14.385 Recitation 10

Paul Schrimpf

November 21, 2008

1 Local Linear Quantile Regression

All the nonparametric methods that we have talked about for mean regression also work for quantile regression. The steps for showing consistency, calculating MSE, and finding the asymptotic distribution for nonparametric quantile regression are similar to the steps for showing the same things for mean regression. Here, we will analyze local linear quantile regression.

1.1 Setup

Let $q(x) = Q_Y(\tau|x)$ be the τth quantile of y given x. Define the local linear quantile regression estimator as:

$$\hat{q}(x) \arg\min_{\alpha} \min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(y_i - \alpha - (x_i - x)\beta) K\left(\frac{x_i - x}{h}\right) \tag{1}$$

The first order condition for this problem is

$$0 = \frac{1}{n} \sum_{i=1}^{n} \left(\tau - \mathbf{1}(y_i - \alpha - (x_i - x)\beta)\right) K\left(\frac{x_i - x}{h}\right) \tag{2}$$

Under appropriate conditions (the kernel has certain properties and a LLN applies)

$$\frac{1}{n} \sum_{i=1}^{n} \left(\tau - \mathbf{1}(y_i - \alpha - (x_i - x)\beta)\right) K\left(\frac{x_i - x}{h}\right) = E\left[\mathbf{1}(y_i - \alpha - (x_i - x)\beta) K\left(\frac{x_i - x}{h}\right)\right] + O(1/(nh)) \tag{3}$$

1.2 Consistency

Let's verify Fischer consistency by showing that this expection is zero at $\alpha = q(x)$. This calculation will also give us the bias as a function of h. As in previous kernel situations, the key will be write the expectation as an integral, make a change of variables, and then take a Taylor expansion.

$$
\begin{aligned}
E\left[\mathbf{1}(y_i - \alpha - (x_i - x)\beta)K\left(\frac{x_i - x}{h}\right)\right] =& \int_X F_{y|x}\left(\alpha + (x_i - x)\beta\right) K\left(\frac{x_i - x}{h}\right) f_x(x_i)dx_i && (4)\\
& \text{change of variables } x_i = x + hu_i \\
=& h\int_U F_{y|x}\left(\alpha + hu_i\beta\right) K(u_i) f_x(x + hu_i)du_i && (5)\\
& \text{let } y_i = q(x_i) + \epsilon_i \\
=& h\int_U F_\epsilon\left(\alpha + hu_i\beta - q(x + hu_i)\right) K(u_i) f_x(x + hu_i)du_i && (6)\\
& \text{expand around } hu = 0 && (7)\\
=& h\int_U [\tau - F_\epsilon(\alpha - q(x)) - hu_i f_\epsilon(\alpha - q(x))(\beta - q'(x)) - \\
& - h^2\frac{1}{2}u_i\left(f_\epsilon'(\alpha - q(x))(\beta - q'(x))q'(x)' + \nabla^2 q(x) f_\epsilon(\alpha - q(x))\right)u_i' + o(h^2 u_i u_i')] \\
& \times\left[f_x(x) + hu_i f_x'(x) + \frac{1}{2}h^2 u_i f_x"(x)u_i' + o(h^2 u_i u_i')\right]K(u)du \\
& && (8)\\
& \text{kernel has mean } 0 && (9)\\
=& h(\tau - F_\epsilon(\alpha - q(x)) f_x(x)) + \\
& + h^3\int u'uK(u)du[(\tau - F_\epsilon(\alpha - q(x)))\frac{1}{2}tr(f_x"(x)) - \\
& - f_x(x)\frac{1}{2}tr\left(f_\epsilon'(\alpha - q(x))(\beta - q'(x))q'(x)' + \nabla^2 q(x) f_\epsilon(\alpha - q(x))\right) + (\beta - q'(x \\
& && (10)
\end{aligned}
$$

If we evaluate this at $\alpha_0 = q(x)$ and $\beta_0 = q'(x)$, we have:

$$
h^3\frac{1}{2}\left(\int u'uK(u)du\right) f_x(x)tr(\nabla^2 q(x)) f_\epsilon(0) \to 0 \text{ as } h \to 0 \tag{11}
$$

Importantly, this implies both $\hat{\beta}$ and $\hat{\alpha}$ are consistent.

