Some Variable Selection and Regularization Methodological Approaches in Quantile Regression with Applications

by

Innocent Mudhombo

A thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy
in the subject
Statistics
at the
UNIVERSITY OF SOUTH AFRICA
SUPERVISOR: Prof. E. Ranganai
Year: 2022

Abstract

The importance of robust variable selection and regularization as solutions to the collinearity influential high leverage points' adverse effects in a quantile regression (QR) setting cannot be overemphasized, just as the diagnostic tools that identify these high leverage points. In the literature, researchers have dealt with variable selection and regularization quite extensively for penalized *QR* that generalizes the well-known least absolute deviation (*LAD*) procedure to all quantile levels. Unlike the least squares (*LS*) procedures, which are unreliable when deviations from the Gaussian assumptions (outliers) exist, the *QR* procedure is robust to Y-space outliers. Although *QR* is robust to response variable outliers, it is vulnerable to predictor space data aberrations (high leverage points and collinearity adverse effects), which may alter the eigen-structure of the predictor matrix. Therefore, in the literature, it is recommended that the problems of collinearity and high leverage points be dealt with simultaneously. In this thesis, we propose applying the ridge regression procedure ($RIDGE$), *LASSO*, elastic net (E-NET), adaptive *LASSO*, and adaptive elastic net (AE-NET) penalties to weighted *QR* (WQR) to mitigate the effects of collinearity and collinearity influential points in the *QR* setting. The new procedures are the penalized WQR procedures i.e., the *RIDGE* penalized WQR (WQR-$RIDGE$), the *LASSO* penalized WQR (WQR-*LASSO*), the E-NET penalized WQR (WQR-E-NET) and the adaptive penalized *QR* procedures (the adaptive *LASSO* penalized *QR* (*QR-ALASSO*) and adaptive E-NET penalized *QR* (*QR*-AE-NET procedures and their weighted versions). The penalized WQR procedures are based on the computationally intensive high-breakdown minimum covariance determinant (MCD) determined weights and the adaptive penalized *QR* procedures are based on the *RIDGE* penalized WQR (WQR-*RIDGE*) estimator based adaptive weights. Under regularity conditions, the adaptive penalized procedures satisfy oracle properties. Although adaptive weights are commonly based on the *RIDGE* regression (RR) estimator in the *LS* setting when regressors are collinear, this estimator may be plausible for the symmetrical distributions at the ℓ_1-estimator (RQ at $\tau = 0.50$) rather than at extreme quantile levels. We carried out simulations and applications to well-known data sets from the literature to assess the finite sample performance of these procedures in variable selection and regularization due to the robust weighting formulation and adaptive weighting construction. In the collinearity-enhancing point scenario under the t-distribution, the WQR penalized versions outperformed the unweighted procedures with respect to average shrunken zero coefficients and correctly fitted models. Under the Gaussian and t-distributions, at predictor matrices with collinearity-reducing points, the weighted regularized procedures dominate the prediction performance (WQR-*LASSO* forms best). In the collinearity-inducing and collinearity-reducing points scenarios under the Gaussian distribution, the adaptive penalized procedures outperformed the non-adaptive versions in prediction. Under the t-distribution, a similar performance pattern is depicted as in the Gaussian scenario, although the performance of all models is adversely affected by outliers. Under the t-distribution, the *QR-ALASSO* and WQR-*ALASSO* procedures performed better in their respective categories.
Keywords:
weighted quantile regression; adaptive *LASSO* penalty; penalty; adaptive E-NET penalty; regularization; Penalization; collinearity inducing point; collinearity hiding point; collinearity influential points

DECLARATION

I declare that SOME VARIABLE SELECTION AND REGULARIZATION METHODOLOGICAL APPROACHES IN QUANTILE REGRESSION WITH APPLICATIONS is my own original work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

Signature:..........................

Innocent Mudhombo

Date: .24/01/2023.......................

ACKNOWLEDGEMENTS

I extend my sincere appreciation to the examiners. Their valuable insights, constructive criticisms, and rigorous examination of my work have contributed immensely to the refinement and quality of this thesis.

My heartfelt thanks to my esteemed supervisor, Prof. Edmore Ranganai. His extensive knowledge, insightful feedback, and unwavering commitment to excellence have been instrumental in shaping this thesis. His mentorship and encouragement have inspired me on this journey.

Furthermore, I am indebted to the Department of Statistics staff for their continuous support. The scholarly environment enriched my understanding and broadened my horizons. I also appreciate Vaal University of Technology for their financial and moral support.

My deepest gratitude to my family and friends for their unwavering support and constant encouragement throughout this journey. A special mention goes to my wife, Rudo Mudhombo, for her unwavering support.

Thank you all for being an integral part of this academic journey and for your unwavering commitment to academic excellence.

Publications

The list of article publications from the thesis is as follows:

1. Edmore Ranganai and Innocent Mudhombo. Variable selection and regularization in quantile regression via minimum covariance determinant based weights. Entropy, 23(1):33, 2021.

2. Innocent Mudhombo and Edmore Ranganai. Robust Variable Selection and Regularization in Quantile Regression Based on Adaptive-*LASSO* and Adaptive *E-NET*. Computation, 10(11):203, 2022.

3. Unpublished article on adaptive weights.

Acronyms and Abbreviations

ABBREVIATION	DESCRIPTION	ABBREVIATION	DESCRIPTION
BIC	Bayesian information criterion	ABE	augmented backward elimination
SVD	singular value decomposition	BE	backward elimination
AIC	Akaike information criterion	CIM	collinearity influential measure
CN	condition number	AR-LASSO	adaptive robust least absolute shrinkage and selection operator
CV	cross validation		
COD	coefficient of determination	CQR	composite quantile regression
COSSO	component selection and smoothing operator	CQR-LASSO	composite quantile regression least absolute shrinkage and selection operator
CVIF	corrected variance inflation factor	DCA	difference convex algorithm
		DRGP	diagnostic robust generalized potentials
E-NET	elastic net	DVIF	difference based variance inflation factor
		EAOCM	eigensystem analysis of correlation matrix
ER	elemental regression	ES	elemental set
EVD	eigenvalue decomposition		
GLM	generalized linear model	GP	generalized potential
		GSIF	generalized standard error inflation factor
LAD	least absolute deviation	LAD-E-NET	least absolute deviation elastic net
LAD-RIDGE	least absolute deviation ridge	LAD-LASSO	least absolute deviation least absolute shrinkage and selection operator
LARS	least angle regression	LLA	local linear approximation algorithm
LASSO	least absolute shrinkage and selection operator	LAD-ALASSO	least absolute deviation adaptive least absolute shrinkage and selection operator
LS	least squares	LS-AE-NET	least squares adaptive elastic net
LS-ARIDGE	least squares adaptive ridge	LS-ALASSO	least squares adaptive least absolute shrinkage and selection operator
LS-E-NET	least squares elastic net	LS-RIDGE	least squares ridge
LS-LASSO	least squares least absolute shrinkage and selection operator	LS-post-LASSO	least squares post least absolute shrinkage and selection operator
MAD	median absolute deviation	MAE	mean absolute error
MBE	mean bias error	MCD	minimum covariance determinant
MD	Mahalanobis distance	MSE	mean squared error
MVE	minimum volume ellipsoid	PLSE	perturbed least squares estimator
QR	quantile regression	QR-AE-NET	adaptive elastic net penalized quantile regression

ABBREVIATION	DESCRIPTION	ABBREVIATION	DESCRIPTION
QR-E-NET	elastic net penalized quantile regression	QR-ALASSO	adaptive least absolute shrinkage and selection operator penalized quantile regression
QR-RIDGE	ridge penalized quantile regression	QR-ARIDGE	adaptive ridge penalized quantile regression
RD	robust distance	RPCA	robust principal components analysis
RMSE	root mean squared error	RQ	regression quantile
SCAD	smoothly clipped absolute deviation	QR-LASSO	least absolute shrinkage and selection operator quantile regression
SIF	standard error inflation factor		
SURE	seemingly unrelated regression equations	VD	variance decomposition
VIF	variance inflation factor	WCQR	weighted composite quantile regression
WLAD	weighted least absolute deviation	WLAD-E-NET	weighted least absolute deviation elastic net
WCQR-LASSO	weighted composite quantile regression least absolute shrinkage and selection operator	WLAD-LASSO	weighted least absolute deviation least absolute shrinkage and selection operator
WLAD-RIDGE	weighted least absolute deviation ridge	WLS-AE-NET	weighted least squares adaptive elastic net
WLS-ALASSO	weighted least squares adaptive least absolute shrinkage and selection operator	WLS-LASSO	weighted least squares least absolute shrinkage and selection operator
WLS-ARIDGE	weighted least squares adaptive ridge	WLS-E-NET	weighted least squares elastic net
WLS-RIDGE	weighted least squares ridge	WQR-AE-NET	weighted adaptive elastic net quantile regression
WQR-ALASSO	weighted adaptive least absolute shrinkage and selection operator quantile regression	WQR-LASSO	weighted least absolute shrinkage and selection operator quantile regression
WQR-ARIDGE	weighted adaptive ridge quantile regression	WQR-E-NET	weighted elastic net quantile regression
WQR-RIDGE	weighted ridge quantile regression	WQRR	weighted quantile ridge regression
WRR	weighted ridge regression	WVIF	weighted variance inflation factor
Tol	tolerance ratio	Fused LASSO	fused least absolute shrinkage and selection operator

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Variable selection and regularization procedures in multiple regression analysis have been dealt with extensively in the literature. However, predictor space data aberrations (X-space outliers and collinearity) and response (Y-space) outliers continue to pose almost insurmountable challenges to these procedures. As a consequent, variable selection and regularization are topical in recent years. In the literature, the least squares (LS) regression is susceptible to all these data aberrations, with some solutions proffered by alternative robust procedures. In the presence of Y-space outliers, robust procedures, such as quantile regression (QR) (Koenker & Basset 1978), have been suggested. QR has the advantage of providing more information about the conditional distribution of the response variable Y given the predictors X at each quantile level and models the conditional quantiles $(Q_{Y/X}(\tau))$ over the entire range of quantiles $\tau \in (0,1)$. Although QR and other least absolute deviation (LAD) based procedures are robust in the presence of Y-space outliers, they are susceptible to high leverage points and collinearity influential points (high leverage points which are collinearity inducing and/or reducing points). The weighted *LAD-LASSO* (*WLAD-LASSO*) (Arslan 2012) mitigates against high leverage point influences, while the adaptive *RIDGE* (*ARIDGE*) (Frommlet

& Nuel 2016), adaptive *E-NET* (*AE-NET*) (Zou & Zhang 2009), *etc*, mitigate against the effects of collinearity. In the *QR* scenario, weighted *QR* (*WQR*) (Salibián-Barrera & Wei 2008), adaptive *QR LASSO* (*QR-ALASSO*) (Wu & Liu 2009), *etc*, are fairly robust in the presence of high leverage (collinearity influential) points and collinearity, respectively. The adaptive and weighted *QR* procedures inherit good properties from their non-adaptive and unweighted counterparts, respectively (see Chapters 2 and 3 for detailed discussions), in addition to enhancing robustness.

Many variable selection and/or regularization procedures are suggested in the literature, including subset selection. According to Breiman (1995), subset selection procedures are unstable for variable selection, especially in high dimensional scenarios. Penalization procedures have been suggested as alternatives to proffer solutions to subset selection procedures' shortcomings in the *LS* and *QR* scenarios. Regularization techniques include the ridge regression (RIDGE) (Hoerl & Kennard 1970), the least absolute shrinkage and selection operator (*LASSO*) (Tibshirani 1996), elastic net (Zou & Hastie 2005) and the extended versions of these three procedures, amongst others.

The broken adaptive ridge (*BAR*) method has been suggested in the literature (see Dai et al. 2018, 2020). The reweighted ℓ_2-penalization based *BAR*, which estimates regression coefficient patterns and yields estimates that have oracle properties. An Oracle property is the ability of a method to select true non-zero coefficients and estimate their values accurately. Oracle property states that a regression estimator converges to the true underlying coefficient values with probability approaching one as sample size approaches infinity (see Wang et al. 2007). According to Dai et al. (2018), its asymptotic consistency has yet to be thoroughly investigated. Dicker et al. (2013) proposed the seamless-L_0 (*SELO*) penalty, which closely resembles the L_0 penalty. The *SELO* penalized *LS* approach, which performs better than other widely used penalized *LS* procedures, is asymptotically normal and always chooses the right model. By merging the non-concave penalized likelihood and the pseudo-score methods, Lin & Lv (2013) presented a regularization method that simultaneously performs variable selection and estimation. This procedure by Lin & Lv (2013) performs better in high-dimensional situations and is based on the smooth integration of counting and absolute deviation $(SICA)$ penalty.

The existence of outlying points poses a threat to the stability of parameter estimation and ultimately to the reliability of resultant estimation/prediction and inference. Outlying points comprise Y-space outliers (outliers) and X-space outliers (high leverage points). When high leverage points influence the collinearity structure of the design matrix, they are called collinearity influential points. High leverage points that mask or create collinearity are collinearity hiding or collinearity inducing, respectively. In this thesis, we answer three research questions: (i) How many variables do we select or how do we penalize them? (ii) How do we discard the unimportant variables? (iii) How do we mitigate against collinearity and high leverage points in them (see Kendall cited in Farrar & Glauber 1967)?

1.1 Penalization in Quantile Regression

Consider the linear regression model

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + e_i,\ i \in [1:n] \tag{1.1}$$

with intercept term $\beta_0 = 0$, where y_i is the i^{th} response observation, $\boldsymbol{x}_i'$ is the i^{th} row of the design matrix $\boldsymbol{X}$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters to be estimated from the data, and e_i is a random error term, with cumulative distribution function F $(e_i \sim F)$.

The *LS* limits its inquiry only to the conditional mean ($E(Y|X)$), hence the need for another method that proffers diversity. *QR* offers an alternative to the conditional mean among stochastic response variables and their predictors. Conditional *QR* involves minimizing the asymmetric version of absolute errors by reformulation of the optimization problem as a parametric linear program (see Koenker & Basset 1978). Violations of normality assumptions are very common in real-life data situations, such as thick-tailed distributions in the error terms. *QR* is one of many procedures that is robust to Y-space outliers, therefore suitable for thick tailed distributions scenarios, as well as asymmetric ones.

The Koenker & Basset (1978) *QR* is based on an optimization problem solved by linear programming techniques by assuming the error term follows a cumulative distribution ($\varepsilon_i \sim F$). We write the *QR* minimization problem as:

$$\hat{\boldsymbol{\beta}}(\tau) = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)|, \;\; i \in [1:n], \tag{1.2}$$

where

$$\rho_\tau(u) = \begin{cases} \tau.u, \; if \; u \geq 0 \\ (\tau - 1).u, \; if \; u < 0 \end{cases}$$

denotes the check function, $u = y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)$ denotes residuals at $\tau \in (0,1)$ *RQ* levels, and $\boldsymbol{\beta}(\tau)$ is the coefficient. If we let $Y_1, Y_2, ..., Y_n$ (Y_is are ordered) be *iid* with continuous and strictly increasing distribution function F, then $F^{-1}(\tau) = inf\{y|F(y) \geq \tau\}$. The conditional quantile function of Y given the covariate X, is then given by $Q_{Y|X}(\tau) = \beta_0 + F^{-1}(\tau) + \boldsymbol{x}'\boldsymbol{\beta}$, where $\boldsymbol{\beta}(\tau)$ is estimated by $\hat{\boldsymbol{\beta}}(\tau) = \begin{pmatrix} \hat{\beta}_0 + F^{-1}(\tau) \\ \hat{\boldsymbol{\beta}} \end{pmatrix}$ with $\hat{\beta}_0 + F^{-1}(\tau)$ corresponding to the intercept and $\hat{\boldsymbol{\beta}}$ corresponds to $(\beta_1, \beta_2, ..., \beta_p)'$.

Consider variable selection and regularization procedures in *QR*. The minimization problem gives the penalized *QR* problem

$$\hat{\boldsymbol{\beta}}(\tau) = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + \lambda \Sigma_{j=1}^{p} |\theta|, for\ i \in [1:n], j \in [1:p], \tag{1.3}$$

where λ is the tuning parameter, the second term is the penalty term. We have a *LASSO* penalty when $\theta = \beta_j$ and *RIDGE* penalty when $\theta = \beta_j^2$ (see Chapter 3 for more details).

In the regression problem, we face a dilemma of choosing the best subset of predictor variables. Some predictors might also be redundant due to data aberrations, such as high leverage points and collinearity, etc.

1.1.1 Motivation for Regularization

In the literature, the least squares (*LS*) methods are known to be sensitive to violations of the Gaussian assumptions, and data aberrations in the design space. Data aberrations in the predictor space (X-space outliers) are referred to as high leverage points and in the response space (Y-space outliers) are outliers. High leverage points can either induce or hide collinearity and are called collinearity influential points (see simulations Chapter 5 for more details). However, not all high leverage points are collinearity influential points.

The collinearity phenomenon occurs when at least two predictor variables are nearly dependent, thus, they contain almost the same information (redundancy). In the literature, collinearity is known to have adverse effects in multiple regression analysis. While sources of collinearity are numerous, it is well-known that some high leverage points alter the eigen-structure of the design matrix, thereby masking or inducing collinearity. These adverse effects include wrong signs of parameter estimates, erroneous interpretation of parameter estimates, and estimates with disproportionate large variances, amongst others (Hoerl & Kennard 1970).

We summarize some procedures and their limitations in the presence of these data aberrations:

- Most Statisticians prefer the *LS* estimator because of its good prediction performance and interpretable results under Gaussian assumptions. However, the following challenges have been identified: (i) collinearity in predictor variables is difficult to detect due to the masking effect of collinearity influential points and (ii) *LS* estimator works well when the number of observations is greater than the number of predictors ($n > p$). When $n < p$, the *LS* estimator fails.

- Non-penalized *QR* is generally inconsistent in high-dimensional data scenarios, especially when $p \geq n$, which is addressed by introducing penalties to the *QR* procedure (Belloni & Chernozhukov 2011). Jiang et al. (2012) extended the Zou & Yuan (2008) composite quantile regression (CQR), which lacked optimality, to a weighted CQR ($WCQR$) version based on data-driven efficient weights. Although $WCQR$ is suitable for heavy tailed distributions, its weighting strategy does not down-weigh high leverage points.

1.1.2 Rationale of Variable Selection and Regularization in Robust Quantile Regression

Many scholars have suggested some remedies for ill-conditioned design matrix data challenges. In the *LS* setting, the inverse of the scatter matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is critical in finding a regression solution. If $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ does not exist, then a regression solution is infeasible. In heavy tailed distributions and asymmetric ones scenarios, robust alternatives to the *LS* were proposed in the regression literature, such as elemental regression (ER) (Ranganai 2007) and quantile regression (RQ). The ER is based on the elemental sets (ESs) (see Chapter 2). Regression Quantiles (RQs) are optimal solutions to a Linear Programming (LP) problem (Koenker & Basset 1978), and these optimal solutions correspond to specific ERs (Ranganai et al. 2014). Although QR is reasonably robust to outliers (response outliers), it is susceptible to high leverage points (predictor space outliers). At the optimal solution, a RQ is given by $\boldsymbol{X}_J^{-1}\boldsymbol{Y}_J$, where J denotes a $p+1$ dimensional ES. Hence, if $\boldsymbol{X}_J^{-1}$ does not exist, the QR solution is infeasible. The harmful effects of collinearity can be worse at the RQ levels, since RQs function influences are bounded in the response variable but unbounded in the predictor space. So, unlike the LS which are both susceptible to outliers and high leverage points, the QR procedure is robust to outliers but very amenable to high leverage points, hence collinearity influential points.

In regression, four scenarios can be considered: (i) the number of observations exceed that of predictor variables ($n > p$), but too many variables (need for variable reduction and variable selection), (ii) the number of predictor variables exceed that of observations ($p > n$), (iii) collinearity and (iv) collinearity influential points (collinearity hiding and collinearity inducing points). Collinearity influential observations, X-space outliers (high leverage points) that hide or induce collinearity (Mason & Gunst 1985). The last challenge (iii) is our focus of attention, and we deal with it extensively in this thesis using robust regularization techniques. Outliers in the response variable (heavy-tailed distributed error terms) are catered by QR, since RQs' influence functions are bounded in the Y-space.

In summary, the following are some motivations for regularization in QR:

- In the LS case, the inquiry is only limited to the conditional mean ($E(Y/X)$), whereas in QR, it is extended to the conditional quantiles $Q_{Y/X}(\tau)$ to all $\tau \in (0,1)$ quantile levels.

- In addition to mathematical tractability, the *LS* performs well under normality assumptions. However, in real life data, very often normality assumptions are violated due to error term distributions thicker than the normal distribution, hence the need for robust techniques such as *QR*.

- We take advantage of the robustness of *QR* in the Y-space and some penalized procedures, which are also robust in the X-space to suggest *QR* regularization procedures, with superior performance in variable selection and prediction in addition to interpretability of the *QR* model.

- Regularization procedures have been extended to the *QR* scenario with their accrued properties from the *LAD* penalized procedures in Chapter 4.

1.2 Overview of the Thesis

In this section, we give an overview of this thesis as well as present our contributions to variable selection and regularization procedures in a *QR* setting which are robust to both X-space and Y-space data aberrations. We first explore the literature on variable selection and regularization and then suggest more robust ones.

The approach to this thesis is two pronged namely; (1) variable selection and regularization in the weighted *QR* scenario via *MCD* based weights in pursuit mitigating against collinearity influential points' adverse effects in variable selection and regularization, i.e., we propose penalized weighted *QR* procedures, and (2) use *QR* based adaptive weights instead of *LS-RIDGE*-based adaptive weights in the literature in suggesting weighted and unweighted penalized *QR*. The proposed procedures are attractive because of their robustness in the presence of outlying points (both in X-space and heavy-tailed distributions).

The objectives of the thesis are to:

(i) suggest robust variable selection and regularization techniques and (ii) compare the performance of the suggested procedures through simulation studies, as well as well-known data from the literature.

1.2.1 Contributions

In this thesis, we suggest regularized *QR* procedures with *LASSO*, adaptive *LASSO* (*ALASSO*), *E-NET* and adaptive *E-NET* (*AE-NET*) penalties. We first propose the weighted penalized *QR* (*WQR*) procedures. These penalized *WQR* procedures are based on the robust weights ω_i. The weights ω_i are based on the minimum covariance determinant (*MCD*). Secondly, we suggest adaptive penalized *QR* procedures (both weighted and unweighted). The adaptive penalty is based on the proposed *WQR-RIDGE* regression parameter estimates. The variable selection approaches are robust in the Y and X-spaces. We carry out simulation studies to test the applicability and performance of these procedures. In summary:

- We extend the *WLAD-LASSO* (see Arslan 2012, Norouzirad et al. 2018) to our proposed regularized weighted *QR* (*WQR*) procedures. The penalized *WQR* procedures are based on the *RIDGE*, *E-NET*, *ALASSO* and *AE-NET* penalties. These regularized *WQR*/*QR* procedures are local estimators, unlike the *LS* estimator one, which is global.

- We propose weighted *QR-RIDGE* (*WQR-RIDGE*) based adaptive weights instead of the ridge regression (*RR*) ones, suggested in the literature. The *WQR-RIDGE* based weights have the advantage of having different adaptive weights at each *RQ* level. The weights based on the *QR* based estimator are preferred to the *RR* based ones since they are locally based and not global like the *RR* one. The penalized procedures *QR-LASSO* and *QR-E-NET* are then extended to the *QR-ALASSO* and *QR-AE-NET* methods, respectively. In turn, adaptive procedures *QR-ALASSO* and *QR-AE-NET* are further extended to *WQR-ALASSO* and *WQR-AE-NET* procedures by the same criterion.

- We take advantage of the robust weights based on the computationally intensive *MCD* due to increase in computer power to formulate our proposed penalized *WQR*s.

- We carry out a comprehensive simulation studies on penalized *WQR*s, adaptive penalized *WQR*s and adaptive penalized *QR*s in the presence of collinearity, high leverage points, collinearity influential points and heavy-tailed error term distributions as well as apply these procedures to well-known data sets from the literature with these inherent data aberrations and established the efficacy of these procedures.

The rest of the thesis is organized as presented next. In Chapter 2, we review the literature on model selection, variable selection, regularization and robust *QR* procedures. Chapter 3 discusses regularization procedures and variable selection in the *LS* scenario. Chapter 4 discusses variable selection in *QR* and suggests *WQR-RIDGE* based adaptive weights and new variable selection and regularization procedures (both weighted and adaptive) in a *QR* setting. We carry out simulation studies and give results of the simulations in Chapter 5. We conclude the thesis by a discussion and make recommendations in Chapter 6.

1.3 Notation

The notation used throughout this thesis is introduced in this section for the reader's reference.

Bold faced letters are vectors and matrices.

Table 1.1: Notation: Penalized Quantile Regression

Symbol	Description
n	sample size
p	number of predictors
$\boldsymbol{Y}$	the $n \times 1$ response vector
y_i	the i^{th} observation of the response vector $\boldsymbol{Y}$, $i \in [1:n]$.
$\boldsymbol{X}$	the $n \times p$ predictor design matrix, excluding the constant term column.
$\boldsymbol{x}_i'$	the i^{th} row of the design matrix $\boldsymbol{X}$, $i \in [1:n]$.
β_0	the LS intercept term
$\boldsymbol{\beta}$	the $p \times 1$ slope vector of parameters (excludes β_0)
e_i	i^{th} random error term, $i \in [1:n]$
$\rho_\tau(u)$	the check function given by $\begin{cases} \tau.u, \; if \; u \geq 0 \\ (\tau-1).u, \; if \; u < 0 \end{cases}$
ε_i	i^{th} residual, $i \in [1:n]$
τ	quantile level
$\boldsymbol{\beta}(\tau)$	QR coefficient.
F	the distribution function of ordered iid Y_is
$F^{-1}(\tau)$	the inverse function of F given by $inf\{y \vert F(y) \geq \tau\}$.
$Q_{Y\vert X}(\tau)$	the conditional quantile function of Y given the covariate X given by $\beta_0 + F^{-1}(\tau) + \boldsymbol{x}'\boldsymbol{\beta}$
$\hat{\boldsymbol{\beta}}(\tau)$	estimate of $\boldsymbol{\beta}(\tau)$ estimated by $\begin{pmatrix} \hat{\beta}_0 + F^{-1}(\tau) \\ \hat{\boldsymbol{\beta}} \end{pmatrix}$ with $\hat{\beta}_0 + F^{-1}(\tau)$ corresponding to the intercept and $\hat{\boldsymbol{\beta}}$ corresponds to $(\beta_1, \beta_2, ..., \beta_p)'$.
λ	the tuning parameter

Symbol	Description
$\hat{\boldsymbol{\mu}}^*$	the sample mean vector
$\hat{\boldsymbol{S}}$	the a $p \times p$ sample covariance matrix
$\boldsymbol{t}_n$	the center of MVE covering at least half of the observations
$\hat{\boldsymbol{C}}$	a $p \times p$ matrix representing shape of the ellipsoid
$\hat{\boldsymbol{\mu}}$	the sample mean of smallest ellipsoid containing half of observations
$\hat{\boldsymbol{\Sigma}}$	the $p \times p$ sample covariance matrix
$RD(x_j)$	MCD robust distance
χ^2_η	Chi-square distribution with η degrees of freedom
ω_i	robust weight based on the MCD robust distance for $i \in [1:n]$
$\hat{\boldsymbol{\beta}}^W(\tau)$	WQR coefficient
α	mixing parameter
ω_j	j^{th} RR based adaptive weight $1/\lvert\beta\rvert^\gamma$, $j \in [1:p]$
λ_j	j^{th} adaptive weight $\omega_j\lambda$, $j \in [1:p]$

Symbol	Description
$\hat{\boldsymbol{\beta}}^{R}(\tau)$	*QR-RIDGE* coefficient
$\hat{\boldsymbol{\beta}}^{WR}(\tau)$	*WQR-RIDGE* coefficient
$\hat{\boldsymbol{\beta}}^{L}(\tau)$	*QR-LASSO* coefficient
$\hat{\boldsymbol{\beta}}^{WL}(\tau)$	*WQR-LASSO* coefficient
$\hat{\boldsymbol{\beta}}^{EN}(\tau)$	*QR-E-NET* coefficient
$\hat{\boldsymbol{\beta}}^{WEN}(\tau)$	*WQR-E-NET* coefficient
$\hat{\boldsymbol{\beta}}^{WR}$	*WRR* coefficient
$\hat{\beta}_j^{WR}$	j^{th} entry of $\hat{\boldsymbol{\beta}}^{WR}$, for $j \in [1:p]$
$\tilde{\omega}_j$	$\hat{\beta}_j^{WR}$ based adaptive weight , for $j \in [1:p]$
$\hat{\beta}_j^{WR}(\tau)$	j^{th} entry of $\hat{\boldsymbol{\beta}}^{WR}(\tau)$, for $j \in [1:p]$
$\check{\omega}_j$	$\hat{\beta}_j^{WR}(\tau)$ based adaptive weight, for $j \in [1:p]$.

Symbol	Description
$\hat{\boldsymbol{\beta}}^{AL}(\tau)$	*QR-ALASSO* coefficient
$\hat{\boldsymbol{\beta}}^{WAL}(\tau)$	*WQR-ALASSO* coefficient
$\hat{\boldsymbol{\beta}}^{AE}(\tau)$	*QR-AE-NET* coefficient
$\hat{\boldsymbol{\beta}}^{WAE}(\tau)$	*WQR-AE-NET* coefficient
$\bigotimes$	Kronecker product
$\boldsymbol{\Omega}$	diagonal matrix $diag(\omega_1, \omega_2, ..., \omega_n)$
$\Psi_{ni}(t)$	$\int_0^t \sqrt{n}(F_i(s/\sqrt{n}) - F_i(0))ds$ is a convex function for each n and i, for $i \in [1:n]$
$F_i(t)$	$P(\varepsilon_i \leq t)$, is the distribution function of ε_i, for $j \in [1:p]$ and time t
$r_\tau(\boldsymbol{Z})$	denotes the approximate error function $Q_{Y\|Z}(\tau) - \boldsymbol{X}'\boldsymbol{\beta}(\tau)$
$\boldsymbol{D}$	a diagonal matrix whose entries are eigenvalues

Table 1.2: Notation: Regularization

Symbol	Description
$\hat{\boldsymbol{\beta}}^{R}$	*RIDGE* coefficient
$\hat{\boldsymbol{\beta}}_{LS}$	*LS* coefficient
$\hat{\boldsymbol{\beta}}^{L}$	*LASSO* coefficient
$\hat{\boldsymbol{\beta}}^{EN}$	*E-NET* coefficient
$\hat{\boldsymbol{\beta}}^{AL}$	*ALASSO* coefficient
ω_j^*	adaptive weight $\lvert\hat{\beta}_j^{EN}\rvert^{-\gamma}$, $j \in [1:p]$
$\hat{\boldsymbol{\beta}}^{AEN}$	*AE-NET* coefficient
$\hat{\boldsymbol{\beta}}^{FL}$	Fused *LASSO* coefficient
β_{j-1}	$(j-1)^{th}$ parameter, $j \in [2:p]$
$\hat{\boldsymbol{\beta}}^{GL}$	Grouped *LASSO* coefficient
κ_j	denotes the j^{th} elliptical norm of $\boldsymbol{\beta}$, $j \in [1:p]$
$\hat{\boldsymbol{\beta}}^{CAP}$	composite absolute penalties slope
$y(t_i)$	denotes a time dependent response variable, where $t_i \in [t_1,...,t_n]$, $i \in [1:n]$
$\varepsilon(t_i)$	time-dependent error term, where $i \in [1:n]$
$\beta_j(t_i)$	time-dependent parameter estimate, $i \in [1:n]$
$w(t_s,t_i)$	time-dependent weight, where $i \in [1:n]$, $s \in [1:n]$
$\hat{\boldsymbol{\beta}}^{SL}$	smoothed *LASSO* criterion coefficient

Table 1.3: General Notation

Symbol	Description
S_t	train set
S_v	validation set
$\boldsymbol{H}$	is the the hat matrix $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$
h_{ii}	the i^{th} diagonal element of the hat matrix $\boldsymbol{H}$, given by $\boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i$, $i \in [1:n]$
$\boldsymbol{H}_{(i)}$	is the hat matrix with i^{th}observation deleted, given by $\boldsymbol{X}(\boldsymbol{X}_{(i)}'\boldsymbol{X}_{(i)})^{-1}\boldsymbol{X}'$, $i \in [1:n]$
$\boldsymbol{X}_c$	relative to the clean subset
$\boldsymbol{H}_c$	hat matrix $\boldsymbol{X}(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}'$ relative to the clean subset
h_{c_i}	the i^{th} leverage value relative to the clean subset $\boldsymbol{X}_c$, given by $\boldsymbol{x}_i(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{x}_i'$, $i \in [1:n]$ ($\boldsymbol{X}$ without high leverage points)
v_j	the weight based on h_{c_i}, $i \in [1:n]$
$\tilde{\boldsymbol{X}}$	$n \times p$ design matrix $\boldsymbol{X}$ standardized to correlation form

$\boldsymbol{R}$	correlation matrix $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$, where $\tilde{\boldsymbol{X}}$ is a normalized matrix
r_{ij}	$(ij)^{th}$ element of the correlation matrix $\boldsymbol{R}$, $i \in [1:n]$, $j \in [1:p]$
R_j^2	j^{th} coefficient of determination (COD), $j \in [1:p]$
Tol_j	j^{th} tolerance ratio given by $1 - R_j^2$, where $j \in [1:p]$
$\boldsymbol{I}$	identity matrix
β_j	j^{th} parameter
θ	parameter value, either β_j or β_j^2
$\check{\mathbf{X}} = (\mathbf{1}, \mathbf{X})$	an $n \times (p+1)$ predictor matrix including a column of 1s
$\mathbf{1}_n$	column of 1s
$\check{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}')'$	$(p+1) \times 1$ parameter vector (p is the number of predictors)
$\mathbf{e} \sim N_n(\mathbf{0}_n, \sigma^2 \mathbf{I}_n)$	an $n \times 1$ vector of errors
$\mathbf{0}_n$	vector of zeros
$\check{\mathbf{X}}_J$	a $(p+1) \times (p+1)$ submatrix of $\check{\mathbf{X}}$
$\mathbf{Y}_J$	a $(p+1) \times 1$ subvector of $\mathbf{Y}$
J	denotes the $p+1$ dimensional ES
$\hat{\check{\boldsymbol{\beta}}}_J$	$(\check{\mathbf{X}}_J'\check{\mathbf{X}}_J)^{-1}\check{\mathbf{X}}_J'\mathbf{Y}_J = \check{\mathbf{X}}_J^{-1}\mathbf{Y}_J$, estimated vector of coefficients
$\check{\mathbf{X}}_J^{-1}$	the inverse of the matrix $\check{\mathbf{X}}_J$
$\hat{y}_{iJ}$	the i^{th} elemental predicted value, $\check{\mathbf{x}}_i'\hat{\check{\boldsymbol{\beta}}}_J$
ε_{iJ}	the i^{th} elemental predicted residual (EPR) $y_{iJ} - \check{\mathbf{x}}_i'\hat{\check{\boldsymbol{\beta}}}_J$, for, $i \notin J$
κ_j	λ_j/λ_p is a singular value or eigenvalue based condition index, where λ_p is the minimum singular value and λ_p is the maximum singular value, for $j \in [1:p]$

Chapter 2

Literature Review

In this chapter we review and discuss the literature on variable selection and regularization for non-robust and robust scenarios. We further review the use of robust criteria in variable selection and regularization, the idea of robustness, some computational aspects of *QR*, collinearity and influential point diagnosis. We justify the extension of the robust idea of the *LAD* to the proposed *QR* regularization procedures. Extensions of existing robust methods have made the robust field rich in diversity in variable selection and parameter estimation. Also, a review of robust distances is given as a precursor to the construction of robust weights and finally, a brief review of elemental sets method is given, since a *RQ* corresponds to an *ES*.

2.1 Review of Variable Selection and Regularization Procedures

The variable selection procedures include the significance criterion, the change-in-estimate criterion, the forward stagewise procedure, stepwise forward selection procedure, stepwise backward selection procedure, as well as their hybrid versions and the regularization (penalization and shrinkage) methods. Remedies of collinearity's adverse effects have been extensively researched in some of these variable selection procedures. The significance criterion, information criterion and the stepwise backward procedures are examples of criteria preferred when collinearity is present (see Mantel 1970, cited in Heinze et al. (2018)), and have become integrated in standard statistical software. We briefly explain some of these procedures in subsequent paragraphs.

Significance criterion is the most popular practical variable selection procedure that uses hypothesis testing methods to choose variables. Different linear regression models are fit and compared using the likelihood ratio test. In logistic regression and survival analysis (Cox regression model), the step-up (score test) or step-down (Wald test) tests select variables (Heinze et al. 2018).

The second variable selection method is the information criterion, which selects a model from a set of plausible models. This criterion penalizes a model for complexity. Examples of information criterion include the Akaike information criterion (AIC) (Akaike 1974) and the Bayesian information criterion (BIC) (Schwarz 1978). The penalty factor of BIC is usually larger than the AIC, and BIC chooses more parsimonious models (Heinze et al. 2018).

