Session 1
Quantile, M-quantile & Expectile Regression

Nikos Tzavidis (University of Southampton)
Ray Chambers (University of Wollongong)
James Dawber(University of Wollongong)
Nicola Salvati (University of Pisa)

Pisa, July, 2016

Outline

- Defining robustness
- Estimating the centre (location) of a distribution
- Quantiles, M-quantiles & expectiles as location parameters
- Regression quantiles, M-quantiles & expectiles
- Asymmetric Laplace Distribution
- Quantile/M-quantile regression: A likelihood perspective
- Examples
- *R* Software

Defining Robustness

- Statistical inference is based on assumptions about the underlying distribution of the observations
- Although assumptions are never exactly true, some statistical models are more sensitive to small deviations from the assumptions than others
- An estimator is robust if it has the following features:
    - Reasonably efficient and unbiased
    - Small deviations from the model assumptions do not substantially impair the performance of the model
    - Somewhat larger deviations will not invalidate the model completely

Robust Regression

- Following Huber (1981) we will interpret robustness as insensitivity to small deviations from the assumptions the model imposes
- In particular, we are interested in distributional robustness
- In this context, robust refers to the shape of a distribution specifically, when it differs from the theoretically assumed distribution
- Although conceptually distinct, distributional robustness and outlier resistance are, for practical purposes, synonymous here

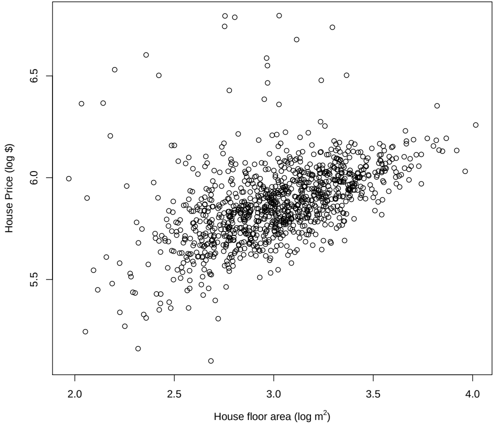
Estimating the Centre of a Distribution

- In order to explain how robust regression works it is helpful to start with the simple case of robust estimation of a parameter at the centre of the distribution
- Consider a set of independent observations y_i, $i = 1, ..., n$, and the linear model (special case of a GLM with identity link function)

$$y_i = \mu_i + \epsilon_i$$

- If the underlying distribution of ϵ is normal, the sample mean is the maximally efficient estimator.
- What if normality does not hold?

Estimating the Centre of a Distribution



$$\epsilon_i = y_i - \mu_i$$

Ordinary Least Squares (OLS)

- OLS popular estimation method $\rightarrow$ No assumption of normality
- OLS is not robust to outliers. It can produce misleading results if unusual cases go undetected. Even a single case can have a significant impact on the regression fit
- The efficiency of the OLS regression can be hindered by heavy-tailed distributions and outliers
- Residual diagnostics can be used. But, once they are found, what shall we do?
- Consider reformulating the model e.g. use transformations. If these do not help, consider using robust regression

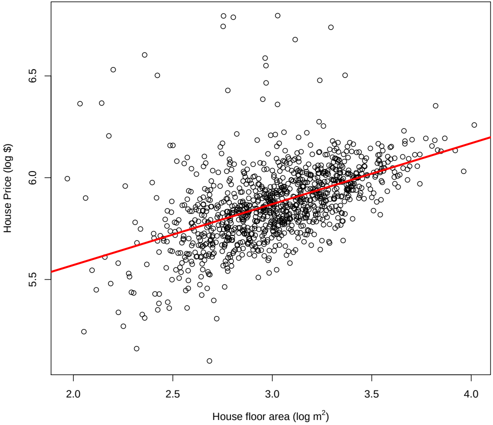
Estimating the Centre of a Distribution - OLS in Detail

- The mean, μ, is derived by minimising the least-squares objective function

$$min \sum_{i=1}^{n} \rho_{LS}(\epsilon_i) = \sum_{i=1}^{n} \rho_{LS}(y_i - \mu_i) = \sum_{i=1}^{n} (y_i - \mu_i)^2$$

- The derivative of the objective function with respect to ϵ gives the influence function which determines the influence of the observations
- In this case the influence is 2ϵ i.e. proportional to ϵ

