Supplementary Materials for Efficient Quantile Regression for Heteroscedastic Models

Yoonsuh Jung

Department of Statistics, University of Waikato, Hamilton 3240, New Zealand

yoonsuh@waikato.ac.nz

Yoonkyung Lee and Steven N. MacEachern

Department of Statistics, The Ohio State University, Columbus, Ohio 43210

yklee@stat.osu.edu and snm@stat.osu.edu

September, 2014

The material in this file is also available as

Jung, Y., MacEachern, S.N., and Lee, Y. (2010). Window width selection for ℓ_2-adjusted quantile regression, *Technical Report 835*, Department of Statistics, The Ohio State University.

Abstract

Quantile regression provides estimates of a range of conditional quantiles. This stands in contrast to traditional regression techniques, which focus on a single conditional mean function. Quantile regression in the finite sample setting can be made more efficient and robust by rounding the sharp corner of the loss. The main modification generally involves an asymmetric ℓ_2 adjustment of the loss function around zero. The resulting modified loss has qualitatively the same shape as Huber's loss when estimating a conditional median. To achieve consistency in the large sample case, the range of ℓ_2 adjustment is controlled by a sequence which decays to zero as the sample size increases. Through extensive simulations, a rule is established to decide the range of modification. The simulation studies reveal excellent finite sample performance of modified regression quantiles guided by the rule.

KEYWORDS: Case indicator; check loss function; penalization method; quantile regression

1 Introduction

Quantile regression has emerged as a useful tool for providing estimates of conditional quantiles of a response variable Y given values of a predictor X. It allows us to estimate not only the center but also the upper and lower tails of the conditional distribution of interest. Due to its ability to capture full distributional aspects, rather than only the conditional mean, quantile regression has been widely applied. Koenker & Bassett (1978) and Bassett & Koenker (1978) consolidate a foundation for quantile regression. This foundation is extended to non-*iid* residuals in the linear model by He (1997) and Koenker & Zhao (1994). The loss function that defines quantile regression is called the check loss. The check loss has an asymmetric v-shape and becomes symmetric for the median. Lee, MacEachern & Jung (2007) introduced a new version of quantile regression where the check loss function is adjusted by an asymmetric ℓ_2 penalty to produce a more efficient quantile estimator. Initially, the modification of the loss function arises from including case-specific parameters in the model.

An additional penalty for the case specific parameters creates an adjustment of the check loss function over an interval. See Lee et al. (2007) for more details. The purpose of this paper is to provide a rule for determining the length of the interval of adjustment in the check loss function. To obtain a consistent estimator, the modification must vanish as the sample size grows. A brief theoretical review of ℓ_2 adjusted quantile regression is given in Section 2. In Section 3, extensive simulations are performed to develop a rule which will provide guidance on implementation of the modified procedure. The performance of the rule is demonstrated in Section 4 through simulation and real data. Discussion and potential extensions appear in Section 5.

2 Overview of ℓ_2 Adjusted Quantile Regression

To estimate the qth regression quantile, the check loss function ρ_q is employed:

$$\rho_q(r) = \begin{cases} qr & \text{for } r \geq 0 \\ -(1-q)r & \text{for } r < 0. \end{cases} \tag{1}$$

We first consider a linear model of the form $y_i = x_i^\top \beta + \epsilon_i$, where the ϵ_i's are *iid* from some distribution with qth quantile equal to zero. The quantile regression estimator $\hat{\beta}$ is the minimizer of

$$L(\beta) = \sum_{i=1}^{n} \rho_q(y_i - x_i^\top \beta). \tag{2}$$

To treat the observations in a systematic fashion, Lee et al. (2007) introduce case-specific parameters γ_i which change the linear model to $y_i = x_i^\top \beta + \gamma_i + \epsilon_i$. From the fact that this is a super-saturated model, $\gamma = (\gamma_1, ..., \gamma_n)^\top$ should be penalized. Together with the case-specific parameters and an additional penalty for γ, the objective function to minimize given in (2) is modified to be

$$L(\beta, \gamma) = \sum_{i=1}^{n} \rho_q(y_i - x_i^\top \beta - \gamma_i) + \frac{\lambda_\gamma}{2} J(\gamma), \tag{3}$$

where $J(\gamma)$ is the penalty for γ and λ_γ is a penalty parameter. Since the check loss function is piecewise linear, the quantile regression estimator is inherently robust. For improving efficiency, an ℓ_2 type penalty for the γ is considered. As detailed in Lee et al. (2007), desired invariance suggests an asymmetric ℓ_2 penalty of the form $J(\gamma_i) := \{q/(1-q)\}\gamma_{i+}^2 + \{(1-q)/q\}\gamma_{i-}^2$. With the $J(\gamma_i)$, let us examine the minimizing values of the γ_i, given β. First, note that $\min_\gamma L(\hat{\beta}, \gamma)$ decouples to minimization over individual γ_i. Hence, given $\hat{\beta}$ and a residual $r_i = y_i - x_i^\top \hat{\beta}$, $\hat{\gamma}_i$ is now defined to be

$$arg \min_{\gamma_i} \mathcal{L}_{\lambda_\gamma}(\hat{\beta}, \gamma_i) := \rho_q(r_i - \gamma_i) + \frac{\lambda_\gamma}{2} J(\gamma_i), \tag{4}$$

and is explicitly given by

$$-\frac{q}{\lambda_\gamma} I\big(r_i < -\frac{q}{\lambda_\gamma}\big) + r_i I\big(-\frac{q}{\lambda_\gamma} \leq r_i < \frac{1-q}{\lambda_\gamma}\big) + \frac{1-q}{\lambda_\gamma} I\big(r_i \geq \frac{1-q}{\lambda_\gamma}\big).$$

