Quantile and Expectile Regressions

Arthur Charpentier (Université de Rennes 1)

Erasmus School of Economics, May 2017

http://freakonometrics.hypotheses.org
Econometrics vs Machine Learning

As claimed in Freedman (2005), Statistical Models, in econometrics, given some probability space \((\Omega, \mathcal{A}, \mathbb{P})\), assume that \(y_i\) are realization of i.i.d. variables \(Y_i\) (given \(X_i = x_i\)) with distribution \(F_{m_i}\) ("conditional distribution story" or "causal story"). E.g.

\[
Y|X = x \sim \mathcal{N}(x^T \beta, \sigma^2) \quad \text{or} \quad Y|X = x \sim \mathcal{B}(m(x)) \quad \text{where} \quad m(x) = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}}
\]

Then solve ("maximum likelihood" framework)

\[
\hat{m}(\cdot) = \arg\max_{m(\cdot) \in \mathcal{F}} \left\{ \sum_{i=1}^{n} \log f_{m(x_i)}(y_i) \right\}
\]

where \(\log \mathcal{L}\) denotes the log-likelihood, see Haavelmo (1944) The Probability Approach in Econometrics.
Econometrics vs Machine Learning

In machine learning (“explanatory data story” in Freedman (2005), Statistical Models), given some dataset \((x_i, y_i)\), solve

\[ m^*(\cdot) = \arg\min_{m(\cdot) \in \mathcal{F}} \left\{ \sum_{i=1}^{n} \ell(m(x_i), y_i) \right\} \]

for some loss functions \( \ell(\cdot, \cdot) \). There is no probabilistic model \textit{per se}.

But some loss functions have simple and interesting properties, e.g.

\[ \ell_s(x, y) = 1_{x > s \neq y} \]

for classification, associated with missclassification dummy (with cutoff threshold \(s\)),

\[
\begin{array}{c|cc}
& y_i = 0 & y_i = 1 \\
\hline
m(x_i) \leq s & n_00 & n_01 \\
m(x_i) > s & n_{10} & n_{11} \\
\end{array}
\]

\[ \hat{m}_s(\cdot) = \arg\min_{m(\cdot) \in \mathcal{F}} \{n_{01} + n_{01}\} \]
OLS Regression, $\ell_2$ norm and Expected Value: $\ell_2(x, y) = [x - y]^2$

Let $\mathbf{y} \in \mathbb{R}^n$, $\overline{y} = \text{argmin}_{m \in \mathbb{R}} \left\{ \sum_{i=1}^{n} \frac{1}{n} [y_i - m]^2 \right\}$. It is the empirical version of

$$
\mathbb{E}[Y] = \text{argmin}_{m \in \mathbb{R}} \left\{ \int [y - m]^2 dF(y) \right\} = \text{argmin}_{m \in \mathbb{R}} \left\{ \mathbb{E}\left[ \|Y - m\|_{\ell_2} \right] \right\}
$$

where $Y$ is a random variable.

Thus, $\text{argmin}_{m(\cdot): \mathbb{R}^k \to \mathbb{R}} \left\{ \sum_{i=1}^{n} \frac{1}{n} [y_i - m(\mathbf{x}_i)]^2 \right\}$ is the empirical version of $\mathbb{E}[Y|X = \mathbf{x}]$.

See Legendre (1805) *Nouvelles méthodes pour la détermination des orbites des comètes* and Gauss (1809) *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.*
Median Regression, $\ell_1$ norm and Median: $\ell_1(x, y) = |x - y|$

Let $y \in \mathbb{R}^n$, median[$y$] $\in \arg\min_{m \in \mathbb{R}} \left\{ \sum_{i=1}^{n} \frac{1}{n} |y_i - m| \right\}$. It is the empirical version of

$$\text{median}[Y] \in \arg\min_{m \in \mathbb{R}} \left\{ \int |y - m| \, dF(y) \right\} = \arg\min_{m \in \mathbb{R}} \left\{ \mathbb{E} [\|Y - m\|_{\ell_1}] \right\}$$

where $Y$ is a random variable, $\mathbb{P}[Y \leq \text{median}[Y]] \geq \frac{1}{2}$ and $\mathbb{P}[Y \geq \text{median}[Y]] \geq \frac{1}{2}$.

$$\arg\min_{m(\cdot): \mathbb{R}^k \to \mathbb{R}} \left\{ \sum_{i=1}^{n} \frac{1}{n} |y_i - m(x_i)| \right\}$$

is the empirical version of $\text{median}[Y|X = x]$.