1.3 Bias

To find the bias of $\hat{\alpha}$ we simply expand the first order condition around α_0. Since the first order condition is not differentiable, we expand the approximation to the FOC. Let $\widehat{FOC}$ denote the true first order condition, and FOC denote the asymptotic approximation. Then

$$
\begin{aligned}
0 =& \widehat{FOC}(\hat{\alpha}, \hat{\beta}) \\
=& FOC(\hat{\alpha}, \hat{\beta}) + O_p(1/(nh)) \\
O_p(1/(nh)) =& FOC(\alpha_0, \beta_0) + (\hat{\alpha} - \alpha_0)\frac{\partial FOC}{\partial \alpha} + (\hat{\beta} - \beta_0)\frac{\partial FOC}{\partial \beta} && (12)
\end{aligned}
$$

From the above we know that $FOC(\alpha_0, \beta_0) = h^3\frac{1}{2}\left(\int u'uK(u)du\right) f_x(x)tr(\nabla^2 q(x)) f_\epsilon(0) + o(h^3)$, and $\hat{\beta} - \beta_0 = O_p(1/(nh))$. Also, by inspecting, (reffocBias), $\frac{\partial FOC}{\partial \beta} = o(h^3)$. From the same equation, we have:

$$
\frac{\partial FOC}{\partial \alpha} = h\left(f_\epsilon(0) f_x(x) + h^2 \text{stuff}\right) \tag{13}
$$

Thus,

$$
\begin{aligned}
(\hat{\alpha} - \alpha_0) =& \frac{h^2\frac{1}{2}\left(\int u'uK(u)du\right) f_x(x)tr(\nabla^2 q(x)) f_\epsilon(0)}{f_\epsilon(0) f_x(x) + o(h^2)} + O_p(1/(nh))\frac{o(h^3)}{o(h^3)} + O_p(1/(nh)) && (14)\\
=& h^2\frac{1}{2}tr(\nabla^2 q(x))\left(\int u'uK(u)du\right) + O_p(1/(nh)) && (15)
\end{aligned}
$$

1.4 Variance

To find the asymptotic variance, we will need to assume that

$$\widehat{FOC}(\hat{\alpha},\hat{\beta}) - \widehat{FOC}(\alpha_0,\beta_0) = FOC(\hat{\alpha},\hat{\beta}) - FOC(\alpha_0,\beta_0) + o_p(1/\sqrt{nh^d}) \tag{16}$$

This would follow from stochastic equicontinuity of the first order condition. The theory section of problem set 3 showed that it holds, but the argument is long, so I won't repeat it here. With this assumption, we have:

$$\begin{aligned}
0 =& \widehat{FOC}(\hat{\alpha},\hat{\beta}) && (17)\\
=& \widehat{FOC}(\alpha_0,\beta_0) + FOC(\hat{\alpha},\hat{\beta}) - FOC(\alpha_0,\beta_0) o_p(1/\sqrt{nh^d}) && (18)\\
=& \widehat{FOC}(\alpha_0,\beta_0) + (\hat{\alpha}-\alpha_0)\frac{\partial FOC}{\partial \alpha} + (\hat{\beta}-\beta_0)\frac{\partial FOC}{\partial \beta} + o(h^3) + o_p(1/\sqrt{nh^d}) && (19)\\
=& \widehat{FOC}(\alpha_0,\beta_0) + (\hat{\alpha}-\alpha_0)(hf_x(x)f_\epsilon(0) + o(h^3) + O_p(1/(nh))o(h^3) + o(h^3) + o_p(1/\sqrt{nh^d}) && (20)\\
\sqrt{nh^d}(\hat{\alpha}-\alpha_0) =& \frac{\sqrt{nh^d}\widehat{FOC}(\alpha_0,\beta_0)}{hf_x(x)f_\epsilon(0)} + \sqrt{nh^d}o(h)^3 + o_p(1) && (21)
\end{aligned}$$

