

# Nonparametric Quantile Estimation: A Geometric Framework for Laplacian Manifold Regularization

Ying ZHANG

<sup>1</sup>Post-doctoral Research Station, Economics and Management School, Wuhan University, Wuhan, China

<sup>2</sup>Post-doctoral Work Station, China Huarong Asset Management Corporation, Beijing, China

**ABSTRACT:** In this article we consider the general problem of utilizing both labeled and unlabeled data to improve quantile regression accuracy, and derive a nonparametric algorithm to compute the entire regularization path of the quantile estimator. We transform the optimization problem of the proposed approach into the quadratic optimization with linear constraint conditions and dimensionality reduction, and illustrate the finite sample behavior of the new approach.

**KEYWORD:** Nonparametric; Quantile Regression; Manifold Regularization; Semi-supervised Learning

## 1 INTRODUCTION

In recent years semi-supervised learning has become increasingly important and popular in machine learning and has numerous practical applications in data analysis including dimensionality reduction, visualization, clustering and classification, which incorporates labeled and unlabeled data in a general-purpose learner and exploits the underlying information of these data to improve efficiency in estimation and prediction for regression problems.

Compared with the mean regression method, quantile estimation is typically concerned with estimating the full family of conditional quantile functions. This method can give us more complete statistical information of response models. And it has been widely used in many real applications, including reference charts in medicine (Cole & Green 1992; Heagerty & Pepe 1999), survival analysis (Yang 1999; Koenker & Geling 2001), and economics (Hendricks & Koenker 1992; Koenker & Hallock 2001). In this paper we focus on deriving an efficient algorithm with utilizing both labeled and unlabeled data to improve quantile regression accuracy, because it is different for the amounts of valid data to be obtained at different distribution level, and especially in some extreme situation, some valid data are scarce in the real world.

Denote the domains of  $x$  and  $y$  by  $X$  and  $Y$ , respectively. Suppose that there is a probability distribution  $P$  on  $X \times Y$  and the estimation function  $\hat{f}$  is generated based on this distribution. Then for any  $\tau \in (0,1)$ , the  $\tau$ -quantile of  $y$  is given

by the infimum over  $\mu$  for which  $\Pr\{y \leq \mu|x\} = \tau$ .

Suppose that we have labeled samples by  $\{(x_1, y_1), \dots, (x_m, y_m)\}$ , with input  $x_i \in X$  and output  $y_i \in Y$ , and unlabeled samples by  $\{x_{m+1}, \dots, x_n\}$ , with only input  $x_i \in X$ . In the case where  $m = n$ , Li et al. (2007) suggested estimation of the conditional quantile  $f_\tau(x)$  can be obtained by the algorithm

$$\min_{f \in H_K} \sum_{i=1}^n \rho_\tau(y_i - f(x_i)) + \frac{\lambda}{2} \|f\|_{H_K}^2$$

where  $H_K$  is a Reproducing Kernel Hilbert Space (RKHS) on  $X$ , which is generated by a positive definite kernel  $K(x_i, x_j)$  and includes the entire family of smoothing splines, and many others, such as polynomial and radial basis. Here  $\lambda > 0$  controls the balance between the smoothness of the fit and its fidelity to the data.  $\rho_\tau(r)$  is the so-called “check function” (Koenker & Bassett 1978)

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

To incorporate labeled and unlabeled data to exploit estimation of the conditional quantile  $f_\tau(x)$ , we consider the extensive cases where  $m < n$ , and derive the nonparametric algorithm of utilizing the manifold regularization framework for semi-supervised learning

$$R(f_\tau) = \min_{f_\tau \in H_K} \sum_{i=1}^m \rho_\tau(y_i - f_\tau(x_i)) + \gamma_A \|g\|_K^2 + \gamma_I \|g\|_I^2 \quad (1)$$

where  $f_\tau = g + \alpha_0$ ,  $g \in H_K$ ,  $\alpha_0 \in \mathbb{R}$  is a scalar unregularized term. Using the representer theorem of Belkin et al. (2006), the solution to (1) has a finite form

$$\hat{f}_\tau(x) = \alpha_0 + \sum_{i=1}^n \alpha_i K(x_i, x). \quad (2)$$

in terms of labeled and unlabeled samples.