See Boscovich (1757) De Litteraria expeditione per pontificiam ditionem ad dimetiendos duos meridiani and Laplace (1793) Sur quelques points du système du monde.
Quantiles and Expectiles, $\ell(x, y) = R(x - y)$

Consider the following risk functions

$$R^q_\tau(u) = u \cdot (\tau - 1(u < 0)), \quad \tau \in [0, 1]$$

with $R^q_{1/2}(u) \propto |u| = \|u\|_1$, and

$$R^e_\tau(u) = u^2 \cdot (\tau - 1(u < 0)), \quad \tau \in [0, 1]$$

with $R^e_{1/2}(u) \propto u^2 = \|u\|_2^2$.

$$Q_Y(\tau) = \arg\min_m \{E(R^q_\tau(Y - m))\}$$

which is the median when $\tau = 1/2$,

$$E_Y(\tau) = \arg\min_m \{E(R^e_\tau(X - m))\}$$

which is the expected value when $\tau = 1/2$. 
Elicitable Measures

“elicitable” means “being a minimizer of a suitable expected score”

$T$ is an elicatable function if there exits a scoring function $S : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ (see functions $\ell(\cdot, \cdot)$ introduced above) such that

$$T(Y) = \arg\min_{x \in \mathbb{R}} \left\{ \int_{\mathbb{R}} S(x, y) dF(y) \right\} = \arg\min_{x \in \mathbb{R}} \{ \mathbb{E}[S(x, Y)] \text{ where } Y \sim F. \}$$

see Gneiting (2011) Making and evaluating point forecasts.

The mean, $T(Y) = \mathbb{E}[Y]$ is elicited by $S(x, y) = \|x - y\|_{\ell_2}^2$

The median, $T(Y) = \text{median}[Y]$ is elicited by $S(x, y) = \|x - y\|_{\ell_1}$

The quantile, $T(Y) = Q_Y(\tau)$ is elicited by $S(x, y) = \tau(y - x)_+ + (1 - \tau)(y - x)_-$

The expectile, $T(Y) = E_Y(\tau)$ is elicited by $S(x, y) = \tau(y - x)_+^2 + (1 - \tau)(y - x)_-^2$
Quantiles and Expectiles (Technical Issues)

One can also write empirical quantiles and expectiles as

**quantile:** \( \text{argmin} \left\{ \sum_{i=1}^{n} \omega_{q}^{q}(\varepsilon_i) \left| y_i - q_i \right| \right\} \) where \( \omega_{q}^{q}(\epsilon) = \begin{cases} 1 - \tau & \text{if } \epsilon \leq 0 \\ \tau & \text{if } \epsilon > 0 \end{cases} \)

**expectile:** \( \text{argmin} \left\{ \sum_{i=1}^{n} \omega_{e}^{e}(\varepsilon_i) (y_i - q_i)^2 \right\} \) where \( \omega_{e}^{e}(\epsilon) = \begin{cases} 1 - \tau & \text{if } \epsilon \leq 0 \\ \tau & \text{if } \epsilon > 0 \end{cases} \)

Expectiles are unique, not quantiles...

Quantiles satisfy \( \mathbb{E}[\text{sign}(Y - Q_Y(\tau))] = 0 \)

Expectiles satisfy \( \tau \mathbb{E}[(Y - E_Y(\tau))_+] = (1 - \tau) \mathbb{E}[(Y - E_Y(\tau))_-] \)

(those are actually the first order conditions of the optimization problem).
Expectiles as Quantiles (Interpretation)

For every $Y \in L^1$, $\tau \mapsto E_Y(\tau)$ is continuous, and strictly increasing

if $Y$ is absolutely continuous, \[ \frac{\partial E_Y(\tau)}{\partial \tau} = \frac{\mathbb{E}[|X - E_Y(\tau)|]}{(1 - \tau)F_Y(E_Y(\tau)) + \tau(1 - F_Y(E_Y(\tau)))} \]