Plugging $\hat{\gamma}$ in (4) produces the ℓ_2 adjusted check loss,

$$\rho_q^\gamma(r) = \begin{cases} (q-1)r - \frac{q(1-q)}{2\lambda_\gamma} & \text{for} \quad r < -\frac{q}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}\frac{1-q}{q} r^2 & \text{for} \quad -\frac{q}{\lambda_\gamma} \leq r < 0 \\ \frac{\lambda_\gamma}{2}\frac{q}{1-q} r^2 & \text{for} \quad 0 \leq r < \frac{1-q}{\lambda_\gamma} \\ qr - \frac{q(1-q)}{2\lambda_\gamma} & \text{for} \quad r \geq \frac{1-q}{\lambda_\gamma}. \end{cases} \tag{5}$$

In other words, ℓ_2 adjusted quantile regression finds β that minimizes $L_{\lambda_\gamma}(\beta) = \sum_{i=1}^n \rho_q^\gamma(y_i - x_i^\top \beta)$. Note that the modified check loss is continuous and differentiable everywhere. The interval of quadratic adjustment is $(-q/\lambda_\gamma, (1-q)/\lambda_\gamma)$, and we refer to the length of this interval $1/\lambda_\gamma$ as the "window width". When the λ_γ is properly chosen, the modified procedure will enjoy its advantage to the full. The next section addresses how to set a good rule for selection of λ_γ.

3 Simulation Study

To develop a rule and obtain a consistent estimator, we first consider λ_γ of the form $\lambda_\gamma := c_q n^\alpha / \hat{\sigma}$ where c_q is a constant depending on q, n is the sample size, α is a positive constant, and $\hat{\sigma}$ is a robust scale estimate of the error distribution. Theorem 2 in Lee et al. (2007) suggests that for $\alpha > 1/3$, the modified quantile regression is asymptotically equivalent to the standard quantile regression. However, for optimal finite sample performance, we will consider a range of α values. We use 1.4826·MAD (Median Absolute Deviation) as a robust scale estimator $\hat{\sigma}$. The form of the rule suggests that c_q should be scale invariant and depend only on the targeted quantile q.

In this section, choice of the window width will be investigated by simulation. Throughout the simulation, the linear model $y_i = \beta_0 + x_i^\top \beta + \epsilon_i$ is assumed. Following the simulation setting in Tibshirani (1996), $x^\top = (x_1, ..., x_8)$ is generated from a multivariate normal distribution with mean (0, ..., 0) and variance Σ, where $\sigma_{ij} = \rho^{|i-j|}$ with $\rho = 0.5$. The true coefficient vector β is taken to be (3, 1.5, 0, 0, 2, 0, 0, 0). Various distributions are considered for ϵ_i, including normal, t, shifted log-normal, shifted gamma, and shifted exponential error distribution. In each distribution, ϵ_i is assumed to be *iid* with median zero and variance 9 (except when the ϵ_i follows the standard normal distribution). For the t distributions, 2.25, 5, and 10 degrees of freedom are used, maintaining a variance of 9.

Several values of α were tried. After examining the results, a decision was made to set α equal to 0.3. This makes α to be independent of sample size. Thus we search only for c_q. Sample sizes range from 10^2 to 10^4, and various quantiles from 0.1 to 0.9 are considered. To gauge the performance of ℓ_2 adjusted quantile regression with λ_γ, define mean squared error (MSE) of the estimated quantile $X^\top \hat{\beta} + \hat{\beta}_0$ at a new X as

$$\begin{aligned} MSE &= E^{\hat{\beta},X} ||(X^\top \hat{\beta} + \hat{\beta}_0) - (X^\top \beta + \beta_0)||^2 \\ &= E^{\hat{\beta},X} \{(\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) + (\hat{\beta}_0 - \beta_0)^2\} \\ &= E^{\hat{\beta}} \{(\hat{\beta} - \beta)^\top \Sigma (\hat{\beta} - \beta) + (\hat{\beta}_0 - \beta_0)^2\}. \end{aligned} \tag{6}$$

MSE is integrated across the distribution of a future X. The distribution of the future X is normal with mean (0, ..., 0) and variance Σ. In the simulation, MSE is approximated by a Monte Carlo estimate over 500 replicates, $\widehat{MSE} = 500^{-1} \sum_{i=1}^{500} ((\hat{\beta}^i - \beta)^\top \Sigma (\hat{\beta}^i - \beta) + (\hat{\beta}_0^i - \beta_0)^2)$, where $\hat{\beta}^i$ and $\hat{\beta}_0^i$ are the estimates of β and the intercept β_0 for the i^{th} replicate, respectively. With fixed α, the window width $(\hat{\sigma}/(c_q n^\alpha))$ is a function of the constant c_q only. Thus by varying c_q, an 'optimal' window width which provides the smallest MSE can be obtained. The optimal window widths, found by a grid search, are shown in Figure 1 for various error distributions.

Each panel of Figure 2 shows a typical shape of the MSE curve as a function of window width. In general, MSE values begin to decrease as we increase the window width from zero until it hits its minimum, and increase thereafter due to an increase in bias. However, when estimating the median with normally distributed errors, MSE decreases as the window width increases. This is not surprising, given the optimality properties of least squares regression for normal theory regression. The comparisons between sample mean and sample median can be explicitly found under the t error distributions using different degrees of freedom. The benefit of the median relative to the mean is greater for thicker tailed distributions. We observe that this qualitative behavior carries over to the optimal window width. Thicker tails lead to shorter optimal windows, as shown in Figure 1.

3.1 Development of a Rule

Under each error distribution mentioned above, the 'optimal' constants which yield smallest $\widehat{MSE}$ are found at the quantiles 0.1, 0.2, ..., 0.9. First, omitting the median, log of the optimal constant $\log(c_q)$ from the standard normal error is regressed on q to suggest a possible relationship. A significant linear relationship exists. The fitted values from this regression were used to produce values for c_q. These values were then applied to the other error distributions. However, the rule obtained from the normal distribution led to poor $\widehat{MSE}$ values when applied to skewed error distributions. This is due to the overestimation of the window width or equivalently, underestimation of c_q near the median. As we can see in Figure 2, too large a window may lead to a huge MSE.

