Nonparametric Quantile Regression

Thomas Wiemann
University of Chicago

TA Discussion # 5
Econ 31740

February 14, 2022

Outline

- Quantile Regression under Misspecification
- Nonparametric Quantile Regression
  - Local Polynomial Quantile Regression
  - Penalized Quantile Regression
- Application to Abrevaya (2006)

Methods discussed today are implemented in Julia with code readily available on GitHub (link to `MyMethods.jl`). Code reproducing the figures can be downloaded here.

Recall the quantile regression (QR) population problem. We have

$$\beta(\tau) := \arg\min_{\beta \in \mathbb{R}^d} E\left[\rho_\tau(Y - X^\top \beta)\right], \quad (1)$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}\{u \leq 0\})$ is the check function for $\tau \in (0, 1)$.

Define the QR specification error as

$$\Delta_\tau(X, \eta) := X^\top - \mathcal{Q}_\tau(Y|X), \quad (2)$$

and let ϵ_τ be the quantile-specific residual

$$\epsilon_\tau := Y - \mathcal{Q}_\tau(Y|X), \quad (3)$$

with conditional density $f_{\epsilon_\tau|X}(\cdot)$.

Quantile Regression under Misspecification (Contd.)

Theorem 1 of Angrist et al. (2006):

Suppose that (i) the conditional density $f_{Y|X}(y)$ exists a.s., (ii) $E[Y]$, $E[\mathcal{Q}_\tau(Y|X),]$ and $E\|X\|$ are finite, and (iii) $\beta(\tau)$ uniquely solves the quantile regression population problem (1). Then

$$\beta(\tau) = \underset{\beta \in \mathbb{R}^d}{\arg\min}\, E\left[\omega_\tau(X,\beta)\Delta_\tau^2(X,\beta)\right] \tag{4}$$

where

$$\begin{aligned} \omega_\tau(X,\beta) &= \int_0^1 (1-u) f_{\epsilon_\tau|X}\left(u\Delta_\tau(X,\beta)\right) du \\ &= \int_0^1 (1-u) f_{Y|X}\left(uX^\top\beta + (1-u)\mathcal{Q}_\tau(Y|X)\right) du \geq 0. \end{aligned} \tag{5}$$

Quantile Regression under Misspecification (Contd.)

Theorem 1 of Angrist et al. (2006) shows that QR has an interpretation similar to the best linear predictor interpretation of of linear (mean) regression. In particular, it shows that QR is the linear predictor that minimizes a weighted sum of squared misspecification errors, where the weights are the average conditional density on the line from $X^\top\beta$ to $\mathcal{Q}_\tau(Y|X)$. Note that multiplication by $(1-u)$ implies higher weight on points on the line closer to $\mathcal{Q}_\tau(Y|X)$.

Angrist et al. (2006) refer to $\omega_\tau(X,\beta)$ as *importance weights*. In addition to the importance weights, the distribution of X also determines how the sum of squared misspecification errors in (4) is weighted:

$$\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \int \omega_\tau(x,\beta)\Delta_\tau^2(x,\beta)d\Pi(x), \tag{6}$$

where Π is the distribution function of X. Let π denote the corresponding density function, then the overall QR weights are given by

$$\omega_\tau(x,\beta)\pi(x). \tag{7}$$

Figure 1: Parametric and Nonparametric Quantile Regression Results

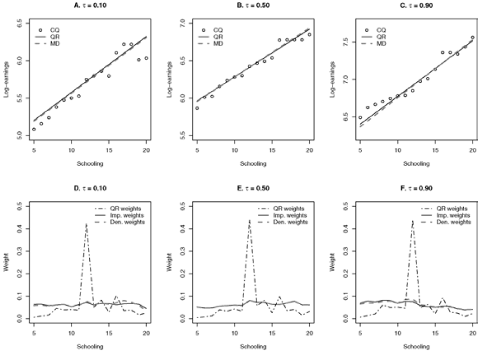


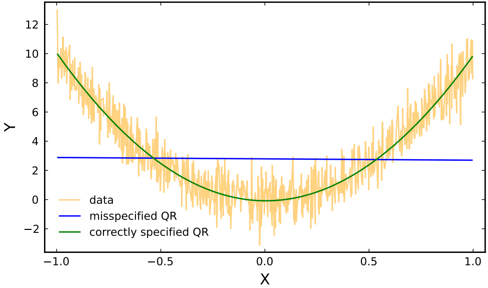
FIGURE 1.—Conditional quantile function and weighting schemes in the 1980 Census (for U.S.-born black and white men aged 40–49). Panels A–C plot the conditional quantile function, the linear quantile regression fit, and Chamberlain's minimum distance fit for log earnings given years of schooling. Panels D–F plot the QR weighting function (histogram × importance weights), the importance weights, and the density weights.

