

Integrated Conditional Moment testing of quantile regression models

Herman J. Bierens¹, Donna K. Ginther

Department of Economics, Pennsylvania State University, 608 Kern Graduate Building, University Park, PA16802, and Tilburg University, The Netherlands (E-mail: hbierens@psu.edu. URL: <http://econ.la.psu.edu/~hbierens/>.)
Research Department, Federal Reserve Bank of Atlanta, 104 Marietta Street, NW, Atlanta, GA 30303, USA

Abstract. In this paper we propose a consistent test of the linearity of quantile regression models, similar to the Integrated Conditional Moment (ICM) test of Bierens (1982) and Bierens and Ploberger (1997). This test requires re-estimation of the quantile regression model by minimizing the ICM test statistic with respect to the parameters. We apply this ICM test to examine the correctness of the functional form of three median regression wage equations.

Key words: Quantile regression; Test for linearity; Integrated conditional moment test; Wage equations

JEL classification: C12, C21, C52, J31

1. Introduction

Median and quantile estimation methods have recently been applied to economic models because these methods impose fewer restrictions on the data than widely-used mean regressions. The linear median regression model assumes that the conditional median of the dependent variable y is a linear function of the vector x of independent variables. The median regression model is particularly suitable if the conditional distribution of the y variable is fat-tailed, or if

¹ Previous versions of this paper have been presented by the first author at the University of Pennsylvania, the Econometric Society European Meeting 1997, Toulouse, Johns Hopkins University, and the conference on Economic Applications of Quantile Regression in Konstanz, Germany. The constructive comments of the co-editor, Bernd Fitzenberger, and a referee, are gratefully acknowledged.

the lowest and/or highest values of y are truncated or misreported. The latter may occur if y is a measure of income, because respondents in the highest income group will often be reluctant or unwilling to reveal their true income. Also, the median regression model may serve as an alternative for a Tobit model if the conditional distribution of the error of the underlying latent model is symmetric but non-normal.

To the best of our knowledge, the only paper in the econometrics literature that addresses the problem of consistent testing of the functional form of quantile regression models is Zheng (1998). Zheng's approach is based on weighted kernel regression estimation. In this paper we propose an alternative consistent test of the linearity of the median regression model, similar to the Integrated Conditional Moment (ICM) test of Bierens (1982) and Bierens and Ploberger (1997). This test can easily be extended to more general quantile regression models. This test requires re-estimation of the median regression model by minimizing the ICM test statistic with respect to the parameters.

Although median and quantile regression has not been used as extensively as OLS in the empirical literature, recent papers have used this method to estimate wage equations and the conditional wage distribution. (See for example, Chamberlain 1994, Buchinsky 1994, 1995, 1997, and Poterba and Rueben 1994). However, the applicability of quantile regression is not limited to labor economics. For example, Chernozhukov and Umantsev (2001) estimate and analyze the conditional market risk of an oil producers stock price as a function of the key economic variables, using quantile regression, and discuss specification tests as well.

In order to show that not only in theory but also in practice the ICM test is able to detect misspecification of quantile regressions models, we shall apply the ICM test to examine the functional form of three wage equations that has been estimated previously by quantile regression methods, using a sample of 28,155 male workers taken from the March 1988 Current Population Survey (CPS). All computations have been done using the econometrics software package *EasyReg*, written by the first author. (*EasyReg* is an interactive Windows 95/98/NT freeware program, which is downloadable from web page <http://econ.la.psu.edu/~hbierens/EASYREG.HTM>. The data is downloadable from web page <http://econ.la.psu.edu/~hbierens/MEDIAN.HTM>)

In the discussion of the ICM test for quantile regression models we will focus on the median regression case. In section 2 of the appendix we show that only a minor change in the ICM testing protocol is required to cover more general quantile regression models as well.

2. Median regression models, and LAD estimation

Consider a random sample $(y_j, x_j) \in \mathbb{R} \times \mathbb{R}^k$, $j = 1, \dots, n$, where the dependent variable y_j is related to the vector x_j of explanatory variables, possibly including a constant 1, by the median regression model

$$y_j = \theta_0^T x_j + \varepsilon_j, \quad \text{where } P[\varepsilon_j > 0 | x_j] = P[\varepsilon_j < 0 | x_j]. \quad (1)$$

As is well known, under some regularity conditions, in particular the condi-

tions that the error term ε_j is continuously distributed with zero median and is independent of x_j , and that $E(x_j^T x_j) < \infty$, the parameter vector θ_0 can be estimated consistently by the Least Absolute Deviation (LAD) estimator $\hat{\theta}_{LAD} = \operatorname{argmin}_{\theta} \sum_{j=1}^n |y_j - \theta^T x_j|$, and $\sqrt{n}(\hat{\theta}_{LAD} - \theta_0) \rightarrow N_k(0, (1/[2f(0)]^2)Q^{-1})$ in distribution, where f is the density of ε_j and $Q = \operatorname{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n x_j x_j^T$. See Koenker and Bassett (1978) and the references therein.

The median regression model is equivalent to

$$u_j(\theta) = I(y_j - \theta^T x_j > 0) - I(y_j - \theta^T x_j < 0), \quad E[u_j(\theta_0)|\tilde{x}_j] = 0 \text{ a.s.}, \quad (2)$$

where $\theta_0 \in \Theta \subset \mathbb{R}^k$ with Θ the parameter space, $I(\cdot)$ is the indicator function, and

Definition 1. \tilde{x}_j is an m -vector of components of x_j such that the σ -Algebra generated by x_j is equal to the σ -Algebra generated by \tilde{x}_j .

For example, if $x_j = (1, z_j, z_j^2)^T$ with z_j a scalar random variable then $\tilde{x}_j = z_j$.

