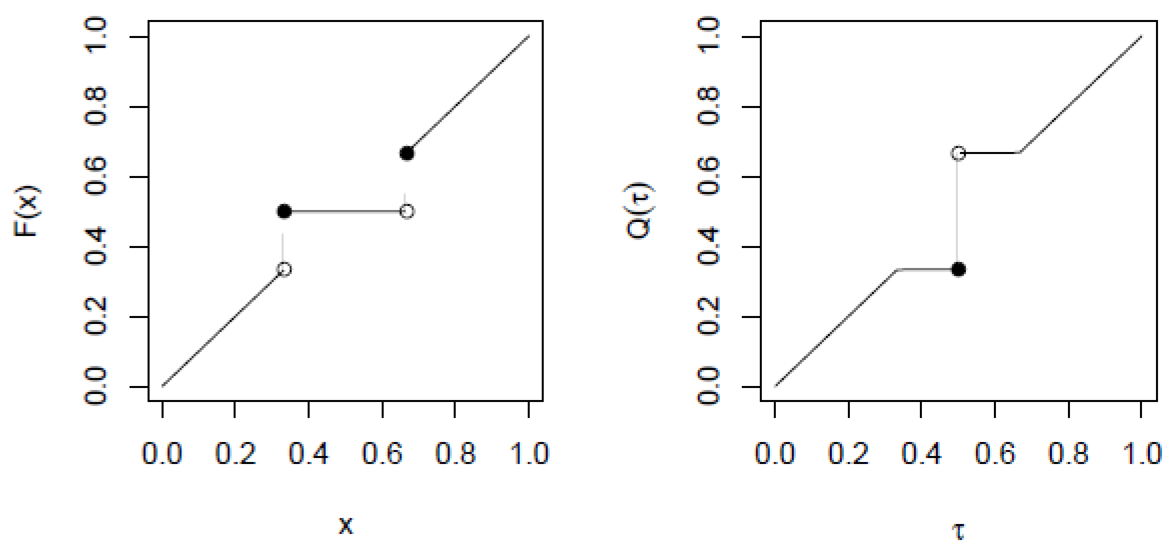
Quantile Regression

Ruibin Xi

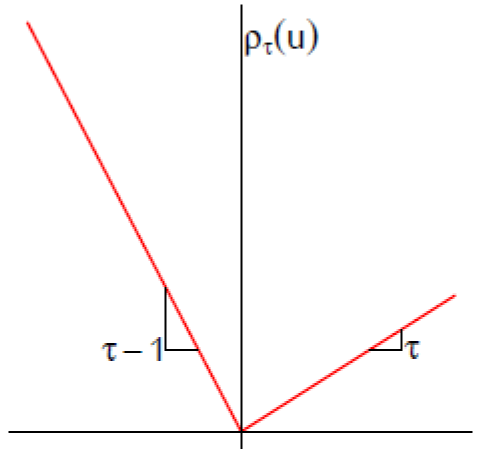
Univariate Quantile

- Given a real-valued random variable, X, with distribution function F, we define the τth quantile of X as

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x \mid F(x) \geqslant \tau\}.$$



The Check function



- We define a loss function

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

- Note that if τ=0.5, $\rho_\tau(u) = |u|/2$
- Quantiles solve a simple optimization problem

$$\hat{\alpha}(\tau) = \text{argmin}\, \mathbb{E}\, \rho_\tau(Y - \alpha)$$

The Check function

- We seek to minimize

$$E\rho_\tau(X - \hat{x}) = (\tau - 1)\int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x})dF(x).$$

- Differentiating w.r.t. $\hat{x}$, we have

$$0 = (1 - \tau)\int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{\infty} dF(x) = F(\hat{x}) - \tau.$$

$$\hat{x} = F^{-1}(\tau)$$

Quantile Regression

- The unconditional quantile solves

$$\alpha_\tau = \mathrm{argmin}_\alpha E\rho_\tau(Y - \alpha)$$

- The conditional quantile solves

$$\alpha_\tau(x) = \mathrm{argmin}_\alpha E_{Y|X=x}\, \rho_\tau(Y - \alpha(X))$$

- Similarly, assume $\alpha_\tau(x) = \beta^T x$, we have the sample version of the problem

$$\mathrm{argmin}_\beta \sum \rho_\tau(y_i - \beta^T x_i)$$

Computation of Quantile Regression (1)

- Linear QR: solve

$$\min_{b \in \mathcal{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T b)$$

