/h, (X_i - X_j)/h)$, so that its statistic exhibit a $nh^{q/2}$ rate of convergence under H_0 . By contrast we smooth only on one of the covariates, while the smoothing on the other covariates is replaced by

suitable weights determined by the function $\psi(\cdot)$. Consequently, as will be shown, our statistic has a $nh^{1/2}$ rate of convergence under the null hypothesis irrespective of the dimension of Z . The test statistic of Bierens and Ginther (2001) has a n rate of convergence, but its asymptotic distribution is not pivotal. As acknowledged by the authors, this yields some practical difficulties. The asymptotic distribution involves the density of the error term at zero, which could in principle be estimated nonparametrically, but this may yield unreliable results due to the sensitivity of the estimator to the smoothing parameter. Other authors rely on bootstrapping for obtaining critical values of cumsum-type tests. This however may not provide an accurate approximation for small or moderate sample sizes, and might become unpractical for very large data sets. Our proposal is an hybrid approach that combines the advantages of existing procedures.

The statistic v_n^2 is an estimator of the variance of $nh^{1/2}I_n(\beta_0)$ conditional on the Z_i under H_0 . It does not consistently estimate the conditional variance of a properly centered version of $nh^{1/2}I_n(\hat{\beta})$ under the alternative hypothesis in general, and it may overestimate this conditional variance causing some loss in power for the test. One could build a more robust estimator by adapting the ones proposed by Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005) in a mean regression context. But we decided in favor of v_n^2 as it is easier to compute and yields a well-behaved and powerful test in our simulations.

In practice, to obtain scale invariance of our test statistic, we recommend that the observations on each component of Z are scaled by an indicator of dispersion, such as their sample standard deviation. While we do not formally consider this in our theoretical analysis, it can be easily seen that this does not affect our results. When there are several continuous covariates, one should choose one as the W variable on which smoothing is applied. If one suspects a potential deviation from the null that is localized around some values of a particular W , one can expect a better power if the test statistic smoothes on this variable. In our simulations, we investigate the influence of the choice of W for the power performances of the test. In our empirical example, we find that the covariate on which we smooth does not matter much for the test's outcome.

From our theoretical study, the function $\psi(\cdot)$ as well as the kernel $K(\cdot)$ should possess an almost everywhere positive and integrable Fourier transform. This is true in particular for (products of) the triangular, normal, Laplace, logistic, and Student densities. Alternatively, one can choose $\psi(x) = g(\|x\|)$ where $g(\cdot)$ is any of the above univariate functions. While the outcome of the test may depend on the choice of the kernel $K(\cdot)$, this influence is expected to be limited as usual in nonparametric estimation. For the choice of the function $\psi(\cdot)$, unreported simulations results, as well as the experiments of Lavergne et al. (2015) for a test of significance of covariates in a nonparametric regression, reveal that it is not so.

Our following theory does not allow for a data-driven choice of the bandwidth h . It seems likely that the procedures of Horowitz and Spokoiny (2002) and Guerre and Lavergne (2005) could be adapted to our setup, but this is outside the scope of the present paper. In our simulations, we adopt a rule-of-thumb approach, and we let the bandwidth vary so as to evaluate the sensitivity of the test's performances to its choice.

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, the functions $K(\cdot)$ and $\psi(\cdot)$, and the bandwidth h .

Assumption 2.1. (a) *The random vectors $(\varepsilon_1, Z_1)', \dots, (\varepsilon_n, Z_n)'$ are independent copies of the random vector $(\varepsilon, Z)' \in \mathbb{R}^{1+q}$. The conditional τ th 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 $f_\varepsilon(\cdot | z)$ of ε given $Z = z$ is uniformly bounded. There exists $a > 0$ such that $f_\varepsilon(\cdot | z)$ is differentiable on $(-a, a)$ for any z with $|f'_\varepsilon(0 | z)| \leq C < \infty$. Moreover, the derivatives $f'_\varepsilon(\cdot | z)$ satisfy a uniform Hölder continuity condition, that is there exist positive constants C_2 and c independent of z such that $|f'_\varepsilon(u_1 | z) - f'_\varepsilon(u_2 | z)| \leq C_2 |u_1 - u_2|^c$, $\forall |u_1|, |u_2| \leq a$.*

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

(b) *The matrix*

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

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*

$$\begin{aligned} \left\| \frac{\partial}{\partial \beta} g(z; \beta) \right\| &\leq A(z), & \left\| \frac{\partial}{\partial \beta} g(z; \beta) \frac{\partial}{\partial \beta'} g'(z; \beta) \right\| &\leq D(z) \quad \text{for any } \beta, \\ \left\| \frac{\partial}{\partial \beta} g(z; \beta_1) - \frac{\partial}{\partial \beta} g(z; \beta_2) \right\| &\leq B(z) \|\beta_1 - \beta_2\| \quad \text{for any } z, \beta_1, \beta_2. \end{aligned}$$

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

Our assumptions combine standard assumptions for parametric quantile regression estimation and ones specific to our lack-of-fit test. Among the latter, the conditions on the error term ε in Assumption 2.1 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(a) does not impose the correctness of the parametric quantile model but ensures that the parameter estimator defined by (2.3) has a well-defined probability limit. 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 $q : \mathbb{R} \rightarrow \mathbb{R}$ monotone or of bounded variation, see e.g. van der Vaart and Wellner (1996, Section 2.6). If there exists $\beta \in B$ such that $g(Z, \beta)$

is squared integrable, then Assumption 2.2(d) follows from 2.2(c). Restrictions on the bandwidth from Assumption 2.3 are compatible with optimal choices for nonparametric estimation, see e.g. Härdle and Marron (1985), for regression checks, see Guerre and Lavergne (2002), or for quantile checks, see Horowitz and Spokoiny (2002). The following theorem states the asymptotic validity of our test.

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

Asymptotic pivotalness is obtained for our test statistic because some amount of smoothing is involved. However it is customary to observe that the normal asymptotic approximation is not very precise for small to moderate sample sizes. Hence bootstrap is usually relied upon to obtain more accurate critical values. In cumsum-type tests, the statistic is not asymptotically pivotal, and bootstrap should be used for any sample size. However, bootstrapping an asymptotically pivotal statistic is expected to provide a more accurate approximation. Results of Li and Wang (1998), who show that the bootstrap moments match the original moments at a rate which depends on the smoothing parameter, suggest that bootstrap approximation is more accurate when there is some smoothing. Even if bootstrap is used in both approaches, we can expect a better size control when bootstrapping a smooth test. Lavergne and Patilea (2012) report some simulations results illustrating this feature.

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

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

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

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

This condition ensures that our sequence of models defined by (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} . To our knowledge, this result on the behavior of $\widehat{\beta}$ under local alternatives is new. He and Zhu (2003) only considered the case $r_n = n^{-1/2}$ while Zheng (1998) assumed \sqrt{n} convergence to some fixed β^* under H_{1n} . Our next result states that these local alternatives can be detected whenever $r_n^2 n h^{1/2} \rightarrow \infty$. 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 $r_n^2 n h^{1/2} \rightarrow \infty$.*

Our test has more than trivial power against the above local alternative when the rate r_n at which they decrease is at most $\sqrt{nh^{1/2}}$, whether they depend on the variable W on which smoothing occurs, or

on X only. This comes from the rate of convergence of the test statistic under H_0 . By contrast, Zheng's test is consistent against the above local alternative when $r_n^2 n h^{q/2} \rightarrow \infty$, where $q = \dim(Z)$, and a cumsum-type test detects those with $r_n^2 n \rightarrow \infty$. So smoothing entails some loss of power against these local alternatives. Of course, one could consider different kind of local alternatives. Since smooth tests are typically powerful against a localized or oscillating alternative, one can expect that our test will be powerful against such a deviation from the null hypothesis if it depends on W only. We investigate this issue in our simulation experiments.

