



Quantile Regression

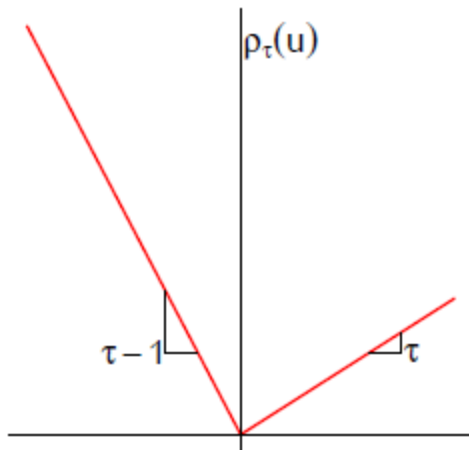
Ruibin Xi



The Check function

- We define a loss function

$$\begin{aligned}\rho_{\tau}(u) &= \begin{cases} \tau u & \text{if } u > 0 \\ (\tau - 1)u & \text{if } u \leq 0 \end{cases} \\ &= u(\tau - I(u < 0))\end{aligned}$$



Optimality Conditions (1)

- Consider the univariate case: the τ th quantile ξ minimizes

$$R(\xi) = \sum_{i=1}^n \rho_{\tau}(y_i - \xi)$$

- $R(\xi)$ is not differentiable, but has directional derivative

Optimality Conditions (2)

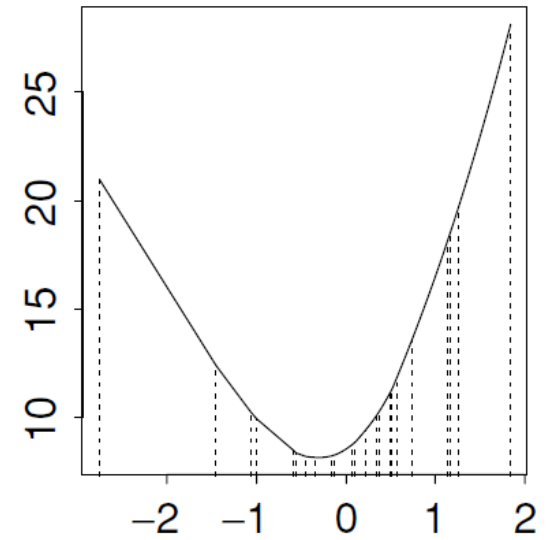
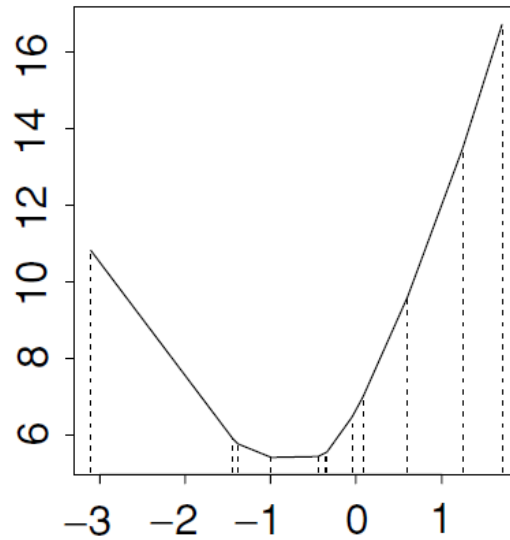
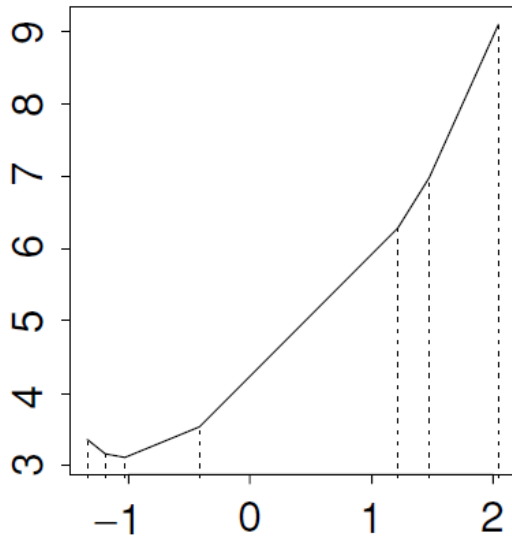
- Right derivative

$$\begin{aligned}R'(\xi+) &= \lim_{h \rightarrow 0_+} (R(\xi + h) - R(\xi))/h \\&= \sum_{i=1}^n \lim_{h \rightarrow 0_+} \frac{\rho_\tau(y - \xi - h) - \rho_\tau(y - \xi)}{h} \\&= \sum_{i=1}^n (I(y_i \leq \xi) - \tau)\end{aligned}$$

