

RESEARCH ARTICLE

M-quantile regression shrinkage and selection via the Lasso and Elastic Net to assess the effect of meteorology and traffic on air quality

M. Giovanna Ranalli¹  | Nicola Salvati²  | Lea Petrella³ | Francesco Pantalone⁴ 

¹Department of Political Science,
University of Perugia, Perugia, Italy

²Department of Economics and
Management, University of Pisa, Pisa,
Italy

³MEMOTEF Department, Sapienza
University of Rome, Rome, Lazio, Italy

⁴Department of Social Statistics and
Demography, University of Southampton,
Southampton, UK

Correspondence

Maria Giovanna Ranalli, Department of
Political Science, University of Perugia,
Via Pascoli, 20, Perugia 06123, Italy.
Email: giovanna.ranalli@unipg.it

Funding information

Università degli Studi di Perugia, Fondo
Ricerca di Ateneo, edizione 2021,
Grant/Award Numbers: ProgettoAIDMIX,
anenhancedmodellintoolforthetransi-
tiontosustainabil-
ity,G.A.No.821105(Horizon2020);
European Commission, Grant/Award
Numbers: 730998, 821105



This article has earned an open data
badge “**Reproducible Research**” for
making publicly available the code
necessary to reproduce the reported
results. The results reported in this article
could fully be reproduced.

Abstract

In this work, we intersect data on size-selected particulate matter (PM) with vehicular traffic counts and a comprehensive set of meteorological covariates to study the effect of traffic on air quality. To this end, we develop an M-quantile regression model with Lasso and Elastic Net penalizations. This allows (i) to identify the best proxy for vehicular traffic via model selection, (ii) to investigate the relationship between fine PM concentration and the covariates at different M-quantiles of the conditional response distribution, and (iii) to be robust to the presence of outliers. Heterogeneity in the data is accounted by fitting a B-spline on the effect of the day of the year. Analytic and bootstrap-based variance estimates of the regression coefficients are provided, together with a numerical evaluation of the proposed estimation procedure. Empirical results show that atmospheric stability is responsible for the most significant effect on fine PM concentration: this effect changes at different levels of the conditional response distribution and is relatively weaker on the tails. On the other hand, model selection allows to identify the best proxy for vehicular traffic whose effect remains essentially the same at different levels of the conditional response distribution.

KEYWORDS

additive models, B-splines, cross-validation, influence function, robust regression

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Biometrical Journal* published by Wiley-VCH GmbH.

1 | INTRODUCTION

Between 2012 and 2015, an innovative system for urban air-quality monitoring was put in place in Perugia (Italy). Aerosol and gas measurements with very high spatiotemporal resolution were obtained through proper instruments integrated on one of the cabins of a public conveyance of the town: the Minimetro. The Minimetro path, about 3 km long, starts from a large suburban parking area, crosses various heavy traffic roads, passes through an urban park, residential areas, tunnels, and climbs a hill to finally reach the old town centre. The different urban scenarios located along the path provide, therefore, information about different environmental situations, while the constant and low speed of the cabins allows for an easy and precise monitoring. Castellini et al. (2014) provide a thorough description of the PMetro project, while Ranalli et al. (2016), Del Sarto, Ranalli, Bakar, et al. (2016), and Del Sarto, Ranalli, Cappelletti, et al. (2016) use functional data analysis and Hierarchical Bayesian spatiotemporal modeling, respectively, to analyze these high-frequency data.

In this paper, we integrate the fine particulate matter (PM) concentration measurements coming from the Minimetro mobile station with data coming from a traffic sampling equipment positioned at one of the intersections of the Minimetro path counting the number of vehicles every 5 min, 24 h/day. We also collect hourly weather conditions data (such as temperature, wind speed, and rainfall) coming from fixed monitoring stations of the Regional Environmental Protection Agency placed in the surrounding of the path, and merge them with the other two data sets. Since data are collected at different time frequency, we convert values to hourly data. Moreover, we reduce the spatial resolution of the data using information from only one point given by the crossroad intersecting the Minimetro path.

For the same data set, Crocchianti et al. (2020) investigate the effect of weather conditions and vehicular traffic by means of additive mixed models while Del Sarto et al. (2019) apply a finite mixture of M-quantile regression models (Alfò et al., 2017). In both papers, an intensive model selection procedure is put in place to select the variable that best measures the effect of vehicular traffic. In fact, traffic can be measured in terms of the number of vehicles passing at the crossroad every hour, at different lag hour or by cumulating them at h hours before the current time point with h ranging on a defined interval. To this end, several models are estimated in Crocchianti et al. (2020) and in Del Sarto et al. (2019), each one with a different traffic-related covariate (all other things being unchanged). The model fitting is evaluated using classical information criteria such as the Bayesian information criterion (BIC), and the model with the lowest BIC is retained. In particular, in Del Sarto et al. (2019) the variable selection step is preliminary and the best proxy for traffic is retained in all subsequent model selection analyses and for all levels (quantiles and M-quantiles) of the conditional distribution of the response given the covariates. This can be a limitation as the effect of traffic can change at different levels of the conditional distribution. In addition, results in Crocchianti et al. (2020) show that the effect of vehicular traffic changes for different fractions of PM, with a larger number of cumulation hours needed to better trace the vehicular effect on smaller particles with respect to larger (coarse) ones. Finally, results from previous analyses are not clear-cut and may be influenced by the hour of the day which is not accounted for in those papers. For all these reasons, in this paper we are interested in investigating how meteorological conditions and vehicular traffic levels affect fine PM concentrations at different levels of the conditional response distribution. Furthermore, we conduct model selection and estimation at the same time, by including all traffic-related variables and the hour of the day as covariates in the model.

When dealing with a problem where the data set contains a large number of covariates, the selection of predictors is a crucial goal. Recently, there has been an active research production on sparse representation of regression problems starting from the work of Tibshirani (1996). In his paper, he introduces the least absolute shrinkage and selection operator (Lasso) as a penalization technique which can simultaneously perform variable selection and parameter estimation. In particular, Lasso estimation represents an L_1 -regularized least squares estimate of the parameters of the model. When categorical predictors are present in the regression model, Lasso may not be completely satisfactory since it only selects individual dummy variables instead of the whole predictor. Moreover, it fails to do grouped selection. In fact, it tends to select one variable from a group and ignore the others. To overcome these problems, several solutions and extensions are proposed in the literature such as the Elastic Net of Zou and Hastie (2005), having an improved performance in particular for correlated predictors. The Elastic Net uses both the L_1 - and the L_2 -norms to penalize the model and usually outperforms the Lasso regularization while enjoying a similar sparsity of representation. In addition, it encourages a grouping effect, where strongly correlated predictors tend to be in or out of the model together.

In this paper, we introduce M-quantile regression shrinkage and selection via the Lasso and Elastic Net to analyze PMetro data. This allows (i) to identify the best proxy for vehicular traffic via model selection and fitting at the same time, while accounting for the hour of the day, (ii) to investigate the relationship between fine PM concentration and the covariates at different M-quantiles of the conditional response distribution, and (iii) to obtain estimates that are robust to outliers. In addition, to account for the clustered-time structure of the data, we introduce in the model a smooth function

of time (day of the year) and we estimate it using penalized B-splines (Eilers & Marx, 1996). This should account for social and generally unmeasured confounders on the distribution of fine PM as proposed also in Bertaccini et al. (2012).

Regularization has been introduced and investigated in the quantile regression framework (Belloni & Chernozhukov, 2011; Koenker, 2004; Li & Zhu, 2008; Petrella & Raponi, 2019; Slawski, 2012; Tang & Lian, 2016; Wang et al., 2012; Yan & Song, 2019; Zhang et al., 2017), also from a Bayesian perspective (Aghamohammadi & Mohammadi, 2017; Alhamzawi et al., 2012; Bernardi et al., 2018; Kyung et al., 2010; Li et al., 2010; Tian et al., 2019). In this paper, we focus on M-quantiles (Breckling & Chambers, 1988). These extend the ideas of M-estimation to a different set of location parameters of the response conditional distribution that lie between quantiles and expectiles. M-quantile regression can be considered as a robustification to outliers of expectile regression (Kneib, 2013; Newey & Powell, 1987) based on influence functions. Moreover, although M-quantiles have a less intuitive interpretation than standard quantiles (Jones, 1994), M-quantile regression offers a number of specific advantages: (i) it allows to trade robustness for efficiency by allowing to select a tuning constant in the influence function; and (ii) it offers computational stability as a wide range of continuous influence functions can be used (Tzavidis et al., 2016).