As an alternative, another rule expressing the relationship between the optimal $\log(c_q)$ and q was developed from the exponential error distribution. The top left plot in Figure 3 shows the relationship between optimal $log(c_q)$ and q. Before fitting a linear model of $log(c_q) = \beta_0 + \beta_1 q + \epsilon$, q greater than 0.5 were converted to $1 - q$, since it was judged desirable to have a rule which will work well for symmetric distributions. The solid line in the top right plot of Figure 3 is the fitted line using all observations, whereas the dashed line is from only observations with $q \geq 0.5$, excluding observations with + mark. The dashed line is accepted as a final rule.

The final rule is compared to the other rules from normal, t, log-normal, and gamma distributions. In Figure 3, the solid lines in the second and third rows represent 'optimal' rules from each distribution mentioned above (developed on quantiles ≥ 0.5) whereas the dashed line is the final rule. Numerical expression of the final rule is given by

$$c_q \approx \begin{cases} 0.5e^{-2.118-1.097q} & \text{for } q < 0.5 \\ 0.5e^{-2.118-1.097(1-q)} & \text{for } q \geq 0.5, \end{cases} \tag{7}$$

where q stands for the qth quantile.

Under various error distributions, the estimated c_q from the rule (7) is employed to gauge its prediction performance. Specifically, $\widehat{MSE}$ values for quantile regression (QR), modified

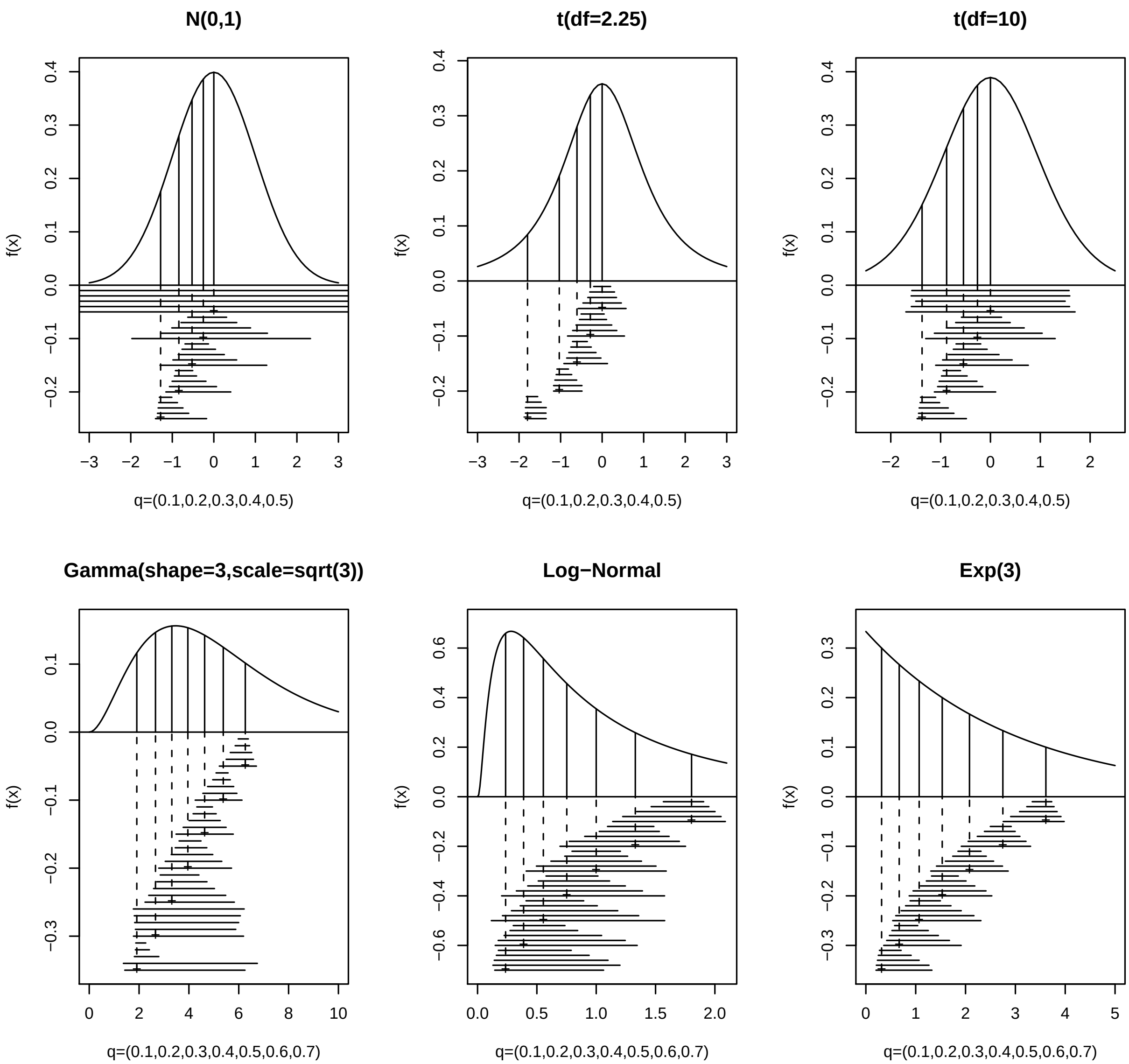

Figure 1: 'Optimal' intervals of adjustment for different quantiles (q), sample sizes (n) and error distributions. The vertical lines in each distribution indicate the true quantiles. The stacked horizontal lines for each quantile are corresponding optimal intervals. The five intervals at each quantile are for $n = 10^2$, $10^{2.5}$, 10^3, $10^{3.5}$ and 10^4.