- Left derivative

$$\begin{aligned}R'(\xi-) &= \lim_{h \rightarrow 0_+} (R(\xi - h) - R(\xi))/h \\&= \sum_{i=1}^n \lim_{h \rightarrow 0_+} \frac{\rho_\tau(y - \xi + h) - \rho_\tau(y - \xi)}{h} \\&= \sum_{i=1}^n (\tau - I(y_i < \xi))\end{aligned}$$

Optimality Conditions (3)



Objective function for $\tau=1/3$; sample size = 7, 12, 23

Optimality Conditions (4)

- A point $\hat{\xi}$ minimizes the objective function if

$$R'(\hat{\xi}+) \geq 0 \quad \text{and} \quad R'(\hat{\xi}-) \geq 0$$

- Let $N = \sum_i I(y_i < \hat{\xi})$, $Z = \sum_i I(y_i = \hat{\xi})$ and $P = \sum_i I(y_i > \hat{\xi})$

- Then

$$N \leq n\tau \leq N + Z$$
$$P \leq n(1 - \tau) \leq P + Z$$

Optimality Conditions (5)

- Linear quantile regression minimizes

$$R(\beta) = \sum_{i=1}^n \rho_{\tau}(y_i - x_i^T \beta)$$

- This function is not differential at the points with zero residuals, but has directional derivative in all directions

Optimality Conditions (6)

- The directional derivative at a direction w is

$$\begin{aligned}\nabla R(\beta, w) &\equiv \frac{d}{dt} R(\beta + tw) \Big|_{t=0} \\ &= \frac{d}{dt} \sum_{i=1}^n (y_i - x_i^\top \beta - x_i^\top tw) [\tau - I(y_i - x_i^\top \beta - x_i^\top tw) < 0] \Big|_{t=0} \\ &= - \sum x_i^\top w \psi_\tau^*(y_i - x_i^\top \beta, -x_i^\top w),\end{aligned}$$

where

$$\psi_\tau^*(u, v) = \begin{cases} \tau - I(u < 0) & \text{if } u \neq 0 \\ \tau - I(v < 0) & \text{if } u = 0. \end{cases}$$

Optimality Conditions (7)

- If

$$\Delta R(\hat{\beta}, w) \geq 0 \text{ for all } w \in \mathcal{R}^p \text{ with } \|w\| = 1$$

then $\hat{\beta}$ minimizes $R(\beta)$

- The basic solution is $b(h) = X(h)^{-1}y(h)$
- To check $b(h)$ is a solution, we must consider

$$\nabla R(b(h), w) = - \sum_{i=1}^n \psi_{\tau}^*(y_i - x_i^{\top} b(h), -x_i^{\top} w) x_i^{\top} w$$

Optimality Conditions (8)

- Let $v = X(h)w$, the optimality is achieved iff

$$0 \leq - \sum_{i=1}^n \psi_{\tau}^*(y_i - x_i^{\top} b(h), -x_i^{\top} X(h)^{-1} v) x_i^{\top} X(h)^{-1} v$$

for all $v \in \mathbb{R}^p$.