With respect to existing literature on M-quantile regression, we introduce the Lasso and the Elastic Net regularization for shrinkage and model selection. In fact, Spiegel et al. (2017) propose the use of nonnegative garrote for semiparametric expectile regression; Zhao and Zhang (2018) propose variable selection in expectile regression using the SCAD (smoothly clipped absolute deviation) penalty function, while Xu et al. (2021) introduce the Elastic Net penalty into expectile regression; Waldmann et al. (2017) perform variable selection for covariates with linear effects in Bayesian expectile regression using Bayesian regularization priors. Yi and Huang (2017) introduce Elastic Net penalization to quantile regression and to Huber loss regression. Our approach is similar to Yi and Huang (2017) as we also consider Huber loss regression, but the latter considers only the central M-quantile. Up to our knowledge, our paper is the first attempt to consider regularized M-quantile methods for the whole set of location parameters and fills in a gap in the literature. In addition, we include a nonparametric term to model departures from linearity of the effect of a covariate on the response and we model it using penalized B-splines. This can be easily extended to additive modeling. To avoid issues in combining the Lasso with penalized predictors (Marra & Wood, 2011), we propose to estimate the coefficients of the B-splines with a penalized least-square procedure similar to that in Pratesi et al. (2009) and outside the Lasso estimation procedure.

The paper is organized as follows. Section 2 describes the data used in the application and Section 3 introduces the methodology. In particular, we introduce the M-quantile with Lasso regularization in Section 3.2, and then we extend the method to the use of Elastic Net in Section 3.3. In Section 3.4, we account for nonlinearities through B-splines. The results of the proposed model to the PMetro data are reported in Section 4. Conclusions and directions for future research are provided in Section 5.

2 | DATA

Data come from the project PMetro, which ran between 2012 and 2015 in order to assess air quality in the town of Perugia (Italy). The response variable is collected through an optical particle counter (OPC) placed on one of the cabins of the Minimetro transportation system (for more details on the performance of the OPC, see Castellini et al., 2014). These cabins travel on a monorail built 5 m above the road level and are driven by a wire rope running at a low speed of about 4–7 m/s. There are seven stations along the path and a single travel is about 20 min long, which is repeated about 40 times per day.

Since PM gravimetric and chemical measurements require long times to achieve reliable results, the OPC measures the particle size distribution in terms of the number of particles in a range of selected aerosol bins. In particular, the concentration measure collected by the OPC is divided into 22 dimensional bins according to the particle diameter, expressed in micrometers (μm , 10^{-6}m), from 0.28 to 10 μm , and we focus on concentration of particles with a diameter lower than 1.10 μm , obtained by summing the first nine dimensional bins.

The predictor variables are used to measure traffic and weather conditions. The latter are measured by fixed monitoring stations of the Regional Environmental Protection Agency while the former are obtained by traffic sampling equipment installed by the Municipality. In particular, the fixed monitoring stations are located along the Minimetro path and provide data about temperature, wind speed, rainfall, total solar radiation, and radon concentration. Traffic is measured by the number of vehicles. These data are collected with different time frequencies and in different locations. In order to exploit all sources of information, the measurement of the variables has been referred to the same time frequency, that is the hour, and to the same location, that is the intersection between the Minimetro path, the fixed monitoring station and the crossroad with traffic equipment. The period considered goes from March 2014 to February 2015, for a total of 201 days

TABLE 1 Descriptive statistics of the data.

Variable	Unit	Range	Mean	SD
Particle concentration	count/m ³	(1968; 615,434)	82,579.99	82,003.59
Particle concentration (log)	-	(7.59; 13.33)	10.93	0.89
Radon concentration	count/min	(2; 2410)	353.49	298.37
Temperature	C	(−4.10; 33.60)	15.28	7.80
Wind speed	m/s	(0; 6.50)	0.74	0.81
Total solar radiation	mW/m ²	(0.20; 77.10)	16.49	18.81
Rainfall	mm	(0; 11)	0.06	0.40
Vehicular traffic	count	(349; 3719)	2414.78	725.93

and 2,608 hourly observations. We use data for a whole year in order to account for seasonal trends. In Table 1, we provide a summary of the variables. Other descriptive statistics of the data are provided in the Supporting Information Appendix (Table A.7) and in Del Sarto et al. (2019).

3 | THE PROPOSED MODELING APPROACH

3.1 | Quantile versus M-quantile regression

When investigating the relationship between a response variable y and p predictors \mathbf{x} , quantile regression (Koenker & Bassett, 1978; Koenker & D'Orey, 1987) allows to analyze the τ th quantile of the conditional distribution of y given \mathbf{x} . This is particularly useful when the average behavior of y given \mathbf{x} , investigated by means of the classical regression analysis, does not give a complete picture of the distribution. In particular, quantile regression assumes that the τ th quantile of y is a linear function of the predictors, in other words

$$Q_\tau(y|\mathbf{x}) = \beta_{0\tau} + \mathbf{x}^T \boldsymbol{\beta}_{1\tau},$$

where $\beta_{0\tau}$ is the intercept and $\boldsymbol{\beta}_{1\tau}$ is the vector of regression coefficients. This model leads to a family of hyperplanes indexed by the value of the corresponding quantile coefficient $\tau \in (0, 1)$. Given a sample of n observations (y_i, \mathbf{x}_i) for $i = 1, \dots, n$, the vector $\boldsymbol{\beta}_\tau = (\beta_{0\tau}, \boldsymbol{\beta}_{1\tau}^T)^T$ is estimated by minimizing

$$\sum_{i=1}^n |r_i(\boldsymbol{\beta}_\tau)| \{(1 - \tau)I(r_i(\boldsymbol{\beta}_\tau) \leq 0) + \tau I(r_i(\boldsymbol{\beta}_\tau) > 0)\},$$

with respect to $\boldsymbol{\beta}_\tau$ through linear programming methods, where $r_i(\boldsymbol{\beta}_\tau) = y_i - \beta_{0\tau} - \mathbf{x}_i^T \boldsymbol{\beta}_{1\tau}$.

M-quantile regression (Breckling & Chambers, 1988) generalizes the quantile (Koenker & Bassett, 1978) and the expectile regression idea of Newey and Powell (1987) through the use of influence functions. Specifically, the M-quantile regression line of order q is defined as the solution

$$MQ_q(y|\mathbf{x}; \psi_q) = \beta_{\psi_0q} + \mathbf{x}^T \boldsymbol{\beta}_{\psi_1q}$$

to

$$\int \psi_q(y - MQ_q) dF(y|\mathbf{x}) = 0,$$

where F denotes the distribution of y given \mathbf{x} underlying the data and ψ_q denotes the influence function associated to the q th M-quantile. The general M-estimator of $\boldsymbol{\beta}_{\psi_q} = (\beta_{\psi_0q}, \boldsymbol{\beta}_{\psi_1q}^T)^T$ is obtained by solving the following set of estimating equations:

$$\sum_{i=1}^n \psi_q(r_i(\boldsymbol{\beta}_{\psi_q}))(1, \mathbf{x}_i^T)^T = \mathbf{0} \quad (1)$$

with respect to $\beta_{\psi q}$. It is assumed that

$$\psi_q(r_i(\beta_{\psi q})) = 2\psi\{\sigma_q^{-1}r_i(\beta_{\psi q})\}\{(1 - q)I(r_i(\beta_{\psi q}) \leq 0) + qI(r_i(\beta_{\psi q}) > 0)\},$$

where $r_i(\beta_{\psi q}) = y_i - \beta_{\psi 0q} - \mathbf{x}_i^T \beta_{\psi 1q}$ and σ_q is the scale.

We consider throughout the paper the Huber (Huber, 1981) influence function, given by $\psi(u) = u$, if $-c \leq u \leq c$, and $\psi(u) = c \text{ sign}(u)$ otherwise, where c is a cutoff constant. For ease of notation, we will drop ψ and use simply β_q instead of $\beta_{\psi q}$.

The role of the cutoff constant c is relevant. In particular, robustness is increased as c decreases, while efficiency is increased as c increases. Note that expectile regression is a particular case of M-quantile regression with $c \rightarrow \infty$, so that for $q = 0.5$, the solution to (1) coincides with ordinary least squares. In this sense, M-quantiles can be seen as a robustification of expectiles. Also, quantile regression can be seen as a particular case of M-quantile regression when the tuning constant c tends to 0. In practice, the tuning constant c is set by the data analyst in order to provide a trade-off between robustness and efficiency. Huber (1981, p. 18) suggests that “good choices are in the range between 1 and 2, say, 1.5.” The default value for c in the `r1m` function of the R package MASS is 1.345, as suggested by Holland and Welsch (1977), which corresponds to 95% of efficiency of the estimates under normality. When the errors are normally distributed, the best choice is to set the tuning constant equal to a large value such as 100, since in this case using a smaller value, such as 1.345, will offer unnecessary robustness at the cost of reduced efficiency of the estimates. If in the diagnostic analysis some outliers are identified, a low value for the tuning constant, such as 1.345, is preferred because resistance against outliers is necessary. Bianchi et al. (2018) propose a data-driven method to select the tuning constant in the Huber loss function using the Generalized Asymmetric Least Informative density. Otto-Sobotka et al. (2019) replace the value of c with a function that represents the changes of the error distribution along the covariates to handle heteroskedastic errors and obtain an estimate of this function from the data.