In epidemiology, the change-in-estimate criterion (see Hosmer Jr et al. 2013) has been preferred. In the change-in-estimate criterion, eliminating a significant predictor from the model results in significant change-in-estimate, and the converse is true for a nonsignificant predictor. The method includes the augmented backward elimination (ABE) procedure as an example. This method selects more variables and less biased coefficients than the backward elimination (BE) procedure.

Forward stepwise regression selects the first predictor variable with the best fit, i.e. the variable with the least sum of squared errors (SSE) (Hesterberg et al. 2008). Subsequently, a predictor that provides the best combination fit with the first one is chosen, and the process continues in that fashion. The forward stepwise regression procedure is susceptible to small changes in data, leading to choosing one variable instead of another, hence unstable.

Backward stepwise regression procedure starts with a larger model, sequentially removing variables that have the least contribution to the fit. It is prudent to note the use of the combination of forward and backward selection procedures, as in Efroymson's procedure (Efroymson 1960).

The forward stagewise procedure, just like its forward stepwise counterpart, selects the first predictor variable with the least *SSE* (best fitting variable). In the next step, the forward stagewise regression selects the variable with the highest correlation with the current residuals, and the procedure continues that way. In comparison, the stagewise procedure produces more stable coefficients compared to its stepwise counterpart. The stagewise procedure is closely related to the boosting algorithm in machine learning (Hesterberg et al. 2008).

Variable selection can be determined by regularization methods. These shrinkage procedures include the *RIDGE* (see Hoerl & Kennard 1970, Hoerl et al. 1975), *LASSO* (Tibshirani 1996) and the least angle regression (*LARS*) (Efron et al. 2004) penalty based procedures as examples. However, the *RIDGE* procedure does not select predictors, as it does not shrink coefficients to zero. The optimum tuning parameter (λ) in *LASSO* and other shrinkage procedures control the penalization strength through cross validation or information criteria (Yu & Feng 2014). The regularization methods are suitable for the low dimensional case ($n > p$), as well as the high dimensional case ($p > n$). These penalization methodologies, as extended to the *QR* scenario, are the focus of this thesis, and we discuss these methods in Chapter 3.

2.1.1 Criteria for Model Performance in Variable Selection

Our interest in this thesis is to determine models with better prediction accuracy that can give precise estimates of coefficients of interest and enhance interpretability of models. The application of error statistics is very popular, and we summarize some of them. Some measures of model performance include (i) mean squared error (MSE), (ii) root mean squared error ($RMSE$), (iii) mean absolute error (MAE) and (iv) median absolute deviation (MAD), *etc*. The $RMSE$ is a frequently used statistic influenced by extreme points and, MAD, though less popular like other robust metrics, is robust in the presence of outlier observations.

We start by presenting the MSE as

$$MSE = (1/n)\Sigma_{i=1}^{n}\varepsilon_i^2 \; for \; i \in [1:n], \tag{2.1}$$

where $\varepsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ is the i^{th} residual value. The extreme point, when squared, unduly increases the MSE value. Finding the square root results in the root mean squared error ($RMSE = \sqrt{MSE}$). The MSE and $RMSE$ are amenable to outliers. Instead of MSE and $RMSE$, the mean absolute error ($MAE = (1/n)\Sigma_{i=1}^{n}|\varepsilon_i|$) is more appropriate. The MAE is robust in the presence of outliers (Rousseeuw & Croux 1993). Another measure used to show model bias (under and over prediction of a model) is mean bias error ($MBE = (1/n)\Sigma_{i=1}^{n}\varepsilon_i$) (see Willmott & Matsuura 2005). We depict the relationship between the three measures by $MBE \leq MAE \leq RMSE$ (Willmott & Matsuura 2005).

Consider a very robust scale estimator MAD given by

$$MAD = 1.4826\left(Median\{\varepsilon_i\} - Median\{\varepsilon_i\}\right), i \in [1:n], \tag{2.2}$$

where $Median\{\varepsilon_i\}$ is the median of test errors, which is basically the middle order statistic for odd n and average of $(n/2)^{th}$ and $(n/2+1)^{th}$ of ordered values for even n. The median is bounded and has the highest breakdown point of 50% (breakdown point measures the least amount of contamination required for a procedure to disintegrate (see Section 2.3)). The median influence function is also bounded (Rousseeuw & Croux 1993). The robustness of this measure makes it suitable for screening data for outliers by the statistic $|x_i - Median\{\varepsilon_i\}|/MAD_n$. The MAD statistic, with a cutoff 2.5 or 3.0, flag spurious data points.

2.1.2 Cross Validation in Regularization Procedures

In the variable selection and regularization techniques, we use some criteria with cross-validation (CV) criteria to select variables. The CV technique selects the optimal tuning parameter or penalty parameter (λ), resulting in selecting the best procedure with the most accurate overall prediction.

Consider a learning method f and data set$D = \{\langle \boldsymbol{x}_j, y_j \rangle\}, \in [1:N]$. Then for predictor vector $\boldsymbol{x}_j$ and response y, the predictive model $M(x)$ is such that $f(D) = M$ and output prediction given by $f(x,D)$. The K-fold CV algorithm $CV(f, D = F_1, ..., F_K)$ (Tsamardinos et al. 2018) in the regularization procedures is given by

---

Algorithm : ***K-Fold Cross-Validation***

Input: training method f , Data matrix $\boldsymbol{D} = \{\langle x_j, y_j \rangle\}, \in [1:N]$ partitioned into equal folds F_i.
Output: Model M, Performance estimation CV, on all folds
(i) Define $\boldsymbol{D}_{(i)} = \boldsymbol{D}$, for $\boldsymbol{F}_i \notin \boldsymbol{D}$ # Data set $\boldsymbol{D}$ is partitioned into K folds F.
$\boldsymbol{D}_{(i)}$ and $\boldsymbol{F}_i$ are referred to as the training set (S_t) and validation set (S_v), respectively.
(ii) I_is are the indexes of $\boldsymbol{F}_i$ # Obtain the indexes of each fold
(iii) $M = f(D)$ # Final Model trained by f on all available data
(iv) $L_{CV} = \frac{1}{K}\Sigma_i^K l(y(I_i), f(Fi, D(i))), i \in [1:K]$ # Performance estimation: learn from $\boldsymbol{D}_{(i)}$, estimate on F_i
train set $\boldsymbol{D}_{(i)}$ is used to fit models, and we test the performance of the model using the validation set F_i
(v) Collect out-of-sample predictions $\Pi = [f(\boldsymbol{F}_1, \boldsymbol{D}_{(1)}); ...; f(\boldsymbol{F}_K, \boldsymbol{D}_{(K)})]$ # Out-of-sample predictions are used by bias-correction methods
(vi) ***Return*** $\langle M, L_{CV}, \Pi \rangle$

In this thesis we use *hqreg R* software program (Yi 2017) which has an inbuilt K-fold cross-validation algorithm for application regularization and regularized *QR* scenarios (see Section 5.1).

2.2 Least Squares and Least Absolute Deviation Variable Selection and Regularization Procedures

In this section, we focus on non-robust (*LS*) and robust (*LAD*) variable selection and regularization procedures. Robust penalized procedures have the advantage of resisting the influence of extreme observations in variable selection and parameter estimation (Jiang et al. 2021). We start by elaborating on the *LS* and *LAD* penalized procedures in the next section.

2.2.1 Least Squares Variable Selection and Regularization Procedures

Regularization techniques exist in the literature, including the James–Stein estimator (James & Stein 1961), ridge regression referred in this thesis as least squares ridge regression (*LS-RIDGE*) (Hoerl & Kennard 1970), non-negative garotte (Breiman 1995), least absolute shrinkage and selection operator (*LS-LASSO*) (Tibshirani 1996), smoothly clipped absolute deviation (*SCAD*) (Fan & Li 2001), *LS* elastic net (*LS-E-NET*) (Zou & Hastie 2005), adaptive *LASSO* (Zou 2006) and the least squares post-*LASSO* (*LS-post-LASSO*) (Belloni & Chernozhukov 2013). The regularization techniques have sparse solutions, except for the *LS-RIDGE* procedure. The literature shows that the *LS-post-LASSO* technique reduced the bias of the *LASSO* estimator. The requirement of $p < n$ in *LS* and James–Stein shrinkage is unimportant for the *LASSO* case. Procedures such as the *LAD-LASSO* (Wang et al. 2007), *ALASSO* proposed by Zou (2006), fused *LASSO* (Tibshirani et al.

2005), group *LASSO* (Yuan & Lin 2006), *LS-E-NET* (Zou & Hastie 2005) and *AE-NET* (Zou & Zhang 2009) are extensions of *LASSO*. The *LASSO* procedure has merits of stable model selection and produces a sparse solution (Wu & Liu 2009). The *E-NET* procedure has similar sparsity of representation to the *LASSO* but often outperform the *LASSO* procedure (Zou & Hastie 2005). The *E-NET* has a competitive advantage over the *LASSO* in that it encourages a grouping effect, where strongly correlated predictors tend to be in or out of the model together. Extreme observations, including high leverage points, more-so collinearity influential points, adversely influence the regularization procedures highly. These penalties have drawbacks of not being robust in the presence of high leverage points (and collinearity influential points). Therefore, in the literature, they have been improved via their adaptive counterparts which we elaborate on further in Subsection 2.2.2 and Chapter 3.

2.2.2 Least Absolute Deviation Variable Selection and Regularization

Regularization procedures based on absolute deviations are robust in the presence of Y-space outliers. Thus, the *LAD* regression is robust to outliers in the response space (heavy-tailed distributions scenarios), unlike the *LS* procedure. The *LAD-LASSO* exploit the robustness of *LAD* in the estimation of parameters and the shrinkage and estimation characteristics, since both the *LAD-LASSO* and the *LAD* are based on absolute deviations, unlike *LS-LASSO* as proposed by Wang et al. (2007). The *LAD-LASSO* estimator also enjoys asymptotic efficiency (oracle property) just like the *LAD* estimator. In *LS*, omitting an important predictor variable will cause a bias in parameter estimates and prediction results. *LAD-LASSO* is a consistent model selection criterion under heavy-tailed distributions, as compared to *LS-LASSO* and other variable selection procedures, such as *AIC*, *BIC*, *etc* that fall short.

Regularization procedures that are robust in the presence of X-space outliers (high leverage points) exist in the literature. Arslan (2012) proposed the *WLAD-LASSO* procedure for robust parameter estimation and variable selection. The *WLAD-LASSO* procedure is superior to the *LAD-LASSO* because it is robust to both the outliers in the predictor space, as well as outliers in the response variable (heavy-tailed distributions). *WLAD-LASSO* inherits robustness in the Y-space from its *LAD-LASSO* counterpart, while its high sensitivity to high leverage point influences is mitigated by the weights applied to the predictor matrix.

In the literature, suggested adaptive regularization methods are robust in the presence of collinearity. One such regularization procedure is the adaptive *LASSO* (*ALASSO*), which was proposed by Zou (2006) to deal with certain situations where *LASSO* is inconsistent for variable selection. *ALASSO* is based on adaptive weights and enjoys oracle properties. Fan et al. (2014) suggested the adaptive robust *LASSO* (*AR-LASSO*) to reduce bias in the procedures by Belloni & Chernozhukov (2011) and Wang et al. (2012). In the *AR-LASSO* scenario, a weight vector of the weighted ℓ_1 penalty is computed. Fan et al. (2014) formally established its asymptotic normality property. We extend these procedures in Chapter 3 to the adaptive procedures namely, adaptive *RIDGE* (*ARIDGE*), *ALASSO*, *AE-NET*, weighted *ALASSO* (*WALASSO*), weighted *AE-NET* (*WAE-NET*), weighted *LAD ALASSO* (*WLAD-ALASSO*) and weighted *LAD AE-NET* (*WLAD-AE-NET*) and in Chapter 4 namely, *QR-ALASSO*, *QR-AE-NET*, weighted *QR-ALASSO* (*WQR-ALASSO*) and weighted *QR-AE-NET* (*WQR-AE-NET*).

2.3 Robustness and Quantile Regression Variable Selection and Regularization

In this section, we give a review of robustness, as well as variable selection and regularization in *QR*.

2.3.1 The Breakdown Point and Quantile Estimation

The robustness of *QR* in the Y-space emanates from the *LAD* estimator, as it is a generalization of *LAD* to all regression quantile (RQ) levels. According to Koenker et al. (2018), we can reduce the residual mean problem of Boscovich to finding the weighted median regression. The *QR* case is an immediate generalization to the median regression case, resulting in the τ^{th} quantile influence function. The conditional mean achieves optimality under the Gaussian law of errors, and the median has a superior performance than the mean when large errors exist. Small contamination of the distribution in the mean scenario at point y contrasts the median scenario, which is bounded by the sparsity at the median. The boundedness property of the median scenario extends to the *QR* case for a finite sparsity.

Koenker (2005) acknowledges Donoho and Huber's sample breakdown point as the most successful notion of global robustness of estimators, since it measures the least amount of contamination required for a procedure to disintegrate (see also Bickel et al. 1982). These high breakdown methods, like the least median squares estimator by Rousseeuw (1984), have become the focus of recent research and achieve asymptotic breakdown point half ($\frac{1}{2}$). However, because of its non-probabilistic formulation, the breakdown point of estimators is still an elusive concept.

Rousseeuw and Hubert in 1999 suggested the use of regression depth in *QR*. They found the approach to be highly robust. The regression depth fits into recent linear programming methods when supported by more explicit rules of choosing edges at each vertex. The dual plot strategy chooses the most favorable direction by finding directional derivatives until the objective function is minimal.

2.3.2 Review of Robust Variable Selection and Regularization in Quantile Regression

Some *LASSO*-penalty (ℓ_1) based *QR*s include the mixed-effect for longitudinal data in estimating random effects by shrinkage approaches (Koenker 2005), the solution path of ℓ_1-penalized *QR* (Li & Zhu 2008) and, *LAD* regression (Wang et al. 2007). The *QR* loss function is not continuous at the origin, which poses challenges on the applicability of oracle properties for the non-concave likelihood (Fan & Li 2001). *QR* variable selection and regularization procedures, such as *SCAD* penalized *QR* and adaptive *LASSO* penalized *QR* proposed by Wu & Liu (2009) have the advantages of oracle properties. The oracle properties of the *SCAD* penalized *QR* and adaptive *LASSO* penalized *QR* procedures are very important in our proposed procedures whose properties we either adopt or extend in this thesis.

Composite quantile regression *LASSO* (*CQR-LASSO*) procedure (Zou & Yuan 2008) is robust just like *CQR* and assumes equal weights for different *RQ*s. The equal weight property of this procedure lacked optimality overall and hence Jiang et al. (2012) proposed the weighted *CQR* (*WCQR*), a procedure that let the data decide efficient weights. Belloni & Chernozhukov (2011) proposed the ℓ_1-penalized *QR* in high dimensional sparse regression models whose estimates are nearly oracle consistent. Wang et al. (2012) considered the non-convex penalized *QR*s in high dimensional settings, which under mild error conditions satisfy oracle properties at a local minima.

2.3.3 Overview of Quantile Regression Computational Aspects

We briefly discuss some *QR* computational aspects in this section. The interior and exterior point procedures are used to search for a distinct *QR* solution set of β_js. As alluded to in Koenker & D'Orey (1987), exterior point procedures and classical parametric linear programming methods are used to find distinct sets of solutions of β_js. The simplex programming method, together with the exterior point procedure, move on the exterior of the constraint set along the edges, searching for this *QR* solution set. Some similar parametric programming methods in the literature are used to determine the *LASSO*-type penalized estimators as well. The number of distinct solutions in exterior point procedures is overwhelming hence the adoption of interior point methods.

The interior point procedures search solution sets from the center of the constraint set towards the edges (Koenker et al. 2018). The development of interior point procedures spans back to 1956, with Frisch's contribution to the logarithmic barrier procedure. Wu & Liu (2009) applied the difference convex algorithm (DCA) to solve the non-convex smoothly clipped absolute deviation (*SCAD*) optimization problem because the penalty function can be decomposed into a difference of two convex functions. The optimization *DCA* for *QR* at every iteration is a more efficient linear programming problem. The local linear approximation algorithm (LLA) proposed by Zou & Li (2008) solves the *SCAD* optimization problem just like the *DCA*. The *LLA* and *DCA* approaches differ in that the *LLA* enforces symmetry in approximating the *SCAD* penalty, hence the *DCA* approach is more attractive than its *LLA* counterpart.

2.4 Review of Collinearity and Collinearity Influential Point Diagnostics

In this section, we explore collinearity and collinearity influential point diagnostics. Consider the design matrix $\boldsymbol{X}$ and normalize it by sample size and standard deviation to unit length by $\tilde{x}_{ij} = (x_{ij} - \bar{x}_j)/S_j$, where $S_j = \Sigma_{i=1}^{n}(x_{ij} - \bar{x}_j)^2,\ i \in [1:n]$ and $j \in [1:p]$ (Midi & Bagheri 2013). The resultant matrix product $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ is a correlation matrix denoted by $\boldsymbol{R}$. As collinearity increase, the correlation matrix $\boldsymbol{R}$ approaches singularity, causing some explosive increase in the elements of the inverse matrix $(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1} = \boldsymbol{R}^{-1}$ (Farrar & Glauber 1967). Consider the partitioned matrix $\tilde{\boldsymbol{X}} = \begin{pmatrix} \tilde{\boldsymbol{X}}_1 & \tilde{\boldsymbol{X}}_2 \end{pmatrix}$, where $\tilde{\boldsymbol{X}}_1$ consists of predictor variables x_k and x_j $k \neq j$ and $\tilde{\boldsymbol{X}}_2$ consists of the remaining predictors. The correlation matrix is then partitioned such that $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{pmatrix}$, where $\boldsymbol{R}_{11}$ and $\boldsymbol{R}_{22}$ are 2×2 and $(p-2) \times (p-2)$ sub-matrices of the correlation matrix, respectively. The inverse of the first term in the partition is given by $\boldsymbol{R}_{11}^{-1} = (\boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21})^{-1}$, where the inversion, the single off-diagonal element of $\boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$ is the partial covariance of x_k and x_j, holding $\tilde{\boldsymbol{X}}_2$ constant.

Some collinearity diagnostics include the correlation matrix of predictors, variance inflation factor (VIF_j), the eigen-system analysis of correlation matrix (*EAOCM*) and the condition number (*CN*). Schaefer et al. (2006) considered the criteria for diagnosing collinearity using coefficient of determination ($COD = R_j^2$) by regressing x_j (the j^{th} predictor) on the other remaining predictors, with a corresponding correlation coefficient $R_j = \Sigma(x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k)/\sqrt{\Sigma(x_{ji} - \bar{x}_j)^2\Sigma(x_{ki} - \bar{x}_k)^2}$, where j and k $(j \neq k)$ are columns of the design matrix $\tilde{\boldsymbol{X}}$. Collinearity exists if the *COD* (and correlation coefficient), $R_j^2 \to 1 (R_j \to \pm 1)$ and sum of squared residuals, $SSE \to \infty$. This criterion is equivalent to considering $CN > 10$ as evidence of collinearity. If we consider the variance of β_j ($Var(\beta_j) = \sigma^2.VIF_j/\Sigma(x_{ji} - \bar{x}_j)^2$), where $VIF_j = (1 - R_j^2)^{-1}$ is the variance inflation factor of the j^{th} coefficient show that collinearity inflates variance of the βs. Note that $1 - R_j$ is the tolerance ratio (Tol_j). A large value of the collinearity diagnostic measure VIF_j ($VIF_j > 10$) shows collinearity exists (O'brien 2007). A $Tol_j \leq 0.20$ point to a potential problem and $Tol_j \leq 0.10$ indicates serious collinearity problems.

Collinearity is related to the conditioning of the design matrix $\tilde{\boldsymbol{X}}$ and the smallest eigenvalue of $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ (see Farrar & Glauber 1967, Silvey 1969). The determinant of the normalized matrix $|\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}|$ (correlation matrix) yields some insights into the degree of collinearity within a predictor variable set on a scale $0 \leq |\boldsymbol{R}| \leq 1$. We note that $|\boldsymbol{R}| \to 0$ means $\tilde{\boldsymbol{X}}$ approach singularity and the converse $|\boldsymbol{R}| \to 1$ imply near orthogonal independent variable set (Farrar & Glauber 1967). Let $\tilde{\boldsymbol{X}}'_{(j)}\tilde{\boldsymbol{X}}_{(j)}$ denote the correlation matrix excluding the j^{th} variable. We then have $r^j = |\boldsymbol{R}_{(j)}|/|\boldsymbol{R}| \to \infty$ indicating the existance of collinearity, hence, the location of singularity within the predictor variable set. The determinant $|\boldsymbol{R}|$, shows existence of collinearity, where r^j gives insight into the location. For the predictor variables X_js, $j \in [1:p]$, the COD of the j^{th} predictor variable when it is regressed on the remaining of the predictors is given by $R_j^2 = 1 - \Delta/\Delta_j$, where $\Delta = |\boldsymbol{R}|$ is the determinant of the correlation matrix of all predictor variables and $\Delta_j = |\boldsymbol{R}_j|$ is the determinant of the correlation matrix of all predictor variables but the j^{th} predictor is left out. We can then use $r^k = R^2_{(k)}/(1 - R^2_{(k)})$ for individual point influence on collinearity, where $0 \leq r^k < \infty$.

The generalized variance inflation factor is given by $GVIF_1 = det(\boldsymbol{R}_{11})/det(\boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21})$ which can be expressed as $GVIF_1 = det(\boldsymbol{R}_{11}) \times det(\boldsymbol{R}_{22})/det(\boldsymbol{R})$, for all predictors and $det(\boldsymbol{R}) = det(\boldsymbol{R}_{22}) \times det(\boldsymbol{R}_{11} - \boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21})$ (see Fox & Monette 1992). We note that $det(\boldsymbol{R}) = 1$ and $det(\boldsymbol{R}) = 0$ for orthogonal and perfectly collinear data, respectively.

In the eigen-system analysis of $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$, at least one small (near-zero) eigenvalue implies near linear dependency between the columns of $\tilde{\boldsymbol{X}}$. In Montgomery et al. (2012), collinearity is examined using a *CN*. Consider the singular values of $\tilde{\boldsymbol{X}}$ to be $d_1, d_2, ..., d_p$ and the eigenvalues of $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ to be $\lambda_1, \lambda_2, ..., \lambda_p$. If we consider the singular value of $\tilde{\boldsymbol{X}}$ and the eigenvalues of $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ such that that $d_1 \geq d_2, ..., \geq d_p$ and $\lambda_1 \geq \lambda_2, ..., \geq \lambda_p$, then, the *CN* of $\tilde{\boldsymbol{X}}$ is given by $\kappa = d_1/d_p = \sqrt{\lambda_1}/\sqrt{\lambda_p}$ (Chatterjee & Hadi 1988*b*). The minimum bound of $\kappa = 1$ is attained when the columns of $\tilde{\boldsymbol{X}}$ are orthogonal. Generally, the diagnostic criterion uses the SVD, which is related to the eigenvalue decomposition (EVD) introduced by Strang (2006). The $n \times p$ design matrix $\tilde{\boldsymbol{X}}$, is decomposed into $\tilde{\boldsymbol{X}} = \boldsymbol{UDV}'$, where $\boldsymbol{V}$ and $\boldsymbol{U}$ are the $p \times p$ matrix of eigenvalues and $n \times p$ matrix of columns associated with p non-zero eigenvalues of $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$, respectively. The matrix $\boldsymbol{D}$, is a diagonal matrix of eigenvalues of $\tilde{\boldsymbol{X}}$, where $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$ and $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$. SVD is intimately related to the theory of diagonalizing a symmetric matrix $\tilde{\boldsymbol{X}}$. The diagonal entries of $\boldsymbol{D}$, are eigenvalues of $\tilde{\boldsymbol{X}}$. The SVD is found by considering an arbitrary real $n \times p$ matrix $\tilde{\boldsymbol{X}}$. This arbitrary real $n \times p$ matrix is such that columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are called left and right singular vectors of $\tilde{\boldsymbol{X}}$, respectively and the diagonal matrix $\boldsymbol{\Omega} : \tilde{\boldsymbol{X}}_{n\times p} = \boldsymbol{U}_{n\times r}\boldsymbol{\Omega}_{r\times r}\boldsymbol{V}'_{r\times p}$, where $r = rank(\tilde{\boldsymbol{X}}) \leq p$. The positive values from the diagonal matrix $\boldsymbol{\Omega}$ are called singular values of $\tilde{\boldsymbol{X}}$. Although using eigenvaues and singular values in diagnostics is equivalent, computationally dealing with the latter is preferable even when $\tilde{\boldsymbol{X}}$ is ill-conditioned.

Sengupta & Bhimasankaram (1997) suggested the collinearity influential measure (CIM), which can identify both collinearity enhancing and reducing points (see also Imon 2002, for the extension to group deletion tools). The leverage collinearity influential point diagnosis tools are based on generalized potentials and are called the diagnostic robust generalized potentials $(DRGP)$ (see Habshah et al. 2009, Bagheri et al. 2012).

2.5 Robust Distances

In this section, we discuss the robust distances. Mahalanobis distance (MD) is the distance between points from the center (centroid) of a data set onto an ellipse (ellipsoid) (Mahalanobis 1936). The MD is superior to its Euclidean distance counterpart, which measures the shortest distance between two points in various ways. Whereas the Euclidean distance is prone to highly correlated variables, the MD is robust in the presence of collinearity. The contribution of each variable scales the MD value according to its variability.

Let $\boldsymbol{x}'_j \in R^p$ be the j^{th} predictor vector produced by an $n \times p$ design matrix $\boldsymbol{X}$ and the sample mean vector $\bar{\boldsymbol{X}} = \hat{\boldsymbol{\mu}}^*$. To investigate aberrant points in the data set, we use the MD. If we denote $\widehat{\boldsymbol{S}} = E(\boldsymbol{X} - \hat{\boldsymbol{\mu}}^*)'(\boldsymbol{X} - \hat{\boldsymbol{\mu}}^*)$ as the covariance matrix, then Mahalanobis distance is given by $MD(\boldsymbol{x}_i) = \sqrt{(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}^*)'\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}^*)},\ for\ i \in [1:n]$. If $\widehat{\boldsymbol{S}} = \boldsymbol{I}$, the MD reduces to Euclidean distance. The MD accounts for the covariance (or correlation) structure of the data (Ghorbani 2019). The sample mean vector and covariance matrices are known to be heavily influenced by outliers in a multivariate scenario, and real outliers might cause small MDs. In this case, outliers remain undetected by a phenomenon called the masking effect. The normality assumptions are often violated, and a solution is proffered by considering robust alternative estimators, such as the robust MD.

A high breakdown robust estimator of multivariate location and scatter called the minimum volume ellipsoid (MVE) estimator was proposed by Rousseeuw (1984) (see also Rousseeuw 1985). The MVE is based on the ellipsoid, with the least volume that covers h of the n cases. The MVE, because of its low bias, is suitable for outlier detection in multivariate data scenarios with the aid of robust distances. The MVE is a reliable procedure for outlier diagnosis, since it is highly resistant to outliers and has an easily accessible computational algorithm.

The MVE estimator is the center and covariance structure of the ellipsoid, with a minimal volume that covers at least h points of the design matrix. The h points are chosen between $[n/2]+1$ and n, where $[x]$ denotes the integer part of x. If we choose $h = [(n+p+1)/2]$, a maximal breakdown value results.

Based on MVE (Rousseeuw 1985) estimates of location and scatter, a robust distance is given by $RD(x_i) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{t}_n)' \widehat{\boldsymbol{C}}_n^{-1} (\boldsymbol{x}_i - \boldsymbol{t}_n)}$, $i \in [1:n]$, where $\boldsymbol{t}_n$ is the center of MVE covering at least half of the observations, and $\widehat{\boldsymbol{C}}_n$ is a $p \times p$ matrix representing shape of the ellipsoid. The main merit of the MVE is its high breakdown point close to $1/2$ and conceptually one of the simplest high breakdown point estimators (Rousseeuw & Leroy 1987).

Another high breakdown robust estimator of center and scatter called the minimum covariance determinant (*MCD*) estimator was proposed by Rousseeuw (1985). The *MCD* estimator is affine equivariant, highly robust estimator of multivariate center and scatter and is very useful in outlier detection. Using the *MCD* estimator was limited until recently due to computer power and less computationally efficient algorithms (see Rousseeuw & Driessen 1999). We have seen the *MCD* method being used in a variety of areas including medicine, finance and chemistry (Rousseeuw & Driessen 1999). The *MCD* estimator has in recent years gained traction in robust multivariate techniques, such as factor analysis and robust principal components analysis (*RPCA*).

Let $\boldsymbol{X} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{ij}, ..., \boldsymbol{x}_{ip})'$ be a data matrix, where $\boldsymbol{x}_{ij}$ is the j^{th} predictor vector in a multivariate location and scatter setting. The *MCD* location estimate $\widehat{\boldsymbol{\mu}}$ is the sample mean vector of the smallest ellipsoid containing half (or $h(n/2 \leq h \leq n)$ observations as defined by the user) of the observations of the design matrix $\boldsymbol{X}$, whereas the *MCD* scatter estimate $\widehat{\boldsymbol{\Sigma}}$ is the covariance matrix multiplied by a consistent factor (Rousseeuw & Hubert 2018).

Based on *MCD* estimator (Rousseeuw 1984) of location and scatter, the robust distance is given by

$$RD(x_i) = \sqrt{\left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}\right)' \widehat{\boldsymbol{\Sigma}}^{-1} \left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}\right)},\ i \in [1:n], \tag{2.3}$$

where $RD(x_i)$ is the *MCD* based robust distance, a modification of the classical Mahalanobis distance. $RD(x_i)^2$ follow a χ^2-distribution with η degrees of freedom. The robust *MCD* based estimator is used to diagnose outliers in multivariate data. The robust distances based on *MCD* are useful in flagging outliers because they are not sensitive to masking effects (Rousseeuw & van Zomeren 1990). The *MCD* based distance is superior to the classical estimator (*MD*), as it is not strongly affected by contamination.

A data set well known in the literature to contain both high leverage points and outliers is the Brainlog data (Rousseeuw & Leroy 1987). We illustrate these inherent data aberrations in *Figure* (2.1). As depicted in *Figure* (2.1), the extreme leverage points pull the *QR* fit towards themselves, resulting in crossing *RQ* lines. On the other hand, *QR* excludes outliers. There is evidence of the effectiveness of the robust distance measure compared to the classical *MD* in detecting both outliers and high leverage points. While the classical ellipsoid is inflated towards the X-outliers 6, 16 and 26, which are dinosaurs with low brain weight and high body weight, the *MCD* one totally excludes them. The boundary ones, 14 and 17, are human and rhesus monkey, respectively. Also, the *distance-distance* plot shows that the robust distance exposes more X and Y-space outliers.

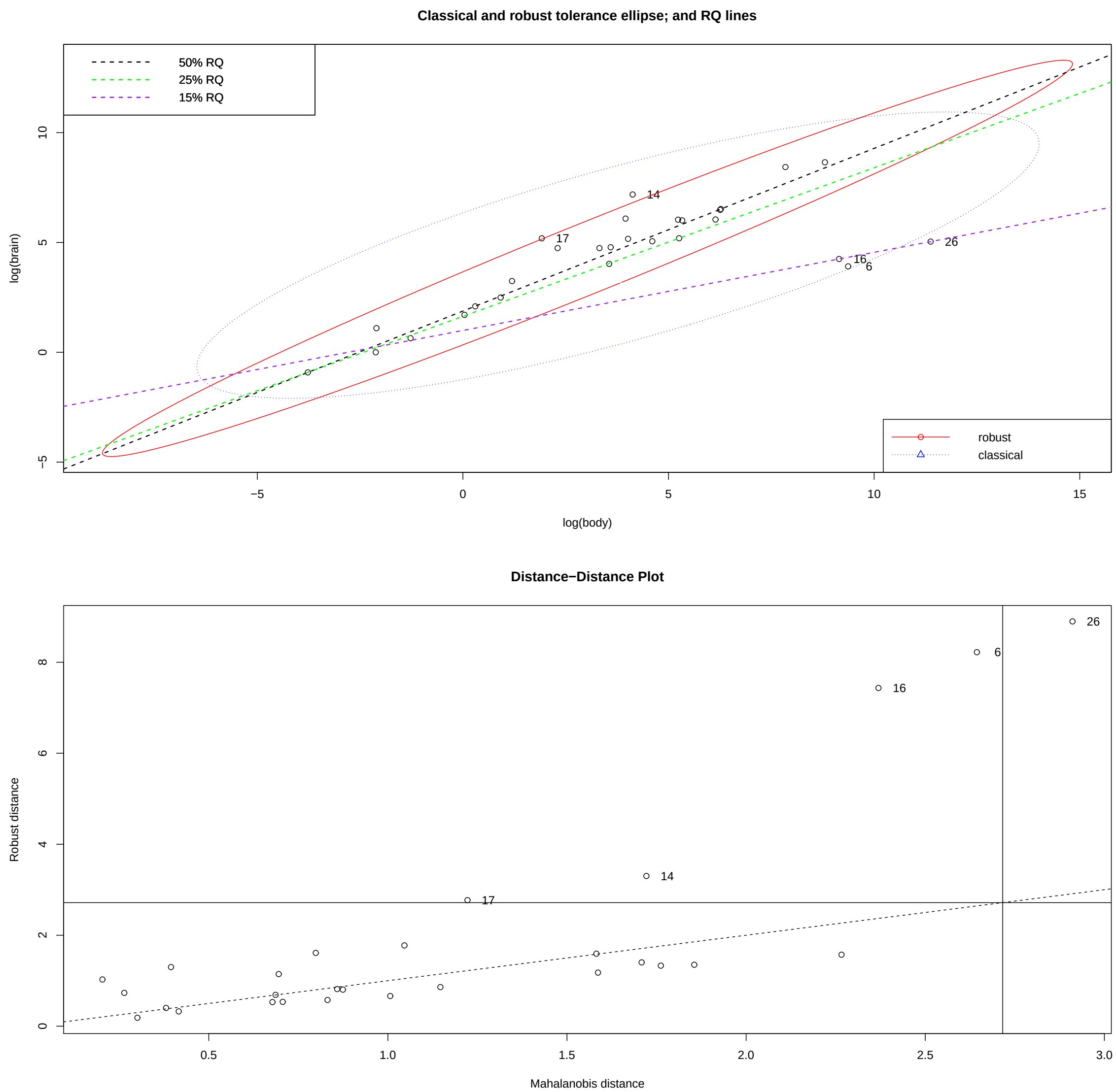


Figure 2.1: Tolerance Ellipses, *RQ* lines and *D-D* plots
Upper Panel: Tolerance ellipse of classical mean and covariance matrix, and that of robust location and scatter matrix based on $\sqrt{\chi^2_{2;0.975}}$; Lower: Robust distance versus classical Mahalanobis distance of the log of the brain of animals with cut-off values $\sqrt{\chi^2_{2;0.975}}$

2.6 Elemental Sets and Elemental Regression

An elemental set (ES) is a subset of the data containing the minimum number of observations $(p+1)$, such that the parameters can be estimated in the model (Smyth & Hawkins 2000). The use of ES methods has been proposed in the literature. These include the elemental set algorithms, such as regression depth and the repeated median (Siegel 1982), where some estimators are found by searching all $C(n,p+1) = \begin{pmatrix} n \\ p+1 \end{pmatrix}$ ESs. The number of ESs tends to be very large in practice. The other applications include detection multiple case data aberrations and the coefficient of determination (*COD*) and Studentized residuals in the *QR* scenario (see Ranganai et al. 2014, Ranganai 2016*b*,*a*). The accuracy of *ES* approximations for regression were investigated (see Rousseeuw 1984, Hawkins et al. 1984, Hawkins & Olive 2002, 1999, Hawkins 1993) and showed excellent approximations to the least median squares (*LMS*), the least trimmed squares (*LTS*) and *LS* criteria. The method has become computationally feasible due to the advent of modern increases in computing power, hence the renewed interest by researchers in *ES*-based methods.