- For $i \in h$, we have $e_i^{\top} = x_i^{\top} X(h)^{-1}$

$$0 \leq - \sum_{i \in h} \psi_{\tau}^*(0, -v_i) v_i - \xi^{\top} v = - \sum_{i \in h} (\tau - I(-v_i < 0)) v_i - \xi^{\top} v$$

where $\xi^{\top} = \sum_{i \in \bar{h}} \psi_{\tau}^*(y_i - x_i^{\top} b(h), -x_i^{\top} X(h)^{-1} v) x_i^{\top} X(h)^{-1}$

Optimality Conditions (9)

- Provided that

$$y_i - x_i^\top b(h) \neq 0 \quad \text{for any } i \notin h$$

the directional derivative condition holds for all v iff it holds for $v = \pm e_k (k = 1, \dots, p)$

- Thus, we have p inequalities for $v = e_i$

$$0 \leq -(\tau - 1) - \xi_i(h) \quad i = 1, \dots, p,$$

for $v = -e_i$

$$0 \leq \tau + \xi_i(h) \quad i = 1, \dots, p$$

Optimality Conditions (10)

Definition 2.1. *We say that the regression observations (y, X) are in general position if any p of them yield a unique exact fit; that is, for any $h \in \mathcal{H}$,*

$$y_i - x_i^\top b(h) \neq 0 \quad \text{for any } i \notin h.$$

Optimality Conditions (11)

Theorem 2.1. *If (y, X) are in general position, then there exists a solution to quantile-regression problem (2.4) of the form $b(h) = X(h)^{-1}y(h)$ if and only if, for some $h \in \mathcal{H}$,*

$$-\tau 1_p \leq \xi(h) \leq (1 - \tau)1_p, \quad (2.5)$$

where $\xi^\top(h) = \sum_{i \in \bar{h}} \psi_\tau(y_i - x_i^\top b(h)) x_i^\top X(h)^{-1}$ and $\psi_\tau = \tau - I(u < 0)$. Furthermore, $b(h)$ is the unique solution if and only if the inequalities are strict; otherwise, the solution set is the convex hull of several solutions of the form $b(h)$.

Optimality Conditions (12)

Theorem 2.2. *Let P , N , and Z denote the proportion of positive, negative, and zero elements of the residual vector $y - X\hat{\beta}(\tau)$. If X contains an intercept, that is, if there exists $\alpha \in \mathbb{R}^p$ such that $X\alpha = \mathbf{1}_n$, then for any $\hat{\beta}(\tau)$, solving (1.19), we have*

$$N \leq n\tau \leq N + Z$$

and

$$P \leq n(1 - \tau) \leq P + Z.$$

Proof of Theorem 2.2

Proof. We have optimality of $\hat{\beta}(\tau)$ if and only if

$$-\sum_{i=1}^n \psi_{\tau}^*(y_i - x_i^{\top} \hat{\beta}(\tau), -x_i^{\top} w) x_i^{\top} w \geq 0$$

for all directions $w \in \mathbb{R}^p$. For $w = \alpha$, such that $X\alpha = 1_n$, we have

$$-\sum \psi_{\tau}^*(y_i - x_i^{\top} \hat{\beta}(\tau), -1) \geq 0,$$

which yields

$$\tau P - (1 - \tau)N - (1 - \tau)Z \leq 0.$$

Similarly, for $w = -\alpha$, we obtain

$$-\tau P + (1 - \tau)N - \tau Z \leq 0.$$

Combining these inequalities and using the fact that $n = N + P + Z$ completes the proof. ■

Robustness (1)

- Influence function

$$F_\varepsilon = \varepsilon\delta_y + (1 - \varepsilon)F.$$

$$IF_{\hat{\theta}}(y, F) = \lim_{\varepsilon \rightarrow 0} \frac{\hat{\theta}(F_\varepsilon) - \hat{\theta}(F)}{\varepsilon}$$

- For the mean

$$\hat{\theta}(F_\varepsilon) = \int y dF_\varepsilon = \varepsilon y + (1 - \varepsilon)\hat{\theta}(F)$$

$$IF_{\hat{\theta}}(y, F) = y - \hat{\theta}(F)$$

- For the median

$$\tilde{\theta}(F_\varepsilon) = F_\varepsilon^{-1}(1/2)$$

$$IF_{\tilde{\theta}}(y, F) = \frac{\text{sgn}(y - F^{-1}(\frac{1}{2}))}{2f(F^{-1}(\frac{1}{2}))}$$

Robustness (2)