## 2 DESCRIPTION OF THE ALGORITHM

The optimization problem (1) is called by us Laplacian quantile regression based on manifold regularization (LapQR). By introducing such a regularizer  $\|g\|_l^2$  to the standard framework of estimating an unknown quantile function  $f_\tau$ , which should reflect the intrinsic structure of marginal distribution  $P_x$ , the graph structure is incorporated into the quantile regression problem. The parameter  $\gamma_A$  controls the complexity of the function in the ambient space, while  $\gamma_l$  controls the complexity of the function in the intrinsic geometry of  $P_x$ . Notice that  $\alpha_0$  isn't regularized in the optimization problem (1), which is inspired by the work of Takeuchi et al. (2006). This ensures that the minimizer of (1) satisfies the following property:

**Quantile Property** Assuming that  $f_\tau$  contains a scalar unregularized term, the minimizer of (1) satisfies:

- 1) The number  $N$  of terms  $m_-$  with  $y_i < f_\tau(x_i)$  is bounded from above by  $\tau m$ .
- 2) The number  $P$  of terms  $m_+$  with  $y_i > f_\tau(x_i)$  is bounded from above by  $(1-\tau)m$ .
- 3) If  $(x, y)$  is drawn iid from a distribution  $\Pr(x, y)$ , with  $\Pr(y|x)$  continuous and the expectation of the modulus of absolute continuity of its density satisfying  $\lim_{\delta \rightarrow 0} E[\varepsilon(\delta)] = 0$ . With probability 1, asymptotically,  $N/m$  equals  $\tau$ .

**Proof.** Without loss of generality, we set  $\hat{f}_\tau = \hat{g} + \hat{\alpha}_0$ . Then for any  $\alpha_0 \in \mathbb{R}$ ,  $R(\hat{g} + \alpha_0)$  has to be minimal when  $\alpha_0 = \hat{\alpha}_0$ , and  $\hat{\alpha}_0$  can be obtained by (3)

$$\min_{\alpha_0 \in \mathbb{R}} \sum_{i=1}^m \rho_\tau(y_i - \hat{g}(x_i) - \alpha_0) + \gamma_A \|\hat{g}\|_K^2 + \gamma_l \|\hat{g}\|_l^2 \quad (3)$$

With respect to  $\alpha_0$ , however, minimizing (3) amounts to finding the  $\tau$ -quantile in terms of  $y_i - \hat{g}(x_i)$ . Application of quantile properties in Takeuchi et al. (2006) proves the first two parts.

For the third part, an analogous reasoning to Proposition 1 of Schölkopf et al. (2000) applies. In a

nutshell, one uses the fact that the measure of the  $-\text{neighborhood}$  of  $f(x)$  converges to 0 for  $\delta \rightarrow 0$ . Moreover, for kernel functions the entropy numbers are well behaved (Williamson et al. 2001). The application of the union bound over a cover of such function classes completes the proof.+

In the empirical case, the penalty term  $\|g\|_l$  can be approximated on the basis of labeled and unlabeled data (Belkin et al. 2006)

$$\|g\|_l^2 = \frac{1}{n^2} \sum_{i,j=1}^n (g(x_i) - g(x_j))^2 W_{ij} = \frac{1}{n^2} g' L g \quad (4)$$

where  $W_{ij}$  are edge weights in the data adjacency graph,  $g = [g(x_1), \dots, g(x_n)]'$ , and  $L$  is the graph Laplacian given by  $L := D - W$ . The diagonal matrix  $D_{ii} = \sum_{j=1}^n W_{ij}$ . In the adjacency graph matrix  $W$ , the bigger weights of  $W_{ij}$  are, the closer distances between sample  $x_i$  and  $x_j$  will be. So minimizing the graph regularizer  $\sum_{i,j=1}^n (g(x_i) - g(x_j))^2 W_{ij}$  is equivalent to finding the function  $g$  which has a similar value for those samples geometrically close enough. In this sense,  $\|g\|_l^2$  is the penalty term that reflects the intrinsic geometric structure of the marginal distribution of the data set.

