Nonparametric M-quantile Regression via Penalized Splines

Monica Pratesi[1], M. Giovanna Ranalli[2], Nicola Salvati[1]
Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Italy[1]
Dipartimento di Economia, Finanza e Statistica, Università degli Studi di Perugia, Italy[2] *giovanna@stat.unipg.it*

Abstract

Quantile regression investigates the conditional quantile functions of a response variables in terms of a set of covariates. M-quantile regression extends this idea by a "quantile-like" generalization of regression based on influence functions. In this work we extend it to nonparametric regression, in the sense that the M-quantile regression functions do not have to be assumed to be linear, but can be left undefined and estimated from the data. Penalized splines are employed to estimate them. This choice makes it easy to move to bivariate smoothing and additive modeling. An algorithm based on penalized iteratively reweighted least squares to actually fit the model is also proposed. Simulation studies are presented that show the finite sample properties of the proposed estimation technique.

Keywords: Robust regression, Iteratively Reweighted Least Squares, Nonparametric smoothing.

1 Introduction

Regression analysis is a standard tool for modeling the relationship between a response variable y and some covariates $\boldsymbol{x}$. It summaries the average behavior of y given $\boldsymbol{x}$ and has been one of the most important statistical methods for applied research for many decades. However, in some circumstances the mean does not give a complete picture of the distribution. It does not consider, for example, the extreme behavior of y conditional on $\boldsymbol{x}$. For this reason, a method that allows for direct modeling of the relationship between the dependent variable and the explanatory variables for these extreme values is needed. In other words, it may be useful to investigate the conditional quantile functions. Such a modeling exercise is referred to as *quantile regression*. M-quantile regression extends this idea by a "quantile-like" generalization of regression based on influence functions (Breckling and Chambers, 1988). For a specified q, in a linear M-quantile regression model $Q_q(\boldsymbol{x}, \psi) = \boldsymbol{x}\boldsymbol{\beta}_\psi(q)$, where ψ denotes the influence function associated with the M-quantile.

Nonparametric smoothing has been usefully applied to quantile regression (see e.g. He, 1997; Takeuchi, Le, Sears and Smola, 2005), but little or no work has been done on extending M-quantile regression with nonparametric modeling. Here we will do so by using Penalized Splines.

The outline of the paper is the following. Section 2 briefly review the M-quantile regression. Section 3 is devoted to Penalized Spline M-quantile regression. Section 4 stretches the simulation study and its results. The attention is on the performance of our method when a single covariate model expresses the true underlying relationship between y and x information, and especially when a bivariate model is considered. This second case is meanly relevant to test the empirical properties of the method when the study variable has a clear spatial pattern as a function of its position in space represented by its geographical coordinates. Section 5 presents and discusses our main findings.

2 M-quantile regression

Quantile regression is a generalization of median regression and has been developed by Koenker and Bassett (1978). In the linear case, quantile regression leads to a family of hyperplanes indexed by the value of the corresponding quantile coefficient $q \in (0, 1)$. For each value of q, the corresponding model $Q_q(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}(q)$ explains how the q^{th} quantile of the conditional distribution of y given $\boldsymbol{x}$ varies with $\boldsymbol{x}$.

The set of regression quantiles parameter estimates satisfies the criterion of minimum sum of absolute asymmetrically weighted residuals: given a sample of n observations, the vector $\boldsymbol{\beta}(q)$ is estimated by minimizing

$$\sum_{i=1}^{n} |r_i[\boldsymbol{\beta}(q)]|\{(1-q)I(r_i[\boldsymbol{\beta}(q)] \leq 0) + qI(r_i[\boldsymbol{\beta}(q)) > 0]\}, \tag{1}$$

where $r_i[\boldsymbol{\beta}(q)] = y_i - \boldsymbol{x}_i\boldsymbol{\beta}(q)$, with respect to $\boldsymbol{\beta}(q)$ by using linear programming methods (Koenker and D'Orey, 1987). However, regression quantile hyper-planes are not comparable with the regression ones based on ordinary least-squares that describe how the mean of y changes with $\boldsymbol{x}$ (Breckling and Chambers, 1988). In fact, the former are based on an absolute deviations criterion, while the latter on a least-squares one.

A generalization of expectation was suggested by Newey and Powell (1987) through the use of *expectile* lines. M-quantile regression extends this idea by a "quantile-like" generalization of regression based on influence functions (Breckling and Chambers, 1988). For a specified q, in a linear M-quantile regression model $Q_q(\boldsymbol{x}, \psi) = \boldsymbol{x}\boldsymbol{\beta}_\psi(q)$, where ψ denotes the influence function associated with the M-quantile. In particular, the general M-estimator of $\boldsymbol{\beta}_\psi(q)$ can be obtained by solving the set of estimating equations:

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

with respect to $\boldsymbol{\beta}_\psi(q)$, assuming that

$$\psi_q(t) = 2\psi\{s^{-1}(t)\}\{(1-q)I(t \leq 0) + qI(t > 0)\}$$

where s is a robust estimate of scale. Robust regression models can be fitted using an Iterative Reweighted Least Squares

algorithm (IRLS) that guarantees the convergence to a unique solution (Kokic et al., 1997).