- Influence function

$$F_\varepsilon = \varepsilon \delta_y + (1 - \varepsilon)F.$$

$$IF_{\hat{\theta}}(y, F) = \lim_{\varepsilon \rightarrow 0} \frac{\hat{\theta}(F_\varepsilon) - \hat{\theta}(F)}{\varepsilon}$$

- Assume $y < F^{-1}(\frac{1}{2})$. If $F_\varepsilon(\tilde{\theta}(F_\varepsilon)) = \frac{1}{2}$

$$\varepsilon + (1 - \varepsilon)F(\tilde{\theta}(F_\varepsilon)) = \frac{1}{2}$$

$$\tilde{\theta}(F_\varepsilon) = F^{-1}\left(\frac{\frac{1}{2} - \varepsilon}{1 - \varepsilon}\right) = F^{-1}\left(\frac{\frac{1}{2} - \varepsilon}{1 - \varepsilon}\right)$$

Robustness (3)

$$\begin{aligned} IF_{\tilde{\theta}}(y, F) &= \lim_{\epsilon \rightarrow 0} \frac{\tilde{\theta}(F_{\epsilon}) - \tilde{\theta}(F)}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{F^{-1}\left(\frac{\frac{1}{2} - \epsilon}{1 - \epsilon}\right) - F^{-1}\left(\frac{1}{2}\right)}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{F^{-1}\left(\frac{1}{2} - \frac{1}{2} \frac{\epsilon}{1 - \epsilon}\right) - F^{-1}\left(\frac{1}{2}\right)}{\epsilon} \\ &= -\frac{1}{2f\left(F^{-1}\left(\frac{1}{2}\right)\right)} \end{aligned}$$

Robustness (4)

- For quantile regression, the influence function is

$$IF_{\hat{\theta}}(y, F) = Q^{-1} x \operatorname{sgn}(y - x^T \hat{\beta}_F(\tau))/2$$

where

$$Q = \int x x^T f(x^T \hat{\beta}_F(x)) dG(x).$$

Robustness (5)

Theorem 2.4. *Let D be a diagonal matrix with nonnegative elements d_i , for $i = 1, \dots, n$; then*

$$\hat{\beta}(\tau; y, X) = \hat{\beta}(\tau; X\hat{\beta}(\tau; y, X) + D\hat{u}, X),$$

where $\hat{u} = y - X\hat{\beta}(\tau; y, X)$.

Interpreting Quantile Regression (1)

- In linear regression

$$E(Y|X = x) = x^\top \beta$$

$$\frac{\partial E(Y|X = x)}{\partial x_j} = \beta_j$$

- In case of dependent x , *e.g.*

$$E(Y|Z = z) = \beta_0 + \beta_1 z + \beta_2 z^2$$

$$\frac{\partial E(Y|Z = z)}{\partial z} = \beta_1 + 2\beta_2 z$$

Interpreting Quantile Regression (2)

- In the transformed model

$$E(h(Y)|X = x) = x^\top \beta,$$

- It is attempting to write

$$\frac{\partial E(Y|X = x)}{\partial x_j} = \frac{\partial h^{-1}(x^\top \beta)}{\partial x_j}$$

Interpreting Quantile Regression (2)

- In the transformed model

$$E(h(Y)|X = x) = x^\top \beta,$$

- It is attempting to write

$$\frac{\partial E(Y|X = x)}{\partial x_j} = \frac{\partial h^{-1}(x^\top \beta)}{\partial x_j}$$

Interpreting Quantile Regression (3)

