WORKING PAPERS

N° 1102

June 2021

"Bootstrapping Quasi Likelihood Ratio Tests under Misspecification"

Pascal Lavergne and Patrice Bertail

Toulouse School of Economics

Bootstrapping Quasi Likelihood Ratio Tests under Misspecification

Patrice Bertail*
MODAL'X, UPL, Université Paris-Nanterre
and
Pascal Lavergne†
Toulouse School of Economics, Université Toulouse Capitole

May 25, 2021

Abstract

We consider quasi likelihood ratio (QLR) tests for restrictions on parameters under potential model misspecification. For convex M-estimation, including quantile regression, we propose a general and simple nonparametric bootstrap procedure that yields asymptotically valid critical values. The method modifies the bootstrap objective function to mimic what happens under the null hypothesis. When testing for an univariate restriction, we show how the test statistic can be made asymptotically pivotal. Our bootstrap can then provide asymptotic refinements as illustrated for a linear regression model. A Monte-Carlo study and an empirical application illustrate that double bootstrap of the QLR test controls level well and is powerful.

Keywords: Hypothesis Testing, Asymptotic Refinements.

---

*This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023- 01).

†Pascal Lavergne acknowledges support from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

Address correspondence: Pascal Lavergne, Toulouse School of Economics, Université Toulouse Capitole, 1 Esplanade de l'Université, 31000 Toulouse FRANCE. Email: pascal.lavergne@ut-capitole.fr

1 Introduction

Our main goal is to develop a consistent bootstrap procedure for quasi likelihood ratio tests with general convex M-estimation under potential model misspecification. As a leading example, consider the linear regression model

$$y_i = x_i'\theta + \varepsilon_i \quad \mathbb{E}\left(\varepsilon_i|x_i\right) = 0 \quad \mathrm{Var}(\varepsilon_i|x_i) = \sigma^2 \qquad i = 1, ..., n$$

where $z_i = (y_i, x_i), i = 1, ..., n$ are independent and identically distributed. Under these assumptions, the natural estimator is least-squares. Tests of restrictions on parameters can be entertained by a standard Wald test. Alternatively, one can built a test on the difference between the least-squares criterion with and without the restrictions. Because this is reminiscent of a likelihood-ratio test, we label such a test a Quasi Likelihood Ratio (QLR) test.[1]

The main advantages of the QLR test are (i) it avoids estimation of the asymptotic covariance matrix of estimators, by contrast to Wald and score tests, (ii) it is transformation equivariant. However, under misspecification, i.e. outside model assumptions, it is generally not asymptotically pivotal due to the potential failure of (the analog of) Bartlett's second identity. In our linear regression example, misspecification can occur because (i) errors are heteroscedastic, and the conditional variance function $\mathrm{Var}\left(\varepsilon|x\right)$ is of unknown form, or (ii) the conditional expectation $\mathbb{E}\left(y|x\right)$ is not linear in x. The study of potentially misspecified models dates back at least to Huber (1967) and Eicker (1967). For conditional ML, Gourieroux, Monfort, and Trognon (1984) have shown that the parameters of the

[1]The terminology does not seem to be settled in the literature. Some authors prefer "distance metric test," see Newey and McFadden (1994), others simply use "likelihood ratio test," see Spokoiny and Zhilova (2015).

conditional mean can be estimated consistently whenever it is well specified even if the likelihood itself is not. In particular, as is well known, ordinary least-squares consistently estimate mean parameters in the presence of heteroscedasticity of unknown form. If the true conditional expectation is nonlinear, the estimator provides the best linear approximation to the true conditional expectation that can be of interest on its own, see White (1980). Building on Eicker (1967), White (1980, 1982) and Royall (1986) have shown how Wald and score statistics can be rendered robust to misspecification. The behavior of likelihood-ratio tests under misspecification is considered by Foutz and Srivastava (1977) and Kent (1982) for unconditional Maximum Likelihood (ML), and Vuong (1989) for conditional ML. Marcellino and Rossi (2008) studied QLR tests in M-estimation.

In this work, we adress two questions: (i) can we design a general consistent bootstrap method for QLR tests under misspecification? (ii) can we obtain second-order correctness? Question (i) has received only a negative answer up to now. Indeed, the issue is intricate because of the non pivotalness of the statistic and the fact that using model assumptions to resample observations cannot replicate the statistic's behavior under misspecification. Recent work on this topic, see below, considers a bootstrap QLR statistic that shifts the original null hypothesis to one that is fulfilled by the data. E.g. when testing $\theta = 0$, the bootstrap statistic is designed to test $\theta = \widehat{\theta}$, where the latter is the original estimator. As the QLR statistic is not pivotal, the bootstrap statistic for the modified null hypothesis is not estimating correctly the distribution of the original statistic under the null if the null hypothesis does not hold. Our simulations shows that in moderate samples this may create large size distorsions even when the nul hypothesis holds. This issue is related to the first guideline for bootstrap hypothesis testing put forward by Hall and Wilson (1991) that "care should be taken to ensure that even if the data might be drawn from a population that fails to satisfy H_0, resampling is done in a way that reflects H_0." Therefore, to obtain bootstrap consistency under misspecification, we use the naive bootstrap and a statistic designed to test the original null hypothesis, but we modify the criterion to be optimized to make the bootstrap statistic behave as if the original null hypothesis H_0 were true. We label our method *bootstrapping QLR under the null hypothesis.* Our method allows to follow the first guideline of Hall and Wilson (1991) without relying on model assumptions, contrary the intuition that "estimation of the appropriate null distribution cannot be done fully nonparametrically, but requires some assumptions about the structure of the underlying family of models," see Tibshirani (1992).

To Question (ii), we provide a partial positive answer. We point out that when testing for a single restriction, a robust QLR statistic, which is asymptotically chi-square under misspecification, easily obtains. This was already noted by Stafford (1996) but seems to have been overlooked. We show that our bootstrap method applied to the robust QLR yields asymptotic second-order correctness in the linear regression model.

We consider as another leading example linear quantile regression, which is particularly interesting because estimation of the asymptotic covariance matrix of estimators must rely on nonparametric density estimation and a QLR test avoids such estimation. Our simulation study consider both mean and quantile regressions and illustrates that imposing the original null hypothesis in the bootstrap world results in a good size control under misspecification. Our study also compare the small sample performance of two competing procedures when testing a single restriction, namely bootstrapping the robust QLR statistic or double bootstrapping the non-robust QLR statistic, see Beran (1988). Our findings suggest that the first procedure may be preferable in moderate samples, because it avoids estimation of Hessian and score variance matrices. This appears in line with the conclusions of Stafford (1996), who states "the use of a model-robust variance estimate for the signed square root, score or Wald statistic, while leaving bias and skewness characteristics relatively unchanged, can increase variability considerably."

We now review the literature on robust bootstrap procedures. Contrary to QLR statistics, Wald and score statistics can be rendered asymptotically pivotal under misspecification. However their small sample behavior are sensitive to the implementation details and the design of the data, see Mackinnon and White (1985) and Chesher and Austin (1991) among others. Bootstrapping robust tests under the null is not often entertained, but can be dealt with a model-dependent method. In linear regression, resampling residuals (possibly obtained under the null) is valid under model assumptions but fail under heteroscedasticity. Wild residual bootstrap has been proposed to deal with heteroscedasticity of unknown form, see Wu (1986), Liu (1988), and Mammen (1992). But Kline and Santos (2012) pointed out that it does not do better than the asymptotic approximation if the conditional expectation is not linear, as linearity of the regression function is imposed when generating bootstrap samples.

Weighted bootstrap methods have also been investigated. One method introduces weights in the function to be optimized for estimation. Its consistency is studied by Bose and Chatterjee (2001), Jin, Ying, and Wei (2001), and Bose and Chatterjee (2003) for general convex M-estimation. Another method introduces weights in the estimating equations, see Chatterjee and Bose (2005). Few work establishes second-order correctness under heteroscedasticity. Bertail and Barbe (1995) provide asymptotic refinements for a general third-order differentiable functional, such as ones of interest in linear regression, however they do not consider the behavior of estimators under misspecification and their results do not apply to quantiles. Das and Lahiri (2019) show second-order correctness of weighted bootstrap M-estimator for the robust t-test in the heteroscedastic linear regression with a non random design, but do not consider mean misspecification. To date, only the naive nonparametric bootstrap of Efron (1979) has been shown to be second-order correct under general misspecification when applied to the robust t-test, see Hall and Horowitz (1996). Our results go in the same direction, provided that we bootstrap QLR under the null. We further consider general QLR tests for general convex non necessarily differentiable criteria under a large variety of misspecifications.

We now review work on bootstrapping QLR tests. Camponovo (2016) establishes higher-order improvements of a block bootstrap test in a dependent data context, when shifting the null hypothesis to one which holds in the bootstrap world. However he assumes the equivalent of second Bartlett's identity (up to an estimable scalar quantity, see his Assumption 3.2), which typically does not hold under misspecification. Spokoiny and Zhilova (2015) study the multiplier bootstrap for a shifted null hypothesis, allowing for a large number of parameters, and show that it may not be valid for misspecified models. Chen and Pouzo (2009) apply a similar method in a semiparametric context. For quantile regression, Angrist, Chernozhukov, and Fernández-Val (2006) rely on subsampling. Lee and Yang (2020) shows consistency of the m out of n bootstraps of QLR tests from M-estimation.

In Section 2, we present our bootstrap method and we detail implementation for linear regression and linear quantile regression. In Section 3, we first revisit previous results on the QLR statistic under model misspecification for a convex but potentially non-differentiable M-estimation criterion. We also show that when testing for a single restriction, a robust QLR statistic, which is asymptotically chi-square under misspecification, easily obtains. We then show consistency of our bootstrapped QLR test and we establish asymptotic higher-order refinements of the bootstrap robust QLR test in the linear regression model.

Section 4 gathers small sample evidence on the behavior of our method compared to existing methods. We then report some empirical results for quantile regression of children birthweight. Section 5 concludes. Section 6 gathers our technical proofs. The Appendix gathers details about our two main examples.

2 Bootstrapping QLR Under the Null

Let $z \in \mathbb{R}^k$ be a random vector with probability distribution P. Consider $q(z, \theta)$ a convex criterion function from $\mathbb{R}^k \times \mathbb{R}^p$ to $\mathbb{R}$ and assume $Q(\theta) = \mathbb{E}\, q(z, \theta)$ admits a unique minimizer $\theta_\dagger$, where $\mathbb{E}$ denotes expectation with respect to P. We consider a general M-estimation setup where, based on a random sample $\{z_i, i = 1, \ldots n\}$ from z, the M-estimator of $\theta_\dagger$ is

$$\widehat{\theta}_n = \arg\min_{\Theta} Q_n(\theta)\,, \qquad Q_n(\theta) = n^{-1} \sum_{i=1}^{n} q(z_i, \theta)\,.$$

The choice of the particular function $q(z, \theta)$ is based on the considered model: it can be a log-density for a parametric model, the squared residual for a mean regression model, or the check function applied to the residual for a quantile regression model, see below. Hence our framework encompasses popular models such as logistic or Poisson regression, generalized linear models, exponential hazard rate regression, and Cox semiparametric regression model. While the assumed model may be misspecified, the M-estimator converges under standard assumptions to

$$\theta_\dagger = \arg\min_{\Theta} Q(\theta)\,, \qquad Q(\theta) = \mathbb{E}\, q(z, \theta)\,. \tag{2.1}$$

We consider testing linear restrictions on the parameter of the form

$$H_0 : \; d' \left(\theta_\dagger - h\right) = \mathbf{0}\,,$$

where d is a known $p \times r$ full rank matrix, $r \leq p$, and h is a known vector in $\mathbb{R}^p$. Depending on what is more convenient, we can also consider the equivalent formulation

$$H_0 : \; \theta_\dagger = H\gamma_\dagger + h \,,$$

where H is a $p \times (p-r)$ full rank matrix such that $d'H = \mathbf{0}$. We study tests based on the QLR statistic

$$\mathrm{QLR}_n = 2\,n\,\left[Q_n(\widehat{\theta}_n^0) - Q_n(\widehat{\theta}_n)\right]\,,$$

where we define the estimator under the null hypothesis as $\widehat{\theta}_n^0 = \arg\min_{\Theta \cap H_0} Q_n(\theta)$ with $\Theta \cap H_0 = \{\theta \in \Theta,\; d'\,(\theta - h) = 0\}$.

We assume that our convex criterion is well approximated by a quadratic function, that is, there exists some score function $D(z_i, \theta)$ and a matrix $A(\theta)$ such that we have, for any t in a neighborhood of $\mathbf{0}$

$$Q_n(\theta_\dagger + t) - Q_n(\theta_\dagger) = n^{-1} \sum_{i=1}^{n} D'(z_i, \theta)t + \frac{1}{2}t'A(\theta)t + R_n(t)\,, \tag{2.2}$$

where $R_n(t)$ is a small reminder term, see our assumptions below. We also assume a similar approximation holds around $\theta_\dagger^0$. Let the empirical score function be the derivative of the quadratic approximation of $Q_n(\cdot)$ with respect to t,

$$S_n(\theta) = n^{-1} \sum_{i=1}^{n} D(z_i, \theta)\,.$$

If $Q_n(\theta)$ is differentiable in θ, this is simply its derivative. If $Q_n(\cdot)$ is not differentiable, $S_n(\theta)$ is a subgradient of $Q_n(\cdot)$ at θ. To bootstrap the QLR statistic, let $Q_n^*(\theta)$ be the criterion based on observations $\{z_i^*, i = 1, \ldots n\}$ resampled with replacement from the original data. We then consider the modified criterion

$$\widehat{Q}_n^*(\theta) = Q_n^*(\theta) - S_n'(\widehat{\theta}_n^0)(\theta - \widehat{\theta}_n^0)\,. \tag{2.3}$$

The bootstrap and constrained bootstrap estimator are defined as

$$\widehat{\theta}_n^* = \arg\min_{\Theta} \widehat{Q}_n^*(\theta) \quad \text{and} \quad \widehat{\theta}_n^{0*} = \arg\min_{\Theta \cap H_0} \widehat{Q}_n^*(\theta) \,.$$

The bootstrap QLR statistic is

$$\mathrm{QLR}_n^* = 2\,n\,\left[\widehat{Q}_n^*(\widehat{\theta}_n^{0*}) - \widehat{Q}_n^*(\widehat{\theta}_n^*)\right] = 2\,n\,\left[Q_n^*(\widehat{\theta}_n^{0*}) - Q_n^*(\widehat{\theta}_n^*) - S_n'(\widehat{\theta}_n^0)(\widehat{\theta}_n^{0*} - \widehat{\theta}_n^*)\right] \,.$$

We label our method *bootstrapping QLR under the null* because it makes the bootstrap estimator $\widehat{\theta}_n^*$ and QLR_n^* behave as if the restrictions were true. Indeed the first-order characterization of a convex function states that

$$Q(\theta) - \nabla_\theta' Q(\theta_\dagger^0)(\theta - \theta_\dagger^0) \geq Q(\theta_\dagger^0) \,.$$

Our modified empirical criterion takes a similar form so that the bootstrap criterion should be approximately minimized at $\widehat{\theta}_n^0$. Our method equates the derivatives of the quadratic approximation of $Q_n^*(\theta)$ at $\widehat{\theta}_n^*$ to $S_n(\widehat{\theta}_n^0)$. Hall and Horowitz (1996) and Andrews (2002) study related “recentering” methods for constructing Wald statistics. The formers use it to account for the non-nullity of the empirical moments in overidentified models estimated by Generalized Method of Moments, the latter because the average score evaluated using block-bootstrapped data can be different from zero. Here we use it instead to account for the fact that the score evaluated at the constrained estimator is not zero.