Estimating the Centre of a Distribution



$\epsilon_i = y_i - \mu_i$

Mean

Minimise $\sum \rho(\epsilon_i)$:

$\rho(\epsilon_i) \propto \epsilon_i^2$

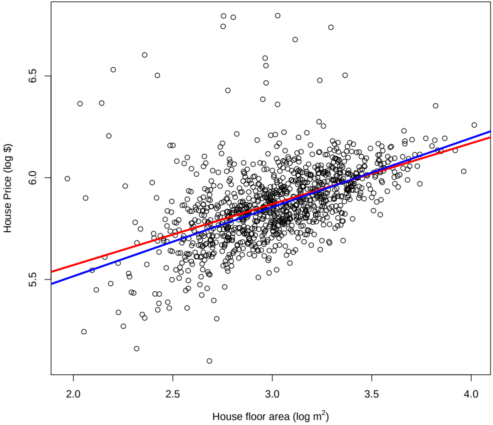
Estimating the Centre of a Distribution

- As an alternative to the mean we now consider the median, μ, as an estimator of the centre of a distribution
- The median is derived by minimising the least absolute-values (LAV) objective function

$$min \sum_{i=1}^{n} \rho_{LAV}(\epsilon_i) = \sum_{i=1}^{n} \rho_{LAV}(y_i - \mu_i) = \sum_{i=1}^{n} |y_i - \mu_i|$$

- Resistant to outliers: The influence of an unusual observation on the median is now bounded

Estimating the Centre of a Distribution



$\epsilon_i = y_i - \mu_i$

Mean

Minimise $\sum \rho(\epsilon_i)$:
$\rho(\epsilon_i) \propto \epsilon_i^2$

Median

Minimise $\sum \rho(\epsilon_i)$:
$\rho(\epsilon_i) \propto |\epsilon_i|$

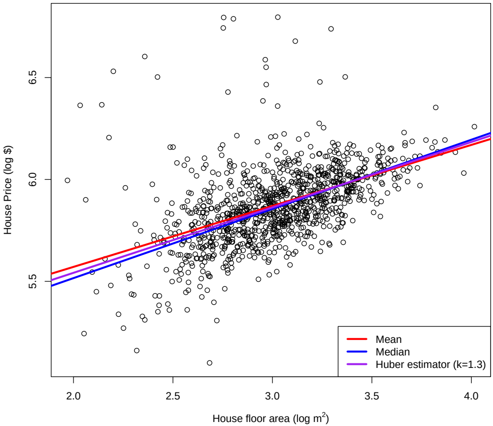
Estimating the Centre of a Distribution

- A good compromise between the efficiency of the least squares and the robustness of the least-absolute values estimators is the use of M-estimation
- An M-estimator for the centre of the distribution, μ, can be defined by using the Huber loss function

$$min \sum_{i=1}^{n} \rho_{Huber}(y_i - \mu_i)$$

- With $\rho_{Huber}(\epsilon_i) = \{ \begin{array}{ll} \frac{1}{2}\epsilon_i^2, & |\epsilon_i| \leq c \\ c|\epsilon_i| - \frac{1}{2}c^2, & |\epsilon_i| > c \end{array}$
- At the centre of the distribution the Huber function behaves like the ρ_{LS} loss, at the extremes it behaves like the ρ_{LAV} loss
- As $c \to 0$, robustness increases

Medians, means and something in between...



$$\epsilon_i = y_i - \mu_i$$

Mean

Minimise $\sum \rho(\epsilon_i)$:
$\rho(\epsilon_i) \propto \epsilon_i^2$

Median

Minimise $\sum \rho(\epsilon_i)$:
$\rho(\epsilon_i) \propto |\epsilon_i|$

M-estimator

Minimise $\sum \rho(\epsilon_i)$:

$$\rho(\epsilon_i) \propto \begin{cases} \epsilon_i^2, & \text{if } |\epsilon_i| < c \\ |\epsilon_i|, & \text{if } |\epsilon_i| \geq c \end{cases}$$