Notes. Figure 1 of Angrist et al. (2006). The minimum deviations estimator of Chamberlain (1994) is defined as the projection of X on $Q_\tau(Y|X)$, which must be estimated non-parametrically in a first step.

Quantile Regression under Misspecification (Contd.)

The results of Angrist et al. (2006) provide a motivation of QR analogous to the best-linear predictor motivation of OLS. Note, however, that there is no guarantee for any linear predictor to well-approximate potentially very non-linear functions, even if it's the best possible linear predictor.

Figure 2: Quantile Regression Estimates ($\tau = 0.5$)



Notes. Quantile regression estimates applied to $n = 1000$ iid samples from the DGP $Y = 10X^2 + \varepsilon$, where $X \sim \mathcal{U}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, 1)$. "Misspecified QR" is a median regression of Y on X. "Correctly specified QR" also includes the regressor X^2.

Nonparametric Quantile Regression

The rest of today's discussion will review basic topics in *nonparametric* quantile regression. The object of interest in these settings is the conditional quantile function $\mathcal{Q}_\tau(Y|X)$, not its best linear predictor (w.r.t. the weighted squared misspecification loss) as is the case in QR.

Nonparametric quantile regression is most simple when X is univariate. We will focus on

- Local Polynomial Quantile Regression, and
- Total Variation Penalized Quantile Regression.

The discussion is primarily based on Chapter 7.1 and 7.2 of Koenker (2005) and references cited therein.

Local Polynomial Quantile Regression

Consider the setting with a continuously distributed scalar outcome Y and a univariate variable X. The object of interest is the conditional quantile function of Y given $X = x$:

$$q_\tau(x) := \mathcal{Q}_\tau(Y|X = x). \tag{8}$$

Suppose we observe a sample $\{(y_i, x_i)\}_{i=1}^n$ with $(y_i, x_i) \overset{iid}{\sim} (Y, X)$. The *local polynomial quantile regression* estimator is defined as

$$\hat{q}_\tau^{lqr}(x) = \hat{\beta}_{0,\tau}(x) \tag{9}$$

where $\hat{\beta}_{0,\tau}(x)$ is the first element of the coefficient vector

$$\hat{\beta}_\tau(x) = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^n w_i(x) \rho_\tau \left(y_i - \sum_{j=0}^p \beta_{j,\tau}(x_i - x)^j \right), \tag{10}$$

with weights $w_i(x) = K\left((x_i - x)/h\right)$ from a positive, symmetric, unimodal kernel function K.

Local Polynomial Quantile Regression

Similar to derivative estimation using local polynomial (mean) regression, derivatives of the conditional quantile function can be estimated by other components of $\hat{\beta}_\tau(x)$. For example, the second component $\hat{\beta}_{1,\tau}(x)$ is an estimate of $\frac{\partial \mathcal{Q}_\tau(Y|X)}{\partial X}\Big|_{X=x}$.

Chaudhuri (1991) analysis asymptotic behavior of a multivariate version of (10) where K is a a uniform kernel with bandwidth decreasing in sample size. In the setting of an additively separable model $Y = g(X) + U$ with exogeneous regressors, he derives rates of convergence of function and derivative estimates. In the univariate case with $p = 2$, the local polynomial quantile regression estimate converges to the conditional quantile function at (essentially) the rate $\mathcal{O}(n^{-\frac{2}{5}})$ and the first derivative converges at $\mathcal{O}(n^{-\frac{1}{5}})$.