Example 1 : Least-Squares Regression. Consider the (pseudo) linear model

$$y_i = x_i'\theta + \varepsilon_i \qquad i = 1, ..., n$$

where $z_i = (y_i, x_i), i = 1, ..., n$ are independent and identically distributed from P. We allow for conditional heteroscedasticity in the error term ε and for correlation between ε

and x. The unique minimizer of $Q(\theta) = \mathbb{E}\,(y - x'\theta)^2/2$ is given by $\theta_\dagger = \mathbb{E}^{-1}(xx')\mathbb{E}\,(x'y)$. The corresponding least-squares estimator is $\widehat{\theta}_n = \mathbb{E}_n^{-1}(xx')\mathbb{E}_n(x'y)$, where $\mathbb{E}_n$ denotes expectation with respect to the empirical distribution of the data P_n. $\widehat{\theta}_n$ is simply the plug-in estimator of $\theta_\dagger$ obtained by replacing P by P_n. Similarly, the constrained minimizer of $Q(\theta)$ is

$$\theta_\dagger^0 = \theta_\dagger - \mathbb{E}^{-1}(xx')d\left(d'\mathbb{E}^{-1}(xx')d\right)^{-1}d'\left(\theta_\dagger - h\right) ,$$

and the constrained estimator is given by

$$\widehat{\theta}_n^0 = \widehat{\theta}_n - \mathbb{E}_n^{-1}(xx')d\left(d'\mathbb{E}_n^{-1}(xx')d\right)^{-1}d'\left(\widehat{\theta}_n - h\right) .$$

Moreover, as $Q_n(\theta)$ is quadratic in θ and since $\widehat{\theta}_n$ minimizes $Q_n(\theta)$, we have

$$\begin{aligned} QLR_n = 2\,n\left[Q_n(\widehat{\theta}_n^0) - Q_n(\widehat{\theta}_n)\right] &= n\left(\widehat{\theta}_n - \widehat{\theta}_n^0\right)\mathbb{E}_n(xx')\left(\widehat{\theta}_n - \widehat{\theta}_n^0\right)' \\ &= \left(d'\left(\widehat{\theta}_n - h\right)\right)'\left(d'\mathbb{E}_n^{-1}(xx')d\right)^{-1}d'\left(\widehat{\theta}_n - h\right) . \end{aligned}$$

It is thus equal to a Wald or Score statistic for testing H_0 up to the residual variance $\widehat{\sigma}^2$.

In the bootstrap world, the modified criterion makes $\widehat{\theta}_n^*$ converging to $\widehat{\theta}_n^0$ as

$$\widehat{\theta}_n^* = \mathbb{E}_n^{*^{-1}}(xx')\left(\mathbb{E}_n^*(xy) + S_n(\widehat{\theta}_n^0)\right) = \mathbb{E}_n^{*^{-1}}(xx')\mathbb{E}_n^*(xy) - \mathbb{E}_n^{*^{-1}}(xx')\mathbb{E}_n(xx')\left(\widehat{\theta}_n - \widehat{\theta}_n^0\right) ,$$

where $\mathbb{E}_n^*$ denotes expectation with respect to the bootstrap distribution P_n^*. The bootstrap QLR statistic is

$$\begin{aligned} QLR_n^* = 2\,n\left[\widehat{Q}_n^*(\widehat{\theta}_n^{0*}) - \widehat{Q}_n^*(\widehat{\theta}_n^*)\right] &= n\left(\widehat{\theta}_n^* - \widehat{\theta}_n^{0*}\right)\mathbb{E}_n^*(xx')\left(\widehat{\theta}_n^* - \widehat{\theta}_n^{0*}\right)' \\ &= \left(d'\left(\widehat{\theta}_n^* - h\right)\right)'\left(d'\mathbb{E}_n^{*^{-1}}(xx')d\right)^{-1}d'\left(\widehat{\theta}_n^* - h\right) . \end{aligned}$$

It is similar to a Wald statistic, but based on an estimator $\widehat{\theta}_n^*$ that behaves as if H_0 was true.

Example 2 : Quantile Regression. Our assumptions allow for a non-differentiable criterion. For quantile regression of order τ, see Koenker and Bassett (1978), $q(y, x, \theta) = \rho_\tau(y - x'\theta)$, with

$$\rho_\tau(u) = |u| \left[(1-\tau)\mathbb{I}(u < 0) + \tau \mathbb{I}(u > 0)\right] = u\left(\tau - \mathbb{I}(u < 0)\right).$$

For the median, $\tau = 1/2$ and $\rho_\tau(u) = \frac{|u|}{2}$. In practice, estimation is performed solving the dual problem

$$\max_t \left\{ y't | - X't = \mathbf{0}, t \in [\tau - 1, \tau]^n \right\},$$

where X is the matrix of observations on covariates, see Koenker (2005). At the optimum, $t_i(\widehat{\theta}_n) = \tau$ if $y_i - x_i'\widehat{\theta}_n$ is positive, $t_i(\widehat{\theta}_n) = \tau - 1$ if it is negative, and the remaining components are determined so that $-X't(\widehat{\theta}_n) = \mathbf{0}$. Theses value are, up to a constant, the rank scores of the quantile regression, see Gutenbrunner and Jureckova (1992). The dual of the restricted quantile regression writes

$$\max_t \left\{ (y - XH)'t | - H'X't = 0, t \in [\tau - 1, \tau]^n \right\},$$

where H is a $p \times (p - r)$ full rank matrix such that $d'H = \mathbf{0}$.

One issue with quantile regression is that the subgradient of the function $\rho_\tau(\cdot)$ can be arbitrarily defined at 0, yielding some indeterminacy. Asymptotically this should be irrelevant, however we have found that in practice, using the rank scores to define $S_n(\cdot)$ gives better empirical results. Namely, we used $S_n(\widehat{\theta}_n) = -X't(\widehat{\theta}_n)/n$ and $S_n(\widehat{\theta}_n^0) = -X't(\widehat{\theta}_n^0)/n$. These fulfill the usual properties encountered for a differentiable criterion, namely $S_n(\widehat{\theta}_n) = \mathbf{0}$ and $H'S_n(\widehat{\theta}_n^0) = \mathbf{0}$. That is, the empirical score is zero for components of the parameter space that are unconstrained. The modified optimization program $\min_\theta \sum_{i=1}^n \rho_\tau\left(y_i^* - x_i^{*\prime}\theta\right) - S_n'(\widehat{\theta}_n^0)\theta$ writes in dual form

$$\max_t \left\{ y^{*\prime} t | - X^{*\prime} t = n\, S_n(\widehat{\theta}_n^0), t \in [\tau - 1, \tau]^n \right\}.$$

Our bootstrap modified criterion can thus be easily computed in practice, even in this non-differentiable case. Generalized linear models may be treated similarly.

3 QLR Test under Misspecification

3.1 Asymptotic Test

We here state results under misspecification with i.i.d. data without imposing differentiability of the criterion. We also study the behavior of the test statistic under local alternatives. We consider the following assumptions.

Assumption A. *(a) $q(z,\theta)$ is convex in θ.*

(b) $Q(\theta) = \mathbb{E}\, q(z,\theta)$ is twice differentiable, the parameter space Θ is an open convex of $\mathbb{R}^p$, and $\theta_\dagger = \arg\min_{\Theta\cap H_0} Q(\theta)$ is unique.

(c) d is full rank and $\theta_\dagger^0 = \arg\min_{\Theta\cap H_0} Q(\theta)$ is unique.

Assumption B. *(a) There exists a vector function $D(z,\theta)$ in $\mathbb{R}^p$ with*

- $\mathbb{E}\, D(z,\theta_\dagger) = 0$.
- *$B(\theta) = \operatorname{Var} D(z,\theta)$ is uniformly bounded and positive definite for θ in a neighborhood of $\theta_\dagger$.*
- *Let $A(\theta) = \nabla_{\theta,\theta'} Q(\theta)$ and $R_n(\theta,t) = Q_n(\theta+t) - Q_n(\theta) - n^{-1}\sum_{i=1}^n D'(z_i,\theta)t$. For any θ in a neighborhood of $\theta_\dagger$, $A(\theta)$ is positive definite, and as $t \to \mathbf{0}$*

$$\mathbb{E}\, R_n(\theta,t) = \frac{1}{2}t'A(\theta)t + o(\|t\|^2)\,, \qquad n\ \operatorname{Var} R_n(\theta,t) = o(\|t\|^2)\,.$$

(b) $A(\theta)$ and $B(\theta)$ are continuous for θ in a neighborhood of $\theta_\dagger$.

We now comment on our assumptions. Assumption A ensures convexity and uniqueness of restricted and unrestricted optima. Convexity ensures that $Q_n(\cdot)$ does not have multiple optima such that $S_n(\theta) = o_p(1)$, which could invalidate our approach, and in particular make the bootstrap test not consistent. Similar phenomenon can yield inconsistency of the score test, see e.g. Freedman (2007). Convexity can be checked in many instances, including linear mean and quantile regression. Twice differentiability of $Q(\cdot)$ holds under standard conditions, for instance in quantile regression when the error term has a density. Condition (c) ensures that our restrictions are not redundant.

The quadratic approximation of $Q_n(\theta + t)$ follows from the following arguments: (i) $D(\cdot, \theta)$ is a (not necessarily unique) subgradient of $q(\cdot, \theta)$, (ii) $\mathbb{E}\, q(Z, \theta + t) - \mathbb{E}\, q(Z, \theta) - \mathbb{E}\, D(z, \theta)'t$ is, by a standard Taylor expansion $\frac{1}{2} t' A(\theta) t + o(\|t\|^2)$ as $t \to \mathbf{0}$, (iii) the remaining centered sample average based on the function is of a small order for $t \to \mathbf{0}$. Assumption B-(a) is true pointwise in t in a neighborhood of $\mathbf{0}$, and together with convexity will be used to obtain an uniform expansion in our proofs. Our supplementary appendix contains details on how our assumptions can be checked for the regression and quantile models.

Theorem 3.1. *For a random sample $\{z_i, i = 1, \ldots n\}$, under Assumptions A and B,*

(a) Under H_0, $\mathrm{QLR}_n \xrightarrow{d} \sum_{j=1}^{r} \lambda_j U_j^2$, where $U = (U_1, \ldots U_r)'$ is a vector of r independent standard normal variables and λ is the vector of r eigenvalues of

$$\left(d' A_\dagger^{-1} d\right)^{-1} d' A_\dagger^{-1} B_\dagger A_\dagger^{-1} d\,, \qquad \text{where } A_\dagger = A\left(\theta_\dagger\right) \text{ and } B_\dagger = B\left(\theta_\dagger\right)\,. \tag{3.4}$$

(b) Under H_A, $\mathrm{QLR}_n / n \xrightarrow{p} c > 0$.

(c) Under $H_{A,n}$: $\theta_{\dagger,n} = H\gamma_\dagger + h + c/\sqrt{n}$ with $d'c \neq \mathbf{0}$, $\mathrm{QLR}_n \xrightarrow{d} \sum_{j=1}^{r} \lambda_j (U_j + \tilde{c}_j)^2$, where λ is the vector of eigenvalues of (3.4) with $\theta_\dagger = H\gamma_\dagger + h$ and at least one $\tilde{c}_j$ is not zero.

When the second Bartlett's identity holds, that is $B_\dagger = A_\dagger$, λ is a vector of ones and the QLR statistic asymptotically follows a centered chi-square distribution with r degrees of freedom. In a correctly specified linear regression, $B_\dagger = \sigma^2 A_\dagger$ for homoscedastic errors with variance σ^2, and we can easily render QLR asymptotically pivotal. In general however, the QLR statistic has a more involved asymptotic distribution under H_0. Our characterization of λ as the eigenvalues of the matrix (3.4) appears to be new, see Lien and Vuong (1987, Lemma 2) for an alternative one in the linear regression model.

3.2 Robust QLR Test

We here focus on an univariate restriction for which one can generally build a robust QLR statistic since λ is scalar. This is of particular empirical relevance for testing the significance of a specific parameter component. Given a consistent estimator $\widehat{\lambda}_n$, the robust QLR statistic is

$$\mathrm{RQLR}_n = \frac{\mathrm{QLR}_n}{\widehat{\lambda}_n} \xrightarrow{d} \chi_1^2 \qquad \text{under } H_0\,,$$

and a test can be entertained using standard critical values. Estimating λ requires consistent estimation of $A_\dagger$ and $B_\dagger$. One can use empirical analogs of the above matrices evaluated at a consistent estimator of $\theta_\dagger$. The quantity $A_\dagger^{-1} B_\dagger A_\dagger^{-1}$ happens to be the asymptotic variance of $\widehat{\theta}_n$, which should be estimated to build the robust Wald and Score statistics. Hence it is as easy (or as difficult) to build the robust QLR statistic than it is to obtain robust Wald or Score statistics.

For testing a single restriction $d'\left(\theta_\dagger - h\right) = 0$ in linear regression, an estimator of λ is easily obtained as

$$\widehat{\lambda}_n = \frac{d' \mathbb{E}_n^{-1}(xx') \Sigma_n \mathbb{E}_n^{-1}(xx') d}{d' \mathbb{E}_n^{-1}(xx') d} \quad \text{with} \quad \Sigma_n = \mathbb{E}_n(xx'(y - x'\widehat{\theta}_n)^2) = n^{-1} \sum_{i=1}^n x_i x_i' (y_i - x_i' \widehat{\theta}_n)^2\,,$$

see Eicker (1967) and White (1982). The numerator is the heteroscedasticity-robust variance estimator provided by most statistical software.