2.4 Bootstrap Critical Values

While the test statistic is asymptotically pivotal under H_0 , the asymptotic approximation of its behavior 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 $\mathbb{I}\{Y \leq g(Z; \beta_0)\} - \tau$ is under H_0 a Bernoulli random variable irrespective of the particular distribution of ε , as noted by Horowitz and Spokoiny (2002). Their proposal is thus to naively (or nonparametrically) bootstrap from the empirical distribution of the residuals. This bootstrap procedure remains asymptotically valid for non identically distributed errors. Alternatively, He and Zhu (2003) note that one could use any continuous distribution with a τ -th quantile equal to 0. While asymptotically valid, these two methods do not explicitly account for potential heteroscedastic errors. As will be shown in simulations, this can yield distorted levels for the test. Thus we will favor the wild bootstrap method for quantile regression introduced by Feng et al. (2011), that for our test works as follows.

1. Let $\hat{\varepsilon}_i = Y_i - g(Z_i; \hat{\beta})$, $1 \leq i \leq n$, and w_1, \dots, 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 $\varepsilon_i^* = w_i |\hat{\varepsilon}_i|$ and $Y_i^* = g(Z_i; \hat{\beta}) + \varepsilon_i^*$ for each $i = 1, \dots, n$.
2. Use the bootstrap data set $\{Y_i^*, Z_i : i = 1, \dots, n\}$ to compute the estimator $\hat{\beta}^*$, the new $U_i^*(\hat{\beta}^*) = \mathbb{I}\{Y_i^* \leq g(Z_i; \hat{\beta}^*)\} - \tau$, 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 $T_n \geq z_\alpha^*$. The following theorem states the asymptotic validity of the bootstrap test.

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

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(T_n^* \leq t \mid Y_1, Z_1, \dots, Y_n, Z_n) - \mathbb{P}(T_n \leq t \mid H_0)| \xrightarrow{p} 0,$$

where $\mathbb{P}(T_n \leq \cdot \mid H_0)$ is the cumulative distribution of T_n under H_0 .

3 Numerical Evidence

3.1 Small Sample Performances

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

$$Y = 1 + W + X + \delta (W^2 + WX + X^2) + \varepsilon, \quad (3.1)$$

where W follows a standard normal distribution, and X independently follows a binomial of size 5 and probability of success 0.5. We also studied a model with many covariates and deviations depending on linear indices, where

$$Y = 1 + W'\alpha + X'\beta + \delta_1 \sqrt{2} \sin(\theta(W_1 + X_1)) + \delta_2 \frac{(W_2 - X_2 + X_3)^2 - 3}{2\sqrt{3}} + \sigma \left(\frac{W_2 + X_2 + X_3}{\sqrt{3}} \right) \eta, \quad (3.2)$$

where $W \in \mathbb{R}^2$ follows a standard multivariate distribution, $X \in \mathbb{R}^3$ independently follows a standard multivariate distribution, $\alpha = (1, -1)'$, $\beta = (1, 1, 1)'$, $\sigma^2(x) = (1 + x^2)/2$ and $\eta = \mathcal{N}(0, 1) - \Phi^{-1}(\tau)$ where Φ is the standard Gaussian c.d.f. and τ is the tested quantile's order, so that the quantile of order τ of η is always 0.

We first consider the comparative performances of the three possible bootstrapping procedures detailed in the last section applied to our test under the null hypothesis, corresponding to $\delta = 0$ for Model (3.1) and $\delta_1 = \delta_2 = 0$ for Model (3.2). The linear quantile regression is estimated using the R package `quantreg`, see Koenker (2015). In computing our statistic, we chose $\psi(\cdot)$ and $K(\cdot)$ as the standard normal density. Figures 1 and 2 report our results based on 5000 replications for Models (3.1) and (3.2) and sample sizes of $n = 100$ and $n = 200$ at nominal level 10%, when the bandwidth is $h = cn^{-1/5}$ with c varying. To check the sensitivity of our results to the error's distribution, we considered in Model (3.1) the three distributions $\mathcal{N}(0, 1)$, $\log \mathcal{N}(0, 1) - 1$ and $\sigma(W) \times (\mathcal{N}(0, 1) - \Phi^{-1}(\tau))$. The three bootstrap methods provide accurate levels for any bandwidth choice when errors are identically distributed, while the use of asymptotic critical values yields large underrejection. In the heteroscedastic case, however, only the wild bootstrap delivers an empirical level close to the nominal one, while naive or uniform bootstrap result in a severely oversized test. These findings hold for the median regression model as well as the lowest decile model, i.e. $\tau = 0.1$. Hence we will use only the wild bootstrap in what follows.

Next, we investigated the power of our test. We compared our test to the one proposed by He and Zhu (2003, hereafter HZ), based on

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

where $\dot{g} = \partial g / \partial \beta$. We also computed the statistic proposed by Zheng (1998), which writes

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

where

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

\tilde{K} is the standard multivariate normal density and $h = n^{-1/(q+4)}$ where q is the total number of covariates. Finally, we computed the statistic from Conde-Amboage et al. (2015) (hereafter CSG), a bootstrapped version of the largest eigenvalue of

$$n^{-2} \sum_{i=1}^n \sum_{j=1}^n \left(\tau - \mathbb{I} [U_i(\hat{\beta}) < 0] \right) \left(\tau - \mathbb{I} [U_j(\hat{\beta}) < 0] \right) \dot{g}'(Z_i, \hat{\beta}) \dot{g}'(Z_j, \hat{\beta}) \sum_{r=1}^n A_{ijr},$$

where A_{ijr} is proportional to the complementary angle between the vectors $(X_i - X_r)$ and $(X_j - X_r)$. For Model (3.2), we performed smoothing on either the first or the second component of W .

We applied the wild bootstrap procedure to compute the critical values of all tests. Figure 3 gathers the power curves of the different tests for Model (3.1) as a function of δ based on 2500 replications for a median and a first decile regression model with $n = 100$, with either standard Gaussian or heteroscedastic Gaussian errors. For the median regression model with normal homoscedastic errors, all tests perform almost similarly. Our test is more powerful for a larger bandwidth, which was expected given our theoretical analysis. For heteroscedastic errors, our test, the smoothing test and CSG test always outperform HZ test. In the linear decile case ($\tau = 0.1$), our test and Zheng's perform similarly while CSG test is less powerful and HZ test is no better than a test with trivial power.

On Figure 4, we considered Model (3.2) with $\delta_2 = 0$ but δ_1 varying and $\theta = 2$. We represent the results for different sample sizes $n = 50, 100, 200$ and 400 . As suggested by a referee, we also considered Zheng's test using a fixed bandwidth, namely $h = 1$ whatever the sample size. By comparison, $n^{-1/5}$ varies from 0.46 to 0.3, and $n^{-1/9}$ varies from 0.65 to 0.51. The CSG and HZ non-smooth tests are not better than a constant test for $n = 100$ and only slightly better for larger sample sizes. Our test performs better when smoothing on W_1 , which is not surprising because the alternative depends only on W_1 and X_1 . But, contrary to what observed in the previous experiments, a larger bandwidth does not improve power, due to the oscillating nature of the alternative. When smoothing on W_2 , our test is less powerful as expected, but its power steadily improves with increasing sample size. Whatever component of W is smoothed, the influence of the bandwidth is much less for our test than for Zheng's test. Finally, the "fixed bandwidth" Zheng test is far to be the best performing.

On Figure 5 we graph the results for Model (3.2) when δ_1 and δ_2 vary together. The considered sample sizes are $n = 100$ or 200 and the frequencies $\theta = 1$ or 2 . For $\theta = 1$, our test performance does not depend on the smoothed covariate we choose. It is however more powerful for a larger bandwidth and in this case, its performance is similar to Zheng's test with a large or fixed bandwidth. For $\theta = 2$, the power curves of all the smooth tests are pretty close, with the exception of our test when we smooth on W_2 with a relatively small bandwidth. The power of CSG test increases with the sample size while the power of HZ test is nearly trivial.