Using (2) and (4), criterion (1) can be rewritten in an equivalent way as

$$\min_{\substack{\alpha_0 \in \mathbb{R} \\ \alpha \in \mathbb{R}^n}} \tau \sum_{i=1}^m \xi_i + (1-\tau) \sum_{i=1}^m \xi_i^* + \gamma_A \alpha' K \alpha + \frac{\gamma_l}{n^2} \alpha' K L K \alpha$$

subject to

$$-\xi_i^* \leq y_i - \sum_{j=1}^n \alpha_j K(x_j, x_i) - \alpha_0 \leq \xi_i, i = 1, \dots, m$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, m$$

where parameter vector  $\alpha = [\alpha_1, \dots, \alpha_n]'$ , and Gram kernel matrix  $K$  is positive definite. Introducing Lagrange multipliers  $\lambda_i, \mu_i, \nu_i, \omega_i$ , and setting the derivatives of  $L$  to 0, for  $i = 1, \dots, m$ , we get

$$\sum_{i=1}^m \lambda_i = \sum_{i=1}^m \mu_i, \mu_i = \tau - \nu_i, \lambda_i = 1 - \tau - \omega_i$$

Then we can get a reduced Lagrangian function:

$$L(\alpha, \lambda, \mu) = \frac{1}{2} \alpha' \left( 2\gamma_A K + 2\frac{\gamma_l}{n^2} K L K \right) \alpha + y'(\mu - \lambda) - \alpha' K E'(\mu - \lambda)$$

where  $y = [y_1, \dots, y_m]'$ ,  $E = [I \ 0]_{m \times n}$ , and  $I$  is an  $m \times m$  identity matrix,  $0$  is an  $m \times (n-m)$  zero

matrix. Setting the derivative of the reduced Lagrangian, we arrive at

$$\frac{\partial L(\alpha, \lambda, \mu)}{\partial \alpha} = 0 \Rightarrow \hat{\alpha} = (2\gamma_A I + 2\frac{\gamma_I}{n^2} LK)^{-1} E'(\hat{\mu} - \hat{\lambda})$$

Since  $0 \leq \mu_i \leq \tau$  and  $0 \leq \lambda_i \leq 1 - \tau$ , we get a new quadratic optimization with linear constraint conditions and dimensionality reduction

$$\min_{(\mu - \lambda) \in \mathbb{R}^m} \frac{1}{2} (\mu - \lambda)' E \left[ (2\gamma_A I_{n \times n} + 2\frac{\gamma_I}{n^2} LK)^{-1} \right]' \times \\ KE'(\mu - \lambda) - y'(\mu - \lambda)$$

subject to

$$\sum_{i=1}^m (\mu_i - \lambda_i) = 0, \mu_i - \lambda_i \in [-(1 - \tau), \tau], i = 1, \dots, m$$

which can be solved by most commercially available packages.

As for  $\hat{\alpha}_0$ , we can substitute it by some points of  $\{y_i - \hat{g}(x_i)\}$  or any value in some interval. In fact, using the Karush-Kuhn-Tucker conditions, we can define the following three sets,

$$E = \{i : y_i - \hat{f}_\tau(x_i) = 0, -(1 - \tau) \leq \mu_i - \lambda_i \leq \tau\},$$

$$L = \{i : y_i - \hat{f}_\tau(x_i) < 0, \mu_i - \lambda_i = -(1 - \tau)\},$$

$$R = \{i : y_i - \hat{f}_\tau(x_i) > 0, \mu_i - \lambda_i = \tau\}.$$

If  $y_i \neq \hat{f}_\tau(x_i)$ ,  $\mu_i - \lambda_i$  are all determined, so do  $\hat{\alpha}_j$  and  $\hat{g}(x_i)$ . Then we only need to focus on the situation of  $y_i = \hat{f}_\tau(x_i)$ . When  $E \neq \emptyset$ , some points  $(x_{i^*}, y_{i^*})$  of labeled data make  $y_{i^*} = \hat{f}_\tau(x_{i^*})$ , we have  $\hat{\alpha}_0 = y_{i^*} - \hat{g}(x_{i^*})$ , then  $\hat{\alpha}_0$  can take the value with

$$\hat{\alpha}_0 = \frac{1}{n_S} \sum_{i^* \in S} (y_{i^*} - \hat{g}(x_{i^*}))$$

where  $S = \{i \in E : -(1 - \tau) < \mu_i - \lambda_i < \tau\}$ , and  $n_S$  denotes the number of elements in  $S$ . When  $E = \emptyset$ , no point  $(x_{i^*}, y_{i^*})$  of labeled data makes  $y_{i^*} = \hat{f}_\tau(x_{i^*})$ , but we also have

$$\hat{\alpha}_0 \in \left[ \max_{i \in L} (y_i - \hat{g}(x_i)), \min_{i \in R} (y_i - \hat{g}(x_i)) \right].$$

When the length of the above interval changes and decreases to 0, we can obtain the approximation value of  $\hat{\alpha}_0$ . Later, we will use the middle point of the above interval as the substitution of  $\hat{\alpha}_0$  when we illustrate LapQR algorithm's efficiency in section4.