Consider the intercept $\beta_0 \neq 0$ and express $Equation$ (1.1) as the linear regression $\mathbf{Y} = \check{\mathbf{X}}'\check{\boldsymbol{\beta}} + \mathbf{e}$, where $\mathbf{Y}$ is an $n \times 1$ response vector, $\check{\mathbf{X}} = (\mathbf{1}, \mathbf{X})$ is an $n \times (p+1)$ predictor matrix including a column of 1s, $\mathbf{1}_n$ is a column of 1s, $\check{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}')'$ is a $(p+1) \times 1$ parameter vector (p is the number of predictors and $\boldsymbol{\beta}$ is a parameter vector excluding β_0), $\mathbf{e} \sim N_n(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$ is an $n \times 1$ vector of errors, $\mathbf{0}_n$ is a vector of zeros and $\mathbf{I}_n$ is an identity matrix. We let $\check{\mathbf{X}}_J$ be a $(p+1) \times (p+1)$ submatrix of $\check{\mathbf{X}}$ and $\mathbf{Y}_J$ be a $(p+1) \times 1$ subvector of $\mathbf{Y}$, then $(\check{\mathbf{X}}_J, \mathbf{Y}_J)$ is J^{th} *ES* (J denotes the $p+1$ dimensional *ES*). The optimal solution to the linear programming problem (*LP*) in $Equation$ 1.2 is a regression quantile (*RQ*) that corresponds to a specific *ES* of size $(p+1)$ (see Koenker & Basset 1978, Koenker 2005). The result of applying the *LS* to the *ES* results in the J^{th} elemental regression (*ER*) coefficient estimator $\hat{\check{\boldsymbol{\beta}}}_J = (\check{\mathbf{X}}_J'\check{\mathbf{X}}_J)^{-1}\check{\mathbf{X}}_J'\mathbf{Y}_J = \check{\mathbf{X}}_J^{-1}\mathbf{Y}_J$, where $\check{\mathbf{X}}_J^{-1}$ is the inverse of the matrix $\check{\mathbf{X}}_J$ ($\check{\mathbf{X}}_J$ is non-singular). For the i^{th} elemental predicted value, $\hat{y}_{iJ} = \check{\mathbf{x}}_i'\hat{\check{\boldsymbol{\beta}}}_J$, the i^{th} elemental predicted residual (*EPR*) is given by $\varepsilon_{iJ} = y_i - \check{\mathbf{x}}_i'\hat{\check{\boldsymbol{\beta}}}_J, for, i \notin J$, where $\hat{\check{\boldsymbol{\beta}}}_J$ is estimated vector of coefficient. According to Ranganai (2016*b*), the *EPR*s and their corresponding sum of squares are related to the *LS* residuals $\varepsilon_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$, for $1 \leq i \leq n$, and the *LS SSE* via an elemental regression weight (*ERW*) (see Rousseeuw 1984).

2.7 Concluding Remarks

In Section 2.1, we reviewed the literature on variable selection procedures, such as the significance criterion, the change-in-estimate criterion, the forward stagewise procedure, stepwise forward selection procedure, the stepwise backward selection procedure and regularization techniques. We also discussed the criteria for model performance, such as *MAE*, *MAD* and *MBE* in Subsection 2.1.1. Cross validation was reviewed in Subsection 2.1.2. We further reviewed the literature on the *LS* and *LAD* variable selection and regularization procedures in section 2.2. We discussed penalized *LS* variable selection procedures in Subsection 2.2.1 and the *LAD* and related robust procedures in Subsection 2.2.2. Regularization techniques include the James–Stein estimator, *LS-RIDGE*, non-negative garotte, *LS-LASSO*, *SCAD*, *LS-E-NET*, *ALASSO* and the *LS-post-LASSO*, just to name a few. We further reviewed the literature on robustness and *QR* variable selection in Section 2.3, where in Subsection 2.3.1, the breakdown point and *QR* were reviewed, with robust criteria (*QR*) reviewed in Subsection 2.3.2, and lastly, the overview of *QR* computational aspects in Subsection 2.3.3. *QR* variable selection procedures include the *SCAD* penalized *QR*, the adaptive *LASSO* penalized *QR*, *ALASSO* penalized *QR* and *CQR-LASSO* procedures, just to mention a few. We reviewed collinearity and collinearity influential point diagnostics in Section 2.4. We further reviewed robust distances as the basis for our suggested weights in Section 2.5. These distances include the *MD*, the *MVE* based distance and the *MCD* based distance. We concluded this chapter by reviewing *ES* and *ER* showing their relationship with *QR* and diagnostics in Section 2.6.

Chapter 3

Variable Selection and Regularization Procedures

Although subset selection has been a topical issue in the 19^{th} century, it tends to be unstable and unsuitable in a high-dimensional data setting (Breiman 1995). Therefore, variable selection and regularization procedures have gained more traction in recent years. Extensive literature on variable selection and regularization proliferated towards the 20^{th} century. In this chapter, we discuss and review some of these variable selection and penalized procedures. The penalized procedures discussed and explored in this chapter serve as the foundation for our proposed regularized and adaptive-regularized *QR* procedures (see Chapter 4).

The rest of the chapter is organized as presented next. Section 3.1 discusses *LS* regularization procedures, with Subsection 3.1.1 reviewing ridge regression (*LS-RIDGE*), Subsection 3.1.2 discussing the *LS-LASSO* procedure and Subsection 3.1.3 discussing the elastic net (LS-E-NET). Section 3.2 reviews the adaptive regularization techniques in the *LS* scenario namely, the adaptive *LASSO* (*LS-ALASSO*) and adaptive LS-E-NET (LS-AE-NET). Sections 3.3 and 3.4 discuss the *LARS* and other regularization procedures, respectively. Lastly, Section 3.5 gives concluding remarks.

3.1 Regularization Procedures in the Least Squares Scenario

Regularization procedures, such as *LS-RIDGE* (see Hoerl & Kennard 1970, Hoerl et al. 1975), *LS-LASSO* (Tibshirani 1996) and *LS-E-NET* (Zou & Hastie 2005) tend to be fairly stable and produce lower prediction errors than subset selection techniques. In the literature, these regularization procedures proffer solutions to the limitations of subset selection procedures. The variable selection and regularization procedures in this section play an important role in our proposed regularized QR procedures (see Chapter 4).

3.1.1 The Ridge Regression

The intention of Hoerl's ridge regression (*LS-RIDGE*) was to overcome collinearity challenges (see Hoerl & Kennard 1970, Hoerl et al. 1975). In the presence of collinearity, the assumption of independence in linear regression is normally violated (Kibria 2003). The *LS* suffers from collinearity challenges, especially in prediction accuracy and interpretation. Although *LS* has low bias and large variance, the prediction accuracy can be improved by some penalized techniques, with *LS-RIDGE* regression suggested in the literature as one of the solutions to the problem of collinearity (collinearity adverse effects). This shrinkage technique introduces a trade-off between bias increase and variance reduction. Ridge regression, though more stable than subset selection, does not set any coefficients to 0 hence does not give easily interpretable results.

The procedure is determined by appending a positive ridge parameter (λ) in the range of $0 <$

$\lambda < 1$ to the design matrix's diagonal entries. Many ridge parameter estimators of λ have been suggested in the literature. This λ minimizes the variance (*MSE* and prediction sum of square (*PRESS*)). The ridge trace is a graphical method used to estimate λ, although McDonald (1980) argues that the procedure is primarily subjective. Hoerl & Kennard (1970) suggested $\lambda = \hat{\sigma}^2/\hat{\boldsymbol{\beta}}^{*2}_{max}$ while Hoerl et al. (1975) suggested $\lambda = p\hat{\sigma}^2/\Sigma_{i=1}^{p}\hat{\beta}_i^{*2} = p\hat{\sigma}^2/\hat{\boldsymbol{\beta}}^{*'}\hat{\boldsymbol{\beta}}^{*}$. There are other ridge parameter estimators (see Lawless & Wang 1976, Hocking et al. 1976, Kibria 2003, Khalaf & Shukur 2005, Alkhamisi et al. 2006, Muniz & Kibria 2009).

LS-RIDGE regression offers the solution to *LS* challenges when the predictor variables are non-orthogonal (Conniffe & Stone 1973). If the matrix $\boldsymbol{X'X}$ is ill conditioned, then the squared distance $E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is large for small eigenvalues. The resultant unstable and inflated coefficients may have incorrect signs.

The linear regression $Equation$ (1.1), with a normally distributed random error term ($\varepsilon_i \sim N(0,1)$), is considered. We consider the penalized *LS* regression with ℓ_2-penalty called *LS-RIDGE* (see Hoerl & Kennard 1970) regression given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{R} = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \lambda\Sigma_{j=1}^{p}\beta_j^2, j \in [1:p],\ i \in [1:n], \tag{3.1}$$

where λ is a positive ridge parameter in the range $0 < \lambda < 1$, β_j is the j^{th} parameter, the second term is the penalty term and other symbols are as defined in $Equation$ (1.1). The optimal ridge parameter is found using the ridge trace.

The *LS-RIDGE* regression estimator is the earliest remedy suggested to deal with collinearity. However, bias and instability are drawbacks that stem from its dependence on λ (Muniz & Kibria 2009). As $\lambda \to \infty$, $\boldsymbol{\beta}^R \to \mathbf{0}$ achieving stability, but the $\boldsymbol{\beta}^R$ estimator becomes biased. On the other hand, as $\lambda \to 0$, $\boldsymbol{\beta}^R \to \boldsymbol{\beta}_{LS}$, this results in unstable but unbiased *LS* parameter estimates. The question to be answered in the *LS-RIDGE* regression is: which value of λ produces the optimal estimator. The optimal value of λ (ridge parameter or bias parameter) is when the system stabilizes with orthogonal characteristics, and the problem of incorrect signs and inflated estimated coefficients variances is solved.

3.1.2 Least Absolute Shrinkage and Selection Operator

Tibshirani, in 1996, proposed the *LS-LASSO*, which retained the good characteristics of subset selection and *LS-RIDGE*. In recent years, *LS-LASSO* has become more popular because its ability to achieve a trade-off between variance and bias. The *LS* estimate obtained by minimizing the residual sum of squared errors (SSE) often has a large variance, hence unsatisfactory (Tibshirani 1996). He argued that the fewer variables exhibiting the strongest effects are determined by the shrinkage method (*LASSO*) than by the *LS-RIDGE* that selects almost all predictors. Models by subset selection are interpretable but extremely variable, resulting in different models from small changes in the data (Tibshirani 1996).

The non-negative garrote (Breiman 1995) directly influenced the development of *LS-LASSO*. Breiman's idea was the minimization of $\Sigma_{i=1}^{n}(y_i - \Sigma_j c_j \boldsymbol{x}_{ij}' \boldsymbol{\beta})^2$ with respect to $\boldsymbol{c} = \{c_j\}$, subject to $c_j \geq 0$ and $\Sigma_{j=1}^{p} c_j \leq \lambda$, where $\boldsymbol{\beta}$ is the usual parameter vector. When $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ in *Equation* (3.1) with standardized $\boldsymbol{x}_i$s, the following conditions are satisfied:

(i) $\Sigma \boldsymbol{x_i}/n = 0$ and

(ii) $\Sigma \boldsymbol{x_i}^2/n = 1$.

The *LS-LASSO* estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}^{L} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + n\lambda \Sigma_{j=1}^{p} |\beta_j|, \; j \in [1:p], \; i \in [1:n], \tag{3.2}$$

where $\lambda \geq 0$ denotes the tuning parameter, which controls the amount of shrinkage, and other terms are defined in previous sections. This regularization procedure has the advantage that the full rank property in the design matrix $\boldsymbol{X}$ is not a requirement.

3.1.3 Elastic Net Regression

In the literature, the elastic net (*LS-E-NET*) procedure has been suggested as a compromise between the *LS-RIDGE* and the *LS-LASSO* procedures with accrued properties. The *LS-E-NET* procedure exploits the advantages of *LS-RIDGE* and *LS-LASSO* regression properties. According to Zou & Hastie (2005), *LS-E-NET* outperforms the *LS-LASSO* in some situations although both have the same oracle properties. The *LS-E-NET* has an advantage of grouping effect on strongly correlated predictors over the *LS-LASSO*, especially when p is much greater than n. The *LS-E-NET* procedure has both the ℓ_1 and ℓ_2 penalties (Zou & Hastie 2005) and is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{EN} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \alpha \lambda_n \Sigma_{j=1}^{p} |\beta_j| + (1-\alpha) \lambda_n \Sigma_{j=1}^{p} \beta_j^2, \; j \in [1:p], \; i \in [1:n], \tag{3.3}$$

where $0 < \alpha < 1$ is a mixing parameter and other terms are as defined in the previous sections. The *LS-E-NET* procedure translates into the *LS-RIDGE* and *LS-LASSO* procedures when $\alpha = 0$ and $\alpha = 1$, respectively. The *LS-LASSO* (ℓ_1)-penalty term encourages the coefficient profiles β_j

to locally level. The application of the *LS-E-NET*-penalty, a strictly convex penalty for $\lambda_n > 0$, yields sparse solutions. The *LS-E-NET* regression's edge over the *LS-LASSO* is better prediction, and the procedure also adequately selects predictors in high-dimensional data $(p > n)$.

3.2 Adaptive Regularization Procedures in the Least Squares Scenario

Non-adaptive penalized procedures can perform poorly in the presence of highly correlated variables in the predictor space (collinearity) (Zou & Zhang 2009). This collinearity problem is often encountered in high-dimensional data, though in this thesis, we focus on high leverage induced or hidden collinearity. Due to the shortcomings of these non-adaptive penalized procedures (*LS-RIDGE*, *LS-LASSO* and *LS-E-NET*), adaptive regularization procedures were suggested in the literature. Adaptive regularization procedures were proffered as solutions to these shotcomings, and the suggested procedures are namely, adaptive *LASSO* (*ALASSO*) (Zou 2006), adaptive *RIDGE* (*ARIDGE*) (Frommlet & Nuel 2016) and adaptive *E-NET* (*AE-NET*) (Zou & Zhang 2009). The *ALASSO* is superior to the *LASSO*, since the latter penalizes coefficient estimates equally. The *ALASSO* also enjoys oracle properties. Just like the *LS-RIDGE*, *LS-LASSO* and *LS-E-NET* procedures, adaptive procedures perform poorly in the presence of collinearity influential points (high leverage points that either induce or reduce collinearity). *SCAD* (Fan & Li 2001) and *MPC* (Zhang et al. 2010) are suggested as solutions for the *LS* case in the literature.

3.2.1 Adaptive LASSO Regression

Another version of *LASSO* called adaptive *LASSO* (*ALASSO*) proposed by Zou (2006) deal with certain situations where *LASSO* is inconsistent for variable selection. *ALASSO* uses adaptive weights for penalizing different parameter estimates in the ℓ_1-penalty. *ALASSO* enjoys the oracle properties and is near minimax optimal. To understand adaptive *LASSO*, one needs to explore the oracle properties, as explained by Fan & Li (2001).

We start by presenting the penalty term for bridge regression, $\Sigma_{j=1}|\hat{\beta}_j|^q$, where $q \geq 1$ (Fu 1998). The adaptive bridge penalty becomes $\Sigma_{j=1}\omega_j|\hat{\beta}_j|^q$, where $\omega_j = 1/|\hat{\beta}|^\gamma$ is the adaptive weight. The *ALASSO* and adaptive *RIDGE* (*ARIDGE*) penalties are special cases of the bridge penalty when $q = 1$ and $q = 2$, respectively (see Zou 2006). In the presence of collinearity and constructing the weights ω_j, Zou (2006) suggested the use of *RIDGE* $\hat{\boldsymbol{\beta}}$ ($\hat{\boldsymbol{\beta}}^R$) as a suitable replacement because it is superior in stability than its *LS* counterpart. Since collinearity severely degrade the performance of the *LASSO* procedure and results in unstable solution paths, Zou & Zhang (2009) proposed an adaptive *E-NET* (*AE-NET*), which is superior to *LASSO*, to deal with the instability due to high-dimensional data (see also Zou & Hastie 2005).

Fan & Li (2001) argue that since *LS-LASSO* uses the same tuning parameter for all parameter estimation, the resultant solution suffers from appreciable bias. Zou (2006) further proposed the modification of this *LS-LASSO* procedure to the *LS* adaptive *LASSO* version *LS-LASSO* given by

$$\hat{\boldsymbol{\beta}}^{AL} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^n (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + n \Sigma_{j=1}^p \lambda_j |\beta_j|, \ j \in [1:p], \ i \in [1:n], \tag{3.4}$$

where $\lambda_j = \omega_j \lambda$ is the adaptive weight. Here, the tuning parameter is different for each regression coefficient, as a result $\hat{\boldsymbol{\beta}}^{AL}$ effectively determines sparse solutions than $\hat{\boldsymbol{\beta}}$. A very important characteristic of the *LS-LASSO* and *LS-ALASSO* estimators is that they can do variable selection and shrinkage simultaneously. They improve prediction accuracy through variance-bias trade-offs. These shrinkage methods, with varying degrees of compliance with oracle properties, are oracle procedures. The oracle properties include identification of the right subset model and optimal estimation rate.

3.2.2 Adaptive Elastic Net Regression

Some shortcomings of the *LS-E-NET* procedure were addressed by its extension to its adaptive version. This variable selection and regularization procedure is known as the adaptive *E-NET* (*AE-NET*) procedure, and it uses $\omega_j^* = |\hat{\beta}_j^{EN}|^{-\gamma}$ as penalty weights, where $\hat{\beta}_j^{EN}$ is the j^{th} *E-NET* estimator (Zou & Zhang 2009). The *AE-NET* penalty function has characteristics inherited from the *ALASSO* and *ARIDGE* penalized procedures. The *AE-NET* estimator of $\boldsymbol{\beta}$ is then defined by

$$\hat{\boldsymbol{\beta}}^{AEN} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \alpha \Sigma_{j=1}^{p} \lambda \omega_j^* |\beta_j| + (1-\alpha) \Sigma_{j=1}^{p} \lambda \omega_j^* \beta_j^2, \; j \in [1:p], \; i \in [1:n], \tag{3.5}$$

where the terms are defined as in previous sections. If $\alpha = 1$ in *Equation* (3.6), then the *AE-NET* reduces to the *ALASSO* and likewise, if $\alpha = 0$, it reduces to the *ARIDGE*. According to Zou & Hastie (2005), the *AE-NET* in the orthogonal design scenario reduces to *ALASSO* regardless, a desirable situation to achieve optimal minimax. The ℓ_2-penalty ensures that the procedure deals with collinearity. However, the *AE-NET* procedure is not robust in the presence of outliers and high leverage points.

3.3 Other Variable Selection and Regularization Procedures

In this section, we summarize a few other variable selection and regularization procedures, and the list is inexhaustive. We aim to expose the reader to many regularization procedures.

Contrasting these procedures is the all-exhaustive procedure called all subsets regression, which considers the size of all variable subsets (Furnival & Wilson 1974). However, this procedure has huge inferential biases. *LASSO* has an affinity for choosing only one predictor from a group of strongly correlated predictors. One might prefer to select the whole group instead, and *LASSO* falls short. In pursuit of small differences in successive coefficients, Tibshirani et al. (2005) proposed a procedure for a sequence of predictors, namely fused *LASSO*. Fused *LASSO* procedure makes use of a mixture of an ℓ_1-penalty on both the coefficients and the difference between adjacent coefficients as follows

$$\hat{\boldsymbol{\beta}}^{FL} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda_2 \Sigma_{j=2}^{p} |\beta_j - \beta_{j-1}|, j \in [1:p],\ i \in [1:n], \tag{3.6}$$

where β_{j-1} is the $(j-1)^{th}$ parameter and other terms are as defined in previous sections.

In grouped *LASSO* (Yuan & Lin 2006), a multi-level factor represented by a set of dummy variables is selected or dropped. Let β_j for groups $j \in [1:J]$ be a sub-vector of the global $\boldsymbol{\beta}$ ($\boldsymbol{\beta}$ from ungrouped variables). We present the grouped *LASSO* as a minimization problem given by

$$\hat{\boldsymbol{\beta}}^{GL} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda \Sigma_{j=1}^{p} |\beta_j|_{\kappa_j},\ j \in [1:p],\ i \in [1:n], \tag{3.7}$$

where κ_j denotes the j^{th} elliptical norm of $\boldsymbol{\beta}$. The group *LASSO* deals with known predictor groups, whereas *E-NET* deals with unknown predictor groups. This is possible after letting β_j

for groups $j \in [1:J]$ be a sub-vector of the global $\boldsymbol{\beta}$. The minimization problem reduces to a *LASSO* procedure if $j = 1$. The group *LASSO*-penalty promotes sparsity in the groups, as compared to the variables involved. According to Lin & Zhang (2006), predictor groups resemble sets of basis functions for smoothing splines, resulting in the component selection and smoothing operator (*COSSO*).

The least angle regression (*LARS*) algorithm (see Efron et al. 2004) and the coordinate descent algorithms for *LASSO* (see Wu & Lange 2008) provided efficient ways of solving the *LS-LASSO*, as compared to the original quadratic program solver. The advantages of the coordinate descent algorithms are that they are simple, fast and can exploit the assumed sparsity of the model.

LARS is a model selection algorithm for a data set with many predictor variables (Efron et al. 2004). Procedures such as all subsets (AS), forward selection (FS) and backward elimination (BE) are some of the procedures in the same family as *LARS*. A modification of *LARS* and its applications to various variable selection methods bring attractive properties, such as computational efficiency. These methods include *LASSO* and forward stagewise linear regression, etc. *LARS* selects predictors with largest absolute correlation with response variable y (Weisberg 2014).

Group *LARS* counter the shortcomings of group-*LASSO*, which could not handle piecewise linear solution paths (Yuan & Lin 2006). The Group *LARS* procedure uses the average squared correlation between the variable group and the current residual in place of the correlation criterion. Other extensions to group *LARS* procedure include the modified group *LARS*, where the average absolute correlation replaces the average squared correlation (Park & Hastie 2006), composite absolute penalties (CAP) approach, a group *LASSO* alike procedure with penalty ℓ_{γ_j}-norm replacing

ℓ_2-norm (Zhao et al. 2009) with a minimization problem given by

$$\hat{\boldsymbol{\beta}}^{CAP} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda \Sigma_{j=1}^{J} (\beta_{\gamma_j})^{\gamma_0},\ j \in [1:p],\ i \in [1:n], \tag{3.8}$$

where γ_j is the *CAP* penalty, $\gamma_j > 1$ for group variable selection.

Time series data are sequential data that vary over time. In contrast to fused-*LASSO*, which deals with sequentially ordered predictor random variables, other procedures deal with naturally ordered response variables (time-ordered data). To show the shortcomings of fused *LASSO* in time ordered data, we deliberate on linear regression with multiple responses as

$$y(t_i) = \boldsymbol{x}_i' \boldsymbol{\beta}(t_i) + \varepsilon(t_i), \forall\ i \in [1:n], \tag{3.9}$$

where $\boldsymbol{x}_i' \in R^n$ is the i^{th} row of the design matrix $\boldsymbol{X}$, $t_1, ..., t_n$ are time points, $y(t_i) \in R^n$ is a time dependent response variable, $\varepsilon(t_i) \in R^n$ is a time-dependent error term, and $\beta(t_i) \in R^p$ is a time-dependent parameter. The application of the $|\beta_j(t_i) - \beta_j(t_{i-1})|$-penalty on fused *LASSO* has challenges of concurrently fitting a model with too many parameters (np, say) for a large sample. Meier & Bühlmann (2008) claim that consecutive time points are more correlated than distant time points. Meier & Bühlmann (2008) then proposed smoothed *LASSO* as a solution to the challenge. Information from different time points is integrated by weight $w(t_s, t_i)$, and the smoothed *LASSO* criterion is then given by

$$\hat{\boldsymbol{\beta}}^{SL} = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{s=1}^{n} w(t_s, t_i)(y(t_s) - \boldsymbol{x}\boldsymbol{\beta}(t_i))^2 + \lambda \Sigma_{j=1}^{p} |\beta_j(t_i)|,\ for\ j \in [1:p],\ i \in [1:n], \tag{3.10}$$

where $\Sigma_{s=1}^{n} w(t_s, t_i) = 1$ is a necessary condition, and the weight $w(t_s, t_i)$ has higher values and guarantees more information on estimates in the neighborhood of time points (t_i). The Turlach et al. (2005) procedure chooses common subsets for the predictor variables for estimating multiple response variables. The smoothed *LASSO* procedure, an extension of *LASSO* is given by

$$\hat{\boldsymbol{\beta}} = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{s=1}^{n}(y(t_s)-\boldsymbol{x}_i\boldsymbol{\beta}(t_i))^2+\lambda\Sigma_{j=1}^{p}max_{i\in[1;n]}|\beta_j(t_i)|,\ for\ j\in[1:p],\ i\in[1:n], \quad (3.11)$$

where $|\beta_j(t_i)|$ is a penalty parameter at the time point t_i.

3.4 Concluding Remarks

This chapter exposed the reader to different variable selection and regularization procedures. These penalized procedures include *LS-RIDGE*, *LS-LASSO*, *LS-E-NET*, *ALASSO*, *AE-NET*, *LARS* and a summary of a few other procedures. The *RIDGE*, *LASSO*, *E-NET* penalties and their adaptive variants discussed in this chapter are used to formulate our proposed penalized *WQR*, adaptive penalized *QR* and adaptive penalized *WQR* variable selection and regularization procedures (see Chapter 4).

In Section 3.1, we gave an overview of regularization procedures in the *LS* scenario. The regularization procedures discussed in detail are the *LS-RIDGE* in Subsection 3.1.1, the *LS-LASSO* in Subsection 3.1.2 and the *LS-E-NET* in Subsection 3.1.3. In Section 3.2, we briefly elaborated on the adaptive penalized procedures namely, the *ALASSO* and *AE-NET*, which we further discussed in Subsections 3.2.1 and 3.2.2, respectively. Finally, we further reviewed the *LARS* method in Section 3.3, summarizing other variable selection and penalized procedures in Section 3.4.

The extensions of *LS-LASSO* have huge advantages over their traditional stepwise deletion and subset selection counterparts. The *LS-LASSO* procedure outperforms its *LS-RIDGE* counterpart, as it sets some coefficients to zero, as compared to *LS-RIDGE*, where almost all coefficients are not zeros. If the predictor variables are too many, the choice of significant predictors is a challenge that remains in the *LS-RIDGE* procedure. The *LS-LASSO* shows predictors with moderate-to-large effects are more visible than in the *LS-RIDGE* procedure. The *LS-E-NET* is a hybrid procedure that inherits the good properties of the *LS-LASSO* and the *LS-RIDGE* procedures. *LARS* improves the calculation of the *LS-LASSO* estimates. The original quadratic program solution method was computationally intensive until its modification by the *LARS* algorithm in calculating all possible *LS-LASSO* estimates. Adaptive procedures are better at dealing with collinearity than the *LS-LASSO* and *LS-E-NET* procedures. Adaptive penalized procedures perform better in the presence of highly correlated variables in the predictor space (collinearity) than the non-adaptive ones.

Chapter 4

Variable Selection and Regularization in Quantile Regression

4.1 Introduction

Variable selection and regularization are very important in data analysis in the *LS* and *QR* scenarios. *QR* estimates the conditional median $Q_{Y/X}(0.5)$ (the *LAD* estimator) among other conditional quantiles (Koenker & Bassett 1978). In this chapter, we discuss existing variable selection and regularization procedures in a *QR* setting. We present our proposed unweighted, weighted and adaptive regularization *QR* procedures. Regularization methods discussed in Chapter 3 are extended to their penalized weighted *QR* (WQR), adaptive penalized *QR* and adaptive penalized WQR counterparts to achieve robustness and desirable variable selection properties.

In the literature, the *LASSO* penalty fails to perform adequately in the presence of high leverage points, and thus collinearity influential points, leading to the suggestion of the smoothly clipped absolute deviation ($SCAD$) (Fan & Li 2001) and minimax concave penalty (MCP) (Zhang et al.

2010) as better options. Furthermore, weight-based procedures were suggested in the literature as remedies for high leverage points and outlier influences in regression. The doubly adaptive penalized procedure, which satisfies both sparseness and robustness properties (Karunamuni et al. 2019) and the *WQR* procedure, which is robust in the presence of high leverage points (Salibián-Barrera & Wei 2008), are suggested in the literature for the *LS* and *QR* scenarios, respectively. In this thesis, we build upon the latter idea to suggest our *WQR* variable selection and regularization procedure, which is robust to data aberrations in both the predictor and response spaces. In this thesis, we suggest penalized *WQR* procedures based on *MCD* based-weights with *RIDGE* (Hoerl & Kennard 1970), *LASSO* (Tibshirani 1996) and *E-NET* (Zou & Hastie 2005) penalties namely, *WQR-RIDGE*, *WQR-LASSO* and *WQR-E-NET* procedures. The *MCD* based weights down-weigh extreme leverage points (see Chapter 2 and Subsection 4.2.1 for more details).

In the literature, *ALASSO* procedures have been suggested as alternatives to the *LASSO* procedure, and these *ALASSO* procedures enjoy the oracle properties that guarantee optimal performance in large samples and high dimensional data (Fan & Peng 2004), not withstanding the computational advantage of the *LASSO* owing to the efficient path algorithm (Zou 2006). In the *LS* scenario, Frommlet & Nuel (2016) suggested the *ARIDGE* procedure for variable selection with interesting formulations of the adaptive weights, and Wu & Liu (2009) suggested an *ALASSO* penalized *QR* (*QR-ALASSO*) procedure for variable selection. Zou and Zhang's *AE-NET* procedure proposed in 2009 (Zou & Zhang 2009), adaptively inherits some good properties from both the *ALASSO* and *ARIDGE* penalty-based procedures. In this thesis, we further extend our suggested penalized *WQR* procedures (*MCD* based ones) to their adaptive versions namely, the *WQR-ALASSO* and *WQR-AE-NET* procedures. The *ALASSO* and *AE-NET* penalties are based on our proposed adaptive weights (*WQR-RIDGE* based adaptive weights) (see Subsection 4.2.1). Just as in the

WLAD scenario (Arslan 2012), the penalized *WQR* procedures are robust in the X-space (predictor space). *LS* $\boldsymbol{\beta}$ estimator used in determining adaptive weights is unsuitable in the presence of collinearity, prompting Zou (2006) to suggest the use of the *RIDGE* $\boldsymbol{\beta}$ as an alternative due to its superiority in the presence of collinearity. Although the use of these $\boldsymbol{\beta}$ estimators and corresponding adaptive weights for the symmetrical distribution may be applicable to the ℓ_1-estimator (*RQ* at $\tau = 0.50$), it may not be applicable at extreme quantile levels in the presence of high leverage (and collinearity influential) points due to the presence of these atypical observations, resulting in *RQ* planes frequently crossing (unequal slope parameter estimates) (see Zhao 2000, Ranganai 2007). Therefore, we use the *WQR-RIDGE* estimator-based adaptive weights. To our knowledge, the *WQR-RIDGE* based adaptive weights have not been applied in a *QR* variable selection scenario. These suggested adaptive weights have the advantage of being robust in the presence of extreme points and adjustable to particular distribution levels, i.e., t_1, t_2 distributions, *etc* and are applicable at all τ *RQ* levels. The suggested adaptive procedures are, thus, the *QR-ALASSO*, *QR-AE-NET*, *WQR-ALASSO* and *WQR-AE-NET* procedures.

The rest of this chapter is summarized as presented next. Section 4.2 discusses penalized *WQR* with Subsections 4.2.1, 4.2.2, 4.2.3 and 4.2.4 details the choice of *MCD* based weights, *RIDGE* penalized *WQR*, *LASSO* penalized *WQR* and *E-NET* penalized *WQR*, respectively. Section 4.3 introduces the adaptive regularized *QR*, with Subsections 4.3.1, 4.3.2 and 4.3.3 discussing adaptive weights choice, adaptive *LASSO* penalized *WQR* and adaptive *E-NET* penalized *WQR*, respectively. In Section 4.4 and Subsection 4.4.1, we discuss asymptotics for *QR* and asymptotics for regularized *QR*, respectively. We conclude the chapter with a summary and concluding remarks in Section 4.6.

4.2 Penalized Weighted Quantile Regression

In this section, we discuss existing variable selection and regularization procedures (unweighted ones) and suggest new weighted ones in a *QR* setting. The new regularized *WQR* methods are robust in the presence of high leverage points and collinearity influential points. We discuss penalized *QR* and suggest penalized *WQR* procedures with *RIDGE* (Hoerl & Kennard 1970), *LASSO* (Tibshirani 1996) and *E-NET* (Zou & Hastie 2005) penalties. The robustness in the X-space is inherited from properties of *WLAD* (Arslan 2012).

The proposed regularized *WQR* procedures are based on the *MCD* robust weights, ω_i described in Subsection 4.2.1. Although the unweighted penalized *QR* procedures are robust in the Y-space, they are not robust in the predictor space. In this section, we extend these unweighted penalized *QR* procedures to penalized *WQR* counterparts, which are robust in the presence of high leverage points and collinearity influential points (high leverage points that induce or reduce collinearity). In the *LS* case, robustness in the X-space is achieved by appropriately chosen weights (see also Hubert & Rousseeuw 1997, Giloni et al. 2006). Robustness in the *QR* setting is achieved in a similar fashion via *WQR*. Thus, the proposed regularization procedures in the *WQR* scenario are namely, *WQR* with ridge penalty (*WQR-RIDGE*), *WQR* with *LASSO* penalty (*WQR-LASSO*) and *WQR* with *E-NET* penalty (*WQR-E-NET*).

4.2.1 Choice of Robust Weights for Weighted Quantile Regression

The hat matrix $h_{ii} = \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i$ is a measure of leverage (see Chatterjee & Hadi 1988*a*). In the literature, h_{ii} has been used as a standard tool for generating weights in weighted *LS* (*WLS*) estimation. The weighting strategy is both mathematically and practically tractable. Since the estimator has a $1/n$ breakdown point, a single high leverage point has the potential to completely dominate the resulting estimates. In addition, such weights may suffer from the masking and swamping effects associated with the *LS*. Contamination in both the predictor and response variables results in a breakdown point of the *LAD* estimator (hence *QR*) of $1/n$ (see Rousseeuw & Leroy 1987), as in the *LS* scenario. Hubert & Rousseeuw (1997) proposed the *WLAD* estimator to circumvent the undesirable effects of both outliers and high leverage points. The weights are given by

$$\omega_i = min\left(1, \frac{p}{RD(x_i)^2}\right), \quad i \in [1:n], \tag{4.1}$$

where the estimator ω_i is based on the computationally intensive high breakdown *MCD* method (Rousseeuw 1985). To avoid the huge computational load associated with the *MCD* based weights Giloni et al. (2006) proposed $v_j = \sqrt{min_j(h_j/h_{c_i})}$, where $h_{c_i} = \boldsymbol{x}_i(\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{x}_i'$ is the i^{th} leverage point relative to the clean subset $\boldsymbol{X}_c$ ($\boldsymbol{X}$ without high leverage points). This thesis uses the robust *MCD* distance $RD(x_j)$ because of an improvement in computer power (efficient algorithms) and its robustness in generalizing the Arslan (2012) *WLAD* concept to all regression quantiles (*RQ*s).

Based on the *MCD* based weighting construction in $Equation$ 4.1, we suggest a WQR estimator. The WQR estimator is given the minimization problem

$$\hat{\boldsymbol{\beta}}^W(\tau) = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^n \omega_i \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)|, \; i \in [1:n], \tag{4.2}$$

where the terms are defined as in $Equation$ (1.2) and $Equation$ (4.1). The *MCD* based weights down-weigh high leverage points, hence collinearity influential points, thereby reducing their influence on parameter estimates and achieving robustness. By using these robust weights, we propose regularized *WQR* procedures based on the *RIDGE*, *LASSO* and *E-NET* penalties namely, the *WQR-RIDGE*, *WQR-LASSO* and *WQR-E-NET*. More details are given in the next few subsections.

4.2.2 Weighted Quantile Regression with a Ridge Penalty

An extensive review of variable selection and regularization in the presence of collinearity has been discussed in Chapter 3. In the presence of outliers (heavy tailed distributions), *LS* procedures fall short due to large sample variances, and *QR* was suggested as an alternative. To mitigate the negative effects of collinearity influential points, Suhail et al. (2019) proposed robust quantile-based *RIDGE* and *RIDGE M*-estimators. They further suggested a new ridge-based QR estimator that chooses an appropriate quantile level to deal with severe collinearity and high error variances (see Suhail et al. 2020) .

Consider the extension of the *QR* procedure in *Equation* (1.2) (Koenker & Basset 1978) by adding the penalty term of the *RIDGE* regression *Equation* (3.1) (Hoerl & Kennard 1970). We have a *RIDGE* penalized *QR* denoted by *QR-RIDGE*. The *QR-RIDGE* procedure is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{R}(\tau) = argmin_{\boldsymbol{\beta}\in R^{p}} \Sigma_{i=1}^{n} \rho_{\tau} |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + \lambda \Sigma_{j=1}^{p} \beta_j^2,\ j \in [1:p],\ i \in [1:n], \tag{4.3}$$

where λ is a positive ridge tuning parameter in the range $\lambda \in (0,1)$, and other terms are as defined in Section 3.2. Many variations of λ have been proposed in the literature (see Hoerl & Kennard 1970, Hoerl et al. 1975, Lawless & Wang 1976, Hocking et al. 1976, Kibria 2003, Khalaf & Shukur 2005).

Based on the weights ω_i in *Equation* (4.1), we propose a *WQR* variable selection and regularization procedure with a *RIDGE*-penalty. The predictor space robustness property is achieved by the incorporated *MCD* weights. The *RIDGE* penalized *WQR* (*WQR-RIDGE*) estimator of $\boldsymbol{\beta}$ is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{WR}(\tau) = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^{n}\omega_i\rho_\tau|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)| + \lambda\Sigma_{j=1}^{p}\beta_j^2,\ j\in[1:p],\ i\in[1:n], \tag{4.4}$$

where the terms are already defined. The tuning parameter shrinks coefficients of predictor variables but not entirely to zero. The procedure is robust in the presence of collinearity influential points, high leverage points and thick tailed distributions.

4.2.3 Weighted Quantile Regression with LASSO Penalty

Consider the *QR* procedure with an ℓ_1-penalty (*LASSO*-penalty (Tibshirani 1996)). The *LASSO* penalized *QR* (*QR-LASSO*) procedure is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{L}(\tau) = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^{n}\rho_\tau|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)| + n\lambda\Sigma_{j=1}^{p}|\beta_j|,\ j\in[1:p],\ i\in[1:n], \tag{4.5}$$

where λ is the tuning parameter that shrinks some coefficients towards zero; the second term is the penalty term, and other terms are as defined in *Equation*s (1.1), (3.2) and (4.1). The *QR-LASSO* procedure may be superior to the *QR-RIDGE* procedure, since coefficients are not entirely shrunk to zero in the latter procedure.

