M-quantile regression in R

Maciej Beręsewicz, Nicola Salvati, Stefano Marchetti

Poznan, 13.10.2016

Poznań University of Economics and Business

Table of contents

M-quantile models

Basic assumptions

- The classical regression model summarizes the behaviour of the mean of a random variable at each point in a set of covariates.
- Quantile regression summarizes the behaviour of different parts of the conditional distribution $f(y|\boldsymbol{x})$ at each point in the set of the $\boldsymbol{x}$'s.
- Each sample unit lie on one and only one regression M-quantile line. The M-quantile of order q for the conditional density of y given $\boldsymbol{x}$.

M-quantile models – notation

Notation

- $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$ – denote the values of a random sample consisting of n units.
- $\boldsymbol{x}_i$ – is the i-th row of a known design matrix $\boldsymbol{X}$
- y_i – is a scalar response variable corresponding to a realization of a continuous random variable with unknown continuous cumulative distribution function.
- In case of small area each sample unit $i \in d$ will lie on one and only one regression M-quantile line. The q-index of this line is the M-quantile coefficient q_{di} of sample unit i.

M-quantile models

A linear regression model for the q conditional quantile of $f(y|\boldsymbol{x})$ is:

$$Q_y(q|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\beta}(q)\,. \tag{1}$$

Estimates of $\boldsymbol{\beta}(q)$ are obtained by minimizing the following function:

$$\sum_{i=1}^{n} |y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}(q)|\{qI(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}(q) > 0) + (1-q)I(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}(q) \leq 0)\}. \tag{2}$$

M-quantile models

The M-quantile regression is a "quantile-like" generalization of regression based on influence functions (M-regression). The M-quantile q of the conditional density $f(y|\boldsymbol{x})$, denoted by m, is defined as the solution to the estimating equation:

$$\int \psi_q(y-m) f(y|\boldsymbol{x})\, dy = 0\,, \tag{3}$$

where ψ_q is an asymmetric influence function that is the derivative of an asymmetric loss function, ρ_q. When a linear relation between M-quantile m and auxiliary variables holds, then the M-quantile regression model is:

$$m_y(q|\boldsymbol{X}) = \boldsymbol{X}\beta_\psi(q)\,. \tag{4}$$

Estimation of M-quantile models parameters

An estimate of $\boldsymbol{\beta}_\psi(q)$ is obtained by minimizing:

$$\sum_{i=1}^{n} \rho_q(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\psi(q)). \tag{5}$$

The minimization of (5) reduces to the following estimating equation:

$$\sum_{i=1}^{n} 2\psi_q(s^{-1}(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_\psi(q))) = 0, \tag{6}$$

where ψ_q is the derivative of the Huber loss function ρ_q, and it is equal to:

$$\psi_q(u) = \begin{cases} q\psi(u) & \text{if } u > 0 \\ (1-q)\psi(u) & \text{if } u \leq 0, \end{cases} \tag{7}$$

with $\psi(u) = uI(|u| \leq c) + \text{sgn}(u)cI(|u| > c)$ and c is a tuning constant. Provided that the tuning constant c is strictly greater than zero, estimates of $\boldsymbol{\beta}_\psi(q)$ are obtained using iterative weighted least squares (IWLS).

Available M-quantile models

- the estimation of the small area distribution function using a non-parametric specification of the conditional MQ of the response variable given the covariates. In particular splies were considered to take into account nonparametric (functional) relation between target variable and given auxiliary variable.
- robust prediction of small area means and distributions

Available M-quantile models

- spatial version of the M-quantile models for small area estimation, for instance Geographically Weighted M-quantile model:

$$Q_y(q|\boldsymbol{X}) = \boldsymbol{X}\beta(u; q), \tag{8}$$

where u denotes location given by spatial coordinates (longitude, latitude). In such setting β vary with u and q and allows for different set of β for each point. Such model is also often refer as spatial non-stationary M-quantile model. Another approach is model that assume global β and local intercept $\delta(u, q)$ that vary across locations:

$$Q_y(q|\boldsymbol{X}) = \boldsymbol{X}\beta(u; q) + \delta(u, q) \tag{9}$$

Available M-quantile models

- constrained M-quantile small area estimators for benchmarking and for the correction of the under/over-shrinkage of small area estimators
- a model-assisted approach and design consistent small area estimators based on the M-quantile small area model

$$\bar{y}_d^{WMQ} = N_d^{-1} \sum_{j \in s_d} w_{dj} y_{dj} + \left(N_d^{-1} \sum_{j \in U_d} \boldsymbol{x}'_{dj} - N_d^{-1} \sum_{j \in U_d} w_{dj} \boldsymbol{x}'_{dj} \right) \widehat{\boldsymbol{\beta}}_{wq_i} \tag{10}$$

Available M-quantile models

- Three-level M-quantile mixed model (Beręsewicz, M., Marchetti, S., Salvati, N., Szymkowiak, M. & Wawrowski, Ł.)
- Logistic M-quantile (mixed) model,
- Poisson M-quantile (mixed) model,
- Semiparametric Poisson M-quantile (mixed) model,
- Disease Mapping via Negative Binomial Regression M-quantiles.