### 3 MODEL SELECTION CRITERION

In practice, one can first use the efficient algorithm in Section 2 to compute the entire solution path, then identify the appropriate value of model parameter that minimizes certain model selection criterion. This avoids the computationally more intensive cross-validation method. In the nonparametric setting, the Schwarz Information Criterion (SIC) is ad hoc and requires considerable further investigation. GACV is an approximation to leave-one-out quantile cross-validation (Yuan 2006), which has the form as

$$GACV = \frac{\sum_{i=1}^m \rho_\tau(y_i - \hat{f}_\tau(x_i))}{m - d\hat{f}_\tau}.$$

Yuan (2006) proposed using the divergence formula

$$\text{div}(\hat{f}_\tau) = \sum_{i=1}^m \frac{\partial \hat{f}_\tau(x_i)}{\partial y_i} = |E|$$

for  $d\hat{f}_\tau$ , which turns out that  $\sum_{i=1}^m \partial \hat{f}_\tau(x_i) / \partial y_i$  can be considered as an estimate of the effective dimension for a general modeling procedure (Stein 1981; Efron 1986; Ye 1998; Meyer & Woodroffe 2000), while  $|E|$  is a convenient estimate for the effective dimension of  $\hat{f}_\tau(x)$ . We use the following form of GACV as model selection criterion

$$GACV = \frac{\sum_{i=1}^m \rho_\tau(y_i - \hat{g}(x_i))}{m - |E|}.$$

To compute the empirical conditional quantile estimator  $\hat{f}_\tau$ , we need to construct a weighted graph for the given point set, and the procedure of estimating  $\hat{f}_\tau$  is stated as follows:

**Input:** Labeled examples  $\{(x_i, y_i)\}_{i=1}^m$  and unlabeled examples  $\{x_j\}_{j=m+1}^n$ .

**Output:** Estimated function  $\hat{f}_\tau : \mathbb{R}^d \rightarrow \mathbb{R}$ .

**Step1:** Construct data adjacency graph with  $n$  nodes using, for example,  $k$  nearest neighbors or a graph kernel. Choose edge weights  $W_{ij}$ , for example, binary weights or heat kernel weights  $W_{ij} = \exp(-\|x_i - x_j\|^2 / 4t)$ .

**Step2:** Choose a kernel function  $K(x, y)$ . Compute Gram matrix  $K_{ij} = K(x_i, x_j)$ .

**Step3:** Compute graph Laplacian matrix:  $L = D - W$  where  $D$  is a diagonal matrix given by  $D_{ii} = \sum_{j=1}^n W_{ij}$ .

**Step4:** Choose  $\gamma_A$  and  $\gamma_I$  using the model selection criterion (e.g., the new GACV).

**Step5:** Compute  $\hat{\alpha}$  and  $\hat{\alpha}_0$  using LapQR algorithm of Section 2.

**Step6:** Output function  $\hat{f}_\tau(x) = \sum_{i=1}^n \hat{\alpha}_i K(x_i, x) + \hat{\alpha}_0$ .

#### 4 EXPERIMENT

For illustrating the finite sample behavior of the proposed LapQR, we generate randomly sample set of size  $n$  with  $m$  labeled observations and  $(n-m)$  unlabeled observations from the following model

$$y(x) = \text{sinc}(x) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2(x)),$$

where  $\sigma(x) = 0.1 \exp(1-x)$ , and  $x \sim U[-1, 1]$ . Since  $\varepsilon$  is normally distributed, the  $\tau$ -quantile of  $\varepsilon$  is given by  $\sigma(x)\Phi^{-1}(\tau)$ , where  $\Phi$  is the cumulative distribution function of the normal distribution with unit variance. This means that the conditional  $\tau$ -quantile function of  $y$  is

$$f_\tau(x) = \text{sinc}(x) + \sigma(x)\Phi^{-1}(\tau).$$

We construct adjacent graph by using  $k$  nearest neighbors with binary weights, and use Gaussian radial basis  $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$ . The parameters are selected by the new GACV outlined in Section 3. We repeat this simulation 50 times at

different values of  $\tau$ : 0.1, 0.25, and 0.5. We also calculate the average amount of quantile  $\hat{f}_\tau(x_i)$  (AQs) which exceeds  $y_i$  with probability close to  $\tau$ , which ensures the estimator satisfies the quantile property, and calculate the average mean square errors (MSEs) and the average prediction errors (PEs) as follows:

