

Grażyna Trzpiot^{*},

SOME TESTS FOR QUANTILE REGRESSION MODELS

Abstracts. We present a specification test for quantile regression models. Although researchers commonly estimate and conduct statistical inference for quantile regression models, rarely do not check the validity of specified models. We present a test for the functional forms of quantile regression models. This test add powers of fitted dependent variables as regressors and check the significance of those added regressors. Additionally we compare nonparametric specification tests, based on kernel functions and bandwidth parameters.

Key words: test, quantile regression model

I. INTRODUCTION

Median and quantile estimation methods have recently been applied to economic models because these methods impost fewer restrictions on the data than mean regression. 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.

Quantile regression has become a standard tool for statistical analysis (Koenker and Bassett, 1978). Although researchers rarely do any check the validity of specified models. In this paper we present a specification test for the functional forms of quantile regression models: quantile regression specification error test. Using the quantile regression estimators instead of the least square estimators, the implementation of the test is similar to regression specification error test by Ramsey (1969) and Ramsey and Schmidt (1976). Additionally to assess the predictive performance of the quantile regression models, we follow Christoffersen's (1998) framework, which is designed for evaluating the accuracy of interval forecasts of quantile.

^{*} Profesor, Akademia Ekonomiczna w Katowicach.

II. QUANTILE REGRESSION AND TEST STATISTIC

Let $\{(y_i, x_i)\}$ for $i = 1, \dots, n$ be an iid sample from the distribution of (y, x) , where $y \in \mathbb{R}$ is used as a regressand and $x \in \mathbb{R}^k$ is used as a regressor vector. Let $F_{y|x}$ be the conditional distribution function of y given x . The τ -th conditional quantile function of y given x is defined as

$$Q_\tau(y|x) = \inf\{y \mid F_{y|x}(y|x) \geq \tau\}.$$

The (linear) quantile regression model is written as

$$Q_\tau(y|x) = x'\beta(\tau) \text{ a.s.}, \quad (1)$$

where $\tau \in (0, 1)$ is a fixed and known quantile of interest, $\beta(\tau)$ is a $k \times 1$ vector of unknown regression coefficients. Note that in general the coefficients $\beta(\tau)$ vary with τ . The conventional quantile regression estimator by Koenker and Bassett (1978) is defined as

$$\hat{\beta}(\tau) = \min_{b \in \mathbb{R}^k} \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - x_i'b) \quad (2)$$

where $\rho_\tau(z) = |\tau - I(z < 0)| \cdot |z|$, is called the check function and I is the indicator function¹.

When $\tau = 0.5$, $\hat{\beta}(\tau)$ corresponds to the least absolute deviation (LAD) estimator. Let $f_{y|x}$ be the conditional density function of y given x . Suppose that the quantile regression model is correctly specified, i.e., the relation (1) holds.

Then under mild regularity conditions (Koenker 2005), we have

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \rightarrow N(0, V(\tau))$$

where

$$V(\tau) = \tau(1-\tau)E[f_{y|x}(x'\beta(\tau))xx']^{-1}E[xx']E[f_{y|x}(x'\beta(\tau))xx']^{-1}.$$

On the other hand, if the linear functional form in (1) is misspecified and the true conditional quantile function is written as $Q_\tau(y|x) = \theta(x, \tau) \neq x'\beta(\tau)$, then Angrist et al. (2006) and Koenker (2005) demonstrate that the quantile regression estimator $\hat{\beta}(\tau)$ converges to

¹ $I[A] = 1$ if A is true, $I[A] = 0$ otherwise

$$\beta^*(\tau) = \min_{b \in R^k} E[\rho_\tau(y_i - x_i' b)] = \min_{b \in R^k} E[w(x, b)(x_i' b - \theta(x, \tau))^2]$$

where

$$w(x, b) = \int_0^1 (1-u) f_{y|x}(\theta(x, \tau) + u(x' b - \theta(x, \tau))) du.$$

Therefore, if the quantile regression model is misspecified, $\hat{\beta}(\tau)$ can be a misleading estimator for the marginal effect of x on the conditional quantile of y (i.e., $\partial Q_\tau(y|x)/\partial x$ for continuous regressors).