“Expectiles have properties that are similar to quantiles” Newey & Powell (1987)

Asymmetric Least Squares Estimation and Testing. The reason is that expectiles of a distribution $F$ are quantiles a distribution $G$ which is related to $F$, see Jones (1994) Expectiles and M-quantiles are quantiles: let

\[ G(t) = \frac{P(t) - tF(t)}{2[P(t) - tF(t)] + t - \mu} \]

where $P(s) = \int_{-\infty}^{s} ydF(y)$.

The expectiles of $F$ are the quantiles of $G$. 
Distortion: Expectiles as Quantiles (Gaussian, Lognormal & Poisson)
Empirical Quantiles & Empirical Expectiles

Consider some i.i.d. sample \{y_1, \cdots, y_n\} with distribution \(F\) (abs. continuous).

\[ Q_\tau = \arg\min \{ \mathbb{E}[\mathcal{R}_\tau^q(Y - q)] \} \quad \text{where} \quad Y \sim F \quad \text{and} \quad \hat{Q}_\tau \in \arg\min \left\{ \sum_{i=1}^{n} \mathcal{R}_\tau^q(y_i - q) \right\} \]

Then as \(n \to \infty\), \(\sqrt{n}(\hat{Q}_\tau - Q_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\tau(1 - \tau)}{f^2(Q_\tau)}\right)\)

\[ E_\tau = \arg\min \{ \mathbb{E}[\mathcal{R}_\tau^e(Y - m)] \} \quad \text{where} \quad Y \sim F \quad \text{and} \quad \hat{E}_\tau = \arg\min \left\{ \sum_{i=1}^{n} \mathcal{R}_\tau^e(y_i - m) \right\} \]

Then as \(n \to \infty\), \(\sqrt{n}(\hat{E}_\tau - E_\tau) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, s_\tau^2\right)\) where, if \(\mathcal{I}_\tau(x, y) = \tau(y - x)_+ + (1 - \tau)(y - x)_-\) (elicitable score for quantiles),

\[ s_\tau^2 = \frac{\mathbb{E}[\mathcal{I}_\tau(E_\tau, Y)^2]}{\left(\tau[1 - F(E_\tau)] + [1 - \tau]F(Eu_\tau)\right)^2}. \]
Quantile Regression (and Heteroskedasticity)

We want to solve, here, \[
\min \left\{ \sum_{i=1}^{n} R^q_{\tau}(y_i - x_i^T \beta) \right\}
\]

\[
y_i = x_i^T \beta + \varepsilon_i \text{ so that } \hat{Q}_{y|x}(\tau) = x^T \hat{\beta} + F_{\varepsilon}^{-1}(\tau)
\]
Optimization Algorithm: Quantile Regression

Simplex algorithm to solve this program. Primal problem is

\[
\min_{\beta, u, v} \left\{ \tau 1^T u + (1 - \tau) 1^T v \right\} \quad \text{s.t.} \quad y = X \beta + u - v, \text{ with } u, v \in \mathbb{R}_+^n
\]

and the dual version is

\[
\max_d \left\{ y^T d \right\} \quad \text{s.t.} \quad X^T d = (1 - \tau) X^T 1 \text{ with } d \in [0, 1]^n
\]


Running time is of order \(n^{1+\delta}k^3\) for some \(\delta > 0\) and \(k = \dim(\beta)\)

(it is \((n + k)k^2\) for OLS, see wikipedia).
Quantile Regression Estimators

OLS estimator $\hat{\beta}^{\text{ols}}$ is solution of

$$\hat{\beta}^{\text{ols}} = \text{argmin} \left\{ \mathbb{E} \left[ (\mathbb{E}[Y|X = x] - x^T\beta)^2 \right] \right\}$$

and Angrist, Chernozhukov & Fernandez-Val (2006) Quantile Regression under Misspecification proved that

$$\hat{\beta}_\tau^q = \text{argmin} \left\{ \mathbb{E} \left[ \omega_\tau(\beta)(Q_\tau[Y|X = x] - x^T\beta)^2 \right] \right\}$$