For quantile regression, $A_\dagger = \mathbb{E}\left(f_\varepsilon(0|x)xx'\right)$ depends on the conditional density of the error term $\varepsilon = y - x'\theta_\dagger$ at 0. Powell (1991) proposes to use the estimator

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y_i - x_i'\widehat{\theta}_n}{h}\right) x_i x_i' \tag{3.5}$$

where $K(\cdot)$ is a density and h a bandwidth. Also $B_\dagger = \mathbb{E}\left((\tau - \mathbb{I}(\varepsilon \leq 0))^2 xx'\right)$ can be estimated by the Eicker-type estimator

$$\frac{1}{n}\sum_{i=1}^{n}\left(\tau - \mathbb{I}(y_i - x_i'\widehat{\theta}_n \leq 0)\right)^2 x_i x_i' ,$$

see Kim and White (2003).

3.3 Bootstrap Test

For our bootstrap test, we need to make two supplementary assumptions. First we need a linear approximation of $S_n(\cdot)$ around $\theta_\dagger^0$. Second, our quadratic approximation should hold around $\theta_\dagger^0$ when the original observations are replaced by z_i^*, which are i.i.d from P_n.

We say that $H_n(t) = o_p(1/\sqrt{n})$ uniformly over $o_p(1)$ neighborhoods of t to mean that for each sequence of random variables u_n of order $o_p(1)$ there exists a sequence of random variables b_n of order $o_p(1)$ such that $\sup_{\|t\|\leq u_n} |H_n(t)| \leq b_n/\sqrt{n}$.

Assumption C. $R_{1n}(t) = n^{-1}\sum_{i=1}^{n} D(z_i, \theta_\dagger^0 + t) - n^{-1}\sum_{i=1}^{n} D(z_i, \theta_\dagger^0)$ *is such that*

$$\mathbb{E}\, R_{1n}(t) = A(\theta_\dagger^0)t + o(\|t\|) \quad \textit{as } t \to \mathbf{0}\,,$$

$$R_{1n}(t) - \mathbb{E}\, R_{1n}(t) = o_p(n^{-1/2}) \qquad \textit{uniformly in } t \textit{ over } o_p(1) \textit{ neighborhoods of } \mathbf{0}\,.$$

Moreover $A(\theta)$ *is positive definite in a neighborhood of* $\theta_\dagger^0$.

Assumption D. *Let $E_{|\mathcal{Z}}$ and $\mathrm{Var}_{|\mathcal{Z}}$ be expectation and variance conditional upon the original sample and $R_n^*(\theta,t) = Q_n^*(\theta+t) - Q_n^*(\theta) - n^{-1}\sum_{i=1}^n D'(z_i^*,\theta)t$. For any θ in a neighborhood of $\theta_\dagger^0$ and as $t \to \mathbf{0}$*

$$\mathbb{E}_{|\mathcal{Z}} R_n^*(\theta,t) = \frac{1}{2}t'A(\theta)t + o_p(\|t\|^2)\,, \qquad n\,\mathrm{Var}_{|\mathcal{Z}}\, R_n^*(\theta,t) = o_p(\|t\|^2)\,.$$

Moreover, $A(\theta)$ and $B(\theta)$ are continuous for θ in a neighborhood of $\theta_\dagger^0$.

The linear approximation in Assumption C follows from the following arguments: (i) the reminder term from a Taylor expansion yields $\mathbb{E}\,D(Z,\theta_\dagger^0+t) - \mathbb{E}\,D(Z,\theta_\dagger^0) - A\left(\theta_\dagger^0\right)t = o(\|t\|)$, and (ii) what is left is a centered empirical process, which under mild conditions is an $o_p(n^{-1/2})$ uniformly in t in a $o_p(1)$ neighborhood of $\mathbf{0}$, see e.g. Sherman (1994). Assumption D may be easily checked by using conditional arguments similar to the ones used to check Assumption B, but replacing P by P_n. For instance, $A(\theta)$ will be replaced by a matrix $A_n(\theta)$ that converges in P-probability to $A(\theta)$.

Our bootstrap test rejects H_0 if $\mathrm{QLR}_n > q_{1-\alpha}^*$, where $q_{1-\alpha}^*$ is the $1-\alpha$ quantile of QLR_n^*. We first establish consistency of our bootstrap procedure.

Theorem 3.2. *For a random sample $\{z_i, i = 1,\dots n\}$, under Assumptions A, B, C, and D,*

(a) Under H_0, $\sup_x \left|\Pr\left(\mathrm{QLR}_n^* \le x \,\middle|\, (y_i,x_i)_{i=1,\dots,n}\right) - \Pr\left(\mathrm{QLR}_n \le x\right)\right| = o_p(1)$.

(b) Let $q_{1-\alpha}^*$ be the conditional quantile of order $(1-\alpha)$ of QLR_n^*. Then under H_A, $\Pr\left(\mathrm{QLR}_n > q_{1-\alpha}^*\right) \to 1$.

(c) Under $H_{A,n}$, $\lim_{n\to\infty}\Pr\left(\mathrm{QLR}_n > q_{1-\alpha}^*\right) > \alpha$.

For an univariate restriction, we can obtain an asymptotically pivotal statistic, so we can hope to obtain asymptotic refinements. The next result confirms this holds for linear regression, irrespective of whether the model is well specified.[2] To our knowledge, the only competing method that has been shown to yield asymptotic refinements under misspecification is the nonparametric bootstrap t-test, see Hall and Horowitz (1996). From our previous results,

$$\mathrm{RQLR}_n = \frac{\mathrm{QLR}_n}{\widehat{\lambda}_n} = n\left(\widehat{\theta}_n - h\right)' d\left(d'\mathbb{E}_n^{-1}(xx')\Sigma_n\mathbb{E}_n^{-1}(xx')d\right)^{-1} d'\left(\widehat{\theta}_n - h\right) ,$$

and similarly

$$\mathrm{RQLR}_n^* = \frac{\mathrm{QLR}_n^*}{\widehat{\lambda}_n^*} = \left(\widehat{\theta}_n^* - h\right)' d\left(d'\mathbb{E}_n^{*-1}(xx')\Sigma_n^*\mathbb{E}_n^{*-1}(xx)d\right)^{-1} d'\left(\widehat{\theta}_n^* - h\right) ,$$

where $\Sigma_n^* = \mathbb{E}_n^*(xx'(y - x'\widehat{\theta}_n^*)^2) = n^{-1}\sum_{i=1}^n x_i^* x_i^{*'}(y_i^* - x_i^{*'}\widehat{\theta}_n^*)^2$.

Theorem 3.3. *Under the assumptions of Theorem 3.2 and*

(a) the Cramer condition $\overline{\lim}_{t\to\infty} |\mathbb{E}\exp\left(i\,t\,d'\mathbb{E}^{-1}(xx')x\,(y - x'\theta_\dagger)\right)| < 1$,

(b) $\mathbb{E}\left(d'\mathbb{E}^{-1}(xx')xy\right)^{12} < \infty$ *and* $\mathbb{E}\left(d'\mathbb{E}^{-1}(xx')xx'\theta_\dagger\right)^{12} < \infty$,

$$\sup_x \left|\Pr\left(\mathrm{RQLR}_n^* \le x \,\middle|\, (y_i, x_i)_{i=1,\ldots,n}\right) - \Pr\left(\mathrm{RQLR}_n \le x\right)\right| = O_p(n^{-3/2}) \qquad \text{under } H_0 .$$

Moreover if $q_{R,1-\alpha}^*$ *is the* $1-\alpha$ *quantile of the bootstrap distribution of* RQLR_n^*,

$$\Pr\left(\mathrm{RQLR}_n \le q_{R,1-\alpha}^*\right) = 1 - \alpha + O(n^{-2}) \qquad \text{under } H_0 .$$

[2] We focus on least-squares regression, as the Edgeworth expansion for quantiles is typically non-standard due to the lattice nature of the subgradient, see e.g. Falk and Janas (1992).

The behavior of the bootstrap robust test under local and global alternatives is similar to the one of the QLR test studied in Theorem 3.2. The Cramer condition automatically holds if some components of x have an absolutely continuous part. It is minimal to ensure, that the influence function of the functional appearing in the quadratic form of QLR_n is non lattice. The condition could be replaced by the sufficient multivariate Cramer condition $\overline{\lim}_{||u||\to\infty}|\mathbb{E}\exp(iu'x(y-x'\theta_\dagger)| < 1$, which is however stronger. Moment conditions of order 12 ensure that the terms in the Edgeworth expansion of RQLR_n depending on moments of order up to 6 match the empirical ones in the bootstrap distribution up to $O_p(n^{-1/2})$ by a standard CLT argument. Such conditions are automatically implied by the existence of higher moments for both y and x. Our results are similar to Hall and Horowitz (1996). With Efron's bootstrap, we believe that the obtained rates cannot be improved as can be inferred from Hall and Mammen (1994).

4 Small Sample Evidence and Application

4.1 Simulations

We focused on inference on the coefficient β_2 in the assumed linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \, .$$

We generated the variable y according to the model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \psi x_{1i} x_{2i} + (1 + l|x_{2i}|)\,\eta_i \, ,$$

where $(\beta_1, \beta_2) = (0, 0)$. The variable x_1 is standard Gaussian, x_2 is independent lognormal with mean 0 and variance 1, and η is independently distributed as a Student law. This specification provides us with an asymmetric covariate and generates observations with high leverage, which can create serious obstacles to heteroscedasticity robust inference as shown by Chesher and Austin (1991). The parameter l controls heteroscedasticity. When $\psi \neq 0$, the linear conditional mean (or quantile) is misspecified. Due to the regressors' independence, β_2 is unchanged under misspecification of the conditional mean or median, which is convenient when studying the tests' behavior for $\psi \neq 0$. We considered values of $l = 0.5$ and $\psi = 0.5$, which corresponds to moderate misspecifications that could go unnoticed.

We ran 20000 simulations with $n = 200$. To speed up computations, we use the warp-speed method proposed by Davidson and MacKinnon (2007) and studied by Giacomini, Politis, and White (2013). Specifically, for each considered hypothesis, we drew one bootstrap and double bootstrap sample for each simulated data, and we used the whole set of bootstrap statistics to compute the bootstrap and double bootstrap p-values associated with each original statistic. We report our results using graphs. The first type of graph draws the errors in rejection probability (ERP), that is the difference between nominal size and the empirical rejection proportion under the null hypothesis $H_0 : \beta_2 = 0$. A perfect test would exhibit an ERP of zero for any nominal size. This gives us a visual way to evaluate whether the null distribution of the test statistic is well approximated by its asymptotic or bootstrap approximation. The second type of graph draws the power curves of each test, where power is evaluated using the same data generating process but testing the null hypothesis $\beta_2 = \beta_2^0$ and varying β_2^0.

Mean Regression We set $\eta \sim t_5$. The QLR statistic was bootstrapped under the null (denoted by QLR0-b) and we used double bootstrap to obtain improvements (denoted by

QLR0-db). We compared them to testing $\beta_2 = \widehat{\beta}_2$ in the bootstrap world with nonparametric resampling (QLR-b and QLR-db). We also considered the asymptotic test based on RQLR, the robust QLR statistic equal to the Wald one. We applied bootstrapping under the null (RQLR0-b) as well as the usual nonparametric bootstrap (RQLR-b), see Hall and Horowitz (1996). When using robust statistics, we estimated the correction factor λ by the HC3 method, as recommended by Long and Ervin (2000) and Cribari-Neto, Ferrari, and Oliveira (2005).

Figure 1 gathers our results for size control. The robust asymptotic test is always undersized at usual nominal sizes, and bootstrap very imperfectly corrects this phenomenon. The bootstrap test QLR-b is always oversized at usual nominal sizes and double bootstrap is moderately helpful. Double bootstrap under the null always yields the best size control. Figure 2 gathers our results in terms of power. Tests that are systematically oversized have a slightly better power, as could be expected, while other tests have similar performances.

Quantile Regression Since most existing bootstrap methods for inference in quantile regression assume a correct model, see Kocherginsky, He, and Mu (2005) for a review, we only compared five tests. The first two are based on robust asymptotic and bootstrap Wald tests, denoted as W and W-b. For standard errors, we used the formula detailed in Section 3.2, and specifically we chose $K(\cdot)$ as the standard normal density and $h = 0.79\, n^{-1/5}\,\mathrm{IQR}$ in (3.5). We also considered the percentile bootstrap, denoted as P-b, which was found to be the best performing method by Tarr (2012). We compared these to our bootstrap under the null QLR test, denoted as QLR0-b, and its double bootstrap version, QLR0-db. We did not look at bootstrap with naive resampling for testing $\beta_2 = \widehat{\beta}_2$ in the bootstrap world, given its poor performances in our previous experiments. We considered several