The median regression model is correctly specified if

$$H_0: \exists \theta_0 \in \Theta: P(E[u_j(\theta_0)|\tilde{x}_j] = 0) = 1. \quad (3)$$

We will test the null hypothesis (3) against the general alternative that (3) is not correct, e.g.

$$H_1: \forall \theta \in \Theta: P(E[u_j(\theta)|\tilde{x}_j] = 0) < 1. \quad (4)$$

3. Integrated conditional moment testing of median regression models

As has been shown by Bierens (1990, 1994 Sec. 3.3) and Bierens and Ploberger (1997), these two hypotheses can be distinguished by using an infinite set of moment conditions of the type $E[u_j(\theta)w_j(\xi)] = 0$, where

$$w_j(\xi) = w(\xi^T \Phi(\tilde{x}_j)), \quad \xi \in \Xi, \quad (5)$$

with Φ a bounded on-to-one mapping, $w(\cdot)$ an analytical function with all but a finite number of derivatives at zero unequal to zero, and Ξ a subset of a Euclidean space with positive Lebesgue measure. Bierens (1990) has shown for the case $w(\cdot) = \exp(\cdot)$ that

Theorem 1. Under H_0 the set $S(\theta) = \{\xi \in \Xi: E[u_j(\theta)w_j(\xi)] = 0\}$ is such that $S(\theta_0) = \Xi$, and for all $\theta \in \Theta \setminus \{\theta_0\}$, $S(\theta)$ has Lebesgue measure zero and is nowhere dense, provided that the parameter θ_0 in (3) is unique. Under H_1 , $S(\theta)$ has Lebesgue measure zero and is nowhere dense for all $\theta \in \Theta$.

Bierens and Ploberger (1997) and Stinchcombe and White (1998) have shown that the same result holds for a much wider class of weight functions $w_j(\cdot)$. In particular, Theorem 1 also holds if we choose

$$w_j(\xi) = \cos(\xi^T \Phi(\tilde{x}_j)) + \sin(\xi^T \Phi(\tilde{x}_j)), \quad \xi \in \mathbb{R}^m, \quad (6)$$

where Φ is the same as in (5). The advantage of this weight function is that it is uniformly bounded: $|w_j(\xi)| \leq \sqrt{2}$. In this paper we will use the weight function (6), for reasons to be explained later.

Denoting

$$\bar{Q}(\theta) = \int |E[u_1(\theta)w_1(\xi)]|^2 d\mu(\xi), \tag{7}$$

where

Assumption 1. $w_j(\xi)$ is defined by (5), μ is a probability measure on Ξ that is absolutely continuous with respect to Lebesgue measure, and Ξ is compact with positive Lebesgue measure,

it follows from Theorem 1 that under H_0 , $\bar{Q}(\theta_0) = 0$, whereas under H_1 , $\inf_{\theta \in \Theta} \bar{Q}(\theta) > 0$. This result suggests to use the Integrated Conditional Moment (ICM) test based on $\inf_{\theta \in \Theta} \hat{Q}(\theta)$, where

$$\hat{Q}(\theta) = \int |\hat{z}(\theta, \xi)|^2 d\mu(\xi) \tag{8}$$

with

$$\hat{z}(\theta, \xi) = (1/n) \sum_{j=1}^n u_j(\theta)w_j(\xi). \tag{9}$$

It follows from the result in Bierens (1990) and Bierens and Ploberger (1997) that under H_0 and Assumptions 1,

$$\sqrt{n}\hat{z}(\theta_0, \xi) \Rightarrow z(\xi), \tag{10}$$

where “ \Rightarrow ” means “converges weakly to” [cf. Billingsley (1968)], and $z(\cdot)$ is a zero mean Gaussian process on Ξ with covariance function

$$\Gamma(\xi_1, \xi_2) = E[z(\xi_1)z(\xi_2)] = E[u_1(\theta_0)^2 w_1(\xi_1)w_1(\xi_2)]. \tag{11}$$

Note that if

Assumption 2. The conditional distribution of y_j given x_j is continuous,

then $P(y_j - \theta_0^T x_j = 0) = 0$, hence $u_j(\theta_0)^2 = 1$ a.s. and consequently

$$\Gamma(\xi_1, \xi_2) = E[w_1(\xi_1)w_1(\xi_2)]. \tag{12}$$

It follows now from Bierens and Ploberger (1997) that under H_0 ,

$$\begin{aligned} F_n(\theta_0) &= \frac{n\hat{Q}(\theta_0)}{(1/n) \sum_{j=1}^n \int w_j(\xi)^2 d\mu(\xi)} \rightarrow \frac{\sum_{i=1}^{\infty} \varepsilon_i^2 \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} \\ &\leq \sup_{m \geq 1} \frac{1}{m} \sum_{i=1}^m \varepsilon_i^2 = \bar{F} \end{aligned} \tag{13}$$

in distribution, where the ε_i 's are independent standard normally distributed, and the λ_i 's are the solutions of the eigenvalue problem $\int \Gamma(\xi_1, \xi_2) \phi_i(\xi_2) \cdot d\mu(\xi_2) = \lambda_i \phi_i(\xi_1)$, with the $\phi_i(\xi)$ the corresponding eigenfunctions. The eigenvalues λ_i are non-negative and real valued, and the eigenfunctions $\phi_i(\xi)$ are real valued and orthogonal with respect to μ .

The statistic $F_n(\theta_0)$ cannot be used as a test statistic, because θ_0 is unknown. If we would plug in a \sqrt{n} -consistent estimator $\hat{\theta}$ for θ_0 , for example the LAD estimator, we need to take the asymptotic distribution of $n(\hat{Q}(\hat{\theta}) - \hat{Q}(\theta_0))$ into account. Under similar conditions as in Powell (1984) and Fitzenberger (1997) it is possible to derive the exact asymptotic null distribution of $n\hat{Q}(\hat{\theta})$, but this asymptotic distribution involves the unknown conditional density at zero of the error term ε_j in model (1). In principle we could estimate this density by a kernel density estimator, but the problem is that the height of the kernel density estimator is very sensitive for the choice of the window width, which renders this approach unreliable. For example, the t -values of the LAD model in Table 3.A below where computed using a kernel density estimator with standard normal kernel and window width $h = s \cdot n^{-0.2}$, where s is the mean of the absolute values of the LAD residuals in deviation of their sample mean. Its value at zero was 0.8853, but when the window width was multiplied by a factor 10 this value reduced to 0.5688! A possible alternative solution is to bootstrap $F_n(\hat{\theta}_{LAD})$. However, for large data sets this may take too much computing time. In particular, the LAD estimation of the model in Table 3.A below, with 18 parameters and 28,155 observations, took about 15 minutes on a Pentium II PC, using *EasyReg*. Thus for this case 1000 bootstraps of $F_n(\hat{\theta}_{LAD})$ would take about 250 hours nonstop! Therefore, we propose the following more practical approach: Choose

$$\hat{F} = \frac{n\hat{Q}(\hat{\theta})}{(1/n) \sum_{j=1}^n \int (w_j(\xi))^2 d\mu(\xi)}, \tag{14}$$

as the test statistic of the ICM test, where

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \hat{Q}(\theta). \tag{15}$$

Then $\hat{F} = F_n(\hat{\theta}) \leq F_n(\theta_0)$, hence the asymptotic inequality in (13) will be preserved:

