Efficient Model Selection in Linear and Non-linear Quantile Regression by Cross-validation

Yoonsuh Jung*, and Steven N. MacEachern

Abstract—Check loss function is used to define quantile regression. In the prospect of cross validation, it is also employed as a validation function when underlying truth is unknown. However, our empirical study indicates that the validation with check loss often leads to choose an over estimated fits. In this work, we suggest a modified or L2-adjusted check loss which rounds the sharp corner in the middle of check loss. It has a large effect of guarding against over fitted model in some extent. Through various simulation settings of linear and non-linear regressions, the improvement of check loss by L2 adjustment is empirically examined. This adjustment is devised to shrink to zero as sample size grows.

Keywords—Cross-validation, Model Selection, Quantile Regression, Tuning Parameter Selection

I. Introduction

MODEL selection is a crucial problem in statistical analysis. As a popular data driven method for model selection, cross validation is described formally in [17] and [5]. [20] proposed cross-validation for selection of the smoothing parameter in smoothing spline regression. Among the many authors who have considered cross validation, [15] provides a rigorous treatment of model selection under the linear model. [4] argued that cross validation is nearly unbiased for the future error rate, but often highly variable, and suggested a new bootstrapping method. [9] provides a thorough review and performs experiments comparing cross validation and bootstrap.

Typically, the loss criterion used for model fitting is employed for model validation. Just as certain modification of a loss criterion could bring robustness or better efficiency in modeling, validation with a modified criterion could lead to selection of a better model. For example, to reduce the influence of outliers in model selection, [13] proposed cross validation with a robust loss function when the squared error loss is used for fitting a linear model. [11] discusses a similar issue in nonparametric regression setting.

In k-fold cross validation, the choice of k is an interesting theme [9]. When k is equal to the sample size, k-fold cross validation is also called leave-one-out cross validation. In the linear model, leave-one-out cross validation is shown to be asymptotically equivalent to AIC [18] and inconsistent [16]. Like AIC, this method is too conservative and suffers from asymptotic difficulties [15], which may lead to a larger and unrealistic model [3].

Y. Jung is a Senior Lecturer in the Department of Mathematics and Statistics, University of Waikato, Hamilton, 3240, New Zealand. e-mail: yoonsuh@waikato.ac.nz, *Corresponding author

S. N. MacEachern is a Professor in the Department of Statistics, The Ohio State University, Columbus, OH, 43210, USA. e-mails: snm@stat.osu.edu

In this work, we mainly focus on cross validation in quantile regression settings. In quantile regression, the check loss function is commonly used for validation as well as fitting. [10] introduced a modified check loss function for fitting a quantile regression model. Here, we consider use of the modified check loss as a validation function for cross validation (CV.M) , and compare its performance to that of cross validation using the check loss (CV). Since the modified check loss is designed to gain the efficiency at the expense of allowing greater bias, the suggested method can be regarded as applying a new version of bias-variance tradeoff to the problem of estimating future error. Since the purpose of this work is not to discuss the choice of k but to examine the difference of two validation functions, k is fixed at 10.

In Section 2, a new validation function is introduced with the details of modification. Simulation studies are in Section 3 to compare the performance of CV and CV.M under various situations of linear and non-linear regression models.

II. Modified Cross Validation Function

The check loss function ρ_q has the form of

$$\rho(u) = \begin{cases} qu & \text{for } u \geq 0 \\ -(1-q)u & \text{for } u < 0. \end{cases} \tag{1}$$

To estimate qth regression quantile, [10] rounded in the middle of ρ_q as follows.

