Quantile Regression (similar tools used for duration models)

- Distributional models

Let Y be a random variable with df $F_Y(y) = Pr[Y \leq y]$.
The quantile function is $Q_Y(\tau)$ where

$$Q_Y(\tau) = \inf\{y : Pr[Y \leq y] \geq \tau\}$$

Examples:

$\tau = \frac{1}{2} \Rightarrow Q_Y(\frac{1}{2}) =$ median of Y

$\tau = 0.1 \Rightarrow Q_Y(0.1) = 1^{st}$ decile

$\tau = 0.25 \Rightarrow Q_Y(0.25) = 1^{st}$ quartile

$\tau = 0.01 \Rightarrow Q_Y(0.01) = 1^{st}$ centile of Y

$Q_Y(\tau|X)$ quantile regression estimates this

$F_Y(y|X)$ duration models estimate this

- $1 - F_Y(y|X)$ survival function

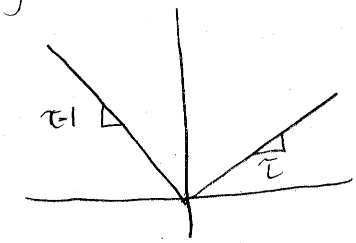
Univariate τ^{th} quantile

$$\hat{\alpha}(\tau) \equiv \underset{\alpha \in \mathbb{R}}{\text{argmin}} \sum_{i=1}^{n} \underbrace{\rho_\tau}_{\text{check penalty function}} (y_i - a)$$

- here, we can just estimate by sorting, but we can also express this as an optimization problem.
- ρ_τ asymmetrically penalizes deviations from y_i

$\rho_\tau(u) = (\tau - I(u<0))u = \tau u^+ + (1-\tau)u^-$, $u^+ = \max\{0,u\}$, $u^- = -\min\{u,0\}$

$\Rightarrow \rho_\tau(y_i - a) = (\tau - I(y_i < a))(y_i - a)$



Regression τ^{th} quantile:

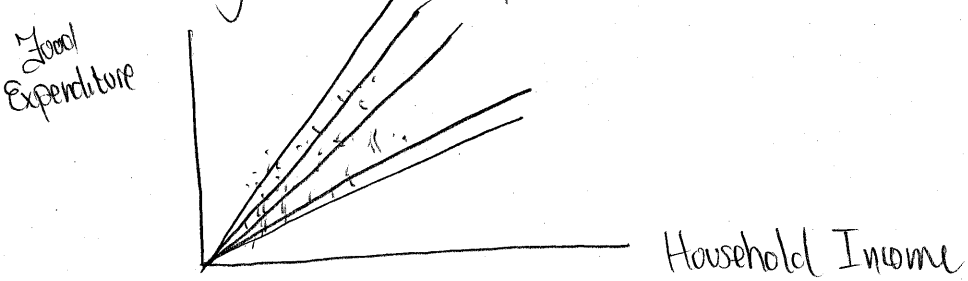
$$\hat\beta(\tau) = \arg\min_b \sum_{i=1}^n \rho_\tau(y_i - x_i' b)$$

Want to get $Q_Y(\tau|X) \approx X'\beta(\tau)$

- $\beta(\tau)$ can depend on τ
  - ie smoking can affect $\tau=\frac{1}{2}$ birthweights much differently than $\tau = 0.05$ birthweights

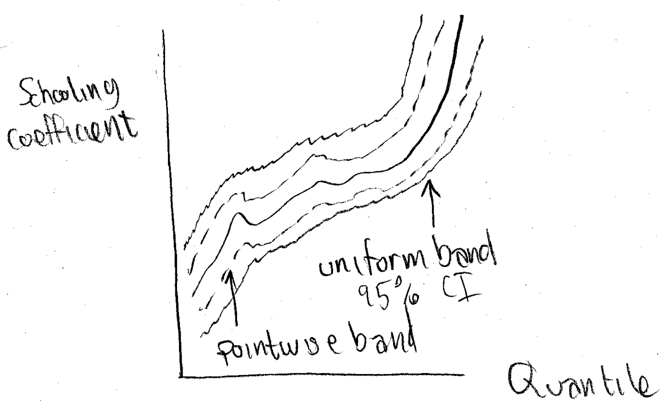
The estimate of the τ'th conditional quantile is $x_i'\hat\beta(\tau)$

- Heterogeneous responses to changes in income.



Angrist-Chernozhukov

$\underbrace{y_i}_{\text{log earning}} = \underbrace{X_i}_{\text{education}} B$



- $$\sup_{\tau \in [0.01, 0.99]} \left| \frac{\hat{\beta}(\tau) - \beta(\tau)}{se(\hat{\beta}(\tau))} \right|$$
- need asymptotic distribution to exactly characterize this to compute uniform CI band

- alternatively, use bootstrapping and find maximal t-statistic.