In the literature, weights have been used in a *QR* scenario by Jiang et al. (2012) in a procedure called the weighted composite *QR* (*WCQR*), with *LASSO*-penalty (*WCQR-LASSO*) for nonlinear regression models with many parameters. Because the Jiang et al. (2012) weighting strategy does not down-weigh high leverage points, we suggest the *MCD*-based weights instead (see the application in Arslan 2012). In this section, we suggest a *LASSO*-regularized *WQR* (*WQR-LASSO*) procedure based on the weights ω_i (see *Equation* (4.1)). The *WQR-LASSO* procedure is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{WL}(\tau) = argmin_{\tilde{\beta} \in R^p} \Sigma_{i=1}^{n} \omega_i \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + n\lambda \Sigma_{j=1}^{p} |\beta_j|, \; j \in [1:p], \; i \in [1:n], \tag{4.6}$$

where the tuning parameter λ is as defined in *Equation* 4.5. In addition to being robust in the response space, i.e., heavy-tailed distribution (like *WCQR-LASSO*), *WQR-LASSO* is robust in the predictor space (high leverage points and collinearity influential points).

4.2.4 Weighted Quantile Regression with Elastic Net Penalty

In this section, we present the *E-NET* penalized *QR* procedure (*QR-E-NET*), a *QR* version of the Zou & Hastie (2005) *E-NET* procedure and our suggested unregularized and regularized versions of it. The *QR-E-NET* is best suitable for applications with unidentified groups of predictors (see Zou & Hastie 2005, Su & Wang 2021). The *E-NET* penalized *QR* (*QR-E-NET*) procedure given is by

$$\hat{\boldsymbol{\beta}}^{EN}(\tau) = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + \alpha\lambda \Sigma_{j=1}^{p} |\beta_j| + (1-\alpha)\lambda \Sigma_{j=1}^{p} \beta_j^2, \; j \in [1:p], \; i \in [1:n], \tag{4.7}$$

where $0 \leq \alpha \leq 1$ is the tuning parameter, and the other terms are as defined in previous sections. Special cases of *QR-E-NET* are *QR-RIDGE* ($\alpha = 0$) and *QR-LASSO* ($\alpha = 1$). Ordinarily, the

E-NET procedure performs better than its *RIDGE* and *LASSO* procedures in some situations (see also Zou & Hastie 2005, for the *LS* scenario). The procedure is adversely influenced by high leverage points hence collinearity influential points.

We propose *WQR* variable selection and regularization procedure denoted by *WQR-E-NET* as a remedy to collinearity influential points (see also Salibián-Barrera & Wei 2008). We apply the weighting scheme ω_i discussed in Section 4.1 to our new proposed method. This procedure is a generalization of the *WLAD* (Arslan 2012) procedure to the *QR* scenario. The *WQR-E-NET* procedure has both the *LASSO* and *RIDGE* penalties. The suggested *WQR-E-NET* estimator is a solution to the minimization problem

$$\hat{\boldsymbol{\beta}}^{WEN}(\tau) = argmin_{\tilde{\beta}\in R^p} \Sigma_{i=1}^{n} \omega_i \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + \alpha\lambda \Sigma_{j=1}^{p} |\beta_j| + (1-\alpha)\lambda\, \Sigma_{j=1}^{p} \beta_j^2,\ j\in[1:p],\ i\in[1:n], \tag{4.8}$$

where α is the mixing parameter. Other terms are defined in previous sections. *WQR-RIDGE* and *WQR-LASSO* are special cases when $\alpha = 0$ and $\alpha = 1$, respectively.

4.3 Adaptive Regularized Quantile Regression

In this section, we discuss existing and suggested adaptive regularized *QR* procedures. We propose penalized *QR* procedures with *ALASSO* and *AE-NET* penalties. The *ALASSO* and *AE-NET* penalties are based on the proposed adaptive weights (*WQR-RIDGE* based weights) (see Section 4.3.1). The penalized weighted *QR* procedures inherit X-space robustness property from the use of the *WLAD* procedure in the generalization to the *QR* scenario.

We suggest adaptive weights $\breve{\omega}_j$, based on the *WQR-RIDGE* (denoted by *WQRR* in adaptive weights formula) parameter estimates (see Section 4.2). The *WQR-RIDGE* parameter based weights have the advantage of being specific to a particular *RQ* in contrast to the global *LS-RIDGE* one. We use these suggested robust adaptive weights $\breve{\omega}_j$ to propose new adaptive variable selection and regularization procedures in the *QR* and *WQR* scenarios. The adaptive penalized *QR* procedures are adaptive *QR* with *LASSO* penalty (*QR-ALASSO*), adaptive *QR* with *E-NET* penalty (*QR-AE-NET*), adaptive *WQR* with *LASSO* penalty (*WQR-ALASSO*) and adaptive *WQR*, with an *E-NET* penalty (*WQR-AE-NET*). The collinearity adverse effects are remedied through the adaptive *QR* procedures, a property inherited from the *LS* counterparts. The adverse effects of collinearity influential points and high leverage points are also remedied by the proposed procedures.

4.3.1 Choice of Adaptive Weights for Regularized Quantile regression

The bridge penalty $\Sigma_{j=1}^{p}|\hat{\beta}_j|^q$ (Fu 1998) is a generalized penalty for most penalized procedures. For adaptive weights, ω_js, the adaptive penalty is given by $\Sigma_{j=1}^{p}\omega_j|\hat{\beta}_j|^q$. Special cases of the adaptive bridge penalty are the *ALASSO* ($q=1$) and *ARIDGE* ($q=2$) penalties. In the literature, *RIDGE* $\boldsymbol{\beta}$ (solution to *Equation* 3.1) is used instead of the *LS* $\boldsymbol{\beta}$ in the presence of collinearity (Zou 2006).

Consider the *MCD* based weights, ω_is and construct a robust weighted *RIDGE* regression (*WRR*) estimator

$$\hat{\boldsymbol{\beta}}^{WR} = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^{n}\omega_i(y_i-\boldsymbol{x}_i'\boldsymbol{\beta})^2+\lambda\Sigma_{j=1}^{p}\beta_j^2, j\in[1:p],\ i\in[1:n], \tag{4.9}$$

where $\lambda(\lambda>0)$ is a ridge parameter in the range $\lambda\in(0,1)$ and other terms are explained in *Equation* (3.1). The *WRR* procedure inherits its ability to deal with collinearity from the Hoerl

& Kennard (1970) *LS-RIDGE* regression. Although this procedure is robust in the presence of collinearity, it fails to optimally shrink coefficient estimates, which might not be feasible. We suggest $\hat{\beta}_j^{WR}$ based adaptive weights taking advantage of its robustness in the predictor space. The adaptive weights have the same construction as that of the Zou & Zhang (2009) *E-NET* adaptive weights. We present the $\hat{\beta}_j^{WR}$ based adaptive weights as

$$\tilde{\omega}_j = \left(\left|\hat{\beta}_j^{WR}\right| + 1/n\right)^{-1}, \tag{4.10}$$

where $\hat{\beta}_{WRR_j}$ (see *Equation* (4.9)) is j^{th} element of parameter vector estimate $\hat{\beta}_j^{WR}$ and $1/n$ is added to avoid dividing by a near zero value. These adaptive weights can be seen as a special case of Frommlet & Nuel (2016)'s adaptive weight $\left(\left|\hat{\beta}_j^{WR}\right|^{\gamma} + \delta^{\gamma}\right)^{(\vartheta-2)/\gamma}$ when $\vartheta = 1$, $\delta = 1/n$ and $\gamma = 1$. The adaptive weights are data-dependent and have the advantage of being robust in the presence of high leverage points. Equation (4.10) reduces to $\omega_j = \left(\left|\hat{\beta}_j^{R}\right| + 1/n\right)^{-1}$ when $\omega_i = 1$ (the unweighted *RR* scenario).

The robust weight, $\tilde{\omega}_j$ may be suitable for the symmetric distribution rather than asymmetric ones at extreme quantile levels in the presence of collinearity influential points (high leverage points). Rather than using the $\hat{\beta}_j^{WR}$ based adaptive weights, we propose a $\hat{\beta}_j^{WR}(\tau)$ (see *Equation* (4.4)) based one (*WQR-RIDGE* estimator of $\boldsymbol{\beta}$). The weights have the same construction as that of Zou & Zhang (2009)'s *E-NET* adaptive weights. Consider the new *QR* based adaptive weights $\breve{\omega}_j$ (see a similar construction in Frommlet & Nuel 2016) as

$$\breve{\omega}_j = \left(\left|\hat{\beta}_j^{WR}(\tau)\right| + 1/n\right)^{-\gamma}, \tag{4.11}$$

where $\hat{\beta}_j^{WR}(\tau)$ is j^{th} parameter estimate of a *WQR-RIDGE* estimator at τ quantile level (see *Equation* (4.4)). The advantage of these adaptive weights, $\breve{\omega}_j$ is that they are different at each τ quantile level. The weights, $\breve{\omega}_j$ are robust, and we use them in the proposed adaptive *QR* procedures in Subsections 4.3.2 and 4.3.3.

4.3.2 Weighted Quantile Regression with an Adaptive LASSO Penalty

We suggest adaptive *LASSO* penalized *QR* and *WQR* procedures denoted by *QR-ALASSO* and *WQR-ALASSO*, respectively, as a remedy to data aberrations in the X-space (see also Zou 2006, for the *LS* adaptive version). The *QR-ALASSO* procedure penalizes predictor variable coefficients at different quantile levels using the $\hat{\beta}_j^{WR}(\tau)$ based adaptive weights, $\breve{\omega}_j$. The *QR-ALASSO* procedure is an extension of *LASSO* penalty based one that performs better. The *QR-ALASSO* procedure is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{AL}(\tau) = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^n \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + \lambda_n \Sigma_{j=1}^p \breve{\omega}_j |\beta_j|, \ for\ j \in [1:p],\ i \in [1:n], \tag{4.12}$$

where $\breve{\omega}_j$ is the adaptive weight and the tuning parameter $\lambda_j = \lambda \breve{\omega}_j$ (instead of $\tilde{\omega}_j = 1/|\tilde{\beta}_j|^\gamma$, for $\gamma > 0$). Other terms are as defined in previous sections. This procedure shrinks predictor coefficients to zero differently, as the tuning parameter is no-longer constant (λ) but varying (λ_j) for $j \in [1:p]$. The asymptotic properties and conditions of the *QR-ALASSO* are the same as those of the *QR-LASSO*. The *QR-ALASSO* procedure is $\sqrt{n}$-consistent.

The robust *MCD* weights, ω_i and adaptive weights, $\breve{\omega}_j$ are used to formulate the *WQR-ALASSO* variable selection procedure, which is an extension of *QR-LASSO*. The procedure inherits some properties of *WLAD-LASSO*, such as robustness in the presence of high leverage points (and collinearity influential points) due to robustly chosen *MCD* based weights, as in *WQR* (Giloni et al. 2006, Ranganai 2007). The *WQR-ALASSO* estimator is a solution to the minimization problem

$$\hat{\boldsymbol{\beta}}^{WAL}(\tau) = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^{n}\omega_i\rho_\tau|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)| + \Sigma_{j=1}^{p}\lambda_j\breve{\omega}_j|\beta_j|,\ for\ j\in[1:p],\ i\in[1:n], \quad (4.13)$$

where terms are as defined in previous sections. The *WQR-ALASSO* procedure is robust in the predictor and response spaces. Its robustness in the presence of high leverage points and collinearity influential points is a property inherited from the *ALASSO* case and the the robust *MCD* based weights.

4.3.3 Weighted Quantile Regression with an Adaptive Elastic Net Penalty

Zou & Zhang (2009) suggested the adaptive *E-NET* (*AE-NET*) procedure that inherits some good properties from both *ALASSO* and *ARIDGE* penalty based procedures. Based on the adaptive weights, $\breve{\omega}_j$ we propose the adaptive *E-NET QR* (*QR-AE-NET*) procedure. The *QR-AE-NET* estimator of $\boldsymbol{\beta}$ is a solution to the minimization problem

$$\hat{\boldsymbol{\beta}}^{AE}(\tau) = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^{n}\rho_\tau|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)| + \alpha\lambda\Sigma_{j=1}^{p}\breve{\omega}_j|\beta_j| + (1-\alpha)\lambda\Sigma_{j=1}^{p}\breve{\omega}_j\beta_j^2,\ j\in[1:p],\ i\in[1:n], \quad (4.14)$$

where the double penalty function $\alpha\Sigma_{j=2}^{p}\lambda_j\tilde{\omega}_j|\beta_j| + (1-\alpha)\Sigma_{j=1}^{p}\lambda_j\tilde{\omega}_j\beta_j^2$ is free to swing from the $\alpha = 0$ (*ARIDGE* penalty) to the other extreme when $\alpha = 1$ (*ALASSO* penalty) controlled by the mixing parameter $\alpha \in [0,1]$. In this thesis, we use $\alpha = 0.50$ for the *QR-AE-NET* scenario.

Lastly, using the adaptive weights $\breve{\omega}_j$ and the *MCD* based weights, ω_i, we propose a weighted version of *QR-AE-NET* (*WQR-AE-NET*). The *WQR-AE-NET* procedure is given by the minimization problem

$$\hat{\boldsymbol{\beta}}^{WAE}(\tau) = argmin_{\boldsymbol{\beta} \in R^p} \Sigma_{i=1}^{n} \omega_i \rho_\tau |y_i - \boldsymbol{x}_i' \boldsymbol{\beta}(\tau)| + \alpha \lambda \Sigma_{j=1}^{p} \breve{\omega}_j |\beta_j| + (1-\alpha) \lambda \Sigma_{j=1}^{p} \breve{\omega}_j \beta_j^2, \; j \in [1:p], \; i \in [1:n], \tag{4.15}$$

where the terms are as defined in previous sections. The mixing parameter $\alpha \in [0,1]$ results in *WQR-ARIDGE*$(\alpha = 0)$ and *WQR-ALASSO*$(\alpha = 1)$ as special cases of the *WQR-E-NET* construction. Just like its unweighted penalized *QR* counterpart, *WQR-AE-NET* inherits desired optimal minimax bound from *ALASSO* (Zou 2006). The *WQR-AE-NET* variable selection and regularization procedure is robust in the presence of outliers, collinearity and collinearity influential points (high leverage points).

4.4 Asymptotics for Quantile Regression

The behavior of the model estimates and inferences in large samples can be understood using the asymptotic theory (large sample theory) of QR analysis. Koenker & Bassett (1978), in one of their seminal papers on the asymptotic theory of QR, showed the consistency and asymptotic normality of the QR estimator under mild conditions. Other authors have since considered asymptotics, namely Kato et al. (2012) in panel quantile regression models and Galvao et al. (2020) on the asymptotic normality of quantile regression with fixed effects, just to mention a few.

Although *RQ*s can be represented clearly in a finite sample distribution, there are challenges computationally and assumptions considerations hence the need for asymptotic distributions in *QR*. In the literature, local linearization and central limit theorem approximation methods are important statistical inference tools, although they are said to be necessary compromises. In the asymptotic scenario, the adequacy of asymptotic approximations is evaluated by higher expansions and resampling methods aided by Monte Carlo simulations. This asymptotic theory imposes the necessary discipline and consistency in the formulation of statistical procedures. Heteroscedasticity and other non-location shift covariate properties in regression cause inconsistencies in which conditional quantile functions proffer a straight-forwardly interpretable objective for statistical analysis.

Consistency is a minimal prerequisite for *QR* to work, therefore, we briefly discuss it. Given $X = x$, the τ^{th} conditional quantile function of Y takes the form $Q_{Y|X}(\tau) = g(x, \beta(\tau))$ for the non-linear *QR* case and $Q_{Y|X}(\tau) = \boldsymbol{x}'\boldsymbol{\beta}(\tau)$ for the linear *QR* case (Koenker et al. 2018).

If we start from a random sample $(y_1, y_2, ..., y_n)$, which is F distributed, the necessary conditions for the univariate sample quantile $\hat{\beta}_n(\tau) = arg\ min_{\beta \in R} \Sigma_{i=1}^{n} \rho_\tau(y_i - \beta)$ imply that $\hat{\beta}_n(\tau) \to \beta(\tau)$ as $n \to \infty$. This works when we have a unique τ^{th} quantile, $\beta(\tau) = F^{-1}(\tau)$ and that $\sqrt{n}(\hat{\beta}_n(\tau) - \beta(\tau)) \to N(0; \omega^2)$, where $\omega^2 = \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}$ and F has density $f(F^{-1}(\tau))$ bounded away from 0 and ∞ near τ.

In addition to the consistency of estimators, we consider convergence. It is prudent to consider how rapidly convergence occurs. Let us say $Y_1, Y_2, ..., Y_n$ are *iid* random variables such that $Y_1 \sim F_1, Y_2 \sim F_2, ..., Y_n \sim F_n$, where F is a distribution function. In the linear *QR* scenario, the τ^{th} conditional quantile function of Y_is given X_is is given by $Q_{Y_i|X_i}(\tau) = F^{-1}_{Y_i|X_i}(\tau) \equiv \boldsymbol{x}_i'\boldsymbol{\beta}(\tau)$ where $P(Y_i < y|x_i) = F_{Y_i|X_i}(y) = F_i(y)$. We discuss the asymptotic behavior of the estimator, $\hat{\boldsymbol{\beta}}_n(\tau) = arg\ min_{\beta \in R^p} \Sigma \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$ using regularity conditions (1) and (2):

(1) $\{F_i\}$ is absolutely continuous. Also, the continuous density, $f_i(\beta)$, is uniformly bounded away from zero and ∞ at the points $\boldsymbol{\beta}_i(\tau), i = 1, 2, ...$

(2) There exists positive definite matrices $\boldsymbol{D}_0$ and $\boldsymbol{D}_1(\tau)$:

$$(a) \qquad lim_{n\to\infty}\frac{\Sigma x_i x_i'}{n} = \boldsymbol{D}_0$$

$$(b)\ lim_{n\to\infty}\frac{\Sigma f_i(\boldsymbol{\beta}_i(\tau))\boldsymbol{x}_i\boldsymbol{x}_i'}{n} = \boldsymbol{D}_1(\tau)$$

$$(c) \qquad max_{i=1,2,\ldots,n}\frac{||\boldsymbol{x}_i||}{\sqrt{n}} \to 0$$

We let matrix $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_2 & \boldsymbol{D}_1 \\ \boldsymbol{D}_1 & \boldsymbol{D}_0 \end{pmatrix} = lim_{n\to\infty}\Sigma_{i=1}^n \begin{pmatrix} f_i^2 & f_i \\ f_i & 1 \end{pmatrix} \bigotimes \boldsymbol{x}_i\boldsymbol{x}_i'$, where $\bigotimes$ is the Kronecker product. Since $\boldsymbol{D}_2$ is positive definite, $\exists$ orthogonal matrix $\boldsymbol{P} : \boldsymbol{P}'\boldsymbol{D}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{D}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_2 - \boldsymbol{D}_1'\boldsymbol{D}_2^{-1}\boldsymbol{D}_1 \end{pmatrix}$ so that we have $\boldsymbol{D}_2 - \boldsymbol{D}_1'\boldsymbol{D}_2^{-1}\boldsymbol{D}_1$ and $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2\boldsymbol{D}_1^{-1} - \boldsymbol{D}_2^{-1}$ being non-negative. Koenker (2005) concludes that if conditions 1 and 2 are met, then $\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}(\tau))$ converges asymptotically in limit to $N(\boldsymbol{0}; \tau(1-\tau)\boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1})$. In the iid noise model with $f_i(\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)) = f(\boldsymbol{\beta}(\tau))$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}(\tau))$ converges asymptotically in limit to $N(\boldsymbol{0}; \omega^2\boldsymbol{D}_0^{-1})$, where $\omega^2 = \frac{\tau(1-\tau)}{f^2(\boldsymbol{\beta}(\tau))}$, $\boldsymbol{D}_1 = lim_{n\to\infty}\Sigma_{i=1}^n f_i\boldsymbol{x}_i\boldsymbol{x}_i'\ and\ \boldsymbol{D}_0 = lim_{n\to\infty}\Sigma_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i'$.

The efficiency of heterogeneous conditional densities of the response variable can be improved by considering the WQR. The parameter estimator is given by $\check{\boldsymbol{\beta}}_n(\tau) = argmin_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^n f_i(\boldsymbol{x}_i'\boldsymbol{\beta})\rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})$. In this scenario under conditions 1 and 2, $\sqrt{n}(\check{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}(\tau))$ converges asymptotically in limit to $N(\boldsymbol{0}; \tau(1-\tau)\boldsymbol{D}_2^{-1}(\tau))$, where $\boldsymbol{D}_2(\tau) = lim_{n\to\infty}\frac{\Sigma f^2(\boldsymbol{\beta}(\tau))\boldsymbol{x}_i\boldsymbol{x}_i'}{n}$ (Koenker 2005).

Given that the conditional quantile model $Q_{Y_i|X_i}(\tau) = g(x_i, \beta_0(\tau))$, that is non-linear in parameters, the non-linear QR estimator is then given by $\hat{\boldsymbol{\beta}}_n(\tau) = arg\ min_{\boldsymbol{\beta}\in R^p}\Sigma_{i=1}^n \rho_\tau(y_i - g(\boldsymbol{x}_i', \boldsymbol{\beta}))$ for $\boldsymbol{\beta} \in R^p$. Here we state the regularity conditions for the non-linear case as follows:

(1) There exists constants k_0, k_1, n_0 and parameters $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \boldsymbol{\beta}$ and $n > n_0$ we have $k_0||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2|| \leq \left(\frac{\Sigma_{i=1}^{n}(g(x_i - \beta_1) - g(x_i, \beta_2))^2}{n}\right)^{\frac{1}{2}} \leq k_1||\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2||$.

(2) There exists positive definite matrices $\boldsymbol{D}_0$ and $\boldsymbol{D}_1(\tau)$: $\dot{g}_i = \frac{\partial g(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}|\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ where $\dot{g}_i$ is the derivative of g_i.

$$(a) \qquad lim_{n\to\infty}\frac{\Sigma \dot{\boldsymbol{g}}_i \dot{\boldsymbol{g}}_i'}{n} = \boldsymbol{D}_0$$

$$(b)\ lim_{n\to\infty}\frac{\Sigma f_i(\boldsymbol{\beta}_i(\tau))\dot{\boldsymbol{g}}_i \dot{\boldsymbol{g}}_i'}{n} = \boldsymbol{D}_1(\tau)$$

$$(c) \qquad max_{i=1,2,\ldots,n}\frac{||\dot{\boldsymbol{g}}_i||}{\sqrt{n}} \to 0$$

Condition (1) guarantees a unique minimum of the objective function at β_0, as in Jureckova & Procházka (1994). Condition (2) is important in determining the behavior of the conditional density of the response variable around the conditional quantile model. Similarly, under conditions (1) and (2), $\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau))$ converges asymptotically in limit to $N(\boldsymbol{0}; \tau(1-\tau)\boldsymbol{D}_1^{-1}\boldsymbol{D}_0\boldsymbol{D}_1^{-1})$, where $\boldsymbol{D}_1(\tau) = lim_{n\to\infty}\frac{\Sigma f(\boldsymbol{\beta}(\tau))\boldsymbol{x}_i\boldsymbol{x}_i'}{n}$, $\boldsymbol{D}_0(\tau) = lim_{n\to\infty}\frac{\Sigma \boldsymbol{x}_i\boldsymbol{x}_i'}{n}$, $\sqrt{n}(\hat{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)) = \frac{\boldsymbol{D}_1^{-1}(\tau)\Sigma_{i=1}^{n}\dot{g}_i\psi_\tau(u_i(\tau))}{n} + o_p(1)$ and $u_i(\tau) = y_i - g(\boldsymbol{x}_i, \boldsymbol{\beta}_0(\tau))$.

4.4.1 Asymptotics for Regularized Quantile Regression

For the penalized *QR* scenario, we consider $P(\varepsilon_i) < 0$ for the linear model and $\boldsymbol{\beta} = (\boldsymbol{\beta}_s, \boldsymbol{\beta}_{p-s})'$ for a regression equation with a partitioned design matrix, where $\boldsymbol{\beta}_s \neq 0$ and $\boldsymbol{\beta}_{p-s} = 0$.

We assume the following theoretical results to be true for asymptotic normality for a suitable choice of λ_n, as stated in Wu & Liu (2009). The two conditions are, thus:

(i) The regression errors $\varepsilon_i s$, are *i.i.d.*, τ^{th} quantile $Q_\tau = 0$ and a continuous, positive density $f(.)$ in a neighborhood of origin zero and F distributed (Pollard 1991). NOTE: $F(0) = \tau$ and

$|f(y)-f(0)| \leq c|y|^{1/2}$, $\forall y$ in the neighborhood of 0 and real quantile $\tau \in (0,1)$.

(ii) Let $\boldsymbol{\Omega} = diag(\omega_1, \omega_2, ..., \omega_n)$, where ω_i for $i \in [1:n]$ are known positive values that satisfy $max\{\omega_i\} = O(1)$.

(iii) Consider the partitioned design matrix $\boldsymbol{X} = (\boldsymbol{X}_s, \boldsymbol{X}_{p-s})'$, such that there exists a positive definite covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = lim_{n\to\infty}\boldsymbol{X}^{*\prime}\boldsymbol{X}^{*}/n$. We can write the covariance sub-matrix $\boldsymbol{\Sigma}_{11} = \frac{1}{n}\boldsymbol{X}_s'\boldsymbol{X}_s$ and the other covariance sub-matrix $\boldsymbol{\Sigma}_{22} = \frac{1}{n}\boldsymbol{X}_{p-s}'\boldsymbol{X}_{p-s}$ for R^s and R^{p-s}, respectively.

Let the solution to a regularization procedure (RP) be $\boldsymbol{\beta}^{RP}$. For $i.i.d.$ random error terms, we state a theorem for asymptotic oracle property in Theorem 4.4.1 (see Wu & Liu 2009, Ranganai & Mudhombo 2021, Mudhombo & Ranganai 2022).

Theorem 4.4.1. *Consider a sample $\{(\boldsymbol{x}_i, y_i), i = 1, ..., n\}$ from a regularization procedure satisfying conditions (i), (iii) and $\boldsymbol{\Omega} = \boldsymbol{I}_n$ (constant weight of 1). If $\sqrt{n}\lambda_n \to 0$ and $n^{(\gamma+1)/2}\lambda_n \to \infty$, then*

1. We have sparsity when $\widehat{\boldsymbol{\beta}}_{p-s}^{RP} = \mathbf{0}$, where RP denotes the regularization procedure;

2. $\sqrt{n}(\widehat{\boldsymbol{\beta}}_s^{RP} - \boldsymbol{\beta}_s^{RP})$ approximates a $N\left(\mathbf{0}, \frac{\tau(1-\tau)\boldsymbol{\Sigma}_{11}^{-1}}{f(0)^2}\right)$.

To extend the oracle results to a non $i.i.d.$ random error scenario, we consider the following assumptions.

(iv) As $n \to \infty$, $max_{i<i<n}\{\boldsymbol{x}_i'\boldsymbol{x}_i/n\} \to 0$.

(v) The random errors $\varepsilon_i s$ are independent with $F_i(t) = P(\varepsilon_i \leq t)$ the distribution function of ε_i. We assume that each $f_i(.)$ is locally linear near zero (with a positive slope), and $F_i(0) = \tau$. Define $\Psi_{ni}(t) = \int_0^t \sqrt{n}(F_i(s/\sqrt{n}) - F_i(0))ds$. which is a convex function for each n and i.

(vi) For each $\boldsymbol{u}$, we assume that $(1/n)\sum_{i=1}^n \Psi_{ni}(\boldsymbol{u}'\boldsymbol{x}) \to \varsigma(\boldsymbol{u})$, where $\varsigma(.)$ is a strictly convex function taking values in $[0,\infty)$.

Corollary 4.4.1.1. *Under Conditions* (ii)*,* (iii) *and* (iv)*, Theorem* 4.4.1 *holds provided the non* $i.i.d.$ *random errors satisfy* (v) *and* (vi) *(see also Knight 1999).*

Remarks.

• *The QR model, a non* $i.i.d.$ *random error model is catered for by* (v) *(Koenker 2005). See proofs on-line (Wu & Liu 2009).*

Some asymptotics are not considered in this study. These include asymptotics under dependent conditions, such as autoregression, *ARMA* models, *ARCH*-like models and extremal *QR*.

4.5 Concluding Remarks

This chapter explored and proposed different types of penalized *QR* and *WQR* variable selection procedures. In Section 4.2, we proposed penalized *WQR* procedures. We first discussed the choice of robust weights ω_i based on the computationally intensive high breakdown *MCD* method in Subsection 4.2.1. In Subsections 4.2.2, 4.2.3 and 4.2.4, we suggested the ω_i based *WQR* procedures namely, *WQR-RIDGE*, *WQR-LASSO* and *WQR-E-NET* procedures, respectively. The new penalized *WQR* procedures are robust in the X and Y-spaces (an inherited property from *WLAD*). We further suggested adaptive penalized *QR* procedures in Section 4.3 by first suggesting adaptive weights $\breve{\omega}_j$ based on $\hat{\beta}_j^{WR}(\tau)$ in Subsection 4.3.1. Using the adaptive weights suggested in Subsection 4.3.1, we further suggested weighted and unweighted adaptive penalized *QR* procedures namely, *QR-ALASSO* and *WQR-ALASSO* in Subsection 4.3.2 and *QR-AE-NET* and *WQR-AE-NET* in Subsection 4.3.3. Because of the carefully chosen adaptive weights $\breve{\omega}_j$, the proposed adaptive penalized WQR procedures are robust in the presence of collinearity and high leverage points. In Section 4.4, we discussed asymptotics for penalized QR procedures.

Chapter 5

Simulation Studies and Application to Well-known Data from the Literature

In this chapter, we perform simulation studies to investigate the finite sample performance of regularized *QR* and *WQR* procedures. The simulation studies are divided into four categories:

(i) a comparison of regularized *QR* procedures *QR-LASSO*, *QR-RIDGE* and *QR-E-NET* against their respective weighted versions *WQR-LASSO*, *WQR-RIDGE* and *WQR-E-NET*.

(ii) a comparison of non-adaptive penalized *QR* procedures *QR-LASSO* and *QR-E-NET* against their respective adaptive versions *QR-ALASSO* and *QR-AE-NET*.

(iii) a comparison of unweighted adaptive penalized *QR* procedures *QR-ALASSO* and *QR-AE-NET* against their respective weighted adaptive versions *WQR-ALASSO* and *WQR-AE-NET*,

and

(iv) an "omnibus" comparison of all penalized procedures.

The robust weights for *WQR* procedures are based on *MCD* based robust distances, and the adaptive weights for adaptive penalized *QR* procedures are based on the $\hat{\beta}_j^{WR}(\tau)$ estimates.

5.1 Simulation Studies Design

The results of the simulation studies are summarized in terms of the average number of correctly/incorrectly fitted zero coefficients $\{\beta_j : j = 3,4,6,7,8\}$ (From $\boldsymbol{\beta} = (3,1.5,0,0,2,0,0,0)'$ used in this thesis as the true $\boldsymbol{\beta}$, zeros in positions $3,4,6,7,8$ are referred to as correctly fitted zeros if estimated as zero and incorrectly as otherwise.), percentage of correctly fitted models and the *MAD* of test errors (a measure of dispersion). When we consider the median of test errors $Median\{\varepsilon_i\}$, the *MAD* of test errors is given by $MAD = 1.4826\,(Median\{\varepsilon_i\} - Median\{\varepsilon_i\})\,, i \in [1:n]$. All simulations are applied at $\tau \in (0.25;0.50)$ *RQ* levels and $n \in (50;100)$ sample sizes. For brevity, the results for $n = 100$ are left out.

We consider the predictor design matrices with six X-space data aberrations comprising high leverage points, collinearity influential points and collinearity coupled with different error term distributions ($D1$-$D6$). The error term distributions comprise the normal distribution with varying tail thickness determined by the error variance and the t-distribution with different tail thickness determined by the different degrees of freedom. We consider six design matrix scenarios and constructed the first five as in Ranganai et al. (2014) while the sixth one as in Arslan (2012) namely:

• $D1$: From $N(0,1)$, generate and orthogonalize an $n \times p$ design matrix $\boldsymbol{X}$. The orthogonal design matrix $\boldsymbol{X}$ satisfies the condition, $\boldsymbol{X}'\boldsymbol{X} = n\boldsymbol{I}$. We first generate the $n \times p$ data, $\boldsymbol{W}$, where $w_{ij} \sim N(0,1)$ with $i \in [1:n]$ and $j \in [1:p]$. The singular value decomposition (SVD) of this design matrix $\boldsymbol{W}$ is given by $\boldsymbol{W} = \boldsymbol{UDV}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal with the diagonal entries of $\boldsymbol{D}$ corresponding to the eigenvalues of $\boldsymbol{W}$ so that $\boldsymbol{X} = \sqrt{n}\boldsymbol{U}$. Then $\boldsymbol{X}'\boldsymbol{X} = n\boldsymbol{I}$, since $\boldsymbol{U}$ is orthogonal.

• $D2$-has collinearity inducing point, i.e., $D1$, with observation having the largest Euclidean distance from the center of the design space moved 10 units in the predictor space.

• $D3$-has collinearity hiding point, i.e., $D1$, with observations having the largest and second largest Euclidean distance from the center of the design space moved 10 units in the predictor space.

• $D4$-has collinearity inducing point, i.e., $D1$, with observation having the largest Euclidean distance from the center of the design space moved 100 units in the predictor space.

• $D5$-the collinearity hiding point scenario, i.e, $D1$, with observations having the largest and second largest Euclidean distance from the center of the design space moved 100 units in the predictor space.

• $D6$-the heavy tailed distribution scenario (contaminated t-distribution design scenario), with 10% contaminated cases. The design matrix is the partitioned matrix $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$, where $\boldsymbol{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{V})$, $\boldsymbol{X}_1 \in R^{(n-m)\times p}$, $\boldsymbol{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{I})$ and $\boldsymbol{X}_2 \in R^{m\times p}$ for $n = (50; 100)$ and $m = (5; 10)$. $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are the uncontaminated and contaminated parts of $\boldsymbol{X}$, respectively. The $(ij)^{th}$ entry of the covariance matrix $\boldsymbol{V}$ (v_{ij}) is determined by the exponential decay $0.5^{|j-i|}$, where $i \in [1:8]$ and $j \in [1:8]$ are rows and columns of $\boldsymbol{V}$, respectively. The covariance matrix $\boldsymbol{I}$, is a diagonal matrix of ones, and the mean vectors are given by $\boldsymbol{\mu_1} = (0,0,0,0,0,0,0,0)'$ and $\boldsymbol{\mu_2} = (1,1,1,1,1,1,1,1)'$ (see Arslan 2012, for similar construction). The collinearity in $D6$ is due to the covariance structure, $\boldsymbol{V}$, and the mean shift in $\boldsymbol{X}_2$ introduces high leverage points to the design matrix.

The error term distributions considered are $e \sim N(\mu, \sigma^2)$, for $(\mu, \sigma) \in ((0;1),(0;3))$ and $e \sim t_d$ for $d \in (1;6)$. The response vector $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')'$ is generated as in Arslan (2012), i.e.,

• $\boldsymbol{Y}_1 = \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \sigma\boldsymbol{e}$, $\boldsymbol{e} \sim N(0, \sigma^2)$, $\sigma \in (1;3)$ and $\boldsymbol{Y}_2 = \boldsymbol{X}_2'\boldsymbol{\beta}_2$ for $D1$-$D5$.

• $\boldsymbol{Y}_1 = \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \sigma\boldsymbol{e}$, $\boldsymbol{e} \sim t_d$, $d \in (1;6)$, $\sigma \in (0.5;1)$ and $\boldsymbol{Y}_2 = \boldsymbol{X}_2'\boldsymbol{\beta}_2$ for $D6$.

In all the cases, $\boldsymbol{\beta}_1 = (3, 1.5, 0, 0, 2, 0, 0, 0)'$ and $\boldsymbol{\beta}_2 = (2, 1, 0, 3, 1.5, 0, 1, 0)'$.

Remarks. *The design matrix, $D1$, is used as baseline for comparisons with $D2$-$D6$ scenarios.*

Figure 5.1 depicts a schematic representation of collinearity influential points for design scenarios $D2$-$D5$. The row panels 1 and 3 show the collinearity inducing cases in $D2$ and $D4$, and row panels 2 and 4 show the collinearity reducing cases in $D3$ and $D5$. One extreme leverage point in the collinearity inducing scenario has an undue influence on the rest of the data. However, adding another high leverage point (scenarios $D3$ and $D5$) reduces the influence of the first extreme leverage point. As points move further away from rest of data (*e.g.*, predictor value (observation) $10\boldsymbol{x}_i'$ to predictor value (observation) $100\boldsymbol{x}_i'$), the collinearity-inducing and hiding effects increase.