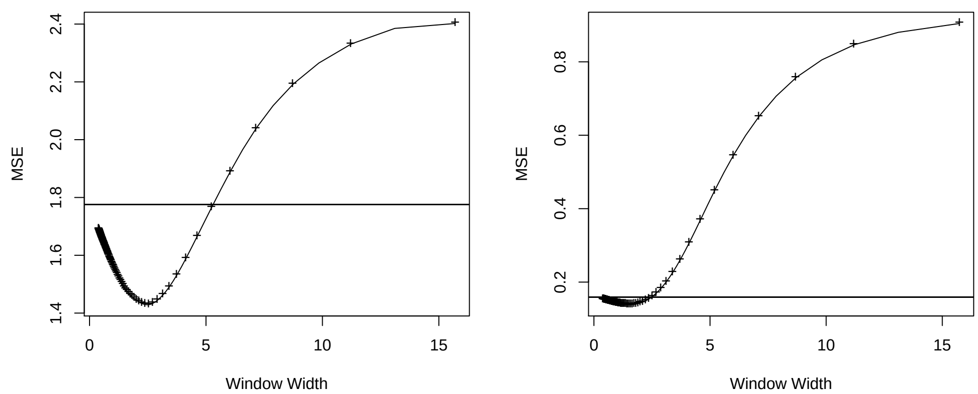

Figure 2: $\widehat{MSE}$ values evaluated at one hundred points marked with '+' and connected by a smoothing spline. The smallest and largest window widths in each plot correspond to the window width approximately 5% and 98% of data in it, respectively. The residual distribution is the t (df=10) distribution, sample sizes are 10^2 (left panel) and 10^3 (right panel), and the 0.2 quantile is estimated. The horizontal lines represent the $\widehat{MSE}$ values from the standard quantile regression.

quantile regression with optimal c_q (OPT), and modified quantile regression with c_q chosen by the final rule (QR.M) are compared. Figures 6 through 11 show the behavior of QR, OPT, and QR.M in terms of $\widehat{MSE}$. Overall, QR.M handily outperforms standard quantile regression. Surprisingly enough, the version of finite sample performance for this modified quantile regression is often nearly optimal. This near-optimality extends across a range of residual distributions.

In practice, the robust linear modeling procedure, `rlm(MASS)` in R package is ready to be utilized. Equipped with the derivative of (5), the modified estimators can be obtained from the `rlm` function by specifying q and the corresponding rule c_q. Since the `rlm` function internally uses re-scaled MAD for the method of scale estimation, the estimate of the scale parameter in λ_γ is automatically obtained.

4 Application to Engel's Data

Engel's data consists of the household food expenditure and household income from 235 European working-class households in the 19th century. Taking the log of food expenditure as a response variable, we investigate the relation between log of food expenditure and log of household income. In Figure 4, Engel's data is plotted after transformation of both variables. Superimposed on the scatter plot are the fitted lines from quantile regression (QR), and modified quantile regression (QR.M) using the rule developed in Section 3. Although the two methods display quite similar fitted lines, Figure 5 reveals the difference between QR