$$\rho^M(u) = \begin{cases} qu - \frac{q(1-q)}{2\lambda_q} & \text{for} \quad \frac{1-q}{\lambda_q} \leq u \\ \frac{\lambda_q}{2}\frac{q}{1-q}u^2 & \text{for} \quad 0 \leq u < \frac{1-q}{\lambda_q} \\ \frac{\lambda_q}{2}\frac{1-q}{q}u^2 & \text{for} \quad -\frac{q}{\lambda_q} \leq u < 0 \\ (q-1)u - \frac{q(1-q)}{2\lambda_q} & \text{for} \quad u < -\frac{q}{\lambda_q}, \end{cases} \tag{2}$$

where λ_q is a tuning parameter for the size of adjustment that depends on the qth target quantile. We call this adjustment a 'window'. Since it is desirable to have a scale invariant window that diminishes as sample size increases, we set $\lambda_q := c_q n^{\alpha}/\hat{\sigma}$, where n is the sample size, α is a positive constant, and $\hat{\sigma}$ is a robust scale estimate of the error distribution. [6] suggest a heuristic rule,

$$c_q \approx \begin{cases} 0.5e^{-2.118-1.097q} & \text{for } q < 0.5 \\ 0.5e^{-2.118-1.097(1-q)} & \text{for } q \geq 0.5, \end{cases} \tag{3}$$

with fixed $\alpha = 0.3$, which will be used for model validation. But, note that the length of adjustment for validation differs from that for fitting. For k-fold cross validation, $(k-1)$ folds are used for model fitting, and only one fold is used for validation. Thus the window width, which is based on the sample size in the validation set, is larger than the window

width used to fit a modified quantile regression. Notice that the modified check loss can be employed in both fitting and validation. To separate the effect of the loss function on fitting from that on validation, we restrict our attention to the effect of modification of the loss for validation only, thus for fitting purpose, check loss is always applied.

First, observe that the modified check loss is a consistent validation function. When $\alpha > 0$, we immediately see that $\rho^M(u) \to \rho(u)$ for every u. Furthermore, the absolute difference between the check loss and modified check loss is uniformly bounded. That is, $sup_u|\rho^M(u) - \rho(u)| = \frac{q(1-q)}{\lambda_\gamma}$. When q=1/2, the validation function in (2) becomes Huber's function. The performances of absolute deviation, squared loss and Huber's loss function as a validation function are compared in [11] for nonparametric kernel regression. A simulation study in the next section investigates the finite sample performance of cross validation with the modified check loss function.

III. Simulations

The behavior of the modified check loss function in (2) as a validation criterion is considered under various statistical procedures. The modified check loss is compared to the check loss for validation under the three settings of 1) the linear model, 2) regression splines, and 3) smoothing splines.

A. Validation under the Linear Model

Three linear models that are adapted from [19] which represent sparse case (β=(5, 0, 0, 0, 0, 0, 0, 0)), intermediate case (β=(3, 1.5, 0, 0, 2, 0, 0, 0)), and dense case β=(0, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0). Assuming $y = \beta_0 + x^\top\beta + \epsilon$, we generated $x = (x_1, \ldots, x_8)^\top$ from a multivariate normal distribution with mean zero and standard deviation 1. The correlation between x_i and x_j was set to $\rho^{|i-j|}$ with $\rho = 0.5$. We consider selection of a linear model among all possible 256 $(= 2^8)$ subset models.

The standard quantile regression models with smallest prediction error are selected by 10-fold cross validation based on two loss functions; check loss (CV) and modified check loss (CV.M). The scale parameter, σ, is estimated from the full least squares model. To gauge the performance of CV and CV.M, mean squared error (MSE) is defined as,

$$\begin{aligned} MSE &= E^{\hat\beta,X}||(X^\top\hat\beta + \hat\beta_0) - (X^\top\beta + \beta_0)||^2 \\ &= E^{\hat\beta,X}\{(\hat\beta - \beta)^\top X^\top X(\hat\beta - \beta) + (\hat\beta_0 - \beta_0)^2\} \\ &= E^{\hat\beta}\{(\hat\beta - \beta)^\top\Sigma(\hat\beta - \beta) + (\hat\beta_0 - \beta_0)^2\}, \end{aligned} \tag{4}$$

which is approximated by a Monte Carlo estimate from 1000 data sets:

$$\widehat{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}((\hat\beta^i - \beta)^\top\Sigma(\hat\beta^i - \beta) + (\hat\beta_0^i - \beta_0)^2).$$