- Let

$$e_i = y_i - x_i^T b$$

$$u_i = \max\{e_i, 0\}$$

$$v_i = \max\{-e_i, 0\}$$

$$e_i = u_i - v_i$$

Computation of Quantile Regression (2)

- We have

$$
\begin{aligned}
\rho_\tau(y_i - x_i^T b) &= \rho_\tau(e_i) \\
&= \rho_\tau(u_i - v_i) \\
&= \tau u_i + (\tau - 1)(-v_i) \\
&= \tau u_i + (1 - \tau) v_i
\end{aligned}
$$

- Minimizing $\sum_{i=1}^{n} \rho_\tau(y_i - x_i^T b)$ is equivalent to minimizing

$$\sum_{i=1}^{n} [\tau u_i + (1 - \tau) v_i]$$

with constraints

$$u_i \geq 0, \quad v_i \geq 0, \quad y_i - x_i^T b = u_i - v_i$$

Computation of Quantile Regression (3)

- If we write

$$u = (u_1, \cdots, u_n)^T \quad v = (v_1, \cdots, v_n)^T$$

- the above problem can be written as

$$\min \left\{ \tau e_n^\top u + (1-\tau) e_n^\top v \,|\, y - Xb = u - v,\ b \in \mathbb{R}^p,\ (u, v) \in \mathbb{R}_+^{2n} \right\}$$

- Note that this is a linear programming (LP) problem

Computation of Quantile Regression (4)

- For the LP problem

$$\min\left\{\tau e_n^\top u + (1-\tau)e_n^\top v \,|\, y - Xb = u - v,\ b \in \mathbb{R}^p,\ (u,v) \in \mathbb{R}_+^{2n}\right\}$$

- the minimum can be obtained at the vertices of the feasible region
- The solution can be written as

$$b(h) = X(h)^{-1} y(h)$$

where h is a p-element subset of {1,…,n}.

Quantile Treatment effect

Lehmann (1974) proposed the following general model of treatment response:

> *"Suppose the treatment adds the amount* $\Delta(x)$ *when the response of the untreated subject would be* x*. Then the distribution* G *of the treatment responses is that of the random variable* $X + \Delta(X)$ *where* X *is distributed according to* F*."*

Lehmann QTE as a QQ-plot

Doksum (1974) defines $\Delta(x)$ as the "horizontal distance" between F and G at x, *i.e.*