The advantages of M-quantile regression models are (a) the simplicity of the algorithm used to fit the model and (b) the great flexibility in modeling by using a wide range of influence functions (i.e. Huber function, Huber proposal 2, Hampel function). A drawback for all quantile-type fitted regression plans is the phenomenon of quantile crossing and it is due to model misspecification, collinearity or huge outlying values. He (1997) proposed a restricted version of regression quantile that avoids the occurrence of crossing while maintaining sufficient modeling flexibility. Another method to overcome this problem is described in Koenker (1984) by forcing proper ordering of the percentile curves. The author considered parallel quantile planes for linear models, but they do not cater to the needs of heteroscedastic models.

3 Penalized Spline M-quantile Regression

3.1 The method

Nonparametric regression is a popular technique that extends linear regression by relaxing the assumption of a pre-specified functional relationship between the mean value of y and the covariates $\boldsymbol{x}$. Such relationship does not have to be assumed linear or polynomial, but only an unknown smooth function. Techniques like Kernels, Local polynomials or Smoothing Splines can then be used to learn this function from the data. Nonparametric regression is a well-established branch of statistical modeling and a wide choice of books and monographs are available; Hastie, Tibshirani and Friedman (2001) is a good introductory guide.

Smoothing has been usefully applied to quantile regression (see e.g. He, 1997; Takeuchi et al., 2005), but little or no work has been done on extending M-quantile regression with nonparametric modeling. Here we will do so by using Penalized Splines. Penalized splines are now often referred to as p-splines and have been recently brought up to attention by Eilers and Marx (1996). P-splines provide an attractive smoothing method for their simplicity of implementation, being a relatively straightforward extension of linear regression, and flexibility to be incorporated in a wide range of modeling contexts. Ruppert, Wand and Carroll (2003) provide a thorough treatment of p-splines and their applications.

Let us first consider only smoothing with one covariate x_1; we will then move to bivariate smoothing and semiparametric modeling. Given an influence function ψ, a nonparametric model for the q^{th} quantile can be written as $Q_q(x_1, \psi) = \tilde{m}_{\psi,q}(x_1)$, where the function $\tilde{m}_{\psi,q}(\cdot)$ is unknown and, in the smoothing context, usually assumed to be continuous and differentiable. Here, we will assume that it can be approximated sufficiently well by the following function

$$m_{\psi,q}[x_1; \boldsymbol{\beta}_\psi(q), \boldsymbol{\gamma}_\psi(q)] = \beta_{0\psi}(q) + \beta_{1\psi}(q)x_1 + \\ + \ldots + \beta_{p\psi}(q)x_1^p + \sum_{k=1}^{K} \gamma_{k\psi}(q)(x_1 - \kappa_k)_+^p, \qquad (3)$$

where p is the degree of the spline, $(t)_+^p = t^p$ if $t > 0$ and 0 otherwise, κ_k for $k = 1, \ldots, K$ is a set of fixed knots, $\boldsymbol{\beta}_\psi(q) = (\beta_{0\psi}(q), \beta_{1\psi}(q), \ldots, \beta_{p\psi}(q))^T$ is the coefficient vector of the parametric portion of the model and $\boldsymbol{\gamma}_\psi(q) = (\gamma_{1\psi}(q), \ldots, \gamma_{K\psi}(q))^T$ is the coefficient vector for the spline one. The latter portion of the model allows for handling nonlinearities in the structure of the relationship. If the number of knots K is sufficiently large, the class of functions in (3) is very large and can approximate most smooth functions. In particular, in the p-splines context, a knot is placed every 4 or 5 observations at uniformly spread quantiles of the unique values of x_1. For large datasets, this rule-of-thumb can lead to an excessive number of knots (and therefore parameters), so that a maximum number of allowable knots, say 35, may be recommended. Note that, on the contrary, the degree of the spline does not have to be particularly large: it is usually taken to be between 1 and 3. The spline model (3) uses a truncated polynomial spline basis to approximate the function $\tilde{m}_{\psi,q}(\cdot)$. Other bases can be used; in particular we will later use radial basis functions to handle bivariate smoothing. More details on bases and knots choice can be found in Ruppert et al. (2003, Chapters 3 and 5).

Given the large number of knots, model (3) can be overparametrized and the resulting approximation would look too wriggly. The influence of the knots is limited by putting a constraint on the size of the spline coefficients: typically $\sum_{k=1}^{K} \gamma_{k\psi}^2(q)$ is bound by some constant, while the parametric coefficients $\boldsymbol{\beta}_\psi(q)$ are left unconstrained. Therefore, estimation can be accommodated by mimicking penalization of an objective function and solving the following set of $(1+p+K)$ estimating equations

$$\sum_{i=1}^{n} \psi_q(y_i - \boldsymbol{x}_i\boldsymbol{\beta}_\psi(q) - \boldsymbol{z}_i\boldsymbol{\gamma}_\psi(q))(\boldsymbol{x}_i, \boldsymbol{z}_i)^T + \lambda \begin{bmatrix} \mathbf{0}_{(1+p)} \\ \boldsymbol{\gamma}_\psi(q) \end{bmatrix} = \mathbf{0}, \qquad (4)$$

where $\boldsymbol{x}_i$ here is the i-th row of the $n \times (1+p)$ matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^p \end{bmatrix}, \qquad (5)$$

while $\boldsymbol{z}_i$ is the i-th row of the $n \times K$ matrix

$$\boldsymbol{Z} = \begin{bmatrix} (x_{11} - \kappa_1)_+^p & \cdots & (x_{11} - \kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_{1n} - \kappa_1)_+^p & \cdots & (x_{1n} - \kappa_K)_+^p \end{bmatrix}, \qquad (6)$$

and λ is a Lagrange multiplier that controls the level of smoothness of the resulting fit.