3.2 | M-quantile regression shrinkage and selection via the Lasso

Suppose we collect sample data on predictors in matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ and sample data on the response variable in vector $\mathbf{y} \in \mathbb{R}^n$. The Lasso-type estimator for M-quantile regression coefficients at $q \in (0, 1)$ can be defined as

$$\hat{\beta}_q = \left(\hat{\beta}_{0q}, \hat{\beta}_{1q}^T \right)^T = \arg \min_{\beta_q} \{L(\mathbf{X}, \mathbf{y}; \beta_q) : \beta_{0q} \in \mathbb{R}, \beta_{1q} \in \mathbb{R}^p\} \text{ subject to } \sum_{j=1}^p |\beta_{jq}| \leq t_q, \tag{2}$$

where the loss function L is given by

$$L(\mathbf{X}, \mathbf{y}; \beta_q) = \sum_{i=1}^n \psi_q(r_i(\beta_q)/\sigma_q) \tag{3}$$

with $r_i(\beta_q) = y_i - \beta_{0q} - \mathbf{x}_i^T \beta_{1q}$, and σ_q is the scale of the residuals that can be estimated by MAD (mean absolute deviation). Here, $t_q \geq 0$ is the tuning parameter that controls the amount of shrinkage applied to the estimates. Setting the derivatives of (3) with respect to β_q to zero yields the following system of equations:

$$[\mathbb{1}_n \mathbf{X}]^T \mathbf{W}_q (\mathbf{y} - \beta_{0q} \mathbb{1}_n - \mathbf{X} \beta_{1q}) = \mathbf{0}_{1+p}, \tag{4}$$

that can be written also as

$$([\mathbb{1}_n \mathbf{X}]^T \mathbf{W}_q \mathbf{X}) \beta_{1q} = [\mathbb{1}_n \mathbf{X}]^T \mathbf{W}_q (\mathbf{y} - \beta_{0q} \mathbb{1}_n), \tag{5}$$

where $\mathbb{1}_n$ is an n vector of ones, $\mathbf{0}_{1+p}$ is a $p + 1$ vector of zeros, $\mathbf{W}_q = \text{diag}(w_{iq})$, and $w_{iq} = \psi_q(r_i(\beta_q))/2r_i(\beta_q)$. As it is customary in M-quantile regression, the weighted normal equations (4) and (5) require an iterative procedure because the weights depend on the regression coefficients. In particular, the iterative procedure can be summarized as follows:

1. Start with an initial value for $\hat{\beta}_q$, compute residuals $r_i(\hat{\beta}_q)$ and weight matrix \mathbf{W}_q ;
2. Given $\hat{\beta}_{0q}$ and \mathbf{W}_q , estimate $\hat{\beta}_{1q}$ from Equation (5);
3. Recompute \mathbf{W}_q and estimate $\hat{\beta}_{0q}$ from Equation (4);
4. Alternate steps 2 and 3 until convergence.

Note that through this procedure we obtain a solution for the unconstrained case, that is, to (3) only without satisfying constraints on the sum of the absolute values of the coefficients in (2). In order to get the Lasso estimates, let $\delta_\ell, \ell = 1, 2, \dots, 2^p$ be the p -tuples $(\pm 1, \pm 1, \dots, \pm 1)$. Then, condition $\sum_{j=1}^p |\beta_{jq}| \leq t_q$ is equivalent to $\delta_\ell^T \beta_{1q} \leq t_q, \forall \ell$. For a given β_{1q} , let $E = \{\ell : \delta_\ell^T \beta_{1q} = t_q\}$ and $S = \{\ell : \delta_\ell^T \beta_{1q} < t_q\}$, and denote by \mathbf{G}_E the matrix whose rows are δ_ℓ for $\ell \in E$. Now the previous unconstrained procedure can be updated in order to find a solution to the system of equations subject to the Lasso constraint. In particular, step 2 should be modified as follows:

2. Given $\hat{\beta}_{0q}$ and \mathbf{W}_q , estimate $\hat{\beta}_{1q}$ from Equation (5) subject to $\mathbf{G}_E \beta_{1q} \leq t_q \mathbf{1}_e$ where e denotes the number of rows of \mathbf{G}_E , by using the constrained M-quantile procedure proposed by Fabrizi et al. (2012); if $\sum_{j=1}^p |\beta_{jq}| > t_q$, add ℓ to the set E where $\delta_\ell = \text{sign}(\hat{\beta}_{1q})$.

Note that the procedure must always converge in a finite number of steps since one element is added to the set E at each step, and there is a total of 2^p elements. For $q = 0.5$, Yi and Huang (2017) consider a semismooth Newton coordinate descent algorithm to obtain parameter estimates. We compare the final estimates obtained with this alternative estimation procedure in a simulation study in Supporting Information Appendix A.2. The Lasso parameter t_q can be estimated by cross-validation or generalized cross-validation (GCV).

The covariance matrix of the Lasso estimates of the regression coefficients may be approximated following the approach in Tibshirani (1996). In particular, by writing the penalty $\sum_{j=1}^p |\beta_{jp}|$ as $\sum_{j=1}^p \beta_{jp}^2 / |\beta_{jp}|$, the Lasso estimate in (3) may be approximated by a ridge regression of the type

$$\hat{\beta}_q = ([1_n \mathbf{X}]^T \mathbf{W}_q [1_n \mathbf{X}] + \hat{\lambda}_q \mathbf{C})^{-1} [1_n \mathbf{X}]^T \mathbf{W}_q \mathbf{y}, \quad (6)$$

where \mathbf{W}_q is the diagonal matrix with weights from the last iteration, $\hat{\lambda}_q$ is that value for which $\sum_{j=1}^p |\hat{\beta}_{jp}| = \hat{t}_q$, and $\mathbf{C} = \text{diag}[0, 1/|\hat{\beta}_{1q}|, \dots, 1/|\hat{\beta}_{pq}|]$ is the penalty matrix. Then, a sandwich-type estimator can be written as

$$\hat{V}(\hat{\beta}_q) = ([1_n \mathbf{X}]^T \mathbf{W}_q [1_n \mathbf{X}] + \hat{\lambda}_q \mathbf{C})^{-1} [1_n \mathbf{X}]^T \mathbf{W}_q [1_n \mathbf{X}] ([1_n \mathbf{X}]^T \mathbf{W}_q [1_n \mathbf{X}] + \hat{\lambda}_q \mathbf{C})^{-1} \hat{\sigma}_q^2, \quad (7)$$

where $\hat{\sigma}_q^2$ is an estimate of the scale, such as the MAD $\hat{\sigma}_q = \text{median}|r_i(\hat{\beta}_q)|/0.6745$. Note that by construction this method provides zero variance estimates when the estimated coefficients are exactly shrunk to zero. In addition, variance estimates in (7) do not account for the variability due to the selection of the tuning parameter. We investigate the magnitude of the underestimation for these variance estimates in simulated data sets. In particular, we compare them with bootstrap-based variance estimates, as this issue becomes more relevant as we extend the methodology to Elastic Net penalization and penalized B-splines. Bootstrap overcomes both above-mentioned issues, but can be much computationally intensive, particularly for a relatively large data set as the one we deal with. See Supporting Information Appendix A.1.

3.3 | Extension to Elastic Net

In the previous section, we have introduced M-quantile regression with Lasso regularization, which allows us to do continuous shrinkage and automatic variable selection simultaneously. Nevertheless, there are some situations in which the Lasso does not perform well. In particular, (i) when $n < p$, it selects at most n variables; (ii) in case of a group of highly correlated variables, Lasso selects just one of them; (iii) when $n > p$ and there is high correlation between predictors, the prediction performance is poor compared to ridge regression (Zou & Hastie, 2005). In order to improve in such scenarios, Zou and Hastie (2005) propose the Elastic Net, which introduces a new penalization given by a convex linear combination between the Lasso and the Ridge penalization. We extend these concepts here to M-quantile regression.