$$F(x) = G(x + \Delta(x)).$$

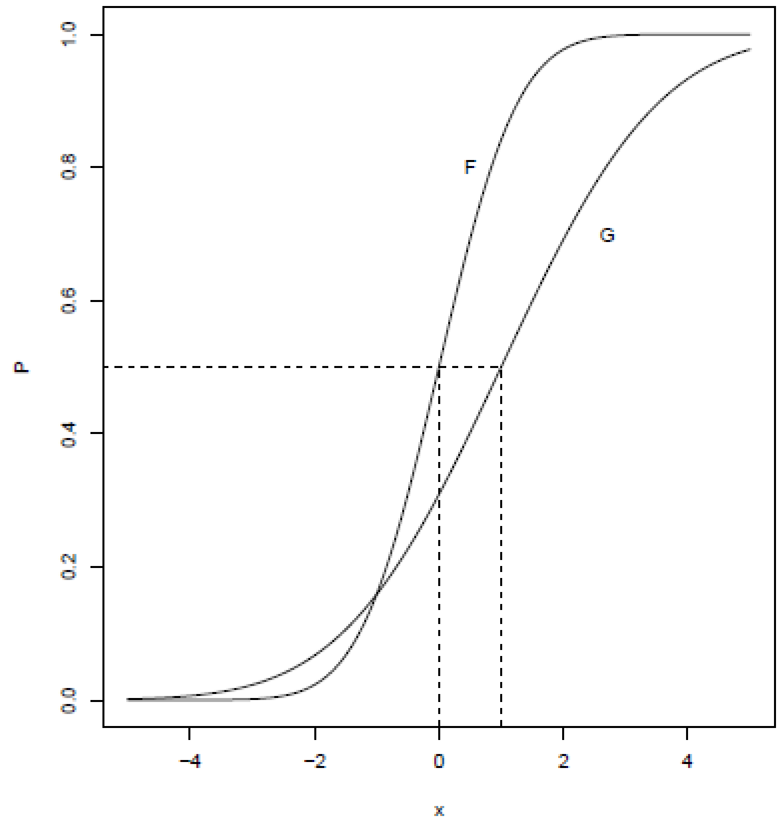
Then $\Delta(x)$ is uniquely defined as

$$\Delta(x) = G^{-1}(F(x)) - x.$$

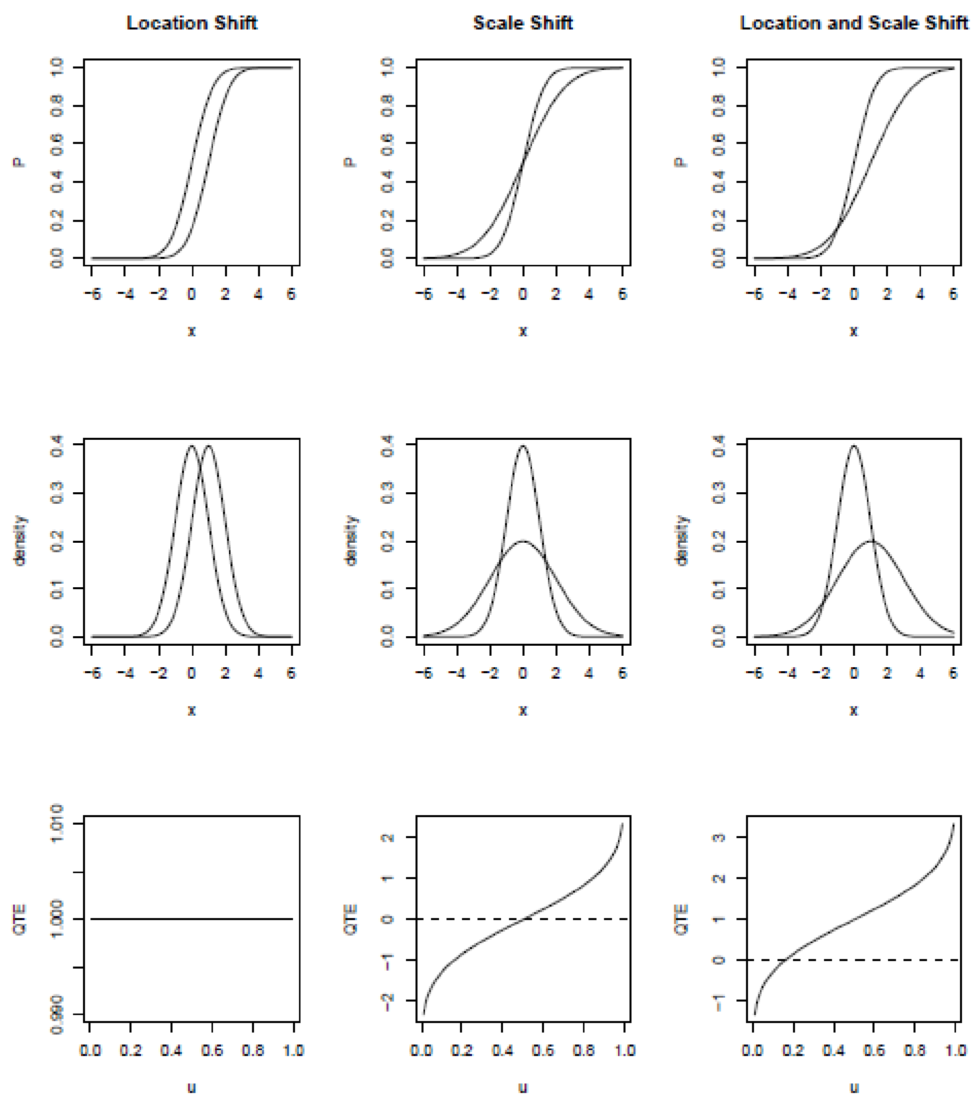
This is the essence of the conventional QQ-plot. Changing variables so $\tau = F(x)$ we have the quantile treatment effect (QTE):