Note that $\widehat{MSE}$ values from CV and CV.M are equal only when the two methods choose the same model over every replicate. $\widehat{MSE}$ from CV ($\widehat{MSE}_{CV}$) and $\widehat{MSE}$ from CV.M ($\widehat{MSE}_{CV.M}$) gauge the relative accuracy of the two validation functions in assessing the fitted models. Table I shows the percentage reduction under three linear models, along with approximate confidence intervals for the corresponding theoretical percentage reductions. The variance of the asymptotic distribution of the above quantity is approximated through the delta method, which suggests the use of

$$\widehat{Var}(\frac{\bar Y}{\bar X}) \approx \frac{1}{n}\left[\frac{\bar y^2}{\bar x^4}s_x^2 + \frac{1}{\bar x^2}s_y^2 - \frac{2\bar y}{\bar x^3}\hat\rho s_x s_y\right].$$

TABLE I
Point estimates and approximate 95% confidence intervals for percentage reduction in mean MSE, based on 1000 replicates with n=500, at selected quantiles. Sparse, intermediate and dense cases are considered.

	q=.1	q=.25	q=.5	q=.8
Sparse	3.794	9.328	16.451	7.640
	(2.602,4.986)	(7.750,10.906)	(14.811,18.091)	(6.098,9.182)
Intermediate	1.333	3.854	8.246	3.417
	(0.411,2.256)	(2.716,4.992)	(6.822,9.570)	(2.332,4.503)
Dense	0.556	1.036	2.298	0.530
	(0.190,0.921)	(0.365,1.708)	(1.739,2.858)	(0.072,0.987)

The various simulation settings in Table I reveal that the modified validation function outperforms the traditional check loss function for all three linear models. As we move from the dense case to the sparse case, the advantage of CV.M over CV grows. When the true regression coefficients are all substantial and non-zero, the two methods are equivalent up to simulation variation (results are omitted). There is certainly a lower bound under which $\widehat{MSE}_{CV}$ and $\widehat{MSE}_{CV.M}$ cannot go. Given a data set, $\widehat{MSE}$ is defined as the minimum MSE of all possible subset models. Treating $\widehat{MSE}_{min}$ as an achievable base, we define the excess (*excess* MSE) to be $\widehat{MSE}_{CV} - \widehat{MSE}_{min}$ and $\widehat{MSE}_{CV.M} - \widehat{MSE}_{min}$ for the two validation functions. We then compute the percentage reduction in *excess* MSE. Table II presents the results. Again, we see that CV.M provides a more accurate assessment of the linear quantile regression model than CV does.

From Tables I and II, the modified check loss function presents its potential as a validation function for linear quantile regression. In subsequent sections, we turn our attention to the performance of CV.M for nonlinear models.

In addition to the performance in terms of MSE, the extent of disagreement between the models selected by CV and CV.M is also of interest. Table III summarizes the number of agreements and disagreements among 1000 replicates. Disagreement tends to increase as the probability density at the target quantile increases, and as the scenario moves from 'dense' to 'sparse'.

B. Validation under Quantile Regression Splines

Two simple simulation settings (adapted and modified from [21]) are considered to observe the performance of CV and CV.M under nonparametric quantile regression. Regression spline models with a natural spline basis expansion are fitted to the simulated data. A reasonable range of the number of interior knots is considered, and the number of knots is selected by 10-fold cross validation using either check loss or modified check loss. Data sets were simulated from

TABLE II
POINT ESTIMATES AND APPROXIMATE 95% CONFIDENCE INTERVALS FOR PERCENTAGE REDUCTION IN MEAN *excess* MSE, BASED ON 1000 REPLICATES WITH N=500, AT SELECTED QUANTILES. SPARSE, INTERMEDIATE AND DENSE CASES ARE CONSIDERED.

