Escanciano, J. C., & Goh, S. C. (2014). Specification analysis of linear quantile models. *Journal of Econometrics*, 178, pp. 495-507.

DOI:10.1016/j.jeconom.2013.07.006

Specification Analysis of Linear Quantile Models

J. C. Escanciano
Indiana University

S. C. Goh[*]
The University of Wisconsin-Milwaukee

Abstract

This paper introduces a nonparametric test for the correct specification of a linear conditional quantile function over a continuum of quantile levels. These tests may be applied to assess the validity of post-estimation inferences regarding the effect of conditioning variables on the distribution of outcomes. We show that the use of an orthogonal projection on the tangent space of nuisance parameters at each quantile index both improves power and facilitates the simulation of critical values via the application of a simple multiplier bootstrap procedure. Monte Carlo evidence is included, along with an application to an analysis of the empirical relationship between age and individual earnings in the U.S.A.

Keywords: Quantile regression, specification tests, empirical processes, wild bootstrap.

JEL classification: C12, C31, C52

[*]Corresponding author. Department of Economics, The University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201. Phone: +1 (414) 229 3395. Fax: +1 (414) 229 3860. E-mail: goh@uwm.edu

1 Introduction

Let Y be a random variable, and X a d-dimensional random vector. Consider the continuum of conditional probability restrictions given by

$$P\left[Y \leq X^\top \theta_0(\alpha) \middle| X\right] = \alpha, \qquad \alpha \in \mathcal{A}, \tag{1.1}$$

where $\mathcal{A}$ is a compact subset of $(0,1)$, $X^\top$ denotes the transpose of X, and $\theta_0(\cdot)$ is a measurable unknown function from $\mathcal{A}$ to a compact subset Θ of $\mathbb{R}^d$. As such, (1.1) involves a continuum of conditional quantile models that are each linear in a vector of parameters $\theta_0(\alpha)$. These models generalize conventional linear regression models identified by a median restriction. Estimators of $\theta_0(\alpha)$ in the context of (1.1) were pioneered in econometrics by Koenker and Bassett (1978). Their methodology has proven to be quite popular in recent years (e.g., Koenker and Hallock, 2001; Koenker 2005 and references cited therein).

This paper develops omnibus tests for the correct specification of the conditional quantile function $X^\top\theta_0(\cdot)$ in (1.1) over $\mathcal{A}$. The hypothesis is that (1.1) holds with probability one for some $\theta_0(\cdot)$ in the corresponding parameter space, while the alternative is simply the negation of the null. We aim to construct consistent tests, i.e., tests that reject with probability tending to one in large samples if the model is misspecified. Such tests are important in applications because the conclusions of any post-estimation inferences based on the fitted quantile model will be sensitive to the implicit assumption that the conditional quantile function $X^\top\theta_0(\cdot)$ is correctly specified for all quantiles $\alpha \in \mathcal{A}$. In particular, if the conditional α'-quantile of Y for some $\alpha' \in \mathcal{A}$ is incorrectly assumed to have the form $X^\top\theta_0(\alpha')$, then estimators of $\theta_0(\alpha')$ will result in missleading inferences of the marginal effect of X on the α'-quantile of Y; see Angrist, Chernozhukov and Fernández-Val (2006).

While omnibus tests of the validity of a linear-in-parameters conditional quantile function against unspecified alternatives have already been developed in a number of different papers, the analysis has to date been mostly limited to a single quantile, generally taken without loss of generality to be the median. See e.g., the papers of Zheng (1998), Bierens and Ginther (2001), Horowitz and Spokoiny (2002) and Whang (2006a, 2006b). Horowitz and Lee (2009) develop a specification test for the more general case where X is possibly endogenous and the single-quantile restriction holds conditionally on a vector of instruments. It is straightforward in theory to extend the testing approach that we propose to the case where X is endogenous and the "structural" quantile function $X^\top\theta_0(\alpha)$ is identified by a vector of instruments satisfying the conditional rank invariance condition of Chernozhukov and Hansen (2005). An extension of this sort, however, is complicated in practice by the apparent unavailability of a computationally convenient estimator of the null nuisance parameter $\theta_0(\alpha)$ in the case of a structural quantile model. In particular, an extension to the case of endogenous X would require in practice estimates of the null nuisance parameters $\theta_0(\alpha_j)$ over a grid of quantiles $\{\alpha_j\}$ that becomes progressively dense in the sample size.

To the best of our knowledge, essentially the only proposal to date for omnibus specification tests of the functional form of a conditional quantile model over a continuum of quantiles is given in Escanciano and Velasco (2010). These authors considered tests of possibly nonlinear dynamic quantile models implemented using subsampling. A specialization of an approach recently proposed by Rothe and

Wied (2012) and based on estimates of conditional *distribution* functions might also be considered in this context.

In this paper, we consider an approach in the specific framework of linear quantile regression models for independently and identically distributed (iid) data. What appears distinctive to our approach is the specific treatment of the nuisance parameters in the testing problem. In contrast to Escanciano and Velasco (2010), we acknowledge our lack of knowledge of $\theta_0(\cdot)$ in testing for (1.1), which can be accounted for by an orthogonal projection of a certain weight function into the so-called tangent space of nuisance parameters at each fixed quantile $\alpha \in \mathcal{A}$. The result of this projection is a test with improved power properties. The improvement in power of such projections has been noticed before in different contexts. Neyman (1959) first applied this idea in the context of fully parametric models. A more recent extension of this idea to the semiparametric context has been proposed by Bickel et al. (2006).

To illustrate the main ideas in the simplest possible terms, consider an example in a fully parametric case where one observes a random sample $\{W_i : i = 1, \ldots, n\}$ from the population of W with density f_{θ_0} satisfying the moment restriction $E[m(W, \theta_0)] = 0$, where θ_0 is an unknown nuisance parameter with values in $\Theta \subset \mathbb{R}^d$. Suppose $m(W, \cdot)$ is continuously differentiable at θ_0, with bounded derivative, and that a $\sqrt{n}-$consistent estimator of θ_0 is available, say θ_n. Furthermore, assume that $E[m^2(W, \theta_0)] < \infty$ and that $E[|s_\theta(W, \theta_0)|^2] < \infty$, where s_θ denotes the derivative of the log density with respect to θ. To test the moment restriction, it is natural to base a test on the sample analog of moments, i.e., on the statistic $\hat{R}_n = n^{-1/2} \sum_{i=1}^n m(W_i, \theta_n)$.

Using a standard Taylor expansion, a uniform law of large numbers and the generalized information equality (i.e., $E[\partial m(W, \theta_0)/\partial\theta] = -E[m(W, \theta_0)s_{\theta_0}(W)]$), we obtain the expansion

$$\hat{R}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(W_i, \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)'\, E[m(W, \theta_0)s_{\theta_0}(W)] + o_P(1). \tag{1.2}$$

It is clear that the asymptotic distribution of $\hat{R}_n$ generally depends on that of the estimator θ_n, which complicates its approximation by bootstrap methods. But if the moment function $m(W, \theta_0)$ is orthogonal to the score s_{θ_0}, i.e., if $E[m(W, \theta_0)s_{\theta_0}(W)] = 0$, then estimation of θ_0 has no asymptotic impact on $\hat{R}_n$. Moreover, similar arguments to those used by Neyman (1959) show that a test satisfying such an orthogonality condition is optimal. See also Bickel et al. (2006).

The major innovation in our paper involves a by-product of an orthogonality condition of the form $E\left[m(W, \theta_0)s_{\theta_0}(W)\right] = 0$ that appears not to have been noticed in the existing literature. We exploit the fact that orthogonality enables a simple multiplier bootstrap approximation of the resulting test. That is, we exploit the fact that if $\{V_i\}_{i=1}^n$ is a sequence of iid random variables with zero mean, unit variance and independent of the sequence $\{W_i\}_{i=1}^n$, then by the same arguments leading to (1.2) above we have

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(W,\theta_n)V_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(W,\theta_0)V_i - \sqrt{n}\,(\theta_n-\theta_0)'\,E[m(W,\theta_0)s_{\theta_0}(W)V_i] + o_P(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(W,\theta_0)V_i + o_P(1)\\ &\to_d N(0,\Omega) \qquad (1.3)\end{aligned}$$

for $\Omega = E[m^2(W,\theta_0)]$. It follows that the orthogonality condition $E[m(W,\theta_0)s_{\theta_0}(W)] = 0$ is critical for consistency of this resampling scheme. In particular, the limiting distribution of $\hat{R}_n$ will depend on that of θ_n if $E[m(W,\theta_0)s_{\theta_0}(W)] \neq 0$.

In the present context of quantile regression, these ideas turn out to be quite important in a practical sense, given the popularity of subsampling schemes in this context. Subsampling-based methods tend both to be more computationally intensive than the multiplier bootstrap scheme described above and to be particularly sensitive to the subjective choice of the subsample size. In sum, we believe that the methodology proposed in our paper is a useful extension of existing methods for testing the linearity of conditional quantiles, both in terms of the improvement in power it offers, and also in terms of the relative simplicity of implementation it offers via a multiplier bootstrap.

A natural alternative to the omnibus testing approach that we advocate is a "directional" one in which the researcher specifies a direction of departure from the hypothesized model. For instance, consider augmenting an existing vector of regressors X with polynomial or other nonlinear transformations of certain components of X. The significance of the additional coefficients in the resulting augmented quantile regression over a relevant continuum of quantiles is interpreted as evidence against correct specification of the original linear quantile regression model in X; see Otsu (2009). The implementation of such a test can be carried out using any of the three approaches outlined in Koenker and Machado (1999). This approach is simply an extension of the classical RESET approach of Ramsey (1969) to the setting of quantile regression and shares the same limitation of inconsistency that it possesses in the original setting of mean regression (Bierens, 1982). That is, there are uncountably many misspecifications that cannot be detected with such an approach; we provide a formal discussion of this point below in Section 3.2. On the other hand, directional tests are relatively simple to apply and often have good power properties if the specified direction of departure turns out to imply a model that is "close" in an appropriate sense to the true model. As such, we believe that the omnibus and directional approaches should be viewed as complements, rather than as substitutes. We show that typical testing strategies based on the significance of coefficients corresponding to nonlinear transformations of certain regressors are not optimal against local alternatives in the sense discussed below in Section 3.2. We also derive optimal directional tests based on the same empirical process on which our proposed omnibus test is based.

The remainder of this paper is organized as follows. In Section 2 we introduce the weighted empirical processes that constitute the basis upon which the new testing procedure is developed. In Section 3.2 we study the asymptotic distribution of the proposed tests under the null as well as under fixed and local alternatives. Section 3.2 also investigates optimal directional tests and their connection with our omnibus test. Section 3.3 discusses the use of a multiplier-bootstrap technique to approximate the asymptotic distributions of test statistics under the null as well as associated issues of implementation. Section 4 summarizes the results of Monte Carlo experiments designed to assess the finite-sample performance of our proposed testing procedures. Section 5 illustrates the applicability of the tests proposed here in the context of an empirical analysis of individual age–earnings profiles using U.S. labor-market data. Mathematical proofs appear in the appendix.

Throughout this paper the symbol C is a generic constant that may change from one expression to another. The indicator of an event A is denoted by $1(A)$, i.e., $1(A) = 1$ if A occurs, and zero otherwise.

2 The Testing Procedure

We consider testing the specification of the linear quantile model given above in (1.1). As such, the null hypothesis H_0 is given by $H_0 : E\left[\psi_\alpha(W, \theta_0) | X\right] = 0$ almost surely (a.s.) for some $\theta_0(\alpha) \in \Theta$ and for all $\alpha \in \mathcal{A}$, where $W \equiv (Y, X^\top)^\top$ and

$$\psi_\alpha(W, \theta_0) \equiv 1\left(Y - X^\top \theta_0(\alpha) \leq 0\right) - \alpha. \tag{2.1}$$

The alternative H_1 is the negation of H_0. The null implies the moment restriction $E[\psi_\alpha(W, \theta_0) w(X)] = 0$ for all measurable functions $w(X)$ such that $E[|w(X)|] < \infty$ and all $\alpha \in \mathcal{A}$. Given a random sample $\{W_i \equiv (Y_i, X_i^\top)^\top\}_{i=1}^n$ of size n, it seems natural to construct test statistics based on the sample analog

$$S_n(w, \alpha) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i, \theta_n) w(X_i), \tag{2.2}$$

where $\theta_n(\alpha)$ is a $\sqrt{n}-$consistent estimator of $\theta_0(\alpha)$. An obvious example of such an estimate is the regression α-quantile developed by Koenker and Bassett (1978), henceforth referred as the QR estimator. This estimator is defined as any solution $\theta_n(\alpha)$ of

$$\arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \rho_\alpha\left(Y_i - X_i^\top \theta\right), \tag{2.3}$$

where $\rho_\alpha(u) \equiv u\left(\alpha - 1\{u \leq 0\}\right)$ is the so-called "check" function.