$$\delta(\tau) = \Delta(F^{-1}(\tau)) = G^{-1}(\tau) - F^{-1}(\tau).$$

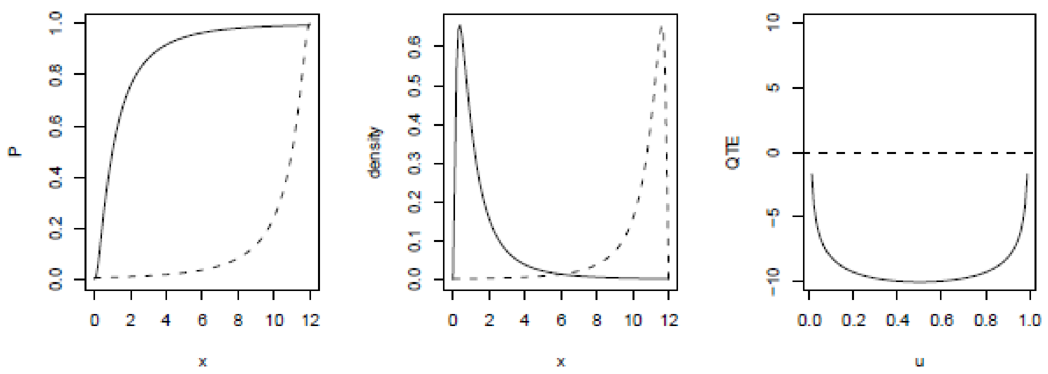
Lehmann-Doksum QTE



Lehmann-Doksum QTE



An asymmetric example



Treatment shifts the distribution from right skewed to left skewed making the QTE U-shaped.

QTE via quantile regression

The Lehmann QTE is naturally estimable by

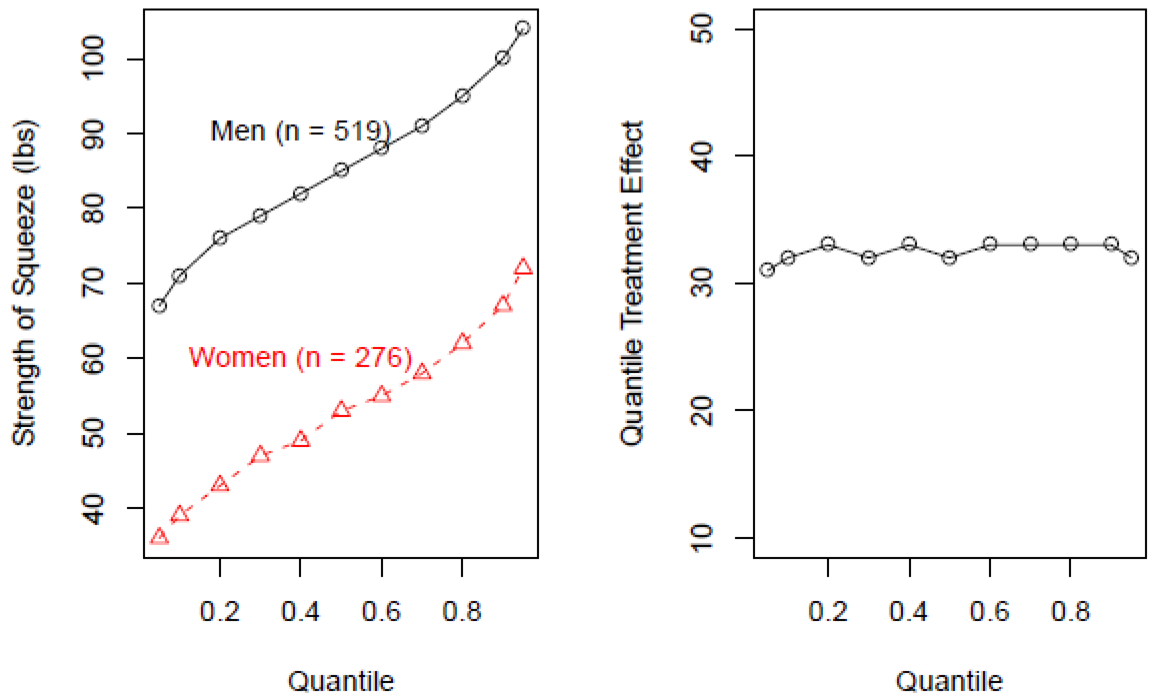
$$\hat{\delta}(\tau) = \hat{G}_n^{-1}(\tau) - \hat{F}_m^{-1}(\tau)$$

where $\hat{G}_n$ and $\hat{F}_m$ denote the empirical distribution functions of the treatment and control observations, Consider the quantile regression model