Re-cap: Defining the Centre (Location) of a Distribution

- Generally, a location parameter, μ, of $F(y)$ is defined by

$$min \int \rho\left(\frac{y-\mu}{\sigma}\right) F(dy)$$

- A natural estimator of μ, $\hat{\mu}$ is defined by

$$min \int \rho\left(\frac{y-\hat{\mu}}{\sigma}\right) \hat{F}(dy)$$

- σ scale for achieving scale invariance
- $\rho(\epsilon) = \epsilon^2 \rightarrow$ Mean
- $\rho(\epsilon) = |\epsilon| \rightarrow$ Median
- $\rho(\epsilon) = Huber \rightarrow$ M-estimator

M-Estimation: An applied viewpoint

$$min \sum_{i=1}^{n} \rho_{Huber}\left(\frac{y_i - \mu}{\sigma}\right)$$

- Denote the derivative of ρ_{Huber} by ψ_{Huber} and recall that $\epsilon_i = (\frac{y_i - \mu}{\sigma})$

$$\psi_{Huber} = \begin{cases} \epsilon_i, & \text{if } |\epsilon_i| < c \\ csgn(\epsilon_i), & \text{if } |\epsilon_i| \geq c \end{cases}$$

- Estimation equation to be solved

$$\sum_{i=1}^{n} \psi_{Huber}\left(\frac{y_i - \mu}{\sigma}\right) = 0$$

M-Estimation: An applied viewpoint

$$\sum_{i=1}^{n} \psi_{Huber}\left(\frac{y_i - \mu}{\sigma}\right) = 0$$

- Solved with iterative weighted least squares
- Define $w_i = \frac{\psi_{Huber}(\epsilon_i)}{\epsilon_i}$
- Huber weights

$$w_i = \begin{cases} 1, & \text{if} |\epsilon_i| < c \\ \frac{c}{|\epsilon_i|}, & \text{if} |\epsilon_i| \geq c \end{cases}$$

- Estimating equation

$$\sum_{i=1}^{n} w_i(y_i - \mu) = 0$$

M-Estimation: An applied viewpoint

$$\sum_{i=1}^{n} w_i(y_i - \mu) = 0$$

- Solution $\mu = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$
- But w_i depends on μ. Use iterative algorithm
- Start with initial value of $\mu^{(h)}$ and compute $w_i^{(h)}$
- Update $\mu \rightarrow \mu^{(h+1)} = \frac{\sum_{i=1}^{n} w_i^{(h)} y_i}{\sum_{i=1}^{n} w_i^{(h)}}$
- Solution when μ linear function of set of parameters β

$$\beta^{(h+1)} = (x^T w^{(h)} x)^{-1} x^T w^{(h)} y$$

- Iterate until convergence

Quantiles as Location Parameters

- The q quantile of $F(y)$,$\mu_y(q)$, can be defined as the 'location' of the transformed distribution that weights observations below $\mu_y(q)$ by $(1-q)$ and observations above $\mu_y(q)$ by q

Generalising Quantiles

Definition

Let $0 < q < 1$. The q location parameter of $F(y)$ corresponding to loss function ρ_q is the value $\mu_y(q)$ that satisfies

$$min \int \rho_q\left(\frac{y - \mu_y(q)}{\sigma_q}\right) F(dy),$$

- $\rho(\epsilon) = \epsilon^2 \rightarrow$ Expectiles; $\rho(\epsilon) = |\epsilon| \rightarrow$ Quantiles
- $\rho(\epsilon) = Huber \rightarrow$ M-quantiles

M-quantiles in detail (Breckling & Chambers,1988, Biometrika)

- Define by ψ_q the derivative of the Huber loss function
- Huber proposal 2 (influence function)

$$\psi_q(\epsilon) = \begin{cases} -(1-q)c & \epsilon < -c \\ (1-q)\epsilon & -c \leqslant \epsilon < 0 \\ q\epsilon & 0 \leqslant \epsilon < c \\ qc & c \leqslant \epsilon \end{cases}$$

$$\int \psi_q\left(\frac{y-\mu_y(q)}{\sigma_q}\right) F(dy) = 0$$

- Good compromise between efficiency and robustness

Regression Quantiles

(Koenker & Bassett, 1978, Econometrica; Koenker & Hallock JEP, 2001)

- Extend the idea to quantiles of conditional distributions
- Let $F(y|x)$ denote the distribution of y given x. The q regression quantile of y at $x = x$ is then $\mu_y(q|x = x)$, where $F(\mu_y(q|x = x)) = q$