The null hypothesis is likely to hold when the process $S_n(w, \cdot)$ is "close" in an appropriate sense to zero for a sufficiently large number of weights w. A related approach was used by Escanciano and Velasco (2010) and has the appealing property of delivering consistent tests. This approach, however, does not acknowledge that $\theta_0(\alpha)$ is a nuisance parameter in the testing procedure and leads to tests with limiting distributions that depend on the estimator used. We provide sufficient conditions below to indicate the validity of an expansion similar to (1.2) in the present context, i.e.,

$$S_n(w, \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i, \theta_0) w(X_i) + \sqrt{n}\left(\theta_n(\alpha) - \theta_0(\alpha)\right)' E[\delta(X, \theta(\alpha)) w(X)] + o_P(1), \tag{2.4}$$

where $\delta(X_i, \theta(\alpha)) = f\left(X_i^\top \theta(\alpha) \middle| X_i\right) X_i$ and $f\left(y | X\right)$ denotes the conditional density of Y given X, evaluated at $Y = y$. We could also write the moment $E[\delta(X, \theta(\alpha)) w(X)]$ in terms of scores as

$E[\psi_\alpha(W,\theta_0)w(X)s_{\alpha,\theta_0}(W)]$, where $s_{\alpha,\theta_0}(W)=\psi_\alpha(W_i,\theta_0)\delta(X_i,\theta(\alpha))/(\alpha(1-\alpha))$, but it is more convenient to work with functions of covariates alone. To make the limiting distribution of $S_n(w,\cdot)$ invariant to that of θ_n we consider weights satisfying $E[\delta(X,\theta(\alpha))w(X)]\equiv 0$.

Power considerations suggest the use of an infinite number of weights. In particular, we note that tests based on a finite number of weights are inconsistent. These notably include tests based on the significance of additional regressors constructed by taking nonlinear transformations of existing regressors; see Section 3.2 for further details.

By an appropriate measure-theoretic argument, it can be shown that H_0 can be characterized as an uncountable number of unconditional moment restrictions given by

$$E[\psi_\alpha(W,\theta_0)1(X\leq x)]=0 \tag{2.5}$$

for $x\in\mathbb{R}^d$ and all $\alpha\in\mathcal{A}$. A joint desire to ensure the asymptotic invariance of the test to θ_n and to exploit an asymptotically increasing number of moment restrictions for the sake of power motivates consideration of tests based on functionals of the empirical process

$$\hat{R}_n(x,\alpha)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_n)\left(1(X_i\leq x)-\hat{D}_n^\top(x,\theta_n(\alpha))\hat{\Delta}_n^{-1}(\alpha)\hat{\delta}_{i\alpha}\right), \tag{2.6}$$

where $\hat{D}_n(x,\theta_n(\alpha))=n^{-1}\sum_{i=1}^n\hat{\delta}_{i\alpha}1(X_i\leq x)$; $\hat{\Delta}_n(\alpha)=n^{-1}\sum_{i=1}^n\hat{\delta}_{i\alpha}\hat{\delta}_{i\alpha}^\top$ and

$$\hat{\delta}_{i\alpha}=\hat{\delta}_n(X_i,\theta_n(\alpha))\equiv\hat{f}_{i\alpha}X_i, \tag{2.7}$$

where $\hat{f}_{i\alpha}=\hat{f}_h\left(X_i^\top\theta_n(\alpha)\middle|X_i\right)$ denotes a consistent estimator of $f\left(X_i^\top\theta(\alpha)\middle|X_i\right)$. A convenient choice of density estimator was recently proposed by Escanciano and Goh (2012), and is given by

$$\hat{f}_{i\alpha}=\hat{f}_h\left(X_i^\top\theta_n(\alpha)\middle|X_i\right)\equiv\frac{1}{mh_m}\sum_{j=1}^{m}K\left(\frac{X_i^\top\theta_n(\alpha)-X_i^\top\theta_n(\alpha_j)}{h_m}\right), \tag{2.8}$$

where $K(\cdot)$ denotes a kernel function, $\{\alpha_j\}_{j=1}^m$ is a random sample from the uniform distribution in $\mathcal{A}$ with $m\to\infty$ as $n\to\infty$, and h_m denotes a possibly data-dependent bandwidth parameter satisfying regularity conditions described below. The conditional density estimator in (2.8), whose form is directly motivated by the restrictions imposed by the null hypothesis, has a critical advantage over alternative conditional density estimators in Rosenblatt (1969). In particular, as shown in Escanciano and Goh (2012), the rate of convergence of $\hat{f}_h\left(y|X_i\right)$ corresponds to that of the Rosenblatt estimator with univariate regressors, regardless of the dimension of X. Furthermore, the estimator does not have random denominators, which, in addition to simplifying the analysis, results in test statistics less empirically sensitive to the choice of bandwidth than would otherwise be the case.

In what follows, we proceed to investigate tests based on continuous functionals of the empirical process $\hat{R}_n(\cdot,\cdot)$ as given above in (2.6). An example that we use in simulations and the empirical example toward the end of this paper is the Cramér–von Mises-type functional, i.e.,

$$CvM_n\equiv\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}\left|\hat{R}_n(X_i,\alpha_j)\right|^2, \tag{2.9}$$

where $\{\alpha_j\}_{j=1}^m$ is the sample generated in the estimation of $f\left(X_i^\top\theta(\alpha)\middle|\, X_i\right)$. This test statistic is computed as $CvM_n = m^{-1}n^{-2}\sum_{j=1}^m \psi_{jn}^\top \hat{H}_j G G^\top \hat{H}_j^\top \psi_{jn}$, where $\psi_{jn} = (\psi_{\alpha_j}(W_1, \theta_n(\alpha_j)), \ldots, \psi_{\alpha_j}(W_n, \theta_n(\alpha_j)))^\top$. Here $\hat{H}_j \equiv I_n - \hat{\delta}_{\cdot j}\left(\hat{\delta}_{\cdot j}^\top \hat{\delta}_{\cdot j}\right)^{-1}\hat{\delta}_{\cdot j}$, where I_n denotes the $n\times n$ identity matrix and $\hat{\delta}_{\cdot j}$ is an $n \times p$ matrix with ith row denoted by $\hat{\delta}_{ij}\left(X_i, \theta_n(\alpha_j)\right) \equiv \hat{f}_{i\alpha_j} X_i$, where $\hat{f}_{i\alpha}$ is as given above in (2.8). Finally, G denotes the $n \times n$ matrix with (i,j)th element given by $g_{ij} \equiv 1\left(X_i \le X_j\right)$. Our proposed test rejects for realized values of CvM_n that take values in the right tail of its asymptotic null distribution, which is developed below in Section 3.1.

3 Asymptotic Results

In what follows we establish the limiting distribution of the quantile-weighted empirical process $\hat{R}_n(\cdot,\cdot)$ given above in (2.6) under both the null hypothesis H_0 and under sequences of alternatives local to H_0 in an appropriate sense. We show how optimal directional tests can be constructed based on $\hat{R}_n(\cdot,\cdot)$. In addition, we establish the consistency of a multiplier bootstrap approximation to the asymptotic null distribution of continuous functionals of $\hat{R}_n(\cdot,\cdot)$.

We show in Theorem 3.1 below that the empirical process $\hat{R}_n(x,\alpha)$ given above in (2.6) is uniformly asymptotically equivalent, with respect to the supremum norm on $\mathbb{R}^d \times \mathcal{A}$, to the process

$$R_{n0}(x,\alpha) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\alpha(W_i,\theta_0)\Pi_\alpha 1(X_i \le x), \tag{3.1}$$

where, henceforth, for a measurable function $w(X)$ we define the orthogonal projection operator $\Pi_\alpha w(X) \equiv w(X_i) - D^\top(w,\theta_0(\alpha))\Delta^{-1}(\alpha)\delta(X_i,\theta_0(\alpha))$, where $\delta(X_i,\theta_0(\alpha)) = f\left(X^\top\theta_0(\alpha)\middle|\, X\right)X_i^\top$,

$$D(w,\theta_0(\alpha)) = E\left[\delta\left(X,\theta_0(\alpha)\right)w(X)\right] \tag{3.2}$$

and where

$$\Delta(\alpha) = E\left[\delta\left(X,\theta_0(\alpha)\right)\delta^\top\left(X,\theta_0(\alpha)\right)\right]. \tag{3.3}$$

In Section 3.2 below we show that any test based on $\hat{R}_n(x,\alpha)$ has improved power over tests based on the process $S_n(x,\alpha) \equiv S_n(1\left(\cdot \le x\right),\alpha)$ given above in (2.2) when confronted with $\sqrt{n}$-local alternatives. In addition, the fact that the process $R_{n0}(x,\alpha)$ does not depend asymptotically on $\theta_n(\alpha)$ enables a convenient multiplier bootstrap scheme for simulating critical values, as we show below in Section 3.3.

3.1 Asymptotic null distribution

The limiting null distributions of the tests proposed here are the limiting distributions of continuous functionals of $\hat{R}_n(\cdot,\cdot)$ under H_0. To derive asymptotic results we consider the following notation and assumptions. For a generic function $g : \mathcal{Z} \to \mathbb{R}$, define $\|g\|_{\mathcal{Z}} \equiv \sup_{z\in\mathcal{Z}}|g(z)|$. We study the weak convergence of $\hat{R}_n(\cdot,\cdot)$ and related processes as elements of $l^\infty(\mathcal{T})$, the space of real-valued functions that are uniformly bounded on $\mathcal{T}$, where $\mathcal{T} \equiv [-\infty,+\infty]^d \times \mathcal{A}$. The space $l^\infty(\mathcal{T})$ is furnished with the supremum norm $\|\cdot\|_{\mathcal{T}}$; let $\mathcal{B}_{d_\infty}$ denote the corresponding Borel σ-algebra. Let $\Rightarrow$ denote weak convergence on $(l^\infty(\mathcal{T}), \mathcal{B}_{d_\infty})$ in the sense of Hoffmann–Jørgensen. See e.g., Dudley (1999, p. 94) or van der Vaart and Wellner (1996, Definition 1.3.3). Let $\mathcal{X}_X$ denote the support of X and define $\mathcal{Y}_\mathcal{A} = \{x^\top \theta_0(\alpha) : x \in \mathcal{X}_X, \alpha \in \mathcal{A}\}$.

Regularity conditions underlying the derivation of the asymptotic distribution of our proposed test statistics under the null are given as follows.

Assumption 1 (Data-generating process) *(i) $\left\{W_i \equiv (X_i^\top, Y_i)^\top\right\}_{i=1}^n$ is a sequence of iid random $(d+1)$-variates; (ii) the conditional densities $\left\{f\left(y|\, x\right) :\, x \in \mathbb{R}^d\right\}$ are uniformly bounded, from above and below on $y \in \mathcal{Y}_\mathcal{A}$ and $x \in \mathcal{X}_X$, with uniformly bounded derivatives with respect to y; (iii) for all $\alpha \in \mathcal{A}$, the matrix $\Delta(\theta(\alpha)) \equiv E\left[\delta(X, \theta(\alpha))\delta(X, \theta(\alpha))^\top\right]$ is nonsingular in a neighborhood of $\theta(\alpha) = \theta_0(\alpha)$; and (iv) $E[\|X\|^2] < \infty$.*

Assumption 2 (Estimator of the nuisance parameter) *The following conditions are satisfied: (i) Θ is compact and $\theta_0(\alpha)$ belongs to its interior; and (ii) $\|\theta_n - \theta_0\|_\mathcal{A} = o_P(n^{-1/2})$.*

Assumption 3 (Kernel and bandwidth) *(a) $K(u)$ satisfies the following conditions: (i) $K(\cdot)$ is of bounded variation; (ii) $\|K\|_\infty \equiv \sup_u |K(u)| < \infty$; (iii) $\int_{-\infty}^{\infty} K(u)du = 1$; (iv) $K(\cdot)$ satisfies a Lipschitz condition on $\mathbb{R}$; and (v) $K(\cdot)$ is of second order, i.e., $\int_{-\infty}^{\infty} uK(u) = 0$, $\int_{-\infty}^{\infty} u^2K(u) = \mu_{2K}$ for some $\mu_{2K} \in (0, \infty)$ and $\int_{-\infty}^{\infty} (K(u))^2\, du = B$ for some $B \in (0, \infty)$. (b) The possibly data-dependent bandwidth h satisfies $P(a_m \le h \le b_m) \to 1$, for deterministic sequences of positive numbers a_m and b_m such that $b_m \to 0$ and $a_m^{d+1} m / \log m \to \infty$, as $n \to \infty$.*

Assumption 1 is standard. The regression quantile estimator of Koenker and Bassett (1978) satisfies Assumption 2. We note that Assumption 3 allows for the use of the most popular smoothing kernels in empirical practice, including in particular the Gaussian kernel. Data-driven bandwidths are also encompassed by Assumption 3.

