Nonparametric Series Quantile Regression: Modeling, Estimation and Inference[1]

Alexandre Belloni*, Victor Chernozhukov†, and Iván Fernández-Val‡
*Duke University, abn5@duke.edu
†Massachussets Institute of Technology, vchern@mit.edu
‡Boston University, ivanf@bu.edu

Abstract

In this paper we develop the nonparametric QR series framework, covering many regressors as a special case, for performing inference on the entire conditional quantile function and its linear functionals. In this framework, we approximate the entire conditional quantile function by a linear combination of series terms with quantile-specific coefficients and estimate the function-valued coefficients from the data. We develop large sample theory for the empirical QR coefficient process, namely we obtain uniform strong approximations to the empirical QR coefficient process by conditionally pivotal and Gaussian processes, as well as by gradient and weighted bootstrap processes. We apply these results to obtain estimation and inference methods for linear functionals of the conditional quantile function, such as the conditional quantile function itself, its partial derivatives, average partial derivatives, and conditional average partial derivatives. We demonstrate the practical utility of these results with a numerical example calibrated to a demand for gasoline empirical application.

Keywords: Quantile regression series processes, uniform inference.

1 Introduction

Quantile regression (QR) is a principal regression method for analyzing the impact of covariates on outcomes, particularly when the impact might be heterogeneous. This impact is characterized by the conditional quantile function and its functionals [1, 3, 4]. For example, we can model the log of the individual demand for some good, Y, as a function of the price of the good, the income of the individual, and other observed individual characteristics X and an unobserved preference U for consuming the good, as

$$Y = Q(X, U),$$

where the function Q is strictly increasing in the unobservable U. With the normalization that $U \sim \text{Uniform}(0, 1)$ and the assumption that U and X are independent, the function $Q(X, u)$ is the u-th conditional quantile of Y given X, i.e. $Q(X, u) = Q_{Y|X}(u|X)$. This function can be used for policy analysis. For example, we can determine how changes in taxes for the good could impact demand heterogeneously across individuals.

In this paper we develop the nonparametric QR series framework for performing inference on the entire conditional quantile function and its linear functionals. In this

[1] This paper is an extract from Belloni, Chernozhukov, and Fernández-Val [2]. The material is based upon work supported by the National Science Foundation under Grant No. 1060889. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

framework, we approximate the entire conditional quantile function $Q_{Y|X}(u|x)$ by a linear combination of series terms, $Z(x)'\beta(u)$. The vector $Z(x)$ includes transformations of x that have good approximation properties such as powers, trigonometrics, local polynomials, or B-splines. The function $u \mapsto \beta(u)$ contains quantile-specific coefficients that can be estimated from the data using the QR estimator of Koenker and Bassett [5]. As the number of series terms grows, the approximation error $Q_{Y|X}(u|x) - Z(x)'\beta(u)$ decreases, approaching zero in the limit. By controlling the growth of the number of terms, we can obtain consistent estimators and perform inference on the entire conditional quantile function and its linear functionals. The QR series framework also covers as a special case the so called many regressors model, which is motivated by many new types of data that emerge in the new information age, such as scanner and online shopping data.

Notation. In what follows, S^{m-1} denotes the unit sphere in $\mathbb{R}^m$. For $x \in \mathbb{R}^m$, we define the Euclidian norm as $\|x\| := \sup_{\alpha \in S^{m-1}} |\alpha' x|$. For any two real numbers a and b, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Calligraphic letters are used to denote the support of interest of a random variable or vector. For example, $\mathcal{U} \subset (0, 1)$ is the support of U, $\mathcal{X} \subset \mathbb{R}^d$ is the support of X, and $\mathcal{Z} = \{Z(x) \in \mathbb{R}^m : x \in \mathcal{X}\}$ is the support of $Z = Z(X)$. The relation $a_n \lesssim b_n$ means that $a_n \leq C b_n$ for a constant C and for all n large enough. The operator E denotes the expectation with respect to the probability measure P, $\mathbb{E}_n$ denotes the expectation with respect to the empirical measure, and $\mathbb{G}_n$ denotes $\sqrt{n}(\mathbb{E}_n - E)$.