A Linear Model for Regression Quantiles

$$\mu_y(q|x) = x^T \beta_q$$

- Computation: Koenker and D'Orey (1987)
- *R: quantreg* library
- *Stata: qreg* command
- Estimation of standard errors via bootstrap
- A number of bootstrap options available in *quantreg*

Quantile Regression - A Likelihood Perspective (Yu & Moyeed, 2001, Statistics and Probability Letters)

Quantile Regression - A Likelihood Perspective

- Minimization of the ρ_{LAV} is equivalent to the maximization of a likelihood function formed by combining independently distributed Asymmetric Laplace densities

$$y \sim \ ALD(\mu, \sigma, q),$$

with pdf

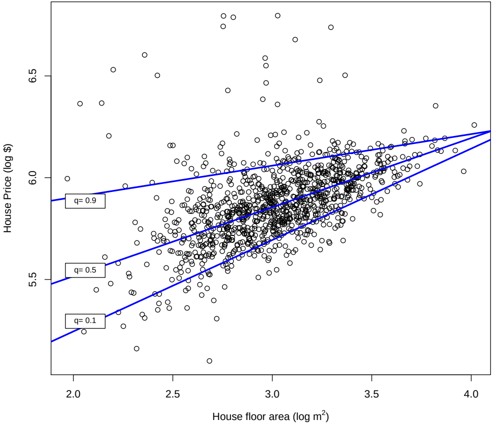
$$f(y|\mu, \sigma, q) = \frac{q(1-q)}{\sigma} exp\left(-\rho_q\left(\frac{y-\mu}{\sigma}\right)\right)$$

- For fixed q the location parameter μ is modelled as a function of x.
- $\mu = x^T \beta_q$

Quantiles

$$\rho_q(\epsilon) = \left[(1-q)\mathbb{I}_{(\epsilon \leq 0)} + q\mathbb{I}_{(\epsilon > 0)}\right] \rho(\epsilon)$$

$$\epsilon = y - \mu$$



Quantiles

Minimise $\sum \rho_q(\epsilon)$
where $\rho(\epsilon) \propto |\epsilon|$

Regression M-Quantiles

For fixed q and influence function ψ given by Huber's proposal 2, compute $\hat{\beta}(q)$ by

$$\sum_{i=1}^{n} \psi_q(\epsilon_{iq}) x_i = 0$$

- Residuals: $\epsilon_{iq} = (y_i - \mu_y(q|x_i;\psi_q))/\sigma_q$
- $\sigma_q \rightarrow$ scale parameter
- $\sigma_q = median|\epsilon_{iq}|/0.6745$
- $\psi_q(\epsilon_{iq}) = 2\psi(\epsilon_{iq})\{q\mathbb{I}(\epsilon_{iq} > 0) + (1-q)\mathbb{I}(\epsilon_{iq} \leqslant 0)\}$

Regression M-Quantiles - Estimation

(Chambers and Tzavidis, 2006, Biometrika)

- Fitting: Use iteratively re-weighted least squares (IRLS)
- Function QRLM in R available from the authors
- Estimation of standard errors (Bianchi & Salvati, Comm. in Stat., 2014)

$$\widehat{Var}(\hat{\beta}_q) = \hat{\sigma}_q \frac{\hat{E}(\psi_q^2(\epsilon))}{\hat{E}(\psi_q'(\epsilon))^2}(x^T x)^{-1}$$

M-Quantile Regression - A Likelihood Perspective (Bianchi, Fabrizi, Salvati & Tzavidis, 2014)

M-Quantile Regression - A Likelihood Perspective

- Minimization of the ρ_{Huber} is equivalent to the maximization of a likelihood function formed by combining independently distributed Asymmetric Least Information densities (ALID)

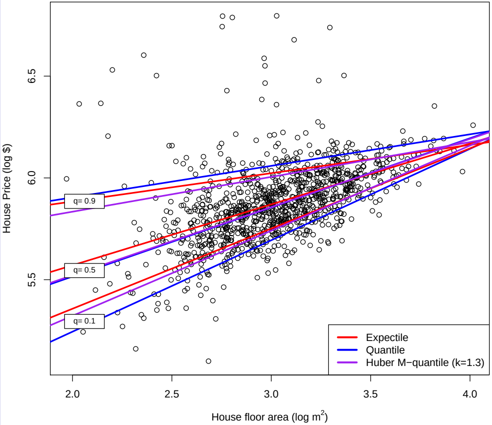
$$y \sim \ ALID(\mu, \sigma, q, c),$$

$$f(y|\mu, \sigma, q, c) = \frac{1}{\sigma B} exp\left[-\rho_{Huber}\left(\frac{y-\mu}{\sigma}\right)\right],$$

