

Support vector expectile regression using IRWLS procedure[†]

Kook-Lyeol Choi¹ · Jooyong Shim² · Kyungha Seok³

¹²³Department of Data Science, Inje University

Received 5 June 2014, revised 24 June 2014, accepted 1 July 2014

Abstract

In this paper we propose the iteratively reweighted least squares procedure to solve the quadratic programming problem of support vector expectile regression with an asymmetrically weighted squares loss function. The proposed procedure enables us to select the appropriate hyperparameters easily by using the generalized cross validation function. Through numerical studies on the artificial and the real data sets we show the effectiveness of the proposed method on the estimation performances.

Keywords: Asymmetrically weighted squares loss function, check function, expectile, generalized cross validation function, iteratively reweighted least squares procedure, quantile regression.

1. Introduction

Quantile regression introduced by Koenker and Bassett (1978) is gradually involving into an ensemble of practical statistical methods for estimating and conducting inference about models for conditional quantile functions. Quantile regression is an increasingly popular method for estimating the quantiles of a distribution conditional on the values of covariates, which allows both the location and the spread of the response to be studied by obtaining not only medians but also lower and higher quantiles. Quantiles are robust against the influence of outliers and, taken several at a time, they give a more complete picture of the conditional distribution than a single estimate of the center. Just as classical linear regression methods based on minimizing sum of squared residuals enable one to estimate a wide variety of models for conditional mean functions, quantile regression methods offer a mechanism for estimating models for the conditional median function, and the full range of other conditional quantile functions. By supplementing the estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, quantile regression is capable of providing a more complete statistical analysis of the stochastic relationships among

[†] This work was supported by the 2013 Inje University research grant.

¹ Professor, Institute of Statistical Information, Department of Data Science, Inje University, Gyungnam 621-749, Korea.

² Adjunct professor, Institute of Statistical Information, Department of Data Science, Inje University, Gyungnam 621-749, Korea.

³ Corresponding author: Professor, Institute of Statistical Information, Department of Data Science, Inje University, Gyungnam 621-749, Korea. E-mail: statskh@inje.ac.kr

random variables. The introductions and current research areas of the quantile regression can be found in Koenker and Hallock (2001), Yu *et al.* (2003) and Hwang (2010).

In general, quantile regression is based on an asymmetric L_1 norm, but the L_2 norm also can be made asymmetric in the same way. Newey and Powell (1987) proposed an asymmetric L_2 norm in a regression setting. In their paper they define the asymmetric least squares estimation for expectile and show that expectile is monotonically increasing and is location and scale equivariant. Schnabel and Eilers (2009) proposed the least asymmetrically weighted squares regression combined with P-splines (PsExR) to estimate expectile regression functions. Wang *et al.* (2011) proposed the support vector machine (SVM) approach for the nonlinear estimation of expectile regression function.

The SVM, firstly developed by Vapnik (1995, 1998), is being used as a new technique for regression and classification problems. SVM is gaining popularity due to many attractive features, and promising empirical performance. SVM was initially developed to solve classification problems but recently it has been extended to the domain of regression problems. SVM is based on the structural risk minimization (SRM) principle, which minimizes an upper bound on the expected risk unlike ERM minimizing the error on the training data. By minimizing this bound, high generalization performance can be achieved. In particular, for the SVM regression case SRM results in the regularized ERM with the ϵ -insensitive loss function. The introductions and overviews of recent developments of SVM regression can be found in Vapnik (1995, 1998), Smola and Schölkopf (1998), Wang (2005) and Shim and Hwang (2013). Training an SVM requires the solution to a quadratic programming (QP) optimization problem. But QP problem presents some inherent limitations which results in computational difficulty especially for the large data sets. Platt (1999), Flake and Lawrence (2002) developed the sequential minimal optimization algorithm which divides the QP problem into a series of small QP problems to avoid such computational difficulty. Perez-Cruz *et al.* (2000) proposed iteratively reweighted least squares (IRWLS) algorithm for SVR by transforming the Lagrangian function into sum of quadratic terms by defining associated weights of predicted errors.