We establish the asymptotic distribution of $\hat{R}_n(\cdot, \cdot)$ under H_0. The proof, which is given in Appendix A.2, proceeds in two steps. The first step involves showing that $\hat{R}_n(\cdot, \cdot)$ is asymptotically equivalent under H_0 to the process $R_{n0}(\cdot, \cdot)$ as given above in (3.1). In the second step we analyze the weak convergence under H_0 of the process $R_{n0}(\cdot, \cdot)$.

Theorem 3.1 *Let Assumptions 1–3 hold. Then, under H_0, we have $\left\|\hat{R}_n - R_{n0}\right\|_\mathcal{T} = o_P(1)$.*

Theorem 3.1 indicates that $\hat{R}_n(\cdot, \cdot)$ behaves like $R_{n0}(\cdot, \cdot)$ in large samples; in particular, the limiting behavior of $\hat{R}_n(\cdot, \cdot)$ does not depend on θ_n. The limiting distribution of $\hat{R}_n(x, \alpha)$ is given in the following result.

Corollary 3.1 *Under the conditions of Theorem 3.1 we have $\hat{R}_n \Rightarrow R_\infty$, where R_∞ denotes a Gaussian process with mean zero and covariance function $(\min\{\alpha_1, \alpha_2\} - \alpha_1\alpha_2) \cdot E\left[\Pi_{\alpha_1} 1(X_i \le x_1)\Pi_{\alpha_2} 1(X_i \le x_2)\right]$.*

Now consider the Cramér–von Mises statistic based on $\hat{R}_n(\cdot, \cdot)$ as given above by CvM_n in (2.9). It follows from Corollary 3.1, the Continuous Mapping Theorem and the Glivenko–Cantelli Theorem that the asymptotic distribution of CvM_n is characterized by the convergence $CvM_n \Rightarrow \int_{\mathcal{T}} |R_\infty(x,\alpha)|^2 \, dF_X(x) d\alpha$, where F_X denotes the marginal distribution of X. The justification for replacing the empirical distributions of $\{X_i\}_{i=1}^n$ and $\{\alpha_j\}_{j=1}^m$ by the limiting distributions follows from the arguments in Chang (1990).

3.2 Asymptotic power and directional tests

We proceed to consider the power properties of tests based on continuous functionals of $\hat{R}_n(\cdot,\cdot)$. For purposes of comparison, we also investigate optimal directional tests and their relation to the omnibus test that we propose.

We first derive the asymptotic distribution of $\hat{R}_n(\cdot,\cdot)$ under a certain sequence of local alternatives converging to the null at a parametric rate $n^{-1/2}$. In particular, we consider the data-generating process for the sequence of local alternatives ($c \equiv c_n = n^{-1/2}$) given by

$$H_{1,c}(b): \; E\left[\psi_\alpha\left(Y - X^\top \theta_0(\alpha)\right) \middle| X\right] = cb(X,\alpha) + r_n(X,\alpha) \text{ a.s.}, \tag{3.4}$$

for some $\theta_0(\alpha) \in \Theta$ and all $\alpha \in \mathcal{A}$. We require the functions $b(\cdot,\cdot)$ and $r_n(\cdot,\cdot)$ to satisfy the conditions of the following assumption.

Assumption 4 (Local alternatives) *For each $n = 1,2,3,\ldots$; $\{W_{i,n} \equiv (X_{i,n}^\top, Y_{i,n})^\top\}_{i=1}^n$ is a sequence of arrays of iid variables satisfying (3.4) such that:*

1. *$E\left[\sup_{\alpha\in\mathcal{A}} |b(X,\alpha)|\right] < \infty$ and $E\left[\sqrt{n}\sup_{\alpha\in\mathcal{A}} |r_n(X,\alpha)|\right] = o(1)$.*
2. *The function $b(X,\cdot)$ is continuous in $\mathcal{A}$, a.s.*

The "non-centrality parameter" under sequences of local alternatives given by $H_{1,n}(b)$ in (3.4) above takes the form

$$D_b(x,\alpha) \equiv E\left[b(X,\alpha)\Pi_\alpha 1(X_i \le x)\right], \tag{3.5}$$

as indicated by the following result. Note that $D_b(x,\alpha) \equiv 0$ if $b(X,\alpha)$ is proportional to (i.e., collinear with) $\delta\left(X, \theta_0(\alpha)\right)$. We implicitly assume that the uniformity conditions in Assumption 1 are also uniform in n, e.g. the conditional density f is uniformly bounded in $n \ge 1$. For simplicity, we also assume that the distribution of $\{X_{i,n}\}_{i=1}^n$ does not vary with n under the local alternatives $H_{1,n}(b) \equiv H_{1,c_n}(b)$.

Theorem 3.2 *Under the conditions of Assumptions 1–4 and under the sequence of local alternatives given by (3.4), we have*

$$\hat{R}_n \Rightarrow R_\infty + D_b, \tag{3.6}$$

where R_∞ is as given above in the statement of Corollary 3.1, and where D_b is as given above in (3.5).

We now discuss optimal directional tests for testing $H_0: \; c = 0$ vs $H_1: \; c \neq 0$ in (3.4) for a fixed $\alpha \in \mathcal{A}$, where an optimal test is defined as one that is asymptotically most powerful and unbiased (AUMPU) as described in Choi et al. (1996). Some standard computations (e.g., Bickel et al. 2006, Section 2.2) show that the optimal test of $H_0 : c = 0$ against $H_1 : c \neq 0$ rejects for large absolute values of $n^{-1/2}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_0)\,\Pi_\alpha b(X,\alpha) = \int b(x,\alpha)R_{n0}(dx,\alpha)$. The feasible optimal test replaces $R_{n0}(\cdot,\cdot)$ by $\hat{R}_n(\cdot,\cdot)$ and uses a consistent estimate of the efficient information. From our results it follows that the feasible test has the same limiting distribution as the infeasible one. Hence, optimal directional tests can be easily constructed as empirical integrals of our basic process $\hat{R}_n(\cdot,\cdot)$.

We proceed to show that the naive testing strategy based on verifying the significance of an additional regressor constructed via a nonlinear transformation of an existing regressor is not optimal. This is shown in the context of the following simple example. For $(X,Y)^\top$ taking values in $\mathbb{R}^2$, fix $\alpha \in \mathcal{A}$ and suppose that the conditional distribution of Y given X satisfies the expansion $P\left[1\left(Y \le \theta_{01}(\alpha) + X\theta_{02}(\alpha) + c_nX^2\right)\middle| X\right] = \alpha + c_nX^2 f\left(\theta_{01}(\alpha) + X\theta_{02}(\alpha)\middle| X\right) + r_n(X,\alpha)$, where $c_n \downarrow 0$ and where $\sqrt{n}\sup_{\alpha\in\mathcal{A}}|r_n(X,\alpha)| = o(c_n)$. The optimal test for $H_0 : c_n = 0$ against $H_1 : c_n \neq 0$ rejects for large absolute values of

$$\begin{aligned}&\int x^2 \hat{f}_h\left(\theta_{n1}(\alpha) + x\theta_{n2}(\alpha)\middle| x\right)\hat{R}_n(dx,\alpha)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_n)\left(X_i^2\hat{f}_{i\alpha} - \hat{D}_n^\top\hat{\Delta}_n^{-1}(\alpha)\hat{\delta}_{i\alpha}\right),\end{aligned} \tag{3.7}$$

where $\theta_n(\alpha) = (\theta_{n1}(\alpha),\theta_{n2}(\alpha))^\top$ and where $\hat{D}_n = n^{-1}\sum_{i=1}^{n}\hat{\delta}_{i\alpha}X_i^2\hat{f}_{i\alpha}$.

Let $\tilde{\theta}_n(\alpha) \equiv (\theta_{n1}(\alpha),\theta_{n2}(\alpha),\theta_{n3}(\alpha))^\top$ be the QR estimator under the alternative, and set $e_3 = (0,0,1)^\top$. It is well known (e.g., Koenker, 2005) that the QR estimator satisfies a Bahadur expansion of the form $\sqrt{n}\left(\tilde{\theta}_n(\alpha) - \theta_0(\alpha)\right) = n^{-1/2}\sum_{i=1}^{n}D^{-1}(X,\theta_0(\alpha))X_i\psi_\alpha(W_i,\theta_0) + o_p(1)$, where the matrix $D(X,\theta_0(\alpha))$ is as given above in (3.2).

The significance test rejects for large absolute values of $e_3^\top\tilde{\theta}_n(\alpha)$, and as such is not optimal for a test of $H_0 :\ c_n = 0$ against the two-sided alternative. This is analogous to the case of the classical t-test under conditional heteroskedasticity. The t-test is optimal only if it is based on an efficient estimator. In contrast, the optimal directional test that rejects for large values of the quantity given above in (3.7) does not require an efficient estimator of the nuisance parameter.

Consider now the global power properties of tests based on $\hat{R}_n(\cdot,\cdot)$, including both omnibus and directional variants. We have from the proof of Theorem 3.1 that under any fixed alternative,

$$\frac{1}{\sqrt{n}}\hat{R}_n(x,\alpha) = E\left[\psi_\alpha(W,\theta_0)\left(1(X \le x) - D^\top(x,\theta_0(\alpha))\Delta^{-1}(\alpha)\delta\left(X,\theta_0(\alpha)\right)\right)\right] + o_P(1) \tag{3.8}$$

uniformly in $x \in \mathbb{R}^d$ and $\alpha \in \mathcal{A}$. The interpretation of θ_0 in the case of the QR estimator is investigated in Angrist et al. (2006). In this leading example, the first order conditions of the estimator imply $E\left[\psi_\alpha(W,\theta_0)\delta\left(X,\theta_0(\alpha)\right)\right] = 0$ for all $\alpha \in \mathcal{A}$. It follows that our test is consistent by the equivalence of (2.5) and the null hypothesis.

Similarly, directional tests are shown to be inconsistent, since there are directions of departure

$E\left[\psi_\alpha\left(Y - X^\top\theta_0(\alpha)\right)\middle| X\right] \neq 0$ such that

$$\frac{1}{\sqrt{n}}\int b(x,\alpha)\hat{R}_n(dx,\alpha) = E\left[\psi_\alpha(W,\theta_0)b(X,\alpha)\right] + o_P(1)$$
$$= E\left[E\left[\psi_\alpha\left(Y - X^\top\theta_0(\alpha)\right)\middle| X\right]b(X,\alpha)\right] + o_P(1). \tag{3.9}$$

From this it follows that directional tests are inconsistent against directions $E\left[\psi_\alpha\left(Y - X^\top\theta_0(\alpha)\right)\middle| X\right]$ that are orthogonal to $b(X,\alpha)$, i.e. directions satisfying $E\left[\psi_\alpha(W,\theta_0)b(X,\alpha)\right] = 0$.

In summary, it is possible to define omnibus and optimal directional tests as functionals of the basic process $\hat{R}_n(\cdot,\cdot)$. Omnibus tests have non-trivial local power against all alternatives not collinear to the score $\delta\left(X,\theta_0(\alpha)\right)$ and are also consistent against all fixed alternatives. In contrast, directional tests can be constructed to have optimal power against alternatives characterized by a user-specified direction. At the same time, directional tests are non-robust in the sense of being inconsistent against alternatives that are orthogonal to the specific directional alternatives against which they are constructed to be optimal. Further empirical evidence of this is provided below in Section 4.

3.3 Multiplier bootstrap approximation

The results presented above in Section 3.1 have shown that the asymptotic null distribution of continuous functionals of $\hat{R}_n(\cdot,\cdot)$ and $S_n(\cdot,\cdot)$ are liable to depend in a complex way on the underlying data-generating process. It follows from this that critical values for test statistics based on continuous functionals of $\hat{R}_n(\cdot,\cdot)$ or $S_n(\cdot,\cdot)$ cannot in general be tabulated for more than a few special cases.

In this section we overcome this problem in the case of functionals of the form $T(\hat{R}_n)$ with the assistance of a multiplier bootstrap. This particular resampling scheme has the practical advantage of not requiring the computation of new parameter estimates at each bootstrap replication, unlike the resampling schemes proposed by, e.g., Hahn (1995), Horowitz (1998), Bilias et al. (2000), Sakov and Bickel (2000), He and Hu (2002), Chernozhukov and Fernández-Val (2005), Whang (2006a) and Escanciano and Velasco (2010), among many others. The multiplier bootstrap that we propose is also practically advantageous in that it does not involve any need to select tuning parameters, apart from the number of bootstrap replications. This is in contrast to approaches based on subsampling, where inferences are in general sensitive to a researcher's subjective choice of subsample size.