	q=.1	q=.25	q=.5	q=.8
Sparse	5.078	12.590	22.349	10.355
	(3.494,6.662)	(10.500,14.681)	(20.189,24.508)	(8.283,12.427)
Intermediate	2.608	7.755	16.362	6.795
	(0.810,4.406)	(5.503,10.006)	(13.834,18.890)	(4.666,8.924)
Dense	2.870	6.044	13.154	2.709
	(0.995,4.744)	(2.232,9.856)	(10.108,16.200)	(0.394,5.024)

TABLE III
NUMBER OF AGREEMENTS AND DISAGREEMENTS OF MODELS SELECTED BY CV AND CV.M AMONG 1000 REPLICATES IN EACH SCENARIO OF SPARSE, INTERMEDIATE AND DENSE AT SEVERAL QUANTILES. '=' , '+', AND '−' REPRESENT SELECTION OF THE SAME MODEL, SELECTION OF A BETTER MODEL (IN TERMS OF MSE BY CV.M), AND SELECTION OF A WORSE MODEL BY CV.M RESPECTIVELY.

	Sparse	Intermediate	Dense
	(=,+,–)	(=,+,–)	(=,+,–)
q=0.1	(694,193,113)	(795,130,75)	(914,56,30)
q=0.25	(535,309,156)	(644,222,134)	(841,89,70)
q=0.5	(471,396,133)	(584,284,132)	(810,132,58)
q=0.8	(546,308,146)	(679,202,119)	(866,77,57)

1. Simple quantiles; $y_i = 2 + 2\cos(x_i) + \exp(-4x_i^2) + \epsilon_i$,
2. Smooth "curvy" quantile: $y_i = 2.5 + \sin(2x_i) + 2\exp(-16x_i^2) + 0.5\epsilon_i$.

The covariate X follows a standard normal distribution, independent of ϵ_i. Two error distributions, standard normal and Exp(1), are employed for ϵ_i. True curves at some quantiles are shown in Figure 1 when the errors follow normal distribution. In each case, 2000 replicates of sample size n=200, 500, and 1000 were generated.

MSE is approximated by a Monte Carlo estimate given by

$$\widehat{MSE} = \frac{1}{2000}\sum_{j=1}^{2000}\frac{1}{n}\sum_{i=1}^{n}(f_s(x_i) - \hat{f}_s^j(x_i))^2,$$

where s=1 for the true underlying function in simulation 1, and s=2 for simulation 2. j indicates the jth replicate. Since there is no "true set of knots" in this simulation study, the set of knots which produces smallest MSE for a given data set is regarded as the base model being pursued. Thus, the base model depends on the generated data. The check loss and the modified check loss do sometimes select the base model, although they more typically show large deviations from the base model. CV.M selects the base model more often than CV does across all simulations carried out. For example, under the normal error distribution in the first simulation, the check loss picks the base models about 20% of the time whereas the modified check loss selects them correctly around 25% of the time. In the first simulation, both loss functions show a higher rate of selecting the base model than in the second simulation, probably due to the simpler form of the first curve, as shown in Figure 1. The percentage reduction in MSE and *excess* MSE when switching from check loss to modified check loss is calculated. The scale parameter, σ, is estimated by the standard deviation of the residuals from a mean smoothing spline regression fit.

Fig. 1. True curves for simulation 1 (left) and simulation 2 (right), at q= 0.1, 0.25, 0.5, and 0.8 quantiles, under a standard normal error distribution.

TABLE IV
POINT ESTIMATES AND APPROXIMATE 95% CONFIDENCE INTERVALS FOR PERCENTAGE REDUCTION IN MEAN MSE AND MEAN *excess* MSE UNDER SIMULATION 1, BASED ON 2000 REPLICATES WITH n=200, AT SELECTED QUANTILES. STANDARD NORMAL ERROR AND EXP(1) ARE CONSIDERED.

