M-QUANTILE REGRESSION: DIAGNOSTICS AND PARAMETRIC REPRESENTATION OF THE MODEL

Annamaria Bianchi, Enrico Fabrizi, Nicola Salvati and Nikos Tzavidis

ABSTRACT: M-quantile regression generalizes both quantile and expectile regression using M-estimation ideas. This paper covers several topics related to estimation, model assessment and hypothesis testing that were so far neglected in the many articles about M-quantile regression methods that appeared in recent years.

KEYWORDS: M-quantile regression, small area estimation, cluster test, likelihood ratio-type test.

1 Introduction

Quantile regression (Koenker and Bassett, 1978) represents a useful generalization of median regression whenever the interest is not limited to the description of the average relationship between a target variable and its predictors, but it encompasses the local behaviour of this relationship at different quantiles of the distribution. Similarly, expectile regression (Newey and Powell, 1987) generalizes mean regression by modelling expectiles of the dependent variable distribution.

Breckling & Chambers, 1988 introduce M-quantile (MQ) regression that uses the ideas of M-regression (Huber, 1973) to model the relationship between the dependent variable and its predictors for various M-quantiles of the distribution. Depending on the choice of the loss function, M-quantiles may reduce to ordinary quantiles and expectiles.

Chambers & Tzavidis, 2006 apply M-quantile regression to small area estimation. The distinguishing features of their approach to this application are its being robust and distribution free. Since then, a number of papers on M-quantile regression applied to small areas has been published (Bianchi and
Salvati, 2014, Chambers et al., 2014, Fabrizi et al., 2014). Nonetheless little attention has been paid so far to the assessment of goodness-of-fit and other inferential procedures.

The main objective of this paper is to fill this gap by introducing a pseudo-$R^2$ diagnostic, likelihood-ratio and Wald type tests for linear hypotheses on the regression parameters. We also introduce a parametric representation of the model, which allows us to introduce an estimator of the scale parameter and, in the case of the Huber proposal 2 influence function, of the loss function’s tuning constant. Finally, we introduce a test for the presence of cluster effects in the data. The finite sample properties of the proposed methods are assessed by means of simulation studies.

2 M-quantile regression

Let $y$ be a random variable and $x$ a $p$-dimensional random vector with first component $x_1 = 1$. The observable data $\{(x_i, y_i), i = 1, \ldots, n\}$ are assumed to be a random sample of size $n$ drawn from the population; thus, they are independent and identically distributed random variables. Assuming a linear model, for any $\tau \in (0, 1)$, the M-quantile of order $\tau$ of $y_i$ given $x_i$ is defined by $MQ_\tau(y_i|x_i) = x_i^T \hat{\beta}_\tau$. $\hat{\beta}_\tau \in \Theta \subset \mathbb{R}^p$ is the solution to the population minimization problem

$$\min_{\hat{\beta} \in \Theta} E \left[ \rho_\tau \left( \frac{y_i - x_i^T \hat{\beta}}{\sigma_\tau} \right) \right],$$

where $\rho_\tau(u) = |\tau - I(u < 0)| \rho(u)$ is a continuously differentiable loss function, and $\sigma_\tau$ is a scale parameter that characterizes the distribution of $\varepsilon_i = y_i - x_i^T \beta_\tau$. Since $\rho$ is continuously differentiable and convex, the MQ regression coefficient estimator $\hat{\beta}_\tau$ can be obtained as the solution of the system of equations $\sum_{i=1}^n \psi_\tau \left( \frac{y_i - x_i^T \hat{\beta}}{\sigma_\tau} \right) x_i = 0$, where $\psi_\tau(u) = d\rho_\tau(u)/du = |\tau - I(u < 0)| \psi(u)$, with $\psi(u) = d\rho(u)/du$, and $\hat{\sigma}_\tau$ is an estimator of $\sigma_\tau$. An iterative method is needed here to obtain a solution.