Local Polynomial Quantile Regression

Implementation of local polynomial quantile regression is straightforward using its linear programming formulation. In slight adaptation of the linear program for quantile regression of TA discussion # 3, we have

$$\begin{aligned} \hat{\beta}_\tau(x) \;=\; & \arg\min_{\beta_\tau} \sum_{i=1}^{n} w_i(x)\left(\tau u_i + (1-\tau) v_i\right) \\ & \text{s.t.} \quad u_i - v_i = y_i - \sum_{j=0}^{p} \beta_{j,\tau}(x_i - x)^j, \forall i, \\ & \beta_\tau \in \mathbb{R}^{p+1}, u_i, v_i \geq 0, \forall i. \end{aligned} \tag{11}$$

Notice this straightforwardly allows for a partially linear specification as well by setting the first constraint to

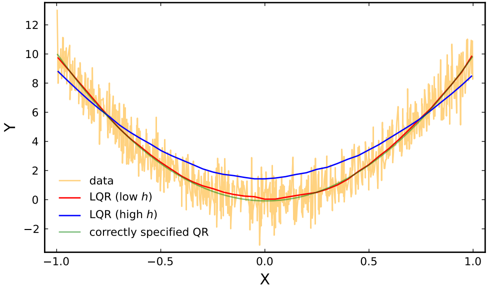
$$u_i - v_i = y_i - \sum_{j=0}^{p} \beta_{j,\tau}(x_i - x)^j - z_i^\top \gamma, \forall i, \tag{12}$$

for some additional variables z_i and decision variables $\gamma \in \mathbb{R}^{\dim Z}$.

Local Polynomial Quantile Regression

I implement local polynomial quantile regression using the linear program formulation (which may very well not be computationally efficient) in Julia as part of the `MyMethods.jl` module (link).

Figure 3: Local Polynomial Quantile Regression Estimates ($\tau = 0.5$)



Notes. Nonparametric quantile regression estimates applied to $n = 1000$ iid samples from the DGP $Y = 10X^2 + \varepsilon$, where $X \sim \mathcal{U}(0, 1)$ and $\varepsilon \sim \mathcal{N}(0, 1)$. "Correctly specified QR" is a median regression of Y on X and X^2. "LQR (low h)" and "LQR (high h)" correspond to local linear quantile regression estimates using the Epanechnikov kernel with bandwidth $h = 0.2$ and $h = 1$, respectively.

Penalized Quantile Regression

Smoothing splines provide an alternative approach to nonparametric quantile regression regression. Considering the same setting with continuously distributed scalar outcome Y, univariate X, and iid data $\{(y_i, x_i)\}_{i=1}^n$, a penalized conditional quantile function estimate is given by the argmin of

$$\min_{q \in C^2} \sum_{i=1}^{n} \rho_\tau \left(y_i - q(x_i)\right) + \lambda \int \left(q''(x)\right)^2 dx, \tag{13}$$

where $\lambda \geq 0$ is a penalty parameter that controls the smoothness of the function q (the smoothest function being a linear function).

Notice that the penalized problem in (13) contains a squared term so that optimization will require quadratic programming. Bosch et al. (1995) provide a corresponding interior point algorithm.

Penalized Quantile Regression

Koenker et al. (1994) consider variations on the roughness-penalization, including penalization using the L_1 norm. In this setting, the problem has a particularly useful formulation in terms of the total variation of q'.

The total variation of an absolutely continuous function $f : [a, b] \to \mathbb{R}$ may be written as

$$V(f) = \sup \sum_{i=1}^{n} |f(x_{i+1}) - f(x_i)| = \int_a^b |f'(x)| \, dx, \tag{14}$$

where the supremum is taken over all partitions $a \leq x_1 < \ldots < x_n \leq b$, and the second equality holds for absolutely continuous functions.

Penalized Quantile Regression

Hence, for functions q with absolutely continuous first derivative we have

$$V(g') = \int_a^b |g''(x)|\, dx. \tag{15}$$

A variant of (13) considered by Koenker et al. (1994) is then

$$\min_q \sum_{i=1}^n \rho_\tau \left(y_i - q(x_i)\right) + \lambda V(q'). \tag{16}$$

Penalized Quantile Regression

To see why this formulation is useful, consider first the problem of finding a continuous function q that minimizes $V(q')$ while interpolating $\{(y_i, x_i)\}_{i=1}^n$. Suppose the data is sorted so that $a \leq x_1 < \ldots < x_n < b$. By the mean value theorem,

$$\exists u_i \in (x_i, x_{i+1}) : \quad g'(u_i) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad \forall i. \tag{17}$$

Then for any function $q : [a, b] \to \mathbb{R}$ with absolutely continuous first derivative, it holds that

$$V(g') \geq \sum_{i=1}^{n-1} \left| \int_{u_i}^{u_{i+1}} g''(x) \right| dx \geq \sum_{i=1}^{n-1} |g'(u_{i+1}) - g'(u_i)| =: V(\hat{g}'), \tag{18}$$

where $\hat{g}$ is the piecewise linear interpolant $\hat{g}(x_i) = y_i$ with knots at x_i.