Error	N(0,1)	N(0,1)	Exp(1)	Exp(1)
	MSE	*excess*MSE	MSE	*excess*MSE
q=0.1	3.579	7.244	4.859	10.515
	(2.267,4.891)	(4.637,9.852)	(2.805,6.914)	(6.219,14.811)
q=0.25	5.469	11.038	4.809	10.717
	(3.863,7.076)	(7.870,14.207)	(2.912,6.705)	(6.608,14.826)
q=0.5	12.934	27.66	3.205	6.955
	(11.074,14.794)	(24.03,31.30)	(0.950,5.460)	(2.135,11.774)
q=0.8	5.605	11.89	8.463	16.728
	(4.156,7.056)	(8.89,14.88)	(6.744,10.181)	(13.463,19.991)
q=0.9	4.828	10.034	5.681	10.698
	(3.370,6.285)	(7.099,12.969)	(4.22,7.14)	(8.024,13.371)

Table IV shows these quantities for the first simulation, when n is 200. Again, the *excess* MSE is obtained by subtracting the MSE of a base model. In every situation considered, modified check loss outperforms check loss.

On average, CV and CV.M select the same model about 75% of the time in the two simulations. Thus the reductions in MSE and *excess* MSE are based on only about 25% of the replicates, which makes the CV.M stand out: Given selection of different models, the reductions in MSE and *excess* MSE are substantial.

C. Validation under Quantile Smoothing Splines

In this section, the performance of the check loss and the modified check loss are compared under quantile smoothing

TABLE V
POINT ESTIMATES AND APPROXIMATE 95% CONFIDENCE INTERVALS FOR PERCENTAGE REDUCTION IN MEAN MSE AND MEAN *excess* MSE BASED ON 1000 REPLICATES WITH n=500, AT SEVERAL QUANTILES. NORMAL ERROR WITH MEAN ZERO, AND STANDARD DEVIATIONS 0.2 IS CONSIDERED.

Reduction	q=0.1	q=0.2	q=0.3
MSE	4.83	6.07	9.77
	(3.27,6.38)	(4.17,7.97)	(7.36,12.18)
excessMSE	13.156	17.86	28.41
	(9.21,17.10)	(12.72,23.00)	(22.55,34.28)
Reduction	q=0.4	q=0.5	
MSE	9.39	10.11	
	(7.09,11.68)	(8.12,12.11)	
excessMSE	29.69	30.14	
	(23.49,35.88)	(25.06,35.22)	

splines. Several versions of quantile smoothing splines are suggested, with slightly different forms. [1] and Jones in the discussion of [2] have proposed estimating a quantile smoothing spline model that minimizes

$$\sum_{i=1}^{n} \rho(y_i - g(x_i)) + \lambda \int (g''(x))^2 dx.$$

[8] suggested use of an ℓ_1 roughness penalty, where the above $(g''(x))^2$ is replaced by $|g''(x)|$. [12] proposed to round out the corner of the check loss in a small interval $(-\epsilon, \epsilon)$ around zero in order to make the loss function differentiable and thus improve computation. This modification is similar in form to our modified check loss. Two essential differences lie in the fact that the adjustment in [12] is chosen to be effectively zero relative to the residuals, and that our modification has an asymmetric window about zero for quantiles other than the median. By adopting [12]'s quantile smoothing splines via the function `qsreg(fields)` in the R package, we compare the fitted models selected by the check loss and the modified check loss. Again, σ needed for the window width of the modified check loss is estimated by the standard deviation of the residuals from the mean smoothing spline regression fit.

Data from sinusoid curve with period 1, along with several error distributions are simulated. That is,

$$y_i = sin(2\pi x_i) + \epsilon_i, i = 1, ..., n,$$

where x_i's are iid from the standard uniform distribution, and ϵ_i's are iid from some error distribution. Error distributions considered are normal, t, shifted gamma, and shifted exponential distribution with median 0, and standard deviation 0.2. The t distributions, with 5 and 10 degrees of freedom, are scaled to have standard deviation 0.2. A fine grid search of the smoothing parameter is conducted to select the 'best' value by CV and CV.M. As in the previous section, the performance of the two validation functions is judged by the percentage reduction in MSE and *excess* MSE. To calculate the *excess* MSE, a base model is defined by the minimum MSE across the fine grid of smoothing parameter value given a data set. The simulation consists of 1000 replicates with sample size 500, 1000 and 2000.