In this paper we propose an IRWLS procedure to solve the QP problem of support vector expectile regression (SVExR) with an asymmetrically weighted squares loss function. The asymmetrically weighted squares loss function can provide the differentiability at 0, which enables to solve QP problem by IRWLS procedure. To select appropriate parameters for the achievement of high generalization performance, a commonly used method is minimizing the cross validation (CV) function. But selecting parameters using CV function is computationally formidable. In SVExR using QP, GCV function cannot be obtained by solving QP directly. But by using IRWLS procedure GCV function can be easily obtained.

The rest of this paper is organized as follows. In Section 2 and 3 we give brief reviews of quantiles, expectiles and SVExR proposed by Wang *et al.* (2011). In Section 4 we propose an IRWLS procedure for SVExR and present the model selection method using GCV function. In Section 5 we perform the numerical studies through examples. In Section 6 we give the conclusions.

2. Quantiles and expectiles

A more complete picture of the conditional distribution can be characterized by expectiles similar to its characterization by quantiles than a single estimate of the center. Expectiles

are similar to quantile but they are determined by tail expectation rather than tail probability. Quantiles are intuitively appealing, but expectiles are easier to compute. The ease of computation will become clear in Section 4.

For a probability density function $f(y)$, the θ th theoretical quantile q_θ of y is defined as a solution to

$$\min_q \left((1 - \theta) \int_{-\infty}^q |y - q| f(y) dy + \theta \int_q^{\infty} |y - q| f(y) dy \right).$$

And the θ th theoretical expectile $e_\theta(y)$ of y is defined as a solution to

$$\min_e \left((1 - \theta) \int_{-\infty}^e (y - e)^2 f(y) dy + \theta \int_e^{\infty} (y - e)^2 f(y) dy \right),$$

which is equivalent to a solution to

$$(1 - \theta) \int_{-\infty}^e (y - e) f(y) dy + \theta \int_e^{\infty} (y - e) f(y) dy = 0.$$

Using the cumulative distribution function $F(x)$ and the partial moment function $G(x) = \int_{-\infty}^x y f(y) dy$, expectile of y , $e_\theta(y)$, can be obtained from the following implicit equation,

$$e_\theta(y) = \frac{(1 - \theta)G(e_\theta(y)) + \theta(E(y) - G(e_\theta(y)))}{(1 - \theta)F(e_\theta(y)) + \theta(1 - F(e_\theta(y)))},$$

and

$$\theta = \frac{G(e_\theta(y)) - e_\theta(y)F(e_\theta(y))}{2(G(e_\theta(y)) - e_\theta(y)F(e_\theta(y))) + (e_\theta(y) - E(y))}.$$

The θ th sample expectile of y is obtained by minimizing the objective function consists of asymmetric weighted squares loss functions such as,

$$\rho_\theta(r) = \theta r^2 I(r \geq 0) + (1 - \theta) r^2 I(r < 0),$$

which is obtained if the observations are assumed to be independently drawn from an asymmetric normal distribution whose probability density function is given as follows;

$$f(y|\theta, \mu, \sigma) = \frac{\sqrt{\theta} + \sqrt{1 - \theta}}{4\sigma\sqrt{\pi\theta(1 - \theta)}} \exp\left(-\frac{1}{\sigma^2}\rho_\theta(y - \mu)\right).$$

Analogous to the well-known QQ (quantile-quantile) plot, EE (expectile-expectile) plot can be one of methods to compare a sample distribution with a theoretical distribution for various values of θ (Schnabel and Eilers, 2009).

3. Support vector expectile regression

Assume that the θ th expectile regression function of the response given $\mathbf{x}_i \in R^p$, $e_\theta(\mathbf{x}_i)$, is linearly related to input vector \mathbf{x}_i as

$$e_\theta(\mathbf{x}_i) = \beta^t \mathbf{x}_i + b \quad \text{for } i = 1, \dots, n,$$

where β and b depend on θ implicitly. β and b are estimated from

$$\min \sum_{i=1}^n u_i (y_i - \beta^t \mathbf{x}_i - b)^2 \text{ for } \theta \in (0, 1),$$

where $u_i = \begin{cases} \theta, & y_i - \beta^t \mathbf{x}_i - b > 0 \\ 1 - \theta, & y_i - \beta^t \mathbf{x}_i - b < 0 \end{cases}$.