In order to avoid such a misspecification problem, it can apply the idea of Ramsey's (1969) test to our quantile regression problem. Classical test focuses on misspecification of the conditional mean function $E[y|x]$, we need new test focus on the conditional quantile function $Q_\tau(y|x) = \theta(x, \tau)$.

If $\theta(x, \tau)$ is sufficiently smooth in $x = (x_1, \dots, x_k)'$, then for a Taylor expansion around $x = 0$ under mild regularity conditions (Koenker, 2005) we have

$$\begin{aligned} Q_\tau(y|x) = & \theta(0, \tau) + \sum_{j=1}^k \frac{\partial \theta(x, 0)}{\partial x_j} x_j + \frac{1}{2!} \sum_{j_1=1}^k \sum_{j_2=1}^k \frac{\partial^2 \theta(x, 0)}{\partial x_{j_1} \partial x_{j_2}} x_{j_1} x_{j_2} \\ & + \dots + \frac{1}{p!} \sum_{j_1=1}^k \dots \sum_{j_p=1}^k \frac{\partial^p \theta(x, 0)}{\partial x_{j_1} \dots \partial x_{j_p}} x_{j_1} \dots x_{j_p} + \dots \end{aligned}$$

Thus, departures from the linear functional form can be assessed by checking the significance of higher-order polynomials of x . However, if the dimension of x is large, it is not practical to include all components that appear in the p -th order polynomial. To avoid this practical problem, we substitute the effect of the p -th order polynomials by the p -th power of the fitted variable $\hat{y}_p = (x' \hat{\beta}(\tau))^p$.

We can approximate conditional quantile function

$$Q_\tau(y|x) \text{ by } x' \beta(\tau) + w_p' \alpha_p(\tau), \quad (3)$$

where $w_p = (\hat{y}_2, \dots, \hat{y}_p)'$ and $\alpha_p(\tau)$ is a $(p-1) \times 1$ parameter vector.

Based on this approximation, the QRESET(p) for the linear specification against the p -th order polynomial is defined as a joint significance test of

$$H_0: \alpha_p(\tau) = 0. \quad (4)$$

III. IMPLEMENTATION OF THE TEST STATISTIC

A procedure to implement the QRESET(p) is described as follows:

1. Compute $\hat{\beta}_p(\tau)$ by (2), and set $\hat{y}_i = x'_i \hat{\beta}_p(\tau)$ for $i = 1, \dots, n$.
2. Using $\{\hat{y}_i\}_{i=1, \dots, n}$, compute

$$(\hat{\beta}_p(\tau), \hat{\alpha}_p(\tau)) = \min_{(b,a) \in \mathbb{R}^k \times \mathbb{R}^{p-1}} \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - x'_i b - w'_{pi} a), \quad (5)$$

and estimate the asymptotic variance of $\hat{\alpha}_p(\tau)$. This computation can be conducted by standard packages for quantile regression.

3. Using $\hat{\alpha}_p(\tau)$ and its asymptotic variance estimator (denoted by $\hat{V}_p(\tau)$), compute the Wald test statistic for (4):

$$W = \hat{\alpha}_p(\tau)' \hat{V}_p(\tau)^{-1} \hat{\alpha}_p(\tau), \quad (6)$$

and conduct the significance test by the $\chi^2_{(p-1)}$ critical value (if $p = 2$, we can use the t-value for $\hat{\alpha}_2(\tau)$).

4. If H_0 is rejected, we suspect misspecification of the linear functional form in (1).

One of the main advantages of the QRESET is its low computational cost: it simply requires an additional quantile regression (5). Compared with existing nonparametric specification tests such as Zheng (1998), the QRESET does not include any kernel function and bandwidth parameter whose choices complicate the practical implementation of nonparametric tests. On the other hand, a disadvantage of the QRESET compared to the nonparametric tests is that the QRESET may lack power when the polynomial approximation in (3) is not accurate. This point will be formalized in the following discussion.