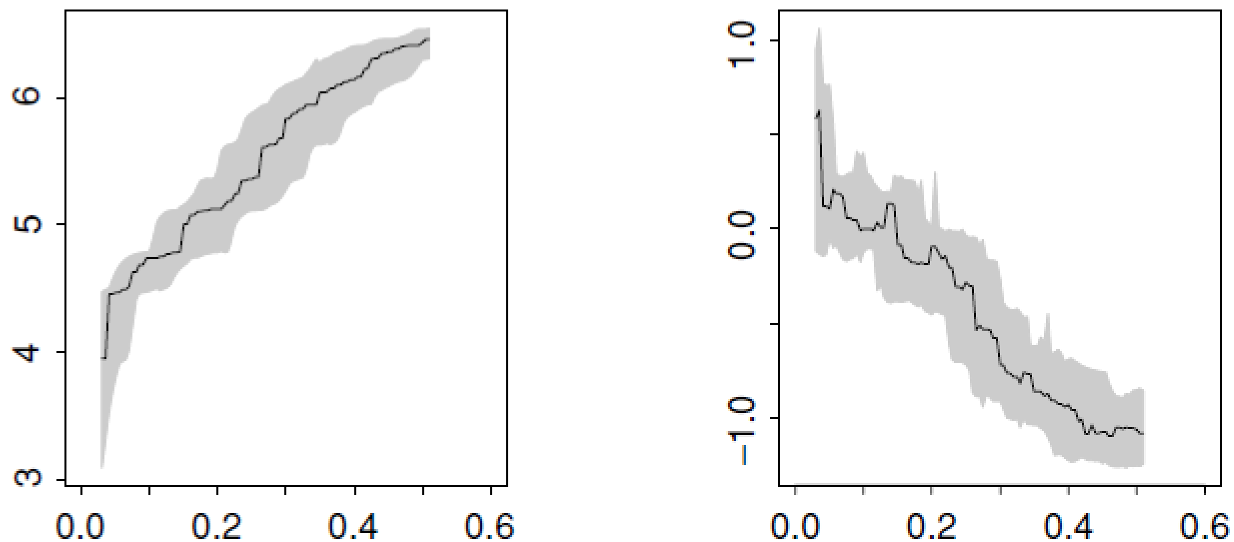
$$Q_{Y_i}(\tau|D_i) = \alpha(\tau) + \delta(\tau)D_i$$

where D_i denotes the treatment indicator

QTE: Strength of squeeze



QTE: guinea pig



Injection of tubercle baccilli to guinea pig

Equivariance of Quantile Regression

Theorem 2.3 (Koenker and Bassett, 1978). *Let A be any $p \times p$ nonsingular matrix, $\gamma \in \mathbb{R}^p$, and $a > 0$. Then, for any $\tau \in [0, 1]$,*

(i) $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$

(ii) $\hat{\beta}(\tau; -ay, X) = -a\hat{\beta}(1 - \tau; y, X)$

(iii) $\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma$

(iv) $\hat{\beta}(\tau; y, XA) = A^{-1}\hat{\beta}(\tau; y, X).$

Equivariance to monotone transformation

For any monotone function h, conditional quantile functions $Q_Y(\tau|x)$ are equivariant in the sense that

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

In contrast to conditional mean functions for which, generally,

$$E(h(Y)|X) \neq h(EY|X)$$

Censoring (1)

- Let y_i^* denote a latent (unobservable) response

$$y_i^* = x_i^\top \beta + u_i \quad i = 1, \ldots, n,$$

where $\{u_i\}$ is i.i.d. from F with density f.

- Due to censoring, we do not observe y_i^*, but we observe

$$y_i = \max\{0, y_i^*\}$$

Censoring (2)

- The model may be estimated by MLE

$$\hat{\beta} = \text{argmin}_\beta \left\{ \prod_{i=1}^{n} (1 - F(x_i^\top \beta))^{\Delta_i} f(y_i - x_i^\top \beta)^{1-\Delta_i} \right\}$$

where $\Delta_i = 1$, if the ith observation is cernsored, and 0 otherwise

- If F is misspecified, inference based on this estimator will be misleading.

Censoring (3)

- We may express the τth conditional quantile function of the observed response y_i as

$$Q_{y_i}(\tau|x_i) = \max\{0, x_i^\top \beta + F_u^{-1}(\tau)\}.$$

- The parameters can be estimated by

$$\min_b \sum_{i=1}^{n} \rho_\tau(y_i - \max\{0, x_i^\top b\})$$

where x_i is assumed to contain an intercept term to absorb the additive effect $F_u^{-1}(\tau)$

Robustness (1)

- Influence function

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

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

- 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

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

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

Robustness (2)

- Influence function

$$F_{\varepsilon} = \varepsilon \delta_{y} + (1 - \varepsilon) F$$

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

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

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

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

Robustness (3)

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

Robustness (4)

- For quantile regression, the influence function is

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

where

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

Robustness (5)

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

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

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