

Prediction of composite indicators using locally weighted quantile regression

Jurga Rukšėnaitė^a, Pranas Vaitkus^b, Povilas Asijavičius^b

^aVilnius Gediminas Technical University,
Saulėtekio ave. 11, LT-10223 Vilnius, Lithuania
jurga.ruksenaite@gmail.com

^bVilnius University,
Naugarduko str. 24, LT-03225 Vilnius, Lithuania
vaitkuspranas@gmail.com

Received: December 18, 2016 / **Revised:** November 17, 2017 / **Published online:** December 14, 2017

Abstract. The main goal of this paper is to improve the existing methods and tools used for solving penalized quantile regression problems. We modified the quantile regression method by implementing the extreme learning machine (ELM) algorithm and features of locally weighted regression. Also, we used different penalty functions. A modified method was used for the one-step-ahead prediction of the composite indicator (CI) of the Lithuanian economy. Our analysis showed that the prediction error of the modified locally weighted quantile regression is smaller in comparison to the other quantile regression.

Keywords: quantile regression, penalty function, extreme learning machine, locally weighted regression, composite indicators.

1 Introduction

Linear regression methods are well known as classical methods; the ordinary least square (OLS) method allows us to assess the linear relationship of variables. This method finds the conditional mean of the response variable. On the other hand, the OLS method is very sensitive to outliers; hence, it provides unstable estimates. Also applicable is the LADR (least absolute deviation regression), which evaluates the median of the response variable. The median is more robust to outliers; hence, this method is more robust to outliers as well. LADR was summarized in [4] when developing the quantile regression. In this way, the conditional quantile of any order was obtained; practically, this fact characterizes the conditional distribution of a response variable. Over the past thirty-five years, quantile regression has become a very important and widely used technique to analyze the whole conditional distribution of a response variable. Such methods are more robust to outliers than that of the linear regression.

The first step of any regression modeling is the selection of variables. In practical problems, there are situations where quite a few variables are important. Usually, most

of them are included in the model, and only then are the most important ones chosen. Insignificant variables should not be included in the model since they complicate interpretation of results and reduce the accuracy of predictions. In order to automate this process, the regularization of the method is often applied. Many versions of regularization are proposed. The L_1 regularization is used in the LASSO (least absolute shrinking and selection operator) method [11]. Later, the LASSO method was improved into the adaptive LASSO method [13]. The nonconcave penalized least squares regression was introduced [2]. This method selects the significant variables and evaluates the estimates of coefficients simultaneously. An example of such a function is SCAD (smoothly clipped absolute deviation) regularization. Also, logarithmic and exponential regularizations were introduced. Later, the MCP (minimax concave penalty) function was developed [12]. In most cases, these above-mentioned penalty functions were used in solving linear regression problems, yet almost analogous methods can be used in quantile regression methods.

This work is an extension of previous work, where the ELM algorithm was combined with the locally weighted regression for the one-step-ahead prediction of the composite indicator (CI) of the Lithuanian economy [8]. An analysis of results showed that the combined method gives a smaller prediction error in comparison to the Levenberg–Marquard, ELM methods or $AR(p)$ process. Also, the analysis of the results based on various accuracy measures suggested that the proposed method may be used for data of rather small sample size and during periods when dynamics of time series may have unexpected changes like during the economical crises and later periods (2008–2010). In spite of the acceptable results, as mentioned in the beginning of the paper, it is known that the OLS method is sensitive to outliers. Hence, the quantile regression was chosen as an alternative (that is, more robust to outliers) for the one-step-ahead prediction of CI.

In practical examples, we will use the CI of the Lithuanian economy that was developed under the methodology described in [7]. The methodology for constructing the CI is based on factor analysis. In this paper, the practical problems of quantile regression are solved using the R package *rqPen* [9]. In this package, different regularization functions are implemented: LASSO, SCAD, and MCP. This tool enables us to solve quantile regression problems using these above-mentioned regularization functions.

The general objective of this research is to modify the existing quantile regression: (i) to develop the locally weighted quantile regression using different regularization functions and the ELM method for the modification of input data; (ii) to check the impact of different regularization functions on the predicted results of CI's.

The structure of this paper is as follows: In Section 1, methodological notes are introduced. The practical implementation of the methods is described in Section 3. Section 4 describes the results, and Section 5 gives concluding remarks.