Theorem 2. *Under Assumptions 1–2 and H_0 , $\limsup_{n \rightarrow \infty} P(\hat{F} > F) \leq P(\bar{F} > F)$ for all non-random $F > 0$.*

Bierens and Ploberger (1997) have shown that

$$P(\bar{F} > 3.23) = 0.10, \quad P(\bar{F} > 4.26) = 0.05 \tag{16}$$

Thus we reject the median regression model at the 10% significance level if $\hat{F} > 3.23$ and at the 5% level if $\hat{F} > 4.26$.

Admittedly, due to the inequality in Theorem 2 the actual size of the test will be smaller than the theoretical size (16), but this is the price we have to

pay for feasibility. The actual size will be smaller than the unknown actual size of the test based on $F_n(\hat{\theta}_{LAD})$, which in its turn is smaller than (16). There is no general answer to the question how much smaller the actual size is: The size distortion involved depends on the actual value of the conditional error density at zero, and on the distribution of the X variables (the latter via the eigenvalues of the covariance function), and therefore varies from case to case.

Note that if we choose the weight function $w_j(\xi)$ as in (6) and the probability measure μ symmetric, then $\int w_j(\xi)^2 d\mu(\xi) = 1$, so that then $\hat{F} = n\hat{Q}(\hat{\theta})$. It is for that reason, and the fact that then $w_j(\xi)$ is uniformly bounded, that we favor the weight function (6). The boundedness of the weight function is important for our applications, because due to the large sample size the integral in (8) has to be computed by Monte Carlo simulation. See section 2 of the appendix

4. Consistency of the ICM test

For the consistency of the ICM test we need to establish that $\liminf_{n \rightarrow \infty} \inf_{\theta \in \Theta} \hat{Q}(\theta) > 0$ under H_1 . A sufficient condition for this is that

$$\hat{Q}(\theta) \rightarrow \bar{Q}(\theta) \text{ a.s.}, \tag{17}$$

uniformly in $\theta \in \Theta$. Therefore, we will now set forth condition for uniform convergence of $\hat{Q}(\theta)$.

Under Assumption 2 it follows from the uniform strong law of large numbers of Jennrich (1969) [cf. Bierens (1994 Sec. 2.7) for details] that

$$\sup_{\xi \in \Xi} |\hat{z}(\theta, \xi) - E[u_1(\theta)w_1(\xi)]| \rightarrow 0 \text{ a.s.}, \tag{18}$$

pointwise in $\theta \in \Theta$, hence $\hat{Q}(\theta) \rightarrow \bar{Q}(\theta)$ a.s., pointwise in $\theta \in \Theta$. If the function $u_j(\theta)$ would be continuous in θ and the parameter space Θ is compact, then by Jennrich’s (1969) uniform strong law of large numbers this result would hold uniformly on Θ as well. However, for the median regression model under review the function $u_j(\theta)$ is discontinuous in the parameters, so that the standard uniform convergence proof [see for example Bierens (1994, Sec. 4.2)] no longer applies. Nevertheless, it can be shown (see section 1 of the appendix) that (18) also holds uniformly on Θ , provided that:

Assumption 3. *The parameter space Θ is compact,*

Assumption 4. *The conditional density $f(y|x)$ of y_j given $x_j = x$ satisfies $E[\sup_y f(y|x_1)] < \infty$.*

Then:

Theorem 3. *Under Assumptions 1–4, $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |\hat{Q}(\theta) - \bar{Q}(\theta)| = 0$ a.s.*

Thus, under the alternative hypothesis (4) we have $\liminf_{n \rightarrow \infty} \hat{Q}(\hat{\theta}) \geq \inf_{\theta \in \Theta} \bar{Q}(\theta) > 0$ a.s., which establishes the consistency of the ICM test.

5. Local power of the ICM test for the case that θ_0 is known

In Bierens and Ploberger (1997) it has been shown that the ICM test of the functional form of conditional expectation models has non-trivial \sqrt{n} local power, which is superior to alternative consistent model misspecification tests based on comparison of the parametric functional form involved with a non-parametric kernel regression model. See the references in Zheng (1998). In this section we will therefore derive the local power properties of the ICM test for median regression models, but only for the special case that θ_0 is known. For the general case (14) we need to derive the limiting distribution and rate of convergence of the ICM estimator (15) under the local alternative, which can be done similar to Bierens and Ploberger (1997), using the conditions in Powell (1984) and Fitzenberger (1997). However, since this ICM estimator is of no particular interest by itself (it only serves an auxiliary role in the ICM test), deriving this limiting distribution is beyond the scope of this paper. Our purpose is to show that for this special case the ICM test has better local power than Zheng’s (1998) test, as a motivation for the use of the ICM test (14).

The local alternative involved is similar to the local alternative considered by Zheng (1998):

$$H_1^L: y_j = \theta_0^T x_j + g(x_j)/\sqrt{n} + \varepsilon_j, \tag{19}$$

where g is a uniformly bounded nonlinear function of x_j such that $P[g(x_j) = 0] < 1$, and ε_j is independent of x_j , with continuous distribution function $F(\cdot)$ satisfying $F(0) = 0.5$. Moreover, let $f(\cdot)$ be the density of ε_j , and assume that $f(0) > 0$ and $f(\cdot)$ is differentiable with uniformly bounded derivative $f'(\cdot)$. Furthermore, we may without loss of generality assume further that $\theta_0 = 0$.