The simulation results from a normal error distribution at several quantiles are shown in Tables V. The results for $q > 0.5$ are similar to those for $q < 0.5$, due to the symmetry of the normal distribution. There are considerable reductions when CV.M is employed, indicating that the averaging effect of CV.M when evaluating the residuals near the target quantile induces better assessment of a fitted model. The results for the other error distributions mentioned above also show reduction in mean MSE and mean *excess* MSE in all cases but are omitted due to page limit.

Overall, quantiles at high density lead to greater reduction in MSE. This tendency holds not only within an error distribution, but also between the distributions. For example, MSE reductions grow as we move from a sharp tail (normal) to thicker tails (t with df=10 and 5) at the tail quantiles. For the same reason, the extent of reductions near the median becomes smaller as we change from the normal distribution to a t with 10 degrees of freedom and further to a t with 5 degrees of freedom.

The same smoothing parameters are selected by CV and CV.M for roughly 50% to 75 % of the replicates. The remaining 25% to 50% of the replicates account for the discrepancy in the assessment of the fitted curve under CV and CV.M. Examination of the cases with a difference shows that the improvement of CV.M is mainly due to not ov The under smoothing with CV is partly demonstrated by the plots of the fitted functions which produce maximum or minimum values of $MSE_{CV.M}/MSE_{CV}$. The Figure 2 shows the selected models by CV and CV.M which yield the minimum value of $MSE_{CV.M}/MSE_{CV}$. In this figure, CV is poor compared to CV.M, and we can clearly observe that CV over fits the data. To the contrary, Figure 3 shows the data sets where CV.M is the poorest, relative to CV. Here, it seems that CV.M sometimes slightly over smooths data. To our eyes, the mild over smoothing in Figure 3 is preferable to the severe under smoothing in the Figure 2. The earlier percentage reduction in MSE calculations confirm this impression. Figure 4 presents the distributions of the estimated smoothing parameter $\hat{\lambda}$ (in log scale) that are centered around $\hat{\lambda}$ from the base model. In Figure 4, only those $\hat{\lambda}$ pairs with different values are included. The long lower tail of the $\hat{\lambda}$ from CV is another indication of under smoothing.

As stated earlier, the benefits of CV.M resulted from the bias-variance tradeoff of estimating future error. This fact is readily verified when MSE value is decomposed to variance and squared bias. Table VI shows the decomposition of variance and squared bias under standard normal error distribution. We can see that the reduction in variance is much larger than the increase in squared bias, thus leading to smaller MSE values. Results for the other error distributions are similar and omitted.

IV. CONCLUSION

It has been shown that cross validation with the check loss in quantile regression is reasonable, but that the technique can be improved by use of the modified check loss function we proposed. The suggested method has demonstrated its superiority in estimating future error through diverse simulations. The advantage is mainly due to the bias-variance tradeoff in estimating future error. This is reflected in the model selection,

Fig. 2. 'Best' cases of the fitted models selected by CV.M when compared with CV under a normal error distribution, with n=500.

Fig. 3. 'Worst' cases of the fitted models selected by CV.M when compared with CV under a normal error distribution, with n=500.

and a range of experiments reveals that the check loss may suffer from over fitting the data, while the modified check loss reduces over fitting thus leading to better model selection.

The same idea can be applied to the other model selection criteria such as BIC, and SIC [14]. As [7] replace the loss function for lack of fit in SIC with check loss, our proposed check function could be also replaced.

Fig. 4. Distribution of $\hat{\lambda}$ from CV and CV.M after subtracting $\hat{\lambda}$ from the base model (in the log scale) under a normal error distribution, with n=500. Only those $\hat{\lambda}$ pairs with different values are included.

TABLE VI
MSE VALUES (MULTIPLIED BY 1000) ARE DECOMPOSED TO VARIANCE (var) AND SQUARED BIAS ($bias^2$), BASED ON 1000 REPLICATES WITH n=500, AT SELECTED QUANTILES. STANDARD NORMAL ERROR IS CONSIDERED.