2 Methodology

In this section, we define our main terms and introduce the methods used in this paper: the quantile regression with regularization, regularization functions, the locally weighted quantile regression, and the locally weighted quantile regression with the ELM modification.

2.1 Quantile regression with regularization

Suppose we have the random variables (r.v.) Y and X . Also suppose the Y 's distribution function with the condition that $X = x$ is $F_Y(y | X = x) = \mathbf{P}(Y \leq y | X = x)$. Then Y 's $\tau \in [0, 1]$ order quantile with condition $X = x$ is defined as a number:

$$Q_{Y|X}(\tau) = F_Y^{-1}(\tau | X = x) = \inf\{y: F_Y(y | X = x) \geq \tau\}.$$

Function $f_\tau(\mathbf{x})$ is called the τ th order (where $0 < \tau < 1$) quantile function if: $\mathbf{P}(Y < f_\tau(\mathbf{x}) | X) = \tau$.

We define the $r = y - f_\tau(\mathbf{x} | \beta)$, then the ‘‘check’’ function:

$$\rho_\tau(r) = \begin{cases} \tau r & \text{if } r > 0, \\ -(1 - \tau)r & \text{otherwise.} \end{cases} \quad (1)$$

Such a ‘‘check’’ function (1) can be used for the evaluation of the quantiles of r.v.

Suppose that Y is r.v. with distribution function $F_Y(y) = \mathbf{P}(Y \leq y)$. Then the quantile of τ order is equal

$$Q_Y(\tau) = \operatorname{argmin}_u \mathbf{E}(\rho_\tau(Y - u)).$$

Having the realization y_1, \dots, y_N of r.v. Y , we evaluate the estimate of τ order quantile of r.v. Y :

$$\begin{aligned} \widehat{Q}_Y(\tau) &= \operatorname{argmin}_u \left\{ \frac{1}{N} \sum_{i=1}^N \rho_\tau(y_i - u) \right\} \\ &= \operatorname{argmin}_u \left\{ \frac{1}{N} \sum_{i=1}^N (\tau(y_i - u)\mathbf{I}(y_i > u) + (1 - \tau)(u - y_i)\mathbf{I}(y_i \leq u)) \right\} \\ &= \operatorname{argmin}_u \widehat{\mathbf{E}}(\rho_\tau(Y - u)). \end{aligned}$$

Moving to the quantile regression, suppose the argument u is dependent on the r.v. X , i.e., $u = f_\tau(X)$. Then we find the Y 's conditional quantile $Q_{Y|X}(\tau)$ of order τ .

Analogically, if we have realizations $\{(\mathbf{x}_i, y_i), i = 1, \dots, N\}$, where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$, we can solve the regression problem between Y and X using the ‘‘check’’ function (1). In this case, we will get the estimate $\widehat{Q}_{Y|X}(\tau)$ of τ order conditional quantile of the response variable. In another way, we have this minimization problem:

$$\min_{f_\tau} \sum_{i=1}^N \rho_\tau(y_i - f_\tau(\mathbf{x}_i | \beta)). \quad (2)$$

When similar minimization problems are solved, regularization method is often used. Often, the regularization functions L_1 and L_2 are used. Then we obtain the LASSO and ridge regression.

Hence, in order to get better results for the quantile regression problem, we shall add the penalty function $J(f_\tau)$:

$$\min_{f_\tau} \sum_{i=1}^N \rho_\tau(y_i - f_\tau(\mathbf{x}_i | \boldsymbol{\beta})) + \lambda J(f_\tau), \quad (3)$$

here $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$, and the regularization parameter λ is carefully chosen.

2.2 Regularization functions

The regularization technique permits us to choose only significant variables for the regression; also, it helps us avoid overfitting. In this paper, we analyzed three different techniques: LASSO, SCAD, MPC.

LASSO method was defined in order to improve the OLS method [11]. Formerly, two main approaches were employed: ridge regression and the exclusion of insignificant variables from the model. Ridge regression gives fewer estimates of stabler coefficients. The second approach rejects irrelevant covariates and makes the results easier to interpret. LASSO combines features of both methods; hence, this method has its drawbacks. The LASSO regularization function shall be defined:

$$J(f_\tau) = \lambda J(\boldsymbol{\beta}) = \lambda \sum_{j=1}^p |\beta_j|,$$

here $\sum_{j=1}^p |\beta_j| \leq t$, t is a parameter that controls the shrinkage of coefficients.