IV. PROPERTIES OF THE TEST STATISTIC

Suppose that $z = (y, x)'$ is continuously distributed. Then the τ -th conditional quantile $Q_\tau(y|x)$ satisfies $\int_{-\infty}^{Q_\tau(y|x)} f_{y|x}(y|x) dy = \tau$, and the quantile regression model (1) is equivalent to the following conditional moment restriction,

$$E[g(z, \beta(\tau))|x] = E[\tau - I\{y - x'\beta(\tau) \leq 0\}|x] = 0 \quad \text{a.s.}, \quad (7)$$

where g is implicitly defined. The null hypothesis of correct specification is written as

$$H_0^* : \Pr\{E[g(z, \beta(\tau))|x] = 0\} = 1, \quad \text{for some } \beta(\tau),$$

and the alternative is that H_0^* is false. Also, note that (7) implies

$$E[h(x)g(z, \beta(\tau))] = 0, \quad \text{for any (measurable) } h. \quad (8)$$

In other words, the conditional moment restriction (7) implies infinitely many unconditional moment restrictions in the forms of (8). If (8) is rejected for some h , we can reject H_0^* as well.

From Koenker (2005), the first-order condition of the minimization problem in (5) is written as

$$\frac{1}{n} \sum_{i=1}^n (x_i', w_{pi}')' (\tau - I\{y_i - x_i' \beta_p(\tau) - w_{pi}' \alpha_p(\tau) \leq 0\}) = o_p(n^{-1/2}), \quad (9)$$

and its population analog is

$$E[(x', w_p')' (\tau - I\{y - x'\beta_p(\tau) - w_p' \alpha_p(\tau) \leq 0\})] = 0. \quad (10)$$

Therefore, the QRESET statistic is considered as the Wald test statistic for the parameter restriction H_0 in (4) under the unconditional moment restriction model (10). From Newey and McFadden (1994), the Wald and GMM distance test statistics are asymptotically equivalent in the standard GMM setup. In addition, since the moment restriction (10) is just identified, the GMM objective function for (10) converges to zero in probability. Therefore, the QRESET statistic is asymptotically equivalent to the (normalized) GMM objective function of the following unconditional moment restriction:

$$E[(x', w_p')' g(z, \beta(\tau))] = 0. \quad (11)$$

Since the QRESET statistic is asymptotically equivalent to the over identification test statistic for the unconditional moment restriction (11), it has the $\chi^2_{(p-1)}$ limiting distribution (Newey and McFadden 1994). Thus, from (8) and (11), the QRESET has nontrivial power against some forms of

misspecification, in which h is written as a p -th order polynomial function.² However, the QRESET is not necessarily powerful against *all* types of h , i.e., the QRESET is inconsistent for the null of necessarily powerful against *all* types of h , i.e., the QRESET is inconsistent for the null of correct specification H_0^* .

To avoid the inconsistency problem for testing H_0^* , we need to introduce a nonparametric approach, such as Zheng (1998) or Bierens and Ginther (2001). However, Zheng's (1998) test requires the choices of a kernel function and a bandwidth parameter, and Bierens and Ginther's (2001) test requires the choice of a weight function. Thus, the implementation of these nonparametric tests can be cumbersome. Also the computational costs of the nonparametric tests can be substantially more expensive than that of the QRESET.

Note that the QRESET can be used to check specification of multiple quantile regression models $Q_{\tau}(y|x) = x'\beta(\tau)$ for $\tau = 1, \dots, m$. In this case, we approximate $Q_{\tau}(y|x)$ by $x'\beta_p(\tau) - w_p \alpha_p(\tau)$ and estimate $\alpha_p(\tau)$ by the quantile regression estimator for each $l = 1, \dots, m$.