Now we assume that the θ th expectile regression function $e_\theta(\mathbf{x}_i)$ is nonlinearly related to input vector \mathbf{x}_i as

$$e_\theta(y|\mathbf{x}_i) = \boldsymbol{\omega}^t \phi(\mathbf{x}_i) + b \text{ for } i = 1, \dots, n,$$

where $\boldsymbol{\omega}$ depends on θ and $\phi(\mathbf{x}_i)$ is a nonlinear feature mapping function, which is used to allow for the case of nonlinear expectile regression, input vectors are nonlinearly transformed into a potentially higher dimensional feature space by a nonlinear mapping function ϕ and then a linear expectile regression is performed there. For this nonlinear quantile regression, solution requires the computations of dot products $\phi(\mathbf{x}_k)^t \phi(\mathbf{x}_l)$, $k, l = 1, \dots, n$, in a potentially higher dimensional feature space. Under Mercer (1909)'s conditions, these demanding computations can be reduced significantly by introducing a kernel function K such that $K(\mathbf{x}_k, \mathbf{x}_l) = \phi(\mathbf{x}_k)^t \phi(\mathbf{x}_l)$. Several choices of kernel functions are possible. Gaussian function is the most frequently used kernel function. The kernel function includes some parameters that characterize its structure. The linear expectile regression can be regarded as special case of nonlinear expectile regression with an identity feature mapping function in nonlinear regression, that is, $K(\mathbf{x}_k, \mathbf{x}_l) = \mathbf{x}_k^t \mathbf{x}_l$, it reduces to linear expectile regression.

Support vector approach for the nonlinear estimation of the θ th expectile regression function $e_\theta(\mathbf{x}_i)$ leads to the optimization problem defined as follows,

$$\min \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^n u_i (y_i - \boldsymbol{\omega}^t \phi(\mathbf{x}_i) - b)^2 \text{ for } \theta \in (0, 1), \tag{3.1}$$

where $C > 0$ is a regularization parameter.

By introducing slack variables ξ_i and ξ_i^* , we can rewrite (3.1) by following optimization problem,

$$\begin{aligned} & \min \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C\theta \sum_{i=1}^n \xi_i^2 + C(1 - \theta) \sum_{i=1}^n \xi_i^{*2} \tag{3.2} \\ & \text{subject to } \begin{cases} y_i - \boldsymbol{\omega}^t \phi(\mathbf{x}_i) - b \leq \xi_i, \\ -y_i + \boldsymbol{\omega}^t \phi(\mathbf{x}_i) + b \leq \xi_i^*, \\ \xi_i \geq 0, \xi_i^* \geq 0, \quad i = 1, \dots, n. \end{cases} \end{aligned}$$

The dual of this problem can be computed straightforwardly using Lagrange multipliers. The dual constraints for ξ_i and ξ_i^* can be combined into one variable. This yields the following quadratic programming problem,

$$\begin{aligned} & \max -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i \\ & - \frac{1}{2C\theta} \sum_{i=1}^n \alpha_i^2 - \frac{1}{2C(1 - \theta)} \sum_{i=1}^n \alpha_i^{*2} \end{aligned}$$

subject to $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0$, $\alpha_i \geq C\theta$ and $\alpha_i^* \geq C(1 - \theta)$. From Karush-Kuhn-Tucker conditions (Kuhn and Tucker, 1951) we can see that $y_i = K_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + b + \frac{\alpha_i}{C\theta}$ for $i \in I_{s_1} = \{i = 1, \dots, n | \alpha_i > 0\}$ and $y_i = K_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + b - \frac{\alpha_i^*}{C(1 - \theta)}$ for $i \in I_{s_2} = \{i = 1, \dots, n | \alpha_i^* > 0\}$ where $K_i = (K(\mathbf{x}_1, \mathbf{x}_i), \dots, K(\mathbf{x}_n, \mathbf{x}_i))$ and $\boldsymbol{\alpha}^{(*)} = (\alpha_1^{(*)}, \dots, \alpha_n^{(*)})^t$. Thus, the optimal value of b is obtained as follows,