2 Model and estimators

The set-up corresponds to a nonparametric series framework:

$$Y = Q_{Y|X}(U|X) = Z'\beta(U) + R(X, U), \quad U|X \sim \text{Uniform}(0, 1) \quad \text{and} \quad \beta(u) \in \mathbb{R}^m,$$

where X is an elementary d-dimensional regressor, and $Z = Z(X)$ is an m-dimensional vector of approximating functions formed on the basis of this regressor with the first component of Z equal to one; the term $R(X, U)$ is the approximation error term, see for instance Newey (1997). The "population" coefficient $\beta(u)$ is defined for each quantile index $u \in (0, 1)$ as the minimizer of

$$\mathbb{E}[\rho_u(Y - Z'\beta)] \tag{2.1}$$

where $\rho_u(z) = (u - 1\{z < 0\})z$ is the check function (Koenker, 2005).

We consider estimation of this model using the entire quantile regression process $\{u \mapsto \widehat{\beta}(u), u \in \mathcal{U}\}$, namely for each $u \in \mathcal{U}$, a compact subset of $(0, 1)$, the estimator $\widehat{\beta}(u)$ solves the empirical analog of equation (2.1)

$$\mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)]. \tag{2.2}$$

We are also interested in various linear functionals of $\widehat{\beta}(\cdot)$, primarily the conditional quantile function $(z, u) \mapsto z'\widehat{\beta}(u)$, and derivatives of this function with respect to some of the elementary regressors.

We make the following primitive assumptions on the data generating process, as $n \to \infty$ and $m = m(n) \to \infty$.

S.1. Data $\{(Y_i, X_i')', 1 \leq i \leq n\}$ are an i.i.d. sequence of real $(1 + d)$-vectors, and $Z_i = Z(X_i)$ is a real m-vector for $i = 1, \ldots, n$.

S.2. The conditional density of the response variable $f_{Y|Z}(y|z)$ is bounded above by $\overline{f}$ and its derivative in y is bounded above by $\overline{f'}$, uniformly in the arguments y and z and in n; moreover, $f_{Y|Z}(z'\beta(u) + R(x,u)|z)$ is bounded away from zero uniformly for all arguments $u \in \mathcal{U}$, z, and n.

S.3. For every m, the smallest eigenvalues of the design matrix $Q_m = E[ZZ']$ and of the Jacobian matrix $J_m(u) = E[f_{Y|Z}(Z'\beta(u) + R(X,u)|Z)ZZ']$ are bounded away from zero uniformly in n.

S.4. The following bound applies: $\max_{i \leq n} \|Z_i\| \leq \zeta(m,d,n) = \zeta_m$.

S.5. The approximation error term $R(X,U)$ is such that $\sup_{x,u\in\mathcal{U}} |R(x,u)| \lesssim m^{-\kappa}$.

Assumptions S.1-S.3 are standard regularity conditions in conditional quantile models. Note that the identification conditions on $f_{Y|Z}$ are at the true conditional quantile which includes the approximation error. Assumption S.4 imposes a uniform bound on the norm of the transformed regressor vector which can grow with the sample size. In many typical basis we have $\zeta_m = \sqrt{m}$. Finally, assumption S.5 introduces a standard bound on the approximation error, see Newey [6] for a discussion.

3 Main Results

The first main result is a uniform rate of convergence. Uniform convergence over $\mathcal{U}$ is achieved at the same rate as for a single quantile.

Theorem 1 (Uniform Convergence Rate for Series QR Coefficients) *Under Condition S, and provided that $\zeta_m^2 m \log n = o(n)$,*

$$\sup_{u\in\mathcal{U}} \left\| \widehat{\beta}(u) - \beta(u) \right\| \lesssim_P \sqrt{\frac{m}{n}}.$$

The second main result is the approximation of the quantile regression process by a conditionally pivotal quantity.