Section 3.2 provides an algorithm to effectively compute $\hat{\boldsymbol{\beta}}_\psi(q)$ and $\hat{\boldsymbol{\gamma}}_\psi(q)$. Once those estimates are obtained, $\hat{m}_{\psi,q}[x_1] = m_{\psi,q}[x_1; \hat{\boldsymbol{\beta}}_\psi(q), \hat{\boldsymbol{\gamma}}_\psi(q)]$ can be computed as our estimate for $Q_q(x_1, \psi)$. The approximation ability of this final estimate will heavily depend on the value of the smoothing parameter λ. Generalized Cross Validation (GCV) has been usefully applied in the context of smoothing splines (Craven and Wahba, 1979) and will be used here too. Details on the criterion are given in Section 3.2.

As we have just dealt with flexible smoothing of quantiles in scatterplots, we can now handle the way in which two continuous variables affect the quantiles of the response without any structural assumptions: $Q_q(x_1, x_2, \psi) = \tilde{m}_{\psi,q}(x_1, x_2)$, i.e. we can deal with *bivariate* smoothing. It is of central interest in a number of application areas as environment and public health. It has particular relevance when geographically referenced responses need to be converted to maps. As seen earlier, p-splines rely on a set of basis functions to handle nonlinear structures in the data. Bivariate smoothing requires bivariate basis functions; Ruppert et al. (2003, Chapter 13) advocate the use of radial basis functions to derive *Low-rank thin plate splines*. In particular, we will assume the following model at quantile q for unit i:

$$m_{\psi,q}[x_{1i}, x_{2i}; \boldsymbol{\beta}_\psi(q), \boldsymbol{\gamma}_\psi(q)] = \beta_{0\psi}(q) + \\ +\beta_{1\psi}(q)x_{1i} + \beta_{2\psi}(q)x_{2i} + \boldsymbol{z}_i\boldsymbol{\gamma}_\psi(q). \quad (7)$$

Here $\boldsymbol{z}_i$ is the i-th row of the following $n \times K$ matrix

$$\boldsymbol{Z} = \left[C(\tilde{\boldsymbol{x}}_i - \boldsymbol{\kappa}_k)\right]_{\substack{1\leqslant i\leqslant n \\ 1\leqslant k\leqslant K}} \left[C(\boldsymbol{\kappa}_k - \boldsymbol{\kappa}_{k'})\right]^{-1/2}_{1\leqslant k\leqslant K}, \quad (8)$$

where $C(\boldsymbol{t}) = ||\boldsymbol{t}||^2 \log||\boldsymbol{t}||$, $\tilde{\boldsymbol{x}}_i = (x_{1i}, x_{2i})$ and $\boldsymbol{\kappa}_k$, $k = 1, \ldots, K$ are knots. The derivation of the $\boldsymbol{Z}$ matrix as in (8) from a set of radial basis functions is lengthy and goes beyond the scope of this paper; Ruppert et al. (2003, Chapter 13) and Kammann and Wand (2003) give a thorough treatment of it. Here, it is enough to notice that the $C(\cdot)$ function is applied so that in the full rank case – i.e. when knots correspond to all the observations – the model for classical bivariate smoothing leads to *Thin plate splines* (see e.g. Green and Silverman, 1994). In addition, the second part of the right hand expression in (8) is a transformation used so that the estimation procedure simplifies; in particular, it can again be written as in (4), with $\boldsymbol{x}_i = (1, \tilde{\boldsymbol{x}}_i)$.

The choice of knots in two dimensions is more challenging than in one. One approach could be that of laying down a rectangular lattice of knots, but this has a tendency to waste a lot of knots when the domain defined by x_1 and x_2 has an irregular shape. In one dimension a solution to this issue is that of using quantiles. However, the extension of the notion of quantiles to more than one dimension is not straightforward. Two solutions suggested in literature that provide a subset of observations nicely scattered to cover the domain are *space filling designs* (see e.g. Ruppert et al., 2003) and the clara algorithm. The first one is based on the maximal separation principle of K points among the unique $\tilde{\boldsymbol{x}}_i$ and is implemented in the `fields` package of the `R` language. The second one is based on clustering and selects K representative objects out of n; it is implemented in the package `cluster` of `R`.

It should be noted, then, that the estimating equations in (4) can be used to handle univariate smoothing and bivariate smoothing by suitably changing the parametric and the spline part of the model, i.e. once the $\boldsymbol{X}$ and the $\boldsymbol{Z}$ matrices are set up. Finally, other continuous or categorical variables can be easily inserted parametrically in the model by adding columns to the $\boldsymbol{X}$ matrix.