$$\hat{b} = \frac{1}{n_{s_1}} \sum_{i \in I_{s_1}} (y_i - K_i(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) - \frac{\hat{\alpha}_i}{C\theta}) + \frac{1}{n_{s_2}} \sum_{i \in I_{s_2}} (y_i - K_i(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) + \frac{\hat{\alpha}_i^*}{C(1 - \theta)}),$$

where n_{s_1} and n_{s_2} are sizes of I_{s_1} and I_{s_2} , respectively.

For given input vector \mathbf{x}_t , the estimator of the θ th expectile regression function is obtained as

$$\hat{e}_\theta(\mathbf{x}_t) = K_t(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) + \hat{b},$$

where $K_t = (K(\mathbf{x}_1, \mathbf{x}_t), \dots, K(\mathbf{x}_n, \mathbf{x}_t))$.

4. Support vector expectile regression using IRWLS procedure

Using IRWLS procedure, we can get a good approximate solution to the optimal problem (3.2). Now (3.2) becomes the optimal problem of obtaining $\boldsymbol{\beta}$ and b to minimize

$$L(\boldsymbol{\beta}, b) = \frac{1}{2} \boldsymbol{\beta}^t K \boldsymbol{\beta} + \frac{C}{2} \sum_i u_i (y_i - K_i \boldsymbol{\beta} - b)^2, \tag{4.1}$$

where K is an $n \times n$ matrix with (i, j) th entities $K(\mathbf{x}_i, \mathbf{x}_j)$ and

$$u_i = \begin{cases} \theta, & y_i - K_i \boldsymbol{\beta} - b > 0 \\ 1 - \theta, & y_i - K_i \boldsymbol{\beta} - b < 0 \end{cases} .$$

Taking partial derivatives of (4.1) with regard to $\boldsymbol{\beta}$ and b , respectively, leads to the optimal values of $\boldsymbol{\beta}$ and b to be the solution to,

$$\begin{aligned} \mathbf{0} &= K \boldsymbol{\beta} - CKU \mathbf{y} + CKUK \boldsymbol{\beta} + CKU \mathbf{1} b \\ 0 &= \mathbf{1}^t U \mathbf{y} - \mathbf{1}^t UK \boldsymbol{\beta} - \mathbf{1}^t U \mathbf{1} b. \end{aligned} \tag{4.2}$$

Here U is the diagonal matrix with entities u_i 's, $\mathbf{y} = (y_1, \dots, y_n)^t$ and $\mathbf{1}^t = (1, \dots, 1)_{1 \times n}$.

The solutions to (4.2) cannot be obtained in a single step since U contain $\boldsymbol{\beta}$ and b therein. Thus, we need to apply IRWLS procedure which starts with initialized values of $\hat{\boldsymbol{\beta}}$ and \hat{b} as follows,

- (i) Calculate U using $r_i = y_i - K_i \hat{\boldsymbol{\beta}}^{(l)} - \hat{b}^{(l)}$.
- (ii) Obtain $\hat{\boldsymbol{\beta}}^{(l+1)}$ and $\hat{b}^{(l+1)}$ from $\begin{pmatrix} \hat{\boldsymbol{\beta}}^{(l+1)} \\ \hat{b}^{(l+1)} \end{pmatrix} = \begin{pmatrix} UK + I/CU \mathbf{1} \\ \mathbf{1}^t UK \mathbf{1}^t U \mathbf{1} \end{pmatrix}^{-1} \begin{pmatrix} U \\ \mathbf{1}^t U \end{pmatrix} \mathbf{y}$.
- (iii) Iterate steps until convergence.