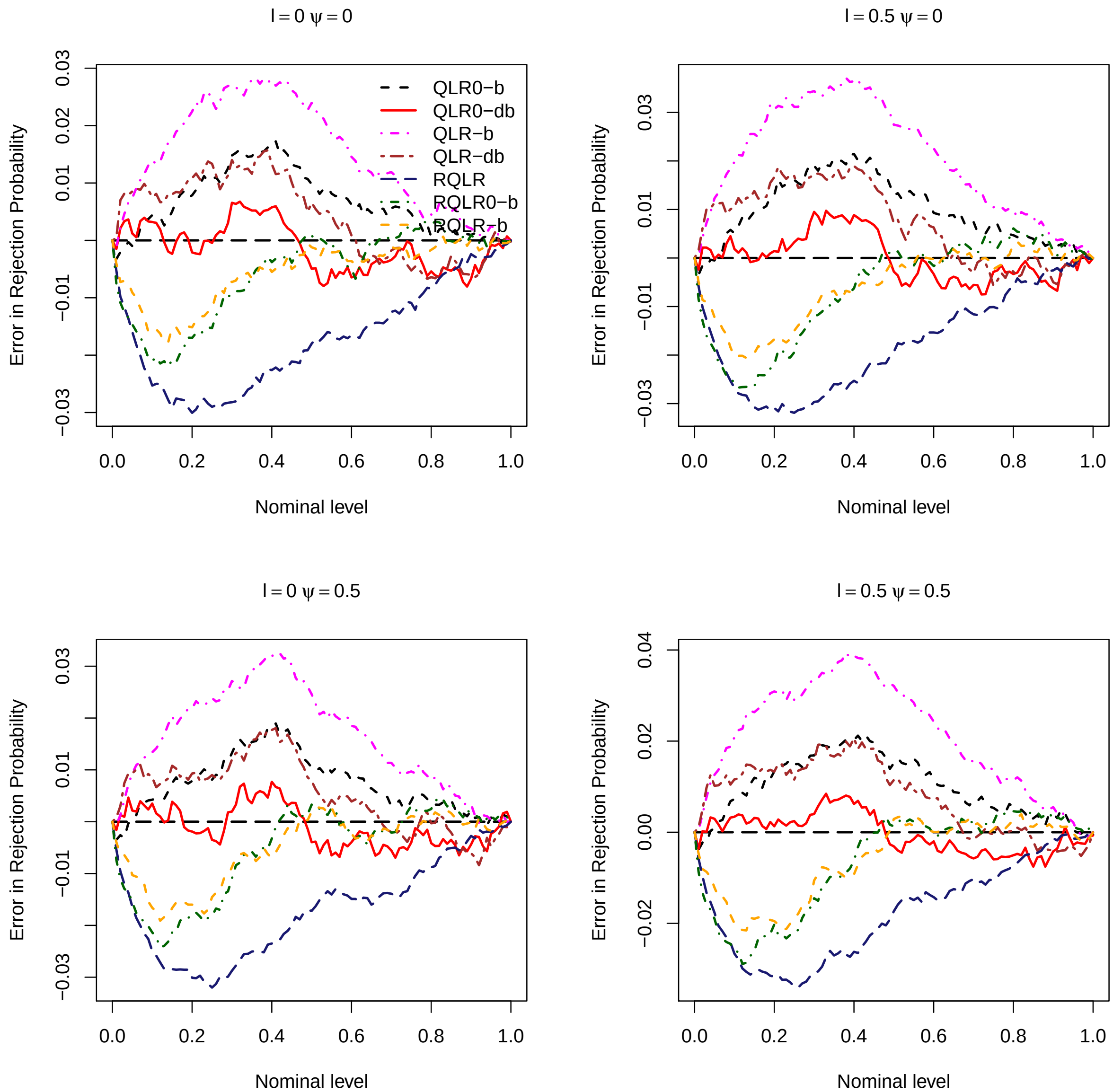

Figure 1: Errors in Rejection Probabilities for mean regression with $\eta \sim t_5$.

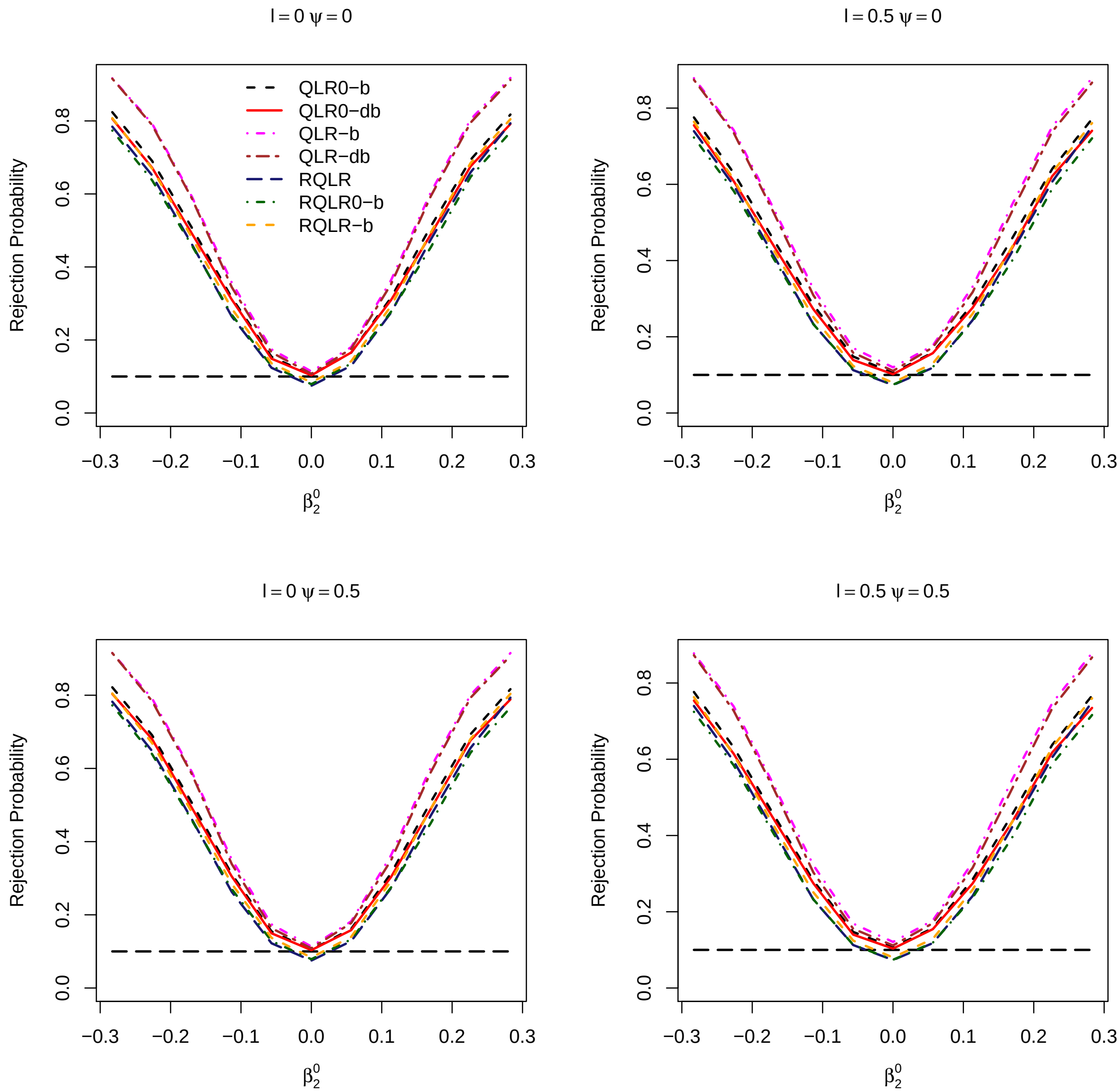

Figure 2: Power curves for mean regression with $\eta \sim t_5$.

setups corresponding to median regression and quantile regression of order $\tau = 0.25$, and η distributed as t_5 or t_1. We report a selection of our results.

Figures 3 and 5 gather our results for size control in median regression, while Figure 7 consider quantile regression of order 0.25. Under misspecification, the asymptotic robust Wald test does not perform well, and bootstrap is not very successful at correcting this phenomenon. The percentile bootstrap performs better, but double bootstrap under the null of the QLR test provides the best size control.

Figures 4, 6, and 7 reports power curves. Misspecification, and in particular heteroscedasticity, has very adverse effects on the tests' performances. Under misspecification, Wald tests can have an erratic behavior, leading to a sometimes non-monotone power. Besides the Wald test, QLR-0db has always highest power under misspecification.

4.2 Empirical Application

We considered some parametric quantile models for children birthweight using data analyzed by Abrevaya (2001) and Koenker (2005), who also gave a detailed data description. We focused on median regression and lower quantiles regression on a subsample of 1089 smoking college graduate white mothers. We considered a model linear in (i) weight gain during pregnancy, (ii) average number of cigarettes per day (denoted as CIGAR), and (iii) dummies indicating whether the child is male and whether the mother is married, and quadratic in mother's age as suggested by Koenker (2005). Maistre, Lavergne, and Patilea (2017) reported that there may be misspecification for lower quantiles, but did not find evidence of misspecification in median regression. In Table 1, we report 90% confidence intervals for the CIGAR parameter using the same methods as in our simulations. We considered 999 bootstrap samples for each potential value of the parameter.

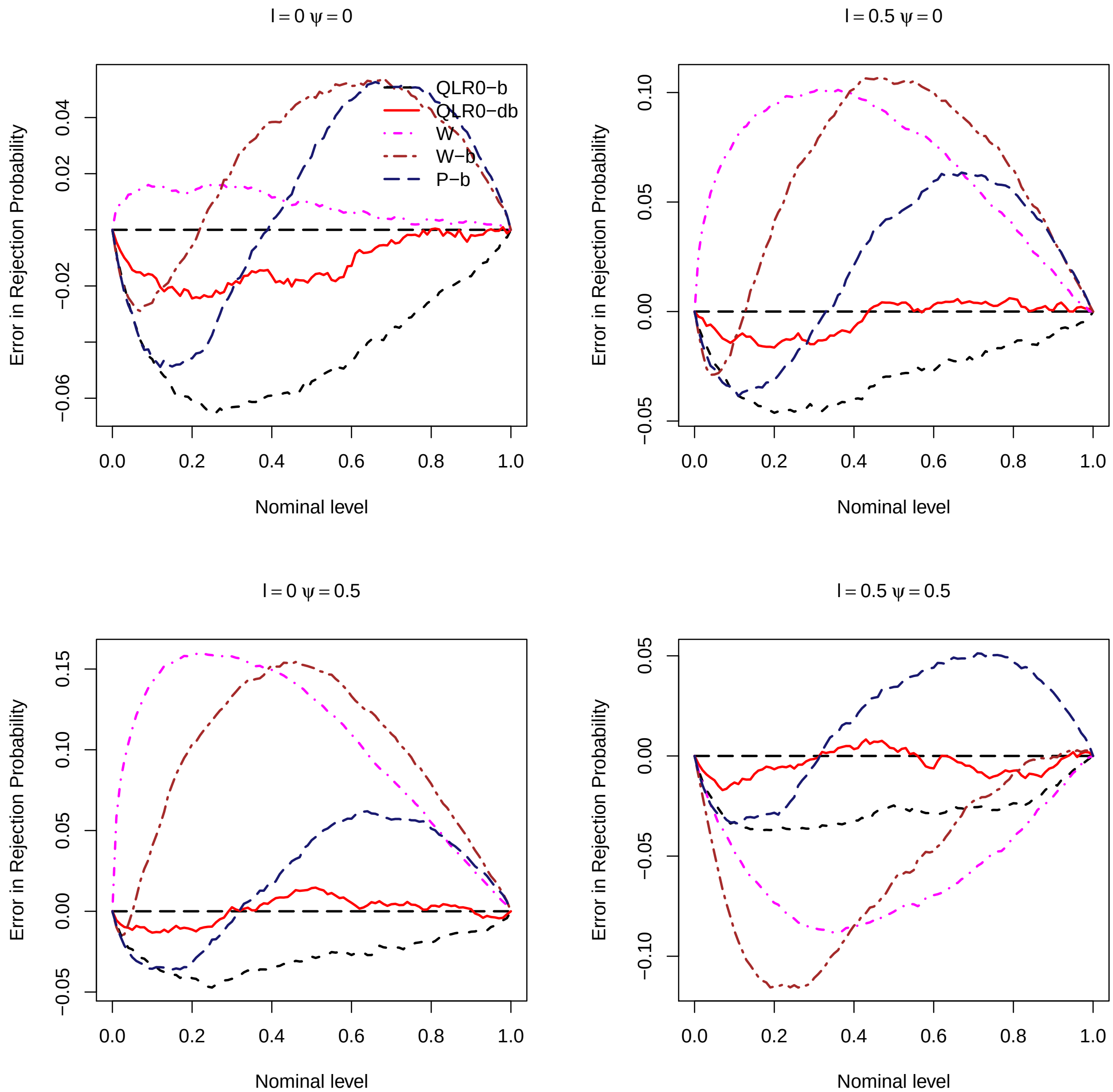

Figure 3: Errors in rejection probabilities for median regression with $\eta \sim t_5$.

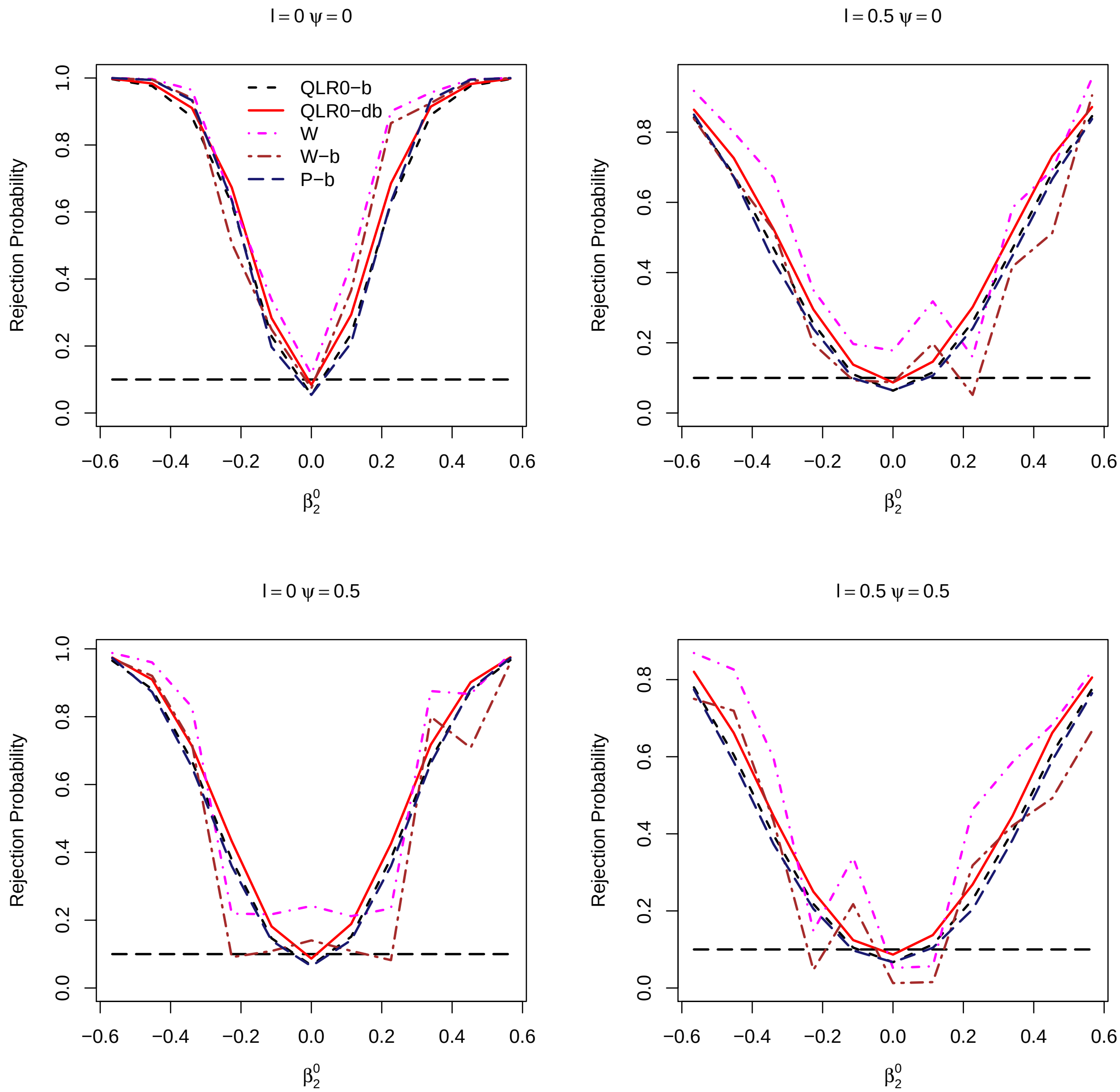

Figure 4: Power curves for median regression with $\eta \sim t_5$.