Under the local alternative (19) with $\theta_0 = 0$ and weight function (6), it follows similarly to Bierens and Ploberger (1997) that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (u_{n,j} w_j(\xi) - E[u_{n,j} w_j(\xi)]) \Rightarrow z_*(\xi) \tag{20}$$

on \mathcal{E} , where $u_{n,j} = I[\varepsilon_j > -g(x_j)/\sqrt{n}] - I[\varepsilon_j < -g(x_j)/\sqrt{n}]$, and $z_*(\xi)$ is a zero mean Gaussian process with covariance function

$$\begin{aligned} \Gamma_*(\xi_1, \xi_2) &= \lim_{n \rightarrow \infty} (E[u_{n,j}^2 w_j(\xi_1) w_j(\xi_2)] - E[u_{n,j} w_j(\xi_1)] E[u_{n,j} w_j(\xi_2)]) \\ &= E[w_j(\xi_1) w_j(\xi_2)] = \Gamma(\xi_1, \xi_2). \end{aligned} \tag{21}$$

Compare (12). Thus, $z_*(\xi)$ has the same distribution as the Gaussian process $z(\xi)$ in (10). The second equality in (21) follows from the fact that $E[u_{n,j}^2 | x_j] = 1$ a.s., and that by the Taylor expansion and the uniform boundedness of $g(\cdot)$, $f'(\cdot)$, and $w_j(\xi)$,

$$\begin{aligned} E[u_{n,j} w_j(\xi)] &= E[(1 - 2F(-g(x_1)/\sqrt{n})) w_j(\xi)] \\ &= (f(0)/\sqrt{n}) E[g(x_1) w_1(\xi)] + O(1/n) \end{aligned}$$

uniformly in ξ . It is now not hard to verify that under the local alternative,

$$n\hat{Q}(0) \rightarrow \int (z_*(\xi) + f(0)E[g(x_1)w_1(\xi)])^2 d\mu(\xi)$$

in distribution. Finally, similarly to Bierens and Ploberger (1997) it can be shown that

$$\int (z_*(\xi) + f(0)E[g(x_1)w_1(\xi)])^2 d\mu(\xi) > \int (z_*(\xi))^2 d\mu(\xi).$$

Therefore, in the case that θ_0 is known the ICM test has non-trivial \sqrt{n} local power, whereas in the same case the test by Zheng (1998) has non-trivial local power only at a slower rate.

For the general case we need to show that $(1/\sqrt{n}) \sum_{j=1}^n (\hat{u}_{n,j} - u_{n,j})w_j(\xi) \Rightarrow z_{**}(\xi)$ on \mathcal{E} for some Gaussian process $z_{**}(\xi)$, where

$$\hat{u}_{n,j} = I[\varepsilon_j > (\hat{\theta} - \theta_0)^T x_j - g(x_j)/\sqrt{n}] - I[\varepsilon_j < (\hat{\theta} - \theta_0)^T x_j - g(x_j)/\sqrt{n}],$$

because then it is guaranteed that under the local alternative, $\text{plim}_{n \rightarrow \infty} \hat{Q}(\hat{\theta}) > 0$. Again, this can be done under similar conditions as in Powell (1984) and Fitzenberger (1997), by mimicking the approach in Bierens and Ploberger (1997).

This comparison of the local power of the ICM test and Zheng's test is only a theoretical comparison, and one may wonder how much the finite sample powers differ. A Monte Carlo study could answer this question, but that is beyond the scope of this paper. Some preliminary unpublished Monte Carlo results by Bernd Fitzenberger suggest that the asymptotic advantage of the ICM test might require very large samples to be present in finite samples.

6. Empirical illustrations

6.1. The data

Linear quantile regression methods have been applied to estimating wage equations and characterizing the conditional distribution of the log of wages (Chamberlain 1994, Buchinsky 1994, 1995, 1997, and Poterba and Rueben 1994). This application of the ICM test uses data from the same source as the above studies, the Current Population Survey (CPS).

This study uses data on males from the 1988 March CPS and uses criteria similar to Buchinsky (1994) to sample the data. The March CPS contains information on previous year's wages, schooling, industry, and occupation. We select a sample of men ages 18 to 70 with positive annual income greater than \$50 in 1992, who are not self-employed nor working without pay. The wage data is deflated by the deflator of Personal Consumption Expenditure for 1992. Our data contains 28,155 observations and has variables for age, years of schooling, years of potential work experience, industry, occupation, and dummy variables for race, region of residence, living in an SMSA, and working part time.

6.2. Wage specifications

As an illustration of the performance of the ICM test we examine three wage equation specifications. The first specification is the traditional Mincer-type model where the log of weekly wages are regressed on a constant, years of schooling, years of potential work experience and its square, and a race dummy. This model is used in most labor papers, including Buchinsky (1994). The second specification adds a cubic and quartic in potential work experience. Murphy and Welch (1990) indicate that a quartic in potential work experience fits the data better than the standard quadratic in experience used in the Mincer model. Finally, we examine a specification similar to that employed by Buchinsky (1997). The log of real weekly wages is regressed on years of schooling and its square, years of potential experience and its square, schooling interacted with experience, and dummy variables for region, SMSA, region interacted with SMSA, part time work, race, and race interacted with schooling, experience, and part time employment. The variable names and their definitions are:

- rwage*: Real weekly wage. Calculated by taking Total Earnings from wages and salaries last year (1987) divided by weeks worked last year. The weekly wage is deflated by the deflator for Personal Consumption Expenditures where 1992 is the base year.
- ed*: Years of schooling. Years of schooling can take on values from zero to 18.
- exper*: Years of experience: age – years of schooling – 6.
- regne*: Dummy variable = 1 if lives in the North East.
- regmw*: Dummy variable = 1 if lives in the Midwest.
- regw*: Dummy variable = 1 if lives in the West.
- smsa*: Dummy variable = 1 if lives in a Standard Metropolitan Statistical Area (SMSA).
- race*: Dummy variable = 1 if black. Only blacks and whites are included in the sample. The other categories are omitted from the sample.
- parttime*: Dummy variable = 1 if worked less than 35 hours a week at job last year.

The other variables in the model are powers or products of these variables.