A schematic representation of collinearity in the $D1$-$D5$ design scenarios induced or reduced by high leverage points is shown in *Table* 5.1. The baseline scenario, $D1$, shows no collinearity among explanatory variables ($r_{ij} \to 0$). Collinearity (correlations in predictor variables) increase as the leverage point moves away from the baseline scenario ($D1$ to any of $D2$ to $D5$). We also notice a shift in signs from the baseline $D1$ to the $D3/D5$ scenarios. The introduction of a high leverage point (collinearity inducing point) in $D2$ saw collinearity increasing among predictor variables, as in Mason & Gunst (1985).

Remarks. *D*2 *and D*4 *have high leverage points that induce collinearity. D*3 *and D*5 *have high leverage points that hide collinearity. D*6 *contains collinearity as well as high leverage points. Collinearity increases as the leverage points move further away from the rest of the data (D*1 *to any of D*2*- D*5*). Collinearity can be hidden by another extreme leverage point (D*3 *and D*5*).*

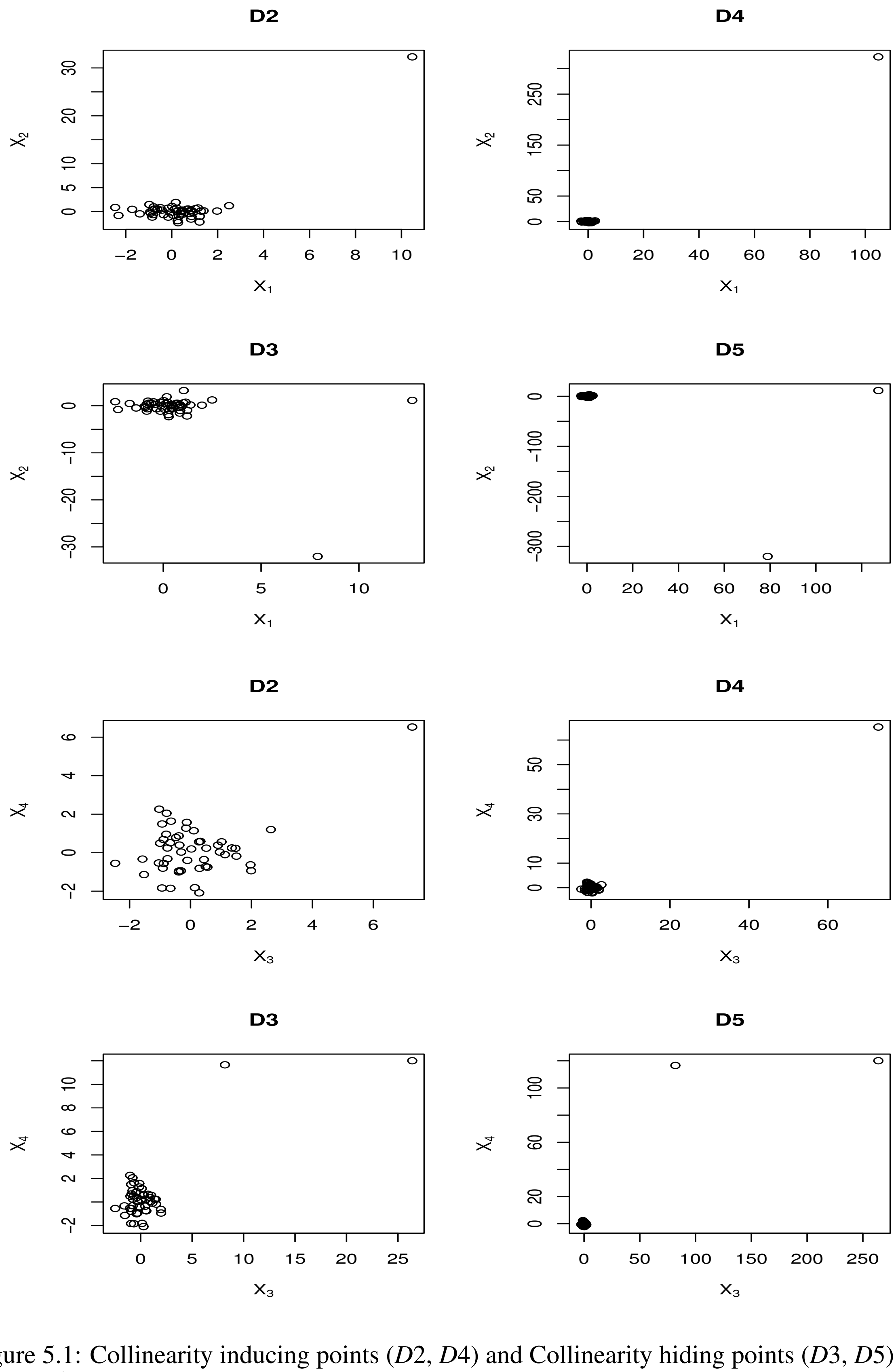


Figure 5.1: Collinearity inducing points ($D2$, $D4$) and Collinearity hiding points ($D3$, $D5$) with i^{th} and/or j^{th} high leverage value(s)

Table 5.1: Collinearity depicted by correlation matrices of predictor variables. $D1$-Orthogonal (baseline), $D2/D4$-with collinearity inducing points and $D3/D5$-with collinearity hiding points.

				$D1$									$D2$									$D4$			
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1.00	0.00	0.01	0.00	0.02	-0.03	0.00	0.01	1.00	0.80	0.59	0.55	0.48	0.73	-0.67	0.71	1.00	1.00	0.99	0.99	0.99	1.00	-1.00	1.00
X_2		1.00	0.00	0.00	0.00	0.00	0.00	0.00		1.00	0.70	0.66	0.55	0.87	-0.79	0.84		1.00	0.99	0.99	0.99	1.00	-1.00	1.00
X_3			1.00	0.00	-0.01	0.01	0.00	0.00			1.00	0.48	0.40	0.64	-0.58	0.61			1.00	0.99	0.98	0.99	-0.99	0.99
X_4				1.00	0.00	-0.01	0.00	0.00				1.00	0.38	0.60	-0.54	0.58				1.00	0.98	0.99	-0.99	0.99
X_5					1.00	0.03	0.00	-0.01					1.00	0.52	-0.46	0.49					1.00	0.99	-0.99	0.99
X_6						1.00	0.01	0.01						1.00	-0.72	0.77						1.00	-1.00	1.00
X_7							1.00	0.00							1.00	-0.70							1.00	-1.00
X_8								1.00								1.00								1.00
													$D3$									$D5$			
X_1									1.00	-0.42	0.85	0.80	-0.66	0.10	-0.88	-0.74	1.00	-0.48	0.97	0.97	-0.86	0.14	-1.00	-0.96
X_2										1.00	-0.23	-0.59	-0.03	-0.83	0.41	0.60		1.00	-0.24	-0.66	-0.02	-0.93	0.43	0.71
X_3											1.00	0.79	-0.79	-0.11	-0.93	-0.71			1.00	0.89	-0.96	-0.12	-0.98	-0.85
X_4												1.00	-0.57	0.29	-0.86	-0.79				1.00	-0.73	0.35	-0.96	-0.99
X_5													1.00	0.31	0.74	0.49					1.00	0.38	0.89	0.68
X_6														1.00	-0.06	-0.32						1.00	-0.08	-0.42
X_7															1.00	0.79							1.00	0.94
X_8																1.00								1.00

The simulated data is split into the training set for model construction and the validation set for model validation in the ratio 80 : 20 (see Section 2.1). We split the data at random to guarantee the two parts have the same distribution. We use the 10-fold cross validation (10-$fold\ \ CV$) criterion to choose an optimal tuning parameter (λ) to achieve the best overall prediction (see also Shao 1993). The tuning parameter value is intimately linked to the accuracy of the predictions made by the models. The penalized *QR* variable selection procedures' performances are found by averaging the errors across different test sets of the data (see Section 2.1). We select the best variable selection and model estimation procedure with the best predictive ability among a class of penalized *QR* procedures. Cross validation methods select penalized *QR* procedures according to the procedure's predictive ability. A *CV* criterion is asymptotically equivalent to many model selection procedures namely, *AIC*, *BIC*, Cp, etc.

In this thesis, we explore the penalized *QR* and WQR procedures using the *hqreg package*, which is a readily available *R*-software program on the website http://cloud.r-project.org/package=hqreg (see Yi 2017). The *hqreg package* implements the semi-smooth Newton coordinate descent algorithm that chooses optimal λ when selecting the best procedure. We perform 10-fold cross validation for the regularization and variable selection using *cv.hqreg*$(\boldsymbol{X}, \boldsymbol{y}, ...)$ and fit the model by *hqreg*$(\boldsymbol{X}, \boldsymbol{y}, ...)$ (see Yi 2017, for the detailed *R* codes).

In the next sections, we present results of simulation studies on the performance of the variable selection and regularized *QR* and *WQR* procedures (non-adaptive and adaptive penalized procedures). These penalized *QR* and *WQR* procedures are *QR-RIDGE*, *QR-LASSO*, *QR-E-NET*, *WQR-RIDGE*, *WQR-LASSO*, W*QR-E-NET*, *QR-ALASSO*, *QR-AE-NET*, *WQR-ALASSO* and *WQR-AE-NET*. We compare and contrast the performance of these penalized procedures with the baseline scenario and, to a larger extent, among themselves. Low *MAD* of test error values, higher percentages of correctly fitted models and a higher/lower average number of correctly/incorrectly fitted zero coefficients indicate better performance. Simulation results are categorized as follows: (i) results of regularized *WQR* procedures (see Section 5.2) and (ii) results of adaptive penalized *WQR* procedures (see Section 5.4). The results of the application of the penalized *QR* and *WQR* methods to real data sets from the literature are presented in Sections 5.3 and 5.5.

The data in this thesis is simulated using the algorithm.

Algorithm : ***Simulated Data***

Input :

Input data set $\{\boldsymbol{x}_i', y_i\}$ # the unweighted data set.

ω_i # the *MCD* based weights of size n.

$\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)$. # $p = 8$.

ε_i # t-distributed or Gaussian distributed error term.

Output: Data set $\{\boldsymbol{x}_i^{*'}, y_i^*\}$

(i) $RD(x_i) = \sqrt{\left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}\right)' \widehat{\boldsymbol{\Sigma}}^{-1} \left(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}\right)}$, $i \in [1:n]$ # Compute the *MCD* based robust distance.

(ii) $\omega_i = min\left(1, \frac{p}{RD(x_i)^2}\right)$, $i \in [1:n]$. # Compute the MCD based weights.

(iii) $\boldsymbol{x}_i^{*'} = \omega_i \boldsymbol{x}_i'$. # i^{th} row of a $n \times p$ matrix $\boldsymbol{X}^*$ (weighted design matrix), and p is the number of predictor variables and n is the sample size.

(iv) $y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$: # The unweighted response.

(v) $y^* = \omega_i y_i$: # The weighted response.

(vi) ***Return*** $\{\boldsymbol{x}_i^{*'}, y_i^*\}$

5.2 Results of Weighted Quantile Regression Regularization Procedures

Baseline scenario (D1)

Table 5.2 displays the simulation results for the well-conditioned predictor matrix $D1$ (baseline scenarios) under the normal distribution and t-distribution with 1 $d.f$ (implying outliers). Under the normal distribution, *LS-LASSO* and *QR-LASSO* perform best in variable (model) selection at $\tau = 0.50$, followed by *QR-E-NET*, though the performance is not significantly different.

Table 5.2: Simulation results for orthogonal case ($D1$) under the normal and t-distributions at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels; bold text indicate better performance.

			$\tau = 0.25$					$\tau = 0.50$				
			med(MAD)	Correctly	No. of Zeros			med(MAD)	Correctly	No. of Zeros		
Distribution	Parameters	Method	Test Error	Fitted	c.zero	inc.zero	Optimal(λ)	Test Error	Fitted	c.zero	inc.zero	Optimal(λ)
$D1-N(.,.)$	$\sigma = 1$	LS-RIDGE						0.16(2.54)	5.00	2.53	0.00	1.49
		LS-LASS0						**-0.02(1.15)**	**90.00**	**4.90**	0.00	0.15
		LS-E-NET						-0.02(1.36)	66.50	4.64	0.00	0.30
		QR-RIDGE	1.28(1.97)	0.00	2.27	0.00	0.12	-0.03(1.99)	1.50	2.33	0.00	0.14
		QR-LASSO	**0.71(1.20)**	**67.50**	**4.56**	0.00	0.04	**0.00(1.15)**	**62.00**	**4.49**	0.00	0.05
		QR-E-NET	0.72(1.25)	18.50	3.59	0.00	0.04	0.01(1.19)	24.00	3.60	0.00	0.04
	$\sigma = 3$	LS-RIDGE						-0.06(4.14)	5.00	2.46	0.00	1.51
		LS-LASS0						**-0.05(3.41)**	**73.00**	**4.86**	0.14	0.25
		LS-E-NET						-0.06(3.54)	44.50	4.27	0.03	0.31
		QR-RIDGE	2.70(4.32)	9.00	3.07	0.03	0.12	-0.04(4.06)	2.50	2.37	0.01	0.12
		QR-LASSO	**2.03(3.60)**	**39.50**	**4.52**	0.38	0.04	**0.01(3.45)**	**40.00**	**4.57**	0.32	0.05
		QR-E-NET	2.18(3.69)	30.50	4.00	0.20	0.04	0.00(3.55)	31.00	3.90	0.11	0.04
$D1-t_d$	$d = 1$ $\sigma = 0.5$	LS-RIDGE						0.24(4.12)	4.00	2.57	0.21	1.61
		LS-LASS0						**0.14(3.31)**	32.00	**4.99**	1.80	0.48
		LS-E-NET						0.15(3.46)	26.00	4.87	1.75	0.43
		QR-RIDGE	2.17(3.21)	3.00	2.33	0.02	0.11	0.02(2.94)	1.50	2.56	0.01	0.13
		QR-LASSO	**1.24(2.16)**	**64.00**	**4.92**	0.72	0.04	**0.02(1.72)**	**64.50**	**4.94**	0.67	0.04
		QR-E-NET	1.44(2.41)	36.50	4.42	0.62	0.03	-0.01(1.94)	32.50	4.33	0.58	0.03
	$d = 1$ $\sigma = 1$	LS-RIDGE						0.19(5.61)	3.00	2.54	0.43	2.15
		LS-LASS0						**0.20(5.38)**	8.50	**4.99**	2.51	1.57
		LS-E-NET						0.22(5.45)	12.00	4.96	2.43	1.33
		QR-RIDGE	3.05(4.30)	2.50	2.37	0.03	0.11	0.02(4.15)	3.50	2.49	0.02	0.12
		QR-LASSO	**2.44(3.80)**	**30.50**	**4.95**	1.57	0.04	**0.02(3.34)**	**33.50**	**4.95**	1.38	0.04
		QR-E-NET	2.62(4.04)	26.00	4.78	1.49	0.03	0.02(3.58)	25.50	4.66	1.27	0.03

[1] $N(.,.)$ denotes normally distributed and t_d denotes t-distribution with d degrees of freedom.

Remarks. *In Table* 5.2*, the five zero coefficients correspond to the set* $\{\beta_j : j = 3,4,6,7,8\}$*, hence the maximum average of correctly shrunk coefficients is* 5*, while the set of correctly selected models is given as a percentage.*

With respect to the percentage of correctly fitted models and the average of correctly fitted zero coefficients, *LS-LASSO* performs best. However, the performance of the respective procedures is decreased at larger σ, with respect to the two metrics. Under the t-distribution scenario, the *QR-LASSO* performs best at $\tau = 0.50$ (followed by *QR-E-NET*), with no marked difference in the median and *MAD* of test error measures, indicating the robustness of the *QR-LASSO* procedure. Just like under the *MAD* metric, *QR-LASSO* performs best at $\tau = 0.50$, followed by *QR-E-NET* with no marked differences in the percentage of correctly fitted models. The three procedures namely, *LS-LASSO*, *QR-LASSO* and *QR-E-NET*, perform more or less equally with respect to the average of correctly fitted zero coefficients. Generally, the respective performances of *QR-LASSO* and *QR-E-NET* procedures at $\tau = 0.25$ and $\tau = 0.50$ are more or less similar.

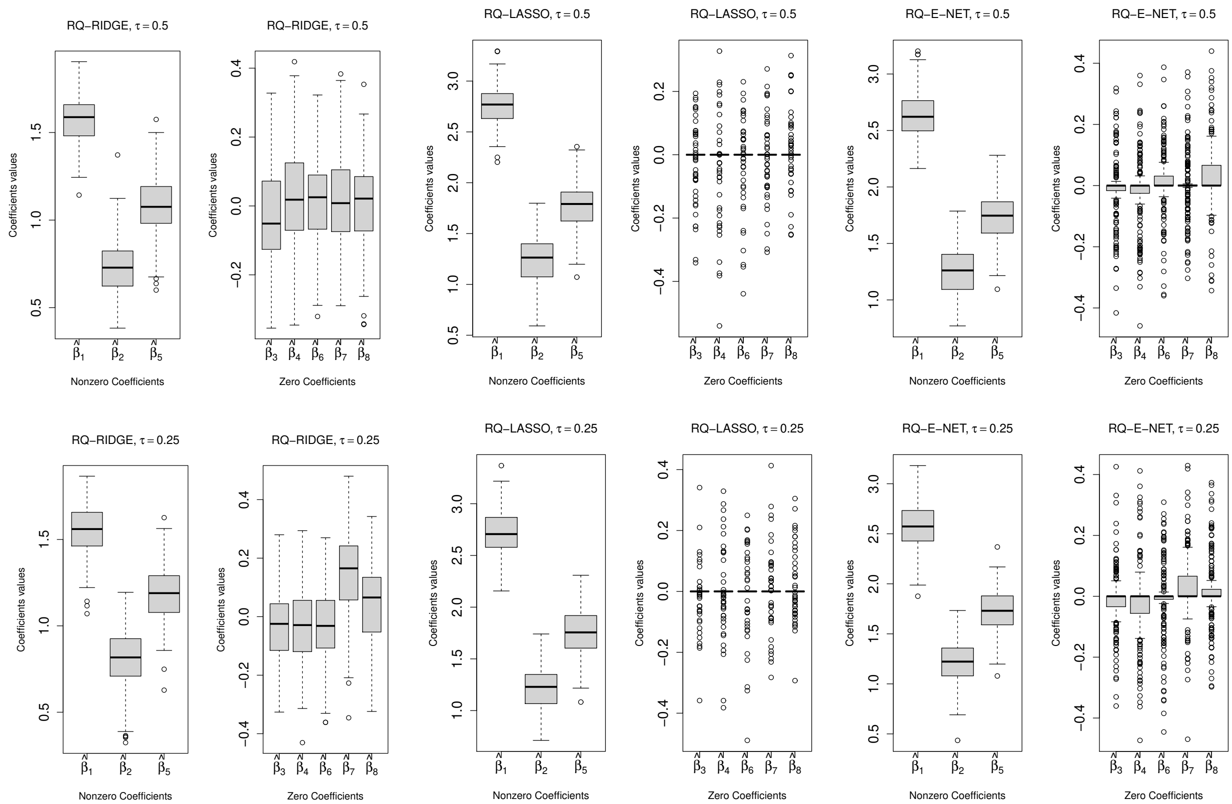


Figure 5.2: Box plots for collinearity inducing points scenario $D1$ under normal distribution with $\sigma = 1$, $d = 1$ at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted scenario).

Remarks. *The non-zero coefficients and zero coefficients are* $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ *and* $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$*, respectively. In Figure* 5.2*, the QR-LASSO and QR-E-NET procedures perform best with respect to correctly shrunk zero coefficients* $\{\beta_j : j = 3, 4, 6, 7, 8\}$*. The QR-RIDGE procedure is the worst performer. Similarly, the QR-LASSO and QR-E-NET procedures also perform best with respect to the maximum average of estimated non-zero coefficients* $\{\beta_j : j = 1, 2, 5\}$.

__Collinearity inducing scenario (D2 and D4)__

Simulation results of the collinearity inducing points scenarios ($D2$ and $D4$) are summarized in *Tables* 5.3 (Gaussian distribution scenarios) and 5.4 (heavy-tailed distribution scenarios), and the

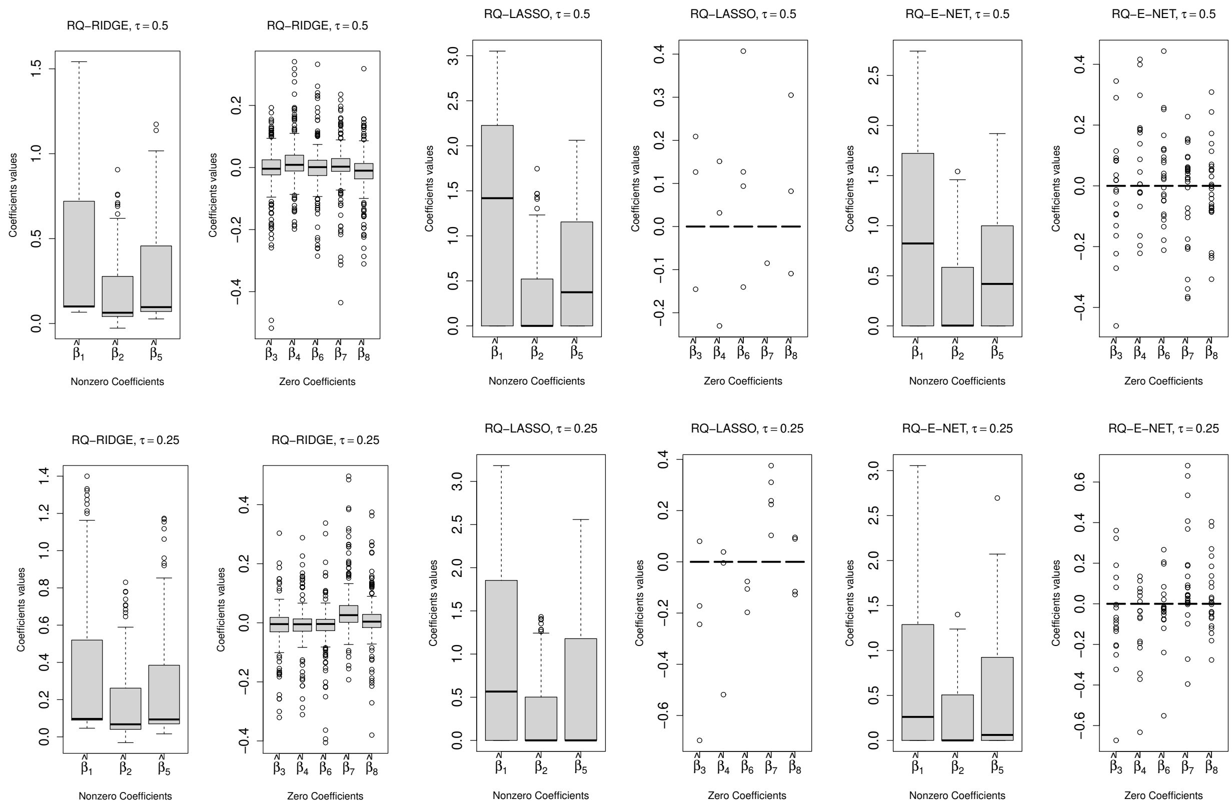


Figure 5.3: Box plots for collinearity inducing points scenario $D1$ under t-distribution with $\sigma = 1$, $d = 1$ at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted scenario).

Remarks. *The non-zero coefficients and zero coefficients are* $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ *and* $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$*, respectively. In Figure* 5.3*, the QR-LASSO and QR-E-NET procedures perform best with respect to correctly shrunk zero coefficients* $\{\beta_j : j = 3, 4, 6, 7, 8\}$*. Similarly, the QR-LASSO and QR-E-NET procedures also perform best with respect to the average of estimated non-zero coefficients* $\{\beta_j : j = 1, 2, 5\}$.

*Figure*s 5.4, 5.5, 5.6 and 5.7 show box plots of shrunken zero coefficients ($\beta_j : j = 3,4,6,7,8$) at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels. Under the Norma distribution scenario, penalized *WQR* procedures perform better than the unweighted ones in prediction (see *MAD* of test errors).

Table 5.3: Simulation results for $D2$ and $D4$ collinearity inducing scenarios under normal distribution at $\tau = 0.25$ and $\tau = 0.50$ quantile levels ($n = 50$); bold text indicate better performance.

			$\tau = 0.25$				$\tau = 0.50$			
			med(MAD)	Correctly	No. of Zeros		med(MAD)	Correctly	No. of Zeros	
Distribution	Parameter	Method	Test Error	Fitted	Correct	Incorrect	Test Error	Fitted	Correct	Incorrect
$D2-N(.,.)$	$\sigma = 1$	QR-RIDGE	**2.77(4.43)**	35.00	3.88	0.00	2.77(4.36)	0.00	2.77	0.00
		QR-LASSO	2.77(4.64)	**81.00**	**5.00**	0.19	2.77(4.38)	9.00	**5.00**	1.05
		QR-E-NET	2.77(4.62)	74.50	4.97	0.03	**2.77(4.36)**	**73.50**	**5.00**	0.27
		WQR-RIDGE	0.95(1.38)	14.00	3.54	0.00	-0.01(1.39)	12.50	3.62	0.00
		WQR-LASSO	**0.55(0.92)**	**92.00**	**4.92**	0.08	**-0.01(0.98)**	**98.00**	**4.99**	0.01
		WQR-E-NET	0.59(0.94)	65.50	4.53	0.00	-0.02(1.05)	84.00	4.84	0.00
	$\sigma = 3$	QR-RIDGE	**2.67(5.10)**	0.00	1.95	0.00	**-0.50(5.08)**	0.00	1.28	0.00
		QR-LASSO	0.81(6.17)	**19.00**	**4.19**	0.47	2.74(6.49)	**23.00**	**3.98**	0.27
		QR-E-NET	1.09(5.91)	12.00	3.32	0.14	-1.55(6.19)	4.50	2.56	0.13
		WQR-RIDGE	1.95(2.97)	13.00	3.32	0.02	-0.05(2.96)	6.00	3.11	0.01
		WQR-LASSO	**1.58(2.60)**	36.00	**4.72**	0.63	**-0.11(2.59)**	43.50	**4.84**	0,57
		WQR-E-NET	1.59(2.64)	**39.50**	**4.45**	0.41	-0.08(2.68)	**47.00**	4.70	0.45
$D4-N(.,.)$	$\sigma = 1$	QR-RIDGE	**1.24(4.39)**	0.00	2.43	0.00	-1.37(4.52)	0.00	1.08	0.00
		QR-LASSO	-1.15(5.63)	**35.00**	**4.08**	0.00	**-1.36(4.51)**	0.00	1.08	0.00
		QR-E-NET	-0.96(5.51)	7.00	2.68	0.00	-1.97(5.68)	0.00	1.94	0.00
		WQR-RIDGE	1.96(2.96)	13.50	3.31	0.02	-0.05(2.97)	6.00	3.11	0.01
		WQR-LASSO	**1.55(2.60)**	**38.50**	**4.76**	0.65	**-0.11(2.59)**	43.50	**4.84**	0.57
		WQR-E-NET	1.59(2.65)	34.00	4.40	0.43	-0.08(2.68)	**47.00**	4.70	0.45
	$\sigma = 3$	QR-RIDGE	3.17(5.08)	0.00	1.99	0.00	**-0.24(5.06)**	0.00	1.29	0.00
		QR-LASSO	**3.19(5.07)**	0.00	**5.00**	3.00	-0.16(5.10)	0.00	**5.00**	3.00
		QR-E-NET	**3.19(5.07)**	0.00	**5.00**	3.00	-0.16(5.10)	0.00	**5.00**	3.00
		WQR-RIDGE	1.68(2.69)	0.50	1.67	0.11	-0.06(2.62)	0.00	1.97	0.08
		WQR-LASSO	**1.42(2.48)**	**15.00**	**4.41**	0.94	**-0.09(2.28)**	**23.00**	**4.56**	0.70
		WQR-E-NET	**1.50(2.48)**	8.50	3.98	0.77	-0.05(2.33)	13.50	4.06	0.49

$N(.,.)$ denotes normally distributed.

Under the Gaussian distribution scenario, the *WQR-LASSO* procedure dominates in prediction, followed by the *WQR-E-NET* procedure in all scenarios. The *WQR-RIDGE* performs the worst in prediction in the penalized *WQR* scenarios. With respect to the percentage of correctly fitted models, the *WQR-LASSO* procedure performs best (followed by *WQR-E-NET*) with *WQR-E-NET* in fewer cases. *WQR-LASSO* performs best with respect to percentage of correctly fitted models at $\tau = 0.25$, except one when *WQR-E-NET* ($\sigma = 3$, $D2$). There is no clear 'winner' in performance with respect to the average correctly/incorrectly fitted zero coefficients at all *RQ* levels. Increasing the magnitude of the collinearity influential point from *D*2 to *D*4 greatly compromises the performance of all procedures in correctly fitting the models. The zero coefficients are correctly shrunk to zero/near zero in most cases, with varying degrees of accuracy.

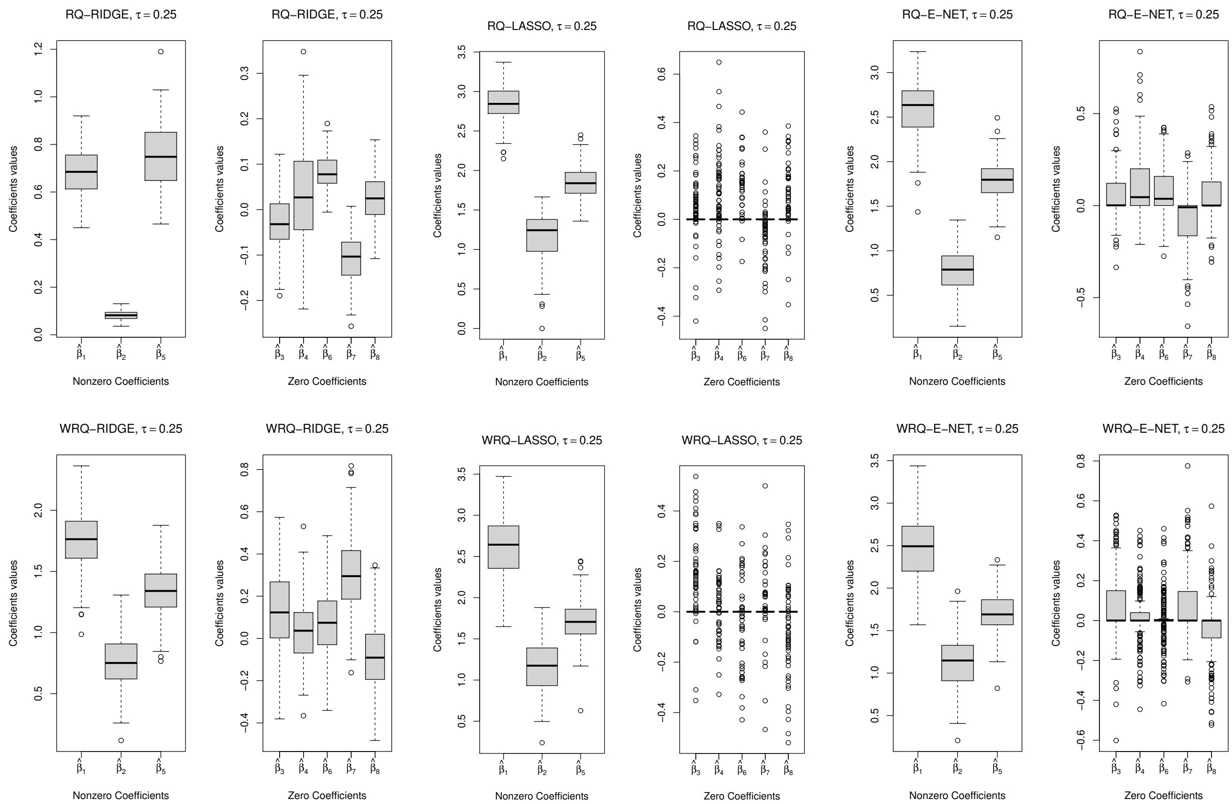


Figure 5.4: Box plots for collinearity inducing points scenario *D*2 under normal distribution with $\sigma = 1$ at $\tau = 0.25$ *RQ* level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Remarks. *The non-zero coefficients and zero coefficients are* $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ *and* $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$*, respectively. In Figure* 5.4*, the LASSO penalized procedures perform best with respect to correctly shrunk zero coefficients* $\{\beta_j : j = 3, 4, 6, 7, 8\}$.

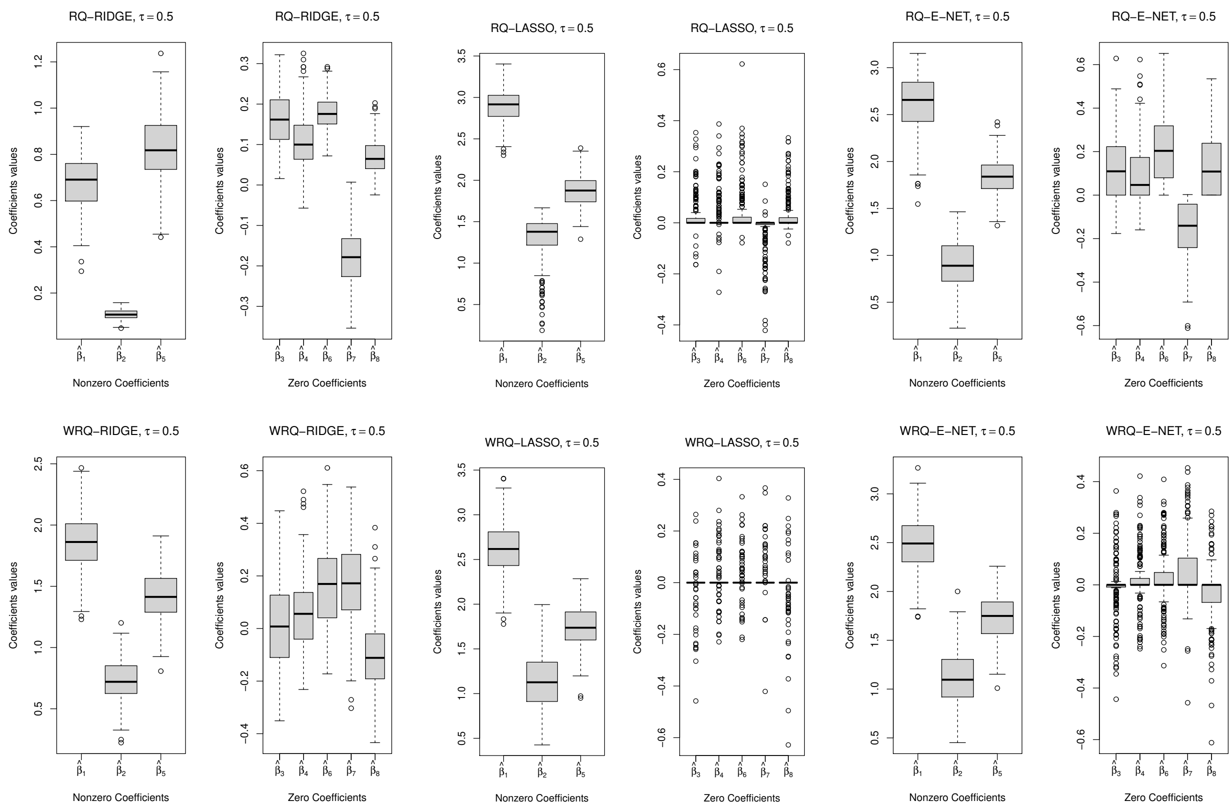


Figure 5.5: Box plots for collinearity inducing points scenario $D2$ under normal distribution with $\sigma = 1$ at $\tau = 0.50$ *RQ* level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Remarks. *In Figure* 5.5*, the LASSO penalized procedures perform best with respect to correctly shrunk zero coefficients* $\{\beta_j : j = 3, 4, 6, 7, 8\}$.