$$MSE = \frac{1}{m} \sum_{i=1}^m (f_\tau(x_i) - \hat{f}_\tau(x_i))^2,$$

$$PE = \frac{1}{m} \sum_{i=1}^m \rho_\tau(y_i - \hat{f}_\tau(x_i)).$$

The results of Tables 1-3 show that there are several trends in the results: the performance of LapQR without randomly unlabeled data tends to yield better performance than that with randomly unlabeled data; the more data are used, the better performance will be; the closer  $\tau$  gets to 0.5, the better performance will be, because there are more valid samples to be used. Figure 1 shows the fitting results by using sample size 400 without randomly unlabeled data at  $\tau=0.5$ .

Table 1. Simulation results of Example 1,  $\tau=0.5$ .

n m/n(%)		50	100	200	400
60	MSE	0.1115 (0.0387)	0.0923 (0.0254)	0.0922 (0.0293)	0.0895 (0.0157)
	AQ	0.4667(0.0997)	0.5333(0.0538)	0.5320(0.0381)	0.4850 (0.0323)
	PE	0.1475(0.0284)	0.1514(0.0294)	0.1466(0.0175)	0.1487(0.0095)
80	MSE	0.0903(0.0351)	0.0510(0.0529)	0.0495(0.0608)	0.0483(0.0428)
	AQ	0.4750(0.0536)	0.4995(0.0493)	0.5133(0.0445)	0.4850(0.0352)
	PE	0.1386(0.0223)	0.1242(0.0254)	0.1128(0.0154)	0.1082(0.0072)
100	MSE	0.0548(0.0202)	0.0396(0.0104)	0.0115(0.0123)	0.0048(0.0024)
	AQ	0.4700(0.0690)	0.5105(0.0283)	0.5325(0.0247)	0.4850(0.0312)
	PE	0.1168(0.0205)	0.1141(0.0178)	0.1072(0.0066)	0.1047(0.0054)
NOTE: The numbers in parentheses are the corresponding standard errors.					

Table 2. Simulation results of Example 1,  $\tau=0.25$ .

n m/n(%)		50	100	200	400
60	MSE	0.1597(0.0252)	0.1433(0.0234)	0.1380(0.0233)	0.1122(0.0151)
	AQ	0.3889(0.1262)	0.3450(0.1236)	0.3222(0.1134)	0.3200(0.0933)
	PE	0.1479(0.0126)	0.1349(0.0137)	0.1462(0.0125)	0.1315(0.0104)
80	MSE	0.1464(0.0249)	0.1217(0.0229)	0.1145(0.0158)	0.1036(0.0128)
	AQ	0.3733(0.1076)	0.3265(0.1032)	0.3325(0.0972)	0.3275(0.0967)
	PE	0.1332(0.0114)	0.1345(0.0124)	0.1337(0.0117)	0.1242(0.0092)
100	MSE	0.1085(0.0220)	0.0963(0.0203)	0.0675(0.0123)	0.0428(0.0124)
	AQ	0.3350(0.0995)	0.3200(0.0683)	0.3025(0.0524)	0.2975(0.0518)
	PE	0.1142(0.0155)	0.1261(0.0139)	0.1346(0.0106)	0.1239(0.0052)
NOTE: The numbers in the parentheses are the corresponding standard errors.					

Table 3. Simulation results of Example 1,  $\tau=0.1$ .

n m/n(%)		50	100	200	400
60	MSE	0.1651(0.0587)	0.1443(0.0545)	0.1246(0.0396)	0.1194(0.0234)
	AQ	0.3700(0.2987)	0.3372(0.2717)	0.3238(0.2248)	0.2529(0.1857)
	PE	0.2063(0.0622)	0.1824(0.0471)	0.1761(0.0472)	0.1528(0.0388)
80	MSE	0.1530(0.0451)	0.1311(0.0329)	0.1232(0.0308)	0.0964(0.0228)
	AQ	0.3589(0.2611)	0.3128(0.2373)	0.2922(0.2312)	0.2445(0.1671)
	PE	0.1964(0.0556)	0.1841(0.0424)	0.1669(0.0374)	0.1492(0.0356)
100	MSE	0.1348(0.0367)	0.1239(0.0304)	0.0925(0.0223)	0.0842(0.0184)
	AQ	0.3200(0.2490)	0.3062(0.2343)	0.2532(0.2542)	0.2050(0.1549)
	PE	0.1864(0.0465)	0.1541(0.0372)	0.1563(0.0326)	0.1397(0.0252)

NOTE: The numbers in the parentheses are the corresponding standard errors.