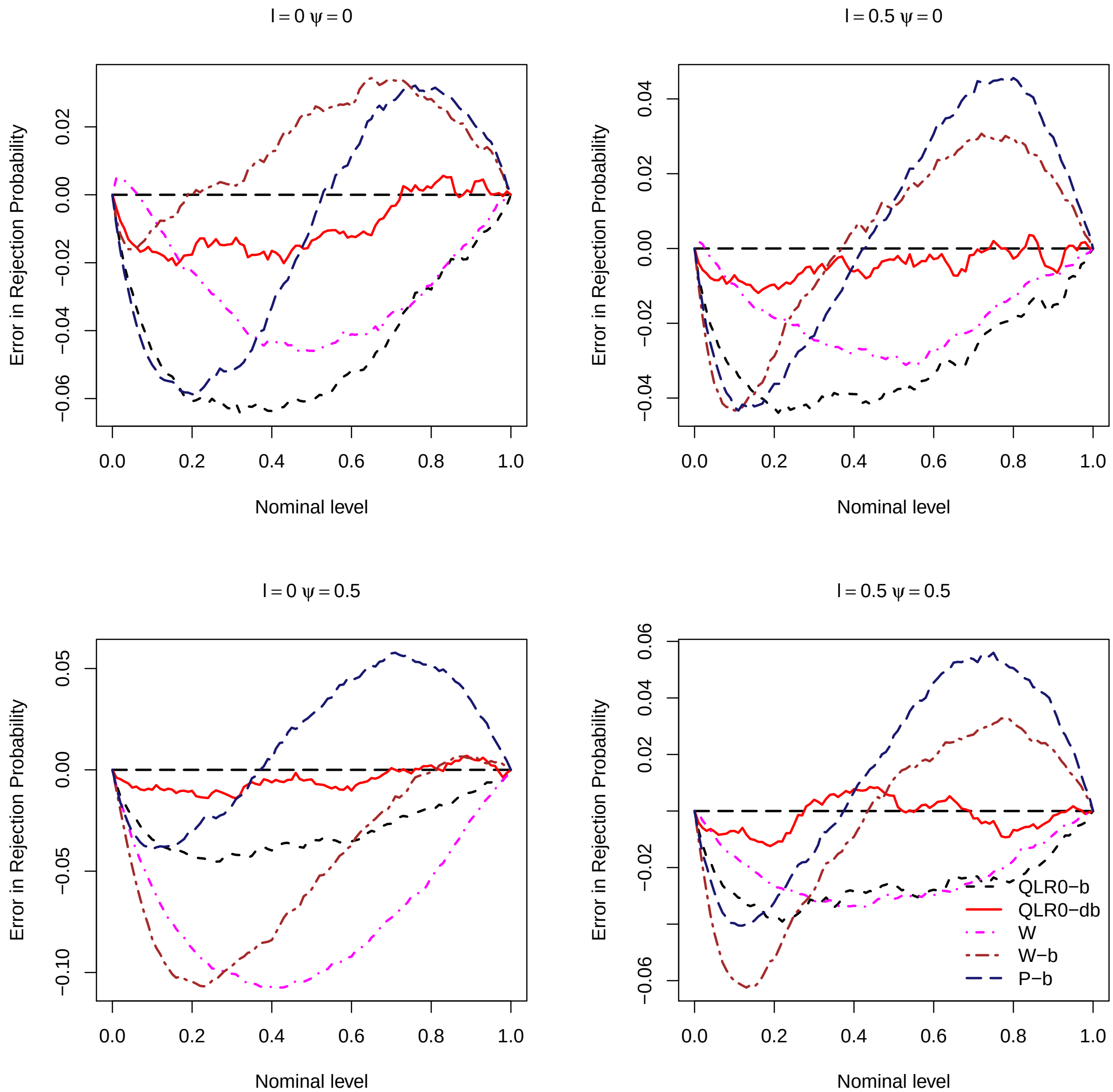

Figure 5: Errors in rejection probabilities for median regression with $\eta \sim t_1$.

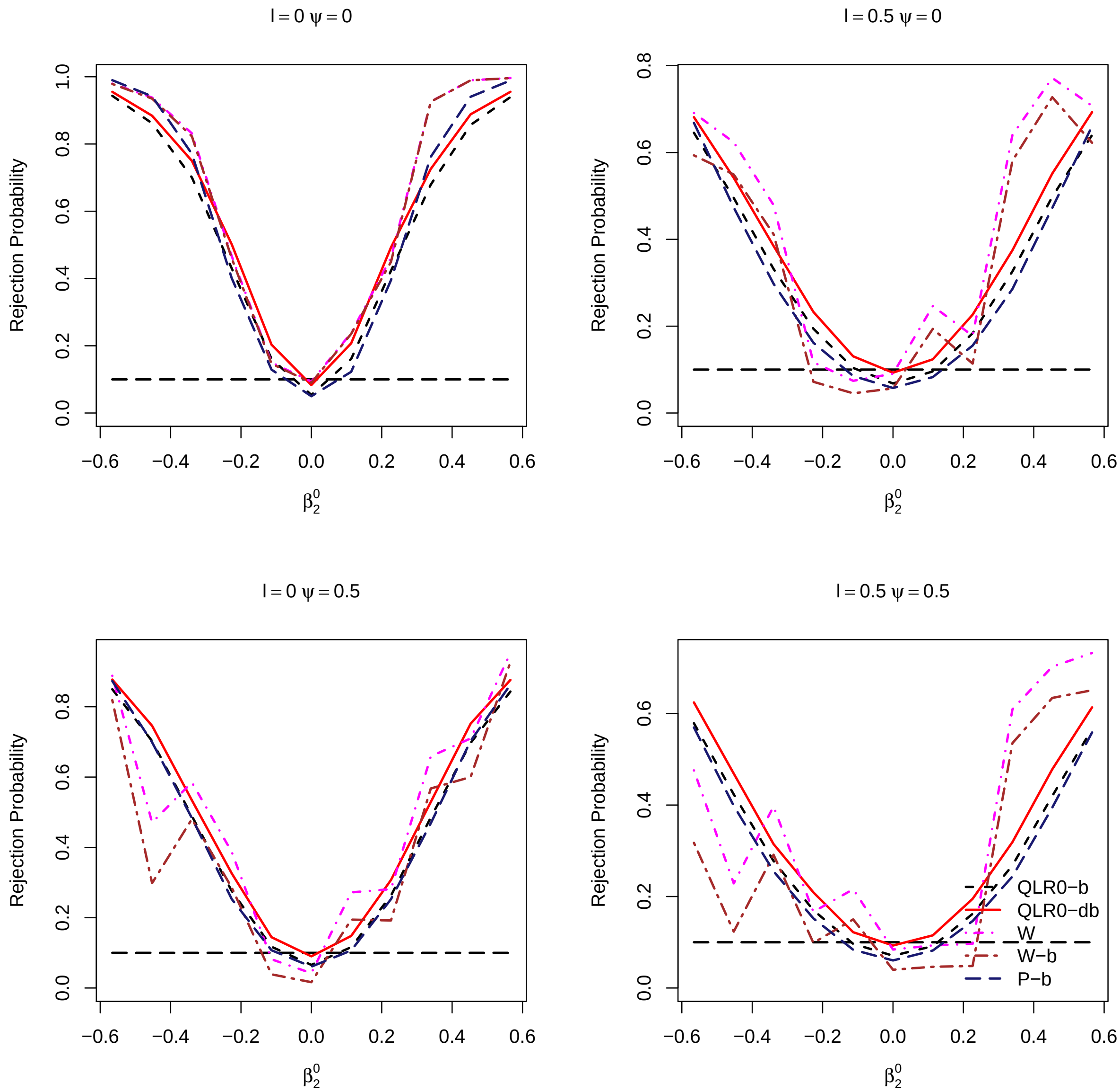

Figure 6: Power curves for median regression with $\eta \sim t_1$.

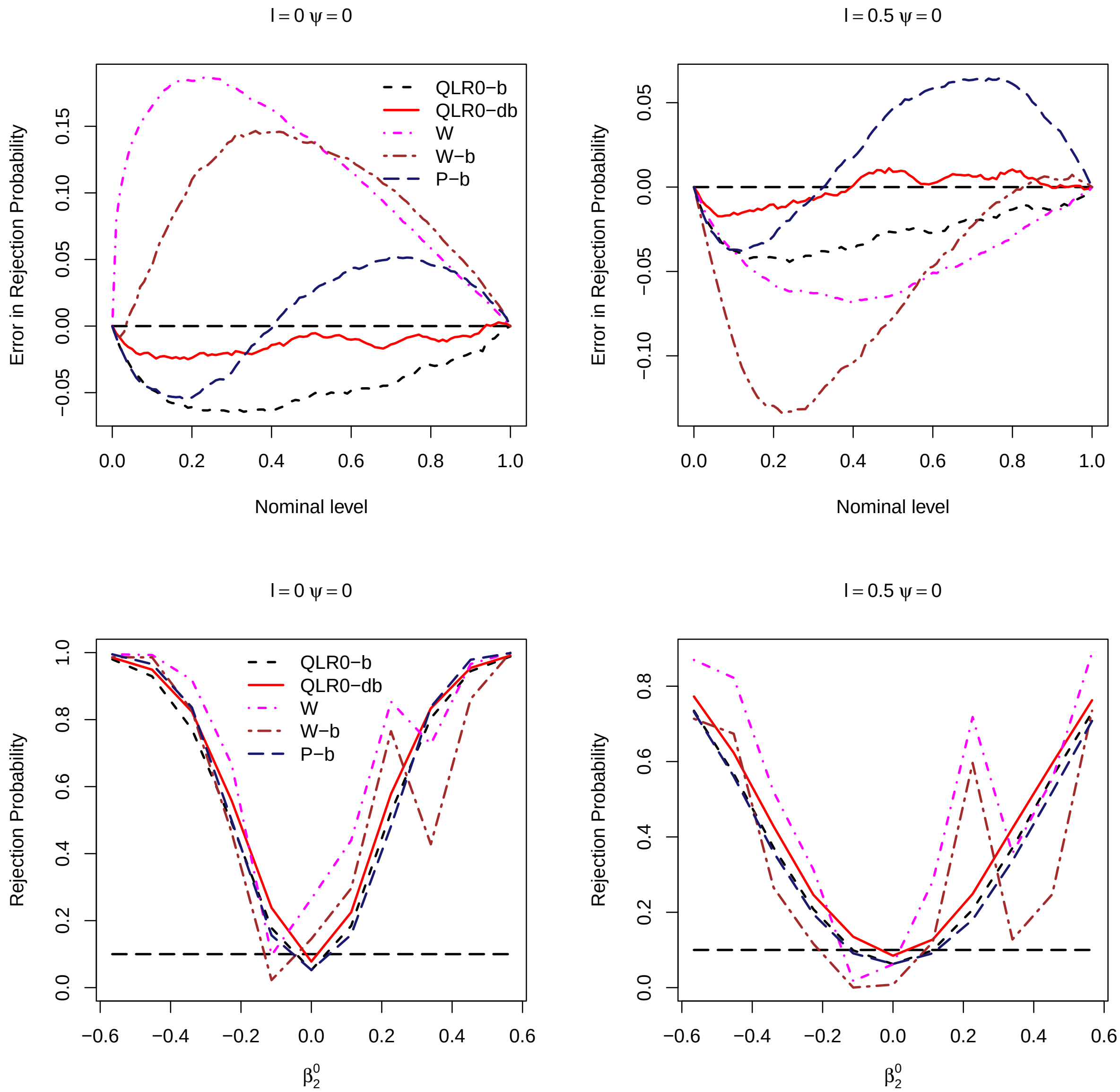

Figure 7: Errors in rejection probabilities and power curves for quantile regression with $\tau = .25$ and $\eta \sim t_5$.

Table 1: 90% Confidence Intervals for CIGAR parameter

	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$
Wald	[-13.83 , 1.49]	[-9.65 , -0.54]	[-9.46 , -1.31]
Wald-b	[-17.09 , 4.67]	[-10.14 , -0.07]	[-9.93 , -0.86]
P-b	[-15.60 , 3.15]	[-10.00 , -0.03]	[-9.83 , -0.94]
QLR0-b	[-15.72 , 1.73]	[-9.86 , 0.23]	[- 8.95 , -0.93]
QLR0-db	[-16.06 , 1.73]	[-9.87 , 0.41]	[-8.95 , -0.74]

For median regression, there are little differences between the outcomes of different methods, but confidence intervals from QLR0-b and QLRO-db are among the shortest, while the bootstrap Wald and percentile intervals are largest. For lower quantiles, the asymptotic Wald-based confidence interval is mostly shortest, while the bootstrap one is much larger. This is coherent with our simulation findings: Wald test does not control size well and can be severely oversized, and the bootstrap makes it more conservative. The difference between confidence intervals can be so large that for $\tau = 0.25$ the Wald and percentile bootstrap signals a significant effect of cigarettes consumption on birthweight, while the confidence intervals based on QLR indicates a non-significant effect. For $\tau = 0.1$, our bootstrap method delivers tighter intervals than competing bootstrap methods, which is likely related to its superior power performances under misspecification as observed in simulations.

5 Conclusion

We have proposed a simple bootstrap method for quasi likelihood-ratio tests that is consistent even under misspecification. A key advantage of QLR tests is that they do not necessitate to estimate a robust-covariance matrix: this can be difficult, as in quantile regression, and can severely affect size control and power performances. We found that our method yield in practice rejection probabilities that are close to nominal levels in small samples, and double bootstrapping under the null the non-robust QLR is preferable to relying on a robust version for size control as well as power.