Table 5.4 summarizes the simulation results for the t-distribution case in the presence of collinearity inducing points ($D2$ and $D4$). The results given in *Tables* 5.4 are also presented graphically in *Figure* 5.8 as an alternative illustration. The *QR-LASSO* performs best among the unweighted procedures with respect to prediction (see *MAD* of test errors). The *WQR-LASSO* performs the best with respect to percentage of correctly fitted models and the average of correctly fitted zero coefficients. In all pairwise comparisons, generally, the weighted penalized

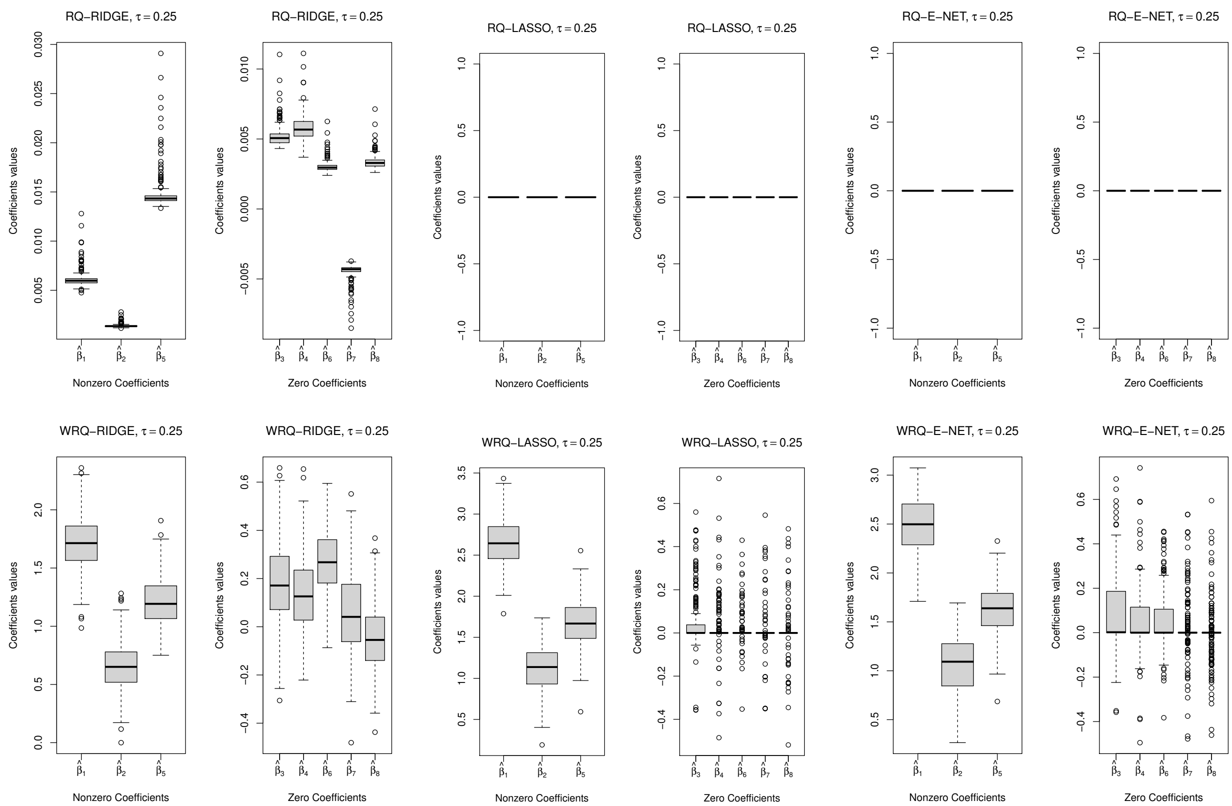


Figure 5.6: Box plots for collinearity inducing points scenario $D4$ under normal distribution with $\sigma = 1$ at $\tau = 0.25$ *RQ* level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Remarks. *In Figure* 5.6*, the LASSO penalized procedures perform best with respect to correctly shrunk zero coefficients* $\{\beta_j : j = 3, 4, 6, 7, 8\}$*. All unweighted penalized QR procedures performed poorly in terms of correctly shrinking zero coefficients. The penalized WQR procedures perform better than the unweighted penalized procedures.*

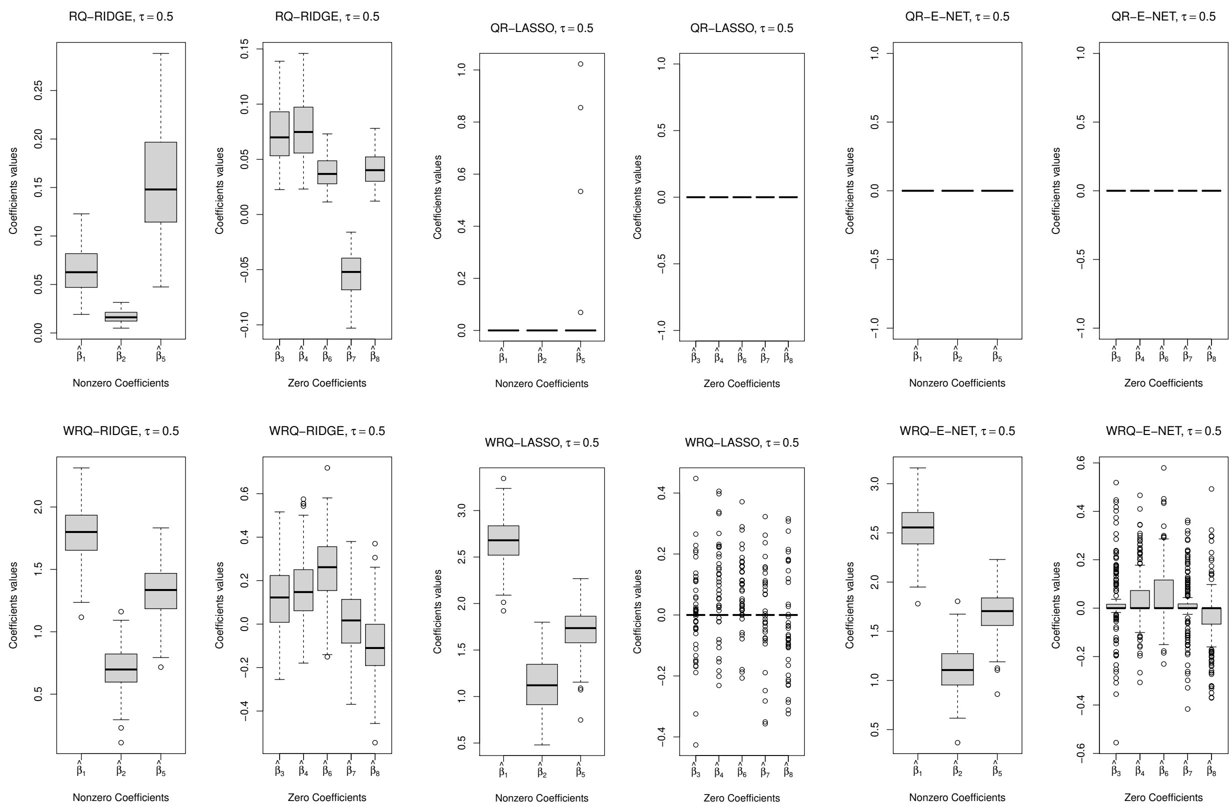


Figure 5.7: Box plots for collinearity inducing points scenario $D4$ under normal distribution with $\sigma = 1$ at $\tau = 0.50$ *RQ* level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Remarks. *In Figure* 5.7, *the LASSO penalized procedures perform best with respect to correctly shrunk zero coefficients* $\{\beta_j : j = 3, 4, 6, 7, 8\}$. *All unweighted penalized QR procedures performed poorly in terms of correctly shrinking zero coefficients. The penalized WQR procedures perform better than the unweighted penalized procedures.*

Figure 5.8: Performance at $D2$ and $D4$ under the t_6 and t_1 distributions with the *RIDGE* and *E-NET* on the *LHS* and *RHS* of *LASSO*, respectively; Upper panel: Model/Variable selection showing the proportion of correct models and the average of correct/incorrect βs selected; Lower panel: Prediction metrics.

*QR*s outperform the unweighted ones with respect to the average of correctly shrunk zero coefficients. With respect to the percentage of correctly fitted models, the weighted procedures dominate most of the time, with the exception of the *WQR-RIDGE* procedure, which performs equally with its unweighted counterpart. Most penalized procedures perform poorly, except *WQR-LASSO* with respect to percentage of correctly fitted models.

Table 5.4: Simulation results for $D2$ and $D4$ collinearity inducing scenarios under t-distribution at $\tau = 0.25$ and $\tau = 0.50$ quantile levels ($n = 50$); bold text indicate better performance.

			$\tau = 0.25$				$\tau = 0.50$				
			Med(MAD)	Correctly	No. of Zeros		med(MAD)	Correctly	No. of Zeros		
Distribution	Parameter	Method	Test Error	Fitted	Correct	Incorrect	Test Error	Fitted	Correct	Incorrect	
$D2-t_d$	$d=1,\sigma=0.50$	QR-RIDGE	3.03(4.31)	0.00	0.99	0.03	0.06(4.38)	0.00	0.95	0.00	
		QR-LASSO	**1.38(2.68)**	1.00	**3.79**	0.76	**-0.02(2.06)**	0.50	**3.19**	0.54	
		QR-E-NET	1.86(3.13)	0.00	1.95	0.56	0.00(2.78)	0.00	1.96	0.39	
		WQR-RIDGE	**1.81(3.85)**	0.00	1.05	0.00	**0.51(3.82)**	0.00	0.87	0.00	
		WQR-LASSO	1.64(4.18)	**47.00**	**4.69**	0.63	0.73(4.13)	**53.50**	**4.66**	0.40	
		WQR-E-NET	1.71(4.03)	12.00	3.81	0.46	0.66(4.02)	4.00	3.28	0.33	
	$d=1,\sigma=1$	QR-RIDGE	3.56(5.33)	0.00	0.96	0.10	-0.05(5.23)	0.00	0.91	0.01	
		QR-LASSO	2.21(5.97)	3.50	**4.73**	1.74	**-0.10(4.01)**	1.00	**3.64**	1.22	
		QR-E-NET	**2.95(4.68)**	0.00	2.47	1.03	0.02(4.49)	0.00	2.37	0.85	
		WQR-RIDGE	2.00(3.92)	0.00	0.91	0.01	**0.27(3.99)**	0.00	0.64	0.00	
		WQR-LASSO	**1.64(2.69)**	23.50	**4.96**	1.70	0.43(4.19)	**35.50**	**4.72**	0.98	
		WQR-E-NET	2.96(4.69)	0.00	2.49	1.06	0.38(4.05)	11.00	3.85	0.81	
$D4-t_d$	$d=1,\sigma=0.50$	QR-RIDGE	3.38(5.06)	0.00	0.79	1.00	0.03(5.14)	0.00	0.82	1.00	
		QR-LASSO	**2.15(3.59)**	0.00	**3.00**	1.81	**-0.01(4.10)**	0.00	**3.18**	1.96	
		QR-E-NET	3.41(4.83)	0.00	1.09	1.07	0.07(4.99)	0.00	1.36	1.25	
		WQR-RIDGE	1.80(3.85)	0.00	1.04	0.00	0.51(3.82)	0.00	0.87	0.00	
		WQR-LASSO	1.65(4.13)	**45.00**	**4.68**	0.64	0.71(4.15)	**51.50**	**4.64**	0.41	
		WQR-E-NET	1.70(4.00)	12.00	3.80	0.47	0.68(4.01)	5.50	3.30	0.33	
	$d=1,\sigma=1$	QR-RIDGE	3.74(5.66)	0.00	0.71	1.00	0.00(5.67)	0.00	0.76	1.00	
		QR-LASSO	2.65(4.53)	0.00	3.10	1.85	-0.05(5.03)	0.00	3.51	2.29	
		QR-E-NET	3.70(5.41)	0.00	1.16	1.12	-0.06(5.51)	0.00	1.60	1.41	
		WQR-RIDGE	2.01(3.94)	0.00	0.93	0.01	0.26(3.99)	0.00	0.65	0.00	1
		WQR-LASSO	1.97(4.10)	23.00	**4.78**	1.49	0.43(4.15)	**36.50**	**4.73**	0.97	
		WQR-E-NET	2.00(4.03)	15.50	**4.35**	1.27	0.39(4,02)	10.50	3.83	0.79	
$D2-t_d$	$d=6,\sigma=0.50$	QR-RIDGE	2.62(3.26)	0.00	1.00	0.00	-0.02(3.37)	0.00	1.00	0.00	
		QR-LASSO	0.54(0.91)	3.50	3.19	0.00	0.02(0.85)	3.00	2.95	0.00	
		QR-E-NET	0.75(1.31)	0.00	1.94	0.00	0.02(1.14)	0.00	1.82	0.00	
		WQR-RIDGE	1.37.3.71)	0.00	1.56	0.00	1.02(3.72)	0.00	1.31	0.00	
		WQR-LASSO	1.32(4.23)	**47.50**	**4.26**	0.00	1.08(4.24)	**52.00**	**4.25**	0.00	
		WQR-E-NET	1.20(4.12)	4.50	2.96	0.00	1.01(4.15)	3.50	2.82	0.00	
	$d=6,\sigma=1$	QR-RIDGE	2.62(3.54)	0.00	1.00	0.00	0.05(3.61)	0.00	1.00	0.00	
		QR-LASSO	0.97(1.63)	3.00	2.98	0.02	0.03(1.55)	1.50	2.80	0.00	
		QR-E-NET	1.28(2.04)	0.00	1.81	0.04	0.04(1.90)	0.00	1.76	0.01	
		WQR-RIDGE	1.46(3.75)	0.00	1.31	0.00	0.82(3.70)	0.00	1.03	0.00	
		WQR-LASSO	1.48(4.09)	**43.50**	**4.13**	0.01	0.98(4.12)	**46.00**	**4.14**	0.00	
		WQR-E-NET	1.43(4.00)	6.50	3.10	0.00	0.96(4.00)	5.00	2.88	0.00	
$D4-td$	$d=6,\sigma=0.50$	QR-RIDGE	3.16(4.45)	0.00	0.97	1.00	0.09(4.48)	0.00	1.00	1.00	
		QR-LASSO	1.65(2.67)	0.00	2.69	1.49	0.01(2.61)	0.00	2.76	1.27	
		QR-E-NET	3.20(4.26)	0.00	1.00	1.00	0.05(4.39)	0.00	1.02	1.02	
		WQR-RIDGE	1.36(3.71)	0.00	1.55	0.00	1.02(3.72)	0.00	1.30	0.00	
		WQR-LASSO	1.32(4.23)	**49.00**	**4.27**	0.00	1.06(4.25)	**51.50**	**4.29**	0.00	
		WQR-E-NET	1.21(4.13)	4.50	2.96	0.00	1.00(4.14)	4.50	2.86	0.00	
	$d=6,\sigma=1$	QR-RIDGE	3.20(4.61)	0.00	0.91	1.00	0.07(4.63)	0.00	0.98	1.00	
		QR-LASSO	1.83(2.83)	0.00	2.62	1.44	0.07(3.21)	0.00	2.81	1.38	
		QR-E-NET	3.16(4.36)	0.00	0.99	1.00	0.04(4.42)	0.00	1.02	1.01	
		WQR-RIDGE	1.46(3.74)	0.00	1.35	0.00	0.81(3.70)	0.00	1.03	0.00	
		WQR-LASSO	1.47(4.12)	**43.00**	**4.12**	0.01	0.98(4.11)	**45.00**	**4.13**	0.00	
		WQR-E-NET	1.46(3.97)	5.50	3.02	0.00	0.96(4.01)	7.00	2.83	0.00	

t_d denotes t-distribution with d degrees of freedom.

Collinearity reducing scenario (D3 and D5)

In *Table* 5.5, we examine the performance of regularized *QR* procedures (weighted and unweighted) in the collinearity reducing points scenarios $D3$ and $D5$ under the Gaussian distribution. The zero coefficients ($\beta_3, \beta_4, \beta_6, \beta_7, \beta_8$) are penalized to zero/near zero as expected (see *Figures* 5.9, 5.10, 5.11 and 5.12). The *WQR-LASSO* performs the best in terms of the *MAD* of test errors, followed by *WQR-E-NET*, though they perform more or less equally in $D5$.

Table 5.5: Simulation results for $D3$ and $D5$ collinearity reducing points scenarios under normal distribution at $\tau = 0.25$ and $\tau = 0.50$ quantile levels ($n = 50$); bold text indicate better performance.

			$\tau = 0.25$				$\tau = 0.50$			
			med(MAD)	Correctly	No. of Zeros		med(MAD)	Correctly	No. of Zeros	
Distribution	Parameter	Method	Test Error	Fitted	Correct	Incorrect	Test Error	Fitted	Correct	Incorrect
$D3-N(.,.)$	$\sigma = 1$	QR-RIDGE	3.22(4.65)	0.50	2.38	0.00	-0.10(4.66)	0.00	1.02	0.00
		QR-LASSO	3.29(4.65)	0.00	**5.00**	3.00	0.04(4.68)	0.00	**5.00**	2.86
		QR-E-NET	**3.29(4.65)**	0.00	**5.00**	3.00	**-0.04(4.64)**	0.00	**5.00**	2.87
		WQR-RIDGE	0.92(1.20)	4.00	2.74	0.00	0.00(1.09)	2.50	2.74	0.00
		WQR-LASSO	**0.45(0.76)**	**62.50**	**4.40**	0.00	**-0.01(0.75)**	**72.00**	**4.64**	0.00
		WQR-E-NET	0.48(0.80)	33.50	3.88	0.00	-001(0.78)	34.50	3.99	0.00
	$\sigma = 3$	QR-RIDGE	**3.38(5.38)**	0.50	1.92	0.00	**-0.22(5.38)**	0.00	1.25	0.01
		QR-LASSO	3.34(5.40)	0.00	**4.99**	2.85	-0.16(5.44)	0.00	**5.00**	2.66
		QR-E-NET	3.34(5.39)	0.00	**4.98**	2.80	-0.15(5.40)	0.00	4.99	2.63
		WQR-RIDGE	1.88(2.67)	3.00	2.63	0.04	-0.03(2.54)	3.00	2.67	0.06
		WQR-LASSO	**1.40(2.33)**	**36.00**	**4.55**	0.49	**-0.03(2.24)**	34.50	**4.65**	0.51
		WQR-E-NET	1.55(2.45)	26.50	4.14	0.41	-0.07(2.33)	29.50	4.34	0.42
$D5-N(.,.)$	$\sigma = 1$	QR-RIDGE	3.22(4.65)	0.50	2.38	0.00	-0.10(4.66)	0.00	1.02	0.00
		QR-LASSO	3.29(4.65)	0.00	**5.00**	3.00	**-0.03(4.64)**	0.00	**5.00**	2.97
		QR-E-NET	3.29(4.65)	0.00	**5.00**	3.00	-0.03(4.65)	0.00	**5.00**	3.00
		WQR-RIDGE	1.71(2.09)	1.50	2.54	0.01	0.00(1.62)	2.50	3.20	0.02
		WQR-LASSO	**1.57(2.02)**	1.00	**4.94**	1.33	**0.00(1.07)**	**48.50**	**4.66**	0.21
		WQR-E-NET	**1.54(2.01)**	0.50	4.93	1.34	0.00(1.16)	32.00	4.35	0.16
	$\sigma = 3$	QR-RIDGE	3.38(5.39)	0.50	1.92	0.00	**-0.22(5.39)**	0.00	1.25	0.01
		QR-LASSO	3.43(5.39)	0.00	**5.00**	3.00	-0.14(5.41)	0.00	**5.00**	2.97
		QR-E-NET	3.43(5.39)	0.00	**5.00**	3.00	-0.14(5.41)	0.00	**5.00**	3.00
		WQR-RIDGE	2.16(2.90)	2.00	2.51	0.02	-0.03(2.71)	2.00	2.57	0.06
		WQR-LASSO	**1.95(2.77)**	2.00	**4.93**	1.80	**-0.03(2.54)**	1.00	**4.88**	1.48
		WQR-E-NET	2.09(2.87)	8.00	4.85	1.59	-0.01(2.66)	4.50	4.72	1.33

$N(.,.)$ denotes normally distributed.

Although *WQR-LASSO* performs the best under $D3$ at $\tau = 0.25$, the performance of all penalized procedures is very poor in $D5$ at all quantile levels. However, the unweighted methods are superior in terms of the average of correct zero coefficients, except for the *RIDGE* penalized procedure, which fluctuates in performance. The performance of the unweighted methods is generally weak throughout in terms of the percentage of correctly fitted models.

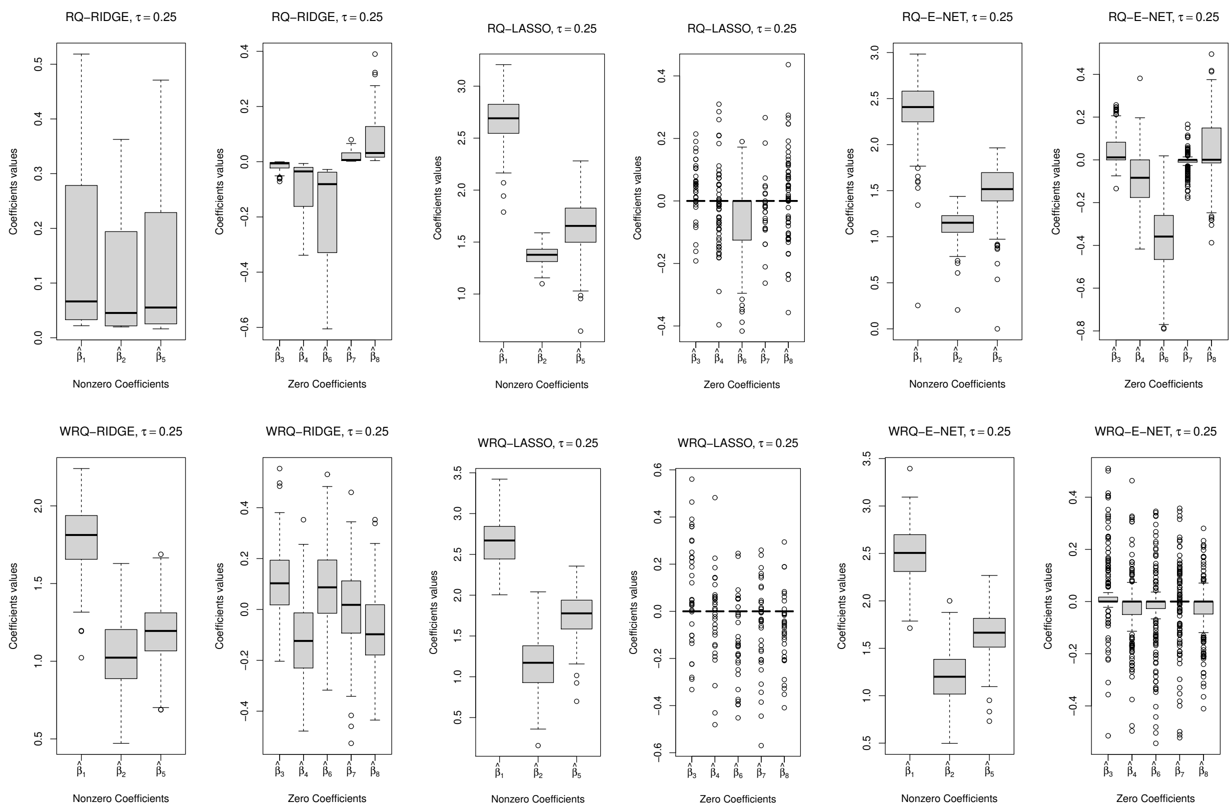


Figure 5.9: Box plots for $D3$-normal distribution with $\sigma = 1$. Collinearity hiding points at $\tau = 0.25$ quantile level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

In *Table* 5.6, we examine the performance of regularized *QR* procedures (both weighted and unweighted) in the collinearity reducing points scenarios $D3$ and $D5$ under the t-distribution. The results given in *Tables* 5.6 are also presented graphically in *Figure* 5.13 as an alternative illustration. Under the t-distribution scenario, the weighted regularized procedures dominate the prediction performance, with *WQR-LASSO* performing the best, followed by *WQR-E-NET*. With respect to average number of correctly fitted zero coefficients and percentage of correctly fitted models, the *WQR-LASSO* procedure performs the best. In the unweighted scenario, *QR-LASSO* outperforms other unweighted versions across all metrics.

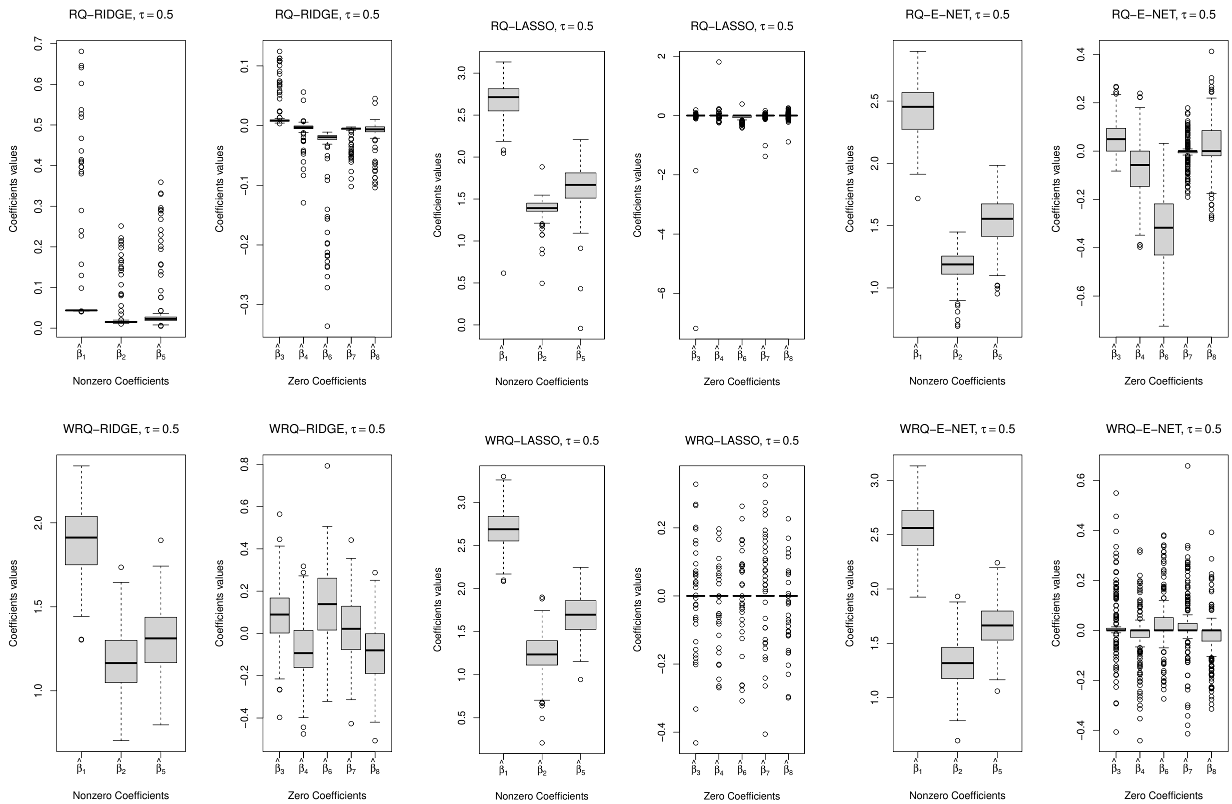


Figure 5.10: Box plots for $D3$ under normal distribution . Collinearity hiding points at $\tau = 0.50$ quantile level with $\sigma = 1$. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Heavy-tailed distribution scenario (D6)

Table 5.7 summarizes the simulation results of penalized QR procedures (weighted and unweighted) for the design matrix $D6$ under the t-distribution, where the collinearity in the design matrix is generated by the exponential decay $v_{ij} = \rho^{|j-i|}$, coupled with high leverage points. The results given in *Tables* 5.7 are also presented graphically in *Figure* 5.14 as an alternative illustration. In the pairwise comparisons, the unweighted procedures outperformed the weighted ones in prediction, with the *WQR-LASSO* dominating the weighted penalized procedures and the *QR-LASSO* dominating the weighted penalized ones. The *WQR-LASSO* performs best with the remaining metrics (percentage of correctly fitted models and average of correctly fitted zero coefficients).

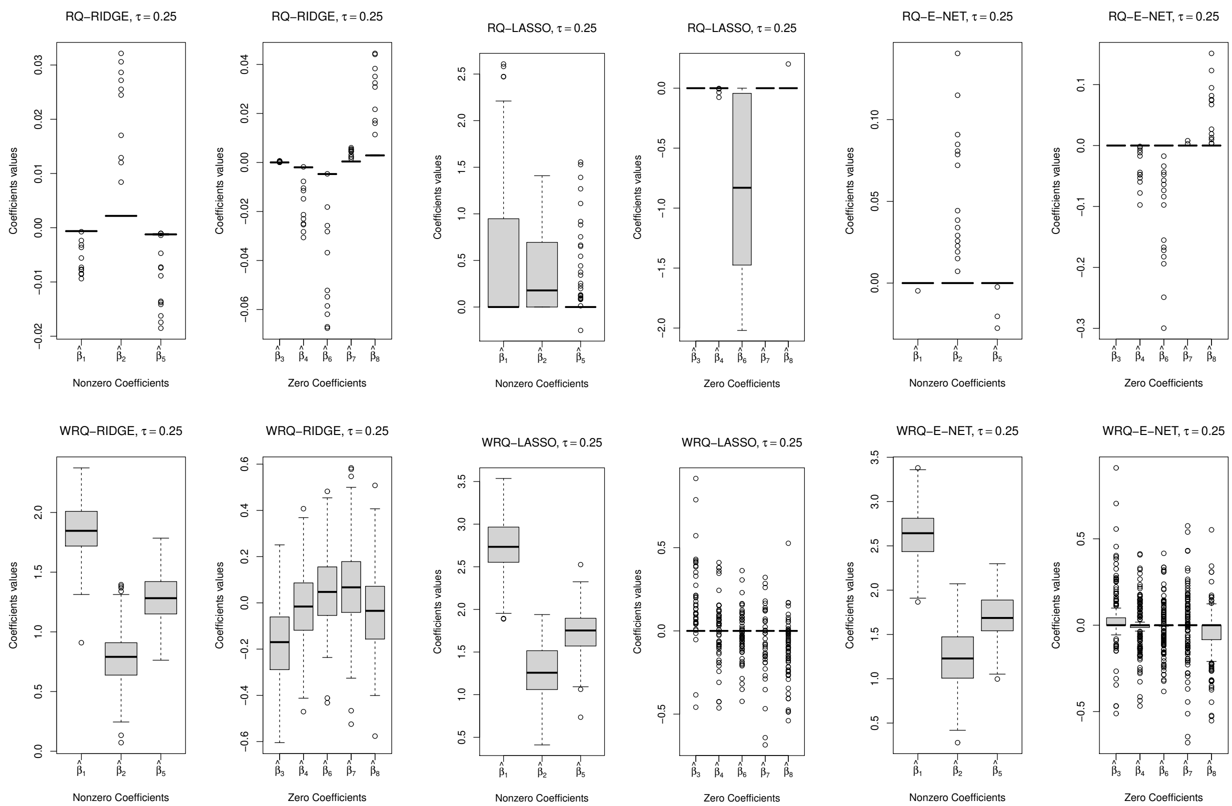


Figure 5.11: Box plots for $D5$ under normal distribution with $\sigma = 1$. Collinearity hiding points at $\tau = 0.25$ quantile level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

The zero coefficients $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)$ are correctly penalized to zero/near zero as expected (see *Figures* 5.15 and 5.16).

5.3 Applications of Penalized Weighted Quantile Regression to Well-known Data Sets from the Literature

In this section, we apply our penalized procedures to well-known data sets from the literature and check their performance in terms of correctly shrinking the correct zero coefficients. We also use the coefficient estimation biases to measure the performance of these penalized procedures.

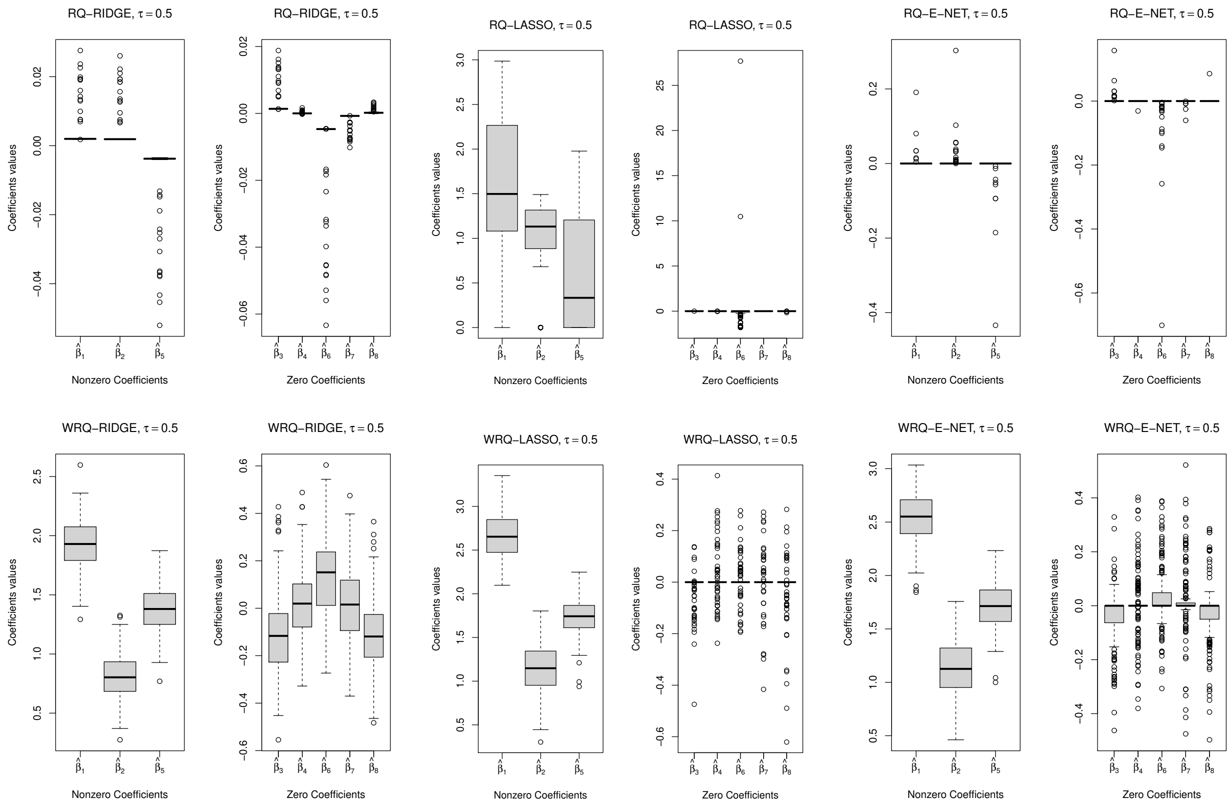


Figure 5.12: Box plots for $D5$ under normal distribution with $\sigma = 1$. Collinearity hiding points at $\tau = 0.50$ quantile level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

The two real data sets considered are the Hawkins et al. (1984) and the Hocking and Pendleton (Hocking & Pendleton 1983) data sets.

5.3.1 Hawkins, Bradu and Kass Data Set

We employ (Hawkins et al. 1984) artificial data with outlying points. The artificial data consist of 75 observations composed of three predictor variables (X_1, X_2, X_3) and the response variable, Y.

Table 5.6: Simulation results for $D3$ and $D5$ collinearity reducing points scenarios under t-distribution with d=(1 *and* 6) degrees of freedom at $\tau = 0.25$ and $\tau = 0.50$ quantile levels; bold text indicate better performance.