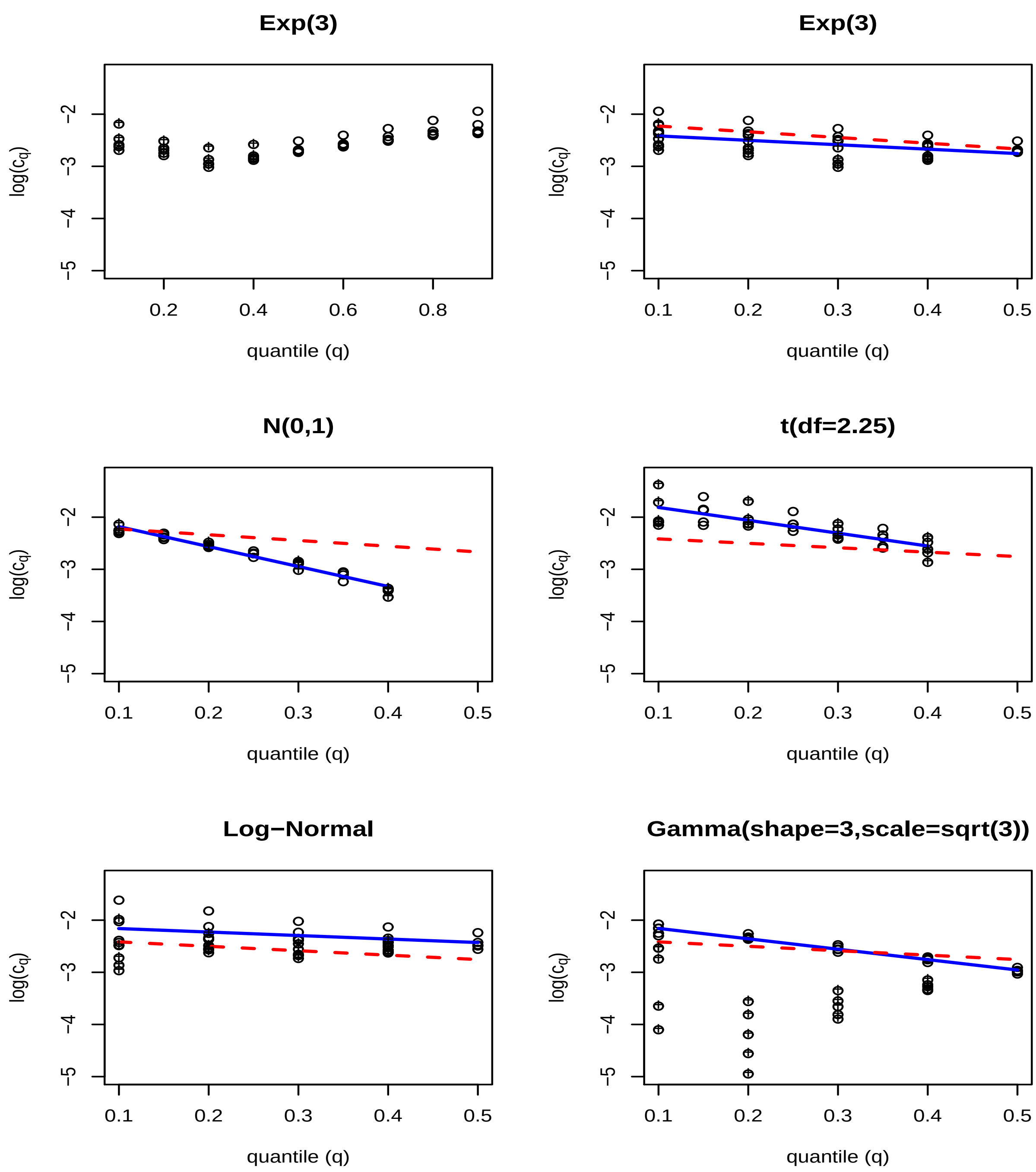

Figure 3: Top left: Relationship between optimal $\log(c_q)$ and quantile from the exponential distribution. Top right: Left plot is folded in half at $q = 0.5$. Circles with a + mark are from the left fold (quantile < 0.5) and the others are from the right fold (quantiles ≥ 0.5). The solid line is the fitted line using all observations whereas the dashed line excludes observations with a + mark (final rule). Solid lines in the middle and bottom plots are the rules corresponding to normal, t, log-normal, and gamma distributions compared to the final rule (dashed line).

and QR.M. We note that these fitted lines from modified quantile regression do not across over the range of log(Household income) in the data. This is partly due to the averaging effect of the ℓ_2 adjustment to the check loss function.

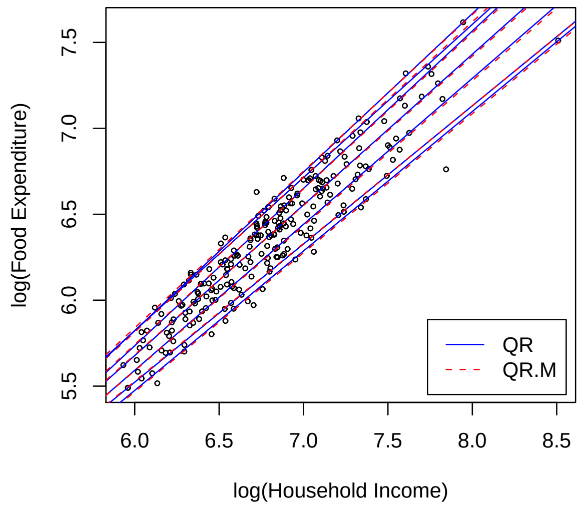

Figure 4: Superimposed on the scatter plot are the 0.05, 0.1, 0.25, 0.5, 0.75, 0.90, 0.95 standard quantile regression (solid, blue) lines, and modified quantile regression (dashed, red) lines for Engel's data after log transformation of both response and predictor variables.

5 Conclusion

We have shown how case-specific indicators can be utilized in the context of quantile regression through regularization of their parameters. The simulation studies suggest a simple rule to select the regularization parameter for the case-specific parameters. The behavior of the newly developed rule is excellent under both symmetric and asymmetric error distributions at any conditional quantile, regardless of the sample size. The analysis of Engel's data also reveals that the modified procedure is less prone to crossing estimates of quantiles than is quantile regression (this is confirmed in further investigation not presented here). For large sample behavior, details of theoretical results and conditions regarding consistency properties are given in Lee et al. (2007). In terms of computation, modified quantile regression requires only slight adjustment to existing software. The simulated and real data analyses have shown the potential of ℓ_2 adjusted quantile regression and the rule for selecting the window width. Finally, we wish to point out a possible direction where our research can be extended. As Koenker & Zhao (1994) and Koenker (2005) considered heteroscedastic models in quantile regression, the scope of our modified quantile regression procedure can