SCAD function was defined in [2]. This regularization function is symmetric and nonconcave on the interval $(0, \infty)$:

$$p_\lambda(\beta_j) = \begin{cases} \lambda |\beta_j| & \text{if } |\beta_j| \leq \lambda, \\ -\left(\frac{|\beta_j|^2 - 2a\lambda|\beta_j| + \lambda^2}{2(a-1)}\right) & \text{if } \lambda < |\beta_j| \leq a\lambda, \\ \frac{(a+1)\lambda^2}{2} & \text{if } |\beta_j| > a\lambda. \end{cases}$$

Hence, the SCAD regularization function is differentiable in the interval $(-\infty, 0) \cup (0, +\infty)$, but its derivative outside of the interval $[-a\lambda, a\lambda]$ is equal to zero. SCAD should give unbiased estimates of coefficients and exclude insignificant variables.

In [2], there is a recommendation to use $\alpha = 3.7$, while λ is often chosen by cross-validation or other methods.

MCP function can be used in solving linear regression problems. The estimates obtained using MCP are accurate and almost unbiased. This algorithm can be used in solving the large dimension multiple regression [12]. MCP regularization can be defined:

$$p(t; \lambda) = \lambda \int_0^{|t|} \left(1 - \frac{x}{\gamma\lambda}\right)_+ dx = \left(\lambda|t| - \frac{t^2}{2\gamma}\right) \mathbf{I}(|t| < \lambda\gamma) + \frac{\lambda^2\gamma}{2} \mathbf{I}(|t| \geq \lambda\gamma),$$

here $\gamma > 0$ and $\lambda > 0$ are regularization parameters.

Regularization function $p(t; \lambda)$ with parameter t is nonincreasing in the interval $(-\infty, 0)$ and nondecreasing in the interval $(0, \infty)$, its derivative $p(t; \lambda)'_t$ is continuous in the above-mentioned intervals.

2.3 Locally weighted quantile regression

The goal of the regression is to evaluate the function m in $\mathbf{E}(Y | X) = m(X)$ having the response variable Y and independent variable X .

In the linear dependency between Y and X and having realization of these random variables $\{(\mathbf{x}_i, y_i), i = 1, \dots, N\}$, the linear regression is defined as $y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \epsilon_i$, here $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ and ϵ_i is error. In this case, $y_i = m(\mathbf{x}_i) + \epsilon_i$, where $m(\mathbf{x}_i) = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$. Coefficients β_1, \dots, β_p will be evaluated solving the minimization problem

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^N (y_i - \boldsymbol{\beta}^T \mathbf{x}_i)^2. \quad (4)$$

In general, the results of such a minimization may be improved. The regression results obtained are usually applied to evaluate the prediction regarding the latest values of the independent variable $\hat{y}_{N+1} = \boldsymbol{\beta}^T(\mathbf{x}_{N+1})$. If \mathbf{x}_{N+1} is different from other \mathbf{x}_i , the $\boldsymbol{\beta}^T$ obtained may be insufficiently accurate.

Hence, in this paper, the locally weighted regression shall be used [5]. It modifies expression (4), which gives weights to the errors ω_i :

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^N \omega_i (y_i - \boldsymbol{\beta}^T \mathbf{x}_i)^2.$$

For locally weighted regression, the weights ω_i are determined depending on how “close” \mathbf{x}_i is to the new query \mathbf{x}_q (further $\mathbf{x}_q = \mathbf{x}_{N+1}$). Often, the Gaussian kernel function K is used:

$$K(\mathbf{x}_i, \mathbf{x}_q) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_q\|^2}{2\sigma^2}\right),$$

$$\|\mathbf{x}_i - \mathbf{x}_q\|^2 = \sqrt{(\mathbf{x}_i - \mathbf{x}_q)^2},$$

where $\sigma^2 \in \mathbb{R}$ is arbitrary.