The method advocated in this section involves approximating the asymptotic distribution of a continuous functional $T(\hat{R}_n)$ with that of $T(\hat{R}_n^*)$, where $\hat{R}_n^*(\cdot,\cdot)$ is a simple multiplier-bootstrap approximation of $\hat{R}_n(\cdot,\cdot)$ given by $\hat{R}_n^*(x,\alpha) \equiv n^{-1/2}\sum_{i=1}^n \psi_\alpha(W_i,\theta_n(\alpha))\left(1(X_i \leq x) - \hat{D}_n^\top(x,\theta_n(\alpha))\hat{\Delta}_n^{-1}(\alpha)\hat{\delta}_{i\alpha}\right)V_i$, where $\{V_i\}_{i=1}^n$ is a sequence of iid random variables with zero mean, unit variance and bounded support and that are also independent of the sequence $\{W_i\}_{i=1}^n$. An example involves iid Bernoulli variates $\{V_i\}$ with

$$P\left(V = 1 - \varphi\right) = \varphi/\sqrt{5} \tag{3.10}$$
$$P\left(V = \varphi\right) = 1 - \varphi/\sqrt{5}, \tag{3.11}$$

where φ is the so-called "golden ratio", i.e. $\varphi = \left(\sqrt{5}+1\right)/2$. See Mammen (1993) for motivation on the appearance of φ in (3.10)–(3.11).

The bootstrap empirical distribution of $T(\hat{R}_n^*)$, i.e., $\hat{F}_n^*\left(t|\left\{W_i\right\}_{i=1}^n\right) = P\left[T\left(\hat{R}_n^*\right) \leq t\Big| \left\{W_i\right\}_{i=1}^n\right]$ is shown to be a consistent estimate of the asymptotic null distribution function of $T(\hat{R}_n)$, i.e., $F_\infty(t) = P\left[T\left(R_\infty\right) \leq t\right].$ In this case, the null hypothesis will be rejected at the τ-level of significance when $T\left(\hat{R}_n\right) \geq c_{n,\tau}^*$, where $c_{n,\tau}^*$ is such that $\hat{F}_n^*\left(c_{n,\tau}^*\big| \left\{W_i\right\}_{i=1}^n\right) = 1-\tau$.

It is also possible to use bootstrap p-values in this context. For example, the null could be rejected whenever $p_n^* < \tau$, where $p_n^* = P\left[T\left(\hat{R}_n^*\right) \geq T\left(\hat{R}_n\right)\Big| \left\{W_i\right\}_{i=1}^n\right]$. The bootstrap-based test based on $T(\hat{R}_n^*)$ is clearly valid if $\hat{F}_n^*$ is a consistent estimator of F_∞ at each continuity point of F_∞. In the case of consistency a.s., we use the notation $T\left(\hat{R}_n^*\right) \xrightarrow{d} T\left(R_\infty\right)$ a.s. See van der Vaart and Wellner (1996) for details.

Theorem 3.3 *Suppose the conditions of Assumptions 1 to 3 hold. Then for any continuous functional* $T(\cdot)$, $T\left(\hat{R}_n^*\right) \xrightarrow{d} T\left(R_\infty\right)$ *a.s.*

It is straightforward to show that Theorem 3.3 implies the consistency of our multiplier-bootstrap test against all alternatives not collinear with the score, provided that T is such that $T(f) = 0 \Leftrightarrow f = 0$ a.s. Moreover, it can be proved that our bootstrap-based test preserves the asymptotic local power properties of $T(\hat{R}_n)$. Details have been omitted in order to economize on space.

Finally, it should be noted that the multiplier-bootstrap scheme described here is not applicable to tests based on functionals of the form $T(S_n)$, where $S_n(\cdot,\cdot)$ is the non-projected empirical process given in (2.2). The reason for this was noted in the introduction, namely, the dependence of the corresponding limiting process $S_\infty(\cdot,\cdot)$ on the asymptotic distribution of $\theta_n(\cdot)$. From this it follows that the same multiplier-bootstrap scheme described above with reference to functionals of the form $T(\hat{R}_n)$ under H_0 will not be consistent when applied to functionals of the form $T(S_n)$ under H_0 (e.g., Beran and Ducharme, 1991). If critical values for tests based on $T(S_n)$ are required, a more conventional approach based on subsampling is recommended (e.g., Chernozhukov and Fernández-Val, 2005). An example is presented in the context of the simulation experiments discussed in the following section of the paper.

4 Numerical Evidence

This section presents the results of a series of Monte Carlo experiments designed to evaluate the finite-sample performance of a Cramér–von Mises test of the linearity hypothesis. In particular, the performance of the test based on the empirical process $\hat{R}_n(\cdot,\cdot)$ given above in (2.6), with critical values simulated using the multiplier bootstrap procedure described in Section 3.3, is compared to the same test based on the more conventional process $S_n(\cdot,\cdot)$ given above in (2.2) with $1(X_i \leq x)$ replacing $w(X_i)$ and where the critical values are simulated by subsampling. In addition, we also compare the performance of our proposed omnibus testing approach based on the empirical process $\hat{R}_n(\cdot,\cdot)$ with a "directional" test based on the significance of a nonlinear transformation of an existing regressor in an augmented quantile regression.

We first evaluate the sensitivity of the rejection rate under the null of the test based on $\hat{R}_n(\cdot,\cdot)$ to the bandwidth h used to implement the conditional density estimator embedded in the specification of the underlying testing process. This is compared to the empirical probabilities of rejection under the null of the same test based on the empirically infeasible process $R_{n0}(\cdot,\cdot)$ given above in (3.1); from this, one can get an idea of the overall effect of the bandwidth on the probability of type-I error. In addition, we compare the power performance of the tests based on the three empirical processes $\hat{R}_n(\cdot,\cdot)$, $R_{n0}(\cdot,\cdot)$ and $S_n(\cdot,\cdot)$. This allows for an evaluation of the finite-sample benefits of the orthogonal projection embedded in $\hat{R}_n(\cdot,\cdot)$ and $R_{n0}(\cdot,\cdot)$.

We consider the data-generating process given by

$$Y_i = X_{1i} + X_{2i} + c\sigma_i^{\frac{3}{2}} + u_i,\ i = 1,\dots,n; \tag{4.1}$$

where $\sigma_i \equiv X_{1i}^2 + X_{2i}^2 + X_{1i}X_{2i}$. The random variables X_{1i}, X_{2i} and u_i are taken to be iid $N(0,1)$ and mutually independent. The hypothesis given above in (1.1) corresponds to the model given in (4.1) with $c = 0$. It follows that the model for the conditional quantile function under the null is simply the quantile-regression model given by $F^{-1}_{Y_i|X_i}(\alpha) = \tilde{X}_i^\top \theta_0(\alpha)$, where $\alpha \in \mathcal{A}$, $\tilde{X}_i = (1, X_{1i}, X_{2i})^\top$, $\theta_0(\alpha) = (\Phi^{-1}(\alpha), 1, 1)^\top$ and where $\Phi^{-1}(\alpha)$ is the quantile function of a standard normal random variable.

We used a sample size of $n = 100$ and considered a subinterval of quantiles given by $\mathcal{A} = [0.1, 0.9]$. The number of Monte Carlo replications was set to 1000, and the number of sequences of bootstrap multipliers generated for each Monte Carlo replication was set to 200.

The estimator $\theta_n(\alpha)$ of the null parameter vector will be taken throughout to be the regression α-quantile of Koenker and Bassett (1978). The asymptotic properties of this estimator have been extensively investigated. In particular, it is known to satisfy the conditions of Assumption 2; see e.g., Gutenbrunner and Jurečková (1992, Theorem 1).

We consider an approximation of the Cramér–von Mises test defined in (2.9) over a grid of $m = 30$ evenly spaced points in the interval $[0.1, 0.9]$. Denote by $\{\alpha_j\}_{j=1}^m$ the points in the grid, with $0.1 = \alpha_1 < \cdots < \alpha_m = 0.9$. As mentioned above, we considered three versions of the Cramér–von Mises test. Each version of the test is based on one of the three empirical processes $\hat{R}_n(\cdot,\cdot)$, $R_{n0}(\cdot,\cdot)$ and $S_n(\cdot,\cdot)$ as given above in (2.6), (3.1) and (2.2), respectively.

We note that the Cramér–von Mises test statistic based on the empirically feasible testing process $\hat{R}_n(\cdot,\cdot)$ can be computed as in (2.9). The bootstrap analog of the Cramér–von Mises test statistic CvM_n is denoted by $CvM_n^* \equiv m^{-1}\sum_{j=1}^m n^{-2}\psi_{jn}^{*\top}\hat{H}_j GG^\top \hat{H}_j^\top \psi_{jn}^*$, where $\psi_{jn}^* = (V_1\psi_{\alpha_j}(W_1, \theta_n(\alpha_j)), \dots, V_n\psi_{\alpha_j}(W_n, \theta_n(\alpha_j))$ and where $V_1, \dots, V_n$ are iid Bernoulli random variates generated in accordance with the scheme given above in (3.10)–(3.11). The sampling distribution of CvM_n^* may be simulated, and the sampling distribution of CvM_n accordingly estimated, by generating many replications of the multiplier sequence $\{V_i\}$ and computing CvM_n^* once for each of the replicated multiplier sequences.

On the other hand, the Cramér–von Mises test statistic based on the empirically infeasible testing process $R_{n0}(\cdot,\cdot)$ is given by $CvM_{n0} \equiv m^{-1}\sum_{j=1}^m n^{-2}\psi_j^\top H_j GG^\top H_j^\top \psi_j$, where for $j = 1,\dots,m$, $\psi_j = (\psi_{\alpha_j}(W_1, \theta_0(\alpha_j)), \dots, \psi_{\alpha_j}(W_n, \theta_0(\alpha_j)))^\top$. Here $H_j \equiv I_n - \delta_j(\delta_j^\top\delta_j)^{-1}\delta_j^\top$, where δ_j is the $n \times 3$ matrix whose ith row is given by $\delta(X_i, \theta_0(\alpha_j))^\top = \phi\left(\Phi^{-1}(\alpha_j)\right)\phi\left(X_{1i}\right)\phi\left(X_{2i}\right)X_i^\top$. The bootstrap approximation to CvM_{n0} is *mutatis mutandis* the same as its empirically feasible counterpart CvM_n^*.

The performance of tests of the linearity hypothesis based on CvM_n and CvM_{n0} are compared to Cramér–von Mises tests based on the naive process $S_n(\cdot, \cdot)$ given above in (2.2). A test of this form is based on the statistic $CvM_{nb0} \equiv m^{-1} \sum_{j=1}^{m} n^{-2} \psi_j^\top G G^\top \psi_j$.

Recall from the discussion at the end of Section 3.3 that the multiplier bootstrap scheme is inapplicable to tests based on the process $S_n(\cdot, \cdot)$. As such, critical values for the Cramér–von Mises test based on $S_n(\cdot, \cdot)$ are simulated by subsampling (e.g., Chernozhukov and Fernández-Val, 2005; Escanciano and Velasco, 2010). In particular, 200 subsamples were generated for each iteration of the Monte Carlo algorithm. The size of each subsample was determined with reference to the suggestion of Sakov and Bickel (2000) applied to the m-out-of-n bootstrap applied to the distribution of univariate sample quantiles; in particular, the subsample size was set to $b = \lfloor \lambda n^{2/5} \rfloor$, where $\lambda > 0$. We experimented with a number of different settings for the leading constant λ and report in Tables 1 and 2 below only the results for the value of λ leading to the most accurately sized test, which for a sample of $n = 100$ yields an optimal subsample size of $b = 18$. Naturally, it is empirically infeasible to choose the subsampling size optimally in this way as the true data-generating process is unknown.

We also consider the relative small-sample performance of a "directional" approach to testing the hypothesis $H_0 : c = 0$ in the context of the data-generating process given above in (4.1). In particular, for $\alpha \in \mathcal{A}$, consider the model

$$F^{-1}_{Y_i|X_{1i},X_{2i}}(\alpha) = \beta_0(\alpha) + X_{1i}\beta_1(\alpha) + X_{2i}\beta_2(\alpha) + X_{1i}^2\beta_3(\alpha),\ i = 1, \ldots, n. \tag{4.2}$$

The particular directional testing approach to a test of $H_0 : c = 0$ in (4.1) above involves the sampling behavior of a Cramér–von Mises functional applied to a Wald statistic for testing the hypothesis $H_0^{(D)} : \beta_3(\cdot) \equiv 0$ in the context of the model in (4.2). Using the same lattice $\{\alpha_j\}$ considered above of $m = 30$ evenly-spaced quantiles in the interval $\mathcal{A} = [0.1, 0.9]$, we consider the test statistic

$$\begin{aligned} &CvM_{nbD} \\ &\equiv \frac{n}{m} \sum_{j=1}^{m} \Big(R\hat{\beta}_n(\alpha_j) - R\beta(\alpha_j) \Big)^\top \Big[\alpha_j (1 - \alpha_j) R \hat{D}_{n1}^{-1}(\alpha_j) D_{n0} \hat{D}_{n1}^{-1}(\alpha_j) R^\top \Big]^{-1} \Big(R\hat{\beta}_n(\alpha_j) - R\beta(\alpha_j) \Big), \end{aligned} \tag{4.3}$$

where $R = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$, $\beta(\alpha)$ is the vector of parameters of the quantile-regression model appearing above in (4.4), $\hat{\beta}_n(\alpha)$ is the regression α-quantile estimator of $\beta(\alpha)$, and for $\tilde{X}_i^\dagger \equiv \begin{pmatrix} 1 & X_{1i} & X_{2i} & X_{1i}^2 \end{pmatrix}^\top$, $D_{n0} = n^{-1} \sum_{i=1}^{n} \tilde{X}_i^\dagger \tilde{X}_i^{\dagger\top}$ and $\hat{D}_{n1}(\alpha) = n^{-1} \sum_{i=1}^{n} \tilde{X}_i^\dagger \tilde{X}_i^{\dagger\top} \hat{f}_{i\alpha}$, where $\hat{f}_{i\alpha}$ denotes the conditional density estimator given above in (2.8). In this connection, the density estimates $\hat{f}_{i\alpha_j}$ were implemented using the bandwidth $h = m^{-1/5}$.