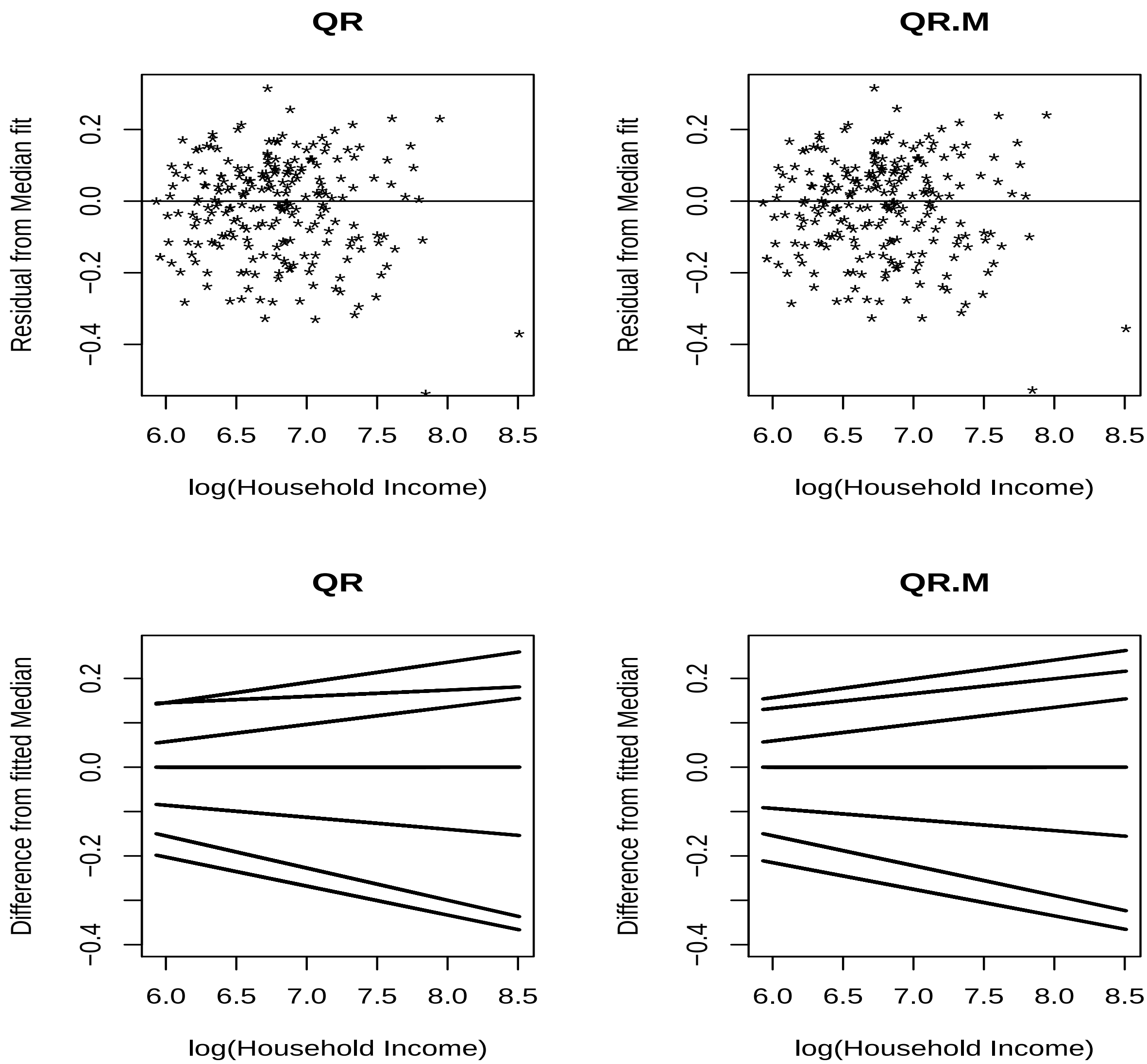

Figure 5: Top: Residuals from a median fit via QR and QR.M. Bottom: Differences between fitted median line and the fitted quantiles at q=0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95.

be expanded to include non-*iid* error models.

References

Bassett, G. & Koenker, R. (1978). Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association* **73**(363): 618–622.

He, X. (1997). Quantile curves without crossing, *The American Statistician* **51**(2): 186–192.

Koenker, R. (2005). *Quantile Regression*, Cambridge U. Press.

Koenker, R. & Bassett, G. (1978). Regression quantiles, *Econometrica* **46**(1): 33–50.

Koenker, R. & Zhao, Q. (1994). L-estimation for linear heteroscedastic models, *Journal of Nonparametric Statistics* **3**(3): 223–235.

Lee, Y., MacEachern, S. N. & Jung, Y. (2007). Regularization of case-specific parameters for robustness and efficiency, *Technical Report No.799, The Ohio State University* .

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B* **58**(1): 267–288.

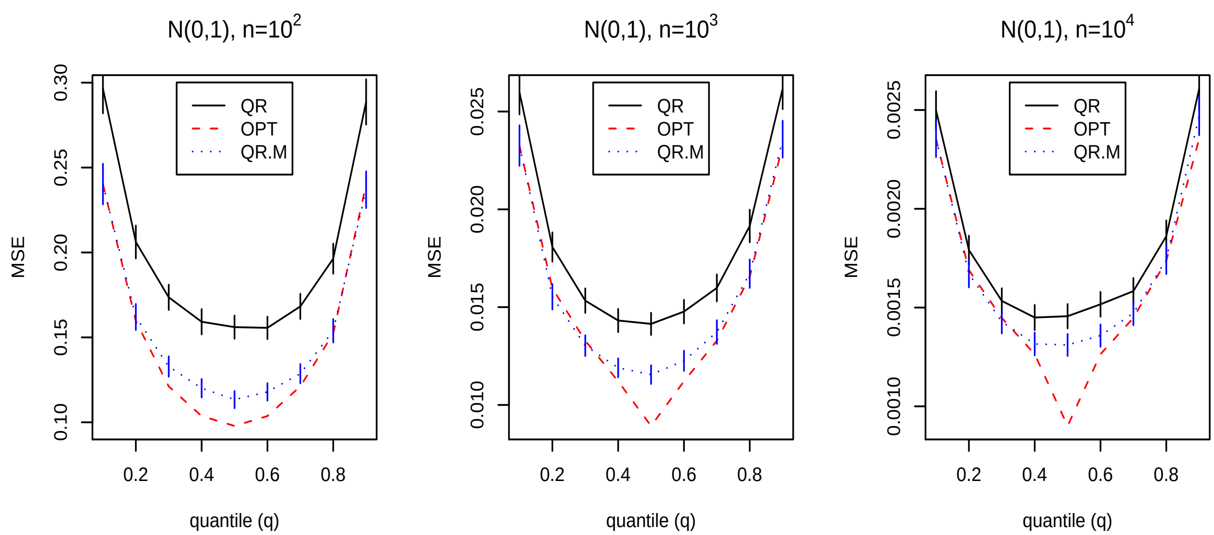

Figure 6: $\widehat{MSE}$ values from quantile regression (QR), modified quantile regression with optimal window width (OPT), and modified quantile regression using the rule (QR.M) under a standard normal error distribution.