We define the Elastic Net estimator for M-quantile regression coefficients at $q \in (0, 1)$ as

$$\hat{\beta}_q = \left(\hat{\beta}_{0q}, \hat{\beta}_{1q}^T \right)^T = \underset{\beta_q}{\operatorname{argmin}} \{ L(\mathbf{X}, \mathbf{y}; \beta_q, \lambda_{1q}, \lambda_{2q}) : \beta_{0q} \in \mathbb{R}, \beta_{1q} \in \mathbb{R}^p \}, \tag{8}$$

where the loss function is given by

$$L(\mathbf{X}, \mathbf{y}; \beta_q, \lambda_{1q}, \lambda_{2q}) = \sum_{i=1}^n \psi_q(r_i(\beta_q)/\sigma_q) + \lambda_{1q} \sum_{j=1}^p |\beta_{jq}| + \lambda_{2q} \sum_{j=1}^p \beta_{jq}^2. \tag{9}$$

Solution to (8) can be achieved by a Lasso-type optimization problem. In fact, following lemma 1 of Zou and Hastie (2005), we define a new *augmented* data set of dimension $n^* = n + p$ by

$$\mathbf{X}_{(n+p) \times p}^* = (1 + \lambda_{2q})^{-1/2} \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda_{2q}} \mathbf{I}_p \end{bmatrix} \quad \text{and} \quad \mathbf{y}_{(n+p)}^* = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_p \end{bmatrix},$$

where \mathbf{I}_p is an identity matrix of dimension p . Let $\gamma_q = \lambda_{1q}/\sqrt{1 + \lambda_{2q}}$ and $\beta_q^* = \sqrt{1 + \lambda_{2q}}\beta_q$. Then, the solution to (9) can be obtained equivalently using

$$L(\mathbf{X}^*, \mathbf{y}^*; \beta_q^*, \gamma_q) = \sum_{i=1}^{n^*} \psi_q(r_i(\beta_q^*)/\sigma_q) + \gamma_q \sum_{j=1}^p |\beta_{jq}^*| \tag{10}$$

with $r_i^*(\beta_q^*) = y_i^* - \beta_{0q}^* - \mathbf{x}_i^{*T} \beta_{1q}^*$. In this way, we can solve the Elastic Net penalization as a Lasso-type problem and find

$$\hat{\beta}_q^* = \underset{\beta_q^*}{\operatorname{argmin}} \{ L(\mathbf{X}^*, \mathbf{y}^*; \beta_q^*, \gamma_q) \},$$

by means of the iterative procedure illustrated in the previous section on the augmented data set. We can approximate the covariance matrix of the estimated coefficients $\hat{\beta}_q^*$ with Elastic Net penalization in (8) using the equivalence in (10), so that

$$\hat{V}(\hat{\beta}_q^*) = ([1_{n+p} \mathbf{X}^*]^T \mathbf{W}_q^* [1_{n+p} \mathbf{X}^*] + \hat{\gamma}_q \mathbf{C}^*)^{-1} [1_{n+p} \mathbf{X}^*]^T \mathbf{W}_q^* [1_{n+p} \mathbf{X}^*] ([1_{n+p} \mathbf{X}^*]^T \mathbf{W}_q^* [1_{n+p} \mathbf{X}^*] + \hat{\gamma}_q \mathbf{C}^*)^{-1} \hat{\sigma}_q^2, \tag{11}$$

where 1_{n+p} is an $n + p$ vector of ones, \mathbf{W}_q^* is the $n + p$ diagonal matrix with weights from the last iteration, $\hat{\gamma}_q$ is the final estimate of the shrinkage factor, and $\mathbf{C}^* = \operatorname{diag}[0, 1/|\hat{\beta}_{1q}^*|, \dots, 1/|\hat{\beta}_{pq}^*|]$ is the penalty matrix.

A first set of final estimates $\hat{\beta}_{1qn}$, which we call naive estimates, is given by

$$\hat{\beta}_{1qn} = \frac{1}{\sqrt{1 + \hat{\lambda}_{2q}}} \hat{\beta}_{1q}^*. \tag{12}$$

Naive estimates in (12) can suffer from *double shrinkage* (Zou & Hastie, 2005). For this reason, as it is suggested in Zou and Hastie (2005), we will use as final estimates

$$\hat{\beta}_{1q} = (1 + \hat{\lambda}_{2q}) \hat{\beta}_{1qn}. \tag{13}$$

The estimate of the covariance matrix of the final estimates in (13) can be obtained simply by

$$\hat{V}(\hat{\beta}_{1q}) = (1 + \hat{\lambda}_{2q}) \hat{V}(\hat{\beta}_{1q}^*).$$

The shrinkage parameter λ_{2q} can be selected by cross-validation or by means of GCV if the data set is very large. In particular, the GCV criterion can be written as

$$GCV(\lambda_{2q}) = \frac{\|(I - \mathbf{H}_{\lambda_{2q}}^*)\mathbf{y}^*\|^2}{1 - \theta \text{tr}\{\mathbf{H}_{\lambda_{2q}}^*\}/n^*}, \quad (14)$$

where

$$\mathbf{H}_{\lambda_{2q}}^* = [\mathbb{1}_{n+p} \mathbf{X}^*]([\mathbb{1}_{n+p} \mathbf{X}^*]^T \mathbf{W}_q^* [\mathbb{1}_{n+p} \mathbf{X}^*] + \hat{\gamma}_q \mathbf{C}^*)^{-1} [\mathbb{1}_{n+p} \mathbf{X}^*]^T \mathbf{W}_q^*$$

is similar to a hat matrix based on an approximated ridge regression-type estimator like that in (6) so that its trace provides an approximated value for the effective number of degrees of freedom used. Traditional GCV is obtained when parameter θ is set equal to 1. Values of θ larger (smaller) than one penalize more (less) complex fits and, therefore, provide more (less) shrinkage.

3.4 | Accounting for nonlinearities via B-splines

So far we have assumed a linear M-quantile regression model, that is,

$$MQ_q(y_i | \mathbf{x}_i; \psi_q) = \beta_{0q} + \sum_{j=1}^p \beta_{jq} x_{ij}. \quad (15)$$

In some circumstances, we can face nonlinear relationships between the response and one or more predictors. Therefore, we introduce the possibility to handle nonlinearities. For ease of notation, we treat the case in which only one covariate is included in the model nonparametrically. Extension to the case where more variables are included nonparametrically is straightforward. In particular, we allow a predictor z to have a nonlinear relationship with the response y and we extend model (15) to the following model:

$$MQ_q(y_i | \mathbf{x}_i, z_i; \psi_q) = \beta_{0q} + \sum_{j=1}^p \beta_{jq} x_{ij} + f_q(z_i),$$

where $f_q(z_i)$ represents the unknown functional relationship between y and z at M-quantile q . Here, we assume that the latter can be well approximated by means of the following polynomial spline of degree l ,

$$f_q(z_i) = \sum_{k=1}^K \alpha_{kq} B_k^{(l)}(z_i), \quad (16)$$

where $B_k^{(l)}(z_i)$ are B-splines basis functions (Eilers & Marx, 1996), α_{kq} are the corresponding coefficients, and K is the number of knots. If the number of knots K is sufficiently large, the class of functions in (16) can approximate most smooth functions. Using a large number of knots, however, can lead to an unstable fit. In order to overcome this problem, the influence of the number of knots is limited by choosing K to be reasonably high and by putting a constraint on the variation of the spline coefficients. Using B-splines, a penalty of differences with order l is applied to all of the neighboring parameters. Let $\mathbf{D}^{(l)}$ denote a difference matrix of order l , then the penalty matrix is $\mathbf{P} = \mathbf{D}^{(l)T} \mathbf{D}^{(l)}$ and provides a penalty made of squared l th order differences in the sequence of coefficients. Let \mathbf{B} be the $n \times K$ matrix of basis functions with $\mathbf{B}_i = (B_1^{(l)}(z_i), \dots, B_K^{(l)}(z_i))^T$ on its i th row, and $\boldsymbol{\alpha}_q = (\alpha_{1q}, \dots, \alpha_{Kq})^T$ be the K -vector of corresponding coefficients, then we can extend the Elastic Net loss function in Equation (9) to encompass penalized splines through the following loss function:

$$L(\mathbf{X}, \mathbf{B}, \mathbf{y}; \boldsymbol{\beta}_q, \boldsymbol{\alpha}_q, \lambda_{1q}, \lambda_{2q}, \lambda_{3q}) = \sum_{i=1}^n \psi_q(r_i(\boldsymbol{\beta}_q, \boldsymbol{\alpha}_q)/\sigma_q) + \lambda_{1q} \sum_{j=1}^p |\beta_{jq}| + \lambda_{2q} \sum_{j=1}^p \beta_{jq}^2 + \lambda_{3q} \boldsymbol{\alpha}_q^T \mathbf{P} \boldsymbol{\alpha}_q, \quad (17)$$

where $r_i(\beta_q, \alpha_q) = y_i - \beta_{0q} - \mathbf{x}_i^T \beta_{1q} - \mathbf{B}_i^T \alpha_q$ and λ_{3q} is a parameter that controls the level of smoothness of the spline part of the model. Of course, Lasso regularization can be seen as a particular case in which λ_{2q} is set to zero.