Following Chernozhukov and Fernández-Val (2005), the sampling distribution of CvM_{nbD} in (4.3) is estimated by subsampling. Following the earlier application of subsampling to the omnibus test based on the process $S_n(\cdot, \cdot)$, we again experiment with various subsample sizes to find the one that yielded the most accurately sized test, relative to a nominal level of 5%. This empirically infeasible process of experimentation resulted in our using a subsample size of $b = 23$. We used 200 subsamples of size $b = 23$ for each of the 1000 Monte Carlo replications considered to generate a usable approximation of the sampling distribution of the subsampled bootstrap analog of CvM_{nbD}.

Table 1 displays rejection probabilities under the null of Cramér–von Mises tests based on $\hat{R}_n(\cdot,\cdot)$, $R_{n0}(\cdot,\cdot)$ and $S_n(\cdot,\cdot)$ and also for the directional approach based on the statistic CvM_{nbD} given above in (4.3). In the context of Table 1, the shift parameter c appearing in the specification of the maintained data-generating process as given in (4.1) above was set to zero. The nominal size of these tests was set to 5%. The test based on $\hat{R}_n(\cdot,\cdot)$ was implemented using three different settings of the bandwidth h used to construct the conditional density estimator $\hat{f}_{i\alpha_j}$ appearing above in (2.8). In particular, the settings $h = \kappa n^{-1/5}$ ($k = .5, 1, 1.5$) were used. It is seen that the empirical rejection probabilities of the test based on $\hat{R}_n(\cdot,\cdot)$ do not exhibit much sensitivity to the particular bandwidth setting used.

Table 2 focuses on the power performance of the three Cramér–von Mises tests based on $\hat{R}_n(\cdot,\cdot)$, $R_{n0}(\cdot,\cdot)$ and $S_n(\cdot,\cdot)$ and also for the directional test based on CvM_{nbD} as given above in (4.3). The nominal size of these tests was again set to 5%. The test based on $\hat{R}_n(\cdot,\cdot)$ was implemented using the bandwidth $h = n^{-1/5}$. The value of the constant c appearing above in (4.1) was taken to range from -.3 to .3, inclusive, in increments of .1. As one might expect, larger values of $|c|$ are seen to imply higher power for the three tests considered. The tests based on $\hat{R}_n(\cdot,\cdot)$ and $R_{n0}(\cdot,\cdot)$ are seen to dominate the more conventional test based on $S_n(\cdot,\cdot)$ for all data-generating processes considered with $c \neq 0$. In addition, the power performance of the directional test based on CvM_{nbD} dominates the omnibus tests based on $\hat{R}_n(\cdot,\cdot)$ and $R_{n0}(\cdot,\cdot)$. This is to be expected, since X_{1i}^2 is highly correlated with the direction of departure from the null model.

TABLES 1 AND 2 ABOUT HERE

We next consider a sequence of alternatives to the linear model implied by (4.1) under $H_0 : c = 0$ that does not involve the presence of X_{1i}^2 as c moves away from the origin. In particular, for (X_{1i}, X_{2i}, u_i) as given above in (4.1), we consider the sequence of alternatives indexed by $c \neq 0$ where

$$Y_i = X_{1i} + X_{2i} + c\sigma_{iD}^{\frac{3}{2}} + u_i,\ i = 1, \ldots, n; \tag{4.4}$$

and where $\sigma_{iD} = X_{2i}^2 + X_{1i}X_{2i}$. In the context of the model given in (4.4) for $c = 0, \pm.1, \pm.2, \pm.3$, Table 3 compares rejection probabilities of the Cramér–von Mises tests based on $\hat{R}_n(\cdot,\cdot)$ and $R_{n0}(\cdot,\cdot)$ to those exhibited by the directional test based on CvM_{nbD} given above in (4.3). Each of the tests considered in Table 3 was implemented in precisely the same way in which they were implemented in Table 2. From Table 3 it is clear that the omnibus approach represented by CvM_n and CvM_{n0} exhibits good power. On the other hand only weak power is exhibited by the directional test based on the observed significance of $\beta_3(\cdot)$ in the augmented quantile regression model in (4.2).

TABLE 3 ABOUT HERE

The simulations presented here have indicated that our proposals for multiplier bootstrap-based tests involving the processes $\hat{R}_n(\cdot,\cdot)$ or $R_{n0}(\cdot,\cdot)$ perform favorably when compared to more conventional specification tests based on the process $S_n(\cdot,\cdot)$ and using simulated critical values derived from subsampling. The size performance of the empirically feasible Cramér–von Mises test based on $\hat{R}_n(\cdot,\cdot)$ that we propose does not appear to be sensitive to the choice of bandwidth used to estimate the score. In addition, the results presented in Table 2 indicate a substantial improvement in power associated with our proposals when compared to a more conventional approach involving the process $S_n(\cdot,\cdot)$ and subsampling. Given the relative popularity of subsampling for inference involving the quantile-regression process, the results presented in Tables 1 and 2 suggest the desirability of less computationally demanding inference methodologies for the quantile-regression process involving some variant of the wild bootstrap.

Finally, the results presented in Tables 2 and 3 lend support to the theoretical discussion in Section 3.2 regarding directional tests based on the observed significance of additional regressors constructed by taking nonlinear transformations of existing regressors. In particular, the directional test based on the statistic given above in (4.3) offers excellent power, as shown in Table 2, against the specific family of alternatives implied by the additional regressor appearing in the augmented quantile regression in (4.2). On the other hand, the directional test exhibits weak power against alternatives that are not collinear with the directional alternatives implied by the additional regressor. This contrasts with the power of the omnibus Cramér–von Mises tests that we develop in this paper. These have high power against the two sets of alternatives considered in Tables 2 and 3.

5 Empirical Example: Returns to Experience in the U.S. Labor Market

We consider the empirical relationship between individual earnings and age. An individual's age is taken as a proxy for experience of the labor market. The relationship between earnings and age has been the subject of an extensive literature in labor economics since at least the time of the seminal studies of Heckman and Polachek (1974) and Mincer (1974). In particular, Mincer (1974) found that individual rates of pay typically increase with age, but at a diminishing rate over time. This stylized fact is largely responsible for the empirical convention of modelling the natural logarithm of individual earnings as a quadratic function of age.

In what follows, we consider the question of whether the commonly adopted quadratic specification of the earnings–age relationship is supported across a broad range of quantiles in the distribution of hourly wages. This has potential implications for the vast literature on the effect on earnings of factors such as gender, race, schooling or certain institutional features of a labor market. The general convention in these studies is to use the quadratic specification to control for the confounding effects of labor-market experience on earnings. In addition, the quadratic specification has been used frequently to model individual earnings growth over a career and to describe the evolution of earnings over time and across demographic groups at a given moment in time. Both of these types of study rely crucially on the assumption that individual earnings are adequately approximated as a quadratic in age. Although the tendency to date has been to approximate mean earnings as a quadratic in age, in what follows we consider the more general problem of approximating each quantile in a continuum of earnings quantiles as a quadratic in age. This has the potential of allowing for different models of the earnings–age relationship to apply to different regions of the earnings distribution. For example, one might in this context be able to entertain models of career earnings growth that differ substantively between workers classified as being high-income or low-income at a given point in time.

Specifically, we apply the omnibus testing approach developed above to assess the fit of the quantile-regression model

$$F^{-1}_{\log Y|X}(\alpha) = \beta_0(\alpha) + \beta_1(\alpha)\cdot X + \beta_2(\alpha)\cdot X^2,\ \alpha \in \mathcal{A}, \tag{5.1}$$

where Y and X denote hourly earnings and age, respectively, where $\mathcal{A}$ denotes one of $[.10, .90]$, $[.10, .50]$ or $[.50, .90]$ and where $\beta_0(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are unknown parameters.

We also use the same approach to evaluate the fit of the quartic specification

$$F^{-1}_{\log Y|X}(\alpha) = \beta_0(\alpha) + \beta_1(\alpha)\cdot X + \beta_2(\alpha)\cdot X^2 + \beta_3(\alpha)\cdot X^3 + \beta_4(\alpha)\cdot X^4,\ \alpha \in \mathcal{A}, \tag{5.2}$$

where $\beta_0(\cdot)$, $\beta_1(\cdot)$, $\beta_2(\cdot)$, $\beta_3(\cdot)$ and $\beta_4(\cdot)$ are unknown parameters, and where Y, X and $\mathcal{A}$ denote the same quantities appearing in (5.1). This comparison has some bearing on the results presented by Murphy and Welch (1990), who argued that the usual quadratic specification of the relationship between mean earnings and age results in highly biased estimates of average growth rates in earnings over a worker's career. In addition, Murphy and Welch (1990) pointed out that biases in estimates of earnings growth of a worker's career may be propagated in studies primarily concerned with the effects of factors such as gender, race or schooling on earnings, and in which the confounding effects of experience are assumed to be controlled for by a quadratic in age. Murphy and Welch (1990) went on to present evidence supporting a quartic specification of the average earnings–age relationship as an accurate and parsimonious alternative.

The fits of the quadratic and quartic quantile-regression specifications were assessed in the context of a recent cross section of the American labor market. The sample in question was taken from the March 2006 supplement to the Current Population Survey and was further restricted to male wage earners possessing at least a Bachelor's degree not currently serving in the armed forces. This resulted in a sample size of $n = 705$. The results appear in Table 4, which displays p-values of Cramér–von Mises tests of the quadratic and quartic log-wage equations given above in (5.1) and (5.2), respectively. These tests were based on the functional given above in (2.9), and followed the same implementation used in the simulation experiments presented above in Section 4. In particular, the unknown parameters appearing in (5.1) and (5.2) were estimated using the appropriate regression α-quantiles, while the continuum $\mathcal{A} \in \{[.10, .90], [.10, .50], [.50, .90]\}$ was approximated using a grid of 30 evenly spaced points between the corresponding endpoints. The distribution of the Cramér–von Mises statistic was approximated using 200 sequences of bootstrap multipliers, where the multipliers were also generated in accordance with the Bernoulli distribution summarized above in (3.10)–(3.11). The conditional density estimate appearing in the expression for CvM_n in (2.9) was computed using (2.8) with a Gaussian kernel. This density estimate was implemented using three different settings of bandwidth given by $h = \kappa n^{-1/5}$, where $\kappa = .5$, $\kappa = 1.0$ and $\kappa = 1.5$ in succession.

It is clear from Table 4 that the p-values for the tests based on the distribution of CvM_n in (2.9) are not apparently highly sensitive to the choice of bandwidth used to implement the conditional density estimates embedded in the test statistic. It is also clear that the quadratic specification given above in (5.1) apparently provides a poor fit for all earnings quantiles in the range $[.10, .90]$. On the other hand, the quadratic specification cannot be rejected at the 5% level for earnings quantiles in the range $[.10, .50]$.

The pattern observed for the quadratic specification constrasts with the observed fit of the quartic specification over each of the ranges of quantiles considered. In particular, the quartic specification cannot be rejected at the 5% level over any of the earnings quantiles considered, although the fit is apparently better for low earnings quantiles rather than for quantiles in the upper half of the earnings distribution. This inference is observed to hold for each of the three settings of bandwidth used.

TABLE 4 ABOUT HERE

The test results in Table 4 are corroborated by the results in the first panel of Table 5, which compares actual compound annual growth rates in hourly earnings for workers aged 30–59 in the sample with the predicted growth rates implied by the quadratic and quartic specifications. Annual growth rates in hourly earnings were considered for workers at the α-quantiles of the earnings distribution conditional on age, where $\alpha = .10, .30, .50, .70, .90$ in succession. The fourth and fifth columns of Table 5 display the relative biases of the quadratic and quartic predictions of annual growth rates in hourly earnings. These are respectively the ratios of the figures in the second and third columns to the actual growth rates appearing in the first column.

It is seen that the relative biases associated with both the quadratic and quartic specifications are greater for earnings quantiles in the top half of the earnings distribution than for the .10- and .30-quantiles. It is also seen that the quartic specification dominates the quadratic specification in terms of relative bias at all earnings quantiles considered. These results echo the pattern exhibited in Table 4 by the p-values of the specification tests of both models.