We select the $\omega_i = K(\mathbf{x}_i, \mathbf{x}_q)$. $K(\mathbf{x}_n, \mathbf{x}') \rightarrow 1$ when $\|\mathbf{x}_n - \mathbf{x}'\| \rightarrow 0$, and $K(\mathbf{x}_n, \mathbf{x}') \rightarrow 0$ when $\|\mathbf{x}_n - \mathbf{x}'\| \rightarrow \infty$. Finally, we obtain this optimization problem:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^N K(\mathbf{x}_i, \mathbf{x}_q) (y_i - \tilde{\boldsymbol{\beta}}^T \mathbf{x}_i)^2, \quad (5)$$

here $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)^T$. The solution to this problem is similar not very different from that of problem (4). Expression (5) is transformed:

$$\begin{aligned} & \min_{\tilde{\beta}} \sum_{i=1}^N (K(\mathbf{x}_i, \mathbf{x}_q)^{1/2} y_i - K(\mathbf{x}_i, \mathbf{x}_q)^{1/2} \tilde{\beta}^T \mathbf{x}_i)^2 \\ &= \min_{\tilde{\beta}} \sum_{i=1}^N \left(K(\mathbf{x}_i, \mathbf{x}_q)^{1/2} y_i - \sum_{j=1}^p \tilde{\beta}_j K(\mathbf{x}_i, \mathbf{x}_q)^{1/2} x_{ij} \right)^2 \\ &= \min_{\tilde{\beta}} \sum_{i=1}^N \left(z_i - \sum_{j=1}^p \tilde{\beta}_j \xi_{ij} \right)^2, \end{aligned}$$

here $z_i = K(\mathbf{x}_i, \mathbf{x}_q)^{1/2} y_i$ and $\xi_{ij} = K(\mathbf{x}_i, \mathbf{x}_q)^{1/2} x_{ij}$, $i = 1, \dots, N$, $j = 1, \dots, p$.

Analogically, we solve the quantile regression problem. Suppose, we have a quantile regression without regularization:

$$\min_{f_\tau} \sum_{i=1}^N \omega_i \rho_\tau(y_i - f_\tau(\mathbf{x}_i | \beta)) = \min_{f_\tau} \sum_{i=1}^N K(\mathbf{x}_i, \mathbf{x}_q) \rho_\tau(y_i - f_\tau(\mathbf{x}_i | \beta)). \quad (6)$$

After certain transformations, we note: $f_\tau(\mathbf{x}_i | \beta) = \beta_1 x_{i1} + \dots + \beta_p x_{ip} = \beta^T \mathbf{x}_i$, $\tilde{y}_i = K(\mathbf{x}_i, \mathbf{x}_q) y_i$, $\tilde{\mathbf{x}}_i = K(\mathbf{x}_i, \mathbf{x}_q) \mathbf{x}_i$. In the case of a linear regression where $\omega_i = K(\mathbf{x}_i, \mathbf{x}_q)$, we find that the solution of (6) is equivalent to the problem:

$$\min_{\beta} \sum_{i=1}^N \rho_\tau(\tilde{y}_i - \beta^T \tilde{\mathbf{x}}_i). \quad (7)$$

Hence, the locally weighted regression method is reduced to problem (7). This modification gives a more accurate estimate of β ; in this way, we get a more accurate prediction of the response variable.

2.4 Locally weighted quantile regression with ELM modification

ELM is a widely used method based on the idea of a single hidden layer of feed-forward neural networks (SLFNs) [3, 10].

In our case, ELM is used for the modification of independent variables. Data \mathbf{x}_i will be changed to a linear transformation with random weights [3], and the activation function (sigmoid function) will be chosen. Let us define the linear quantile regression problem:

$$\min_{\beta} \sum_{i=1}^N \rho_\tau(y_i - \beta^T \mathbf{x}_i).$$

Hence, once we have \mathbf{x}_i , the new pseudovariables \mathbf{z}_i are created:

$$z_{ij} = \varphi(\omega_j^T \mathbf{x}_i), \quad j = 1, \dots, m, \quad (8)$$

here $\boldsymbol{\omega}_j = (\omega_{1j}, \dots, \omega_{pj})^T$ are randomly generated weights by ELM, φ is a sigmoid function that $\varphi(u) = \varphi_1(u)$ or $\varphi(u) = \varphi_2(u)$, where φ_1 and φ_2 are defined:

$$\begin{aligned} \varphi_1 : \mathbb{R} &\rightarrow [0, M] \text{ such that } \varphi_1(u) = M/(1 + e^{-hu}), \text{ here } h > 0 \text{ is a parameter.} \\ \varphi_2 : \mathbb{R} &\rightarrow [-M, M] \text{ such that } \varphi_2(u) = M(e^u - e^{-u})/(e^u + e^{-u}). \end{aligned}$$

We have modified data $\mathbf{z}_i = (z_{i1}, \dots, z_{im})$, $i = 1, \dots, N$. Now, we have the m new covariates composed of previous p covariates. The number m is arbitrary. Exactly these new data will be used in further analysis.