Method	CV	CV.M	CV	CV.M	CV	CV.M
	var	var	$bias^2$	$bias^2$	MSE	MSE
q=0.1	2.080	1.949	0.136	0.142	2.216	2.090
q=0.2	1.556	1.427	0.078	0.086	1.633	1.513
q=0.3	1.297	1.173	0.079	0.080	1.375	1.254
q=0.4	1.229	1.108	0.062	0.066	1.290	1.173
q=0.5	1.205	1.032	0.065	0.076	1.270	1.108
q=0.6	1.218	1.093	0.069	0.073	1.287	1.166
q=0.7	1.292	1.133	0.077	0.085	1.368	1.218
q=0.8	1.497	1.350	0.090	0.101	1.588	1.451
q=0.9	2.028	1.906	0.125	0.126	2.152	2.032

REFERENCES

[1] Bloomfield, P., Steiger, W.: Least Ablolute Deviations: Theory, Applications, and Algorithms. Boston, Birkhäuser (1983)

[2] Cole, T.: Fitting smoothed centile curves to reference data. Journal of the Royal Statistical Society, Series A (Statistics in Society) 151(3), 385–418 (1988)

[3] Efron, B.: Estimating the error rate of a prediction rule: Improvement on cross-validation. Journal of the American Statistical Association 78(382), 316–331 (1983)

[4] Efron, B., Tibshirani, R.: Improvements on cross-validation: The .632+ bootstrap method. Journal of the American Statistical Association 92(438), 548–560 (1997)

[5] Geisser, S.: The predictive sample reuse method with applications. Journal of the American Statistical Association 70(350), 320–328 (1975)

[6] Jung, Y., MacEachern, S.N., Lee, Y.: Window width selection for ℓ_2 adjusted quantile regression. Tech. rep., Department of Statistics, The Ohio State University (2009)

[7] Koenker, R.: Additive models for quantile regression: model selection and confidence bandaids. CeMMAP working papers CWP33/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies (Nov 2010), http://ideas.repec.org/p/ifs/cemmap/33-10.html
[8] Koenker, R., Ng, P., Portnoy, S.: Quantile smoothing splines. Biometrika 81(4), 673–680 (1994)
[9] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. pp. 1137–1143. Morgan Kaufmann (1995)
[10] Lee, Y., MacEachern, S.N., Jung, Y.: Regularization of case-specific parameters for robustness and efficiency. Statistical Science 27(3), 350 – 372 (2012)
[11] Leung, D.H.Y.: Cross-validatoin in nonparametric regression with outliers. The Annals of Statistics 33(5), 2291–2310 (2005)
[12] Nychka, D., Gray, G., Haaland, P., Martin, D., O'Connell, M.: A nonparametric regression approach to syringe grading for quality improvement. Journal of the American Statistical Association 90(432), 1171–1178 (Dec 1995)
[13] Ronchetti, E., Field, C., Blanchard, W.: Robust linear model selection by cross-validation. Journal of the American Statistical Association 92(439), 1017–1023 (1997)
[14] Schwarz, G.: Estimating the dimension of a model. The Annals of Statistics 6, 461–464 (1978)
[15] Shao, J.: Linear model selection by cross-validation. Journal of the American Statistical Association 88(422), 486–494 (1993)
[16] Shao, J.: An asymptotic theory for linear model selection. Statistica Sinica 7, 221–264 (1997)
[17] Stone, M.: Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B 36(2), 111–147 (1974)
[18] Stone, M.: An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. Journal of the Royal Statistical Society, Series B 39(1), 44–47 (1977)
[19] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58(1), 267–288 (1996)
[20] Wahba, G., Wold, S.: A completely automatic french curve: fitting spline functions by cross validation. Communications in Statistics 4, 1–17 (1975)
[21] Yu, K., Jones, M.: Local linear quantile regression. Journal of the American Statistical Association 93(441), 228–237 (1998)