3.2 The algorithm

Let us rewrite the set of estimating equations in (4) as follows

$$\sum_{i=1}^{n} \psi_q(y_i - \boldsymbol{u}_i\boldsymbol{\eta}_\psi(q))\boldsymbol{u}_i^T + \lambda\boldsymbol{G}\boldsymbol{\eta}_\psi(q) = \boldsymbol{0}, \quad (9)$$

where $\boldsymbol{u}_i = (\boldsymbol{x}_i, \boldsymbol{z}_i)$, $\boldsymbol{\eta}_\psi(q) = (\boldsymbol{\beta}_\psi(q)^T, \boldsymbol{\gamma}_\psi(q)^T)^T$ and $\boldsymbol{G} = \text{diag}\left\{\boldsymbol{0}_{(1+p)}, \boldsymbol{1}_K\right\}$. If we define the weight function $w(e) = \psi(e)/e$ and let $w_i = w(e_i)$, then (9) can be written as

$$\sum_{i=1}^{n} w_i(y_i - \boldsymbol{u}_i\boldsymbol{\eta}_\psi(q))\boldsymbol{u}_i^T + \lambda\boldsymbol{G}\boldsymbol{\eta}_\psi(q) = \boldsymbol{0}. \quad (10)$$

Solving this set of estimating equations is a penalized weighted least squares problem in which weights, residuals and coefficients depend one upon another. Further, the value of the smoothing parameter λ has to be chosen. The GCV criterion to be optimized (minimized) to this end is the following

$$GCV(\lambda) = \frac{\sum_{i=1}^{n}\left\{(\boldsymbol{I} - \boldsymbol{S}_\lambda)\boldsymbol{y}\right\}_i}{\left(1 - n^{-1}\theta\text{tr}(\boldsymbol{S}_\lambda)\right)^2},$$

where $\boldsymbol{S}_\lambda$ is the smoother matrix associated with $\hat{m}_{\psi,q}[\boldsymbol{u}_i]$, i.e. $\hat{m}_{\psi,q}[\boldsymbol{u}_i] = \boldsymbol{S}_\lambda\boldsymbol{y}$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, and θ is a constant that penalizes additional degrees of freedom given by the trace of the smoother matrix.

An iterative solution to this problem through *Iteratively Reweighted Penalized Least Squares*, IRPLS, is here proposed. In what follows we will consider fixed the influence function ψ and the quantile of interest q; we will then drop suffixes and indexes for ease of notation when this will not lead to ambiguity.

1. Select initial estimates $\boldsymbol{\eta}^0$.
2. At each iteration t, calculate residuals $e_i^{(t-1)} = y_i - \boldsymbol{u}_i\boldsymbol{\eta}^{(t-1)}$ and associated weights $w_i^{(t-1)}$ from the previous iteration.
3. Optimize the $GCV(\lambda)$ criterion over a grid of λ values and obtain λ^*.
4. Calculate the new weighted penalized least squares estimates as

$$\boldsymbol{\eta}^t = \left[\boldsymbol{U}^T\boldsymbol{W}^{(t-1)}\boldsymbol{U} + \lambda^*\boldsymbol{G}\right]^{-1}\boldsymbol{U}^T\boldsymbol{W}^{(t-1)}\boldsymbol{y},$$

where $\boldsymbol{U} = \{\boldsymbol{u}_i\}_{i=1,\ldots,n}$ and $\boldsymbol{W}^{(t-1)} = \text{diag}\left\{w_i^{(t-1)}\right\}$ is the current weight matrix.

Iterate steps 2, 3 and 4 until convergence. `R` code that implements this algorithm is available from the authors.

4 Simulation studies

In this section we report on some Monte Carlo simulation studies carried out to investigate the performance of the p-splines M-quantile regression – PSPL– as compared to standard linear M-quantiles – LIN. We first report on simulations with a single covariate and then move to the bivariate case.

4.1 A single covariate

The following four models are used to generate the true underlying relationship between the covariate x and the response variable y:

Linear. $m(x) = 1 + 2(x - 0.5)$;

Exponential. $m(x) = \exp(6x)/400$;

Cycle. $m(x) = 2\sin(2\pi x)$;

Jump. $m(x) = 1 + 2(x - 0.5)I(x \leqslant 0.5) + 0.5I(x > 0.5)$.

The first model represents a situation in which LIN is based on the true model and PSPL may be too complex and overparametrized. The second and third models define an increasingly more complicate structure of the relationship between y and x, while the last one is a discontinuous function for which both LIN and PSPL are misspecified. More in detail, $n = 200$ x values are generated from a Uniform distribution in $[0, 1]$; y values are generated at each replicate by adding errors to the signals defined above. Two different settings are considered: Gaussian errors with mean 0 and standard deviation 0.4 and Cauchy errors with location parameter 0 and scale parameter 0.05. The first setting is considered as a situation of "regularly" noisy data with a signal-to-noise ratio of about 2 for all signals, but the Exponential function for which it is only about 0.4. The second one, on the contrary, defines a situation of more noisy data with the likely presence of extreme and outlying observations. This provides a 4×2 design of simulations.