Then the multiple QRESET(p) can be defined as the Wald test for the joint hypothesis

$$H_0: \alpha_p(\tau_1) = \dots = \alpha_p(\tau_m) = 0.$$

By applying Koenker (2005) we can derive the asymptotic property of the Wald test.

It can be suggest to use the QRESET as a practical complement to nonparametric tests particularly in small samples. Also can be recommend to use combinations of the QRESETs with different orders to detect various types of misspecifications.

Advantages of the QRESET over the nonparametric specification tests are:

(i) since the QRESET is free from the choices of kernel functions and bandwidth parameters, the QRESET is easier to implement than nonparametric tests, and

(ii) in small sample sizes (typically fewer than 100), the QRESET can be more powerful than Zheng's (1998) test.

On the other hand, a disadvantage of the QRESET is the fact that the QRESET is not necessarily consistent for all types of misspecification. The QRESET is derived from a polynomial approximation of the conditional quantile function, so, it may have poor power performance when this polynomial approximation is inaccurate. Based on these arguments it can be suggest to use the QRESET as a complement to the nonparametric specification tests particularly in small samples.

² Based on the notion of the information matrix test, Kim and White (2002) considered a different choice for $h(\cdot)$

V. DYNAMIC QUANTILE TEST

To assess the predictive performance of the models under consideration, we follow Christoffersen's (1998) framework, which is designed for evaluating the accuracy of out-of-sample interval forecasts VaR. Defining $H_t = \mathbf{I}(r_t < -\text{VaR}_t)$, Christoffersen (1998) terms the sequence of VaR forecasts efficient with respect to F_{t-1} if

$$E(H_t | F_{t-1}) = \alpha \quad (12)$$

which, by applying iterated expectations, implies that H_t is uncorrelated with any function of a variable in the information set available at $t - 1$. If Equation (12) holds, then VaR violations will occur with the correct conditional and unconditional probability, and neither the forecast for VaR_t nor that for H_t could be improved.

Equation (12) is stronger than correct conditional coverage; it suggests that any $x_{t-1} \in F_{t-1}$ be uncorrelated with H_t . In particular, Engle and Manganelli (2004) remark that conditioning violations on the VaR for the period itself is essential. To illustrate this point, they left $\{\text{VaR}_t\}_{t=1}^T$ be a sequence of iid random variables such that

$$\text{VaR}_t = \begin{cases} K, & \text{with probability } 1 - \alpha \\ -K, & \text{with probability } \alpha \end{cases} \quad (13)$$

As with all asymptotically motivated inferential procedures, the actual size of the tests for finite samples can deviate from their nominal sizes. Lopez (1997) examines the size of unconditional and conditional coverage tests via simulation, as well as their power against various model misspecifications. For a sample size of 500 observations, he finds both tests to be adequately sized. Even for such a small sample, power appears to be reasonable. For the LR_{uc} test, for example, he reports that, for α values of 0,05 or smaller, if the true data generating process is conditionally heteroskedastic, then power is well above 60% for wrong distributional assumptions for the innovations. In general, the tests have only moderate power when volatility dynamics are closely matched but power increases under incorrect innovation distributions, especially further out in the tails.

For K very large and conditioning also on VaR_t , the violation sequence exhibits correct conditional coverage but, conditional on VaR_t , the probability of a violation is either almost zero or almost one. None of the above tests has power against this form of inefficiency. To operationalize equation (12), one

can, similar to Christoffersen (1998) and Engle and Manganelli (2004), regress H_t on a judicious choice of explanatory variables in F_{t-1} , for example:

$$H_t = \lambda_0 + \sum_{i=1}^p \beta_i H_{t-1} + \beta_{p+1} V \hat{\alpha} R_t + \mu_t$$

where, under the null hypothesis:

$$H_0: \lambda_0 = \lambda \text{ and } \beta_i = 0, \text{ for } i = 1, \dots, p+1.$$

In vector notation, we have:

$$H - \lambda \mathbf{1} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$u_t = \begin{cases} -\alpha, & \text{with probability } 1-\alpha \\ 1-K, & \text{with probability } \alpha \end{cases}$$

where $\beta_0 = \lambda_0 - \lambda$ and $\mathbf{1}$ is a conformable vector of ones.