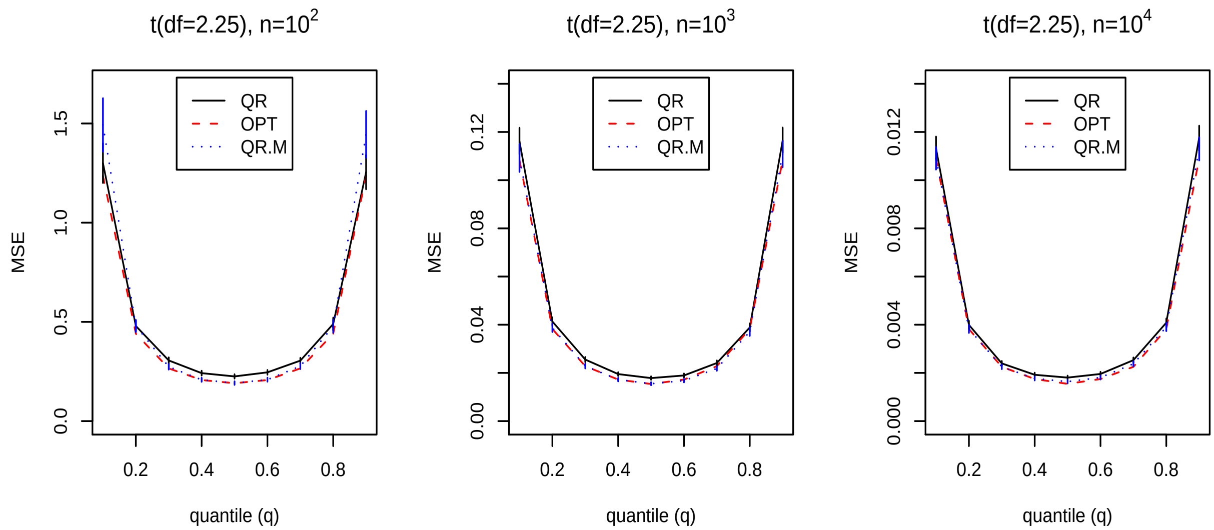

Figure 7: $\widehat{MSE}$ values from quantile regression (QR), modified quantile regression with optimal window width (OPT), and modified quantile regression using the rule (QR.M) under t (df=2.25) error distribution.

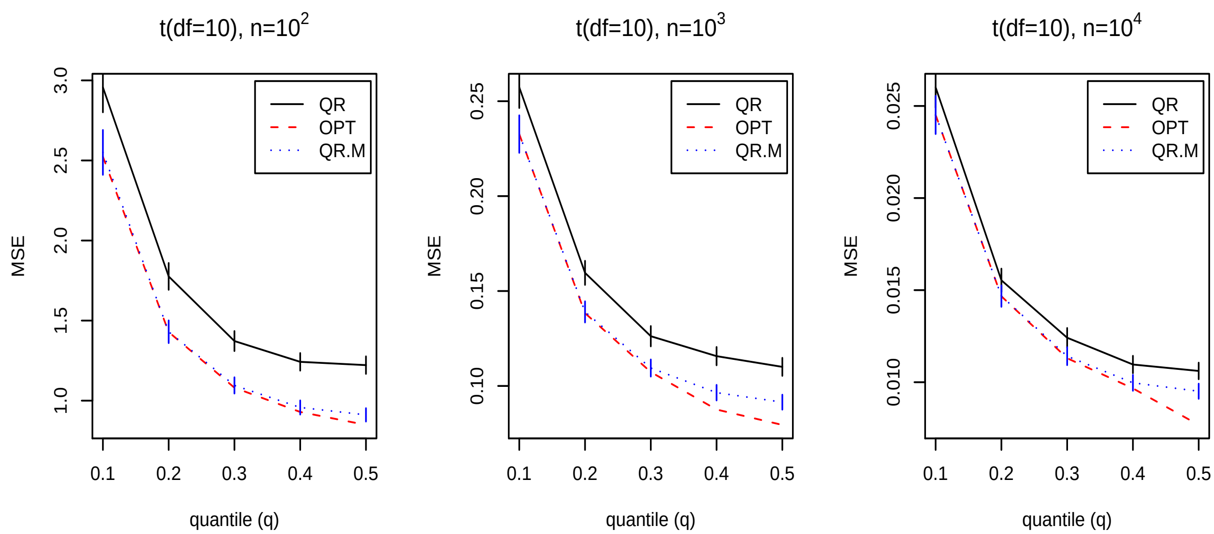

Figure 8: $\widehat{MSE}$ values from quantile regression (QR), modified quantile regression with optimal window width (OPT), and modified quantile regression using the rule (QR.M) under a t (df=10) error distribution.