For each simulation and each of the $R = 1000$ replicates, LIN and PSPL parameter estimates and response estimates at observed x points are calculated at the deciles using the Huber 2 influence function. In addition, for PSPL a truncated linear bases is used, i.e. $p = 1$, with $K = 39$ knots set at x quantiles; the smoothing parameter λ has been chosen via GCV with $\theta = 2$; this means that each additional degree of freedom used to approximate the underlying signal is penalized twice. It is common to use a value of θ between 1 and 3.

For each technique the following quantities are computed at each quantile to compare performances.

MCEV. Monte Carlo Expected Value, defined for each i and each q as $R^{-1}\sum_{r=1}^{R}\hat{m}^r_{\psi,q}[x_i]$;

MASE. Mean Average Squared Errors, defined for each q as

$$(Rn)^{-1}\sum_{i=1}^{n}\sum_{r=1}^{R}(\hat{m}^r_{\psi,q}[x_i] - m_{\psi,q}[x_i])^2.$$

Figure 1 shows the MCEV for both LIN and PSPL for all simulations, together with the true value of the signal for the nine deciles investigated. LIN works well in the linear case for all quantiles; PSPL, on the other hand, seems to work well with more complicated structures and is able to capture even the Cycle signal with a Cauchy error. These findings were somehow expected and are supported by Table 1. The first part of it reports the values of the ratios of LIN MASEs to the PSPL ones. Large gains in efficiency of PSPL over LIN are shown for the more complicated structures as expected. In the Linear case, the performance of the two methods is similar in the Gaussian case, while for Cauchy errors losses of efficiency are shown by PSPL.

4.2 Bivariate case

Let x_1 and x_2 take uniformly spread values in the interval $[-1, 1]$ to form a grid of $n = 256$ points. Two model surfaces have been considered:

Plane. $m(x_1, x_2) = 0.5x_1 + 0.2x_2$;

Mountain. $m(x_1, x_2) = \cos\sqrt{(1.2\pi x_1)^2 + (1.2\pi x_2)^2}$.

Figure 2 shows the perspective plots of these two models. Response values are generated at each simulation replicate by adding errors to the surfaces introduced.

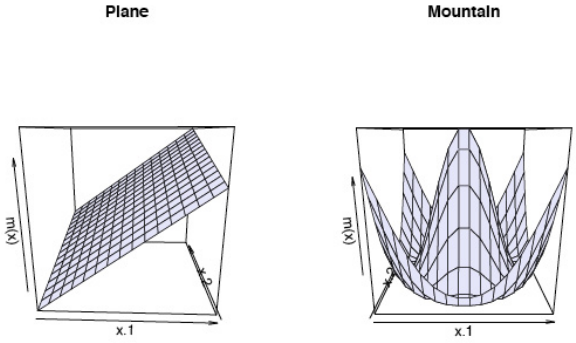

Figure 2: Perspective plots of the two models used in the bivariate simulation studies.

As in the previous section, two settings are considered: Gaussian errors with mean 0 and standard deviation 1 and Cauchy errors with location parameter 0 and scale parameter 1. Signal to noise ratios for the Gaussian settings take values of 0.11 for the Plane surface and 0.26 for the Mountain one; these represent less good-quality datasets compared to the univariate case. This becomes especially true for the Cauchy errors distribution case. A 2×2 design of simulations is therefore set up. For each of the $R = 1000$ replicates LIN and PSPL parameters and surface estimates have been computed; in particular, PSPL uses the radial basis mentioned in Section 3.1 with $K = 50$ knots laid down on a regular grid. The performance quantities computed for the two techniques are the same as those explored for the univariate case.

Plots of MCEV for all cases would be too space consuming and are not reported here, although available from the authors. Here we report only those for the Plane and Mountain with Gaussian errors simulations and a subset of quantiles. Figures 3 and 4 are arranged with quantiles on rows and, respectively, the true surface, LIN and PSPL MCEVs on columns. Biases look negligible in all cases except for the LIN approximation of the Mountain surface as expected.

The second part of Table 1 reports MASE ratios for all quantiles and the four simulations. Gains in efficiency for