3 Diagnostics

The MQ-model can be partitioned as $MQ_\tau(y_i|x_i) = x_i^T \hat{\beta}_{\tau 1} + x_i^T \hat{\beta}_{\tau 2}$, where $\hat{\beta}_\tau = (\hat{\beta}_{\tau 1}, \hat{\beta}_{\tau 2})^T$, $\hat{\beta}_{\tau 1}$ is a $(p - k) \times 1$ vector and $\hat{\beta}_{\tau 2}$ is a $k \times 1$ $(0 < k < p)$ vector. Interest focuses on tests of the null hypothesis $H_0 : \hat{\beta}_{\tau 2} = \mathbf{0}$. Let $\hat{\beta}_\tau$ denote the MQ-estimator of the full model and let $\hat{\beta}_\tau = (\hat{\beta}_{\tau 1}, \mathbf{0}^T)^T$ denote the MQ-estimator under $H_0$. 
When \( \beta_{\tau} \) restricts to include only the intercept, a goodness-of-fit measure may be defined as \( R^2_\tau = 1 - \sum_{i=1}^{n} \rho_\tau \left( \frac{y_i - x_i^T \hat{\beta}_\tau}{\hat{\sigma}_\tau} \right) / \sum_{i=1}^{n} \rho_\tau \left( \frac{y_i - x_i^T \tilde{\beta}_\tau}{\tilde{\sigma}_\tau} \right) \).

This measure is the natural analog of the well-known \( R^2 \) used in ordinary least squares. It varies between 0 and 1 and it represents a local measure of goodness-of-fit for a specified MQ regression (at a fixed \( \tau \)). Under assumptions given in Bianchi & Salvati, 2015, we provide (asymptotically consistent) likelihood ratio-type and Wald test statistics for testing \( H_0 \).

4 Parametric representation

We introduce a useful parametrization for M-quantile regression. Given a loss function \( \rho_\tau \), we can define a random variable with density function

\[
f_\tau(y; \mu_\tau, \sigma_\tau) = \frac{1}{\sigma_\tau B_\tau} \exp \left[ -\rho_\tau \left( \frac{y - \mu_\tau}{\sigma_\tau} \right) \right], \quad -\infty < y < +\infty, \tag{1}
\]

where \( B_\tau = \int_{-\infty}^{+\infty} \exp \left[ -\rho_\tau (y) \right] dy < +\infty \). We show that \( \mu_\tau \) represents the M-quantile of the distribution and that under distribution (1) and when \( \mu_{\tau i} = MQ_\tau(y_i | x_i) = x_i^T \hat{\beta}_\tau \), the estimates of the unknown regression parameters \( \hat{\beta}_\tau \) and the scale \( \sigma_\tau \) may be obtained by maximising the corresponding log-likelihood function. In the case of the Huber proposal 2 loss function, the density \( f_\tau \) is called Asymmetric Least Information (ALI) distribution. The ALI distribution depends on the tuning constant \( c \). We propose to interpret \( c \) as a parameter of the density \( f_\tau \) and estimate \( c \) together with \( \hat{\beta}_\tau \) and \( \sigma_\tau \) by maximising the log-likelihood function. A similar approach can be applied to special other cases associated to different loss functions and tuning constants.

5 LR-type test for the presence of clustering effects

Cluster effects are typical of the analysis of hierarchical data resulting from sampling in two or more stages. The observable data can then be denoted as \( \{(x_{ij}, y_{ij}), i = 1, \ldots, n_j; j = 1, \ldots, d\} \), where \( d \) is the number of the primary units drawn from the population and \( n_j > 0 \) is the number of secondary (individual) units drawn from each primary unit. In this context, individual level covariates can fail to account for the specificity of units belonging to the same primary unit. Linear mixed models are a popular tool for the analysis of this type of data, in which the clustering is modelled by means of random effects.
Here we propose a new approach based on M-quantile regression and the characterization of clustering of observations first introduced by Chambers & Tzavidis, 2006. The definition of group-specific MQ-coefficients \( \tau = (\tau_1, \ldots, \tau_D)^T \) differs from that proposed by Chambers & Tzavidis, 2006. Within each group, \( \tau_j \) is assumed to uniquely solve

\[
\min_{\tau} E \left[ \rho \left( \frac{y_{ij} - x_{ij}^T \beta_\tau}{\sigma} \right) \bigg| j \right],
\]

i.e. \( \tau_j \) is the minimizer of the specified objective function within group \( j \).

For MQ models, testing for the presence of clustering effects is equivalent to testing whether the group-specific MQ-coefficients are all equal, that is \( H_0 : \tau_j = 0.5 \forall j = 1, \ldots, D \) against \( H_A : \tau_j \neq 0.5 \) for at least one \( j \). The value 0.5 corresponds to the global minimizer. The vector of estimated MQ-coefficients \( \hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_D)^T \) is defined as the empirical counterpart of (2). Based on this estimator, we propose an asymptotic test for testing \( H_0 \) based on the \( \chi^2 \) distribution.

References