- In case of quantile regression, for any increasing function h

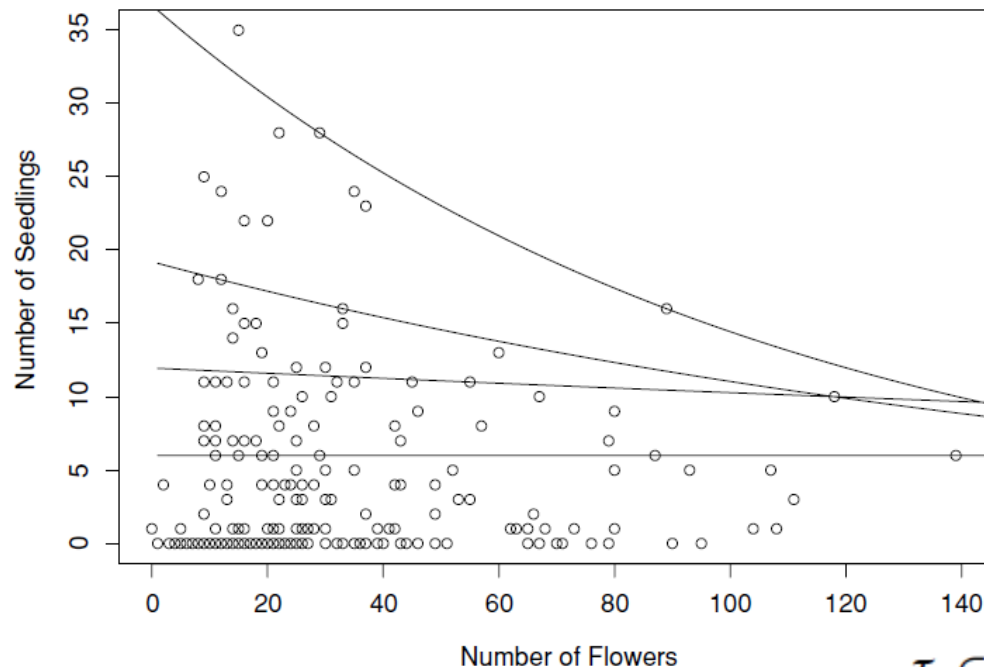
$$Q_{h(Y)}(\tau|X = x) = h(Q_Y(\tau|X = x))$$

- If $Q_{h(Y)}(\tau|X = x) = x^\top \beta(\tau)$, we have

$$\frac{\partial Q_Y(\tau|X = x)}{\partial x_j} = \frac{\partial h^{-1}(x^\top \beta)}{\partial x_j}$$

An example-Glacier lily(1)

- Cade et al. considered models for the prevalence of glacier lily seedling
- Response: number of lily seedling in 256 2X2m quadrats
- Covariate: number of flowers in the same area