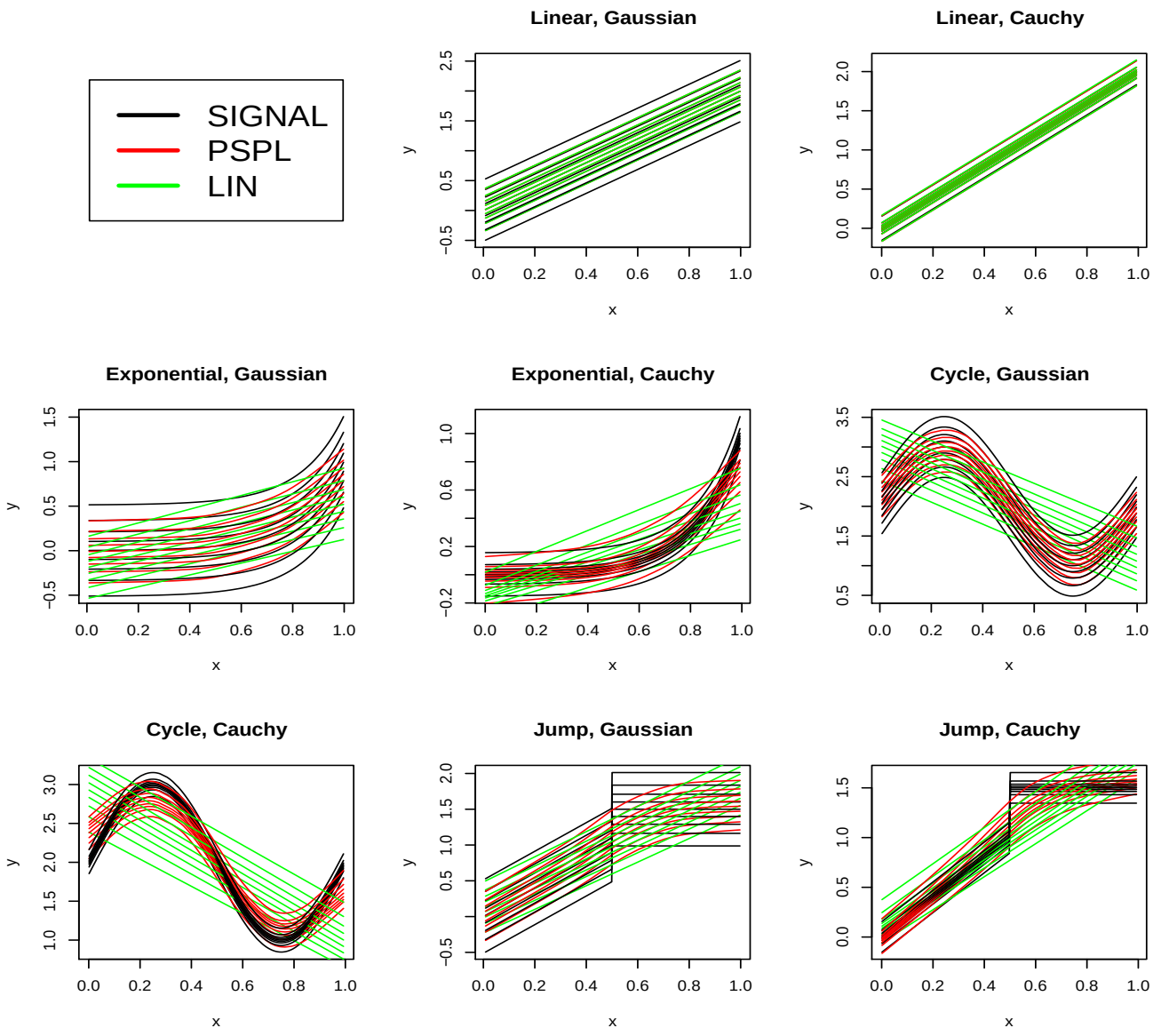

Figure 1: MCEV for LIN and PSPL together with the true quantile functions for all univariate simulation studies.