6.3. Practical implementation of the ICM test

The ICM tests have been conducted using the weight function functions

$$w_{n,j}(\xi) = \cos(\xi^T \Phi(\tilde{x}_{n,j})) + \sin(\xi^T \Phi(\tilde{x}_{n,j})) \tag{22}$$

(cf. (6)), where $\Phi(x)$ is a vector valued function with components $\arctan(x_i)$, and $\tilde{x}_{n,j}$ is the vector of instruments \tilde{x}_j , in deviation of their sample means and standardized by their sample standard errors. See Bierens (1982, 1990) for the reason for the latter. We recall that only the untransformed regressors need to be used as instruments. Thus, products and powers of explanatory variables are excluded from the list of instruments, because the σ -Algebra generated by these instruments is the same as the σ -Algebra generated by all regressors. The

instrumental variables involved in the Tables 1.B–3.B below are in italic. Moreover, the chosen probability measure μ in the ICM test is the uniform probability measure on the hypercube $\mathcal{E} = [-c, c] \times \cdots \times [-c, c]$, where c is either 1, 5 or 10. The reason for choosing different values of c is the following: With weight function (22) and the uniform probability measure μ involved, the test statistic of the ICM test takes the form

$$\begin{aligned} \hat{F}(c) &= (2c)^{-m} \min_{\theta \in \Theta} \int_{-c}^c \cdots \int_{-c}^c |(1/\sqrt{n}) \sum_{j=1}^n u_j(\theta) w_{n,j}(\xi)|^2 d\xi \\ &= \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n u_i(\theta) u_j(\theta) \prod_{k=1}^m \frac{\sin(c[\arctan(\tilde{x}_{n,i}(k)) - \arctan(\tilde{x}_{n,j}(k))])}{c[\arctan(\tilde{x}_{n,i}(k)) - \arctan(\tilde{x}_{n,j}(k))]}, \end{aligned} \tag{23}$$

where m is the number of instruments, and $u_j(\theta)$ is defined by (2). See (37) in the appendix for the second equality. It follows now from (23) that

$$\hat{F}(0) = \min_{\theta \in \Theta} [(1/\sqrt{n}) \sum_{j=1}^n u_j(\theta)]^2 \rightarrow 0 \tag{24}$$

in distribution, and $\hat{F}(\infty) = 0$. Thus, if we choose c too small or too large, the power of the test will be negatively affected. Actually, the choice of c is crucial for the finite sample power of the test. One might think of choosing c as to maximize the test statistic (23). However, it is not clear whether the convergence in distribution results under the null hypothesis hold uniformly in c . More research is needed to investigate this issue, which is beyond the scope of this paper.

The LAD estimators were used as starting values for computing the ICM test statistics. Since the signs of the large median residuals will likely not change, the minimization of the ICM objective function has been conducted by adjusting the 10% smallest (in absolute value) residuals only, using the downhill simplex method of and Nelder and Mead (1965) (see also Press et al. 1989, pp. 289–293), which do not require the use of derivatives, and after each iteration round we have checked whether the signs of the other 90% residuals have changed. Only one iteration round sufficed.

The integral of the ICM statistic has been computed by Monte Carlo simulation, using 1000 random drawings from the uniform distribution on \mathcal{E} , because the sample size is too large to compute this integral numerically by (23).

A more detailed description of this testing protocol will be given in section 2 of the appendix.

6.4. Median regression estimation and test results

The t -values of the LAD estimators are based on the assumption that the error ε_j is independent of the explanatory variables², as in Koenker and Bassett

² This excludes heteroskedasticity of the errors. Admittedly, this assumption may not be realistic for wage equations. If so, the t -values reported in Tables 1.A–3.A are biased. However, asymptotically the ICM test is not affected by heteroskedasticity. See Assumptions 2 and 4.

Table 1.A. LAD estimation results for the standard Mincer type model

<i>X</i>	estimates	<i>t</i> -values
race	-0.251165	-18.132
ed	0.093462	68.514
exper	0.076289	80.844
exper ²	-0.001274	-62.568
1	4.279231	208.134

Table 1.B. ICM estimation and test results for the standard Mincer type model

<i>X</i>	<i>c</i> = 1	<i>c</i> = 5	<i>c</i> = 10
<i>race</i>	-0.251183	-0.251832	-0.251853
<i>ed</i>	0.093469	0.093711	0.093718
<i>exper</i>	0.076294	0.076492	0.076498
exper ²	-0.001274	-0.001277	-0.001277
1	4.279546	4.290609	4.290952
ICM test	4.224736	27.54178	18.57165
Critical values		10%: 3.23	5%: 4.26

Table 2.A. LAD estimation results for the quartic model

<i>X</i>	estimates	<i>t</i> -values
race	-0.245609	-18.003
ed	0.095481	70.784
exper	0.166344	49.568
exper ²	-0.008562	-30.016
exper ³	0.000201	23.086
exper ⁴	-0.000002	-20.445
1	4.005403	181.053

Table 2.B. ICM estimation and test results for the quartic model

<i>X</i>	<i>c</i> = 1	<i>c</i> = 5	<i>c</i> = 10
<i>race</i>	-0.245609	-0.246033	-0.246020
<i>ed</i>	0.095481	0.095646	0.095641
<i>exper</i>	0.166344	0.166631	0.166622
exper ²	-0.008562	-0.008577	-0.008577
exper ³	0.000201	0.000201	0.000201
exper ⁴	-0.000002	-0.000002	-0.000002
1	4.005398	4.012319	4.401210
ICM test	2.425816	11.26330	7.969391
Critical values		10%: 3.23	5%: 4.26

Table 3.A. LAD estimation results for the Buchinsky type model

<i>X</i>	estimates	<i>t</i> -values
<i>ed</i>	0.137425	17.115
<i>exper</i>	0.086972	43.398
<i>regne</i>	0.046971	2.173
<i>regmw</i>	-0.013488	-0.786
<i>regw</i>	0.042601	2.337
<i>smsa</i>	0.134946	9.895
<i>race</i>	-0.267068	-3.566
<i>parttime</i>	-0.934852	-69.697
<i>exper</i> ²	-0.001063	-48.122
<i>ed</i> ²	-0.000397	-1.516
<i>ed</i> * <i>exper</i>	-0.001785	-16.719
<i>regne</i> * <i>smsa</i>	0.055559	2.314
<i>regmw</i> * <i>smsa</i>	0.097281	4.795
<i>regw</i> * <i>smsa</i>	0.004530	0.212
<i>race</i> * <i>ed</i>	0.000753	0.149
<i>race</i> * <i>exp</i>	0.000678	0.648
<i>race</i> * <i>parttime</i>	0.142179	3.473
<i>l</i>	3.788416	57.192