The remaining three panels in Table 5 present the relative biases of the quadratic and quartic quantile-regression specifications for the same set of earnings quantiles, but grouped according to whether the workers were aged 30–39, 40–49 or 50–59. It is seen that the actual annual growth rates in hourly earnings over the various earnings quantiles considered do not evolve in parallel fashion across the three different age ranges. For example, the workers at the tenth percentile of the earnings distribution over the age range 30–39 or 50–59 experience annual decreases in earnings, while their counterparts aged 40–49 experience fairly large annual increases. In contrast, median earnings workers aged 30–39 experience very slight annual decreases, while their counterparts aged 40–49 or 50–59 have slight annual increases. This suggests that the functional form of the earnings–age relationship may differ across quantiles. In addition, the relative bias of the quadratic and quartic specifications tends to be lowest for workers aged 40–49. The quartic fit also tends to dominate the quadratic fit in terms of relative bias for workers in the 30–39 and 50–59 age ranges.

In summary, the quadratic and quartic specifications both generate biased predictions of annual growth rates in hourly earnings across the various earnings quantiles and age ranges considered. As was suggested earlier by the test results in Table 4, this bias is occasionally severe, particularly in the case of the quadratic specification applied to earnings quantiles in the upper half of the distribution. Indeed, Table 5 suggests that the quadratic and quartic models may both be too simplistic to capture differing trajectories of career earnings growth between different quantiles of the conditional distribution of hourly earnings given age. On the other hand, it may be desirable in certain applications to select a parsimonious, if biased, parametric specification of the earnings–age relationship across all earnings quantiles in the range $[.10, .90]$. In this connection, Tables 4 and 5 both indicate that if a strict comparison is in order, then the quartic specification is generally preferable to the quadratic.

TABLE 5 ABOUT HERE

6 Conclusion

This paper has proposed a class of tests for the hypothesis of linearity in parameters of functions that are identified by conditional quantile restrictions. We have argued that these tests provide a potentially useful diagnostic tool for empirical researchers interested in the effect of conditioning variables on the distribution of some outcome variable of interest. We exploit various results from empirical process theory to derive the asymptotic behavior of our proposed test procedures. In a manner analogous to approaches taken by Neyman (1959) and Bickel et al. (2006), we have also shown how the use of an orthogonal projection on the tangent space of nuisance parameters at each quantile both improves power and facilitates the simulation of critical values via a simple multiplier bootstrap procedure. This principle applies generally, but is particularly convenient for quantile restrictions as discussed here. This is because the most popular resampling method for quantile specifications involves subsampling, which is both more computationally intensive than the multiplier bootstrap scheme we advocate, and also inconvenient from the point of view of requiring the subjective choice of subsample size. Simulation evidence and an empirical example concerning the specification of the earnings–age relationship in the U.S. labor market illustrate the feasibility of our approach in data sets of moderate size.

A Appendix

A.1 Preliminary results

We begin with an important result of Chen et al. (2003) that allows for the bounding of entropy numbers and the verification of stochastic equicontinuity for processes indexed by both Euclidean and function-valued parameters. In this connection, define a generic function class $\mathcal{H} \equiv \{w \to m(w, \theta, g) : \theta \in \Theta, g \in \mathcal{G}\}$, where Θ and $\mathcal{G}$ are generic Banach spaces with associated norms $\|\cdot\|_\Theta$ and $\|\cdot\|_\mathcal{G}$, respectively. Recall that the covering number $N(\epsilon, \Theta, \|\cdot\|_\Theta)$ of Θ is the minimal number N for which there exist ϵ-neighborhoods $\{\{\theta : \|\theta - \theta_j\|_\Theta \le \epsilon\}, \|\theta_j\|_\Theta < \infty, j = 1, \ldots, N\}$ covering Θ. A bracket $[l_j, u_j]$ is the set of elements $\theta \in \Theta$ such that $l_j \le \theta \le u_j$. The covering number with bracketing $N_{[\cdot]}(\epsilon, \Theta, \|\cdot\|_\Theta)$ is the minimal N for which there exist ϵ-brackets $\{[l_j, u_j] : \|l_j - u_j\|_\Theta \le \epsilon, \|l_j\|_\Theta, \|u_j\|_\Theta < \infty, j = 1, \ldots, N\}$ covering Θ. An envelope function G for the class $\mathcal{G}$ is a measurable function such that $G(x) \ge \sup_{g \in \mathcal{G}} |g(x)|$. Define the L_2 norm $\|f\|_2^2 \equiv E[f^2(W)]$ and the entropy number

$$J_{[\cdot]}(\delta, \mathcal{G}, \|\cdot\|_2) \equiv \int_0^\delta \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{W}, \|\cdot\|_2)} d\varepsilon. \quad \text{(A.1)}$$

Other definitions of concepts from empirical process theory may be found in e.g., van der Vaart and Wellner (1996).

Lemma A.1 (Chen et al. (2003, Theorem 3)) *Assume that*

$$E\left[\sup_{\theta_2:\,\|\theta_1-\theta_2\|_\Theta<\delta}\ \sup_{g_2:\,\|g_1-g_2\|_{\mathcal{G}}<\delta}|m(W,\theta_1,g_1)-m(W,\theta_2,g_2)|^2\right]\le K\delta^s \tag{A.2}$$

for some constant $s\in(0,2]$. Then the following hold:

1. *For any $\epsilon>0$, $N_{[\cdot]}(\epsilon,\mathcal{H},\|\cdot\|_2)\le N\left(\left[\frac{\epsilon}{2K}\right]^{2/s},\Theta,\|\cdot\|_\Theta\right)\times N\left(\left[\frac{\epsilon}{2K}\right]^{2/s},\mathcal{G},\|\cdot\|_{\mathcal{G}}\right)$.*

2. *If in addition to (A.2) holding for Θ denoting a compact subset of $\mathbb{R}^k$ for some k, we have $\int_0^\infty\sqrt{\log N\left(\epsilon^{2/s},\mathcal{G},\|\cdot\|_{\mathcal{G}}\right)}d\epsilon<\infty$, the empirical process $\left\{M_n(\theta,g)\equiv n^{-1}\sum_{i=1}^n m\left(W_i,\theta,g\right):\ \theta\in\Theta,g\in\mathcal{G}\right\}$ is asymptotically stochastically equicontinuous, i.e., for any sequence of positive constants $\delta_n=o(1)$,*

$$\sup_{\|\theta_1-\theta_2\|_\Theta\le\delta_n,\,\|g_1-g_2\|_{\mathcal{G}}\le\delta_n}\|M_n(\theta_1,g_1)--M(\theta_1,g_1)-M_n(\theta_2,g_2)+M(\theta_2,g_2)\|=o_P(n^{-1/2}), \tag{A.3}$$

where $M(\theta_1,g_1)\equiv E\left[m(W_i,\theta_1,g_1)\right]$.

The following Lemma is implicit in Section 2.10.3 of van der Vaart and Wellner (1996).

Lemma A.2 (Product class) *Let $\mathcal{F}$ and $\mathcal{G}$ be classes of functions with squared-integrable envelopes F and G, respectively. Then*

$$N(2\epsilon\|FG\|_2,\mathcal{F}\cdot\mathcal{G},\|\cdot\|_2)\le N(\epsilon\|F\|_2,\mathcal{F},\|\cdot\|_2)\times N(\epsilon\|G\|_2,\mathcal{G},\|\cdot\|_2)$$

and

$$N_{[\cdot]}(C\epsilon,\mathcal{F}\cdot\mathcal{G},\|\cdot\|_2)\le N_{[\cdot]}(\epsilon,\mathcal{F},\|\cdot\|_2)\times N_{[\cdot]}(\epsilon,\mathcal{G},\|\cdot\|_2).$$

We now state a weak convergence theorem that is useful in dealing with estimation effects in test functionals involving the non-smooth summands $\psi_\alpha(W_i,\theta_n)$. This result is of independent interest. With some abuse of notation, define $\psi_\alpha(W,b)\equiv 1\left(Y-X^\top b(\alpha)\le 0\right)-\alpha$. Given a sequence $\{W_{i,n}\equiv(X_{i,n}^\top,Y_{i,n})^\top\}_{i=1}^n$ of iid arrays for each n, define the weighted empirical process

$$V_n(\gamma)\equiv n^{-1/2}\sum_{i=1}^n\left(\psi_\alpha(W_{i,n},b)-E\left[\psi_\alpha(W_{i,n},b)|\,X_{i,n}\right]\right)w_n(X_{i,n},\gamma), \tag{A.4}$$

which is indexed by $\gamma\equiv(b,\alpha,x)\in\Gamma:=\mathcal{B}\times\mathcal{A}\times\mathbb{R}^d$, where $\mathcal{B}$ is a class of bounded, Lipschitz $\mathbb{R}^d$-valued functions on $\mathcal{A}$, and $\mathcal{A}$ is a compact subset of $(0,1)$. Define the class $\mathcal{W}_n\equiv\{w_n(X_{i,n},\gamma):\gamma\in\Gamma\}$, with a measurable envelope $W_n\equiv\sup_{w_n\in\mathcal{W}_n}|w_n(\cdot)|$ satisfying the Lindeberg condition, where for each $\varepsilon>0$, $E[W_n^2]=O(1)$ and $E[W_n^2 1(W_n>\varepsilon\sqrt{n})]=o(1)$.

Consider the following pseudo-metric for $\rho(\gamma_1, \gamma_2) \equiv \|\theta_1 - \theta_2\|_{\mathcal{A}} + |\alpha_1 - \alpha_2| + |F_X(x_1) - F_X(x_2)|$, where $\gamma_j \equiv (\theta_j, \alpha_j, x_j) \in \Gamma$, $j = 1, 2$. Assume that w_n is such that for a sufficiently small $\delta > 0$,

$$\sup_{\rho(\gamma,\gamma_1)<\delta} \|w_n(\cdot, \gamma) - w_n(\cdot, \gamma_1)\|_2 \leq C\delta. \tag{A.5}$$

Furthermore, we require the following assumption on $\mathcal{W}_n$:

Assumption 5 *The class $\mathcal{W}_n$ satisfies the previous conditions and is such that $J_{[\cdot]}(\epsilon_n, \mathcal{W}_n, \|\cdot\|_2) \to 0$ for every $\epsilon_n \downarrow 0$.*

Theorem A.3 *Under Assumptions 1 and 5, the process $V_n(\gamma)$ is $\rho-$stochastically equicontinuous.*

Proof. For $(b, \alpha) \in \mathcal{B} \times \mathcal{A}$, define $h(W, b, \alpha) \equiv \psi_\alpha(W, b) - E[\psi_\alpha(W, b)|\, X]$ and the class $\mathcal{H} \equiv \{w \to h(w, b, \alpha) : (b, \alpha) \in \mathcal{B} \times \mathcal{A}\}$. Fix $(b_1, \alpha_1) \in \mathcal{B} \times \mathcal{A}$. Set $\delta \in (0, 1)$. By the triangle inequality, we have the following uniformly in $n \geq 1$:

$$\begin{aligned}
&E\left[\sup_{b:\, \|b_1-b\|_{\mathcal{A}}<\delta^2} \sup_{\alpha:\, |\alpha_1-\alpha|<\delta^2} |h(W_{i,n}, b_1, \alpha_1) - h(W_{i,n}, b, \alpha)|^2\right] \\
&\leq CE\left[\sup_{\alpha:\, |\alpha_1-\alpha|<\delta^2} \sup_{b:\, \|b_1-b\|_{\mathcal{A}}<\delta^2} |\psi_{\alpha_1}(W_{i,n}, b_1) - \psi_\alpha(W_{i,n}, b)|^2\right] \\
&+ CE\left[\sup_{\alpha:\, |\alpha_1-\alpha|<\delta^2} \sup_{b:\, \|b_1-b\|_{\mathcal{A}}<\delta^2} |E[\psi_{\alpha_1}(W_{i,n}, b_1)|\, X] - E[\psi_\alpha(W_{i,n}, b)|\, X]|^2\right] \\
&\leq CE\left[\left(F\left(X^\top b_1(\alpha_1) + C|X|\delta^2 \middle|\, X\right) - F\left(X^\top b_1(\alpha_1) - C|X|\delta^2) \middle|\, X\right)\right)\right] \\
&+ CE\left[|X|^2 \sup_{\alpha:\, |\alpha_1-\alpha|<\delta^2} \sup_{b:\, \|b_1-b\|_{\mathcal{A}}<\delta^2} |b_1(\alpha_1) - b_2(\alpha_2)|^2\right] \\
&\leq C\delta^2.
\end{aligned} \tag{A.6}$$

Apply Lemma A.1, Lemma A.2 and a slight modification of van der Vaart and Wellner (1996, Theorem 2.7.11) to obtain $J_{[\cdot]}(\delta_n, \mathcal{H}\cdot\mathcal{W}_n, \|\cdot\|_2) \to 0$ for every $\delta_n \downarrow 0$. The desired conclusion follows from the proof of van der Vaart (1998, Theorem 19.28).