3.2 Empirical Illustration

Our empirical example allows us to compare the results of different tests in a practical situation where several covariates are present and also to investigate whether the choice of the continuous variable on which smoothing is performed in our test matters in practice. We considered some parametric quantile models for children birthweight using data analyzed by Abrevaya (2001) and Koenker and Hallock (2001), who also gave a detailed data description. We focused on median regression and the 10th percentile quantile regression. Models are estimated and tested on a subsample of 1168 smoking college graduate mothers. Results are gathered in Table 1. We first considered a simple model which is linear in weight gain during pregnancy (WTGAIN) and average number of cigarettes per day (CIGAR), and quadratic in age (AGE) as suggested by the analysis of Koenker and Hallock (2001). We standardize all explanatory variables and we implement our test in turn with age, the average number of cigarettes, or the weight gain as the W variable. Other details are identical to what was done in our simulations. None of the tests detects misspecification in quantile models at a 10% nominal level, that is all the corresponding p -values are larger than 0.1.

We then 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 and CSG tests do not reject the model at either quantiles, while Zheng's test detects a misspecified median regression model only for the smaller bandwidth ($c = 1$) at 10% level. Our test detects a misspecified median regression model at 10% level, irrespective of the the considered bandwidth when we smooth on the weight gain as well as when we smooth on the age with bandwidth constant $c = 2$ or when smoothing on the average number of cigaretttes with $c = 1$. This shows that it can be more powerful than its competitors, especially in a practical situation where the number of covariates is large. This empirical exercise hence illustrates that our new test, beside existing procedures, is a valuable addition to the practitioner toolbox.

Table 1: Application: point estimates, standard deviations (in parentheses) and tests p -values

	$\tau = 0.5$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.1$
CIGAR	-5.05 (2.3)	-8.36 (3.53)	-5.07 (2.36)	-8.07 (3.25)
WTGAIN	7.69 (1.32)	14.96 (1.2)	8.31 (1.31)	15.91 (1.4)
AGE	43.6 (50.59)	133.67 (30.11)	78.59 (45.85)	117.62 (48.42)
AGESQ	-0.84 (0.81)	-2.23 (0.5)	-1.38 (0.72)	-1.94 (0.82)
BOY			137.22 (34.35)	-5.22 (47.33)
BLACK			-177.78 (75.09)	-124.18 (69.17)
MARRIED			21.62 (48.39)	41.75 (54.66)
NOVISIT			-211.62 (406.72)	-275.15 (112.5)
HZ	0.255	0.373	0.429	0.296
CSG	0.829	0.913	0.371	0.771
Zheng's test c=1	0.395	0.950	0.050	0.843
Zheng's test c=2	0.560	0.980	0.125	0.575
Our test c=1 W=AGE	0.812	0.930	0.130	0.990
Our test c=2 W=AGE	0.756	0.965	0.077	0.854
Our test c=1 W=CIGAR	0.131	0.988	0.081	0.499
Our test c=2 W=CIGAR	0.311	0.973	0.108	0.390
Our test c=1 W=WTGAIN	0.276	0.710	0.059	0.797
Our test c=2 W=WTGAIN	0.266	0.960	0.049	0.815

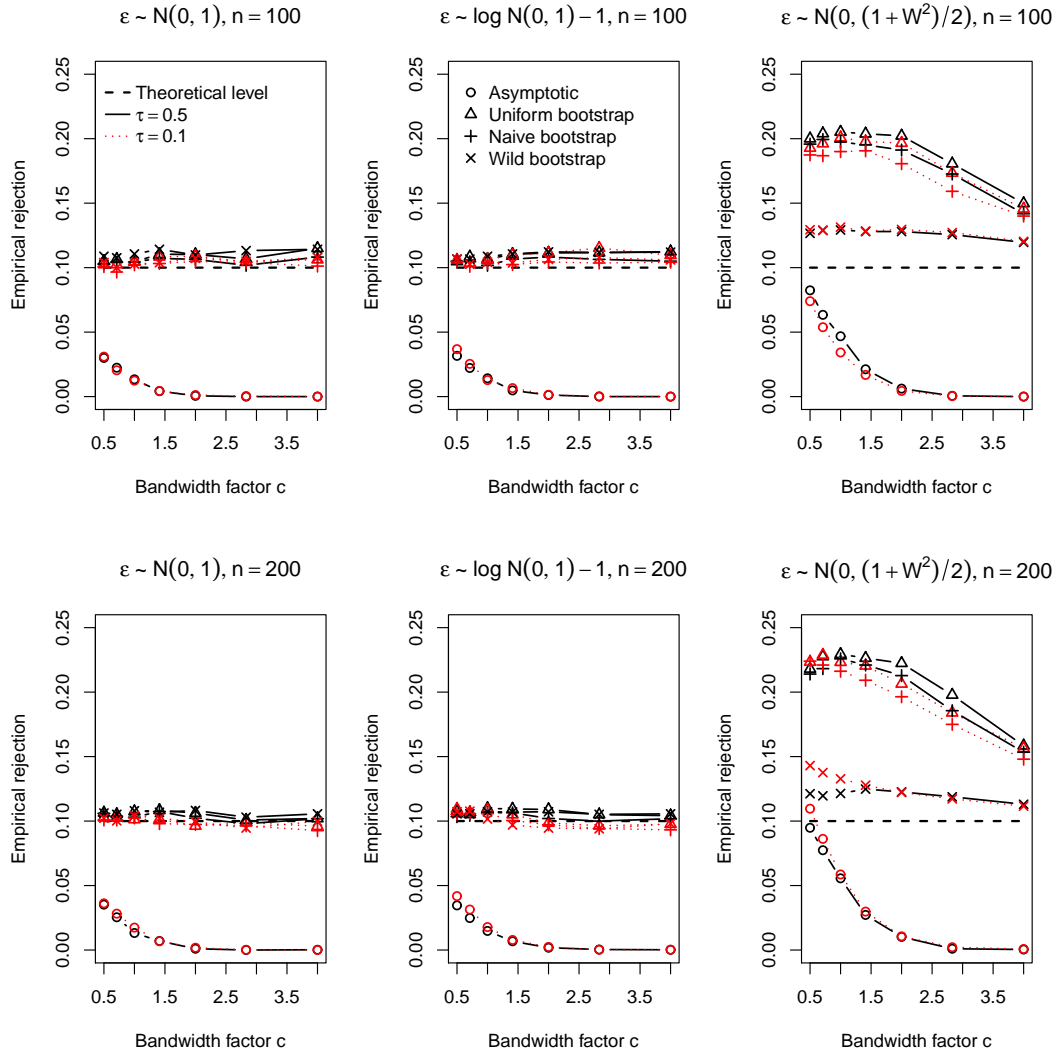


Figure 1: Empirical rejections for model (3.1) under H_0 as a function of the bandwidth, $n = 100$ and 200