- The location parameter (M-quantiles)μ are modelled by using x, $\mu = x^T \beta_q$

Quantiles, expectiles and something in between...

$$\rho_q(\epsilon) = \left[(1-q)\mathbb{I}_{(\epsilon \leq 0)} + q\mathbb{I}_{(\epsilon > 0)}\right] \rho(\epsilon) \quad \epsilon = y - \mu$$



Quantile

Minimise $\sum \rho_q(\epsilon)$ where
$\rho(\epsilon) \propto |\epsilon|$

Expectile

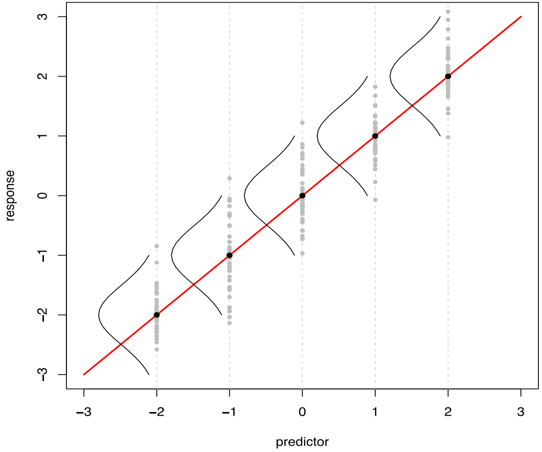
Minimise $\sum \rho_q(\epsilon)$ where
$\rho(\epsilon) \propto \epsilon^2$

Huber M-quantile

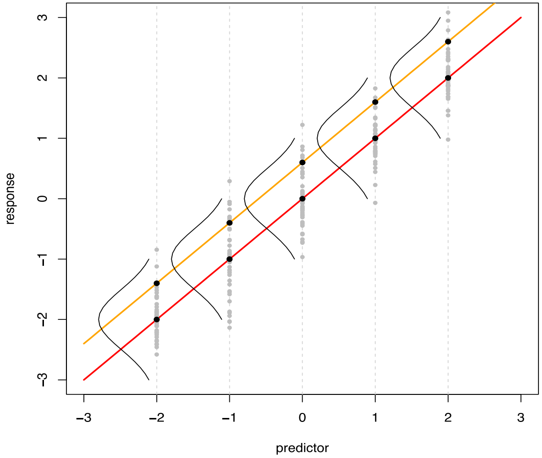
Minimise $\sum \rho_q(\epsilon)$ where

$$\rho(\epsilon) \propto \begin{cases} \epsilon^2, & \text{if } |\epsilon| < c \\ |\epsilon|, & \text{if } |\epsilon| \geq c \end{cases}$$

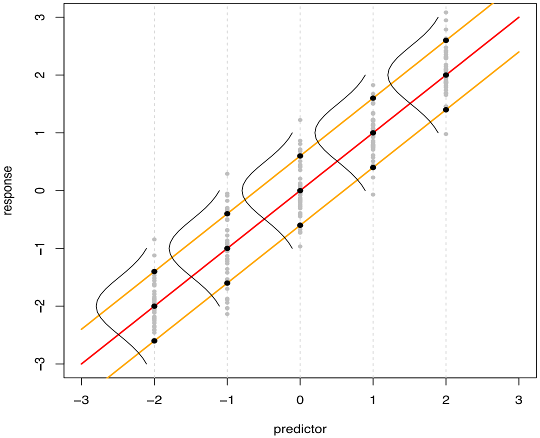
Quantile, M-quantile & Expectile Regression



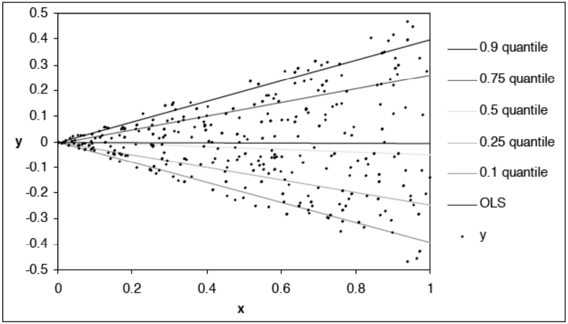
Quantile, M-quantile & Expectile Regression