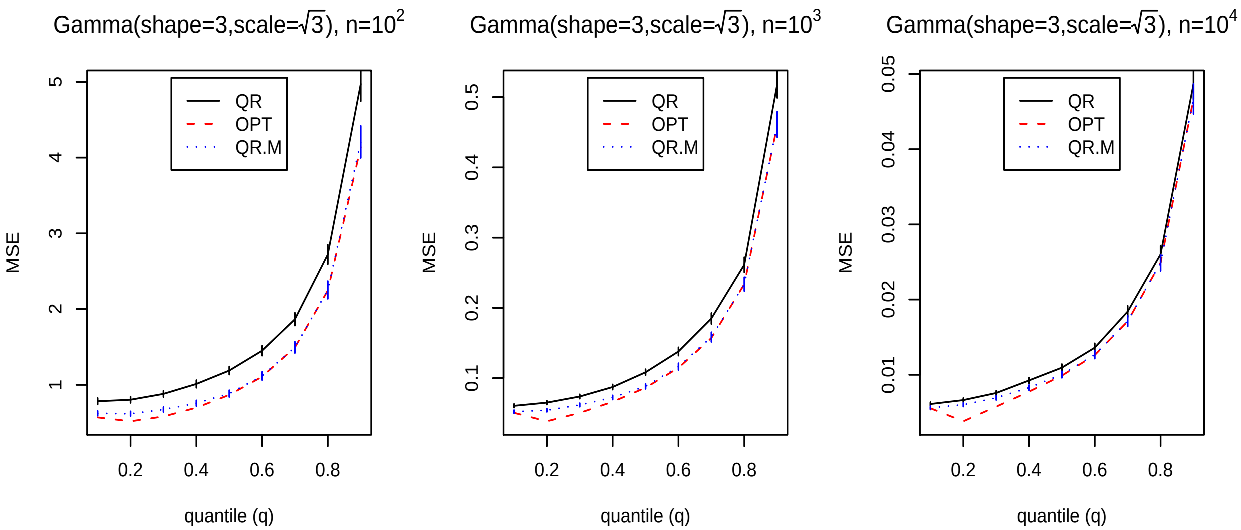

Figure 9: $\widehat{MSE}$ values from quantile regression (QR), modified quantile regression with optimal window width (OPT), and modified quantile regression using the rule (QR.M) under a gamma $(3,\sqrt{3})$ error distribution.

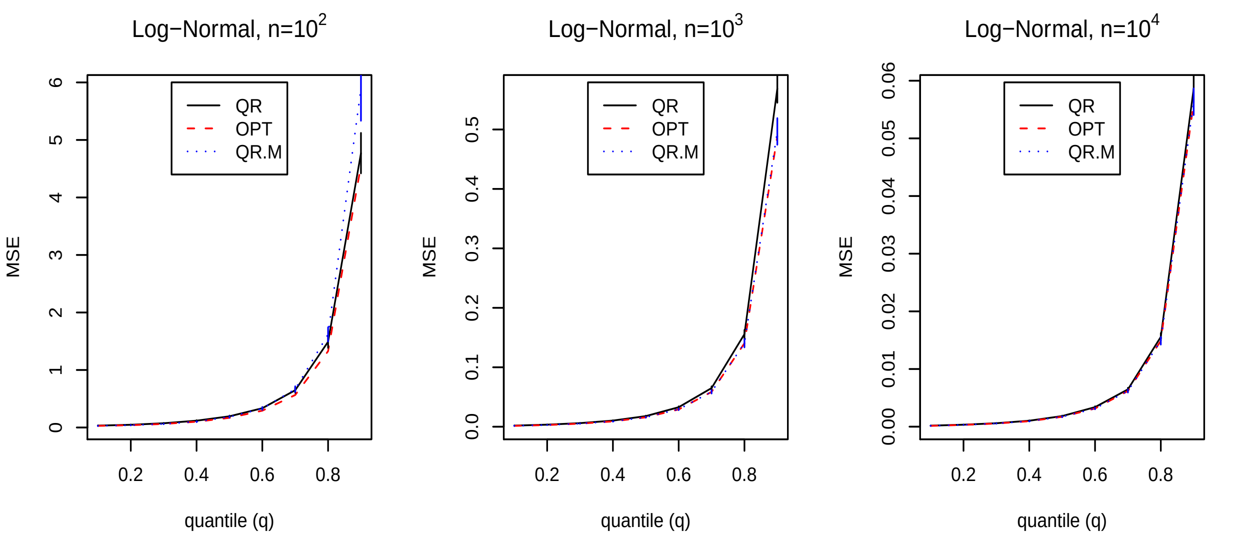

Figure 10: $\widehat{MSE}$ values from quantile regression (QR), modified quantile regression with optimal window width (OPT), and modified quantile regression using the rule (QR.M) under a log-normal error distribution.

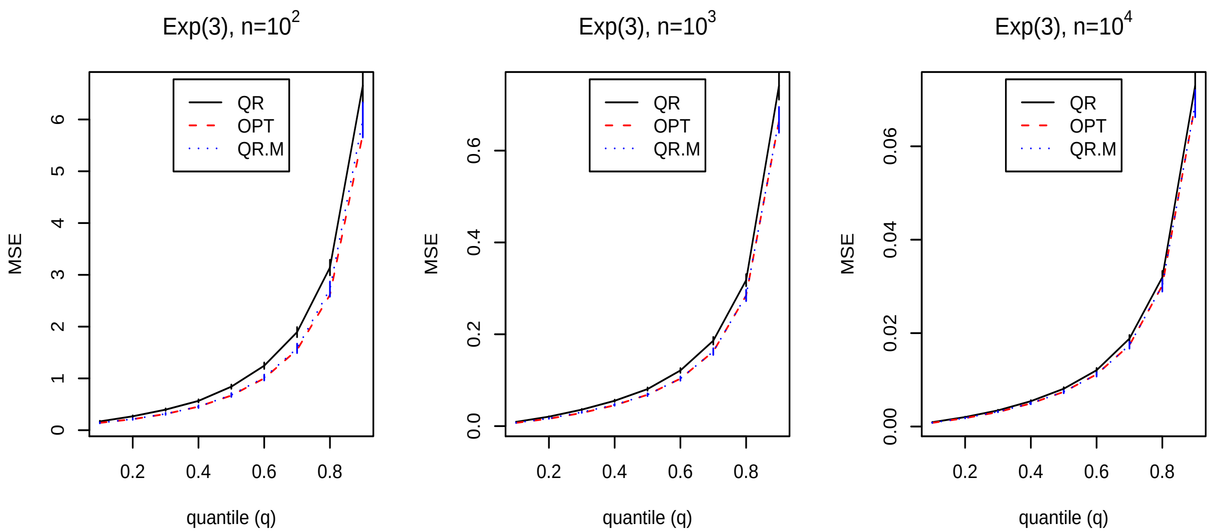

Figure 11: $\widehat{MSE}$ values from quantile regression (QR), modified quantile regression with optimal window width (OPT), and modified quantile regression using the rule (QR.M) under an exponential (3) error distribution.