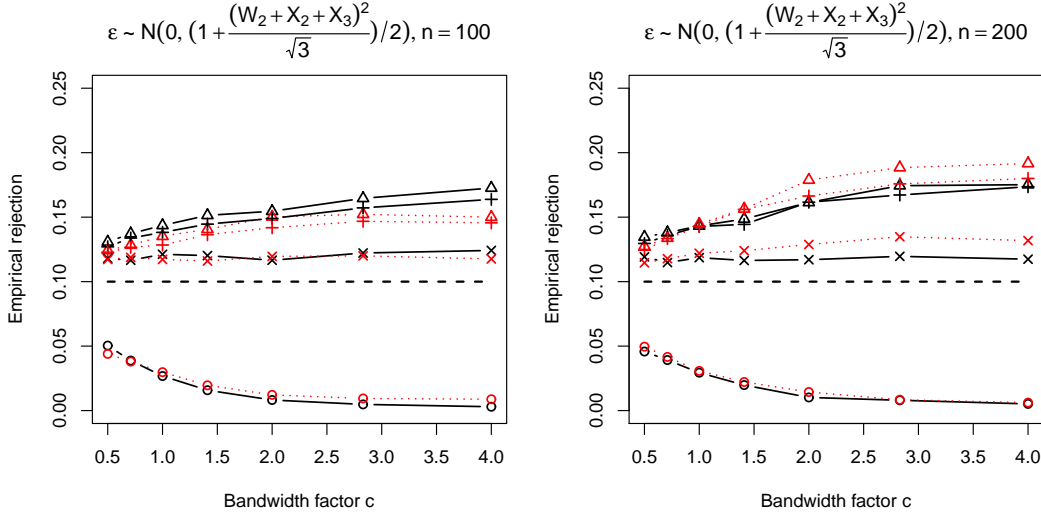


Figure 2: Empirical rejections for model (3.2) under H_0 as a function of the bandwidth, $n = 100$ and 200

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, \dots, 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). Below, 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, $F_\varepsilon(\cdot | z)$ is the conditional distribution function of ε given $Z = z$, such that $F_\varepsilon(0 | \cdot) \equiv \tau$. Below C, C_1, C_2, \dots 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

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

Let us introduce some simplifying notation:

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

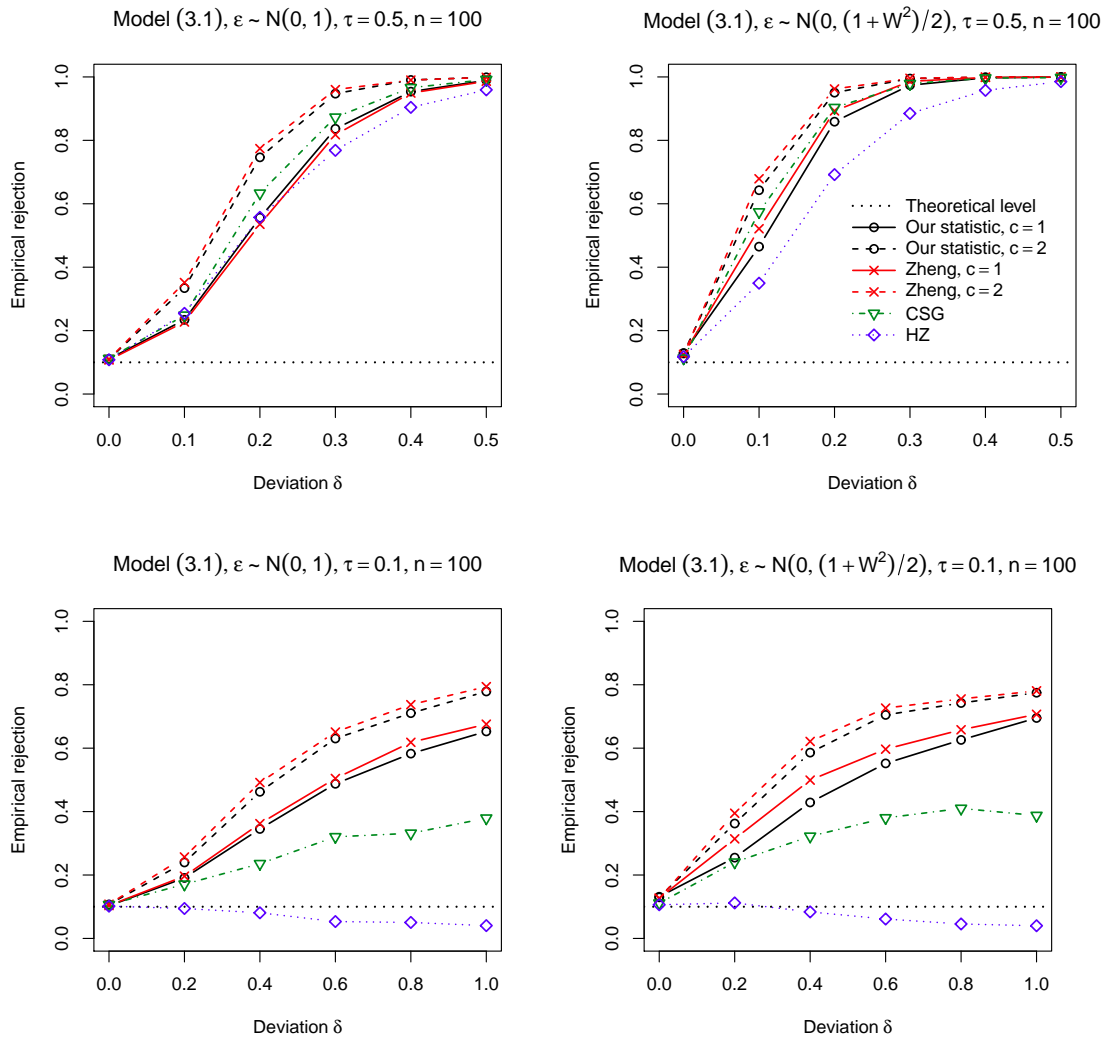


Figure 3: Power curves for model (3.1).

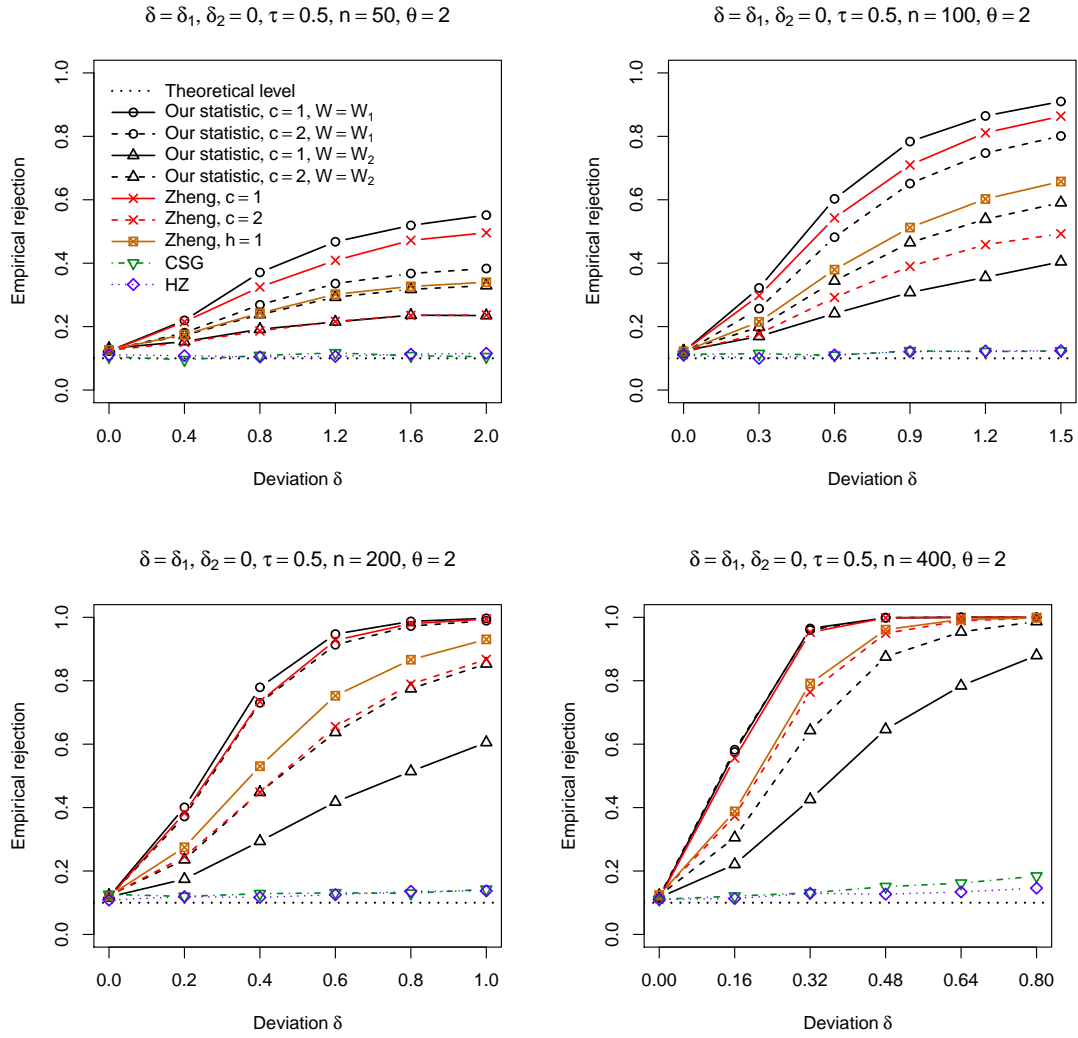


Figure 4: Power curves for model (3.2), $\theta = 2, \delta_2 = 0$.