6 Proofs

We first recall two useful results.

Theorem 6.1 (Andersen and Gill (1982, Theorem II.1))**.** *Let $Q_n(\cdot)$ be a sequence of random convex functions defined on an open convex Θ such that $Q_n(\theta) \xrightarrow{p} Q(\theta)$ for any $\theta \in \Theta$. Then for any compact subset K of Θ, $\sup_{\theta\in K}|Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$.*

Theorem 6.2 (Hjort and Pollard (1993, Basic Corollary))**.** *Suppose $A_n(s)$ is convex and can be represented as $\frac{1}{2}s'Vs + U_n's + C_n + r_n(s)$, where V is symmetric and positive definite, U_n is stochastically bounded, C_n is arbitrary, and $r_n(s) = o_p(1)$ as $n \to \infty$ for each s. Then α_n, the argmin of $A_n(s)$, is only $o_p(1)$ away from $\beta_n = -V^{-1}U_n$, the argmin of $\frac{1}{2}s'Vs + U_n's$. If also $U_n \xrightarrow{d} U$ then $\alpha_n \xrightarrow{d} -V^{-1}U$.*

Corollary 6.3. *Under the previous assumptions, we have as $n \to \infty$,*

$$\min_s A_n(s) = -\frac{1}{2}U_n'V^{-1}U_n + C_n + o_p(1)$$

.

Proof. The function $A_n(s) - U_n's - C_n$ is convex and converges in probability to $\frac{1}{2}s'Vs$ for each s. By Theorem 6.1, the convergence is uniform on bounded sets. Hence, using $\alpha_n = \beta_n + o_p(1)$ and $U_n = O_p(1)$,

$$\begin{aligned}\min_s A_n(s) = A_n(\alpha_n) &= \frac{1}{2}\alpha_n' V\alpha_n + U_n'\alpha_n + C_n + o_p(1)\\ &= \frac{1}{2}\beta_n' V\beta_n + U_n'\beta_n + C_n + o_p(1) = -\frac{1}{2}U_n'V^{-1}U_n + C_n + o_p(1)\,. \quad \square\end{aligned}$$

Proof of Theorem 3.1. Theorem 6.1 together with our i.i.d. assumption and A yield that $\mathrm{QLR}_n / n = Q_n(\widehat{\theta}_n^0) - Q_n(\widehat{\theta}_n)$ converges to $Q(\theta_\dagger^0) - Q_n(\theta_\dagger)$. Under H_A, this is a positive real constant.

We now deal with Parts (a) and (c) of the Theorem. Consider now the local alternatives

$$H_{A,n}:\ \theta_{\dagger,n} = H\gamma_\dagger + h + c\,n^{-1/2}\,.$$

and let $A_{\dagger,n} = A(\theta_{\dagger,n})$. Let $\theta_{\dagger,n}^0 = \arg\min_{\Theta\cap H_0} \mathbb{E}\,q(z,\theta)$, which depends on n since the expectation is taken under the sequence $\theta_{\dagger,n}$. The function $Q(\theta) = \mathbb{E}\,q(z,\theta)$ is convex, and the convex function $C^0(u) = C(Hu - c) = n\,[Q(\theta_{\dagger,n} + (Hu - c)/\sqrt{n}) - Q(\theta_{\dagger,n})]$ is minimized at $Hu - c = \sqrt{n}(\theta_{\dagger,n}^0 - \theta_{\dagger,n})$. For each fixed u, as $n \to \infty$, we have

$$\begin{aligned}C^0(u) &= \frac{1}{2}(Hu - c)'A_{\dagger,n}(Hu - c) + o(1)\\ &= -c'A_{\dagger,n}'Hu + \frac{1}{2}u'H'A_{\dagger,n}Hu + \frac{1}{2}c'A_{\dagger,n}c + o(1)\,.\end{aligned}$$

By Corollary 6.3, the minimum is reached at $u^* = -H(H'A_{\dagger,n}H)^{-1}H'A_{\dagger,n}c + o(1)$, and

$$\sqrt{n}(\theta_{\dagger,n}^0 - \theta_{\dagger,n}) = Hu^* - c = -A_{\dagger,n}^{-1/2}M_H A_{\dagger,n}^{1/2}c + o(1)\,,$$

with $M_H = \mathbf{I} - A_{\dagger,n}^{1/2} H (H' A_{\dagger,n} H)^{-1} H' A_{\dagger,n}^{1/2}$. Hence $\|\theta^0_{\dagger,n} - \theta_{\dagger,n}\| = O(n^{-1/2})$. Since $\widehat{\theta}^0_n$ converges to $\theta^0_{\dagger,n}$ by convexity of $Q_n(\cdot)$ and Theorem 6.1, $\widehat{\theta}^0_n$ belongs to any sufficiently small fixed neighborhood of $\theta_{\dagger,n}$ with probability tending to 1 as n grows to infinity.

Consider the convex function $C_n(s) = n\,[Q_n(\theta_{\dagger,n} + s/\sqrt{n}) - Q_n(\theta_{\dagger,n})]$, which is minimized at $\sqrt{n}(\widehat{\theta}_n - \theta_{\dagger,n})$. Let $\Delta_n = n^{-1/2} \sum_{i=1}^n D(z_i, \theta_{\dagger,n})$, which is $O_p(1)$ by Assumption B. Then

$$C_n(s) = \Delta_n' s + n\, R_n(s/\sqrt{n})\,.$$

From Assumption B, for each s, as $n \to \infty$ we get

$$\mathbb{E}\, n\, R_n(s/\sqrt{n}) = \frac{1}{2} s' A_{\dagger,n} s + o(1), \qquad \operatorname{Var} n\, R_n(s/\sqrt{n}) = o(1)\,.$$

Hence for each s

$$C_n(s) = \Delta_n' s + \frac{1}{2} s' A_{\dagger,n} s + o_p(1)\,.$$

By Corollary 6.3, $\min_s C_n(s) = -\frac{1}{2}\Delta_n' A_{\dagger,n}^{-1} \Delta_n + o_p(1)$.

The convex function $C_n^0(u) = C_n(Hu - c) = n\,[Q_n(\theta_{\dagger,n} + (Hu-c)/\sqrt{n}) - Q_n(\theta_{\dagger,n})]$ is minimized at $Hu - c = \sqrt{n}(\widehat{\theta}^0_n - \theta_{\dagger,n})$. By arguments similar to the ones used above, for each fixed u, we obtain as $n \to \infty$,

$$\begin{aligned} C_n^0(u) &= \Delta_n'(Hu - c) + \frac{1}{2}(Hu-c)' A_{\dagger,n} (Hu - c) + o_p(1) \\ &= (\Delta_n - A_{\dagger,n} c)'\, Hu + \frac{1}{2} u' H' A_{\dagger,n} H u - \Delta_n' c + \frac{1}{2} c' A_{\dagger,n} c + o_p(1)\,. \end{aligned}$$

By Corollary 6.3, we have

$$\begin{aligned}\min_u C_n^0(u) &= -\frac{1}{2}\left(\Delta_n - A_{\dagger,n}c\right)' H(H'A_{\dagger,n}H)^{-1}H'(\Delta_n - A_{\dagger,n}c) - \Delta_n' c + \frac{1}{2}c'A_{\dagger,n}c + o_p(1)\\ &= -\frac{1}{2}\left(\Delta_n - A_{\dagger,n}c\right)' H(H'A_{\dagger,n}H)^{-1}H'(\Delta_n - A_{\dagger,n}c)\\ &\quad + \frac{1}{2}\left(\Delta_n - A_{\dagger,n}c\right)' A_{\dagger,n}^{-1}\left(\Delta_n - A_{\dagger,n}c\right) - \frac{1}{2}\Delta_n' A_{\dagger,n}^{-1}\Delta_n + o_p(1)\\ &= \frac{1}{2}\left(\Delta_n - A_{\dagger,n}c\right)' A_{\dagger,n}^{-1/2} M_H A_{\dagger,n}^{-1/2}(\Delta_n - A_{\dagger,n}c) - \frac{1}{2}\Delta_n' A_{\dagger,n}^{-1}\Delta_n + o_p(1)\,.\end{aligned}$$

Therefore we obtain

$$\begin{aligned}\mathrm{QLR}_n = \min_u C_n^0(u) - \min_s C_n(s) &= \frac{1}{2}\left(\Delta_n - A_{\dagger,n}c\right)' A_{\dagger,n}^{-1/2} M_H A_{\dagger,n}^{-1/2}(\Delta_n - A_{\dagger,n}c) + o_p(1)\\ &= \left(B_{\dagger,n}^{-1/2}(\Delta_n - A_{\dagger,n}c)\right)' B_{\dagger,n}^{1/2} A_{\dagger,n}^{-1/2} M_H A_{\dagger,n}^{-1/2} B_{\dagger,n}^{1/2}\left(B_{\dagger,n}^{-1/2}(\Delta_n - A_{\dagger,n}c)\right) + o_p(1)\,.\end{aligned}$$

Let $\theta_\dagger = \lim_n \theta_{\dagger,n} = H\gamma_\dagger + h = \theta_\dagger^0$. By Assumption B-(b), $B_{\dagger,n} \to B_\dagger = B(\theta_\dagger^0)$ and $A_{\dagger,n} \to A_\dagger = A(\theta_\dagger^0)$. Hence $B_{\dagger,n}^{-1/2}(\Delta_n - A_{\dagger,n}^{1/2}c) \xrightarrow{d} N(-B_\dagger^{-1/2}A_\dagger c, \mathbf{I})$. QLR_n is thus asymptotically distributed as $\sum_{j=1}^p \lambda_j\left(U_j + \tilde{c}_j\right)^2$, where the λ_j are the eigenvalues of $G = B_\dagger^{1/2} A_\dagger^{-1/2} M_H A_\dagger^{-1/2} B_\dagger^{1/2}$, U_j are independent standard normal, and the $\tilde{c}_j$ are the scalar products of $-B_\dagger^{-1/2}A_\dagger c$ with the corresponding eigenvectors of G. G has rank r, which is the rank of M_H. Moreover, $M_H = P_d = A_\dagger^{-1/2} d\left(d'A_\dagger^{-1}d\right)^{-1} d'A_\dagger^{-1/2}$, since P_d and $P_H = \mathbf{I} - M_H$ are orthogonal projection matrices that annihilate each other and $\mathrm{rank}(P_d) + \mathrm{rank}(P_H) = p$. The eigenvalues of G are thus the ones of $P_d A_\dagger^{-1/2} B_\dagger A_\dagger^{-1/2}$, which are also the ones of $\left(d'A_\dagger^{-1}d\right)^{-1} dA_\dagger^{-1} B_\dagger^1 A_\dagger^{-1} d$.

Under H_0, we have $c = \mathbf{0}$ and Part (a) follows. Under $H_{A,n}$, we want to show that at least one of the $\tilde{c}_j$ corresponding to a non-zero eigenvalue is not zero. But

$$GB_\dagger^{-1/2} A_\dagger c = B_\dagger^{1/2} A_\dagger^{-1/2} P_d A_\dagger^{1/2} c \neq \mathbf{0}$$

as soon as $d'c \neq \mathbf{0}$. $\square$

Proof of Theorem 3.2. We first study $S_n(\widehat{\theta}_n^0)$. The convex function $C_n^0(u) = C_n(Hu) = n\left[Q_n(\theta_\dagger^0 + Hu/\sqrt{n}) - Q_n(\theta_\dagger^0)\right]$ is minimized at $Hu = \sqrt{n}(\widehat{\theta}_n^0 - \theta_\dagger^0)$. For each fixed u,

$$C_n^0(u) = \Delta_n^{0\prime} Hu + \frac{1}{2} u' H' A_\dagger^0 Hu + o_p(1)\,,$$

where $A_\dagger^0 = A(\theta_\dagger^0)$ and $\Delta_n^0 = n^{-1/2}\sum_{i=1}^n D(z_i, \theta_\dagger^0)$. By Theorem 6.2, $\sqrt{n}(\widehat{\theta}_n^0 - \theta_\dagger^0) = -H(H'A_\dagger^0 H)^{-1} H'\Delta_n^0 + o_p(1)$. From Assumption C,

$$\begin{aligned}\sqrt{n} S_n(\widehat{\theta}_n^0) - \Delta_n^0 &= \sqrt{n} A_\dagger^0 (\widehat{\theta}_n^0 - \theta_\dagger^0) + o_p(1) \\ &= -A_\dagger^0 H (H' A_\dagger^0 H)^{-1} H' \Delta_n^0 + o_p(1) = O_p(1).\end{aligned}$$

Let $\Delta_n^{0*} = n^{-1/2}\sum_{i=1}^n D_i(z_i^*, \theta_\dagger^0)$. From Gine and Zinn (1990), $\Delta_n^{0*} - \Delta_n^0$ has conditionally on the initial sample the same asymptotic distribution in probability as $\Delta_n^0 - n^{1/2}\mathbb{E}\, D(z, \theta_\dagger^0)$, which is gaussian by Assumption B. It is thus bounded in probability, and it is also the case for

$$\Delta_n^{0*} - \sqrt{n} S_n(\widehat{\theta}_n^0) = \Delta_n^{0*} - \Delta_n^0 + \Delta_n^0 - \sqrt{n} S_n(\widehat{\theta}_n^0)\,.$$

Consider the convex function $\widehat{C}_n(s) = n\left[\widehat{Q}_n(\theta_\dagger^0 + s/\sqrt{n}) - \widehat{Q}_n(\theta_\dagger^0)\right]$, which is minimized at $\sqrt{n}(\widehat{\theta}_n^* - \theta_\dagger^0)$. From Assumption D, for each fixed s,

$$\widehat{C}_n(s) = \left(\Delta_n^{0*} - \sqrt{n} S_n(\widehat{\theta}_n^0)\right)' s + \frac{1}{2} s' A_\dagger^0 s + o_p(1)\,.$$

By Corollary 6.3,

$$\min_s \widehat{C}_n(s) = -\frac{1}{2}\left(\Delta_n^{0*} - \sqrt{n} S_n(\widehat{\theta}_n^0)\right)' (A_\dagger^0)^{-1} \left(\Delta_n^{0*} - \sqrt{n} S_n(\widehat{\theta}_n^0)\right) + o_p(1)\,.$$