Theorem 2 (Strong Approximations to the QR Process by a Pivotal Coupling) *Under Condition S, $m^3\zeta_m^2 \log^7 n = o(n)$, and $m^{-\kappa+1}\log^3 n = o(1)$, the QR process is uniformly close to a conditionally pivotal process, namely*

$$\sqrt{n}\left(\widehat{\beta}(u) - \beta(u)\right) = J_m^{-1}(u)\mathbb{U}_n(u) + r_n(u),$$

where

$$\mathbb{U}_n(u) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(u - 1\{U_i \leq u\}), \tag{3.3}$$

where $U_1, \ldots, U_n$ are i.i.d. Uniform$(0,1)$, independently distributed of $Z_1, \ldots, Z_n$, and

$$\sup_{u\in\mathcal{U}} \|r_n(u)\| \lesssim_P \frac{m^{3/4}\zeta_m^{1/2}\log^{3/4} n}{n^{1/4}} + \sqrt{m^{1-\kappa}\log n} = o(1/\log n).$$

Under the appropriate growth conditions stated above, the preceding theorem has established that

$$\sqrt{n}(\widehat{\beta}(u) - \beta(u)) = J_m^{-1}(u)\mathbb{U}_n(u) + o_P(1),$$

where the matrix-valued functions $u \mapsto J_m(u)$ are uniformly nonsingular, and the process $\mathbb{U}_n(u)$ is as defined in (3.3). The approximation by a pivotal quantity is very important since it enables us to perform inference without resorting to normal approximations. Indeed, we can obtain the distribution of $\mathbb{U}_n(u)$ readily via Monte-Carlo simulation and can consistently estimate the matrix $J_m(u)$ (see Theorem 4). Then we can carry out inference based on the empirical quantile regression process $u \mapsto \widehat{\beta}(u)$. We refer to this approach as the pivotal method.

The pivotal method is closely related to another approach to inference, which we refer to here as the gradient bootstrap method. This approach was previously introduced by Parzen, Wei and Ying [7] for parametric models with fixed dimension. We extend it to a considerably more general nonparametric series framework. The main idea is to generate draws $\widehat{\beta}^*(\cdot)$ of the QR process as solutions to QR problems with gradients perturbed by a pivotal quantity $\mathbb{U}^*_n(\cdot)/\sqrt{n}$. In particular, let us define a gradient bootstrap draw $\widehat{\beta}^*(u)$ as the solution to the problem

$$\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)] - \mathbb{U}^*_n(u)'\beta/\sqrt{n}, \tag{3.4}$$

for each $u \in \mathcal{U}$, where $\mathbb{U}^*_n(\cdot)$ is defined in (3.3). The problem is solved many times for independent draws of $\mathbb{U}^*_n(\cdot)$, and the distribution of $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ is approximated by the empirical distribution of the bootstrap draws of $\sqrt{n}(\widehat{\beta}^*(\cdot) - \widehat{\beta}(\cdot))$.

Next we turn to a strong approximation based on a sequence of Gaussian processes.

Theorem 3 (Strong Approximation to the QR Process by a Gaussian Coupling) *Under conditions S.1-S.4 and $m^7\zeta_m^6 \log^{22} n = o(n)$, there exists a sequence of zero-mean Gaussian processes $G_n(\cdot)$ with a.s. continuous paths, that has the same covariance functions as the pivotal process $\mathbb{U}_n(\cdot)$ in (3.3) conditional on $Z_1, \ldots, Z_n$, namely,*

$$E[G_n(u)G_n(u')'] = E[\mathbb{U}_n(u)\mathbb{U}_n(u')'] = \mathbb{E}_n[Z_iZ_i'](u \wedge u' - uu'), \text{ for all } u \text{ and } u' \in \mathcal{U}.$$