(under weak conditions) where

$$\omega_\tau(\beta) = \int_0^1 (1 - u) f_{y|x}(u x^T\beta + (1 - u)Q_\tau[Y|X = x])du$$

$\hat{\beta}_\tau^q$ is the best weighted mean square approximation of the true quantile function, where the weights depend on an average of the conditional density of $Y$ over $x^T\beta$ and the true quantile regression function.
Quantile Regression Estimators

Under weak conditions, $\hat{\beta}^q_\tau$ is asymptotically normal:

$$\sqrt{n}(\hat{\beta}^q_\tau - \beta^q_\tau) \xrightarrow{D} \mathcal{N}(0, \tau(1-\tau)D^{-1}_\tau \Omega_x D^{-1}_\tau),$$

where

$$D_\tau = \mathbb{E}[f_\varepsilon(0)XX^T] \text{ and } \Omega_x = \mathbb{E}[X^TX].$$

hence, the asymptotic variance of $\hat{\beta}$ is

$$\hat{\text{Var}}[\hat{\beta}^q_\tau] = \frac{\tau(1-\tau)}{[f_\varepsilon(0)]^2} \left(\frac{1}{n} \sum_{i=1}^n x_i^T x_i\right)^{-1}$$

where $f_\varepsilon(0)$ is estimated using (e.g.) an histogram, as suggested in Powell (1991) Estimation of monotonic regression models under quantile restrictions, since

$$D_\tau = \lim_{h \downarrow 0} \mathbb{E}\left(\frac{1(|\varepsilon| \leq h)}{2h} XX^T\right) \sim \frac{1}{2nh} \sum_{i=1}^n 1(|\varepsilon_i| \leq h)x_i x_i^T = \hat{D}_\tau.$$
Visualization, $\tau \mapsto \hat{\beta}_\tau^q$

See Abreveya (2001) The effects of demographics and maternal behavior on the distribution of birth outcomes
Visualization, $\tau \mapsto \hat{\beta}^q_{\tau}$

Quantile Regression on Panel Data

In the context of panel data, consider some fixed effect, $\alpha_i$ so that

$$y_{i,t} = x_{i,t}^T \beta + \alpha_i + \varepsilon_{i,t} \text{ where } Q_{\tau}(\varepsilon_{i,t} | X_i) = 0$$

Canay (2011) A simple approach to quantile regression for panel data suggests an estimator in two steps,

- use a standard OLS fixed-effect model $y_{i,t} = x_{i,t}^T \beta + \alpha_i + u_{i,t}$, i.e. consider a within transformation, and derive the fixed effect estimate $\hat{\beta}$

$$ (y_{i,t} - \bar{y}_i) = (x_{i,t} - \bar{x}_{i,t})^T \beta + (u_{i,t} - \bar{u}_i) $$

- estimate fixed effects as $\hat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} (y_{i,t} - x_{i,t}^T \hat{\beta})$

- finally, run a standard quantile regression of $y_{i,t} - \hat{\alpha}_i$ on $x_{i,t}$’s.

See rqpd package.
Quantile Regression with Fixed Effects (QRFE)

In a panel linear regression model, \( y_{i,t} = x_{i,t}^T \beta + u_i + \varepsilon_{i,t} \), where \( u \) is an unobserved individual specific effect.

In a fixed effects models, \( u \) is treated as a parameter. Quantile Regression is

\[
\min_{\beta, u} \left\{ \sum_{i,t} R_q^{\tau}(y_{i,t} - [x_{i,t}^T \beta + u_i]) \right\}
\]

Consider Penalized QRFE, as in Koenker & Bilias (2001) Quantile regression for duration data,

\[
\min_{\beta_1, \ldots, \beta_\kappa, u} \left\{ \sum_{k,i,t} \omega_k R_{\tau_k}^{q_k}(y_{i,t} - [x_{i,t}^T \beta_k + u_i]) + \lambda \sum_i |u_i| \right\}
\]

where \( \omega_k \) is a relative weight associated with quantile of level \( \tau_k \).
Quantile Regression with Random Effects (QRRE)

Assume here that \( y_{i,t} = x_{i,t}^T \beta + u_i + \varepsilon_{i,t}. \)

Quantile Regression Random Effect (QRRE) yields solving

\[
\min_{\beta} \left\{ \sum_{i,t} R_q^\tau(y_{i,t} - x_{i,t}^T \beta) \right\}
\]

which is a weighted asymmetric least square deviation estimator.