The following formula defines the modified quantile regression:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^N \rho_{\tau}(\tilde{y}_i - \boldsymbol{\beta}^T \tilde{\mathbf{x}}_i) + \lambda J(\boldsymbol{\beta}), \quad (9)$$

here, the response variable is unchanged:

$$\tilde{y}_i = K(\mathbf{x}_i, \mathbf{x}_q) y_i. \quad (10)$$

However, we have a new expression of independent variables:

$$\tilde{\mathbf{x}}_i = K(\mathbf{x}_i, \mathbf{x}_q) \mathbf{z}_i, \quad (11)$$

here \mathbf{z}_i is defined as (8). In this way (only data transformation), the local quantile regression with ELM modified covariates is obtained. The assessment of this method is the same as for the initial quantile regression (2).

3 Practical implementation

In this section, previously described methods will be applied for data analysis. The regression will be constructed in two ways: the general quantile regression and locally weighted quantile regression with ELM modified covariates (further – the modified method).

Practical modeling is concentrated on three different values of the order τ ($\tau = 0.05, 0.5, 0.95$). When the variant is $\tau = 0.5$, we deal with the conditional median of the response variable. As we know, the median and mean are the same for the r.v. with a symmetric density function. The quantile of the $\tau = 0.5$ order is the most likely value of the response variable. An interpretation of the quantiles of order $\tau = 0.05$ and $\tau = 0.95$ gives the confidence interval of the response variable (in this case, 90 per cent).

Recently, new indicators (indexes) have been constructed that reflect the changes in economy more precisely than the gross domestic product [8]. Usually, this type of indicator is constructed as a combination of different indicators from various fields. Statistical data are selected regarding economic theory and additional methods (correlation, causality analysis, etc.). Weights are chosen using mathematical methods or including additional sources of information. In this way, so-called CI's are constructed. The advantage of the CI mainly depends on the chosen methodology and on selected statistical data that reflect

the general tendencies of a specific field. In summary, the CI is defined as a mathematical function:

$$\text{CI} = f(X, \omega),$$

here, X is a set of variables that compile CI, ω – weights that are assigned to every variable.

In this paper, we will use the methodology for constructing the CI presented in [8]. For the analysis and prediction, we extended the time period (1998–2014) and reduced the number of variables to $k = 12$. Economical indicators of monthly and quarterly periodicity were used: statistical data of industry, construction, domestic trade, foreign trade, services, the producer price index. We will briefly describe the steps followed in constructing the CI. First of all, a preliminary analysis was applied (outliers were detected, missing values were assigned), and all data were seasonally adjusted. Data of quarterly periodicity were transformed into monthly periodicity. The range of all indicators was transformed into the range $[0, 1]$. Using the factor analysis method, we left only the most important indicators for further analysis. The weights of individual indicators were obtained by using the Nicoletti method [6] from the rotated factor loading matrix.

Hence, we have data $\{(\mathbf{x}_t, y_t), t = 1, \dots, 204\}$, where $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,12})$ is the t value of the covariate, and y_t is the t value of CI. Here we changed the index i to t in order to highlight the importance of time in the time series as the order of data.

In modeling economic indicators, lags of covariates are usually also included in the models. We suppose that

$$y_t = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-d}) + \epsilon_t.$$

In this case, we say that y depends on the last d values of \mathbf{x} . In modeling CI, we deal with four different cases. We will use linear dependency and different lags ($d = 1, 2, 3, 4$):

$$y_t = \beta_0 + \sum_{j=1}^d \beta_j \mathbf{x}_{t-j} + \epsilon_t. \quad (12)$$

We solve problems (12) and obtain the general quantile regression

$$\min_{\beta} \sum_{t=1}^N \rho_{\tau} \left(y_t - \beta_0 - \sum_{j=1}^d \beta_j \mathbf{x}_{t-j} \right) + \lambda J(\beta) \quad (13)$$

and the modified quantile regression method

$$\min_{\beta} \sum_{t=1}^N \rho_{\tau} (\tilde{y}_t - \beta^T \tilde{\mathbf{x}}_t) + \lambda J(\beta). \quad (14)$$

\tilde{y}_t , $\tilde{\mathbf{x}}_t$ and the components are defined in (8), (10), and (11). Here $\beta = (\beta_0, \dots, \beta_d)^T$, and $J(\beta)$ – one of the regularization functions (LASSO, SCAD, or MCP). In methods (13) and (14) with $d = 1, 2, 3, 4$, CI is evaluated without considering the recent values of the

covariates. Then the modified quantile regression will be constructed “around” the recent value of the covariate \mathbf{x}_t , which will be used in forecasting.