Now, we just need to find the distribution of $\sqrt{nh^d}\widehat{FOC}(\alpha_0,\beta_0)$. That is,

$$\sqrt{nh^d}\widehat{FOC}(\alpha_0,\beta_0) = \sqrt{\frac{h^d}{n}}\sum_{i=1}^{n}\left(\tau - \mathbf{1}(\epsilon_i - (x_i - x)q'(x) < 0)\right) K\left(\frac{x_i - x}{h}\right) \tag{22}$$

Assume a CLT applies so that this is normal with mean given by its exectation (which we computed above) and variance equal to

$$\begin{aligned}
var\left[\left(\tau - \mathbf{1}(\epsilon_i - (x_i - x)q'(x) < 0)\right) K\left(\frac{x_i - x}{h}\right)\right] =& var(f(x_i,\epsilon_i)) \\
=& E[var(f(x_i,\epsilon_i)|x_i)] + var[E[f_x,\epsilon_i]|x_i] && (23)\\
& \text{steps like those taken to show consistency} \\
=& h^2\tau(1-\tau)f_x(x)\int K(u)^2 du + 0 && (24)
\end{aligned}$$

Thus, we have:

$$\sqrt{nh^d}(\hat{\alpha}-\alpha_0) \xrightarrow{d} N\left(h^2\frac{1}{2}tr(\nabla^2 q(x))\left(\int u'uK(u)du\right), \frac{\tau(1-\tau)\int K(u)^2}{f_x(x)f_\epsilon(0)^2}\right) \tag{25}$$

and the MSE is

$$MSE(\hat{\alpha}-\alpha_0) = h^4\frac{1}{4}tr(\nabla^2 q(x))^2\left(\int u'uK(u)du\right)^2 + \frac{\tau(1-\tau)\int K(u)^2}{nh^d f_x(x)f_\epsilon(0)^2} \tag{26}$$

With $d = 1$, The MSE minimizing bandwidth will then be $h = O(n^{-1/5})$, which makes the bias squared and variance disappear at the same rate.

2 Semi-Parametric Estimators

2.1 Transformation Model

A transformation model takes the form:

$$\Lambda(y) = x\beta + u \tag{27}$$

where the transformation, Λ, and the distribution of u are unknown. Recall from problem set 3 and recitation 6 that the transformation model is generalization of the mixed proportional hazard model and the accelerated failure time model. In particular, the mixed proportional hazard model can be written as

$$\log z(t) = x\beta - \log(\alpha) + u \tag{28}$$

where $z(t)$ is the integrated hazard function, α is frailty, and u has a Gumbel distribution.

2.1.1 Estimation of β

There are multiple $\sqrt{n}$ consistent estimators of β. Han (1987) describes a maximum rank score estimator of β. β can also be recovered from an estimator of average derivative of $E[y|x]$. We discussed average derivative estimation in class, so let's take that approach. We can write the transformation model as:

$$y = \Lambda^{-1}(x\beta + u) \tag{29}$$

so

$$E[y|x] = E[\Lambda^{-1}(x\beta + u)|x] = g(x) \tag{30}$$

The derivative is

$$\frac{\partial g}{\partial x_k} = E[\Lambda^{-1\prime}(x\beta + u)|x]\beta_k \tag{31}$$

so we can recover β up to scale by looking at

$$\frac{E[w(x)(\partial g)/(\partial x_k)]}{E[w(x)(\partial g)/(\partial x_j)]} = \frac{\beta_k}{\beta_j} \tag{32}$$

where $w(x)$ is any function that depends on x. We could set $w(x) = 1$, and then use the estimator of $E[(\partial g)/(\partial x_k)]$ discussed in lecture. Alternatively, we can follow Powell, Stock, and Stoker (1989) and take $w(x) = f(x)$. Then, integration by parts shows that

$$E[f(x)(\partial g)/(\partial x_k)] = -2E[yf'(x)] \tag{33}$$

The derivative of the density of x can be estimated by kernel methods,

$$\hat{f'}_j(x) = \frac{1}{n}\sum_{i\neq j}\frac{1}{h}K'\left(\frac{x_i - x}{h}\right) \tag{34}$$

where we leave out observation j to simplify analysis of the weighted average derivative estimator, which is

$$E[f(x)\widehat{(\partial g)}/(\partial x_k)] = \frac{-2}{n}\sum y_i\hat{f'}_i(x_i) \tag{35}$$