The convex function $\widehat{C}_n^0(u) = \widehat{C}_n(Hu) = n\left[\widehat{Q}_n(\theta_\dagger^0 + Hu/\sqrt{n}) - \widehat{Q}_n(\theta_\dagger^0)\right]$ is minimized at $Hu = \sqrt{n}(\widehat{\theta}_n^{0*} - \theta_\dagger^0)$. From Assumption D, for each fixed u,

$$\widehat{C}_n^0(u) = \left(\Delta_n^{0*} - \sqrt{n} S_n(\widehat{\theta}_n^0)\right) Hu + \frac{1}{2} u' H' A_\dagger^0 Hu + o_p(1)\,.$$

By Corollary 6.3,

$$\min_u \widehat{C}_n^0(u) = -\frac{1}{2}\left(\Delta_n^{0*} - \sqrt{n}S_n(\widehat{\theta}_n^0)\right)' H' \left(H' A_\dagger^0 H\right)^{-1} H \left(\Delta_n^{0*} - \sqrt{n}S_n(\widehat{\theta}_n^0)\right) + o_p(1)\,.$$

Let $M_H^0 = \mathbf{I} - \left(A_\dagger^0\right)^{1/2} H(H'A_\dagger^0 H)^{-1} H' \left(A_\dagger^0\right)^{1/2}$. Then

$$\begin{aligned}\mathrm{QLR}_n &= \min_u \widehat{C}_n^0(u) - \min_s \widehat{C}_n(s)\\ &= \frac{1}{2}\left(\Delta_n^{0*} - \sqrt{n}S_n(\widehat{\theta}_n^0)\right)' \left(A_\dagger^0\right)^{-1/2} M_H^0 \left(A_\dagger^0\right)^{-1/2}\left(\Delta_n^{0*} - \sqrt{n}S_n(\widehat{\theta}_n^0)\right) + o_p(1)\,.\end{aligned}$$

Now

$$M_H^0 \left(A_\dagger^0\right)^{-1/2}\left[\sqrt{n}S_n(\widehat{\theta}_n^0) - \Delta_n^0\right] = -M_H^0 \left(A_\dagger^0\right)^{1/2} H(H'A_\dagger^0 H)^{-1} H' \Delta_n^0 + o_p(1) = o_p(1)\,,$$

as M_H^0 annihilates $\left(A_\dagger^0\right)^{1/2} H(H'A_\dagger^0 H)^{-1} H'$. This yields

$$\mathrm{QLR}_n = \frac{1}{2}\left(\Delta_n^{0*} - \Delta_n^0\right)' \left(A_\dagger^0\right)^{-1/2} M_H^0 \left(A_\dagger^0\right)^{-1/2}\left(\Delta_n^{0*} - \Delta_n^0\right) + o_p(1)\,.$$

Under H_0, $\theta_\dagger^0 = \theta_\dagger$, and $\Delta_n^* - \Delta_n$ has conditionally on the initial sample the same asymptotic distribution in probability as Δ_n, so QLR_n^* has conditionally on the initial sample the same asymptotic distribution in probability as QLR_n.

Under H_A, QLR_n^* stays conditionally bounded in probability, and so does its conditional $(1-\alpha)$ quantile. Since QLR_n diverges, asymptotic power converges to one.

Under $H_{A,n}$, $\|\theta_{\dagger,n}^0 - \theta_{\dagger,n}\| = O(n^{-1/2})$. Using the continuity of $A(\theta)$ and $B(\theta)$, QLR_n^* is asymptotically, conditional on the initial sample, distributed as $\sum_{j=1}^p \lambda_j U_j^2$ from Theorem 3.1-(a), where λ is the vector of r eigenvalues of

$$\left(d'(A_\dagger^0)^{-1}d\right)^{-1} d'(A_\dagger^0)^{-1} B_\dagger^0 (A_\dagger^0)^{-1} d\,, \qquad \text{where } B_\dagger^0 = B\left(\theta_\dagger^0\right) \qquad \theta_\dagger^0 = H\gamma^\dagger + h\,.$$

The asymptotic distribution of QLR_n is the one of $\sum_{j=1}^p \lambda_j \left(U_j + \tilde{c}_j\right)^2$, with at least one non-zero $\tilde{c}_j$. It first-order stochastically dominates the conditional distribution of QLR_n^*, and the asymptotic test's power is thus non-trivial. □

Proof of Theorem 3.3. We adopt a functional view of the regression parameters, which allows to compute the influence functions of the parameters and to prove the asymptotic validity of the method when standardizing by the correct variance. We then derive the Edgeworth expansions of the bootstrap and original statistics and show that their coefficients coincide up to terms of order $O_p(n^{-1/2})$. Define

$$\gamma(P) \equiv \gamma = \left(d'\mathbb{E}^{-1}(xx')d\right)^{-1} d'\left(\theta_\dagger - h\right) \quad \text{and} \quad \Sigma = \mathbb{E}\left(xx'(y - x'\theta_\dagger)^2\right),$$

where expectations at taken at P Let $\gamma(P_n) \equiv \gamma_n$ and $\gamma(P_n^*) \equiv \gamma_n^*$ be similar quantities defined by replacing $\mathbb{E}$ by $\mathbb{E}_n$ or $\mathbb{E}_n^*$. Then

$$\begin{aligned}
\mathrm{RQLR}_n &= \frac{n\,\gamma_n^2\left(d'\mathbb{E}_n^{-1}(xx')d\right)^2}{d'\mathbb{E}_n^{-1}(xx')\Sigma_n\mathbb{E}_n^{-1}(xx')d} \\
\mathrm{RQLR}_n^* &= \frac{n\,\left(\gamma_n^* - \gamma_n\right)^2\left(d'\mathbb{E}_n^{*-1}(xx')d\right)^2}{d'\mathbb{E}_n^{*-1}(xx')\Sigma_n^*\mathbb{E}_n^{*-1}(xx')d}\,.
\end{aligned}$$

Since $\gamma(P)$ is a smooth functional of moments, its influence function is given by

$$\begin{aligned}
\gamma^{(1)}(y,x) = {} & \left(d'\mathbb{E}^{-1}(xx')d\right)^{-1} d'\mathbb{E}^{-1}(xx')x(y - x'\theta_\dagger) \\
& - d'\left(\theta_\dagger - h\right)\frac{d'\mathbb{E}^{-1}(xx')\left(xx' - \mathbb{E}\,xx'\right)\mathbb{E}^{-1}(xx')d}{\left(d'\mathbb{E}^{-1}(xx')d\right)^2}\,.
\end{aligned}$$

Under H_0, the asymptotic variance of γ_n is

$$\mathrm{Var}\left(\gamma^{(1)}(y,x)\right) = \frac{d'\mathbb{E}^{-1}(xx')\,\Sigma\,\mathbb{E}^{-1}(xx')d}{\left(d'\mathbb{E}^{-1}(xx')d\right)^2} = \omega^2\,.$$

Define the empirical counterpart and its bootstrap version as

$$\omega_n^2 = \frac{d'\mathbb{E}_n^{-1}(xx')\,\Sigma_n\,\mathbb{E}_n^{-1}(xx')d}{\left(d'\mathbb{E}_n^{-1}(xx')d\right)^2} \quad \text{and} \quad \omega_n^{2*} = \frac{d'\mathbb{E}_n^{*-1}(xx')\Sigma_n^*\,\mathbb{E}_n^{*-1}(xx')d}{\left(d'\mathbb{E}_n^{*-1}(xx')d\right)^2}\,.$$

Provided ω^2 exists and is different from zero, $n^{1/2}\omega_n^{-1}\left(\gamma_n - \gamma\right) \xrightarrow{d} N(0,1)$, since this is the ratio of two continuously Hadamard differentiable functionals from van der Vaart (1998, Theorem 20.8). Similarly, $n^{1/2}\omega_n^{*\,-1}\,(\gamma_n^* - \gamma_n) \xrightarrow{d} N(0,1)$ conditionally on the original sample from Gill (1989). As $\mathrm{RQLR}_n = \left(n^{1/2}\omega_n^{-1}\,(\gamma_n - \gamma)\right)^2$ and γ is a smooth functional of moments with regular gradients (influence functions), $n^{1/2}\omega_n^{-1}\,(\gamma_n - \gamma)$ admits an Edgeworth expansion up to $O(n^{-2})$, under moments of order 6 of the gradients and the Cramer condition of Theorem 3.3, see Bhattacharya and Ghosh (1978). Moreover, by symmetry, we have that uniformly in $x \geq 0$

$$\begin{aligned}\Pr\left(n^{1/2}\left|\frac{\gamma_n-\gamma}{\omega_n}\right| \leq x\right) &= \Pr\left(n^{1/2}\frac{\gamma_n-\gamma}{\omega_n} \leq x\right) - \Pr\left(n^{1/2}\frac{\gamma_n-\gamma}{\omega_n} \leq -x\right)\\ &= \Phi(x) - \Phi(-x) \ - \frac{2}{n}\left(\frac{k_2}{2} + \frac{k_4}{24}(x^3-3x)\right.\\ &\left. +\frac{k_6}{72}(x^5 - 10x^3 + 15x)\right)\phi(x) \ + O(n^{-2})\,.\end{aligned}$$

Explicit expressions of the cumulants k_2, k_4 and k_6 are given in Bertail and Barbe (1995, Appendix 2), where the coefficients involve both the influence function $\gamma^{(1)}(y,x)$ of γ and the one of ω^2, see also Withers (1983, 1984) and Hall (1992). The bootstrap distribution of $n^{1/2}\omega_n^{*-1}\,(\gamma_n^* - \gamma_n)$ has the same functional form and thus admits the same Edgeworth expansion with true cumulants replaced by the empirical ones $k_{2,n}$, $k_{4,n}$ and $k_{6,n}$. The result then follows from the fact that $k_{j,n} - k_j = O_P(n^{-1/2})$, for $j = 2,4,6$, which is ensured by the moment conditions of order 12 (since k_6 contains moments of order 6 of the gradient, we need moments of order 12 to ensure that a CLT holds). The result about coverage probability follows from Hall (1986). □

References

Abrevaya, J. (2001): “The Effects of Demographics and Maternal Behavior on the Distribution of Birth Outcomes,” *Empir. Econ.*, 26, 247–257.

Andersen, P. K. and R. D. Gill (1982): "Cox's Regression Model for Counting Processes: A Large Sample Study," *Ann. Statist.*, 10, 1100–1120.

Andrews, D. W. (2002): "Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119–162.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Beran, R. (1988): "Prepivoting Test Statistics - A Bootstrap View of Asymptotic Refinements," *J. Am. Stat. Assoc.*, 83, 687–697.

Bertail, P. and P. Barbe (1995): *The Weighted Bootstrap*, vol. 98 of *Lecture Notes in Statistics*, Springer-Verlag New York.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *Ann. Statist.*, 6, 434–451.

Bose, A. and S. Chatterjee (2001): "Generalised Bootstrap in Non-Regular M-Estimation Problems," *Statistics & Probability Letters*, 55, 319–328.

——— (2003): "Generalized Bootstrap for Estimators of Minimizers of Convex Functions," *J. Stat. Plan. Infer.*, 117, 225–239.

Camponovo, L. (2016): "Asymptotic Refinements of Nonparametric Bootstrap for Quasi-Likelihood Ratio Tests for Classes of Extremum Estimators," *Econom. J.*, 19, 33–54.

Chatterjee, S. and A. Bose (2005): "Generalized Bootstrap for Estimating Equations," *Ann. Statist.*, 33, 414–436.

Chen, X. and D. Pouzo (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *J. Econometrics*, 152, 46–60.

Chesher, A. and G. Austin (1991): "The Finite-Sample Distributions of Heteroskedasticity Robust Wald Statistics," *J. Econometrics*, 47, 153–173.

Cribari-Neto, F., S. L. P. Ferrari, and W. A. S. C. Oliveira (2005): "Numerical Evaluation of Tests Based on Different Heteroskedasticity-Consistent Covariance Matrix Estimators," *J. Stat. Comput. Simul.*, 75, 611–628.

Das, D. and S. Lahiri (2019): "Second Order Correctness of Perturbation Bootstrap M-estimator of Multiple Linear Regression Parameter," *Bernoulli*, 25, 654–682.

Davidson, R. and J. G. MacKinnon (2007): "Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap," *Comput. Statist. Data Anal.*, 51, 3259–3281.

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jacknife," *Ann. Statist.*, 7, 1–26.

Eicker, F. (1967): "Limit Theorems for Regressions with Unequal and Dependent Errors," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif.: University of California Press, 59–82.

Falk, M. and D. Janas (1992): "Edgeworth Expansions for Studentized and Prepivoted Sample Quantiles," *Stat. Probabil. Lett.*, 14, 13–24.

Foutz, R. V. and R. C. Srivastava (1977): "The Performance of the Likelihood Ratio Test When the Model is Incorrect," *Ann. Statist.*, 5, 1183–1194.

Freedman, D. A. (2007): "How Can the Score Test Be Inconsistent?" *The American Statistician*, 61, 291–295.

Giacomini, R., D. N. Politis, and H. White (2013): "A Warp-Speed Method For Conducting Monte Carlo Experimens Involving Bootstrap Estimators," *Economet. Theor.*, 29, 567–589.

Gill, R. D. (1989): "Non- and Semi-Parametric Maximum Likelihood Estimators and the Von Mises Method (Part 1) [with Discussion and Reply]," *Scand. J. Stat.*, 16, 97–128.

Gine, E. and J. Zinn (1990): "Bootstrapping General Empirical Measures," *Ann. Probab.*, 18, 851–869.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681–700.

Gutenbrunner, C. and J. Jureckova (1992): "Regression Rank Scores and Regression Quantiles," *Ann. Statist.*, 20, 305–330.

Hall, P. (1986): "On the Bootstrap and Confidence Intervals," *Ann. Statist.*, 14, 1431–1452.

——— (1992): *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, Springer-Verlag New York.

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-Of-Moments Estimators," *Econometrica*, 64, 891–916.

Hall, P. and E. Mammen (1994): "On General Resampling Algorithms and their Performance in Distribution Estimation," *The Annals of Statistics*, 22, 2011–2030.

Hall, P. and S. R. Wilson (1991): "Two Guidelines for Bootstrap Hypothesis Testing," *Biometrics*, 47, pp. 757–762.

Hjort, N. L. and D. Pollard (1993): "Asymptotics for Minimisers of Convex Processes," *arXiv:1107.3806*, 24.

Huber, P. J. (1967): "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif.: University of California Press, 221–233.