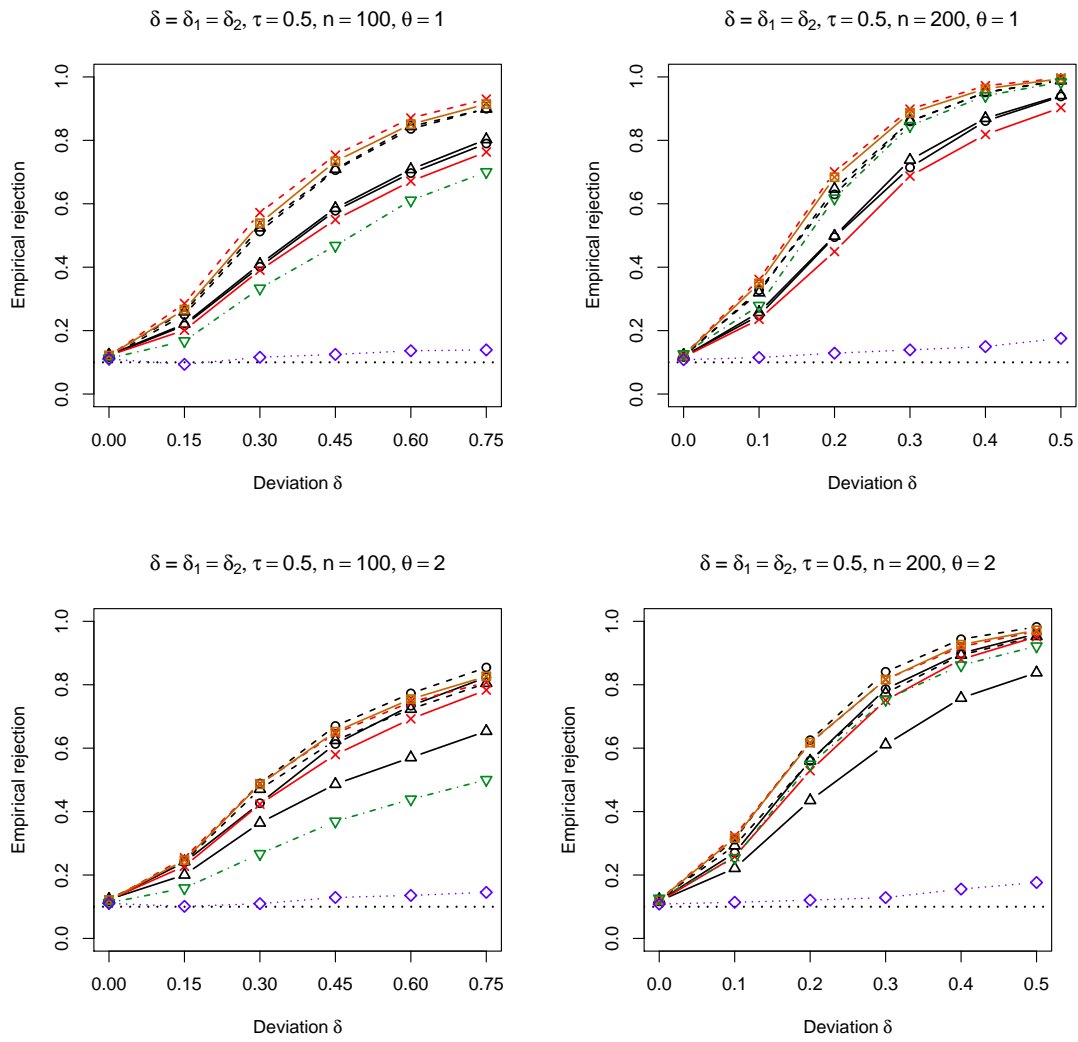


Figure 5: Power curves for model (3.2), $\theta = 1$ and $2, \delta_1 = \delta_2$.

Under \mathcal{H}_0

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

By a Taylor expansion, decompose

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

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

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

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

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

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

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

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 of 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). For the required Euclidean property, see, for instance, Nolan and Pollard (1987) and Pakes and Pollard (1989). Next, if $\dot{g}^{(l)}$ denotes the l th component of the vector of first-order derivatives \dot{g} , $1 \leq l \leq p$, and

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

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

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

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

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

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

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

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 4.2 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 Material.

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 $r_n \rightarrow 0$. 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*

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

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}

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

Let us decompose

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

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

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

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

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

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

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

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

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

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

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

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

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. The fact that the mean of W_{7n}/r_n^2 tends to a positive constant C_1 could be proved by the same arguments as used in Lemma 2.1 to show that $\lim_{h \rightarrow 0} I(h) > 0$ when $\mathbb{P}(\mathbb{E}[U | W, X] = 0) < 1$ (see the Supplementary Material).

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

$$\text{Var}(W_{1n}(\beta_0) | Z_1, \dots, Z_n) = \frac{1}{n^2(n-1)^2} \sum_{i \neq j} F_\varepsilon(-r_n \delta_i | Z_i) [1 - F_\varepsilon(-r_n \delta_i | Z_i)] \\ \times F_\varepsilon(-r_n \delta_j | Z_j) [1 - F_\varepsilon(-r_n \delta_j | Z_j)] h^{-2} K_{h,ij}^2 \psi_{ij}(\mu) \\ \leq \frac{h^{-1}}{16n(n-1)} \left[\frac{1}{n(n-1)} \sum_{i \neq j} h^{-1} K_{h,ij}^2 \psi_{ij} \right]$$

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, $W_{1n}(\beta_0) = o_{\mathbb{P}}(r_n^2)$, provided that $r_n^2nh^{1/2} \rightarrow \infty$. Next, let

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

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 $n^{1/2}r_n^2 \rightarrow \infty$, we have

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

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

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

The remaining terms W_{2n} , W_{3n} and W_{4n} can be treated in the following way. By Hoeffding's decomposition $r_n^{-1}hW_{3n}(\beta) = U_n^2(\beta) + U_n^1(\beta)$ with U_n^1 , U_n^2 degenerate U -processes of order 1 and 2, respectively. In view of Assumption 2.2(d) and the fact that $K(\cdot)$ is bounded, apply Corollary 4 of Sherman (1994) to deduce that $U_n^2(\beta) = O_{\mathbb{P}}(n^{-1})$ uniformly in β . If $K_{h,ij}(\theta) = K_h((X_i - X_j)'\theta)$ and

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

we can write

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

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

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

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

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

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 $U_i(\beta)$ with $U_i^*(\beta) = \mathbb{I}\{Y_i^* \leq g(Z_i; \beta)\} - \tau$ 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 $nh^{1/2}W_n^*(\hat{\beta})/v_n$ *conditionally* upon the original sample; and (b) the $O_p(n^{-1/2})$ rate for $\hat{\beta}^* - \hat{\beta}$, and the negligibility of $W_n^*(\hat{\beta}^*) - W_n^*(\hat{\beta})$ given the original sample. If S_{1n}^* and S_{2n}^* denote bootstrapped statistics, S_{1n}^* is bounded in probability given the sample if