Table 3.B. ICM estimation and test results for the Buchinsky type model

<i>X</i>	<i>c</i> = 1	<i>c</i> = 5	<i>c</i> = 10
<i>ed</i>	0.137446	0.137841	0.138012
<i>exper</i>	0.086984	0.087234	0.087342
<i>regne</i>	0.046978	0.047114	0.047171
<i>regmw</i>	-0.013490	-0.013529	-0.013545
<i>regw</i>	0.042607	0.042730	0.042782
<i>smsa</i>	0.134966	0.135356	0.135521
<i>race</i>	-0.267107	-0.267876	-0.268206
<i>parttime</i>	-0.934991	-0.937682	-0.938834
<i>exper</i> ²	-0.001063	-0.001066	-0.001067
<i>ed</i> ²	-0.000397	-0.000398	-0.000399
<i>ed</i> * <i>exper</i>	-0.001785	-0.001791	-0.001793
<i>regne</i> * <i>smsa</i>	0.055567	0.055727	0.055796
<i>regmw</i> * <i>smsa</i>	0.097295	0.097575	0.097695
<i>regw</i> * <i>smsa</i>	0.004530	0.004543	0.004549
<i>race</i> * <i>ed</i>	0.000753	0.000756	0.000756
<i>race</i> * <i>exp</i>	0.000678	0.000680	0.000681
<i>race</i> * <i>parttime</i>	0.142110	0.142608	0.142784
<i>l</i>	3.788976	3.799885	3.804553
ICM test	1.981098	3.333470	2.292801
Critical values		10%: 3.23	5%: 4.26

(1978). The error density $f(\cdot)$ has been estimated by a kernel density estimator with standard normal kernel and window width $h = s \cdot n^{-0.2}$, where s is the mean of the absolute values of the LAD residuals in deviation of their sample mean. This scaling factor s makes the kernel density estimator location and scale invariant.

The LAD estimation results are presented in Tables 1.A–3.A, and ICM

estimation and test results are presented in Tables 1.B–3.B. First, observe that the ICM estimators are virtually the same as the LAD estimators. We will explain this in section 3 of the appendix. Next, observe that the ICM tests for $c = 1$ do not reject the three models at the 5% significance level, and that the ICM tests for $c = 5$ and $c = 10$ reject the median regressions in Tables 1.A–2.A at the 5% significance level. The differences of the ICM test statistics for $c = 1, c = 5$ and $c = 10$ in Tables 1.B–2.B indicate that the null hypothesis is not true, despite the test results for $c = 1$. Apparently, the value $c = 1$ is too close to zero [cf. (24)]. Only the Buchinsky type median regression in Table 3.A is accepted by the ICM test at the 10% significance level for $c = 1$ and $c = 10$, and at the 5% significance level for $c = 5$.

These empirical illustrations clearly demonstrate the practical applicability and usefulness of the ICM test for quantile regression models.

Appendix

1. Proof of Theorem 3

We can approximate the median regression model by a continuously differentiable model, in various ways. For example, let for arbitrary small $\varepsilon > 0$,

$$\begin{aligned}
 u_j^{(\varepsilon)}(\theta) &= \int_{-\infty}^{(y_j - \theta^T x_j)/\varepsilon} |v| \exp(-v^2/2) dv - 1 \\
 &= (1 - \exp[-0.5(y_j - \theta^T x_j)^2/\varepsilon^2])u_j(\theta),
 \end{aligned}
 \tag{25}$$

and $\hat{z}^{(\varepsilon)}(\theta, \xi) = (1/n) \sum_{j=1}^n \hat{u}_j^{(\varepsilon)}(\theta)w_j(\xi)$. Then it is easy to verify from (2), (7) and (25) that

$$\begin{aligned}
 |\hat{z}(\theta, \xi) - \hat{z}^{(\varepsilon)}(\theta, \xi)| &\leq \frac{1}{n} \sum_{j=1}^n |u_j(\theta)| \exp(-0.5(y_j - \theta^T x_j)^2/\varepsilon^2) \bar{w} \\
 &\leq \frac{1}{n} \sum_{j=1}^n \exp(-0.5(y_j - \theta^T x_j)^2/\varepsilon^2) \bar{w}.
 \end{aligned}
 \tag{26}$$

It follows from Assumption 3 and Jennrich’s (1969) uniform law of large numbers that for any fixed $c > 0$,

$$\begin{aligned}
 \sup_{\theta \in \Theta, \varepsilon \geq c} |(1/n) \sum_{j=1}^n \exp[-0.5(y_j - \theta^T x_j)^2/\varepsilon^2] \\
 - E(\exp[-0.5(y_1 - \theta^T x_1)^2/\varepsilon^2])| \rightarrow 0 \text{ a.s.},
 \end{aligned}
 \tag{27}$$

Moreover,

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta, 0 \leq \varepsilon \leq c} \frac{1}{n} \sum_{j=1}^n \exp(-0.5(y_j - \theta^T x_j)^2 / \varepsilon^2) \\
 & \leq \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^n \exp(-0.5(y_j - \theta^T x_j)^2 / c^2) \\
 & \leq \limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=1}^n \exp(-0.5(y_j - \theta^T x_j)^2 / c^2) \right. \\
 & \quad \left. - E[\exp(-0.5(y_1 - \theta^T x_1)^2 / c^2)] \right| \\
 & \quad + \sup_{\theta \in \Theta} E[\exp(-0.5(y_1 - \theta^T x_1)^2 / c^2)] \\
 & = \sup_{\theta \in \Theta} E[\exp(-0.5(y_1 - \theta^T x_1)^2 / c^2)] \text{ a.s.} \tag{28}
 \end{aligned}$$