Let \( \Sigma = [\sigma_{s,t}(\tau)] \) denote the matrix

\[
\sigma_{i,s}(\tau) = \begin{cases} 
\tau(1 - \tau) & \text{if } t = s \\
\mathbb{E}[1\{\varepsilon_{it}(\tau) < 0, \varepsilon_{is}(\tau) < 0\}] - \alpha^2 & \text{if } t \neq s
\end{cases}
\]

If \( (nT)^{-1} X^T \{I_n \otimes \Sigma T \times T(\tau)\} X \to D_0 \) as \( n \to \infty \) and \( (nT)^{-1} X^T \Omega_f X = D_1 \), then

\[
\sqrt{nT}(\hat{\beta}_q^\tau - \beta_q^\tau) \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, D_1^{-1} D_0 D_1^{-1}\right).
\]
Expectile Regression

We want to solve, here, \( \min_{\beta} \left\{ \sum_{i=1}^{n} R^e_{\tau}(y_i - x_i^T \beta) \right\} \)

see Koenker (2014) Living Beyond our Means for a comparison quantiles-expectiles
Expectile Regression

Solve here \( \min_{\beta} \left\{ \sum_{i=1}^{n} R^e_{\tau}(y_i - x_i^T \beta) \right\} \) where \( R^e_{\tau}(u) = u^2 \cdot (\tau - 1(u < 0)) \)

“this estimator can be interpreted as a maximum likelihood estimator when the disturbances arise from a normal distribution with unequal weight placed on positive and negative disturbances” Aigner, Amemiya & Poirier (1976) Formulation and Estimation of Stochastic Frontier Production Function Models.


Expectiles can (also) be related to Breckling & Chambers (1988) \( M \)-Quantiles.

Comparison quantile regression and expectile regression, see Schulze-Waltrup et al. (2014) Expectile and quantile regression - David and Goliath?
Visualization, $\tau \mapsto \hat{\beta}_\tau^e$

See Abreveya (2001) The effects of demographics and maternal behavior on the distribution of birth outcomes
Expectile Regression, with Random Effects (ERRE)

Quantile Regression Random Effect (QRRE) yields solving

$$\min_{\beta} \left\{ \sum_{i,t} R_{\alpha}^e(y_{i,t} - x_{i,t}^T \beta) \right\}$$

One can prove that

$$\hat{\beta}_{\tau}^e = \left( \sum_{i=1}^n \sum_{t=1}^T \hat{\omega}_{i,t}(\tau) x_{it} x_{it}^T \right)^{-1} \left( \sum_{i=1}^n \sum_{t=1}^T \hat{\omega}_{i,t}(\tau) x_{it} y_{it} \right),$$

where $$\hat{\omega}_{i,t}(\tau) = |\tau - 1(y_{it} < x_{it}^T \hat{\beta}_{\tau}^e)|.$$
Expectile Regression with Random Effects (ERRE)

If $W = \text{diag}(\omega_{11}(\tau), \ldots, \omega_{nT}(\tau))$, set

$$\overline{W} = \mathbb{E}(W), \quad H = X^T \overline{W} X \quad \text{and} \quad \Sigma = X^T \mathbb{E}(W \varepsilon \varepsilon^T W) X.$$ 

and then

$$\sqrt{nT} \{ \hat{\beta}_\tau^e - \beta_\tau^e \} \xrightarrow{\mathcal{L}} \mathcal{N}(0, H^{-1} \Sigma H^{-1}),$$

see Barry et al. (2016) Quantile and Expectile Regression for random effects model.
Application to Real Data

ERRE  QRRE

Intercept

Education

Experience

Experience^2

Time

ERRE  QRRE

Ability

Mother educ.

Father educ.

Broken home

Nbr of siblings