Under the null hypothesis, of equation (12), the regressors should have no explanatory power, that is $H_0: \beta = 0$. Because the regressors are not correlated with the dependent variables under the null hypothesis, invoking a suitable central limit theorem yields

$$\hat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(H - \lambda \mathbf{1}) \sim N(0, (\mathbf{X}'\mathbf{X})^{-1}(\lambda(1-\lambda))),$$

from which Engle and Manganelli (2004) deduce the test statistic

$$DQ = \frac{\hat{\beta}_{LS}' \mathbf{X}' \mathbf{X} \hat{\beta}_{LS}}{\lambda(1-\lambda)} \sim \chi_{p+2}^2. \quad (14)$$

VI. NONPARAMETRIC ESTIMATION OF CONDITIONAL QUANTILES

Denote, as earlier, the τ quantile of the distribution Y given $X = x$ as $Q_y(\tau/X)$ which solves

$$F(Q_y(\tau/X) / x) = \tau \quad (15)$$

where $F(y/x)$ is the conditional cumulative distribution of Y given x evaluated at $Y = y$ an estimate $\hat{Q}(\tau|x)$ can be obtained from the observed pairs (X_i, Y_i) ($i = 1, \dots, n$) by solving (1) after replacing F with some estimate \hat{F} . One choice of \hat{F} , which smoothes over X , is

$$\hat{F}(y|x) = \frac{\sum_{i=1}^n K\{(X_i - x)/h\}I[Y_i \leq y]}{\sum_{i=1}^n K\{(X_i - x)/h\}} \quad (16)$$

where K is a kernel function and I is the indicator function, h is the bandwidth parameter. For chosen h we can use a cross validation approach to minimize the loss function

$$L(h) = \sum_{i=1}^n \rho_\tau(z)(Y_i - \hat{Q}_\tau^{(i)}(X_i)) \quad (17)$$

where $\rho_\tau(z)$ can be interpreted as the loss function (Koenker, R., G. Bassett. (1978)) and $\hat{Q}_\tau^{(i)}$ denotes the estimate of $Q_\tau(X_i)$ using bandwidth h , where observation i has been dropped from a sample.

Equivalently, the nonparametric quantile regression estimator $\hat{Q}_\tau^{(i)}$ can be defined to minimize

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{h^p} K\left(\frac{x - x_i}{h}\right) \rho_\tau(y_i - \hat{Q}_\tau^{(i)}) \quad (18)$$

over all Q_τ .

For nonparametric quantile regression Yu and Jones (1998) suggests the automatic bandwidth selection strategy for smoothing conditional quantiles, which minimizes mean squared error of the conditional quantile functions as follows:

$$h_\tau = h_{mean} \{\tau(1-\tau) / \phi(\Phi^{-1}(\tau))^2\}^{0.5} \quad (19)$$

where ϕ and Φ are the standard normal density and cumulative distribution functions.

VII. CONCLUSION

Compared with existing nonparametric specification tests, the QRESET does not contain kernel functions and bandwidth parameters, and thus is easy to implement. Although the QRESET is not necessarily consistent against all types of misspecifications, simulation results indicate that the QRESET has reasonable size and power properties and can be more powerful than nonparametric specification tests in small samples.

Quantile regression allows us to directly model conditional VaR, utilizing only the pertinent information that determines quantiles of interest. This is contrast with the traditional methods that use information on the central moments of conditional distribution – mean, variance, kurtosis etc. to construct the VaR estimates. From this point of view quantile regression is important for modeling intermediate and extremal conditional VaR.