Quantile, M-quantile & Expectile Regression

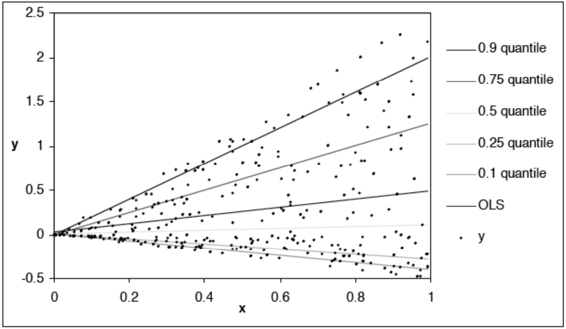


Heteroscedasticity & Symmetry



if errors are **symmetric**, then $\beta_q = -\beta_{1-q} \; \forall q < 0.5$

Heteroscedasticity & Asymmetry



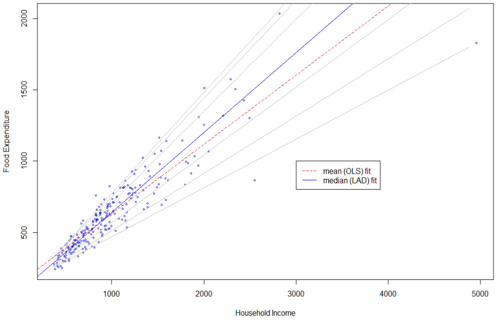
More on Graphical Diagnostics

- Spacing of quantile lines important
- Larger spacing $\Rightarrow$ Longer tail
- Shorter spacing $\Rightarrow$ Shorter tail

An Example

- Data on household income and food expenditure (Koenker & Hallock, JEP, 2001)
- Data from 235 European working-class households
- Plot presents estimated quantile regression lines corresponding to the quantiles 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95 alongside the median and mean fits

Quantile Regression: An Example



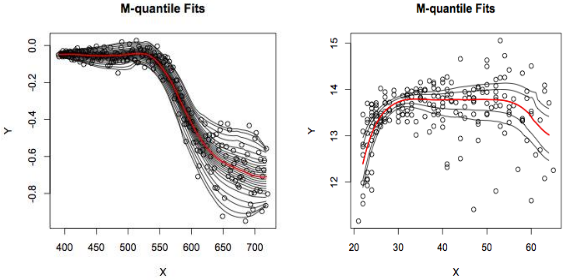
Quantile Regression: Comments on the Example - Koenker & Hallock, JEP, 2001

- Dispersion of food expenditure increases as household income increases
- Spacing of the quantile regression lines shows that the conditional distribution of household food expenditure, given household income, is skewed to the left
- Narrower spacing of the upper quantiles showing a short upper tail
- Wider spacing of the lower quantiles indicating a longer lower tail

Extensions

- Until now we have assumed a linear model for quantiles/M-quantiles
- The scope of quantile/M-quantile regression can be significantly extented by allowing for more complex model specifications
- Non-parametric quantile/M-quantile regression

Non-parametric Quantile / M-quantile Regression



Regression M-quantiles Vs Regression Quantiles

- M-quantile estimation via IWLS. Easier to estimate / convergence
- M-quantile regression allows for more flexibility in modelling. For example, the tuning constant c used in the Huber influence function
- c can be used to trade outlier robustness for efficiency (M-quantile Vs. expectile regression)
- $c \to 0$ quantile regression
- $c \to \infty$ expectile regression
- However, M-quantiles are harder to interpret. Useful for prediction purposes (e.g. small area estimation)

R Packages

Table: Quantile, M-quantile & Expectile Regression

Regression Type	Target Parameter	*R* Package
OLS	Mean	*lm*
Robust	Median	*rlm*,(*MASS*)
Quantile	Quantiles	*rq*,(*quantreg*)
Quantile (ALD)	Quantiles	*lqm*, (*lqmm*)
Bayesian Quantile (ALD)	Quantiles	(*bayesQR*)
M-Quantile, Expectile	M-quantiles/Expectiles	[1]
Expectile	Expectiles	*amlnormal*,[2]

[1] Available from the presenters

[2] *VGAM*