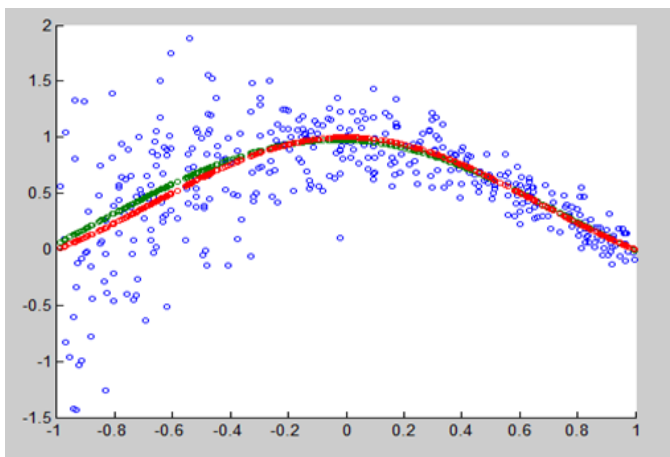


Figure 1 Simulation with 400 observations,  $\tau=0.5$ . The blue circles are 400 observations, the red curve is the fitted quantile function, the green curve is the true quantile function.

## 5 CONCLUSION

In this article we propose an efficient quantile regression algorithm that incorporates labeled and unlabeled data to compute the entire regularization path of the LapQR, we also provide convergence statements and bounds on the quantile estimator. And the experimental results show the feasibility of our new approach and the LapQR is a promising nonparametric method of quantile estimation.

## ACKNOWLEDGEMENTS

This work is supported by Project funded by China Postdoctoral Science Foundation (2014M552068).

## REFERENCES

[1] Belkin, M. et al. 2006. Manifold Regularization: A Geometric Framework for Learning from Examples. *Journal of Machine Learning Research* 7: 2399-2434.  
[2] Cole, T. & Green, P. 1992. Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood.

*Statistics in Medicine* 11: 1305–1319.  
[3] Efron, B. 1986. How Biased Is the Apparent Error Rate of a Prediction Rule?. *Journal of the American Statistical Association* 81: 461–470.  
[4] Heagerty, P. & Pepe, M. 1999. Semiparametric Estimation of Regression Quantiles With Application to Standardizing Weight for Height and Age in U.S. Children. *Journal of the Royal Statistical Society (Ser. C)* 48: 533–551.  
[5] Hendricks, W. & Koenker, R. 1992. Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. *Journal of the American Statistical Association* 87: 58–68.  
[6] Koenker, R. & Bassett, R. 1978. Regression Quantile. *Econometrica* 46: 33-50.  
[7] Koenker, R. & Geling, O. 2001. Reappraising Medfly Longevity: A Quantile Regression Survival Analysis. *Journal of the American Statistical Association* 96: 458–468.  
[8] Koenker, R. & Hallock, K. 2001. Quantile Regression. *Journal of Economic Perspectives* 15: 143–156.  
[9] Li, Y. et al. 2007. Quantile Regression in Reproducing Kernel Hilbert Space. *Journal of the American Statistical Association* 102: 255-268.  
[10] Meyer, M. & Woodroffe, M. 2000. On the Degrees of Freedom in Shape-Restricted Regression. *The Annals of Statistics* 28: 1083–1104.  
[11] Schölkopf, A.J. et al. 2000. New support vector algorithms. *Neural Computation* 12: 416-426.  
[12] Stein, C. 1981. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics* 9: 1135–1151.  
[13] Takeuchi, I. et al. 2006. Nonparametric quantile estimation. *Journal of Machine Learning Research* 7: 1231-1264.  
[14] Williamson, R.C. et al. 2001. Generalization bounds for regularization networks and support vector machines via entropy numbers of compact operators. *IEEE Transaction on Information Theory* 47: 2516–2532.  
[15] Yang, S. 1999. Censored Median Regression Using Weighted Empirical Survival and Hazard Functions. *Journal of the American Statistical Association* 94: 137–145.  
[16] Ye, J. 1998. On Measuring and Correcting the Effects of Data Mining and Model Selection. *Journal of the American Statistical Association* 93: 120–131.  
[17] Yuan, M. 2006. GACV for Quantile Smoothing Splines. *Computational Statistics and Data Analysis* 50: 813-829.