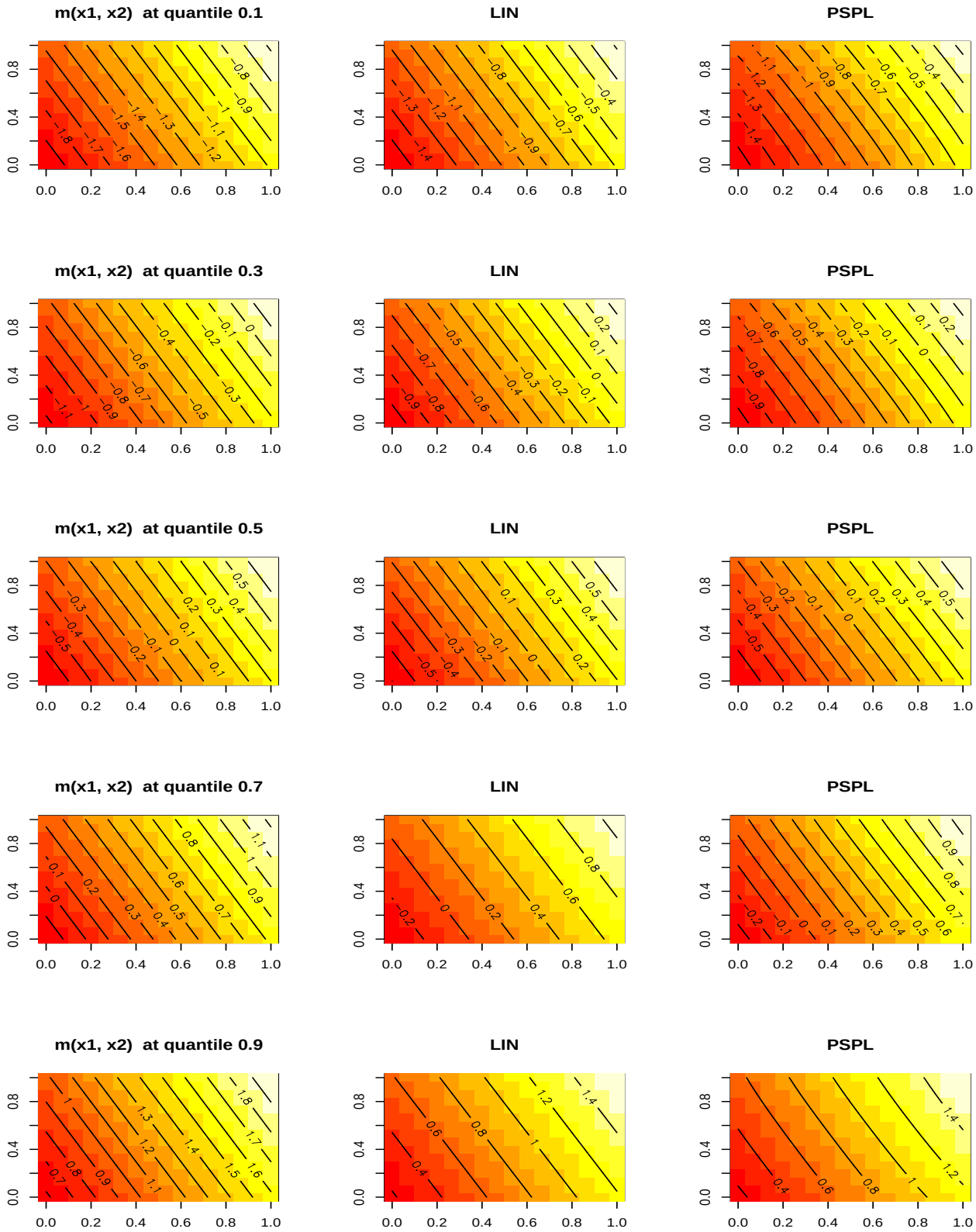

Figure 3: Images of the true quantile function and MCEVs for LIN and PSPL at five quantiles for the Plane with Gaussian errors simulation.

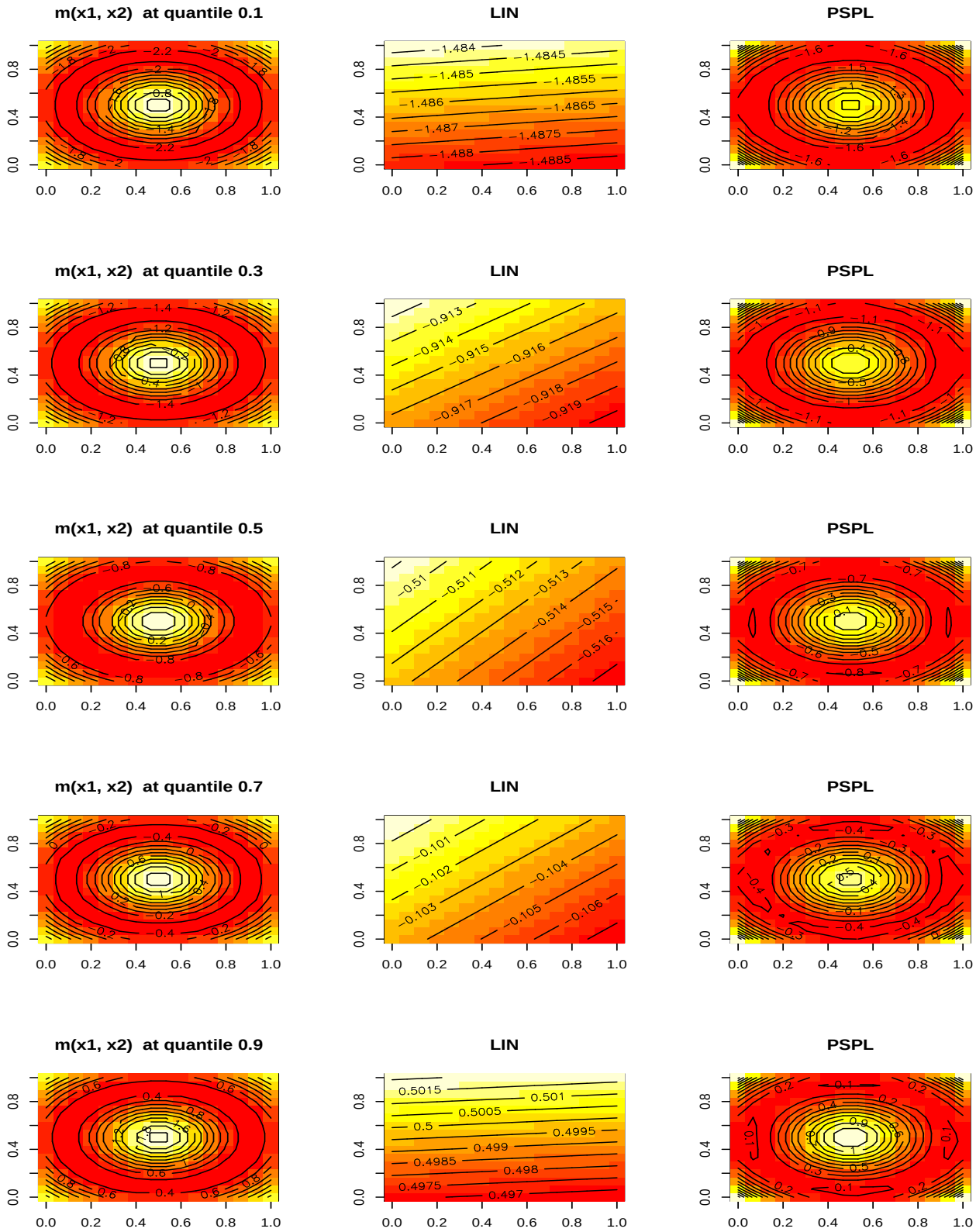

Figure 4: Images of the true quantile function and MCEVs for LIN and PSPL at five quantiles for the Mountain with Gaussian errors simulation.