The solution can be obtained again as a Lasso-type optimization problem using the augmented data set introduced in the previous section and introducing

$$\mathbf{B}_{(n+p) \times K}^* = \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix}.$$

Then the solution to (17) can be obtained using

$$L(\mathbf{X}^*, \mathbf{B}^*, \mathbf{y}^*; \beta_q^*, \alpha_q^*, \gamma_q, \lambda_{3q}) = \sum_{i=1}^{n^*} \psi_q(r_i(\beta_q^*)/\sigma_q) + \gamma_q \sum_{j=1}^p |\beta_{jq}^*| + \lambda_{3q} \alpha_q^{*T} \mathbf{P} \alpha_q^* \tag{18}$$

with $r_i^*(\beta_q^*, \alpha_q^*) = y_i^* - \beta_{0q}^* - \mathbf{x}_i^{*T} \beta_{1q}^* - \mathbf{B}_i^{*T} \alpha_q^*$. Now, consider the following unconstrained version of (18):

$$L(\mathbf{X}^*, \mathbf{B}^*, \mathbf{y}^*; \beta_q^*, \alpha_q^*, \lambda_{3q}) = \sum_{i=1}^{n^*} \psi_q(r_i^*(\beta_q^*)/\sigma_q) + \lambda_{3q} \alpha_q^{*T} \mathbf{P} \alpha_q^*. \tag{19}$$

Setting its derivatives with respect to β_q^* and α_q^* to zero yields the following equivalent two systems of equations:

$$[\mathbb{1}_{n+p} \mathbf{X}^* \mathbf{B}^*]^T \mathbf{W}_q^* (\mathbf{y}^* - \beta_{0q}^* \mathbb{1}_{n+p} - \mathbf{X}^* \beta_{1q}^* - \mathbf{B}^{*T} \alpha_q^*) + \lambda_{3q} \begin{bmatrix} \mathbf{0}_{1+p} \\ \mathbf{P} \alpha_q^* \end{bmatrix} = \mathbf{0}_{1+p+K}, \tag{20}$$

and

$$[\mathbb{1}_{n+p} \mathbf{X}^* \mathbf{B}^*]^T \mathbf{W}_q^* \mathbf{X}^* \beta_{1q}^* = [\mathbb{1}_{n+p} \mathbf{X}^* \mathbf{B}^*]^T \mathbf{W}_q^* (\mathbf{y}^* - \beta_{0q}^* \mathbb{1}_{n+p} - \mathbf{B}^{*T} \alpha_q^*). \tag{21}$$

We propose the following procedure to obtain estimates for the two sets of parameters β_q^* and α_q^* :

1. Start with an initial value for $\hat{\beta}_q^*$ and $\hat{\alpha}_q^*$; compute residuals $r_i^*(\hat{\beta}_q^*, \hat{\alpha}_q^*)$ and weight matrix \mathbf{W}_q^* ;
2. Given $\hat{\beta}_{0q}^*$, $\hat{\alpha}_q^*$, and \mathbf{W}_q^* , estimate $\hat{\beta}_{1q}^*$ from Equation (21) subject to $\mathbf{G}_E \beta_{1q}^* \leq t_q \mathbb{1}_e$ by using the constrained M-quantile procedure proposed by Fabrizi et al. (2012); if $\sum_{j=1}^p |\hat{\beta}_{jq}^*| > t_q$, add ℓ to the set E when $\delta_\ell = \text{sign}(\hat{\beta}_{1q}^*)$;
3. Recompute \mathbf{W}_q^* and update estimates for $\hat{\beta}_{0q}^*$ and $\hat{\alpha}_q^*$ from Equation (20);
4. Alternate steps 2 and 3 until convergence.

Note that in this case rescaling the vector $\hat{\alpha}_q^*$ is not needed because the last p rows of matrix $\mathbf{B}_{(n+p) \times K}^*$ are all zeros.

The tuning parameter t_q is selected at step 2 by means of cross-validation, while the smoothing parameter λ_{3q} is chosen at step 3 via GCV as proposed in Pratesi et al. (2009). Note that Elastic Net regularization pertains only \mathbf{x} -variables, while penalization for B -splines is conducted outside the Elastic Net shrinkage. For this reason, the estimation procedure alternates step 2 (Elastic Net shrinkage) and step 3 (penalized spline estimation) to obtain the overall fit. Simple truncated splines can be accommodated by changing the basis functions in matrix \mathbf{B} and the penalization to ridge shrinkage as $\alpha_q^T \mathbf{P} \alpha_q = \sum_{k=1}^K \alpha_{kq}^2$.

When looking at the Elastic Net estimate in the presence of a spline, we can approximate the variance of $\hat{\beta}_q^*$ and $\hat{\alpha}_q^*$ as follows. Let $\mathbf{U}^* = [\mathbb{1}_{n+p} \mathbf{X}^* \mathbf{B}^*]$, then these estimates may be approximated by a ridge regression type-estimator such as

$$\begin{bmatrix} \hat{\beta}_q^* \\ \hat{\alpha}_q^* \end{bmatrix} = (\mathbf{U}^{*T} \mathbf{W}_q^* \mathbf{U}^* + \mathbf{G})^{-1} \mathbf{U}^{*T} \mathbf{W}_q^* \mathbf{y}^*, \tag{22}$$

where $\mathbf{G} = \text{blockdiag}\{\hat{\gamma}_q \mathbf{C}^*, \hat{\lambda}_{3q} \mathbf{P}\}$, then

$$\hat{V}(\hat{\beta}_q^*, \hat{\alpha}_q^*) = (\mathbf{U}^{*T} \mathbf{W}_q^* \mathbf{U}^* + \mathbf{G})^{-1} \mathbf{U}^{*T} \mathbf{W}_q^* \mathbf{U}^* (\mathbf{U}^{*T} \mathbf{W}_q^* \mathbf{U}^* + \mathbf{G})^{-1} \hat{\sigma}_q^2. \tag{23}$$

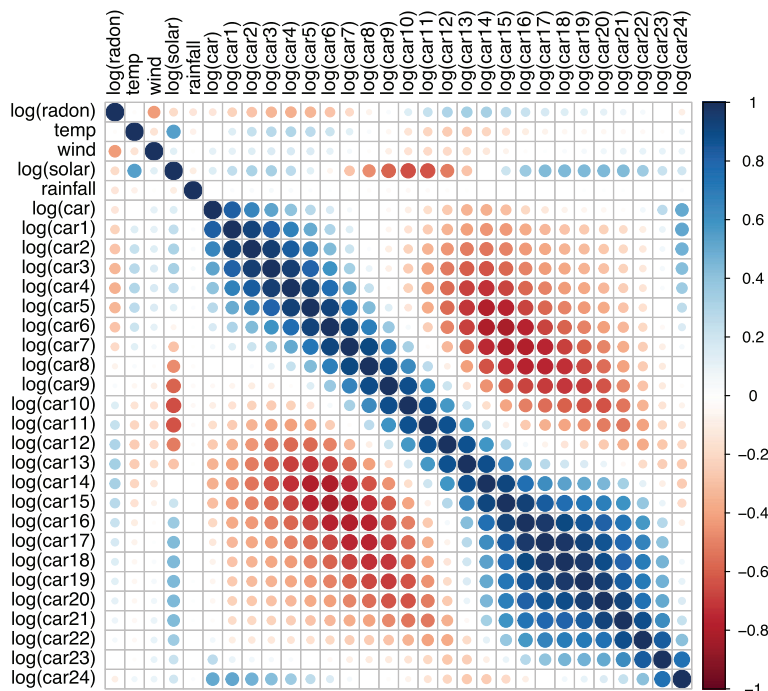


FIGURE 1 Correlation between meteorological and vehicular traffic variables.

The GCV criterion for the smoothing parameter can be written similarly to (14) as

$$GCV(\lambda_{3q}) = \frac{\|(I - \mathbf{H}_{\lambda_{3q}}^*)\mathbf{y}^*\|^2}{1 - \theta \text{tr}\{\mathbf{H}_{\lambda_{3q}}^*\}/n^*}, \tag{24}$$

where

$$\mathbf{H}_{\lambda_{3q}}^* = \mathbf{U}^* \left(\mathbf{U}^{*T} \mathbf{W}_q^* \mathbf{U}^* + \mathbf{G} \right)^{-1} \mathbf{U}^{*T} \mathbf{W}_q^*$$

is a combination of a hat and a smoother matrix based on the same approximated ridge regression-type estimator in (22) so that its trace provides an approximated value for the effective overall (covariate and spline components) number of degrees of freedom used.