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

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

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

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

Lemma 4.2. *Under the assumptions of Theorem 2.4,*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(nh^{1/2}W_n^*(\hat{\beta})/v_n \leq t \mid Y_1, Z_1, \dots, Y_n, Z_n\right) - \Phi(t) \right| \rightarrow 0, \quad \text{in probability.}$$

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 $U_i^*(\hat{\beta}) = \mathbb{I}\{Y_i^* \leq g(Z_i; \hat{\beta}^*)\} - \tau$. Define the martingale array $\{S_{n,m}^*, \mathcal{F}_{n,m}^*, 1 \leq m \leq n, n \geq 1\}$ where $S_{n,1}^* = 0$ and $S_{n,m}^* = \sum_{i=2}^m G_{n,i}^*$ with

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

and $\mathcal{F}_{n,m}^*$ is the σ -field generated by $\{\bar{Z}, w_1, \dots, w_m\}$ where $\bar{Z} = \{Y_1, \dots, Y_n, Z_1, \dots, Z_n\}$ and w_1, \dots, w_n are the bootstrap weights. Thus $nh^{1/2}W_n^*(\hat{\beta}) = S_{n,n}^*$. Next define

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

Recall that

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

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

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

Moreover,

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

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

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

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

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

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 $O_{\mathbb{P}}(n^{-1/2})$ rate for $\hat{\beta}^* - \hat{\beta}$, and the negligibility of $W_n^*(\hat{\beta}^*) - W_n^*(\hat{\beta})$ 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 $(n)_k = n(n-1)\dots(n-k+1)$ and let $\mathbf{i}_k^n = (i_1, \dots, i_k)$ be a k -tuple of distinct integers from the set $\{1, \dots, n\}$. Similarly, $\mathbf{i}_k^{2n} = (i_1, \dots, i_k)$ denotes a k -tuples of distinct integers from $\{1, \dots, 2n\}$. Moreover, a function g on \mathcal{S}^k is called degenerate if for each $i = 1, \dots, k$, and all $s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_k \in \mathcal{S}$, $\mathbb{E}[g(s_1, \dots, s_{i-1}, S, s_{i+1}, \dots, s_k)] = 0$.

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

Lemma 4.3. *Let k be a positive integer and \mathcal{G} a degenerate class of real-valued functions on $\mathbb{R}^{1+q} \times \dots \times \mathbb{R}^{1+q}$. Suppose \mathcal{G} is Euclidean(c,d) for a squared integrable envelope and some $c, d > 0$. Fix $z_1, \dots, z_n \in \mathbb{R}^q$ and let $u_1, \dots, u_n, u_{n+1}, \dots, u_{2n}$ be independent copies of the random variable u . For $i = 1, \dots, n$, let $v_i = (u_i, z_i)$ and $v_{n+i} = (u_{n+i}, z_i)$. Define $g_{\mathbf{i}_k^n}(u_{i_1}, \dots, u_{i_k}) = g(v_{i_1}, \dots, v_{i_k})$ 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*

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

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, \dots, z_n) such that

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

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 4.3, $\sup_{\beta} |n^{-1} \Gamma_n^*(\beta) - \mathbb{E}[\rho_\tau(Y - g(Z; \beta)) | \overline{Z}]|$ is asymptotically negligible given the sample $\overline{Z} = \{Y_1, \dots, Y_n, Z_1, \dots, Z_n\}$. 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

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

indexed by β . Lemma 4.3 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 (S.5) of the Supplementary Materials, that is using the directional derivative of $\Gamma_n^*(\beta)$ at $\widehat{\beta}^*$ along any direction γ , deduce

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

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 ,

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

and for any sample \overline{Z} , the distribution function $F_{\varepsilon^*}(\cdot | \overline{Z})$ 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 $F_{\varepsilon^*}(\cdot | \overline{Z})$ 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 4.3, derive conditional versions of Lemma 1 of Zheng (1998) and of Corollary 4 of Sherman (1994). Here, we only need conditional versions of such results for families of functions having the

Euclidian property for a constant envelope. 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 $\varepsilon > 0$

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

Acknowledgements

S. Maistre acknowledges financial support from the contract Projet d'Actions de Recherche Concertées (ARC) 11/16-039 of the Communauté française de Belgique, granted by the Académie universitaire Louvain. V. Patilea acknowledges support from the chair "New Challenges for New Data" of LCL and Genes.

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 (2001): "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, vol. 26 of *Stud. Empir. Econom.*, 307–324.
- BIERENS, H. J. AND W. PLOBERGER (1997): "Asymptotic theory of integrated conditional moment tests," *Econometrica*, 65, 1129–1151.
- CONDE-AMBOAGE, M., C. SÁNCHEZ-SELLERO, AND W. GONZÁLEZ-MANTEIGA (2015): "A lack-of-fit test for quantile regression models with high-dimensional covariates," *Comput. Statist. Data Anal.*, 88, 128–138.
- ESCANCIANO, J. AND S. GOH (2014): "Specification analysis of linear quantile models," *J. Econometrics*, 178, 495–507.
- FAN, Y. AND Q. LI (2000): "Consistent Model Specification Tests: Kernel-Based Tests versus Bierens' ICM Tests," *Econometric Theory*, 16, 1016–1041.
- 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*, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London.
- HÄRDLE, W. AND E. MAMMEN (1993): "Comparing nonparametric versus parametric regression fits," *Ann. Statist.*, 21, 1926–1947.
- 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.
- (2015): *quantreg: Quantile Regression*, r package version 5.19.
- KOENKER, R. AND J. G. BASSETT (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 (2015): “A significance test for covariates in nonparametric regression,” *Electron. J. Statist.*, 9, 643–678.
- LAVERGNE, P. AND V. PATILEA (2010): “One for All and All for One: Regression Checks With Many Regressors,” Discussion Papers dp08-06, Department of Economics, Simon Fraser University.
- (2012): “One for all and all for one: regression checks with many regressors,” *J. Bus. Econom. Statist.*, 30, 41–52.
- LI, Q. AND S. WANG (1998): “A Simple Consistent Bootstrap Test for a Parametric Regression Function,” *J. Econometrics*, 87, 145–165.
- 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 Series in Statistics, Springer-Verlag, New York.
- 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”

Samuel Maistre*, Pascal Lavergne† and Valentin Patilea‡

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) . Let $K(\cdot)$ and $\psi(\cdot)$ be bounded, even, integrable functions with (almost everywhere) positive, integrable Fourier transforms, and assume $\int_{\mathbb{R}} K(v)dv > 0$. Assume $\mathbb{E}(|U|^2) < \infty$, and define*

$$I(h) = \mathbb{E} [U_1 U_2 h^{-1} K((W_1 - W_2)/h) \psi(X_1 - X_2)] .$$

Then for any $h > 0$,

$$\mathbb{E}[U | W, X] = 0 \text{ a.s.} \Leftrightarrow I(h) = 0.$$

Moreover, if $\mathbb{P}(\mathbb{E}[U | W, X] = 0) < 1$, then $\inf_{h \in (0,1]} I(h) > 0$.

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,

$$\begin{aligned} I(h) &= \mathbb{E} \left[U_1 U_2 \int_{\mathbb{R}} e^{2\pi i t(W_1 - W_2)} \mathcal{F}[K](th) dt \int_{\mathbb{R}^{q-1}} e^{2\pi i \langle s, X_1 - X_2 \rangle} \mathcal{F}[\psi](s) ds \right] \\ &= \int_{\mathbb{R}^{q-1}} \int_{\mathbb{R}} |\mathbb{E}[\mathbb{E}[U | W, X] e^{2\pi i \{tW + \langle s, X \rangle\}}]|^2 \mathcal{F}[K](th) \mathcal{F}[\psi](s) dt ds . \end{aligned}$$

*CREST (Ensaï), France ; Université de Lyon, France and Université catholique de Louvain, Belgium.