For given input vector \mathbf{x}_t , the estimator of the θ th expectile regression function $e_\theta(\mathbf{x}_t)$ is obtained as

$$\widehat{e}_\theta(\mathbf{x}_t) = K_t \widehat{\boldsymbol{\beta}} + \widehat{b}$$

where $K_t = (K(\mathbf{x}_1, \mathbf{x}_t), \dots, K(\mathbf{x}_n, \mathbf{x}_t))$.

The problem of choosing the hyperparameters is ubiquitous in function estimation. Thus, we now illustrate the model selection method which chooses the appropriate hyperparameters of SVExR. The functional structure of SVExR is characterized by the regularization parameter C and the kernel parameter. To choose the hyperparameters of SVExR we first define the cross validation (CV) function as follows,

$$CV(\boldsymbol{\tau}) = \frac{1}{n} \sum_{i=1}^n \widehat{u}_i (y_i - \widehat{e}_\theta^{(-i)}(\mathbf{x}_i))^2,$$

where $\boldsymbol{\tau}$ is the set of parameters, \widehat{u}_i is the final estimate of u_i from full data, $\widehat{e}_\theta^{(-i)}(\mathbf{x}_i)$ is the estimate of $e_\theta(\mathbf{x}_i)$ from data without i th observation, while $\widehat{e}_i = \widehat{e}_\theta(\mathbf{x}_i)$ is the estimate of $e_\theta(\mathbf{x}_i)$ from full data. Since for each candidate of hyperparameter set, n of $\widehat{e}_\theta^{(-i)}(\mathbf{x}_i)$'s should be computed, selecting parameters using CV function is computationally formidable when we use SVQR using QP explained in Section 3. But for SVQR using IRWLS procedure we can find a GCV function which enables much faster model selection than CV function.

Using the leave-one-out lemma (Craven and Wahba, 1979) and the first order Taylor expansion,

$$(y_i - \widehat{e}_\theta^{(-i)}(\mathbf{x}_i)) - (y_i - \widehat{e}_\theta(\mathbf{x}_i)) = \widehat{e}_\theta(\mathbf{x}_i) - \widehat{e}_\theta^{(-i)}(\mathbf{x}_i) \approx \frac{\partial \widehat{e}_\theta(\mathbf{x}_i)}{\partial y_i} (y_i - \widehat{e}_\theta^{(-i)}(\mathbf{x}_i)),$$

we have

$$(y_i - \widehat{e}_\theta^{(-i)}(\mathbf{x}_i)) \approx \frac{y_i - \widehat{e}_\theta(\mathbf{x}_i)}{1 - \frac{\partial \widehat{e}_\theta(\mathbf{x}_i)}{\partial y_i}}. \tag{4.3}$$

Using (4.3) the ordinary cross validation (OCV) function is obtained as

$$OCV(\boldsymbol{\tau}) = \frac{1}{n} \sum_{i=1}^n \widehat{u}_i \left(\frac{y_i - \widehat{e}_\theta(\mathbf{x}_i)}{1 - h_{ii}} \right)^2,$$

where h_{ii} is the diagonal element of the hat matrix \mathbf{H} such that

$$\mathbf{H} = (\mathbf{K}, \mathbf{1}) \left(\begin{matrix} \widehat{U} \mathbf{K} + I/C \widehat{U} \mathbf{1} \\ \mathbf{1}^t \widehat{U} \mathbf{K} \mathbf{1}^t \widehat{U} \mathbf{1} \end{matrix} \right)^{-1} \left(\begin{matrix} \widehat{U} \\ \mathbf{1}^t \widehat{U} \end{matrix} \right) \text{ and } e_\theta(y|\mathbf{x}) = \mathbf{H} \mathbf{y}.$$

By replacing h_{ii} with $trace(\mathbf{H})/n$, we have the GCV function as follows,

$$GCV(\boldsymbol{\tau}) = \frac{n \sum_{i=1}^n \widehat{u}_i (y_i - \widehat{e}_\theta(\mathbf{x}_i))^2}{(n - trace(\mathbf{H}))^2}.$$