Twelve different models are constructed using the general (3) and modified (9) quantile regression by following these steps:

- (i) Data points are chosen: $104 \leq n \leq 204$.
- (ii) The lag d ($d = 1, 2, 3, 4$) and the regularization function $J(\beta)$ are fixed.
- (iii) The τ order ($\tau = 0.05, 0.5, 0.95$) of quantile is chosen.
- (iv) The general (3) and the modified (9) quantile regression for the n first elements of the sample are used (i.e., $N = n$), and $\mathbf{x}_q = (\mathbf{x}_n, \dots, \mathbf{x}_{n-d})$.
- (v) Having \mathbf{x}_q , we predict (one-step-ahead) the values of y_n of the quantile of the τ order using the general and modified methods.
- (vi) These steps are repeated with $\tau = 0.05$, $\tau = 0.5$, and $\tau = 0.95$. The estimates of the general and modified regressions $\hat{y}_n^{(0.05)}$, $\hat{y}_n^{(0.5)}$, $\hat{y}_n^{(0.95)}$, $\hat{\hat{y}}_n^{(0.05)}$, $\hat{\hat{y}}_n^{(0.5)}$, and $\hat{\hat{y}}_n^{(0.95)}$ are obtained. The value in brackets stands for the order of the quantile.

This algorithm is repeated with every n from the interval $[104, 204]$. E.g., predictions of the modified method: $\tau = 0.05$: $\hat{\mathbf{y}}^{(0.05)} = (\hat{y}_{104}^{(0.05)}, \dots, \hat{y}_{204}^{(0.05)})$, and predictions of the general method are obtained, e.g., $\hat{\hat{\mathbf{y}}}^{(0.05)} = (\hat{\hat{y}}_{104}^{(0.05)}, \dots, \hat{\hat{y}}_{204}^{(0.05)})$.

The following measures to verify the accuracy of the models were used:

- Mean absolute error: $\text{MAE}(\hat{\mathbf{y}}, \mathbf{y}) = \sum_{i=1}^T |y_i - \hat{y}_i|/T$.
- Mean absolute percentage error: $\text{MAPE}(\hat{\mathbf{y}}, \mathbf{y}) = \sum_{i=1}^T |(y_i - \hat{y}_i)/y_i|/T \cdot 100$.
- Root mean square error: $\text{RMSE}(\hat{\mathbf{y}}, \mathbf{y}) = (\sum_{i=1}^T (y_i - \hat{y}_i)^2)^{1/2}$.
- The measure M_1 : $M_1(\hat{\mathbf{y}}, \hat{\hat{\mathbf{y}}}, \mathbf{y}) = \sum_{i=1}^T \mathbf{I}(|\hat{y}_i - y_i| < |\hat{\hat{y}}_i - y_i|)/T$. In this case, the $M_1(\hat{\mathbf{y}}, \hat{\hat{\mathbf{y}}}, \mathbf{y})$ indicates the share of all values, where the estimate \hat{y}_i is better than the estimate $\hat{\hat{y}}_i$ in comparison to the original data y_i .
- The measure M_2 : $M_2(\hat{\mathbf{y}}, \hat{\hat{\mathbf{y}}}, \mathbf{y}) = \sum_{i=1}^T \mathbf{I}(\hat{y}_i < y_i < \hat{\hat{y}}_i)/T$. Here, the $M_2(\hat{\mathbf{y}}, \hat{\hat{\mathbf{y}}}, \mathbf{y})$ indicates the share of all values, where the estimate \hat{y}_i falls within the interval between $\hat{\hat{y}}_i$ and $\hat{\hat{\hat{y}}}_i$.