$$\tau \in \{0.75, 0.9, 0.95, 0.99\}$$

Glacier Lily



Glacier Lily

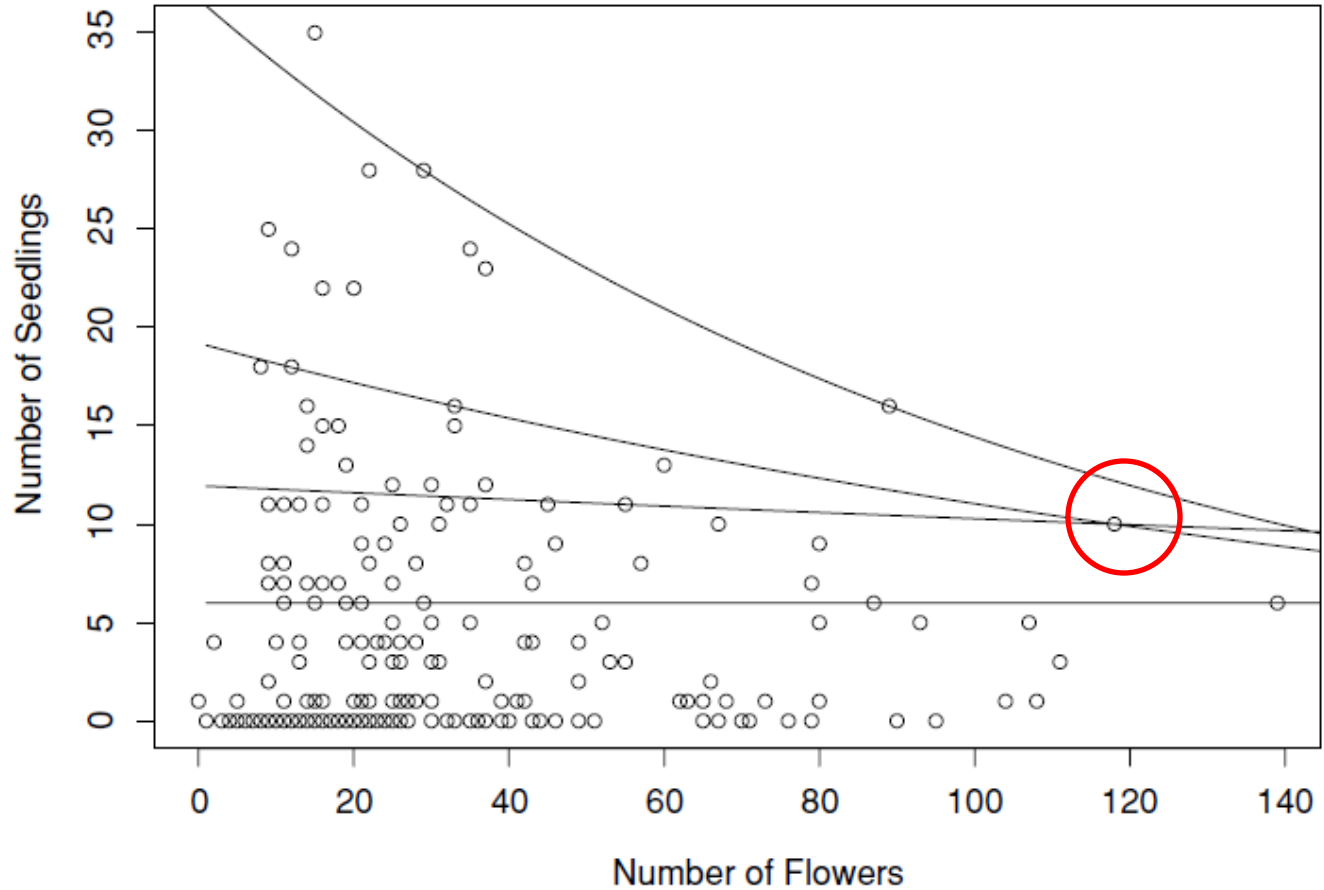


An example-Glacier lily(2)

- There is a strong negative relationship between the number of seedlings and the number of observed flowers in the upper tail of the conditional distribution

Negative slopes for upper regression quantiles were consistent with the explanation provided by Thompson et al. that sites where flowers were most numerous because of lack of pocket gophers (which eat lilies), were rocky sites that provided poor moisture conditions for seed germination; hence seedling numbers were lower.

Crossing Quantiles (1)



Crossing Quantiles (2)

- Let $\bar{x} = n^{-1} \sum x_i$ be the centroid of the design, the estimated conditional quantile function is

$$\hat{Q}_Y(\tau|\bar{x}) = \bar{x}^\top \hat{\beta}(\tau)$$

Theorem 2.5. *The sample paths of $\hat{Q}_Y(\tau|\bar{x})$ are nondecreasing in τ on $[0, 1]$.*

Crossing Quantiles (3)

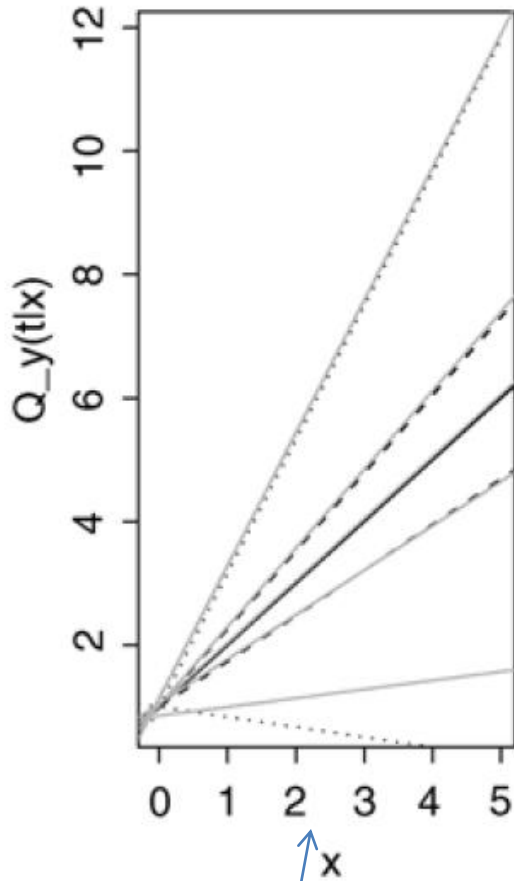
- In case serious crossing quantile, this can be taken as an evidence of misspecification of the covariate effects
- Consider the model