4 | APPLICATION

The proposed modeling approach has been applied to the PMetro data presented in Section 2. Indeed, we are interested in investigating how meteorological conditions and vehicular traffic levels affect fine PM concentrations in Perugia at different levels of the conditional response distribution. Since a large number of traffic predictors is available and multicollinearity can be envisioned among them (see Figure 1), Elastic Net regularization is employed in order to perform variable selection and estimation at the same time. In particular, the response variable is given by fine PM concentration on the log scale, while predictors can be grouped into two sets: meteorological data and vehicular traffic data. The first group includes radon concentration on the log-scale, temperature, wind speed, total solar radiation on log-scale and rainfall; the second group includes a set of 25 count variables, $car_0 - car_{24}$, that measure the number of vehicles that passed at that hour, car_0 , 1 hour before, car_1 , 2 hours before, car_2 , up to 24 hours before, car_{24} . These variables are included on the log-scale. The log-transformation is used to obtain an approximately equal spread of the variables in the model and an approximately linear relationship with the response when possible. Moreover, a set of 13 dummy variables is also added to control for the hour of the day of the observation, $h_7 - h_{20}$, where h_7 denotes 7 a.m. and h_{20} denotes 8 p.m. The reference hour is 6 a.m. Finally, in order to capture unobserved heterogeneity among days, a smooth function of the Julian day is added to the model as in Bertaccini et al. (2012). We expect this variable to have a nonlinear relationship with the response,

while for the other variables a linear relationship is adequate (see results in Crocchianti et al., 2020; Del Sarto et al., 2019). Temperature is the only variable for which some curvature has been detected in similar data sets, but a quadratic term is not found to be significant in this context.

The model can be summarized as follows:

$$\begin{aligned}
 MQ_q\{\log(PM)\} = & \beta_{0q} + \log(radon)\beta_{1q} + temp\beta_{2q} + wind\beta_{3q} + \log(solar)\beta_{4q} + rainfall\beta_{5q} \\
 & + \log(car_0)\beta_{6q} + \log(car_1)\beta_{7q} + \log(car_2)\beta_{8q} + \dots + \log(car_{24})\beta_{30q} \\
 & + h_7\beta_{31q} + h_8\beta_{32q} + \dots + h_{20}\beta_{44q} + f_q(Julian\ day).
 \end{aligned}$$

The model is fitted at different levels of $q = \{0.1, 0.2, \dots, 0.9\}$. The cutoff constant c is set to 1.345 to grant robustness with respect to a few outlying observations. At each level of q , the penalization parameter of the ridge part of the model λ_{2q} is selected by GCV, the tuning parameter for Lasso is selected by means of cross-validation, and the smoothing parameter λ_{3q} for B-splines is selected using GCV, as illustrated in Section 3. For each level of q , the estimation procedure requires 3.5 h on a Intel Xeon CPU E5-2420 @ 1.90 GHz server. In particular, the following estimates are obtained for λ_{2q} at each value of q , $\hat{\lambda}_{2q} = \{0.00, 0.20, 0.25, 0.35, 0.30, 0.30, 0.30, 0.20, 0.05\}$.

Variance estimates are obtained by means of the analytic formula in (23) and by means of bootstrap. In particular, we draw a nonparametric bootstrap sample of the same size randomly with replacement from our original data set. Then, using the bootstrap sample we estimate the model again by using the same configuration as for the original model. That is, we use the final estimates of the tuning parameters \hat{t}_q , $\hat{\lambda}_{2q}$, and $\hat{\lambda}_{3q}$ of the original model. We repeat the procedure 1000 times. We follow this approach to save computing time (Spiegel et al., 2021, also use this type of bootstrap for flexible expectile regression with large data sets). Estimating the tuning parameters in each bootstrap sample would allow to consider also the uncertainty of those. However, with the current algorithm on the original data set, optimizing 1000 times the tuning parameters is prohibitive. Moreover, for B-splines, it cannot be ensured that the estimated curve of the main model always lays inside the bootstrap interval. Numerical illustrations in Supporting Information Appendix A.1 show that the underestimation due to keeping the tuning parameters fixed is acceptable, so that this approach is a viable option when dealing with large data sets.

In Table 2, we report the estimated coefficients at $q = 0.5$, together with standard error estimates obtained using bootstrap and the analytic formula. The meteorological predictors play an important role on fine PM concentration; the sign and magnitude of the estimated coefficients are in line with those from the literature and those obtained in Del Sarto et al. (2019). In particular, radon concentration accounts for the most impact. Radon concentration can be used as a proxy of the planetary boundary layer (PBL), which represents the lowest part of the atmosphere. Its height is important in dispersion of pollutants and particles, since the lower the PBL, the higher the particle concentration. Radon products can be considered as natural tracers of the low PBL layers mixing properties: the higher the radon concentration, the lower the PBL. The estimate of radon concentration takes value 1.059; this implies that if it increases (and hence the PBL decreases) by 10%, then the M -median fine PM concentration increases by approximately 11% (se 3%). Wind is inversely related while total solar radiation is positively related to fine PM, as expected. Rainfall and temperature are discarded from the model. Estimates of the effect of the hour of the day are all retained in the model and all positive, with values that increase over the course of the day, by this providing evidence of an accumulation phenomenon. Finally, we can notice that the model performs heavy selection of vehicular traffic variables. Indeed, among all the 25 vehicular traffic predictors, only seven are retained— $\log(car_0)$, $\log(car_5)$, $\log(car_8)$, $\log(car_{11})$, $\log(car_{17})$, $\log(car_{20})$, and $\log(car_{24})$ —with a positive effect in all cases but for $\log(car_{24})$. Indeed, only the association with the concurrent number of vehicles, car_0 , is strong enough (significant also looking at the corresponding standard error). The negative association detected with car_{24} is only marginally significant and can be likely due to a cyclical effect coming from the day of the week.

Figure 2 displays the B-spline estimate of the effect of the day of the year together with bootstrap-based and analytic 95% confidence bounds. The estimated function follows a seasonal pattern for which there are relatively higher levels of fine PM concentrations during wintertime, reflecting confounders like social processes such as heating, and then decreasing in summertime. In the case in which the data covered a multiyear period, then a periodic spline, such as those illustrated in Eilers and Marx (2010) should be considered to describe the trend within each solar year.

Figure 3 provides estimated coefficients for meteorological covariates at different levels of q , together with the corresponding 95% confidence intervals. Figures A.5 and A.6 in the Supporting Information Appendix show similar plots for vehicular traffic and hour-of-the-day covariates, respectively. The value of the intercept is increasing with q , temperature

TABLE 2 Parameter estimates and corresponding estimated standard errors by bootstrap (se_b) and analytic formula (se_{an}) at $q = 0.50$. Note that the analytical s.e. estimates are by definition zero when the estimated coefficient is zero.

Predictor	Estim.	se_b	se_{an}	Predictor	Estim.	se_b	se_{an}
Intercept	11.218	0.128	0.079	$\log(car_{17})$	0.232	0.119	0.065
$\log(radon)$	1.059	0.033	0.024	$\log(car_{18})$	0.000	0.095	-
temp	0.000	0.001	0.000	$\log(car_{19})$	0.000	0.106	-
wind	-0.165	0.044	0.022	$\log(car_{20})$	0.134	0.139	0.091
$\log(solar)$	0.147	0.038	0.024	$\log(car_{21})$	0.000	0.132	-
rainfall	0.000	0.063	0.000	$\log(car_{22})$	0.000	0.126	-
$\log(car_0)$	0.193	0.103	0.049	$\log(car_{23})$	0.000	0.123	-
$\log(car_1)$	0.000	0.100	-	$\log(car_{24})$	-0.160	0.091	0.066
$\log(car_2)$	0.000	0.072	-	h_7	0.151	0.095	0.125
$\log(car_3)$	0.000	0.061	-	h_8	0.261	0.119	0.148
$\log(car_4)$	0.000	0.047	-	h_9	0.189	0.111	0.177
$\log(car_5)$	0.076	0.056	0.037	h_{10}	0.231	0.124	0.208
$\log(car_6)$	0.000	0.050	-	h_{11}	0.156	0.106	0.241
$\log(car_7)$	0.000	0.052	-	h_{12}	0.185	0.122	0.264
$\log(car_8)$	0.062	0.077	0.036	h_{13}	0.276	0.130	0.278
$\log(car_9)$	0.000	0.053	-	h_{14}	0.316	0.142	0.282
$\log(car_{10})$	0.000	0.056	-	h_{15}	0.407	0.157	0.270
$\log(car_{11})$	0.068	0.057	0.044	h_{16}	0.635	0.161	0.264
$\log(car_{12})$	0.000	0.048	-	h_{17}	0.975	0.163	0.257
$\log(car_{13})$	0.000	0.050	-	h_{18}	1.306	0.178	0.249
$\log(car_{14})$	0.000	0.049	-	h_{19}	1.557	0.211	0.264
$\log(car_{15})$	0.000	0.064	-	h_{20}	1.675	0.258	0.285
$\log(car_{16})$	0.000	0.074	-				