Email: samuel.maistre@uclouvain.be

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

‡CREST (Ensaï), France. Email: patilea@ensai.fr

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

$$\mathbb{E} [\mathbb{E} [U | W, X] e^{2\pi i\{tW+(s,X)\}}] = 0 \quad \forall t, s \Leftrightarrow \mathbb{E} [U | W, X] = 0 \quad \text{a.s.}$$

Next, we consider the case $\mathbb{P}(\mathbb{E} [U | W, X] = 0) < 1$. Let us notice that, by the condition $\mathbb{E} [\mathbb{E}^2 [U | W, X]] < \infty$ and Plancherel's Theorem, the map

$$(t, s) \mapsto \left| \mathbb{E} [\mathbb{E} [U | W, X] e^{2\pi i\{tW+(s,X)\}}] \right|, \quad (t, s) \in \mathbb{R} \times \mathbb{R}^{q-1}, \quad (\text{S.1})$$

is squared integrable. Moreover, the Fourier Transforms $\mathcal{F}[K](\cdot)$ and $\mathcal{F}[\psi](\cdot)$ are bounded, and $\forall t \in \mathbb{R}, \lim_{h \rightarrow 0} \mathcal{F}[K](ht) = \int_{\mathbb{R}} K(v)dv$. Then the Lebesgue Dominated Convergence Theorem implies that the map $h \mapsto I(h)$ is continuous on $(0, 1]$ and

$$\lim_{h \rightarrow 0} I(h) = \int_{\mathbb{R}^{q-1}} \int_{\mathbb{R}} \left| \mathbb{E} [\mathbb{E} [U | W, X] e^{2\pi i\{tW+(s,X)\}}] \right|^2 \mathcal{F}[\psi](s) dt ds \int_{\mathbb{R}} K(v)dv.$$

Thus $I(\cdot)$ could be extended by continuity on $[0, 1]$. Finally, since the map defined in Equation (S.1) is also nonnegative and non identically equal to 0 whenever $\mathbb{E} [U | W, X] \neq 0$, and $\mathcal{F}[\psi](\cdot)$ and $\mathcal{K}[\psi](\cdot)$ are almost everywhere positive, $\lim_{h \rightarrow 0} I(h)$ is necessarily positive, and so is $I(h)$ for any $h \in (0, 1]$. ■

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 $r_n \rightarrow 0$. If $\hat{\beta} = \arg \min_{\beta \in B} \Gamma_n(\beta)$ with $\Gamma_n(\beta) = \sum_{i=1}^n \rho_\tau(Y_i - g(Z_i; \beta))$, then under $\mathcal{H}_0, \hat{\beta} - \beta_0 = O_{\mathbb{P}}(n^{-1/2})$ and under H_{1n} defined in (2.5), $\hat{\beta} - \beta_n = O_{\mathbb{P}}(n^{-1/2})$ where*

$$\beta_n = \beta_0 - r_n^2 [\mathbb{E} [f'_\varepsilon(0 | Z) \dot{g}(Z; \beta_0) \dot{g}'(Z; \beta_0)]]^{-1} \mathbb{E} [f'_\varepsilon(0 | Z) \delta^2(Z) \dot{g}(Z; \beta_0)].$$

Proof. It is easy to check that

$$|\rho_\tau(a - b) - \rho_\tau(a)| \leq |b| \max(\tau, 1 - \tau) \leq |b|. \quad (\text{S.2})$$

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, $\sup_{\beta} |n^{-1}\Gamma_n(\beta) - \mathbb{E}[\rho_{\tau}(Y - g(Z; \beta))]| \rightarrow 0$, 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 $\sup_{\beta \in B} |\Delta_n(\beta)| \rightarrow 0$ in probability, where

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

and $l(u, z; \beta) = u + g(z; \beta_0) - g(z; \beta)$. By inequality (S.2),

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

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

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

indexed by β . First, let us notice that

$$\nu_n(\beta) - \nu_n(\beta_0) = o_{\mathbb{P}}(1) \tag{S.3}$$

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

$$\begin{aligned}
& |\mathbb{E}[\psi_\tau(l(\varepsilon, z; \beta_1) + r_n \delta(z))] - \mathbb{E}[\psi_\tau(l(\varepsilon, z; \beta_2) + r_n \delta(z))]| & (S.4) \\
& \leq |F_\varepsilon(g(z; \beta_1) - g(z; \beta_0) - r_n \delta(z) \mid z) - F_\varepsilon(g(z; \beta_2) - g(z; \beta_0) - r_n \delta(z) \mid z)| \\
& \leq f_\varepsilon(v_n \mid z) |g(z; \beta_1) - g(z; \beta_2)| \\
& \leq CA(z) \|\beta_1 - \beta_2\|
\end{aligned}$$

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 $\{z \mapsto \mathbb{E}[\psi_\tau(l(\varepsilon, z; \beta) + r_n \delta(z))] : \beta \in B\}$ 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 (S.4) and conditions on $|\dot{g}(z; \beta) - \dot{g}(z; \beta_0)|$.

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

$$\begin{aligned}
0 & \leq \lim_{t \rightarrow 0} t^{-1} [\Gamma_n(\hat{\beta} + t\gamma) - \Gamma_n(\hat{\beta})] & (S.5) \\
& = - \sum_{\{Y_i \neq g(Z_i; \hat{\beta})\}} \psi_\tau(Y_i - g(Z_i; \hat{\beta})) \gamma' \dot{g}(Z_i; \hat{\beta}) \\
& \quad + \lim_{t \rightarrow 0} \sum_{\{Y_i = g(Z_i; \hat{\beta})\}} t^{-1} \rho_\tau(g(Z_i; \hat{\beta}) - g(Z_i; \hat{\beta} + t\gamma)) \\
& = - \sum_{\{Y_i \neq g(Z_i; \hat{\beta})\}} \psi_\tau(Y_i - g(Z_i; \hat{\beta})) \gamma' \dot{g}(Z_i; \hat{\beta}) \\
& \quad - \sum_{\{Y_i = g(Z_i; \hat{\beta})\}} \psi_\tau(-\gamma' \dot{g}(Z_i; \hat{\beta})) \gamma' \dot{g}(Z_i; \hat{\beta}) \\
& = -D_{1n}(\hat{\beta}) - D_{2n}(\hat{\beta}).
\end{aligned}$$

By Assumption 2.2, $|D_{2n}(\hat{\beta})|$ is bounded by $\sum_{\{Y_i = g(Z_i; \hat{\beta})\}} A(Z_i)$. As, for any x , the error term u has a continuous law given $Z = z$, the number of observations with $Y_i = g(Z_i; \hat{\beta})$ is bounded in probability as the sample size tends to infinity. On the other hand, the moment

condition on $A(\cdot)$ implies that $\max_{1 \leq i \leq n} A(Z_i) = o_{\mathbb{P}}(n^{1/2})$. As γ is an arbitrary direction, it follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\tau} \left(Y_i - g(Z_i; \hat{\beta}) \right) \dot{g}(Z_i; \hat{\beta}) = o_{\mathbb{P}}(1). \quad (\text{S.6})$$