5. Numerical studies

We illustrate the performance of proposed SVExR using IRWLS procedure through the simulated data under different expectile levels. We use $\theta = 0.1, 0.5, 0.9$ for expectile levels. We compare the proposed SVExR with the nonlinear expectile regression PsExR of Schnabel and Eilers (2009). For comparison we calculate the average and standard error of root mean squared errors (RMSEs) for each estimated expectile regression function as follows:

$$RMSE_e = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{e}_\theta(y_i) - e_\theta(y_i))^2}.$$

Example 5.1 Here x 's are generated from a uniform distribution $U(0, 1)$ and y 's are generated from a normal distribution $N(\sin(2\pi x), 1)$. The true θ th expectile regression function is given as

$$e_{0.1}(x) = -0.8616 + \sin(2\pi x), e_{0.5}(x) = \sin(2\pi x), e_{0.9}(x) = 0.8616 + \sin(2\pi x).$$

Figure 5.1 shows true expectile regression functions (solid lines), the estimated expectile regression functions $\hat{e}_\theta(x_i)$'s by the proposed SVExR (left; dashed lines) and PsExR (right; dashed lines), respectively, which are superimposed on the scatter plots of x versus y . We can see that both methods provide the estimators of expectile regression functions close to the true expectile regression functions. Table 5.1 shows the average and standard error of 100 RMSEs for each expectile level, where the standard error of RMSEs is given in parentheses. As seen from Table 5.1, the proposed SVExR yields the smaller averages of RMSEs for 3 expectile levels but larger standard errors.

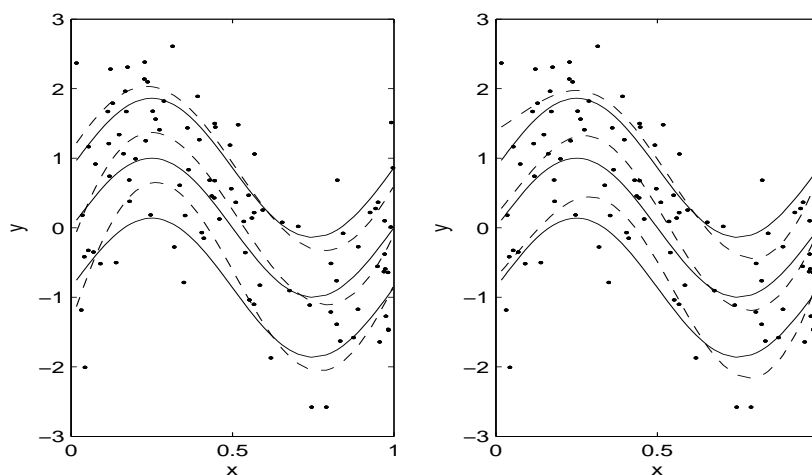


Figure 5.1 True expectile regression functions (solid lines) and their estimates (dashed lines) by the proposed SVExR (left) and PsExR (right) for $\theta = 0.1, 0.5$ and 0.9 in one of 100 data sets

Table 5.1 The average of 100 RMSEs of estimated expectile regression functions for $\theta = 0.1, 0.5$ and 0.9 (standard error is in parenthesis)

θ	0.1	0.5	0.9
SVExR	0.2734 (0.0084)	0.2257(0.0067)	0.2591 (0.0077)
PsExR	0.2816 (0.0080)	0.2423(0.0064)	0.2616 (0.0075)

Example 5.2 In this example expectile regression functions are used to identify girls whose triceps values lie in the tail of underlying distribution (Cole and Green, 1992), where the expectile regression functions are dependent on the age which are standardized for estimation of expectile regression functions. Figure 5.2 shows three expectile regression functions superimposed on the scatter plots. Three expectile regression functions show high triceps values in age 3 followed by a decrease until age 8 and then a second rise. For $\theta = 0.1$ and 0.5 both methods show the similar expectile regression functions but for $\theta = 0.9$ the proposed SVExR show more wiggly patterns. As we know from the data points, the 0.9th expectile regression function by the proposed SVExR beyond age 10 is less smoothed than the other two expectile regression functions, indicating that the variation of data is greater for girls with higher triceps values beyond age 10.