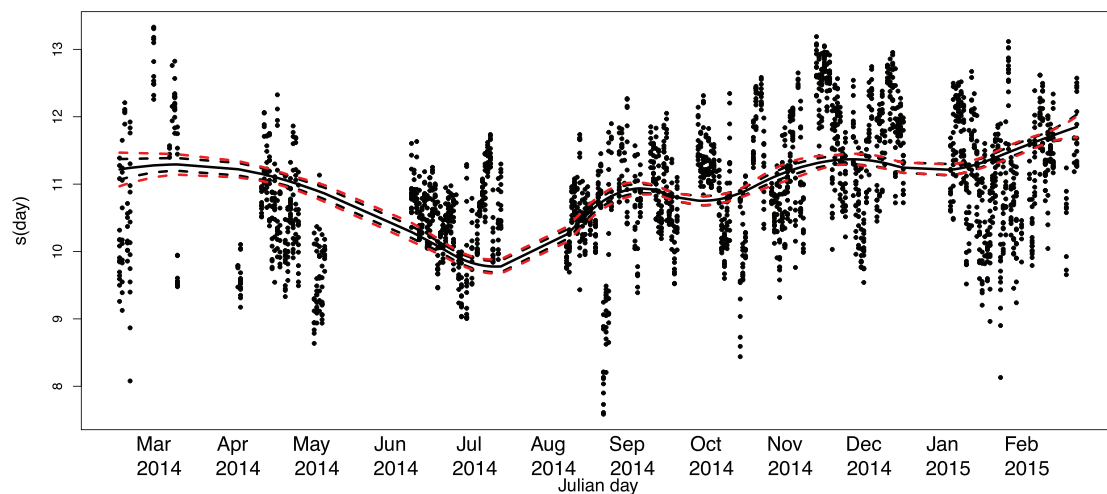
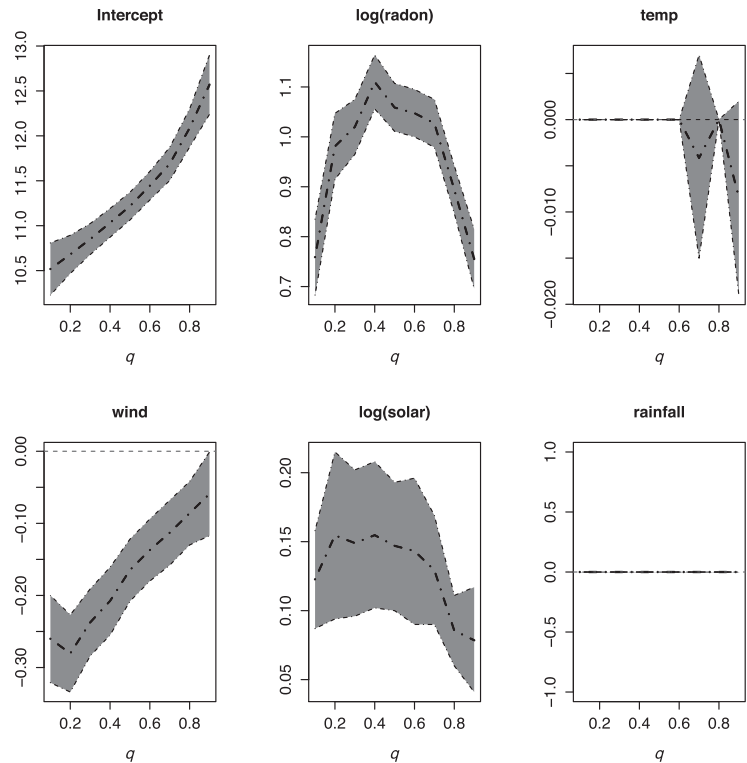


FIGURE 2 Estimate of the effect of the Julian day and corresponding 95% confidence interval at $q = 0.5$. Dashed red: bootstrap-based confidence interval; dashed black: analytic formula.

and rainfall remain nonsignificant along all the values of q . Wind speed is always negative and significantly different from zero, and decreases in absolute value with q . This implies that higher M-quantiles of the distribution of fine PM counts depend less strongly on wind speed.

Radon concentration is always significantly positive and displays an inverse U-shape, with a peak at $q = 0.4$. This, implies that central M-quantiles of the distribution of fine PM counts, between 0.3 and 0.7, depend more strongly on

FIGURE 3 Estimated coefficients for the meteorological predictors at different levels of q , with 95% confidence intervals obtained by analytic standard error (SE) estimates.



radon concentration than the lower and the upper tails. The effect of concurrent vehicular traffic, car_0 , is estimated to be zero for $q = 0.1$, and is significantly positive at all other values of q . The estimates display a relatively constant pattern with respect to q . This implies that the relative impact of an increase in vehicular traffic remains the same for lower and for higher M-quantiles of the distribution of fine PM. The features observed at $q = 0.5$ for the other traffic-related variables remain similar also at the other values of q .

Supporting Information Appendix A.4 reports a set of figures that compare the estimates obtained with the proposed procedure with those obtained using the Elastic Net quantile regression estimates as proposed by Yi and Huang (2017) and implemented in the `hqreg` R function with `method = "quantile"` for $\tau = \{0.1, 0.2, \dots, 0.9\}$. Results appear to be different for some effects and these differences are due to at least two main reasons. First, note that Yi and Huang (2017) approach, as implemented with `hqreg, method = "quantile"`, models the quantiles of the conditional distribution of the response, which can be very different from the M-quantiles, which are essentially robust expectiles. Second, `hqreg` does not allow for nonparametric regression, that is, to approximate the effect of one or more covariates via B-splines. Therefore, for the effect of the calendar day we have included a polynomial term of order 4 in `hqreg`, but the final approximation is essentially linear (see Figure A.7 in the Supporting Information Appendix) and this, of course, changes the effect of the other covariates, such as temperature (see Figure A.8 in the Supporting Information Appendix) and the effect of vehicular traffic (see Figure A.9 in the Supporting Information Appendix). The latter in particular, seems to be better captured by the proposed method as the approach of Yi and Huang (2017) sets the effect of concurrent vehicular traffic, car_0 , always equal to zero and this is less coherent with PM dynamics and previous results on the same data.

5 | CONCLUSIONS

In this paper, we analyze data on air quality measured by concentration of fine PM in order to assess the contribution associated with vehicular traffic. Vehicular traffic can be measured by different variables and can have a different effect at different locations of the response distribution. For this reason, we introduce M-quantile regression models with Lasso and Elastic Net penalizations in order to conduct model estimation and selection at the same time for different levels of the distribution of fine PM and to be robust to outlying observations. We use data for a whole year in order to account for seasonal trends and include meteorological covariates and other confounders to try and isolate the effect of traffic. The

effect of the day of the year is included nonparametrically and estimated from the data by means of penalized B-splines in order to enhance flexibility of the predictive part of the model.

The effect of traffic for this type of data has usually been evaluated in two steps: first, the best proxy is detected via model selection by including one variable at a time, then the effect of such a detected variable is included in the model, and often left unchanged at different locations of the distribution of fine PM concentration. See, for example, Del Sarto et al. (2019). Introducing Lasso and Elastic Net for M-quantile regression allows to perform these two steps simultaneously and to detect possible differences at different levels of q . In addition, using Elastic Net allows to handle predictors that display a relevant collinearity, as it is the case with the data at hand. Including all traffic-related variables, we identify a major driver given by the number of concurrent vehicles. The relative effect of traffic is essentially invariant across values of q , that is, traffic affects equally the low and high M-quantiles of fine PM concentration. As PM concentration is modeled on a log-scale, this implies that the effect is higher in absolute terms for the upper tail of the distribution of fine PM.

With respect to the existing methodological literature, our paper fills in a gap: Yi and Huang (2017) introduce Elastic Net for the Huber loss function for $q = 0.50$, our paper extends it to any value of q . In addition, to better model the data at hand, we allow for the possibility of including covariates nonparametrically via B-splines. Finally, we provide bootstrap-based and analytical variance estimates for regression coefficients and for B-splines function approximations.