Thus $\hat{g}$ is the optimal interpolant w.r.t. the total variation of g'.

Penalized Quantile Regression

It is now easy to see that the solution to (16) is also piecewise linear with knots at the x_i.

Proof by contradiction: Suppose $\tilde{q}$ is the solution to (16) but is not piecewise linear. Construct a linear interpolant $\hat{q}$ with knots at the x_i such that $\tilde{q}(x_i) = \hat{q}(x_i)$. Then by (18), $\hat{q}$ has a smaller total variation penalty but the same function values for all i. Hence $\tilde{q}$ is not the solution to (16).

Penalized Quantile Regression

Given that the solution to the penalized quantile regression problem (16) takes a piecewise linear form with knots at the x_i, construction of a corresponding linear program follows quickly. Let

$$\hat{q}(x) = \alpha_i + \beta_i(x - x_i), \quad x \in [x_i, x_{i+1}], \, \forall i. \tag{19}$$

Then we have

$$\begin{aligned} V(\hat{q}') &= \sum_{i=1}^{n-1} |g'(x_{i+1}) - g'(x_i)| \\ &= \sum_{i=1}^{n-1} |\beta_{i+1} - \beta_i| \\ &= \sum_{i=1}^{n-2} \left| \frac{\alpha_{i+2} - \alpha_{i+1}}{x_{i+2} - x_{i+1}} - \frac{\alpha_{i+1} - \alpha_i}{x_{i+1} - x_i} \right|, \end{aligned} \tag{20}$$

where the last equation sets $\beta_n = \beta_{n-1}$.

Penalized Quantile Regression

The linear program for penalized quantile regression is given by

$$
\begin{aligned}
\min_{\beta_\tau} \; & \sum_{i=1}^{n} (\tau u_i + (1-\tau) v_i) + \lambda \sum_{i=1}^{n-2} \Delta_i \\
\text{s.t.} \; & u_i - v_i = y_i - \alpha_i, \; \forall i = 1, \dots, n, \\
& \frac{\alpha_{i+2} - \alpha_{i+1}}{x_{i+2} - x_{i+1}} - \frac{\alpha_{i+1} - \alpha_i}{x_{i+1} - x_i} \leq \Delta_i, \; \forall i = 1, \dots, n-2 \\
& \frac{\alpha_{i+2} - \alpha_{i+1}}{x_{i+2} - x_{i+1}} - \frac{\alpha_{i+1} - \alpha_i}{x_{i+1} - x_i} \leq -\Delta_i, \; \forall i = 1, \dots, n-2 \\
& \alpha_i \in \mathbb{R}, \Delta_i, u_i, v_i \geq 0, \forall i.
\end{aligned}
\tag{21}
$$

We may amend the first constraint in similar fashion as we did for local polynomial quantile regression to allow for a partially linear specification with additional variables (see equation (12)).

Penalized Quantile Regression

Note that the linear programming formulation in (21) runs into difficulties when the x_i are non-unique. A simple practical remedy is to introduce a new set of constraints that sets the α_i equal for all observations with the same value x_i.

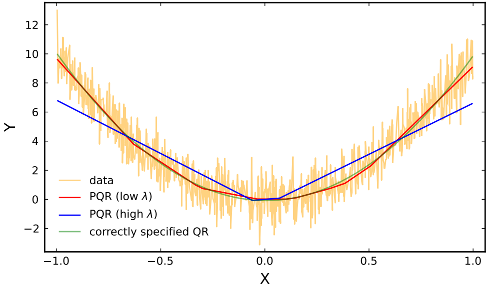
In particular, when x_i are non-unique, introduce the additional constraints

$$\forall i, \quad \alpha_i = \alpha_j, \, \forall j : x_j = x_i. \tag{22}$$

Depending on the specific solver you use in practice, you may still need to make numerical adjustments to the lags $x_{i+1} - x_i$ to avoid division by zero. After introducing constraints (22), you can set zero-valued differences to some finite number. This won't affect the results (up to numerical accuracy).