Furthermore, it follows from Assumption 4 that

$$0 \leq E(\exp[-0.5(y_1 - \theta^T x_1)^2 / \varepsilon^2]) \leq \varepsilon \sqrt{2\pi} E[\sup_y f(y|x_1)] \tag{29}$$

hence

$$\lim_{\varepsilon \downarrow 0} \sup_{\theta \in \Theta} E(\exp[-0.5(y_1 - \theta^T x_1)^2 / \varepsilon^2]) = 0. \tag{30}$$

Combining (27) through (30), it follows that for arbitrary $c > 0$ there exists a number $a > 0$ such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta, 0 \leq \varepsilon \leq a} |\hat{Q}(\theta) - \hat{Q}^{(\varepsilon)}(\theta)| < c \text{ a.s.} \tag{31}$$

and similarly

$$\sup_{\theta \in \Theta, 0 \leq \varepsilon \leq a} |\bar{Q}(\theta) - \bar{Q}^{(\varepsilon)}(\theta)| < c, \tag{32}$$

where

$$\begin{aligned}
 \hat{Q}^{(\varepsilon)}(\theta) &= \int |\hat{z}^{(\varepsilon)}(\theta, \xi)|^2 d\mu(\xi), \\
 \bar{Q}^{(\varepsilon)}(\theta) &= \int |E[u_1^{(\varepsilon)}(\theta)w_1(\xi)]|^2 d\mu(\xi).
 \end{aligned} \tag{33}$$

Moreover, it follows from Jennrich’s (1968) uniform law of large numbers that for fixed $\varepsilon > 0$,

$$\hat{Q}^{(\varepsilon)}(\theta) \rightarrow \bar{Q}^{(\varepsilon)}(\theta) \text{ a.s., uniformly on } \Theta. \tag{34}$$

Combining (31), (32) and (34), and letting $c \downarrow 0$, Theorem 3 follows. Q.E.D.

2. The ICM test algorithm

Given model (1), the first thing we have to do is to select the variables to be included in the vector $\tilde{x}_j = (\tilde{x}_j(1), \dots, \tilde{x}_j(m))^T$ of instruments. For example, in the case of the Buchinsky type model in Table 3.B, \tilde{x}_j consists of the ($m = 8$) italic variables, i.e., the smallest set of variables such that the conditional distribution of y_j given x_j is the same as the conditional distribution of y_j given \tilde{x}_j . Next, we have to standardize \tilde{x}_j , as follows (see Bierens 1982, 1990 for the reason):

$$\tilde{x}_{n,j} = (\tilde{x}_{n,j}(1), \dots, \tilde{x}_{n,j}(m))^T, \text{ where}$$

$$\tilde{x}_{n,j}(i) = \frac{\tilde{x}_j(i) - (1/n) \sum_{j=1}^n \tilde{x}_j(i)}{\sqrt{(1/n) \sum_{j=1}^n \tilde{x}_j(i)^2 - ((1/n) \sum_{j=1}^n \tilde{x}_j(i))^2}}$$

for $i = 1, \dots, m$. Moreover, we have to transform $\tilde{x}_{n,j}$ by a bounded one-to-one mapping $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}^m$. We have chosen $\Phi((x_1, \dots, x_m)^T) = (\arctan(x_1), \dots, \arctan(x_m))^T$.

Denote for $\xi \in \mathcal{E} = \times_{i=1}^m [-c, c]$, with $c = 1, 5$ or 10 ,

$$u_{p,n,j}(\theta_1, \theta) = d_{p,j}(\theta_1)u_j(\theta) + (1 - d_{p,j}(\theta_1))u_j(\theta_1) \tag{35}$$

and $z_{p,n}(\theta_1, \theta, \xi) = (1/n) \sum_{j=1}^n u_{p,n,j}(\theta_1, \theta)w_{n,j}(\xi)$, where $w_{n,j}(\xi)$ is defined by (22), $u_j(\theta)$ is defined by (2), and $d_{p,j}(\theta_1)$ is a dummy variable which takes the value 1 if j belongs the set of the $p\%$ observations with the smallest value of $|y_j - \theta_1^T x_j|$, and zero if not. In the empirical applications we have chosen $p = 10$.

Let μ be the uniform probability measure on \mathcal{E} . Then the integral $\hat{Q}_p(\theta_1, \theta) = \int |z_{p,n}(\theta_1, \theta, \xi)|^2 d\mu(\xi)$ can be approximated by

$$\tilde{Q}_{N,p}(\theta_1, \theta) = \frac{1}{N} \sum_{s=1}^N z_{p,n}(\theta_1, \theta, \xi_s)^2, \tag{36}$$

were ξ_1, \dots, ξ_N are random drawings from the uniform distribution on \mathcal{E} . In the empirical application we have chosen $N = 1000$.

Note that, due to the boundedness of $w_{n,j}(\xi)$, (36) is a mean of bounded random variables $z_{p,n}(\theta_1, \theta, \xi_s)^2$, conditional on the data and given the values of θ_1 and θ . This will boost the performance of the law of large numbers on which this approximation relies. Moreover, it is not hard to verify that for the weight function (22) and the uniform probability measure μ on $\mathcal{E} = \times_{i=1}^m [-c, c]$,

$$\int w_{n,i}(\xi)w_{n,j}(\xi) d\mu(\xi) = \prod_{k=1}^m \frac{\sin(c[\arctan(\tilde{x}_{n,i}(k)) - \arctan(\tilde{x}_{n,j}(k))])}{c[\arctan(\tilde{x}_{n,i}(k)) - \arctan(\tilde{x}_{n,j}(k))]}, \tag{37}$$

hence $\int w_{n,j}(\xi)^2 d\mu(\xi) = 1$.

Now given the initial value $\theta_1 = \tilde{\theta}_{LAD}$, minimize $\tilde{Q}_{N,p}(\theta_1, \theta)$ to θ , using the simplex method of Nelder and Mead (1965), which yields $\theta_2 = \operatorname{argmin}_{\theta} \tilde{Q}_{N,p}(\theta_1, \theta)$. Since the objective function is piecewise constant, the vertexes of the start-simplex should be chosen sufficiently far away from the start point, say by line search in the principle directions up to the point where the objective function becomes unequal to its value in θ_1 .