			$\tau = 0.25$				$\tau = 0.50$ (LAD)			
			Med(MAD)	Correctly	No. of Zeros		med(MAD)	Correctly	No. of Zeros	
Distribution	Parameter	Method	Test Error	Fitted	Correct	Incorrect	Test Error	Fitted	Correct	Incorrect
$D3-t_d$	$d=1, \sigma=0.50$	QR-RIDGE	3.17(5.26)	0.00	2.94	0.02	-0.04(5.22)	0.00	2.94	0.01
		QR-LASSO	**1.38(2.74)**	34.50	**4.57**	0.75	**-0.03(2.06)**	**58.50**	**4.71**	0.46
		QR-E-NET	2.23(3.79)	8.00	4.06	0.85	-0.05(2.30)	3.00	3.56	0.51
		WQR-RIDGE	1.77(2.43)	0.00	4.00	0.01	0.00(1.79)	0.50	2.19	0.00
		WQR-LASSO	**1.00(1.82)**	**48.50**	**4.86**	0.91	**-0.01(1.27)**	**61.00**	**4.91**	0.63
		WQR-E-NET	1.11(1.95)	24.00	4.40	0.85	-0.02(1.36)	37.00	4.40	0.58
	$d=1, \sigma=1$	QR-RIDGE	3.58(5.65)	0.00	2.81	0.11	-0.07(5.62)	0.00	2.94	0.03
		QR-LASSO	**2.69(4.75)**	20.00	**4.79**	1.49	**-0.02(3.55)**	22.00	**4.78**	1.01
		QR-E-NET	3.18(5.28)	5.50	4.58	1.91	-0.06(4.23)	4.00	4.16	1.31
		WQR-RIDGE	2.28(2.98)	1.50	2.69	0.02	0.00(2.63)	1.00	2.49	0.01
		WQR-LASSO	**1.76(2.83)**	24.50	**4.94**	1.70	**-0.02(2.26)**	29.00	**4.95**	1.41
		WQR-E-NET	1.91(2,89)	17.50	4.76	1.66	0.00(2.39)	20.50	4.70	1.36
$D5-t_d$	$d=1, \sigma=0.50$	QR-RIDGE	3.21(5.30)	0.00	2.10	2.00	-0.03(5.33)	0.00	2.19	1.01
		QR-LASSO	**3.09(5.23)**	1.50	**5.00**	2.42	**0.04(4.27)**	**33.50**	**4.99**	1.79
		QR-E-NET	3.15(5.29)	0.00	4.93	2.74	0.00(5.32)	0.00	4.98	2.94
		WQR-RIDGE	1.76(2.42)	0.00	2.46	0.01	0.00(1.82)	0.50	2.23	0.00
		WQR-LASSO	**1.01(1.81)**	**49.50**	**4.86**	0.90	**-0.01(1.28)**	**62.00**	**4.93**	0.63
		WQR-E-NET	1.13(1.98)	26.00	4.40	0.86	-0.01(1.36)	34.00	4.39	0.57
	$d=1, \sigma=1$	QR-RIDGE	**3.61(5.72)**	0.00	2.10	2.00	-0.06(5.72)	0.00	2.21	1.01
		QR-LASSO	3.55(5.80)	1.00	**5.00**	2.55	**-0.04(5.32)**	13.00	**5.00**	2.50
		QR-E-NET	3.56(5.73)	0.00	4.96	2.81	-0.04(5.73)	0.00	4.99	2.95
		WQR-RIDGE	2.29(2.98)	1.50	2.69	0.02	0.00(2.64)	1.50	2.48	0.02
		WQR-LASSO	**1.77(2.83)**	23.50	**4.93**	1.71	**-0.03(2.29)**	29.00	**4.95**	1.43
		WQR-E-NET	1.92(2.89)	16.50	4.76	1.67	0.00(2.38)	21.50	4.72	1.34
$D3-t_d$	$d=6, \sigma=0.50$	QR-RIDGE	2.90(4.83)	0.00	3.00	0.00	-0.05(4.84)	0.00	2.95	0.00
		QR-LASSO	**0.56(0.85)**	**48.50**	**4.23**	0.00	**0.02(0.80)**	**67.00**	**4.54**	0.02
		QR-E-NET	1.12(1.75)	0.50	3.08	0.00	-0.01(0.88)	2.00	2.92	0.00
		WQR-RIDGE	0.95(1.21)	0.00	1.80	0.00	0.00(0.90)	0.00	1.75	0.00
		WQR-LASSO	**0.50(0.63)**	**63.50**	**4.49**	0.00	**0.00(0.54)**	**80.00**	**4.78**	0.00
		WQR-E-NET	0.53(0.65)	14.00	3.50	0.00	0.01(0.55)	21.00	3.64	0.00
	$d=6, \sigma=1$	QR-RIDGE	3.12(4.84)	0.00	3.00	0.00	-0.04(4.83)	0.00	2.99	0.00
		QR-LASSO	**1.06(1.60)**	**39.50**	**3.82**	0.01	**0.00(1.51)**	**55.50**	**4.35**	0.03
		QR-E-NET	1.54(2.37)	5.50	3.27	0.00	-0.02(1.79)	2.00	3.07	0.00
		WQR-RIDGE	1.31(1.53)	0.50	2.16	0.00	0.00(1.31)	0.00	1.93	0.00
		WQR-LASSO	**0.75(1.01)**	**56.00**	**4.35**	0.02	**0.02(0.92)**	**63.50**	**4.53**	0.00
		WQR-E-NET	0.78(1.03)	25.00	3.66	0.00	0.01(0.92)	19.00	3.56	0.00
$D5-t_d$	$d=6, \sigma=0.50$	QR-RIDGE	2.96(5.02)	0.00	2.03	2.00	-0.06(5.02)	0.00	2.01	1.00
		QR-LASSO	**2.96(4.86)**	1.00	**5.00**	2.13	**-0.01(2.02)**	**58.50**	**4.98**	0.94
		QR-E-NET	2.94(4.98)	0.00	4.93	2.78	0.00(4.99)	0.00	4.98	2.93
		WQR-RIDGE	0.96(1.21)	0.00	1.80	0.00	0.00(0.90)	0.00	1.78	0.00
		WQR-LASSO	**0.50(0.63)**	**63.50**	**4.49**	0.00	**0.00(0.53)**	**78.50**	**4.76**	0.00
		WQR-E-NET	0.54(0.65)	14.00	3.48	0.00	0.01(0.55)	20.50	3.66	0.00
	$d=6, \sigma=1$	QR-RIDGE	3.18(4.96)	0.00	2.11	2.00	-0.03(4.99)	0.00	2.07	1.00
		QR-LASSO	**3.08(4.90)**	1.00	**5.00**	2.33	**-0.13(3.83)**	34.00	**4.99**	1.71
		QR-E-NET	3.11(5.00)	0.00	4.93	2.83	0.00(5.00)	0.00	**4.99**	2.93
		WQR-RIDGE	1.31(1.52)	0.50	2.17	0.00	0.00(1.30)	0.50	1.93	0.00
		WQR-LASSO	**0.75(1.01)**	**56.50**	**4.36**	0.02	**0.01(0.91)**	**60.00**	**4.46**	0.00
		WQR-E-NET	0.78(1.03)	25.50	3.66	0.00	0.01(0.92)	19.50	3.58	0.00

t_d denotes t-distribution with d degrees of freedom.

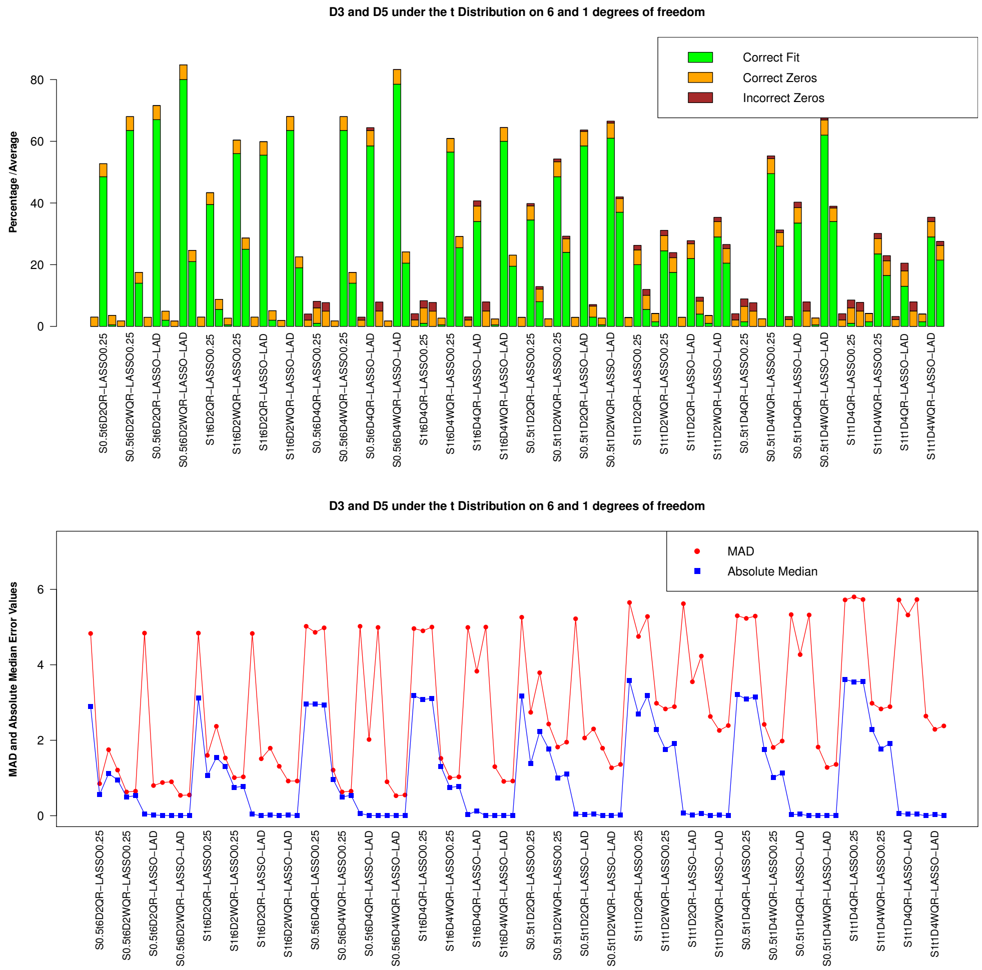
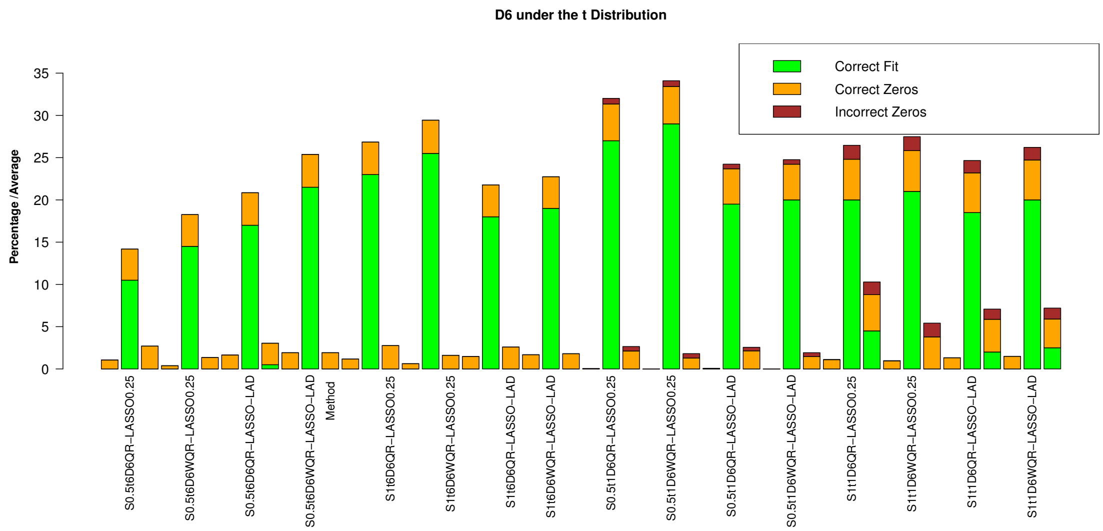


Figure 5.13: Performance at $D3$ and $D5$ under the t_6 and t_1 distributions distributions with the *RIDGE* and *E-NET* on the *LHS* and *RHS*, respectively, with the *RIDGE* and *E-NET* on the *RHS* and *LHS* of *LASSO*, respectively; Upper panel: Model/Variable selection showing the proportion of correct models and the average of correct/incorrect βs selected; Lower panel: Prediction metrics.

Remarks. *The prediction pattern under the t-distributions exhibited under* D3 *and* D5 *in Figure* 5.13 *(lower panel) are different to that exhibited at* D2 *and* D4 *(Figure* 5.8*) in that the MAD of measure is more erratic (but the absolute median error is less erratic) at* D3 *and* D5 *but absolute. In fact, the prediction picture of QR-LASSO and WQR-LASSO at* $\tau = 0.50$ *based on the absolute median error are similar at* D3 *and* D5*, whereas, at* D2 *and* D4*, WQR-LASSO performs better with respect to this measure.*



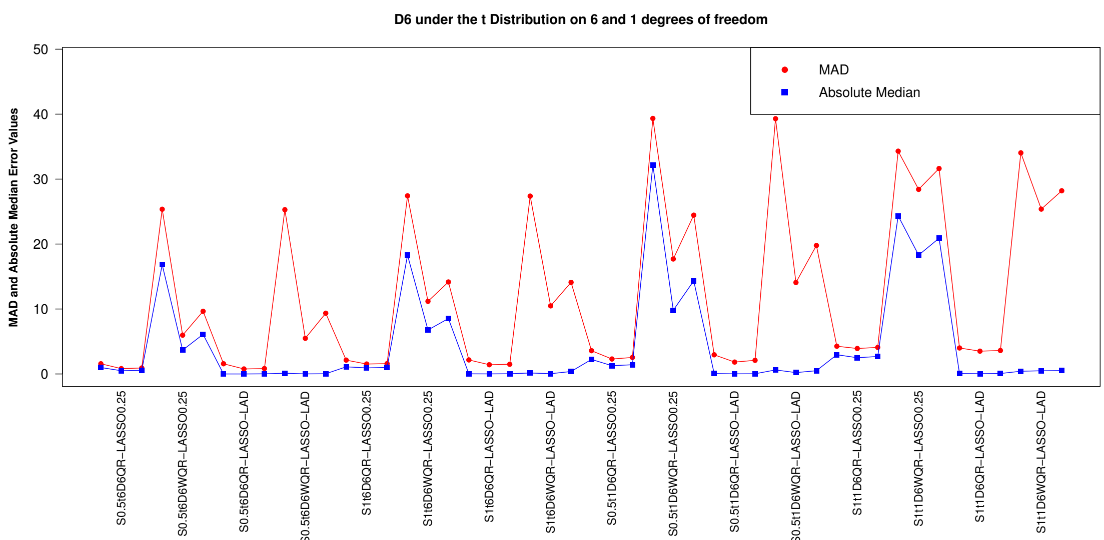


Figure 5.14: Performance at $D6$ under the t_6 and t_1 distributions with the *RIDGE* and *E-NET* on the *LHS* and *RHS* of *LASSO*, respectively; Model/Variable selection showing the proportion of correct models and the average of correct/incorrect βs selected; Lower panel: Prediction metrics.

Remarks. *The prediction pattern under the t-distributions exhibited under D*6 *in Figure* 5.14 *(lower panel) is slightly poorer under the* t_1*-distribution compared to that exhibited under the* t_6*-distribution.*

Table 5.7: Simulation results $D6$ collinearity by $0.5^{|j-i|}$ exponential decay under t-distribution with d=(1 and 6) degrees of freedom at $\tau = 0.25$ and $\tau = 0.50$ quantile levels; bold text indicate better performance.

			$\tau = 0.25$				$\tau = 0.50$				
			Med(MAD)	Correctly	No. of Zeros		med(MAD)	Correctly	No. of Zeros		
Distribution	Parameter	Method	Test Error	Fitted	Correct	Incorrect	Test Error	Fitted	Correct	Incorrect	
			n=50(m=5)								
$D6-t_d$	$d=1, \sigma=0.50$	QR-RIDGE	2.24(3.58)	0.00	0.04	0.00	0.06(2.95)	0.00	0.07	0.00	
		QR-LASSO	**1.27(2.30)**	**27.00**	**4.35**	0.66	**0.02(1.83)**	19.50	**4.18**	0.55	
		QR-E-NET	1.42(2.54)	0.00	2.12	0.53	0.03(2.10)	0.00	2.14	0.42	
		WQR-RIDGE	32.17(39.34)	0.00	0.00	0.00	0.62(39.30)	0.00	0.00	0.00	
		WQR-LASSO	**9.81(17.70)**	**29.00**	**4.42**	0.68	**0.21(14.08)**	20.00	**4.22**	0.54	
		WQR-E-NET	14.34(24.45)	0.00	1.28	0.53	0.50(19.78)	0.00	1.46	0.46	
	$d=1, \sigma=1$	QR-RIDGE	2.96(4.27)	0.00	1.09	0.01	0.06(4.00)	0.00	1.32	0.00	
		QR-LASSO	**2.50(3.92)**	**20.00**	**4.81**	1.65	**0.04(3.51)**	18.50	**4.70**	1.46	
		QR-E-NET	2.69(4.09)	4.50	**4.29**	1.50	0.07(3.61)	2.00	3.87	1.21	
		WQR-RIDGE	24.34(34.30)	0.00	0.96	0.01	0.42(34.04)	0.00	1.49	0.00	
		WQR-LASSO	**18.30(28.42)**	21.00	**4.83**	1.65	**0.49(25.38)**	20.00	**4.72**	1.49	
		WQR-E-NET	20.89(31.63)	0.00	3.79	1.63	0.52(28.20)	2.50	3.40	1.30	
$D6-t_d$	$d=6, \sigma=0.50$	QR-RIDGE	0.99(1.57)	0.00	1.06	0.00	0.00(1.57)	0.00	1.65	0.00	
		QR-LASSO	**0.50(0.82)**	10.50	**3.68**	0.00	**0.00(0.77)**	17.00	**3.85**	0.00	
		QR-E-NET	0.56(0.92)	0.00	2.70	0.00	0.02(0.83)	0.50	2.54	0.00	
		WQR-RIDGE	16.82(25.36)	0.00	0.38	0.00	-0.11(25.29)	0.00	1.92	0.00	
		WQR-LASSO	**3.67(5.97)**	14.50	**3.76**	0.00	**0.01(5.49)**	21.50	**3.88**	0.00	
		WQR-E-NET	6.10(9.65)	0.00	1.35	0.00	-0.03(9.35)	0.00	1.92	0.00	
	$d=6, \sigma=1$	QR-RIDGE	1.10(2.13)	0.00	1.17	0.00	0.02(2.16)	0.00	1.47	0.00	
		QR-LASSO	**0.93(1.53)**	**23.00**	**3.84**	0.00	**0.01(1.42)**	18.00	**3.78**	0.00	
		QR-E-NET	1.01(1.59)	0.00	2.76	0.00	0.03(1.49)	0.00	2.59	0.00	
		WQR-RIDGE	18.29(27.42)	0.00	0.62	0.00	0.15(27.38)	0.00	1.68	0.00	1
		WQR-LASSO	**6.77(11.18)**	**25.50**	**3.93**	0.00	**0.04(10.48)**	19.00	**3.74**	0.00	
		WQR-E-NET	8.53(14.16)	0.00	1.61	0.00	0.38(14.10)	0.00	1.80	0.00	
			n=100(m=10)								
$D6-t_d$	$d=1, \sigma=0.50$	QR-RIDGE	2.11(3.45)	0.00	0.40	0.00	0.05(3.23)	0.00	0.11	0.00	
		QR-LASSO	**1.37(2.47)**	**59.00**	**4.89**	0.75	**0.02(2.00)**	**57.50**	**4.80**	0.68	
		QR-E-NET	1.45(2.61)	6.00	3.48	0.65	0.04(2.20)	1.00	2.63	0.48	
		WQR-RIDGE	24.68(35.82)	0.00	0.14	0.00	0.15(35.98)	0.00	0.02	0.00	
		WQR-LASSO	**10.18(18.26)**	**60.00**	**4.92**	0.78	**0.16(14.77)**	**58.50**	**4.84**	0.70	
		WQR-E-NET	14.09(23.75)	1.00	2.37	0.74	0.34(21.48)	1.00	1.68	0.56	
	$d=1, \sigma=1$	QR-RIDGE	3.13(4.74)	0.00	0.42	0.00	0.09(4.50)	0.00	0.11	0.00	
		QR-LASSO	**2.51(4.12)**	**33.00**	**4.96**	1.54	**0.04(3.64)**	**32.00**	**4.86**	1.32	
		QR-E-NET	2.62(4.25)	10.00	4.04	1.24	0.06(3.81)	4.00	3.41	1.03	
		WQR-RIDGE	27.14(39.88)	0.00	0.00	0.00	0.74(39.76)	0.00	0.05	0.00	
		WQR-LASSO	**18.68(30.49)**	**34.00**	**4.93**	1.46	**0.22(26.78)**	**34.00**	**4.90**	1.36	
		WQR-E-NET	22.41(34.25)	1.00	3.27	1.33	0.49(31.92)	1.00	2.57	1.11	
$D6-t_d$	$d=6, \sigma=0.50$	QR-RIDGE	0.98(1.60)	0.00	0.64	0.00	-0.01(1.74)	0.00	0.20	0.00	
		QR-LASSO	**0.44(0.71)**	**50.00**	**4.26**	0.00	**-0.01(0.72)**	**42.50**	**4.26**	0.00	
		QR-E-NET	0.45(0.74)	1.00	2.14	0.00	0.00(0.76)	0.00	1.72	0.00	
		WQR-RIDGE	20.59(30.57)	0.00	0.00	0.00	0.10(30.78)	0.00	0.00	0.00	
		WQR-LASSO	**3.26(5.23)**	**50.00**	**4.31**	0.00	**-0.05(5.27)**	**51.00**	**4.40**	0.00	
		WQR-E-NET	5.43(8.77)	0.00	1.08	0.00	0.11(9.77)	0.00	0.85	0.00	
	$d=6, \sigma=1$	QR-RIDGE	1.41(2.20)	0.00	0.52	0.00	0.03(2.29)	0.00	0.22	0.00	
		QR-LASSO	**0.84(1.33)**	**41.00**	**4.11**	0.00	**0.01(1.32)**	**33.00**	3.98	0.00	
		QR-E-NET	0.87(1.41)	1.50	2.21	0.00	0.02(1.38)	0.50	1.79	0.00	
		WQR-RIDGE	16.37(22.15)	0.00	0.07	0.00	0.16(32.31)	0.00	0.00	0.00	
		WQR-LASSO	**6.17(9.77)**	**45.50**	**4.15**	0.00	**0.05(9.65)**	**36.50**	**4.06**	0.00	
		WQR-E-NET	8.87(13.96)	0.00	1.09	0.00	0.12(14.27)	0.00	0.74	0.00	

t_d denotes t-distribution with d degrees of freedom.

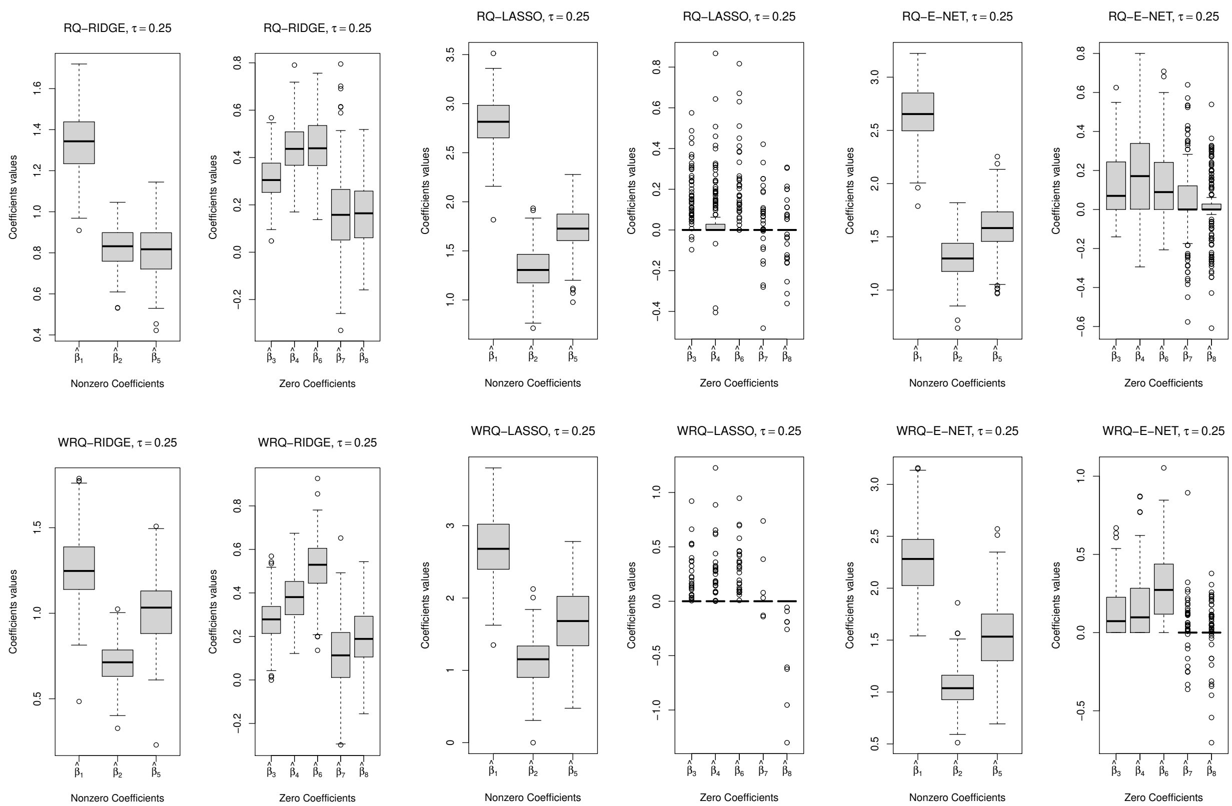


Figure 5.15: Box plots for *D*6 under *t*-distribution with $\sigma = 1$ and *d*=6 degrees of freedom at $\tau = 0.25$ quantile level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

The advantage of this data is that the position of extreme points is known exactly, making it possible to measure procedures effectively. The observations 1-10 are high leverage points (influential points), and points 11-14 are non-influential high leverage points. Artificial data are used to compare the *RIDGE*, *LASSO* and *E-NET* penalized *QR* variable selection procedures (both weighted and unweighted). The partitioned response vector $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')$ generates the response variable in this section, where $\boldsymbol{Y}_1 = \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \varepsilon_1$, $\varepsilon_1 \sim t_1$ for observations 11-75 with $\boldsymbol{\beta}_1 = (2, 2, 0)'$ and $\boldsymbol{Y}_2 = \boldsymbol{X}_2'\boldsymbol{\beta}_2$, where $\boldsymbol{\beta}_2 = (1, 1, 0)'$ for observations 1-10. The data is categorized into full and reduced data sets. The full data set contains collinearity inducing points, while the reduced data does not have collinearity inducing points.

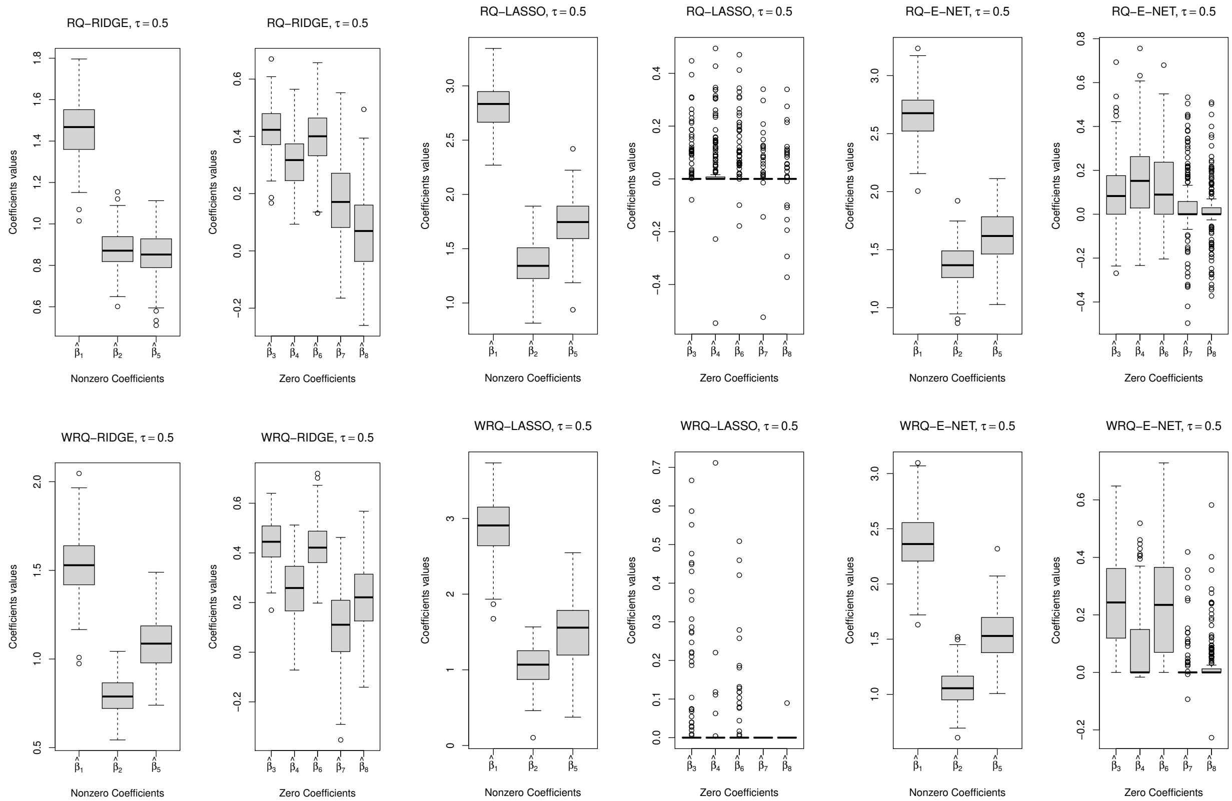


Figure 5.16: Box plots for $D6$ under t-distribution with $\sigma = 1$ and d=6 degrees of freedom at $\tau = 0.50$ quantile level. Vertical panels 1, 2 and 3 are for *QR-RIDGE*, *QR-LASSO* and *QR-E-NET* procedures (unweighted and weighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

We present parameter estimates and biases in *Table* 5.8 with the results of both the full and reduced data scenarios. The zero coefficient β_3 is correctly shrunk to zero for *WQR-LASSO* at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels with very minimal bias. The zero coefficients are shrunk to near zero for the rest of the penalized *QR* procedures. The penalized *WQR* performs better at $\tau = 0.50$ and is equivalent to the non-penalized version under both the *RIDGE* and *E-NET* penalties, since $\lambda = 0$. The *WQR-LASSO* is the best penalized *WQR* procedure at $\tau = 0.25$ when $\lambda \neq 0$ and all zero coefficients are shrunk to zero. Considering all penalties, an improvement in performance is depicted in the reduced data (data without high leverage points-observations 1-14) scenario at all *RQ* levels. In this reduced data scenario, penalized *WQR* procedures outperform their unweighted

Table 5.8: Results for the Hawkins full and reduced data sets at $\tau = 0.25$ and $\tau = 0.50$ quantile levels; bold text indicate better performance.

		β	QR $\hat{\beta}(Bias)$	QR-RIDGE $\hat{\beta}(Bias)$	QR-LASSO $\hat{\beta}(Bias)$	QR-E-NET $\hat{\beta}(Bias)$	WQR-RIDGE $\hat{\beta}(Bias)$	WQR-LASSO $\hat{\beta}(Bias)$	WQR-E-NET $\hat{\beta}(Bias)$
					FULL				
λ				0.00	0.00	0.00	0.00	0.06	0.00
$\tau = 0.50$	intercept	0.00	2.27(-2.27)	2.39(-2.39)	2.39(-2.39)	2.39(-2.39)	**0.11(-0.11)**	**0.00(0.00)**	**0.11(-0.11)**
	X1	2.00	1.39(0.61)	1.45(0.55)	1.45(0.55)	1.45(0.55)	**1.93(0.07)**	**1.93(0.07)**	**1.93(0.07)**
	X2	2.00	1.87(0.13)	1.79(0.21)	1.79(0.21)	1.79(0.21)	**2.01(-0.01)**	**1.97(0.03)**	**2.01(-0.01)**
	X3	0.00	-0.78(0.78)	-0.74(0.74)	-0.74(0.74)	-0.74(0.74)	**-0.09(0.09)**	**0.00(0.00)**	**-0.09(0.09)**
λ				0.00	0.00	0.00	0.50	0.50	0.50
$\tau = 0.25$	intercept	-1.00	1.09(-2.09)	**1.32(-2.32)**	**1.32(-2.32)**	**1.32(-2.32)**	0.18(-1.18)	0.29(-1.29)	0.29(-1.29)
	X1	2.00	1.59(0.41)	**1.48(0.52)**	**1.48(0.52)**	**1.48(0.52)**	0.35(1.65)	0.00(2.00)	0.00(2.00)
	X2	2.00	1.94(0.06)	**1.80(0.20)**	**1.80(0.20)**	**1.80(0.20)**	0.39(1.61)	0.00(2.00)	0.00(2.00)
	X3	0.00	-0.88(0.88)	**-0.76(0.76)**	**-0.76(0.76)**	**-0.76(0.76)**	0.38(-0.38)	0.00(0.00)	0.00(0.00)
					REDUCED				
λ				0.00	0.00	0.00	0.00	0.00	0.00
$\tau = 0.50$	Intercept	0.00	0.74(-0.74)	0.50(-0.50)	0.50(-0.50)	0.50(-0.50)	**0.27(-0.27)**	**0.27(-0.27)**	**0.27(-0.27)**
	X1	2.00	1.86(0.14)	1.84(0.16)	1.84(0.16)	1.84(0.16)	**1.86(0.14)**	**1.86(0.14)**	**1.86(0.14)**
	X2	2.00	1.93(0.07)	1.87(0.13)	1.87(0.13)	1.87(0.13)	**1.96(0.04)**	**1.96(0.04)**	**1.96(0.04)**
	X3	0.00	-0.07(0.07)	-0.03(0.03)	-0.03(0.03)	-0.03(0.03)	**-0.06(0.06)**	**-0.06(0.06)**	**-0.06(0.06)**
λ				0.00	0.00	0.00	0.50	0.33	0.50
$\tau = 0.25$	Intercept	-1.00	0.69(-1.69)	**-0.09(-0.91)**	**-0.09(0.09)**	**-0.09(-0.91)**	1.74(-2.74)	2.22(-3.22)	2.20(-3.20)
	X1	2.00	1.67(0.33)	**1.73(0.27)**	**1.73(0.27)**	**1.73(0.27)**	0.30(1.70)	0.00(2.00)	0.00(2.00)
	X2	2.00	1.90(0.10)	**1.95(0.05)**	**1.95(0.05)**	**1.95(0.05)**	0.48(1.52)	0.00(2.00)	0.10(1.90)
	X3	0.00	-0.08(0.08)	**-0.10(0.10)**	**-0.10(0.10)**	**-0.10(0.10)**	0.22(-0.22)	0.00(0.00)	0.00(0.00)

[1] The intercept $F^{-1}(\tau) + \boldsymbol{\beta}_0$, translate to 0 and -1 under the t_1 error term distribution, at quantile levels $\tau = 0.50$ and $\tau = 0.25$, respectively. The biases are calculated by the difference between the true and estimated parameters ($\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$). They show the ability of our regularization and variable selection procedures in *QR* procedures, to shrink some zero coefficients to zero.

versions.

5.3.2 Hocking and Pendleton Data Set

The performance of weighted regularized QR and WQR procedures in this section is evaluated using bias in estimated coefficients and penalization of zero coefficients in the Hocking and Pendleton data set (Hocking & Pendleton 1983). The data set consists of 26 observations, with a response variable Y and 3 predictor variables (X_1, X_2, X_3) with X_3 created by a linear combination of X_1 and X_2. In the literature, observations 11, 17 and 18 are flagged as outlying points, and this data set contains collinearity hiding points. The response vector $\boldsymbol{Y}$ is generated by $\boldsymbol{Y}_1 = \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \varepsilon_1,\ \varepsilon_1 \sim t_1$, for the first 22 observations and $Y_2 = \boldsymbol{X}_2'\boldsymbol{\beta}_2$ for the remaining 4 observations, such that $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')$, where $\boldsymbol{\beta}_1 = (3, -2, 0)'$ and $\boldsymbol{\beta}_2 = (1, 1, 0)'$. We consider the results of the full data set containing collinearity hiding points (high leverage points that are collinearity influential points) and the reduced data without collinearity hiding point 24.

We summarize the results of the full and reduced Hocking & Pendleton (1983) data set in *Table* 5.9 for assessing the performance of the penalized QR and WQR procedures. The *Table* summarizes the β coefficients and biases (the difference between the true βs and the estimated βs). Unlike at $\tau = 0.25$, the last coefficient is shrunk to zero at $\tau = 0.50$ in the penalized weighted scenarios. In the unweighted versions, the βs are best estimated at $\tau = 0.25$, whereas in the WQR scenario, the βs are best estimated at $\tau = 0.50$ across all penalty functions. When the tuning parameter $\lambda = 0$ in the penalized WQR procedures, the non-penalized WQR procedures are optimal.

Table 5.9: Results for the Hocking full and reduced data sets at $\tau = 0.25$ and $\tau = 0.50$ quantile levels; bold text indicate better performance.

		β	QR $\hat{\beta}(Bias)$	QR-RIDGE $\hat{\beta}(Bias)$	QR-LASSO $\hat{\beta}(Bias)$	QR-E-NET $\hat{\beta}(Bias)$	WQR-RIDGE $\hat{\beta}(Bias)$	WQR-LASSO $\hat{\beta}(Bias)$	WQR-E-NET $\hat{\beta}(Bias)$
					FULL				
λ				0.11	0.06	0.11	0.00	0.00	0.00
$\tau = 0.50$	Intercept	0.00	25.09(-25.09)	24.34(-24.34)	27.63(-27.63)	23.63(-23.63)	**0.36(-0.36)**	**0.36(-0.36)**	**0.36(-0.36)**
	X1	3.00	1.55(1.45)	0.86(2.14)	1.28(1.72)	1.06(1.94)	**2.94(0.06)**	**2.94(0.06)**	**2.94(0.06)**
	X2	-2.00	-2.30(0.30)	-0.86(-1.14)	-2.12(0.12)	-1.21(-0.79)	**-2.08(0.08)**	**-2.08(0.08)**	**-2.08(0.08)**
	X3	0.00	-0.66(0.66)	0.17(-0.17)	-0.49(0.49)	0.00(0.00)	**0.01(-0.01)**	**0.01(-0.01)**	**0.01(-0.01)**
λ				0.00	0.06	0.06	0.00	0.00	0.00
$\tau = 0.25$	Intercept	-1.00	23.53(-24.53)	25.26(-26.26)	30.32(-31.32)	33.13(-34.13)	**-0.08(-0.92)**	**-0.08(-0.92)**	7.62(-8.62)
	X1	3.00	1.19(1.81)	1.09(1.91)	0.56(2.44)	0.30(2.70)	**2.95(0.05)**	**2.95(0.05)**	2.95(0.05)
	X2	-2.00	-1.96(-0.04)	-1.98(-0.02)	-1.70(-0.30)	-1.53(-0.47)	**-2.47(0.47)**	**-2.47(0.47)**	-2.47(0.47)
	X3	0.00	-0.15(0.15)	-0.16(0.16)	0.00(0.00)	0.02(-0.02)	**-0.03(0.03)**	**-0.03(0.03)**	-0.03(0.03)
					REDUCED				
λ				0.00	0.22	0.08	0.00	0.06	0.00
$\tau = 0.50$	Intercept	0.00	-59.31(59.31)	-56.47(56.47)	40.67(-40.67)	8.77(-8.77)	**0.12(-0.12)**	-0.24(0.24)	**0.12(-0.12)**
	X1	3.00	5.78(-2.78)	5.65(-2.65)	0.00(3.00)	2.09(0.91)	**2.88(0.12)**	2.77(0.23)	**2.88(0.12)**
	X2	-2.00	-0.22(-1.78)	-0.32(-1.68)	-1.18(-0.82)	-1.37(-0.63)	**-1.88(-0.12)**	-1.44(-0.56)	**-1.88(-0.12)**
	X3	0.00	2.13(-2.13)	2.05(-2.05)	0.00(0.00)	0.30(-0.30)	**0.07(-0.07)**	0.20(-0.20)	**0.07(-0.07)**
λ				0.00	0.00	0.00	0.00	0.00	0.00
$\tau = 0.25$	Intercept	-1.00	-56.16(55.16)	-59.60(58.60)	-59.61(58.61)	-59.61(58.61)	**-0.37(-0.63)**	**-0.37(-0.63)**	**-0.37(-0.63)**
	X1	3.00	5.67(-2.67)	5.80(-2.80)	5.80(-2.80)	5.80(-2.80)	**3.02(-0.02)**	**3.02(-0.02)**	**3.02(-0.02)**
	X2	-2.00	-0.61(-1.39)	-0.48(-1.52)	-0.48(-1.52)	-0.48(-1.52)	**-2.59(0.59)**	**-2.59(0.59)**	**-2.59(0.59)**
	X3	0.00	1.96(-1.96)	2.13(-2.13)	2.13(-2.13)	2.13(-2.13)	**-0.12(0.12)**	**-0.12(0.12)**	**-0.12(0.12)**

[1] The intercept $F^{-1}(\tau) + \boldsymbol{\beta}_0$, translate to 0 and -1 under the t_1 error term distribution, at quantile levels $\tau = 0.50$ and $\tau = 0.25$, respectively. The biases are calculated by the difference between the true and estimated parameters ($\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$). They show the ability of our regularization and variable selection procedures in *QR* procedures, to shrink some zero coefficients to zero.