Also, $G_n(\cdot)$ approximates the empirical process $\mathbb{U}_n(\cdot)$, namely,

$$\sup_{u\in\mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \lesssim_P o(1/\log n).$$

Consequently, if in addition S.5 holds with $m^{-\kappa+1}\log^3 n = o(1)$,

$$\sup_{u\in\mathcal{U}} \|\sqrt{n}(\widehat{\beta}(u) - \beta(u)) - J_m^{-1}(u)G_n(u)\| \lesssim_P o(1/\log n).$$

Another related inference method is the weighted bootstrap for the entire QR process. Consider a set of weights $h_1, ..., h_n$ that are i.i.d. draws from the standard exponential distribution. For each draw of such weights, define the weighted bootstrap draw of the QR process as a solution to the QR problem weighted by $h_1, \ldots, h_n$:

$$\widehat{\beta}^b(u) \in \arg\min_{\beta\in\mathbb{R}^m} \mathbb{E}_n[h_i\rho_u(Y_i - Z_i'\beta)], \quad \text{for } u \in \mathcal{U}.$$

We show in [2] that the distribution of $\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u))$ is valid for approximating the distribution of the QR process.

In order to implement some of the inference methods, we need uniformly consistent estimators of the Gram and Jacobian matrices. The natural candidates are

$$\widehat{\Sigma}_m = \mathbb{E}_n[Z_iZ_i'], \tag{3.5}$$

$$\widehat{J}_m(u) = \frac{1}{2h_n}\mathbb{E}_n[1\{|Y_i - Z_i'\widehat{\beta}(u)| \le h_n\} \cdot Z_iZ_i'], \tag{3.6}$$

where h_n is a bandwidth parameter, such that $h_n \to 0$, and $u \in \mathcal{U}$. The following result establishes uniform consistency of these estimators and provides an appropriate rate for the bandwidth h_n which depends on the growth of the model.

Theorem 4 (Estimation of Gram and Jacobian Matrices) *If conditions S.1-S.4 and $\zeta_m^2 \log n = o(n)$ hold, then $\widehat{\Sigma}_m - \Sigma_m = o_P(1)$ in the eigenvalue norm . If conditions S.1-S.5, $h_n = o(1)$, and $m\zeta_m^2 \log n = o(nh_n)$ hold, then $\widehat{J}_m(u) - J_m(u) = o_P(1)$ in the eigenvalue norm uniformly in $u \in \mathcal{U}$.*

4 Numerical Example

To evaluate the performance of our estimation and inference methods in finite samples, we conduct a Monte Carlo experiment designed to mimic the gasoline demand empirical example in [2]. We consider the following design for the data generating process:

$$Y = g(X) + \sigma\Phi^{-1}(U), \tag{4.7}$$

where $g(x) = \alpha_0 + \alpha_1 x + \alpha_2 \sin(2\pi x) + \alpha_3 \cos(2\pi x) + \alpha_4 \sin(4\pi x) + \alpha_5 \cos(4\pi x)$, $U \sim U(0,1)$, and Φ^{-1} denotes the inverse of the CDF of the standard normal distribution. The parameters of $g(x)$ and σ are calibrated by applying least squares to the gasoline data set where Y is the logarithm of household gasoline consumption and X is the logarithm of gasoline price.

To analyze the properties of the inference methods in finite samples, we draw 500 samples from the DGP (4.7) with 3 sample sizes, n: $5,001$, $1,000$, and 500 observations. For $n = 5,001$ we fix X to the values in the gasoline data set, whereas for the smaller sample sizes we draw X with replacement from the values in the data set and keep it fixed across samples. We focus on the average quantile elasticity function

$$u \mapsto \theta(u) = \int \partial_x g(x) d\mu(x),$$

over the region $I = [0.1, 0.9]$. We estimate this function using linear, power and B-spline quantile regression with the same number of terms and other tuning parameters as in [2]. Although $\theta(u)$ does not change with u in our design, we do not impose this restriction on the estimators. For inference, we compare the performance of 90% confidence bands for the entire elasticity function. These bands are constructed using the pivotal, Gaussian and weighted bootstrap methods. The interval I is approximated by a finite grid of 91 quantiles $\tilde{I} = \{0.10, 0.11, ..., 0.90\}$.