Jin, Z., Z. Ying, and L. J. Wei (2001): "A Simple Resampling Method by Perturbing the Minimand," *Biometrika*, 88, 381–390.

Kent, J. T. (1982): "Robust Properties of Likelihood Ratio Tests," *Biometrika*, 69, 19–27.

Kim, T.-H. and H. White (2003): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," in *Advances in Econometrics*, Bingley: Emerald (MCB UP), vol. 17, 107–132.

Kline, P. and A. Santos (2012): "Higher Order Properties of the Wild Bootstrap Under Misspecification," *J. Econometrics*, 171, 54–70.

Kocherginsky, M., X. He, and Y. Mu (2005): "Practical Confidence Intervals for Regression Quantiles," *J. Comput. Graph. Statist.*, 14, 41–55.

Koenker, R. (2005): *Quantile Regression*, New York: Cambridge University Press.

Koenker, R. and J. Bassett, Gilbert (1978): "Regression Quantiles," *Econometrica*, 46, pp. 33–50.

Lee, S. M. S. and P. Yang (2020): "Bootstrap Confidence Regions Based on M-Estimators Under Nonstandard Conditions," *Ann. Statist.*, 48, 274–299.

Lien, D. and Q. H. Vuong (1987): "Selecting the Best Linear Regression Model: A Classical Approach," *J. Econometrics*, 35, 3 – 23.

Liu, R. (1988): "Bootstrap Procedures Under Some Non-Iis Models," *Ann. Statist.*, 16, 1696–1708.

Long, J. S. and L. H. Ervin (2000): "Using heteroscedasticity consistent standard errors in the linear regression model," *Am. Stat.*, 54, 217–224.

MacKinnon, J. G. and H. White (1985): "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *J. Econometrics*, 29, 305–325.

Maistre, S., P. Lavergne, and V. Patilea (2017): "Powerful Nonparametric Checks for Quantile Regression," *J. Stat. Plan. Infer.*, 180, 13 – 29.

Mammen, E. (1992): "Bootstrap, Wild Bootstrap, and Asymptotic Normality," *Probab. Theory Rel.*, 93, 439–455.

Marcellino, M. and B. Rossi (2008): "Model Selection for Nested and Overlapping Nonlinear, Dynamic and Possibly Mis-specified Models," *Oxford B. Econ. Stat.*, 70, 867–893.

Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, vol. 4*, ed. by R. Engle and D. McFadden, Cambridge University Press, chap. 36, 2111–2245.

Nolan, D. and D. Pollard (1987): "U-Processes: Rates of Convergence," *Ann. Statist.*, 15, 780 – 799.

Powell, J. (1991): "Estimation of Monotonic Regression Models UUnder Quantile Restrictions," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen, Cambridge University Press.

Praestgaard, J. and J. A. Wellner (1993): "Exchangeably Weighted Bootstraps of the General Empirical Process," *Ann. Probab.*, 21, 2053–2086.

Royall, R. M. (1986): "Model Robust Confidence Intervals Using Maximum Likelihood Estimators," *Int. Stat. Rev.*, 54, 221.

Sherman, R. (1994): "Maximal Inequalities for Degenerate $U-$Processes with Applications to Optimization Estimators," *Ann. Statist.*, 22, 439–459.

Spokoiny, V. and M. Zhilova (2015): "Bootstrap Confidence Sets under Model Misspecification," *Ann. Statist.*, 43, 2653–2675.

Stafford, J. E. (1996): "A Robust Adjustment of the Profile Likelihood," *Ann. Statist.*, 24, 336–352.

Tarr, G. (2012): "Small Sample Performance of Quantile Regression Confidence Intervals," *J. Stat. Comput. Simul.*, 82, 81–94.

Tibshirani, R. (1992): "Bootstrap Hypothesis Testing," *Biometrics*, 48, 969–970.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307–333.

White, H. (1980): "Using Least Squares to Approximate Unknown Regression Functions," *Int. Econ. Rev.*, 21, 149–70.

——— (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, pp. 1–25.

Withers, C. S. (1983): "Expansions for the Distribution and Quantiles of a Regular Functional of the Empirical Distribution with Applications to Nonparametric Confidence Intervals," *Ann. Statist.*, 11, 577–587.

——— (1984): "Asymptotic Expansions for Distributions and Quantiles with Power Series Cumulants," *J. Roy. Stat. Soc. B.*, 46, 389–396.

Wu, C. F. J. (1986): "Jackknife, Bootstrap and Other Resampling Methods in Regression-Analysis - with Discussion," *Ann. Statist.*, 14, 1261–1295.

Appendix

We here briefly check our approximations of Assumptions B, C, and D for our two main examples.

Linear Regression

B We have $q(z,\theta) = (y - x'\theta)^2/2$ and $D(z,\theta) = \partial q(z,\theta)/\partial\theta = -x(y - x'\theta)$. Then

$$\mathbb{E}\, D(z,\theta_\dagger) = -\mathbb{E}\, x(y - x'\theta_\dagger) = 0\,, \qquad B(\theta) = \mathrm{Var}\, D(z,\theta) = \mathbb{E}\, \left(x(y - x'\theta)^2 x\right)\,.$$

We can write

$$\begin{aligned} Q_n(\theta + t) - Q_n(\theta) - n^{-1}\sum_{i=1}^n D'(z_i,\theta)t = R_n(\theta,t) &= n^{-1}\sum_{i=1}^n \frac{1}{2} t' x_i x_i' t \\ \mathbb{E}\, R_n(\theta,t) &= \frac{1}{2} t' \mathbb{E}\,(xx')t \\ n\ \mathrm{Var}\, R_n(\theta,t) &= \frac{1}{4}\,\mathrm{Var}(t'xx't) = o(\|t\|^2) \end{aligned}$$

when x has moments of order at least 4, since we have

$$Var(t'xx't) = E(t'(xx' - E(xx')t)^2 \leq ||t||^2 E(||xx' - E(xx')||^2)||t||^2 \leq 2||t||^4 E||x||^4\,.$$

C As $D(z,\theta+t) - D(z,\theta) = xx't$, $\mathbb{E}\, D(z,\theta+t) - D(z,\theta) = \mathbb{E}\,(xx')\,t$. The process $R_{1n}(t) - \mathbb{E}\, R_{1n}(t)$ is a degenerate U-process of order 1 based on the class of vector functions $\{f(x,t) = xx't,\ t \in \mathbb{R}^p\}$ which is Euclidean, see Nolan and Pollard (1987). Each function is such that $\mathbb{E}\,|f(x,t)| \to 0$ as $t \to 0$. For bounded t, these functions have a squared-integrable envelope as soon as $\mathbb{E}\,(xx')$ exists. From Sherman (1994, Corollary 8), $R_{1n}(\theta,t) - \mathbb{E}\, R_{1n}(\theta,t) = o_p(n^{-1/2})$ uniformly over $o_p(1)$ neighborhoods of 0.

D We have

$$Q_n^*(\theta + t) - Q_n^*(\theta) - n^{-1}\sum_{i=1}^n D'(z_i^*,\theta)t = R_n(\theta,t) = n^{-1}\sum_{i=1}^n \frac{1}{2} w_i t' x_i x_i t\,,$$

where $(w_1, \dots w_n) \sim \text{Mult}_n(n, (n^{-1}, \dots n^{-1}))$. Hence

$$\mathbb{E}_{|\mathcal{Z}} R_n(\theta, t) = n^{-1} \sum_{i=1}^{n} \frac{1}{2} t' x_i x_i' t = \frac{1}{2} t' \mathbb{E}\,(xx')t + o_p(1)$$

$$n \;\text{Var}_{|\mathcal{Z}}\, R_n(\theta, t) \leq \frac{15}{4} n^{-1} \sum_{i=1}^{n} [t' x_i x_i' t]^2 = o_p(\|t\|^2)$$

using that multinomial weights are negatively correlated and with fourth moments bounded by 15, see Praestgaard and Wellner (1993).

Quantile Regression

B For quantile regression of order τ, $q(y, x, \theta) = \rho_\tau(y - x'\theta)$, with $\rho_\tau(u) = u\,(\tau - \mathbb{I}\,(u < 0))$. Without loss of generality, we focus on median regression, i.e. $\tau = 1/2$, and rescale ρ_τ to obtain $q(y, x, \theta) = |y - x'\theta|$ and $D(z, \theta) = -x\,\text{sign}(y - x'\theta)$. Then

$$\mathbb{E}\, D(z, \theta_\dagger) = \frac{\partial Q(\theta_\dagger)}{\partial \theta} = -\mathbb{E}\,\left[x\,\text{sign}(y - x'\theta_\dagger)\right] = 0$$

$$B(\theta) = \text{Var}\, D(z, \theta) = \mathbb{E}\,\left(xx'\,\text{sign}(y - x'\theta_\dagger)\right)\;.$$

We have

$$Q_n(\theta + t) - Q_n(\theta) - n^{-1} \sum_{i=1}^{n} D'(z_i, \theta)t = R_n(\theta, t) = n^{-1} \sum_{i=1}^{n} r_n(z_i, \theta, t)$$

$$\text{where } r_n(z_i, \theta, t) = 2\left(y_i - x_i'(\theta + t)\right)$$

$$\times\left[\mathbb{I}(x_i't \leq y_i - x_i'\theta \leq 0) - \mathbb{I}(0 \leq y_i - x_i'\theta \leq x_i't)\right]\;.$$

Consider

$$\mathbb{E}\,\left(r_n(z, \theta, t)|x\right) = \int_{x'(\theta+t)}^{x'\theta} 2\left(y - x'(\theta + t)\right) f_{y|x}(y|x)\, dy\,.$$

Deriving with respect to $x't$, we obtain

$$
\begin{aligned}
\frac{\partial}{\partial(x't)}\mathbb{E}\,(r_n(z,\theta,t)|x) &= 2\left[F_{y|x}(x'(\theta+t)|x) - F_{y|x}(x'\theta)\right] = 0 \qquad \text{at} \;\; t=0 \\
\frac{\partial^2}{\partial(x't)^2}\mathbb{E}\,(r_n(z,\theta,t)|x) &= 2f_{y|x}(x'(\theta+t)|x) \\
\text{so that} \;\; \mathbb{E}\,r_n(z,\theta,t) &= 2\,t'\mathbb{E}\,\left[x'xf_{y|x}(x'\theta|x)\right]t + o(\|t\|^2)
\end{aligned}
$$

as $t \to 0$ if $f_{y|x}$ is bounded continuous at $x'\theta$ for all x and the above expectation exists.

Consider now

$$
\mathbb{E}\,\left(r_n^2(z,\theta,t)|x\right) = 4\,\mathrm{sign}(x't)\int_{x'\theta}^{x'(\theta+t)}\left(y_i - x_i'(\theta+t)\right)^2 f_{y|x}(y|x)\,dy\,.
$$

Assume that $x't \geq 0$, the case of $x't < 0$ would be dealt with similarly. Deriving with respect to $x't$,

$$
\begin{aligned}
\frac{\partial}{\partial(x't)}\mathbb{E}\,\left(r_n^2(z,\theta,t)|x\right) &= -8\int_{x'\theta}^{x'(\theta+t)}\left(y_i - x_i'(\theta+t)\right) f_{y|x}(y|x)\,dy = 0 \qquad \text{at} \;\; t=0 \\
\frac{\partial^2}{\partial(x't)^2}\mathbb{E}\,\left(r_n^2(z,\theta,t)|x\right) &= 8\left[F_{y|x}(x'(\theta+t)|x) - F_{y|x}(x'\theta)\right] = 0 \qquad \text{at} \;\; t=0 \\
\frac{\partial^3}{\partial(x't)^3}\mathbb{E}\,\left(r_n^2(z,\theta,t)|x\right) &= 8f_{y|x}(x'(\theta+t)|x)\,.
\end{aligned}
$$

This yields $\mathbb{E}\,r_n^2(z,\theta,t) = O(\|t\|^3)$ for any fixed t. Now use

$$
n\;\mathrm{Var}\,R_n(\theta,t) \leq \mathbb{E}\,r_n^2(z,\theta,t) = o(\|t\|^2)\,.
$$

C We have

$$
\begin{aligned}
D(z,\theta+t) - D(z,\theta) &= -2x\left[\mathbb{I}\left(x't \leq y - x'\theta \leq 0\right) - \mathbb{I}\left(0 \leq y - x'\theta \leq x't\right)\right] \\
\mathbb{E}\,D(z,\theta+t) - D(z,\theta) &= 2\mathbb{E}\,x\left[F_{y|x}\left(x'(\theta+t)\right) - F_{y|x}\left(x'\theta\right)\right] \\
&= 2\mathbb{E}\,\left[xx'f_{y|x}(x'\theta|x)\right]t + o(\|t\|)
\end{aligned}
$$

as $t \to 0$. The process $R_{1n}(t) - \mathbb{E}\, R_{1n}(t)$ is a degenerate U-process of order 1 based on the class of vector functions $\left\{ f(x,t) = -2x \left[\mathbb{I}\left(x't \leq y - x'\theta^0_\dagger \leq 0 \right) - \mathbb{I}\left(0 \leq y - x'\theta^0_\dagger \leq x't \right) \right] \right\}$, which is Euclidean. We have $\mathbb{E}\,|f(x,t)| \to 0$ as $t \to 0$. For bounded t, these functions have a squared-integrable envelope as soon as $\mathbb{E}\, xx'$ exists. From Sherman (1994, Corollary 8), $R_{1n}(t) - \mathbb{E}\, R_{1n}(t) = o_p(n^{-1/2})$ uniformly over $o_p(1)$ neighborhoods of 0.

D We have

$$Q_n^*(\theta + t) - Q_n^*(\theta) - n^{-1} \sum_{i=1}^n D'(z_i^*, \theta) t = R_n^*(\theta, t) = n^{-1} \sum_{i=1}^n w_i r_n(z_i, \theta, t) \,,$$

where $(w_1, \ldots w_n) \sim \text{Mult}_n(n, (n^{-1}, \ldots n^{-1}))$. Hence

$$\begin{aligned} \mathbb{E}_{\,|\mathcal{Z}} R_n(\theta, t) &= n^{-1} \sum_{i=1}^n r_n(z_i, \theta, t) = 2\, t' \mathbb{E} \left[x'x f_{y|x}(x'\theta|x) \right] t + o_p(\|t\|^2) \\ n \ \text{Var}_{|\mathcal{Z}}\, R_n(\theta, t) &\leq \frac{15}{4} n^{-1} \sum_{i=1}^n r_n^2(z_i, \theta, t) = o_p(\|t\|^2) \,. \end{aligned}$$