Small area estimation – what is that?

- Given the growing demand for information about poverty indicators at lower levels than NUTS 1 in Poland, there is a pressing need to take advantage of appropriate SAE techniques and data from different statistical sources (EU-SILC, census or administrative registers).
- The SAE methodology has been developed to produce reliable estimates of different characteristics of interest, such as means, counts, quantiles or ratios for domains for which only small samples are available (Rao and Molina, 2016)
- Indirect methods for poverty mapping are based on models that take into account between area variation beyond what is explained by auxiliary variables

M-quanitle for small areas

- Chambers and Tzavidis (2006) extended the use of M-quantile regression models to small area estimation.
- They characterize the conditional variability across the population of interest by the M-quantile coefficients of the population units. Using the M-quantile coefficients it is possible to define an M-quantile small area model:

$$y_{jd} = \mathbf{x}_{jd}^T \beta_\psi(\theta_d) + \epsilon_{jd} \,, \tag{11}$$

where $\beta_\psi(\theta_d)$ is the unknown vector of M-quantile regression parameters for the unknown area-specific M-quantile coefficient θ_d and ϵ_{jd} is the unit level random error term with distribution function G for which no explicit parametric assumptions are being made.

M-quanitle for small areas

- The area-specific M-quantile coefficient θ_d is estimated averaging the M-quantile coefficients of the sample units in area d, so $\hat{\theta}_d = n_d^{-1} \sum_{j=1}^{n_d} \theta_j$.
- Then, $\beta_\psi(\hat{\theta}_d)$ is estimated solving equation (5).
- The predicted values under the M-quantile small area model are $\hat{y}_{jd} = \mathbf{x}_{jd}^T \hat{\beta}_\psi(\hat{\theta}_d)$.

Currently available R codes

These codes were mainly developed by Nicola Salvati, Stefano Marchetti, Nikos Tzavidis and Timo Schmid

- Linear M-quantile regression – `MQRE`;
- Linear M-quantile random effect regression (two-level model) – `mq_function`; a three-level model proposed in Beresewicz, Marchetti, Salvati, Szymkowiak & Wawrowski (submitted);
- Semiparametric M-quantile random effect regression – `npqrlm`;
- Geographically weighted M-quantile random effect regression (only SAE) – `mqgwr.sae`;
- Binomial M-quantile regression – `glm.mq.binom`
- Poisson M-quantile regression – `glm.mq.poisson`

Github repository

- LINK: `DepartmentOfStatisticsPUE/m-quantile-models-codes`
- Structure:
    - Sample project – codes from SAMPLE project developed
    - MQRE
    - BinaryMQ

Instalation of packages

MQRE – M-quantile random effect model

```
devtools::install_github(
'DepartmentOfStatisticsPUE/m-quantile-models-codes',
subdir = 'MQRE')
```

BinaryMQ – Binomial M-quantile model

```
devtools::install_github(
'DepartmentOfStatisticsPUE/m-quantile-models-codes',
subdir = 'BinaryMQ')
```

Bibliography (selected)

- Beręsewicz, M., Marchetti, S., Salvati, N., Szymkowiak, M. & Wawrowski, Ł., The use of a three level M-quantile model to map poverty at LAU 1 in Poland
- Chambers, R., & Tzavidis, N. (2006). M -quantile models for small area estimation. Biometrica, 93(2), 255–268.
- Chambers, R., Chandra, H., Salvati, N., & Tzavidis, N. (2014). Outlier robust small area estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(1), 47-69.
- Pratesi, M., Ranalli, M. G., & Salvati, N. (2009). Nonparametric M-quantile regression using penalised splines. Journal of Nonparametric Statistics, 21(3), 287-304.
- Salvati, N., Tzavidis, N., Pratesi, M., & Chambers, R. (2012). Small area estimation via M-quantile geographically weighted regression. Test, 21(1), 1-28.