Despite the wide range of topics covered in this paper, some research problems remain open. For example, our proposed toolkit use an extension of the algorithm proposed by Tibshirani (1996). Developing algorithms for Lasso and Elastic Net estimators for M-quantile regression coefficients based on gradient descent such as that used in Yi and Huang (2017) for robust regression at $q = 0.5$ remains an open research problem that would improve on computational time considerably. Finally, an efficient implementation of the proposed procedures could be provided in a R package, which could include a main function for model fitting, and a variety of auxiliary functions for summary and plotting. This is left for future research.

ACKNOWLEDGMENTS

The work of Ranalli has been carried out with the support of the project AIDMIX, Fondo di ricerca di Ateneo, edizione 2021, Universita degli Studi di Perugia. The work of Salvati has been carried out with the support of the project InGRID 2 (Grant Agreement N. 730998) and of the project LOCOMOTION (Grant Agreement N. 821105). The authors are grateful to the PMetro project for providing the data and to David Cappelletti for useful discussions.


CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the [supplementary material](#) of this article.

OPEN RESEARCH BADGES

 This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the [Supporting Information](#) section.

This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

M. Giovanna Ranalli  <https://orcid.org/0000-0001-7005-8572>

Nicola Salvati  <https://orcid.org/0000-0002-4160-9387>

Francesco Pantalone  <https://orcid.org/0000-0002-7943-7007>

REFERENCES

- Aghamohammadi, A., & Mohammadi, S. (2017). Bayesian analysis of penalized quantile regression for longitudinal data. *Statistical Papers*, 58(4), 1035–1053.
- Alfò, M., Salvati, N., & Ranalli, M. G. (2017). Finite mixtures of quantile and M-quantile regression models. *Statistics and Computing*, 27(2), 547–570.
- Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive lasso quantile regression. *Statistical Modelling*, 12(3), 279–297.

- Belloni, A., & Chernozhukov, V. et al. (2011). L1-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics*, 39(1), 82–130.
- Bernardi, M., Bottone, M., & Petrella, L. (2018). Bayesian quantile regression using the skew exponential power distribution. *Computational Statistics & Data Analysis*, 126, 92–111.
- Bertaccini, P., Dukic, V., & Ignaccolo, R. (2012). Modeling the short-term effect of traffic and meteorology on air pollution in turin with generalized additive models. *Advances in Meteorology*, 2012, Article 609328:1–16.
- Bianchi, A., Fabrizi, E., Salvati, N., & Tzavidis, N. (2018). Estimation and testing in M-quantile regression with applications to small area estimation. *International Statistical Review*, 86(3), 541–570.
- Breckling, J., & Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4), 761–771.
- Castellini, S., Moroni, B., & Cappelletti, D. (2014). PMetro: Measurement of urban aerosols on a mobile platform. *Measurement*, 49, 99–106.
- Crocchianti, S., Del Sarto, S., Ranalli, M. G., Moroni, B., Castellini, S., Petroselli, C., & Cappelletti, D. (2020). Spatiotemporal correlation of urban pollutants by long-term measurements on a mobile observation platform. *Environmental Pollution*, 268, 115645.
- Del Sarto, S., Marino, M. F., Ranalli, M. G., & Salvati, N. (2019). Using finite mixtures of M-quantile regression models to handle unobserved heterogeneity in assessing the effect of meteorology and traffic on air quality. *Stochastic Environmental Research and Risk Assessment*, 33(7), 1345–1359.
- Del Sarto, S., Ranalli, M. G., Bakar, K. S., Cappelletti, D., Moroni, B., Crocchianti, S., Castellini, S., Spataro, F., Esposito, G., Ianniello, A., & Salvatori, R. (2016). Bayesian spatiotemporal modeling of urban air pollution dynamics. In *Topics on methodological and applied statistical inference* (pp. 95–103). Springer.
- Del Sarto, S., Ranalli, M. G., Cappelletti, D., Moroni, B., Crocchianti, S., & Castellini, S. (2016). Modelling spatio-temporal air pollution data from a mobile monitoring station. *Journal of Statistical Computation and Simulation*, 86(13), 2546–2559.
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89–121.
- Eilers, P. H., & Marx, B. D. (2010). Splines, knots, and penalties. *Wiley interdisciplinary reviews: Computational Statistics*, 2(6), 637–653.
- Fabrizi, E., Salvati, N., & Pratesi, M. (2012). Constrained small area estimators based on M-quantile methods. *Journal of Official Statistics*, 28(1), 89–106.
- Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6(9), 813–827.
- Huber, P. J. (1981). *Robust statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc, New York.
- Jones, M. C. (1994). Expectiles and M-quantiles are quantiles. *Statistics and Probability Letters*, 20, 149–153.
- Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, 13, 275–303.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1), 74–89.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.
- Koenker, R. W., & D'Orey, V. (1987). Computing regression quantiles. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 36(3), 383–393.
- Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis*, 5(2), 369–411.
- Li, Q., Xi, R., & Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3), 533–556.
- Li, Y., & Zhu, J. (2008). L1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1), 163–185.
- Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7), 2372–2387.
- Newey, W., & Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55, 819–847.
- Otto-Sobotka, F., Salvati, N., Ranalli, M. G., & Kneib, T. (2019). Adaptive semiparametric M-quantile regression. *Econometrics and Statistics*, 11, 116–129.
- Petrella, L., & Raponi, V. (2019). Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. *Journal of Multivariate Analysis*, 173, 70–84.
- Pratesi, M., Ranalli, M. G., & Salvati, N. (2009). Nonparametric M-quantile regression using penalised splines. *Journal of Nonparametric Statistics*, 21(3), 287–304. <https://doi.org/10.1080/10485250802638290>
- Ranalli, M. G., Rocco, G., Jona Lasinio, G., Moroni, B., Castellini, S., Crocchianti, S., & Cappelletti, D. (2016). Functional exploratory data analysis for high-resolution measurements of urban particulate matter. *Biometrical Journal*, 58(5), 1229–1247.
- Slawski, M. (2012). The structured elastic net for quantile regression and support vector classification. *Statistics and Computing*, 22(1), 153–168.
- Spiegel, E., Kneib, T., von Gablenz, P., & Otto-Sobotka, F. (2021). Generalized expectile regression with flexible response function. *Biometrical Journal*, 63(5), 1028–1051.
- Spiegel, E., Otto-Sobotka, F., & Kneib, T. (2017). Model selection in semiparametric expectile regression. *Electronic Journal of Statistics*, 11(2), 3008–3038.
- Tang, X., & Lian, H. (2016). Mean and quantile boosting for partially linear additive models. *Statistics and Computing*, 26(5), 997–1008.
- Tian, Y., Shen, S., Lu, G., Tang, M., & Tian, M. (2019). Bayesian lasso-regularized quantile regression for linear regression models with autoregressive errors. *Communications in Statistics - Simulation and Computation*, 48(3), 777–796.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
- Tzavidis, N., Salvati, N., Schmid, T., Flouri, E., & Midouhas, E. (2016). Longitudinal analysis of the strengths and difficulties questionnaire scores of the Millennium Cohort Study children in England using M-quantile random-effects regression. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179, 427–452.

- Waldmann, E., Sobotka, F., & Kneib, T. (2017). Bayesian regularisation in geoadditive expectile regression. *Statistics and Computing*, 27(6), 1539–1553.
- Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, 107(497), 214–222.
- Xu, Q., Ding, X., Jiang, C., Yu, K., & Shi, L. (2021). An elastic-net penalized expectile regression with applications. *Journal of Applied Statistics*, 48(12), 2205–2230.
- Yan, A., & Song, F. (2019). Adaptive elastic net-penalized quantile regression for variable selection. *Communications in Statistics - Theory and Methods*, 48(20), 5106–5120.
- Yi, C., & Huang, J. (2017). Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, 26(3), 547–557.
- Zhang, Y., Lian, H., & Yu, Y. (2017). Estimation and variable selection for quantile partially linear single-index models. *Journal of Multivariate Analysis*, 162, 215–234.
- Zhao, J., & Zhang, Y. (2018). Variable selection in expectile regression. *Communications in Statistics - Theory and Methods*, 47(7), 1731–1746.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B - Statistical Methodology*, 67, 301–320.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ranalli, M. G., Salvati, N., Petrella, L., & Pantalone, F. (2023). M-quantile regression shrinkage and selection via the Lasso and Elastic Net to assess the effect of meteorology and traffic on air quality. *Biometrical Journal*, e2100355. <https://doi.org/10.1002/bimj.202100355>