In all formulas, T stands for the length of time series. In practice, we will use the M_2 to find out how often the actual value falls within the confidence interval. E.g., for the 90 per cent confidence interval, the measure will be calculated: $M_2(\mathbf{y}^{(0.05)}, \mathbf{y}^{(0.95)}, \mathbf{y})$, here $\mathbf{y}^{(\tau)}$ stands for the quantile of the τ order.

4 Results

Finally, after we have performed the modeling with different d and all regularizations, the set of vectors is obtained: $\hat{\mathbf{y}}^{(0.05)}$, $\hat{\mathbf{y}}^{(0.5)}$, $\hat{\mathbf{y}}^{(0.95)}$, $\hat{\hat{\mathbf{y}}}^{(0.05)}$, $\hat{\hat{\mathbf{y}}}^{(0.5)}$, and $\hat{\hat{\mathbf{y}}}^{(0.95)}$. In order to analyze the accuracy of our predictions, the original data $(y_{104}, \dots, y_{204})$ are compared to the estimates of quantiles. The prediction value should be around the mean or the median; hence, the prediction of the median shall show the general view. In the case of 90 per cent, the realization of the random variable (CI) should fall within the confidence interval

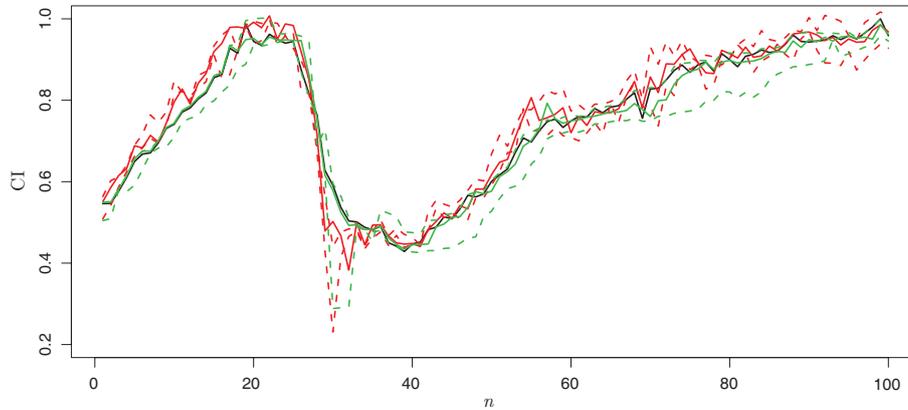


Figure 1. Data with four lags and MCP regularization. (Online version in colour.)

Table 1. Characteristics of accuracy (input data with three lags).

	1:3 'LASSO'	1:3 'SCAD'	1:3 'MCP'
MAE(REZ_OLD)	0.024	0.024	0.024
MAE(REZ_NEW)	0.024	0.015	0.017
MAPE(REZ_OLD)	3.327	3.380	3.370
MAPE(REZ_NEW)	3.493	2.037	2.275
RMSE(REZ_OLD)	0.032	0.033	0.033
RMSE(REZ_NEW)	0.037	0.023	0.025
M1(REZ_NEW, REZ_OLD)	0.530	0.620	0.600
M2(REZ_OLD_lo, REZ_OLD_up)	0.490	0.560	0.560
M2(REZ_NEW_lo, REZ_NEW_up)	0.100	0.910	0.920

Table 2. Characteristics of accuracy (input data with four lags).

	1:4 'LASSO'	1:4 'SCAD'	1:4 'MCP'
MAE(REZ_OLD)	0.026	0.027	0.027
MAE(REZ_NEW)	0.017	0.013	0.012
MAPE(REZ_OLD)	3.635	3.763	3.734
MAPE(REZ_NEW)	2.408	1.742	1.635
RMSE(REZ_OLD)	0.036	0.038	0.037
RMSE(REZ_NEW)	0.023	0.018	0.018
M1(REZ_NEW, REZ_OLD)	0.600	0.650	0.680
M2(REZ_OLD_lo, REZ_OLD_up)	0.510	0.460	0.460
M2(REZ_NEW_lo, REZ_NEW_up)	0.090	0.920	0.910

between the 0.05 and 0.95 quantiles. Hence, we will analyze how well the model was able to assess the confidence interval of the response variable.

The example ($d = 4$ and MCP regularization) of modeling results is presented in Fig. 1 and in Tables 1, 2. In the figure, the black line denotes the median of original data; the red line denotes the median obtained by the general quantile method; the green line –

by the modified method; the dashed red and green lines denote the 0.05 and 0.95 quantiles by the general and modified methods, respectively.