REFERENCES

- Angrist J, Chernozhukov V, Fernández-Val I. (2006), Quantile regression under misspecification, with an application to the US wage structure. *Econometrica* 74:539–563
- Bierens H. J., Ginther D. K. (2001), Integrated conditional moment testing of quadratic regression models. *Empir Econ* 26:307–324
- Chernozhukoy V., Umantsev L.. (2001), *Conditional Value-at-Risk: Aspects of Modeling and Estimation*. *Empirical Economics*, 26, 271–292.
- Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review* 39, 841–862.
- Engle, R. Manganelli F. S.. (2004), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics*, 22, 367–381.
- Ihaka R., Gentleman R. (1996), R: A language for data analysis and graphics. *J Comput Graph Stat* 5:299– 314
- Kim T-H., White H. (2002), Estimation, inference, and specification testing for possibly misspecified quantile regression. Working paper
- Koenker R. (2005), Quantile regression. Cambridge University Press, Cambridge
- Koenker R, Bassett G. (1978), Regression quantiles. *Econometrica* 46:33–50
- Koenker, R. Hallok K. F. (2001), Quantile Regression. *Journal of Economic perspective*, 15,4, 143-156.
- Koenker, R. Portnoy S.. (1997), Quantile Regression. Working Paper 97-0100, University of Illinois at Urbana-Champaign.
- Koenker, R., Bassett G.. (1978), Regression Quantiles. *Econometrica* 46, 33–50.
- Koenker, R., Machado G. (1999), Goodness of fit and related Inference processes for quantile regression, *Journal of American Statistical Association*, 94, 448, 1296-1310.
- Lopez, J. A. (1997), Regulatory Evaluation of Value-at-Risk Models. Staff Report 33, November 1997, Federal Reserve Bank of New York.
- Newey W.K., McFadden D.L. (1994), Large sample estimation and hypothesis testing. In: Engle R.F., McFadden D..L (eds) *Handbook of econometrics*, vol IV. Elsevier, Amsterdam, pp 2213–2245
- Parzen M.I., Wei L., Ying Z. (1994), A resampling method based on pivotal estimating functions. *Biometrika* 81:341–350

- Ramsey J.B. (1969), Tests for specification errors in classical linear least-squares regression analysis. *J R Stat Soc B* 31:350–371
- Ramsey J.B., Schmidt P. (1976), Some further results on the use of OLS and BLUS residuals in specification error tests. *J Am Stat Assoc* 71:389–390
- Trzpiot G. (2007), Regresja kwantylowa a estymacja VaR, *Prace Naukowe AE Wrocław* 1176, 465–471.
- Trzpiot G. (2008), Implementacja metodologii regresji kwantylowej w estymacji VaR, *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania nr 9*, 316 – 323, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
- Trzpiot G. (2009), Estimation methods for quantile regression, *Studia Ekonomiczne* 53, 81–90, *Zeszyty Naukowe AE Katowice*
- Yu and Jones (1998), Local linear quantile regression, *Journal of American Statistical Association*, 93,441, 228–237.
- Zheng JX (1998), A consistent nonparametric test of parametric regression models under conditioning quantile restrictions. *Econom Theory* 14:123–138

Grażyna Trzpiot

WYBRANE TESTY MODELU REGRESJI KWANTYLOWEJ

Przedstawiamy test weryfikujący jakość specyfikacji modelu regresji kwantylowej. Często wyznaczamy modele regresji kwantylowej i przeprowadzamy dalsze wnioskowanie analizując jedynie poziom błędów. Przedstawimy test dla funkcjonalnej formy modelu regresji kwantylowej. Test dopasowuje zmienną zależną jako wyjaśniającą oraz sprawdza istotność wprowadzonej do modelu zmiennej. Dodatkowo porównamy z nieparametryczną specyfikacją modelu wykorzystującą funkcję jądrową oraz przedział parametrów.
Słowa kluczowe: test, model regresji kwantylowej.