If $u_j(\theta_1) = u_j(\theta_2)$ for $d_{p,j}(\theta_1) = 0$ then we are done: $\hat{\theta} \approx \tilde{\theta} = \theta_2$, and the simulated ICM test statistic becomes

$$\hat{F} \approx n\tilde{Q}_{N,p}(\tilde{\theta}, \tilde{\theta}), \tag{38}$$

else we repeat the minimization procedure with θ_1 replace by θ_2 .

It would be better to conduct this algorithm on the basis of the exact integral $\hat{Q}_p(\theta_1, \theta)$, but this will involve $m(n^2 - n)/2$ multiplications of different pairs of (35) and the m factors of the product in (37), plus n squares of (35). Since our sample size is $n = 28155$, the computation of the integral $\hat{Q}_p(\theta_1, \theta)$ will take therefore too long on a regular PC (The computation of each of the three ICM test statistics (38) for the model in Table 3.B took about 8 hours on a Pentium II PC, using *EasyReg*).

It is easy to verify that in order to extend the ICM test to more general quantile regressions, with $P(y_j - \theta_0^T x_j \leq 0 | x_j) = q$, we only need to redefine $u_j(\theta)$ in (2) as

$$u_j(\theta) = \frac{I(y_j - \theta^T x_j > 0) - I(y_j - \theta^T x_j < 0) + 2q - 1}{2\sqrt{q(1 - q)}},$$

where the scaling involved guarantees that under the null hypothesis, $E[u_j(\theta_0)^2] = 1$, and start the minimization of (36) from the corresponding quantile estimator.

3. Why are the LAD and ICM estimators close?

In order to explain why the LAD and ICM estimators in Tables 1–3 are so close, suppose that the correct median regression model is $y_j = g(x_j) + \varepsilon_j$, and that ε_j is independent of x_j , with distribution function $F(\cdot)$ satisfying $F(0) = 0.5$. Moreover, let $H(\cdot)$ be the distribution function of x_j . Since the LAD estimator is actually a method of moment estimator, namely the solution of the moment conditions $(1/n) \sum_{j=1}^n u_j(\theta)x_j = 0$, it converges a.s. to

$$\begin{aligned} \theta_* &= \operatorname{argmin}_{\theta} \|E[u_j(\theta)x_j]\|^2 = \operatorname{argmin}_{\theta} \|E[(1 - 2F(g(x_j) - \theta^T x_j))x_j]\|^2 \\ &= \operatorname{argmin}_{\theta} \iint (1 - 2F(g(x_1) - \theta^T x_1))(1 - 2F(g(x_2) - \theta^T x_2)) \\ &\quad \times W_1(x_1, x_2) dH(x_1) dH(x_2), \end{aligned} \quad (39)$$

where $W_1(x_1, x_2) = x_1^T x_2$, whereas the ICM estimator with weight function $w(\xi^T \Phi(x_j))$ converges to

$$\begin{aligned} \theta_{**} &= \operatorname{argmin}_{\theta} \int \|E[u_j(\theta)w(\xi^T \Phi(x_j))]\|^2 d\mu(\xi) \\ &= \operatorname{argmin}_{\theta} \iint (1 - 2F(g(x_1) - \theta^T x_1))(1 - 2F(g(x_2) - \theta^T x_2)) \\ &\quad \times W_2(x_1, x_2) dH(x_1) dH(x_2), \end{aligned} \quad (40)$$

where $W_2(x_1, x_2) = \int w(\xi^T \Phi(x_1))w(\xi^T \Phi(x_2)) d\mu(\xi)$. Therefore, asymptotically the difference between the LAD estimator and the ICM estimator will not be substantial if the misspecification is rather modest, because the objective functions (39) and (40) only differ with respect to the weight functions W_1 and W_2 .

References

- Bierens HJ (1982) Consistent model specification tests. *Journal of Econometrics* 20:105–134
- Bierens HJ (1990) A consistent conditional moment test of functional form. *Econometrica* 58:1443–1458
- Bierens HJ (1994) Topics in advanced econometrics: Estimation, testing, and specification of cross-section and time series models. Cambridge University Press, Cambridge
- Bierens HJ, Ploberger W (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65:1129–1151
- Billingsley P (1968) Convergence of probability measures. John Wiley, New York
- Buchinsky M (1994) Changes in the U.S. wage structure 1963–1987: Application of quantile regression. *Econometrica*, 62:405–458
- Buchinsky M (1995) Quantile regression box-cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics* 65:109–154
- Buchinsky M (1997) Recent advances in quantile regression models: A practical guide for empirical research. *Journal of Human Resources* (forthcoming)
- Chamberlain G (1994) Quantile regression, censoring, and the structure of wages. In: Sims C (ed.) Proceedings of the Sixth World Congress of the Econometric Society. Cambridge University Press, New York
- Chernozhukov V, Umantsev L (2001) Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* (this issue)
- Fitzenberger B (1997) A guide to censored quantile regression. In: Madala GS, Rao CR (ed.) Handbook of Statistics, Vol. 15. Elsevier, Amsterdam
- Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–1054
- Jennrich RI (1969) Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40:633–643
- Koenker R, Bassett G (1978) Regression quantiles. *Econometrica* 46:33–50

- Murphy KM, Welch F (1990) Empirical age-earnings profiles. *Journal of Labor Economics* 8:202–229
- Nelder JA, Mead R (1965) A simplex method for function minimization. *The Computer Journal* 7:308–313
- Poterba JM, Reuben KS (1994) The distribution of public sector wage premiums: New evidence using quantile regression methods. NBER Working Paper
- Powell JL (1984) Least absolute deviations estimation of the censored regression model. *Journal of Econometrics* 25:303–325
- Press WH., Flannery BP, Teukolsky SA, Vetterling WT (1989) *Numerical recipes (fortran version)*. Cambridge University Press, Cambridge
- Stinchcombe MB, White H (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14:295–325
- Zheng JX (1998) A consistent nonparametric test of parametric regression models under conditioning quantile restrictions. *Econometric Theory* 14:123–138