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

$$\begin{aligned} \nu_n(\beta_0) &= \nu_n(\hat{\beta}) + o_{\mathbb{P}}(1) && [\text{by (S.3)}] \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left[\psi_{\tau} \left(Y_i - g(Z_i; \hat{\beta}) \right) \mid Z_i \right] \dot{g}(Z_i; \hat{\beta}) + o_{\mathbb{P}}(1) && [\text{by (S.6)}] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[F_{\varepsilon} \left(g(Z_i; \hat{\beta}) - g(Z_i; \beta_0) - r_n \delta(Z_i) \mid Z_i \right) - \tau \right] \dot{g}(Z_i; \hat{\beta}) + o_{\mathbb{P}}(1) \\ &= \left\{ \frac{1}{n} \sum_{i=1}^n f_{\varepsilon}(0 | Z_i) \dot{g}(Z_i; \beta_0) \dot{g}'(Z_i; \beta_0) \right\} \sqrt{n} (\hat{\beta} - \beta_0) \\ &\quad - r_n \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n f_{\varepsilon}(0 | Z_i) \delta(Z_i) \dot{g}(Z_i; \beta_0) \right\} \\ &\quad + r_n^2 \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n f'_{\varepsilon}(0 | Z_i) \delta^2(Z_i) \dot{g}(Z_i; \beta_0) \right\} \\ &\quad + o_{\mathbb{P}} \left(\sqrt{n} \|\hat{\beta} - \beta_0\| \right) + o_{\mathbb{P}} \left(r_n^2 \sqrt{n} \right), \end{aligned}$$

where the last equality is based on a local expansions of $F_{\varepsilon}(\cdot | z)$ 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 $f_{\varepsilon}(0 | Z) \delta(Z) \dot{g}(Z; \beta_0)$ has zero mean, we obtain

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

from which the result follows. ■

Lemma 4.3. *Let k be a positive integer and \mathcal{G} a degenerate class of real-valued functions on $\mathbb{R}^{1+q} \times \dots \times \mathbb{R}^{1+q}$. Suppose \mathcal{G} is Euclidean(c,d) for a squared integrable envelope and some $c, d > 0$. Fix $z_1, \dots, z_n \in \mathbb{R}^q$ and let $u_1, \dots, u_n, u_{n+1}, \dots, u_{2n}$ be independent copies of the random variable u . For $i = 1, \dots, n$, let $v_i = (u_i, z_i)$ and $v_{n+i} = (u_{n+i}, z_i)$. Define $g_{\mathbf{i}_k}^n(u_{i_1}, \dots, u_{i_k}) = g(v_{i_1}, \dots, v_{i_k})$ and define $g_{\mathbf{i}_k}^{2n}$ similarly. Suppose that for any k -tuple \mathbf{i}_k , 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

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

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, \dots, z_n) such that

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

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, \dots, z_n arbitrarily.

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

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

with $u_{i_j}^*$ equal to either u_{i_j} or u_{n+i_j} where i_j ranges over the set $\{1, \dots, n\}$, and r is the number of elements $u_{i_1}^*, \dots, u_{i_k}^*$ belonging to $\{u_{n+1}, \dots, u_{2n}\}$. Independently, take a sample $\sigma_1, \dots, \sigma_n$ 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

$$\mathbb{E} \Phi \left(\sup_{\mathcal{G}} \left| \sum_{\mathbf{i}_k^n} g_{\mathbf{i}_k^n}(u_{i_1}, \dots, u_{i_k}) \right| \right) \leq \mathbb{E} \Phi \left(\sup_{\mathcal{G}} \left| \sum_{\mathbf{i}_k^n} \sigma_{i_1} \dots \sigma_{i_k} \tilde{g}(\mathbf{i}_k^n) \right| \right). \quad (\text{S.7})$$

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

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

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

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

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

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

where $D(\epsilon, d_{U_{2n}^k}, \mathcal{G})$ are the *packing numbers* of the set \mathcal{G} with respect to the pseudometric $d_{U_{2n}^k}$, $\delta_n^k = \sup_{\mathcal{G}} \sqrt{U_{2n, z_1, \dots, z_n}^k(g^2)}$ 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) , $1 \leq i \neq j \leq 2n$. Finally, our result follows using the arguments of the Main Corollary of Sherman (1994). ■

References

- DE LA PEÑA, V. H. AND E. GINÉ (1999): *Decoupling. From dependence to independence*, Springer-Verlag, New York.
- 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.
- VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak convergence and empirical processes*, Springer Series in Statistics, Springer-Verlag, New York.

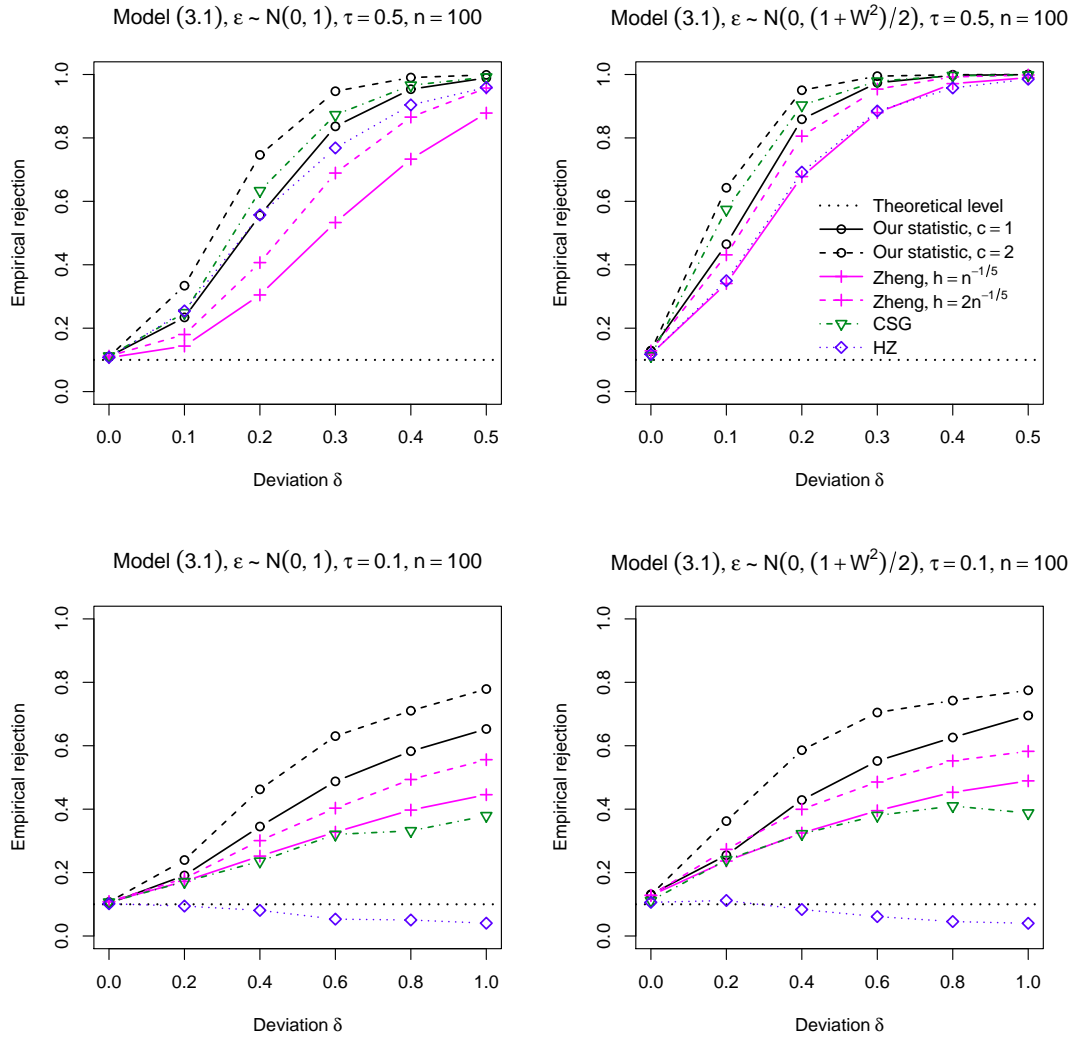


Figure 1: Power curves for model (3.1) with $h = cn^{-1/5}$ for Zheng's statistic.