Powell, Stock, and Stoker show that if f is $\frac{d}{2} + 3$ times differentiable, where d is the dimension of x, and $Nh^{d+2} \to \infty$, while $Nh^{d+4} \to 0$ then

$$\sqrt{n}(E[f(x)\widehat{(\partial g)}/(\partial x_k)] - E[f(x)(\partial g)/(\partial x_k)]) \xrightarrow{d} N(0, V) \tag{36}$$

This implies that our estimate of β is also asymptotically normal and converges at a $\sqrt{n}$ rate

$$\sqrt{n}(\widehat{\frac{\beta_k}{\beta_j}} - \frac{\beta_k}{\beta_j}) \xrightarrow{d} N(0, V_\beta) \tag{37}$$

2.1.2 Estimation of $\Lambda(\cdot)$

Horowitz (1996) developed an estimator of $\Lambda(\cdot)$ and $F_u(\cdot)$. Somewhat surprisingly, both of these functions can be estimated at a $\sqrt{n}$ rate. Consider the cdf of y conditional on $x\beta$:

$$G(y|x\beta) = F_u(\Lambda(y) - x\beta) \tag{38}$$

Differentiating $G(y|z)$ with respect to y and z gives:

$$G_y(y|z) = \Lambda'(y) f_u(\Lambda(y) - x\beta) \tag{39}$$
$$G_z(y|z) = f_u(\Lambda(y) - x\beta) \tag{40}$$

So,

$$\Lambda(y) = \Lambda(y_0) + \int_{y_0}^{y} \int_Z w(z) dz \Lambda'(y) dy \tag{41}$$
$$= \Lambda(y_0) + \int_{y_0}^{y} \int_Z w(z) \frac{G_y(y|z)}{G_z(y|z)} dz dy \tag{42}$$
$$\tag{43}$$

where $w(z)$ is any function that integrates to 1. G_y and G_z. An estimate of Λ can be constructed as

$$\hat{\Lambda}(y) = \int_{y_0}^{y} \int_Z w(z) \frac{\hat{G}_y(y|z)}{\hat{G}_z(y|z)} dz dy \tag{44}$$

Horowitz proposes estimating $\hat{G}_y$ by taking the derivative with respect to z of the kernel regression of $\mathbf{1}(y_i < y)$ on $z_i = x_i\beta$. Local linear or polynomial regression could probably also be used to estimate the derivative directly. $\hat{G}_y(y|z)$ can be estimated using kernel methods as well. Specifically,

$$\hat{G}_y(y|z) = \frac{\sum K_y((y_i - y)/h) K_z((z_i - z)/h)}{\sum K_z((z_i - z)/h)} \tag{45}$$

Although the estimates of both $\hat{G}_y$ and $\hat{G}_z$ converge slower than $\sqrt{n}$, the estimate of $\hat{\Lambda}$ averages over many values of z to achieve a $\sqrt{n}$ rate of convergence. See Horowitz's paper for details. Important conditions are that K_z be a higher-order kernel (ie $\int u^k K_z(u) du = 0$ for some $k > 1$) and that h_y and h_z go to zero faster than the usual optimal rate. Both of these extra conditions are to make the bias of $\hat{\Lambda}$ be $o_p(1/\sqrt{n})$.

2.2 Estimation of $F_u(\cdot)$

The obvious estimator of F_u is the empirical distribution function of $\hat{\Lambda}(y) - x\hat{\beta}$. However, this estimator will not work because $\hat{\Lambda}(y)$ is only consistently estimable in a compact interval, $[y_1, y_2]$. Nonetheless, a simply modification of the EDF works. This estimator is based on $F_u(u) P(y \in [y_1, y_2] | u) = F_u(u | y \in [y_1, y_2]) P(y \in [y_1, y_2])$, which rearranged suggests:

$$\hat{F}_u(u) = \frac{\sum \mathbf{1}(\hat{\Lambda}(y_i) - x_i\hat{\beta} < u) \mathbf{1}(\hat{\Lambda}(y_1) - u \le x_i\beta \le \hat{\Lambda}(y_2) - u)}{\sum \mathbf{1}(\hat{\Lambda}(y_1) - u \le x_i\beta \le \hat{\Lambda}(y_2) - u)} \tag{46}$$