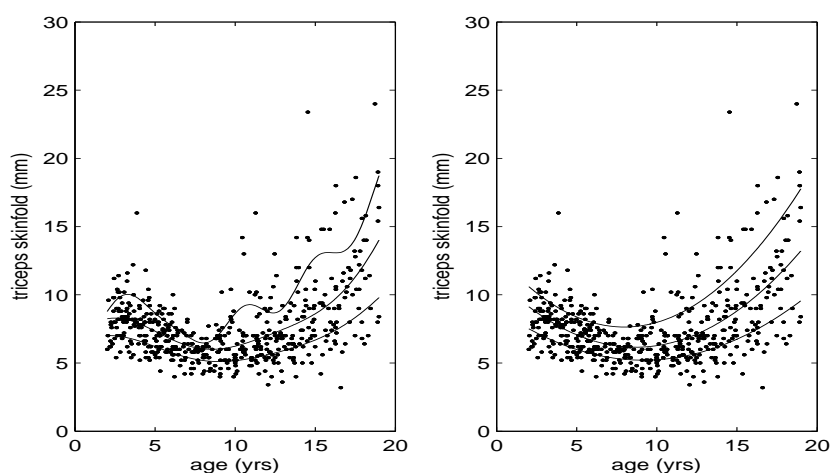


Figure 5.2 The estimated expectile regression functions of triceps skinfold in Gambian girls from age 2 to 18 years by the proposed SVExR (left) and PsExR (right) for $\theta = 0.1, 0.5$ and 0.9

6. Conclusions

In this paper, we dealt with estimating nonlinear expectile regression function based on SVM formulation using IRWLS procedure. We found that the proposed SVExR provides satisfying results in estimating expectile regression functions in the given examples. The proposed SVExR is very applicable for the high dimensional input vectors, which is an advantage to Schnabel and Eilers (2009) with P-splines.

References

- Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, **11**, 1305-1319.
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions : Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics*, **31**, 377-403.
- Flake, G. W. and Lawrence, S. (2002). Efficient SVM regression training with SMO. *Machine Learning*, **46**, 271-290.

- Hwang, C. (2010). M-quantile regression using kernel machine technique. *Journal of the Korean Data & Information Science Society*, **21**, 973-981.
- Koenker, R. and Bassett, G. (1978). Regression quantile. *Econometrica*, **46**, 33-50.
- Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, **40**, 122-142.
- Kuhn, H. and Tucker, A. (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, CA, 481-492.
- Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A*, **209**, 415-446.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819-847.
- Perez-Cruz, F., Navia-Vazquez, A., Alarcon-Diana, P. L. and Artes-Rodriguez, A. (2000). An IRWLS procedure for SVR. In *Proceedings of European Association for Signal Processing, EUSIPO 2000*, Tampere, Finland.
- Platt, J. (1999). Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods-Support Vector Learning*, edited by B. Schölkopf, C. J. C. Burges and A. J. Smola, MIT Press, Cambridge, MA, 185-208.
- Schnabel, S. K. and Eilers, P. H. C. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis*, **53**, 4168-4177.
- Shim, J. and Hwang, C. (2013). Expected shortfall estimation using kernel machines. *Journal of the Korean Data & Information Science Society*, **24**, 625-636.
- Smola, A. and Schölkopf, B. (1998). On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*, **22**, 211-231.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.
- Vapnik, V. N. (1998). *Statistical learning theory*, John Wiley, New York.
- Wang, L. (Ed.) (2005). *Support vector machines: Theory and application*, Springer, New York.
- Wang, Y., Wang, S. and Lai, K. (2011). Measuring financial risk with generalized asymmetric least squares regression. *Applied Soft Computing*, **11**, 5793-5800.
- Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research area. *The Statistician*, **52**, 331-350.