■

Our next results involve uniform convergence rates for kernel estimator $\hat{f}_{i\alpha} \equiv \hat{f}_h\left(X_i^\top \theta_n(\alpha)\middle|\, X_i\right)$. We view $\hat{f}_{i\alpha}$ as a function of θ_n and write

$$\hat{f}_h(x, b) \equiv \frac{1}{mh_m} \sum_{j=1}^{m} K\left(\frac{x^\top \{\theta_0(\alpha) - \theta_0(\alpha_j) + n^{-1/2}b(\alpha) - n^{-1/2}b(\alpha_j)\}}{h_m}\right), \tag{A.7}$$

The proofs of the results below follow along the lines of Lemma 1.5 in the supplemental material of Escanciano et al. (2013), hence their proofs are omitted. For a_m and b_m as in Assumption 3, define $d_m \equiv \sqrt{\left(\log a_m^{-1} \vee \log\log m\right) / (ma_m) + b_m^2}$.

Lemma A.4 *Let Assumptions 1 – 3 hold. Then*

$$\sup_{a_m \leq h \leq b_m} \sup_{\alpha \in \mathcal{A}, b \in \mathcal{B}} \max_{1 \leq i \leq n} \left| \hat{f}_h (X_i, b) - f \left(X_i^\top \theta_0(\alpha) \middle| X_i \right) \right| = O_p \left(n^{-1/2} + d_m \right).$$

Next, consider the asymptotic behavior under H_0 of the process $S_n(x, \alpha) = n^{-1/2} \sum_{i=1}^n \psi_\alpha(W_i, \theta_n) 1(X_i \leq x)$ as defined in (2.2) above.

Lemma A.5 *Under Assumptions 1–3,*

$$\sup_{(x,\alpha) \in \mathcal{T}} \left| S_n(x, \alpha) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(W_i, \theta_0) 1(X_i \leq x) - D^\top(x, \theta_0(\alpha)) \sqrt{n} \left(\theta_n(\alpha) - \theta_0(\alpha) \right) \right| = o_P(1).$$

Proof. Apply Theorem A.3 with $\mathcal{W} = \{\bar{x} \to 1(\bar{x} \leq x) : x \in [-\infty, \infty]^d\}$, which satisfies the conditions of the Theorem by Theorem 2.7.1 in van der Vaart and Wellner (1996). It is easy to show that by Assumption 1, $\theta_n \in \mathcal{B}$ with probability tending to one and $\theta_0 \in \mathcal{B}$. Hence, by the uniform consistency result of Lemma A.4, we conclude (see A.3 and the definition of V_n in (A.4) above) that

$$\sup_{(x,\alpha) \in \mathcal{T}} |V_n(\theta_n, \alpha, w(x)) - V_n(\theta_0, \alpha, w(x))| = o_P(1), \tag{A.8}$$

where $w(x) = 1(X \leq x)$. In turn, (A.8) is equivalent under the null to

$$\begin{aligned} \sup_{(x,\alpha) \in \mathcal{T}} \Bigg| & S_n(x, \alpha) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\alpha(W_i, \theta_0) 1(X_i \leq x) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(E\left[\psi_\alpha(W_i, \theta_0) \middle| X_i \right] - E\left[\psi_\alpha(W_i, \theta_n) \middle| X_i \right] \right) 1(X_i \leq x) \Bigg| = o_P(1). \end{aligned} \tag{A.9}$$

Applying a mean value argument we obtain

$$\begin{aligned} \sup_{(x,\alpha) \in \mathcal{T}} \Bigg| & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(E\left[\psi_\alpha(W_i, \theta_0) \middle| X_i \right] - E\left[\psi_\alpha(W_i, \theta_n) \middle| X_i \right] \right) 1(X_i \leq x) \\ & - \sqrt{n} \left(\theta_n(\alpha) - \theta_0(\alpha) \right)^\top \frac{1}{n} \sum_{i=1}^n f \left(X_i^\top \widetilde{\theta}_n(\alpha) \middle| X_i \right) 1(X_i \leq x) \Bigg| = o_P(1), \end{aligned} \tag{A.10}$$

where $\widetilde{\theta}_n(\alpha)$ is such that $\left| \widetilde{\theta}_n(\alpha) - \theta_0(\alpha) \right| \leq |\theta_n(\alpha) - \theta_0(\alpha)|$ a.s. for each $\alpha \in \mathcal{A}$. By our assumptions, we have uniformly in $(x, \alpha) \in \mathcal{T}$ that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f \left(X_i^\top \widetilde{\theta}_n(\alpha) \middle| X_i \right) 1(X_i \leq x) &= \frac{1}{n} \sum_{i=1}^n f \left(X_i^\top \theta_0(\alpha) \middle| X_i \right) 1(X_i \leq x) + o_P(1) \\ &= D(x, \theta_0(\alpha)) + o_P(1), \end{aligned} \tag{A.11}$$

where the last equality follows from the Glivenko–Cantelli Theorem, since $(x, \alpha) \to f \left(X^\top \theta_0(\alpha) \middle| X \right) 1(X \leq x)$ for $(x, \alpha) \in \mathcal{T}$ is Glivenko–Cantelli by Lemma A.2. Hence, we obtain the expansion

$$\begin{aligned} \sup_{(x,\alpha) \in \mathcal{T}} \Bigg| & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(E\left[\psi_\alpha(W_i, \theta_0) \middle| X_i \right] - E\left[\psi_\alpha(W_i, \theta_n) \middle| X_i \right] \right) 1(X_i \leq x) \\ & + D^\top(x, \theta_0(\alpha)) \sqrt{n} \left(\theta_n(\alpha) - \theta_0(\alpha) \right) \Bigg| = o_P(1), \end{aligned} \tag{A.12}$$

which together with (A.9) shows the desired result.

■

Lemma A.6 *Under Assumptions 1–3,*

$$\sup_{\alpha\in\mathcal{A}}\left|S_{1n}(\alpha)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_0)\delta(X_i,\theta_0(\alpha))-\Delta^\top(\alpha)\sqrt{n}\left(\theta_n(\alpha)-\theta_0(\alpha)\right)\right|=o_P(1),$$

where $S_{1n}(\alpha)=n^{-1/2}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_n)\hat{\delta}_n(X_i,\theta_n(\alpha))$.

Proof. We apply Theorem A.3 with $\mathcal{W}_n=\{\bar{x}\to\hat{f}_h(\bar{x},b)\,\bar{x}:b\in\mathcal{B}\}$. By Lemma A.4 we can take as envelope for the class the function $W_n(X_i)=|X_i|$, which trivially satisfies the conditions of the theorem. Similarly, we have uniformly in $n\geq 1$ that $E\left[\sup_{b:\,\|b_1-b\|_{\mathcal{A}}<\delta^2}\left|\hat{f}_h(X,b_1)\,X-\hat{f}_h(X,b)\,X\right|^2\right]\leq C\delta^2$. Hence, conclude from Lemma A.1 and van der Vaart and Wellner (1996, Theorem 2.7.11) that Assumption 5 holds for this class. The stochastic equicontinuity implies

$$\begin{aligned}
S_{1n}(\alpha)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_0)f\left(X_i^\top\theta_0(\alpha)\middle|X_i\right)X_i\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\alpha(W_i,\theta_0)\{\hat{f}_h(X_i,0)-f\left(X_i^\top\theta_0(\alpha)\middle|X_i\right)\}X_i\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(E\left[\psi_\alpha(W_i,\theta_0)|\,X_i\right]-E\left[\psi_\alpha(W_i,\theta_n)|\,X_i\right]\right)\hat{f}_h\left(X_i^\top\theta_n(\alpha)\middle|X_i\right)X_i+o_P(1)\\
&\equiv S_{11n}(\alpha)+S_{12n}(\alpha)+S_{13n}(\alpha)+o_P(1).
\end{aligned}\tag{A.13}$$

Van der Vaart (1998, Theorem 19.28) applied to the class $\{w\to\psi_\alpha(w,\theta_0)\{\hat{f}_h(x,0)-f\left(x^\top\theta_0(\alpha)\middle|x\right)\}x:\alpha\in\mathcal{A}\}$ yields the stochastic equicontinuity of the process S_{12n}. In addition, for a fixed $\alpha\in\mathcal{A}$,

$$\begin{aligned}
Var\left[S_{12n}(\alpha)\middle|\{\alpha_j\}_{j=1}^{m}\right]&\leq E\left[|X|^2\left\{\hat{f}_h(X_i,0)-f\left(X_i^\top\theta_0(\alpha)\middle|X_i\right)\right\}^2\right]\\
&=o_P(1).
\end{aligned}\tag{A.14}$$

Thus $\sup_{\alpha\in A}|S_{12n}(\alpha)|=o_P(1)$. Finally, simple arguments using Lemma A.4 show that

$$\sup_{(x,\alpha)\in\mathcal{T}}\left|S_{13n}(\alpha)-\Delta^\top(\alpha)\sqrt{n}\left(\theta_n(\alpha)-\theta_0(\alpha)\right)\right|=o_P(1).\tag{A.15}$$

■

Lemma A.7 *Under Assumptions 1–3,* $\sup_{(x,\alpha)\in\mathcal{T}}\left|\hat{D}_n(x,\theta_n(\alpha))-D(x,\theta_0(\alpha))\right|=o_P(1)$.

Proof. The proof follows from Lemma A.4 and the Glivenko–Cantelli Theorem in a routine fashion.

■

Lemma A.8 *Under the conditions of Assumptions 1–3,* $\sup_{\alpha\in\mathcal{A}}\left|\hat{\Delta}_n^{-1}(\alpha)-\Delta^{-1}(\alpha)\right|=o_P(1)$.

Proof. The proof follows from Lemma A.4, the Glivenko–Cantelli Theorem and the continuous mapping theorem.

■

A.2 Proof of Theorem 3.1

We have

$$
\begin{aligned}
&\hat{R}_n(x,\alpha) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_n) \left(1(X_i \le x) - \hat{D}_n(x,\theta_n(\alpha))^\top \hat{\Delta}_n^{-1}(\alpha)\hat{\delta}_n(X_i,\theta_n(\alpha))\right) \\
&= S_n(x,\alpha) - \hat{D}_n(x,\theta_n(\alpha))^\top \hat{\Delta}_n^{-1}(\alpha) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_n)\hat{\delta}_n(X_i,\theta_n(\alpha)) \\
&\equiv S_n(x,\alpha) - \hat{D}_n(x,\theta_n(\alpha))^\top \hat{\Delta}_n^{-1}(\alpha) \cdot S_{1n}(\alpha)\,.
\end{aligned} \tag{A.16}
$$

By Lemmas A.5, A.6, A.7 and A.8 we have that uniformly in $(x,\alpha) \in \mathcal{T}$:

$$
\begin{aligned}
&\hat{R}_n(x,\alpha) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0) 1(X_i \le x) + D^\top(x,\theta_0(\alpha))\sqrt{n}\,(\theta_n(\alpha) - \theta_0(\alpha)) \\
&- D(x,\theta_0(\alpha))^\top \Delta^{-1}(\alpha) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0)\delta(X_i,\theta_0(\alpha)) \\
&- D(x,\theta_0(\alpha))^\top \Delta^{-1}(\alpha) \cdot \Delta(\alpha) \cdot \sqrt{n}\,(\theta_n(\alpha) - \theta_0(\alpha)) + o_P(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0) \left(1(X_i \le x) - D(x,\theta_0(\alpha))^\top \Delta^{-1}(\alpha)\delta(X_i,\theta_0(\alpha))\right) + o_P(1) \\
&= R_{n0}(x,\alpha) + o_P(1).
\end{aligned} \tag{A.17}
$$

Q.E.D.

A.3 Proof of Corollary 3.1

Note that $R_{n0}(x,\alpha) = n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0) \left(1(X_i \le x) - D(x,\theta_0(\alpha))^\top \Delta^{-1}(\alpha)\delta(X_i,\theta_0(\alpha))\right)$. The weak convergence of $R_{n0}(x,\alpha)$ follows from the joint weak convergence of $n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0) 1(X_i \le x)$ and of $n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0)\delta(X_i,\theta_0(\alpha))$ since $D(x,\theta_0(\alpha))$ and $\Delta^{-1}(\alpha)$ are uniformly continuous in $\mathcal{T}$, as guaranteed by our conditions. The joint asymptotic equicontinuity follows from that of the marginals. A standard multivariate central limit theorem implies the convergence of the finite-dimensional distributions. The asymptotic equicontinuity of $n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0) 1(X_i \le x)$ follows from an application of Theorem A.3 applied to the class $\mathcal{B} = \{\theta_0\}$ and $\mathcal{W} = \{X \to 1(X \le x) : x \in [-\infty,\infty]^d\}$. Similarly, the asymptotic equicontinuity of $n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(W_i,\theta_0)\delta(X_i,\theta_0(\alpha))$ follows from the same theorem applied to $\mathcal{B} = \{\theta_0\}$ and $\mathcal{W} = \{(X) \to \delta(X,\theta_0(\alpha)) : \alpha \in \mathcal{A}\}$. Note that for all $\varepsilon \in (0,1)$, we have

$$
\begin{aligned}
&E\left[\sup_{\alpha_2:\,|\alpha_1-\alpha_2|<\varepsilon}|\delta(X,\theta_0(\alpha_1))-\delta(X,\theta_0(\alpha_2))|^2\right]\\
&\le CE\left[\sup_{\alpha_2:\,|\alpha_1-\alpha_2|<\varepsilon}\left|f\left(X^\top\theta_0(\alpha_1)\middle|X\right)-f\left(X^\top\theta_0(\alpha_2)\middle|X\right)\right|^2\right]\\
&\le C\varepsilon^2. \qquad (A.18)
\end{aligned}
$$

Hence, by Lemma A.1 $\mathcal{W}$ satisfies the conditions of the Theorem A.3. Q.E.D.