Table 1: MASE values for LIN for each quantile and simulation study; MASE for PSPL $= 1$.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Univariate									
Linear, Gaussian	1.0	1.0	0.9	0.9	0.8	0.8	0.9	0.9	1.0
Linear, Cauchy	0.9	0.7	0.6	0.5	0.6	0.6	0.7	0.9	1.1
Exponential, Gaussian	1.3	1.8	2.5	3.5	4.0	3.4	2.5	1.8	1.3
Exponential, Cauchy	2.3	4.7	6.8	8.2	9.1	9.9	9.8	7.2	3.0
Cycle, Gaussian	5.5	10.4	17.5	26.1	30.7	26.2	17.6	10.5	5.6
Cycle, Cauchy	5.0	5.4	5.2	5.0	4.9	4.9	5.1	5.2	4.7
Jump, Gaussian	1.2	1.6	1.9	2.2	2.4	2.2	1.9	1.6	1.2
Jump, Cauchy	2.1	2.8	3.0	3.1	3.2	3.3	3.3	3.2	2.6
Bivariate									
Plane, Gaussian	0.9	0.9	0.8	0.7	0.6	0.7	0.8	0.9	0.9
Plane, Cauchy	1.0	0.8	0.7	0.7	0.7	0.7	0.8	0.9	1.0
Mountain, Gaussian	1.3	1.9	2.5	2.9	3.1	2.9	2.5	1.9	1.3
Mountain, Cauchy	1.0	0.9	1.0	1.1	1.1	1.1	1.0	1.0	1.0

PSPL are shown as expected for the Mountain response surface. Such gains are more remarkable for the Gaussian errors distribution. Losses in efficiency are shown for the Plane surface and central quantiles.

5 Conclusions

In this paper we propose an extension to M-quantile regression with nonparametric modeling via penalized splines. This may be particularly useful when the functional form of the relationship between the variable of interest and the covariates is believed to be non linear. The proposed approach, beyond the features of M-quantile models, allows for dealing with undefined functional forms that can be estimated from the data. To fit the model we propose an algorithm based on penalized iteratively reweighted least squares.

Relative performances of nonparametric M-quantile regression are evaluated through Monte Carlo experiments. Results from the simulation studies indicate that this approach works well and competes with the conventional M-quantile regression models. The nonparametric M-quantile regression can be widely used in many important application areas, such as financial and economic statistics and environmental and public health modeling. Currently, following Chambers and Tzavidis (2006), the authors are investigating the use of nonparametric M-quantile models in small area estimation methods in the case in which the functional form of the relationship between the variable of interest and the covariates is left unspecified (Pratesi, Ranalli and Salvati, 2006).

Acknowledgments

The work reported here has been developed under the support of the project PRIN *Metodologie di stima e problemi non campionari nelle indagini in campo agricolo-ambientale* awarded by the Italian Government to the Universities of Florence, Cassino, Pisa and Perugia. The authors would like to thank Ray Chambers for his help and support.

References

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika* **75**, 761–771.

Chambers, R. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika* **93**, 255–268.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik* **31**, 377–403.

Eilers, P. H. C. and Marx, B. D. (1996). Reply to comments on "Flexible smoothing with B-splines and penalties". *Statistical Science* **11**, 115–121.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall Ltd.

Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: with 200 Full-color Illustrations*. Springer-Verlag Inc.

He, X. (1997). Quantile curves without crossing. *The American Statistician* **51**, 186–192.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society - Series C,* **52**, 1–18.

Koenker, R. (1984). A note on l-estimators for linear models. *Statistics and Probability Letters* **2**, 323–325.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

Koenker, R. and D'Orey, V. (1987). Computing regression quantiles. *Biometrika* **93**, 255–268.

Kokic, P., Chambers, R., Breckling, J. & Beare, S. (1997). A measure of production performance. *Journal of Business and Economic Statistics* **10**, 419–435.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**, 819–847.

Pratesi, M., Ranalli, M.G. and Salvati, N. (2006). P-splines M-quantile small area estimation: assessing the ecological conditions of lakes in the Northeastern US. *Proceedings of the Conference on Spatial Data Methods for Environmental and Ecological Processes*, Foggia, September 2006.

Ruppert, D., Wand, M.P. and Carroll, R. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge, New York.

Takeuchi, I., Le, V., Sears, T. and Smola, A. (2005). Nonparametric quantile regression. *Journal of Machine Learning Research* **7**, 1001–1032.