$$y_i = \beta_0 + x_i\beta_1 + (\gamma_0 + \gamma_1x_i)v_i$$

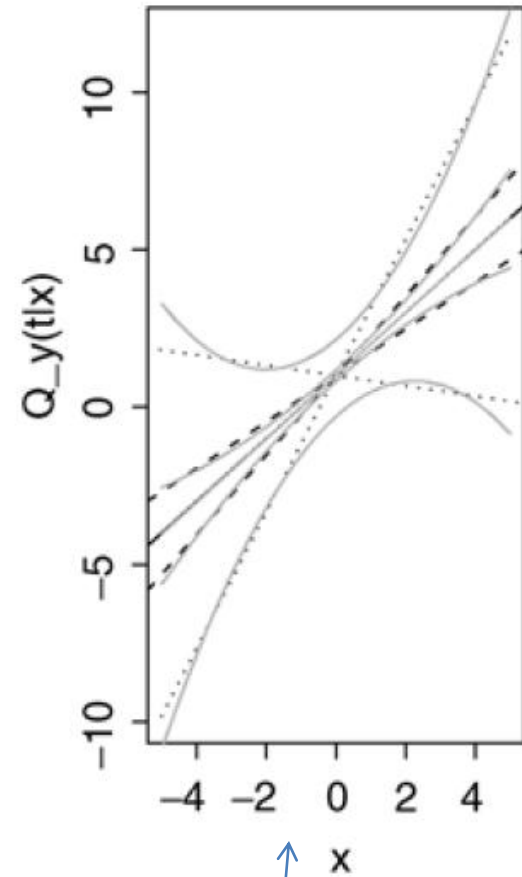
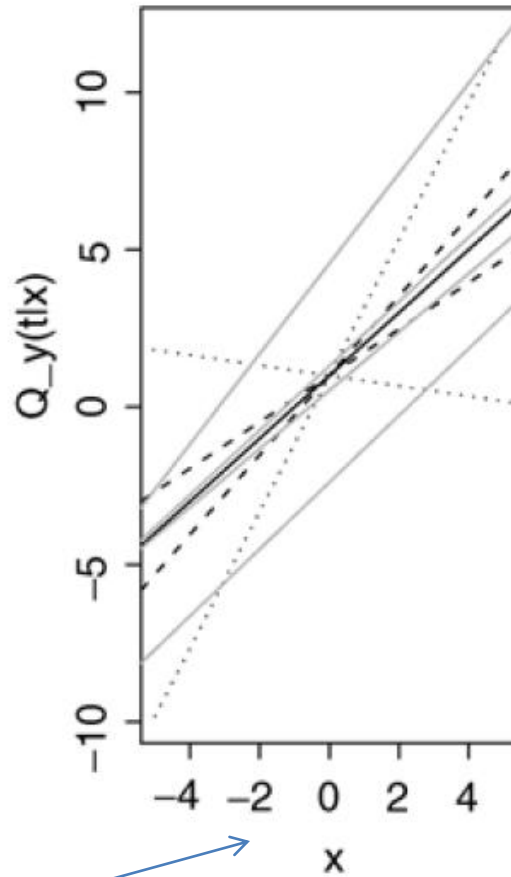
$$Q_Y(\tau|x_i) = \begin{cases} \beta_0 + x_i\beta_1 + (\gamma_0 + \gamma_1x_i)F^{-1}(\tau) & \text{if } \gamma_0 + \gamma_1x_i \geq 0 \\ \beta_0 + x_i\beta_1 + (\gamma_0 + \gamma_1x_i)F^{-1}(1 - \tau) & \text{otherwise.} \end{cases}$$

Crossing Quantiles (4)

$$\beta_0 = 1, \beta_1 = 1, \gamma_0 = 0, \gamma_1 = 1$$



Linear model



Quadratic model