A.4 Proof of Theorem 3.2

The proof of Theorem 3.2 parallels the proof of Theorem 3.1 without any change in the arguments. This is so because Theorem A.3 is valid under the null and local alternatives. The only additional step is to prove that a uniform law of large numbers holds for $n^{-1}\sum_{i=1}^n b(X_i,\alpha)1(X_i\le x)$, which is satisfied under the conditions of Assumption 4 above. Q.E.D.

A.5 Proof of Theorem 3.3

As in Theorem 3.1, we can write $\hat{R}_n^*(x,\alpha)=S_n^*(x,\alpha)-\hat{D}_n(x,\theta_n(\alpha))^\top\hat{\Delta}_n^{-1}(\alpha)\cdot S_{1n}^*(\alpha)$, where $S_n^*(x,\alpha)=n^{-1/2}\sum_{i=1}^n\psi_\alpha(W_i,\theta_n)1(X_i\le x)V_i$ and $S_{1n}^*(\alpha)=n^{-1/2}\sum_{i=1}^n\psi_\alpha(W_i,\theta_n)\hat{\delta}_n(X_i,\theta_n(\alpha))V_i$. The classes of functions $\{(W,V)\to\psi_\alpha(W,\theta)1(X\le x)V:\alpha\in\mathcal{A},\theta\in\mathcal{B},x\in[-\infty,\infty]^d\}$ and $\{(W,V)\to\psi_\alpha(W,\theta)\hat{f}_h(X,\theta)XV:\alpha\in\mathcal{A},\theta\in\mathcal{B}\}$ are P-Donsker, as can be shown by applications of Lemma A.1. It follows from a stochastic equicontinuity argument and the consistency of θ_n and the score that, uniformly in $(x,\alpha)\in\mathcal{T}$ we have the convergences

$$
S_n^*(x,\alpha)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\alpha(W_i,\theta_0)1(X_i\le x)V_i+o_P(1) \qquad (A.19)
$$

and

$$
S_{1n}^*(\alpha)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\alpha(W_i,\theta_0)\delta(X_i,\theta_0(\alpha))V_i+o_P(1). \qquad (A.20)
$$

The rest of the proof follows from the multiplier central limit theorem; see van der Vaart and Wellner (1996, Theorem 2.9.2, p. 179), and the continuous mapping theorem. Q.E.D.

Acknowledgments

We would like to thank Don Andrews, Jushan Bai, Thomas Lemieux, Marcelo Moreira and Kevin Song for their insightful comments. We are also grateful to an Associate Editor and to two referees for their close reading of our paper. All errors are our own.

Escanciano acknowledges the support of the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

REFERENCES

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. and G. R. Ducharme, 1991, Asymptotic theory for bootstrap methods in statistics. Les publications du Centre de recherches mathématiques, Montreal.

Bickel, P. J., Y. Ritov and T. M. Stoker, 2006, Tailor-made tests for goodness of fit to semiparametric hypotheses. Annals of Statistics 34, 721–741.

Bierens, H. J., 1982, Consistent model specification tests. Journal of Econometrics 20, 105-134.

Bierens, H. J. and D. K. Ginther, 2001, Integrated conditional moment testing of quantile regression models. Empirical Economics 26, 307–324.

Bilias, Y., S. Chen and Z. Ying, 2000, Simple resampling methods for censored regression quantiles. Journal of Econometrics 99, 373–386.

Chang, N. M., 1990, Weak convergence of a self-consistent estimator of a survival function with doubly censored data. Annals of Statistics 18, 391–404.

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

Chernozhukov, V. and I. Fernández-Val, 2005, Subsampling inference on quantile regression processes. Sankhyā 67, 253–276.

Chernozhukov, V. and C. Hansen, 2005, An IV model of quantile treatment effects. Econometrica 73, 245–262.

Choi, S., W. J. Hall and A. Schick, 1996, Asymptotically uniformly most powerful tests in parametric and semiparametric models. Annals of Statistics 24, 841-861.

Dudley, R. M., 1999, Uniform central limit theorems. Cambridge University Press, Cambridge, U.K.

Escanciano, J. C. and C. Goh, 2012, Conditional density estimation in linear quantile regression. Unpublished paper.

Escanciano, J. C., D. T. Jacho-Chávez and A. Lewbel, 2013, Uniform convergence of weighted sums of non- and semi-parametric residuals for estimation and testing. Supplemental materials. Unpublished paper.

Escanciano, J. C. and C. Velasco, 2010, Specification tests of parametric dynamic conditional quantiles. Journal of Econometrics 159, 209-221.

Giné, E. and J. Zinn, 1990, Bootstrapping general empirical measures. Annals of Probability 18, 851–869.

Gutenbrunner, C. and J. Jurečková, 1992, Regression rank scores and regression quantiles. Annals of Statistics 20, 305–330.

Hahn, J., 1995, Bootstrapping quantile regression models. Econometric Theory 11, 105–121.

He, X. and F. Hu, 2002, Markov chain marginal bootstrap. Journal of the American Statistical Association 97, 783–795.

Heckman, J. J. and S. Polachek, 1974, Empirical evidence on functional form of the earnings–schooling relationship. Journal of the American Statistical Association 69, 350–354.

Horowitz, J. L., 1998, Bootstrap methods for median regression models. Econometrica 66, 1327–1351.

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

Horowitz, J. L. and V. G. Spokoiny, 2002, An adaptive, rate-optimal test of linearity for median regression models. Journal of the American Statistical Association 97, 822–835.

Koenker, R., 2005, Quantile regression. Cambridge University Press, New York.

Koenker, R. and G. Bassett, 1978, Regression quantiles. Econometrica 46, 33–50.

Koenker, R. and K. F. Hallock, 2001, Quantile regression. Journal of Economic Perspectives 15(4), 143–156.

Koenker, R. and J. A. F. Machado, 1999, Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94, 1296-1310.

Mammen, E., 1993, Bootstrap and wild bootstrap for high-dimensional linear model. Annals of Statistics, 21, 225-285.

Mincer, J., 1974, Schooling, experience and earnings. Columbia University Press, New York.

Murphy, K. M. and F. Welch, 1990, Empirical age–earnings profiles. Journal of Labor Economics 8, 202–229.

Neyman, J., 1959, Optimal asymptotic tests of composite statistical hypotheses, in: U. Grenander, (Ed.), Probability and statistics: The Harald Cramér Volume. Almqvist and Wiksell, Stockholm. pp. 213–234.

Otsu, T., 2009, RESET for quantile regression. Test 18, 381-391.

Ramsey, J. B., 1969, Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society, Series B., 31, 350–371.

Rosenblatt, M., 1969, Conditional probability density and regression estimate, in: P. R. Krishnaiah, (Ed.), Multivariate Analysis, II. Academic Press, New York. pp. 25-31.

Rothe, C. and D. Wied, 2012, Misspecification testing in a class of conditional distributional models. Journal of the American Statistical Association, forthcoming.

Sakov, A. and P. J. Bickel, 2000, An Edgeworth expansion for the m out of n bootstrapped median. Statistics and Probability Letters 49, 217–223.

van der Vaart, A. W., 1998, Asymptotic statistics. Cambridge University Press, Cambridge, U.K.

van der Vaart, A. W. and J. A. Wellner, 1996, Weak convergence and empirical processes: With applications to statistics. Springer-Verlag New York, New York.

Whang, Y.-J., 2006a, Consistent specification testing for quantile regression models, in: D. Corbae, S. N.

Durlauf, and B. E. Hansen, (Eds.), Econometric theory and practice: Frontiers of analysis and applied research. Cambridge University Press, New York. pp. 288–308.

Whang, Y.-J., 2006b, Smoothed empirical likelihood methods for quantile regression models. Econometric Theory 22, 173–205.

Zheng, J. X., 1998, A consistent nonparametric test of parametric regression models under conditional quantile restrictions. Econometric Theory 14, 123–138.

Table 1: Model (4.1): Empirical rejection probabilities over 1000 Monte Carlo replications of 5% tests under $H_0 : c = 0$, $n = 100$, $h = \kappa n^{-1/5}$, 200 multiplier and subsampling bootstrap replications, subsample sizes b

	CvM_n		CvM_{n0}	CvM_{nb0}	CvM_{nbD}
$\kappa = .5$	$\kappa = 1$	$\kappa = 1.5$		$b = 18$	$b = 23$
.0450	.0460	.0430	.0230	.0620	.0510

Table 2: Model (4.1): Empirical rejection probabilities over 1000 Monte Carlo replications of 5% tests, $n = 100$, $h = n^{-1/5}$, 200 multiplier and subsampling bootstrap replications, subsample sizes b

c	CvM_n	CvM_{n0}	CvM_{nb0} $(b = 18)$	CvM_{nbD} $(b = 23)$
-.3	.9200	.9990	.4140	.9790
-.2	.7760	.9560	.2700	.9460
-.1	.4170	.6190	.1340	.7290
0.0	.0460	.0230	.0620	.0510
.1	.3910	.5730	.1120	.7480
.2	.7930	.9710	.2490	.9610
.3	.9280	.9960	.4120	.9800

Table 3: Model (4.4): Empirical rejection probabilities over 1000 Monte Carlo replications of 5% tests, $n = 100$, $h = n^{-1/5}$, 200 multiplier and subsampling bootstrap replications, subsample size b

c	CvM_n	CvM_{n0}	CvM^*_{nbD} ($b = 23$)
-.3	.6590	.7590	.2760
-.2	.5420	.5650	.2150
-.1	.4520	.3050	.1680
0.0	.0460	.0230	.0510
.1	.4620	.3000	.1490
.2	.5850	.5520	.2140
.3	.6630	.7520	.2970

Table 4: Linearity tests over various ranges of quantiles, quadratic and quartic log wage equations; male non-military wage earners with at least a Bachelor's degree; $n = 705$

	p-values, H_0 : quadratic		
Range of quantiles	$h = .5n^{-1/5}$	$h = n^{-1/5}$	$h = 1.5n^{-1/5}$
[.10, .90]	.0348	.0299	.0348
[.10, .50]	.1393	.1443	.1493
[.50, .90]	.0100	.0100	.0149
	p-values, H_0 : quartic		
Range of quantiles	$h = .5n^{-1/5}$	$h = n^{-1/5}$	$h = 1.5n^{-1/5}$
[.10, .90]	.6468	.6070	.5970
[.10, .50]	.8209	.8209	.8259
[.50, .90]	.2090	.2139	.2239

Table 5: Estimates of hourly earnings growth across age ranges and the earnings distribution; male non-military wage earners with at least a Bachelor's degree

Age range	Compound annual growth rates of earnings			Quadratic	Quartic
	Actual rate	Quadratic prediction	Quartic prediction	prediction relative to actual rate	prediction relative to actual rate
30–59:					
.10-quantile	-.0051	.0013	-.0015	-.2639	.2946
.30-quantile	-.0075	.0011	.0003	-.1526	-.0425
median	-.0006	.0026	.0016	-4.1848	-2.4778
.70-quantile	-.0016	.0040	.0039	-2.4785	-2.4013
.90-quantile	-.0043	.0050	.0041	-1.1588	-.9612
30–39:					
.10-quantile	-.0051	.0013	-.0015	-.2639	.2946
.30-quantile	-.0075	.0011	.0003	-.1526	-.0425
median	-.0006	.0026	.0016	-4.1848	-2.4778
.70-quantile	-.0016	.0040	.0039	-2.4785	-2.4013
.90-quantile	-.0043	.0050	.0041	-1.1588	-.9612
40–49:					
.10-quantile	.0131	.0012	-.0036	.0903	-.2721
.30-quantile	.0112	.0010	-.0006	.0909	-.0550
median	.0060	.0023	.0001	.3883	.0093
.70-quantile	-.0009	.0036	.0015	-3.8366	-1.5961
.90-quantile	.0156	.0045	.0015	.2883	.0939
50–59:					
.10-quantile	-.0208	-.0036	-.0044	.1730	.2125
.30-quantile	-.0150	-.0035	-.0037	.2299	.2477
median	.0039	-.0025	-.0019	-.6490	-.4975
.70-quantile	.0028	-.0003	.0035	-.1018	1.2402
.90-quantile	-.0051	.0023	.0063	-.4570	-1.2390