Table 1 reports estimation and inference results averaged across 200 simulations. The true value of the elasticity function is $\theta(u) = -0.74$ for all $u \in \tilde{I}$. Bias and RMSE are the absolute bias and root mean squared error integrated over $\tilde{I}$. SE/SD reports the ratios of empirical average standard errors obtained by delta method to empirical standard deviations. Length gives the empirical average of the length of the confidence band. SE/SD and length are integrated over the grid of quantiles $\tilde{I}$. Cover reports empirical coverage of the confidence bands with nominal level of 90%. Stat is the empirical average of the 90% quantile of the maximal t-statistic used to construct the bands. Table 1 shows that the linear estimator has higher absolute bias than the more flexible power and B-spline estimators, but displays lower rmse, especially for small sample sizes. The analytical standard errors provide good approximations to the standard deviations of the estimators. The confidence bands have empirical coverage close to the nominal level of 90% for all the estimators and sample sizes considered; and weighted bootstrap bands tend to have larger average length than the pivotal and Gaussian bands. In results not reported, we find that gradient bootstrap with 199 repetitions produces similar results to weighted bootstrap.

All in all, these results strongly confirm the practical value of the theoretical results and methods developed in the paper. They also support the empirical example by verifying that our estimation and inference methods work quite nicely in a very similar setting.

Table 1: Finite Sample Properties of Estimation and Inference Methods for Average Quantile Elasticity Function

				Pivotal			Gaussian			Weighted Bootstrap		
						$n = 5,001$						
	Bias	RMSE	SE/SD	Cover	Length	Stat	Cover	Length	Stat	Cover	Length	Stat
Linear	0.04	0.14	1.03	90	0.77	2.64	90	0.77	2.64	91	0.83	2.87
Power	0.00	0.15	1.01	91	0.85	2.65	91	0.85	2.65	93	0.91	2.83
B-spline	0.01	0.15	1.02	91	0.86	2.65	90	0.86	2.65	92	0.93	2.83
						$n = 1,000$						
Linear	0.03	0.30	1.06	89	1.78	2.64	90	1.78	2.64	91	1.96	2.99
Power	0.01	0.33	1.07	92	2.00	2.66	91	2.00	2.65	96	2.23	2.95
B-spline	0.01	0.35	1.05	90	2.07	2.65	90	2.07	2.65	95	2.31	2.95
						$n = 500$						
Linear	0.02	0.44	1.05	90	2.61	2.64	91	2.61	2.64	93	2.85	3.05
Power	0.01	0.52	1.06	92	3.13	2.65	92	3.14	2.66	96	3.47	3.01
B-spline	0.01	0.54	1.06	91	3.26	2.65	91	3.26	2.65	96	3.59	3.00

Notes: 500 repetitions. Simulation standard error for coverage probability is 1%.

References

[1] O. Arias, K. F. Hallock, and Walter Sosa-Escudero. Individual heteroneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics*, 26(1):7–40, 2001.

[2] A. Belloni, V. Chernozhukov, and I. Fernandez-Val. Conditional Quantile Processes based on Series or Many Regressors. *ArXiv e-prints*, May 2011.

[3] M. Buchinsky. Changes in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62(2):405–458, Mar. 1994.

[4] R. Koenker. *Quantile regression.* Cambridge University Press, New York, 2005.

[5] R. Koenker and G. Basset. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[6] Whitney K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.

[7] M. I. Parzen, L. J. Wei, and Z. Ying. A resampling method based on pivotal estimating functions. *Biometrika*, 81(2):341–350, 1994.