If we observe only the graphical results, it is quite difficult to determine which method is the better one. In different periods, one of the method gives more accurate predictions, or predictions are quite similar.

In the tables, REZ_NEW stands for the estimate of the median obtained by the modified method, and REZ_OLD is the estimate of the median obtained by the general method, REZ_OLD_lo, REZ_OLD_up, REZ_NEW_lo, REZ_NEW_up stand for the quantiles of 0.05 and 0.95, respectively. The notation “1:3” means that input data with the lags first, second, and third were used in a particular model.

Better results were obtained by the modified method using input data with three or four lags. MAE, MAPE, RMSE, and other statistics confirmed this fact. We noticed that the widths of the confidence intervals obtained by both methods are similar, but in the case of the modified method with SCAD and MCP regularizations, more than 90 per cent of observed data fall within the constructed confidence interval. In the case of a general method, only 46 per cent of data realizations fall within the interval. More results can be found in [1].

The analysis showed that measures of accuracy are dependent on regularization and the number of lags. The smallest errors of general method was obtained using LASSO regularization, while in the case of the modified quantile regression, the best result was obtained using four lags and MCP regularization. We see that different methods proceed with different accuracy depending on the number of lags. However, the modified method has this advantage: its confidence interval and estimate of the median are more accurate.

Also, the comparative analysis of obtained results and results of previous research [8] was performed. In this paper and in [8], the CI is slightly different. Here the time period is extended to 1998–2014 (in the previous research, 1998–2010), and the number of variables was reduced to $k = 12$ (previously, $k = 28$). The purpose and expected results of the combined method of extreme learning machine and locally weighted regression [8] and locally weighted quantile regression are unequal as well. Regardless of the differences we compared some accuracy measures (RMSE and MAPE) of one-step-ahead predictions. The analysis showed that RMSE and MAPE are slightly smaller of best models of locally weighted quantile regression.

5 Conclusions

In this paper, the quantile regression with penalty function was extended by including local weights and the ELM method for the modification of covariates. This developed modified method was adopted for the one-step-ahead prediction of the CI of the Lithuanian economy. Our analysis indicated that the locally weighted quantile regression with regularization obtains on average better results than the known quantile regression with regularization: the new method enabled us to obtain more accurate conditional medians and predictions of confidence intervals. The best results were obtained by methods with SCAD and MCP regularizations.

References

1. P. Asijavičius, Locally weighted quantile regression, Master thesis, Vilnius University, 2015.
2. J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Stat. Assoc.*, **96**(456):1348–1360, 2001.
3. G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing*, **70**(1–3):489–501, 2006.
4. R. Koenker, G. Bassett, Regression quantiles, *Econometrica*, **46**(1):33–50, 1978.
5. A. Moore, J. Schneider, K. Deng, Efficient locally weighted polynomial regression predictions, in D.H. Fisher (Ed.), *Proceedings of the Fourteenth International Conference on Machine Learning, Nashville, TN, July 08–12, 1997*, Morgan Kaufmann, San Francisco, CA, 1997, pp. 236–244.
6. G. Nicoletti, S. Scarpetta, O. Boylaud, Summary indicators of product market regulation with an extension to employment protection legislation, OECD Working Papers, No. 226, OECD Publishing, Paris, 2000.
7. J. Rukšėnaitė, Impact of factor rotation methods on simulation composite indicators, *Math. Model. Anal.*, **16**(3):418–431, 2011.
8. J. Rukšėnaitė, P. Vaitkus, Prediction of composite indicators using combined method of extreme learning machine and locally weighted regression, *Nonlinear Anal. Model. Control*, **17**(2):238–251, 2012.
9. B. Sherrwood, Penalized quantile regression, rqPen package documentation, 2015.
10. J. Tang, G. Huang, Extreme learning machine for multilayer perceptron, *IEEE Trans. Neural Networks Learn. Syst.*, **27**(4):809–821, 2015.
11. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc., Ser. B*, **1**(58):267–288, 1996.
12. C.H. Zhang, Nearly unbiased variable selection under minimax concave penalty, *Ann. Stat.*, **38**(2):894–942, 2010.
13. H. Zou, The adaptive lasso and its oracle properties, *J. Am. Stat. Assoc.*, **101**(476):1418–1429, 2006.