As before, I implement penalized quantile regression using the linear program formulation in Julia (link).

Figure 4: Penalized Quantile Regression Estimates ($\tau = 0.5$)



Notes. Nonparametric quantile regression estimates applied to $n = 1000$ iid samples from the DGP $Y = 10X^2 + \varepsilon$, where $X \sim \mathcal{U}(0,1)$ and $\varepsilon \sim \mathcal{N}(0,1)$. "Correctly specified QR" is a median regression of Y on X and X^2. "PQR (low λ)" and "PQR (high λ)" correspond to penalized quantile regression estimates with penalty parameter $\lambda = 1$ and $\lambda = 30$, respectively.

Application to Abrevaya (2006)

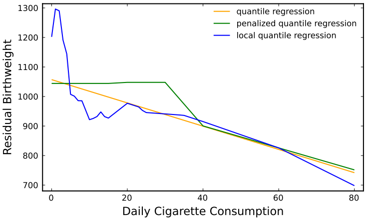
To illustrate the implemented procedures, it's worth revisiting the analysis of birthweight of Abrevaya (2006) that you've seen in Problem set 3. I consider the same data of 9,800 observations and estimate quantile regression of birthweight on mother's daily cigarettes consumption, mother's age and squared-age, weightgain and squared-weightgain of the mother during pregnancy, as well as indicators for male child, and the mother's martial status, race, whether she completed high-school, some college, and completed college, as well as whether she received no prenatal care, or in second trimester, or in third trimester.

The key variables of interest here are the mother's (a) daily cigarette consumption, and (b) age. All other variables are included via partially linear specifications (see, e.g., equation (12)).

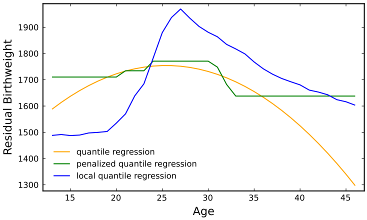
Results

For local linear qunatile regression, I use a Gaussian kernel with adaptively chosen bandwidth. In particular, for all unique values of daily cigarette consumption (or age) j, I set h_j equal to the minimum bandwidth such that the sum of all kernel weights is at least 2000. For the penalized quantile regression, I set $\lambda = 8$ (for no particular reason).

Figure 5: Parametric and Nonparametric Quantile Regression Results



(a) Daily Cigarette Consumption



(b) Age

Notes. Code can be downloaded here. I ran into some perfect collinearity issues with including all controls so don't take these results *too* seriously.

Last Comments

Today's discussion covered some basic approaches to univariate nonparametric quantile regression. Needless to say: there's lots left to explore!

A selection of related topics includes:

- Multivariate quantile regression (Chapter 7.3 of Koenker, 2005, is a good starting point, but there are many "modern" approaches, e.g., Ye and Padilla, 2021; Li et al., 2021)
- Approximately sparse quantile regression (see, e.g., Belloni et al., 2017)

References

Abrevaya, J. (2006). Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics*, 21(4):489–519.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563.

Belloni, A., Chernozhukov, V., and Kato, K. (2017). Thigh-dimensional quantile regression. In Koenker, R., Chernozhukov, V., He, X., and Peng, L., editors, *Handbook of Quantile Regression*, chapter 15. CRC Press.

Bosch, R. J., Ye, Y., and Woodworth, G. G. (1995). A convergent algorithm for quantile regression with smoothing splines. *Computational statistics & data analysis*, 19(6):613–630.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In *Advances in econometrics: sixth world congress*, volume 2, pages 171–209.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local bahadur representation. *The Annals of Statistics*, 19(2):760–777.

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

Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, 81(4):673–680.

Li, D., Li, Q., and Li, Z. (2021). Nonparametric quantile regression estimation with mixed discrete and continuous data. *Journal of Business & Economic Statistics*, 39(3):741–756.

Ye, S. S. and Padilla, O. H. M. (2021). Non-parametric quantile regression via the k-nn fused lasso. *Journal of Machine Learning Research*, 22(111):1–38.