5.4 Results on Adaptive Penalized Quantile Regression

In this section, we present simulation results, discuss them and compare the performance of non-adaptive penalized *QR* procedures namely, *QR-LASSO*, *QR-E-NET*, *WQR-LASSO*, and *WQR-E-NET*, with their adaptive variants namely, *QR-ALASSO*, *QR-AE-NET*, *WQR-ALASSO* and *WQR-AE-NET*. We evaluate the variable selection and prediction performance of these adaptive penalized procedures in the presence of collinearity, high leverage points and collinearity influential points under two different distribution scenarios. The two distribution scenarios are the Gaussian distribution with varying σs and the t-distribution with varying degrees of freedom (d) and σs. The best adaptive penalized *QR* procedure (weighted/unweighted) is determined by the metrics namely, the average number of correctly/incorrectly fitted zero coefficients, the percentage of correctly fitted models and the *MAD* of test errors ($MAD = 1.4826\left(Median\{\varepsilon_i\} - Median\{\varepsilon_i\}\right), i \in [1:n]$, where $Median\{\varepsilon_i\}$ is the median of test errors). Without loss of generality, we only consider non-adaptive and adaptive penalized procedures at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels, with the mixing parameter for the *E-NET* penalty as $\alpha = 0.50$.

Baseline scenario (D1)

We discuss simulation results for the baseline scenario (orthogonal design scenario-$D1$) under the Gaussian distribution and t-distribution (heavy tailed distribution cases with $d = 1$ degree of freedom) scenarios (see *Table* 5.10). The baseline scenario contains no high leverage points and typically no collinearity influential points. The adaptive regularized *QR* procedures namely, *QR-ALASSO* and *QR-AE-NET*, correctly shrink zero coefficients to zero/near zero compared to the non-adaptive versions (see also the box plots of these baseline scenarios in *Figure* 5.17). In the majority of cases, adaptive penalized procedures outperform the non-adaptive versions, with the exception of *QR-LASSO* at $\tau \in (0.25; 0.50)$ *RQ* levels (t-distribution case, $d = 1$, $\sigma = 1$). In the Gaussian and t-distribution cases, *QR-ALASSO* is superior to the *QR-AE-NET* with respect to the *MAD* of test errors metric. Also, *QR-ALASSO* has higher percentages of correctly fitted models and average number of correctly fitted zero coefficients.

Table 5.10: Penalized quantile regression at $D1$ under the normal and t-distributions for $n = 50$ at $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels; bold text indicate better performance.

			$\tau = 0.25$					$\tau = 0.50$				
			median(MAD)	Correctly	No of Zeros			median(MAD)	Correctly	No of Zeros		
Distribution	Parameter	Method	Test Error	Fitted	c.zero	inc.zero	median(λ)	Test Error	Fitted	c.zero	inc.zero	median(λ)
$D1-N(.,.)$	$\sigma = 1$	QR-LASSO	0.72(1.17)	62.00	4.40	0.00	0.04	0.02(1.16)	**65.50**	**4.51**	0.00	0.04
		QR-E-NET	0.75(1.26)	16.50	3.35	0.00	0.03	0.02(1.19)	21.00	3.56	0.00	0.04
		QR-ALASSO	**0.71(1.10)**	**99.50**	**5.00**	0.00	0.03	**0.01(1.13)**	**100.00**	**5.00**	0.00	0.04
		QR-AE-NET	0.75(1.16)	**97.00**	**4.97**	0.00	0.05	0.01(1.17)	**100.00**	**5.00**	0.00	0.06
	$\sigma = 3$	QR-LASSO	2.20(3.60)	44.00	4.39	0.28	0.04	0.09(3.47)	47.50	4.51	0.22	0.05
		QR-E-NET	2.29(3.68)	27.00	3.83	0.14	0.04	0.06(3.62)	32.50	3.99	0.09	0.04
		QR-ALASSO	**2.15(3.38)**	**60.00**	**4.95**	0.46	0.01	**0.06(3.41)**	49.00	4.90	0.55	0.02
		QR-AE-NET	2.28(3.54)	**67.50**	**4.90**	0.28	0.01	0.05(3.51)	**60.50**	**4.85**	0.30	0.03
$D1-t_d$	$d = 1, \sigma = 1$	QR-LASSO	2.32(3.81)	30.50	4.96	1.66	0.04	**0.03(2.91)**	**40.50**	4.96	1.27	0.04
		QR-E-NET	2.55(3.87)	29.50	4.77	1.46	0.03	0.02(3.27)	29.00	4.64	1.10	0.03
		QR-ALASSO	**2.36(3.78)**	11.50	**4.97**	1.95	0.00	-0.07(3.24)	32.50	**4.99**	1.56	0.00
		QR-AE-NET	2.52(3.93)	12.50	4.94	1.86	0.00	-0.05(3.43)	37.00	4.93	1.38	0.00
	$d = 1, \sigma = 3$	QR-LASSO	4.70(7.15)	1.50	**5.00**	2.93	0.06	-0.14(6.89)	1.50	4.99	2.85	0.05
		QR-E-NET	4.72(7.14)	1.50	4.99	2.91	0.05	-0.14(6.91)	1.00	4.97	2.84	0.05
		QR-ALASSO	**4.68(7.12)**	0.50	4.99	2.93	0.00	**-0.14(6.84)**	0.50	**5.00**	2.86	0.00
		QR-AE-NET	4.72(7.17)	0.50	**5.00**	2.94	0.00	-0.13(6.90)	2.00	4.99	2.85	0.00

[1] $N(.,.)$ denotes normally distributed and t_d denotes t-distribution with d degrees of freedom.

Remarks. *The set of zero coefficients corresponds to the set* $\{\beta_j : j = 3,4,6,7,8\}$ *with a maximum average of correctly shrunk coefficients of* 5*. The set of correctly selected models is evaluated as a percentage.*

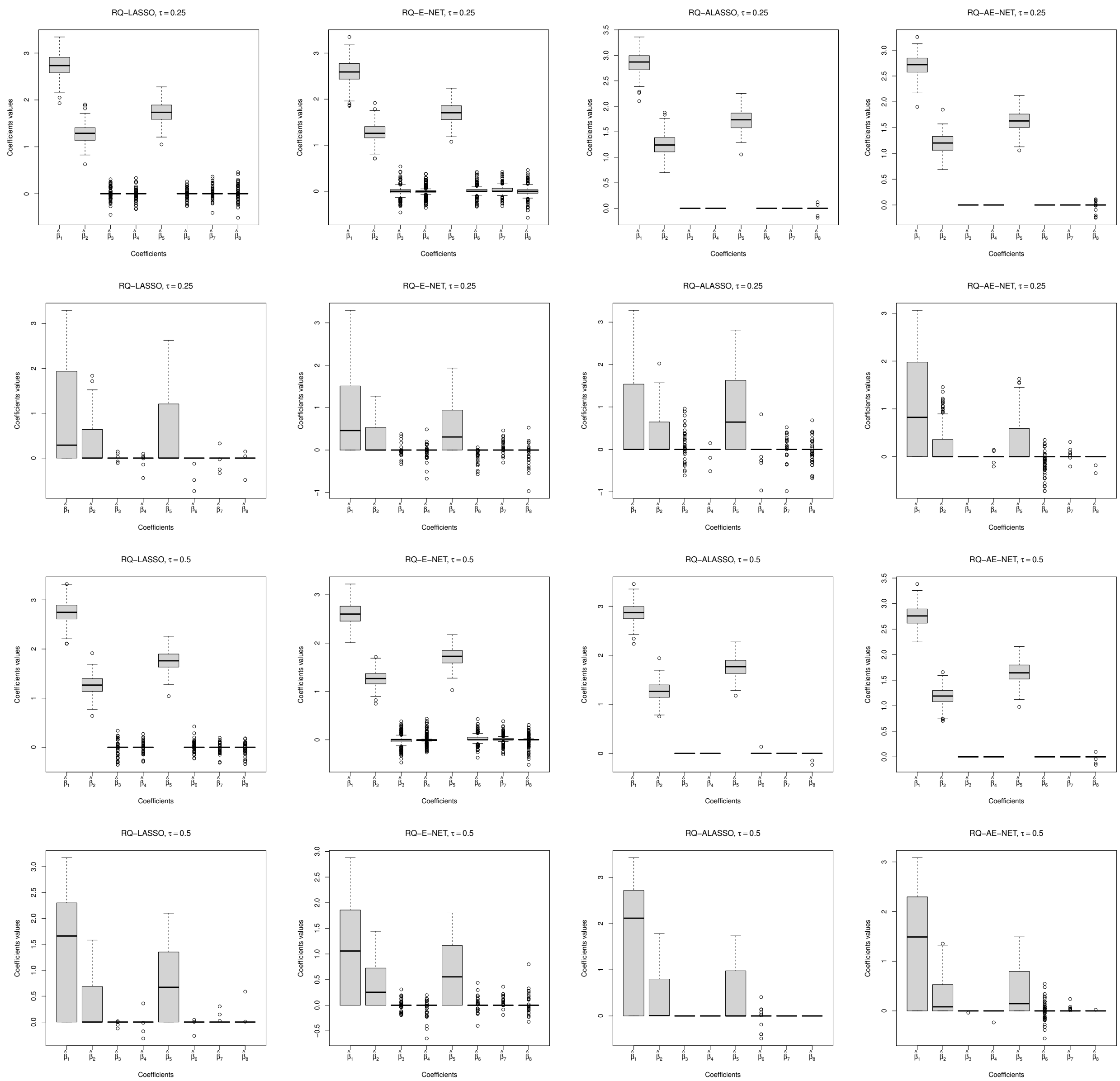


Figure 5.17: Box plots of coefficient values at $D1$ for RQ at $\tau = 0.25$ and $\tau = 0.50$. Horizontal panels 1 and 3 are under the normal distribution with $\sigma = 1$. Horizontal panels 2 and 4 are under the t-distribution with $\sigma = 1$, $d = 1$. All box plots are for unweighted procedures. The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Collinearity inducing scenario ($D2$ ***and*** $D4$)

The results of the collinearity-inducing point scenarios $D2$ and $D4$ are shown in *Table* 5.11. The adaptive regularized QR procedures outperform the non-adaptive ones 100% of the time in pairwise comparisons with respect of correct shrunk zero coefficients (see also Box plots in the *Figures* 5.18 and 5.19). Adaptive penalized procedures outperform the non-adaptive penalized ones, both in correctly fitting the models and with respect to prediction (all above 80% of the time). The weighted scenarios depict a similar performance pattern as the unweighted ones.

The adaptive penalized *WQR* procedures namely, *WQR-ALASSO* and *WQR-AE-NET* outperform the non-adaptive versions of *WQR* procedures with respect to correctly fitting the models and in prediction (more than 60% of the time). In the pairwise comparisons, the *WQR-ALASSO* and *WQR-AE-NET* procedures outperform the non-weighted versions *QR-ALASSO* and *QR-AE-NET* with respect to all metrics in the majority of cases.

Table 5.11: Weighted and unweighted quantile regression at $D2$ and $D4$ for $n = 50$ ($\tau = 0.25$ and $\tau = 0.50$) under the normal distribution; bold text indicate better performance.

			$\tau = 0.25$					$\tau = 0.50$				
			median(MAD)	Correctly	No of Zeros			median(MAD)	Correctly	No of Zeros		
Distribution	Parameter	Method	Test Error	Fitted	c.zero	inc.zero	median(λ)	Test Error	Fitted	c.zero	inc.zero	median(λ)
$D2-N(.,.)$	$\sigma = 1$	QR-LASSO	-1.10(5.65)	38.00	4.00	0.01	0.03	-1.87(5.78)	36.00	3.83	0.01	0.02
		QR-E-NET	-0.93(5.55)	2.50	2.55	0.00	0.02	-1.91(5.72)	0.00	1.94	0.00	0.02
		QR-ALASSO	-1.15(5.66)	**79.00**	**4.80**	0.02	0.01	-1.90(5.82)	**86.50**	**4.87**	0.01	0.02
		QR-AE-NET	**-0.43(4.99)**	**98.00**	**4.98**	0.00	0.02	**-1.64(5.17)**	**88.00**	**4.88**	0.00	0.03
	$\sigma = 3$	QR-LASSO	0.96(6.30)	17.50	4.23	0.43	0.02	-1.41(6.56)	21.00	3.99	0.28	0.03
		QR-E-NET	1.25(5.85)	9.50	3.39	0.16	0.02	-1.34(6.21)	4.50	2.53	0.09	0.02
		QR-ALASSO	0.74(6.53)	46.00	4.88	0.50	0.01	**-1.63(6.66)**	**82.50**	**4.98**	0.18	0.01
		QR-AE-NET	1.43(5.61)	44.00	4.81	0.41	0.01	-0.97(5.73)	45.50	4.60	0.28	0.02
	$\sigma = 1$	WQR-LASSO	-1.48(4.49)	62.50	4.47	0.00	0.03	-2.22(4.24)	66.50	4.57	0.00	0.04
		WQR-E-NET	-1.56(4.61)	11.00	3.34	0.00	0.03	-2.04(4.39)	30.00	3.86	0.00	0.04
		WQR-ALASSO	-1.56(4.62)	**97.50**	**4.98**	0.00	0.02	-2.12(4.50)	**100.00**	**5.00**	0.00	0.02
		WQR-AE-NET	**-1.61(4.19)**	**93.00**	**4.92**	0.00	0.04	**-2.09(4.26)**	**98.50**	**4.99**	0.00	0.04
	$\sigma = 3$	WQR-LASSO	0.50(4.63)	29.50	4.53	0.62	0.03	-1.06(4.49)	36.50	4.69	0.54	0.04
		WQR-E-NET	0.76(4.31)	28.00	4.34	0.54	0.04	-0.99(4.28)	28.00	4.34	0.37	0.05
		WQR-ALASSO	1.50(2.28)	47.50	4.97	0.73	0.01	**0.04(2.27)**	**51.50**	**5.00**	0.59	0.02
		WQR-AE-NET	**1.64(2.38)**	**57.00**	**4.97**	0.49	0.02	**0.03(2.26)**	**67.00**	**4.99**	0.38	0.03
$D4-N(.,.)$	$\sigma = 1$	QR-LASSO	0.97(1.65)	10.00	3.88	0.61	0.02	-0.03(1.40)	5.50	3.49	0.44	0.02
		QR-E-NET	2.74(3.71)	1.00	1.69	0.16	0.01	-0.07(3.56)	0.00	1.06	0.01	0.02
		QR-ALASSO	**0.84(1.37)**	**75.50**	**5.00**	0.34	0.01	**0.01(1.42)**	**83.50**	**4.97**	0.14	0.01
		QR-AE-NET	2.46(3.46)	41.00	4.35	0.13	0.01	-0.01(3.14)	11.00	3.96	0.03	0.01
	$\sigma = 3$	QR-LASSO	1.67(6.06)	3.50	4.07	1.26	0.02	-1.31(6.57)	5.00	3.46	0.64	0.02
		QR-E-NET	2.64(5.30)	0.00	2.25	0.75	0.01	-0.78(5.30)	0.00	1.60	0.36	0.02
		QR-ALASSO	1.40(6.37)	43.50	4.84	0.88	0.01	**-1.25(6.76)**	**81.50**	**5.00**	0.45	0.01
		QR-AE-NET	2.79(5.09)	32.00	4.39	0.53	0.01	-0.48(5.09)	6.00	2.81	0.31	0.01
	$\sigma = 1$	WQR-LASSO	0.47(0.78)	**64.50**	**4.51**	0.00	0.03	**0.00(0.85)**	**67.50**	**4.58**	0.00	0.04
		WQR-E-NET	0.49(0.85)	30.00	3.96	0.00	0.04	0.00(0.90)	33.00	4.05	0.00	0.04
		WQR-ALASSO	**0.48(0.79)**	**96.50**	**4.97**	0.00	0.02	**0.00(0.82)**	**99.50**	**5.00**	0.01	0.02
		WQR-AE-NET	**0.54(0.89)**	**100.00**	**5.00**	0.00	0.04	**0.00(0.80)**	**99.50**	**5.00**	0.01	0.04
	$\sigma = 3$	WQR-LASSO	1.37(2.31)	15.00	4.46	1.09	0.04	-0.04(2.33)	32.50	4.64	0.59	0.04
		WQR-E-NET	1.52(2.38)	35.00	4.29	0.28	0.04	-0.06(2.34)	20.50	4.01	0.42	0.04
		WQR-ALASSO	1.48(2.30)	46.50	4.92	0.58	0.01	-0.04(2.32)	33.50	4.83	0.79	0.01
		WQR-AE-NET	1.62(2.41)	41.00	4.87	0.68	0.02	-0.02(2.30)	32.50	4.83	0.70	0.01

[1] $N(.,.)$ denotes normally distributed.

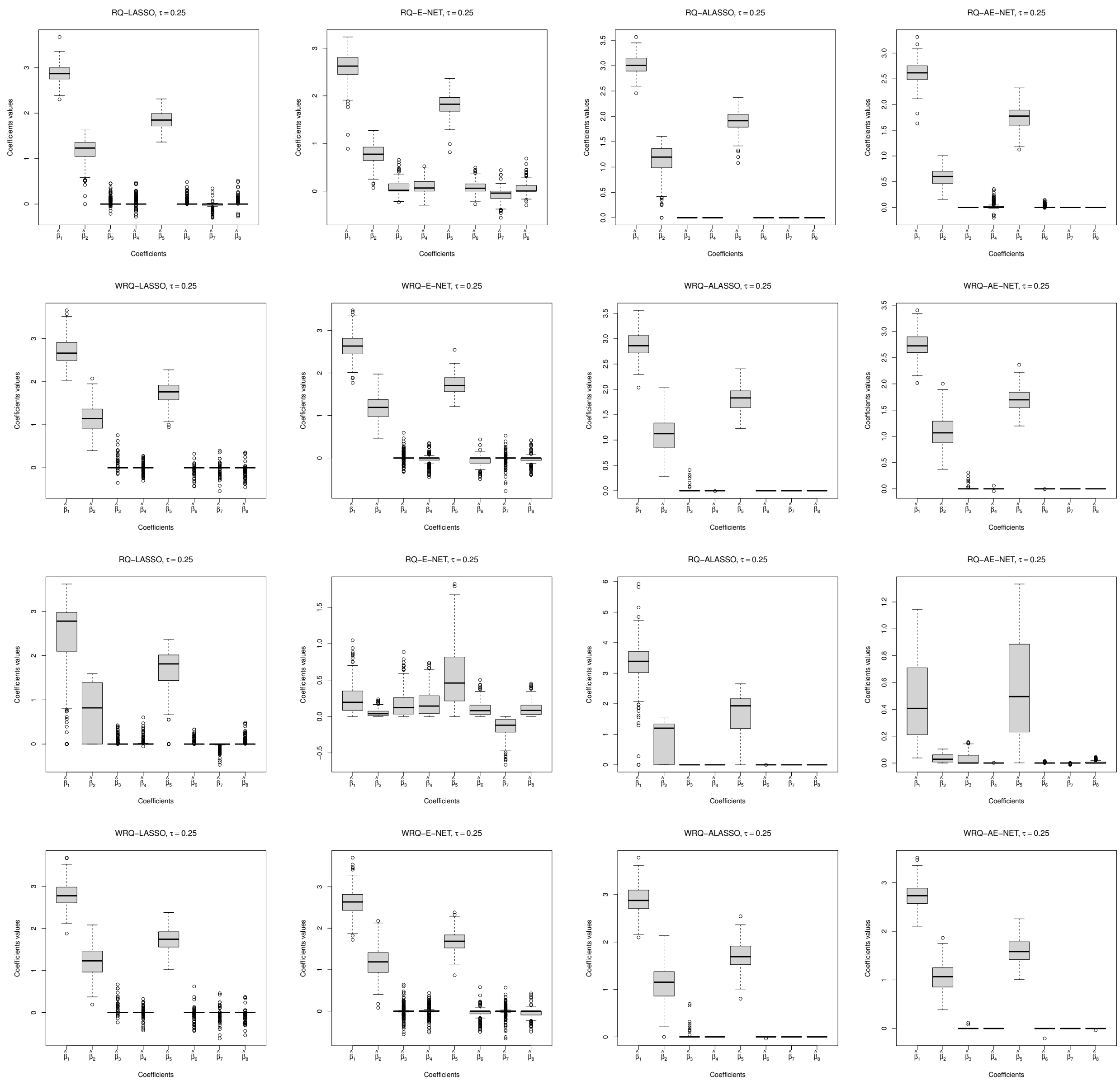


Figure 5.18: Box plots of coefficient values in the $D2$ and $D3$ scenarios at $\tau = 0.25$ *RQ* levels. Horizontal panels 1 and 2 are for $D2$ under the normal distribution with $\sigma = 1$. Horizontal panels 3 and 4 are for $D3$ scenario under the normal distribution with $\sigma = 1$ (both weighted and unweighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

Collinearity reducing scenario (D3 AND D5)

In *Table* 6.12, we show simulation results of variable selection and prediction performance in the presence of collinearity hiding points ($D3$ and $D5$ scenarios) under the normal distribution.

The *QR-ALASSO* and *QR-AE-NET* procedures dominate the non-adaptive penalized *QR* versions *QR-LASSO* and *QR-E-NET* with respect to all metrics.

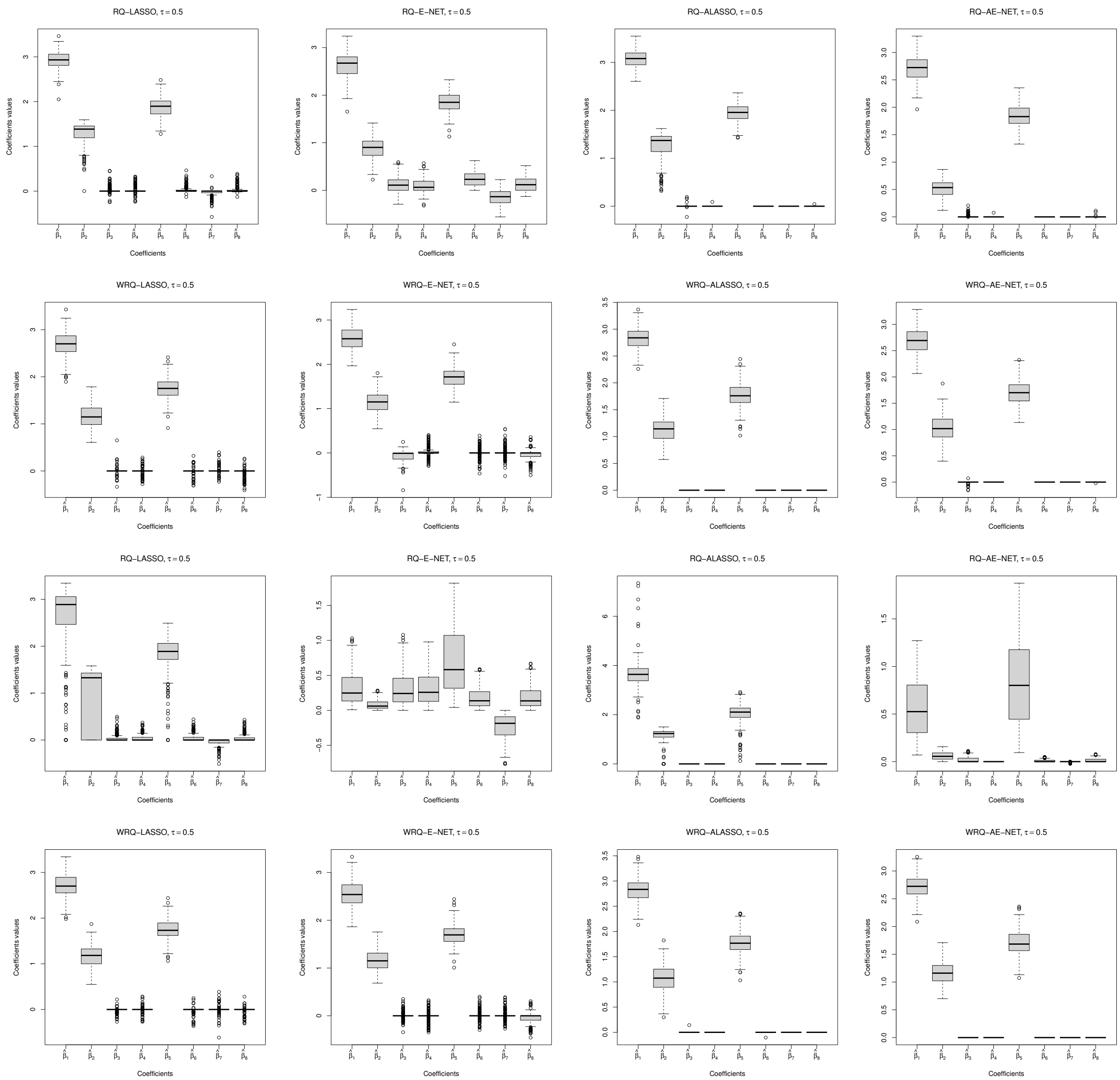


Figure 5.19: Box plots of coefficient values in the $D2$ and $D3$ scenarios at $\tau = 0.50$ RQ levels. Horizontal panels 1 and 2 are for $D2$ under the normal distribution with $\sigma = 1$. Horizontal panels 3 and 4 are for $D3$ scenario under the normal distribution with $\sigma = 1$ (both weighted and unweighted scenarios). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

The two unweighted adaptive procedures (*QR-ALASSO* and *QR-AE-NET*) outperform the non-adaptive ones by 100% and 88% with respect to prediction, by 100% and 50% with respect to correctly fitted models, respectively and by 100% for both with respect to the average

Table 5.12: Weighted and unweighted quantile regression at $D3$ and $D5$ for $n = 50$ ($\tau = 0.25$ and $\tau = 0.50$) under the normal distribution; bold text indicate better performance.

			$\tau = 0.25$					$\tau = 0.50$				
			median(MAD)	Correctly	No of Zeros			median(MAD)	Correctly	No of Zeros		
Distribution	Parameter	Method	Test Error	Fitted	c.zero	inc.zero	median(λ)	Test Error	Fitted	c.zero	inc.zero	median(λ)
$D3-N(.,.)$	$\sigma = 1$	QR-LASSO	**0.65(1.21)**	**60.50**	**4.45**	0.00	0.01	**-0.01(1.19)**	**62.50**	**4.48**	0.00	0.02
		QR-E-NET	0.81(1.46)	12.50	3.48	0.00	0.01	-0.04(1.39)	12.50	3.45	0.00	0.02
		QR-ALASSO	**0.68(1.19)**	**95.00**	**4.97**	0.02	0.01	**-0.03(1.16)**	**100.00**	**5.00**	0.00	0.01
		QR-AE-NET	0.84(1.39)	**81.50**	**4.81**	0.00	0.01	-0.09(1.64)	44.50	4.45	0.00	0.02
	$\sigma = 3$	QR-LASSO	2.01(3.79)	42.50	4.48	0.35	0.01	-0.08(3.61)	53.00	4.54	0.30	0.02
		QR-E-NET	2.47(4.20)	17.50	3.97	0.43	0.01	-0.10(4.02)	11.00	3.92	0.54	0.01
		QR-ALASSO	**2.15(3.64)**	**77.00**	**4.96**	0.29	0.01	**-0.14(3.50)**	**72.00**	**4.88**	0.25	0.00
		QR-AE-NET	2.37(3.95)	45.00	4.44	0.33	0.00	-0.17(4.06)	42.50	4.64	0.50	0.01
	$\sigma = 1$	WQR-LASSO	0.48(0.91)	28.50	3.83	0.00	0.04	**-0.01(0.67)**	**71.50**	**4.65**	0.00	0.04
		WQR-E-NET	0.41(0.77)	15.50	3.59	0.00	0.03	-0.01(0.77)	35.00	3.96	0.00	0.05
		WQR-ALASSO	**0.41(0.71)**	**98.00**	**4.98**	0.00	0.02	**-0.02(0.75)**	**100.00**	**5.00**	0.00	0.03
		WQR-AE-NET	**0.47(0.74)**	**85.00**	**4.85**	0.00	0.04	**-0.01(0.68)**	**95.00**	**4.95**	0.00	0.04
	$\sigma = 3$	WQR-LASSO	1.25(2.12)	29.00	4.39	0.57	0.03	-0.05(2.27)	39.50	4.66	0.48	0.04
		WQR-E-NET	1.46(2.31)	24.50	4.21	0.42	0.04	-0.05(2.30)	31.50	4.38	0.35	0.04
		WQR-ALASSO	0.11(4.96)	39.50	4.82	0.61	0.01	-1.24(4.57)	47.50	4.90	0.58	0.00
		WQR-AE-NET	**0.61(4.50)**	**56.00**	**4.94**	0.52	0.02	-1.11(4.33)	**52.50**	**4.93**	0.47	0.01
$D5-N(.,.)$	$\sigma = 1$	QR-LASSO	2.68(4.14)	12.00	5.00	1.89	0.01	**-0.06(2.34)**	**54.50**	**4.98**	0.80	0.00
		QR-E-NET	3.27(4.67)	0.00	4.94	2.94	0.04	-0.02(4.67)	0.00	4.97	2.91	0.09
		QR-ALASSO	0.93(1.52)	**93.00**	**5.00**	**0.08**	0.00	**-0.02(1.54)**	**89.50**	**5.00**	0.16	0.00
		QR-AE-NET	3.26(4.66)	0.00	5.00	2.90	0.02	-0.01(4.67)	0.00	5.00	2.96	0.05
	$\sigma = 3$	QR-LASSO	3.61(5.43)	1.00	5.00	2.57	0.05	0.08(5.40)	11.50	4.98	2.56	0.01
		QR-E-NET	3.68(5.45)	0.00	4.94	2.87	0.04	-0.03(5.42)	0.00	4.96	2.92	0.09
		QR-ALASSO	2.65(4.27)	41.00	5.00	0.88	0.00	0.02(4.66)	20.50	4.97	2.06	0.00
		QR-AE-NET	3.66(5.46)	0.00	4.97	2.96	0.00	-0.04(5.43)	0.00	4.99	2.95	0.07
	$\sigma = 1$	WQR-LASSO	**0.50(0.82)**	**71.00**	**4.61**	0.00	0.04	0.01(0.70)	68.50	4.61	0.00	0.04
		WQR-E-NET	0.53(0.82)	46.00	4.23	0.00	0.04	0.01(0.69)	34.50	3.95	0.00	0.04
		WQR-ALASSO	**0.51(0.78)**	**97.50**	**4.98**	0.00	0.03	**0.01(0.66)**	**98.50**	**4.99**	0.00	0.02
		WQR-AE-NET	**0.55(0.77)**	**92.50**	**4.92**	0.00	0.04	0.01(0.73)	**97.50**	**4.98**	0.00	0.04
	$\sigma = 3$	WQR-LASSO	1.57(2.39)	36.50	4.64	0.55	0.04	0.04(2.26)	42.00	4.72	0.47	0.04
		WQR-E-NET	1.67(2.46)	27.50	4.24	0.43	0.04	0.03(2.26)	29.00	4.35	0.38	0.05
		WQR-ALASSO	**1.52(2.26)**	**48.50**	**4.94**	0.67	0.01	0.03(2.14)	39.50	4.98	0.73	0.01
		WQR-AE-NET	**1.64(2.31)**	**67.00**	**4.97**	0.33	0.02	**0.03(2.65)**	**86.50**	**5.00**	0.15	0.03

[1] $N(.,.)$ denotes normally distributed.

number of correctly fitted zero coefficients. The performance picture in the weighted scenario is as follows: with respect to prediction, both the *WQR-LASSO* and *WQR-E-NET* procedures outperform their respective unweighted adaptive penalized versions in prediction 63% of the time. These adaptive penalized *WQR* procedures also dominate their non-adaptive versions with respect to correctly fitting models. Despite the dominance of *WQR-AE-NET* when $\sigma = 3$ ($D3$ and $D5$), the *WQR-ALASSO* procedure dominates the weighted penalized criteria. Therefore, our proposed adaptive weights improve the performance of models with respect to all metrics in the unweighted adaptive *QR* scenarios. However, in the weighted scenario, the adaptive weights hamper the performance of the models with respect to prediction.

Heavy-tailed distribution scenarios (D2 AND D3)

Table 5.13 shows the results of variable/model selection and prediction performance in the presence of collinearity inducing and hiding points ($D2$ and $D3$) under the t-distribution (heavy-tailed distribution). For brevity, the results of the heavy-tailed $D4$ and $D5$ scenarios are left out since the results are similar to those of $D2$ and $D3$.

In the collinearity inducing points scenario $D2$ under the t-distribution adaptive penalized *QR* and *WQR* versions outperform the non-adaptive penalized *QR* versions with respect to all metrics. In the pairwise comparisons, the *QR-ALASSO* and *QR-AE-NET* procedures outperform the non-adaptive versions in prediction 88% of the time (both procedures). With respect to correctly fitted models, they (*QR-ALASSO* and *QR-AE-NET*) are superior 100% and 38% of the time, respectively. With respect to correctly shrinking zero coefficients, the unweighted adaptive penalized *QR* procedures outperform the unweighted non-adaptive penalized versions 100% and 88% of the time, respectively. The pairwise comparisons demonstrate the dominance of penalized *WQR* procedures (*WQR-ALASSO* and *WQR-AE-NET*) over non-adaptive penalized versions. They (*WQR-ALASSO* and *WQR-AE-NET*) dominate the non-adaptive penalized *WQR* versions in prediction 100% and 75% of the time, respectively. With respect to correctly fitted models, these penalized *WQR* procedures perform better, that is, 88% of the time compared to the non-adaptive penalized *WQR* ones. In pairwise comparisons, the *WQR-ALASSO* and *WQR-AE-NET* procedures outperform the unweighted versions (*QR-ALASSO* and *QR-AE-NET*) with respect to prediction. In terms of all metrics, the *WQR-ALASSO* overall outperforms all other models.

Table 5.13: Weighted and unweighted quantile regression at $D2$ and $D3$ under heavy tailed t-distribution for $n = 50$ ($\tau = 0.25$ and $\tau = 0.50$); bold text indicate better performance.

			0.25					0.50				
			median(MAD)	Correctly	No of Zeros			median(MAD)	Correctly	No of Zeros		
Distribution	Parameter	Method	Test Error	Fitted	c.zero	inc.zero	median(λ)	Test Error	Fitted	c.zero	inc.zero	median(λ)
$D2-t_d$	$d=1,\sigma=1$	QR-LASSO	2.49(4.51)	1.50	4.26	1.50	0.02	0.00(3.64)	3.00	3.97	1.20	0.02
		QR-E-NET	3.26(5.05)	0.00	2.79	1.32	0.02	-0.01(4.32)	0.00	2.67	0.92	0.02
		QR-ALASSO	2.72(4.55)	0.50	4.50	1.56	0.00	0.02(3.76)	1.50	4.00	1.30	0.00
		QR-AE-NET	3.53(5.28)	0.00	3.37	1.41	0.00	0.05(4.64)	0.00	3.12	1.20	0.00
	$d=1,\sigma=0.5$	QR-LASSO	1.08(2.02)	15.50	4.05	0.54	0.02	-0.02(1.70)	30.00	4.37	0.53	0.03
		QR-E-NET	1.75(3.09)	0.00	1.30	0.20	0.02	0.01(2.49)	0.00	1.64	0.27	0.02
		QR-ALASSO	**1.16(2.11)**	**58.00**	**4.92**	0.71	0.01	**-0.02(1.72)**	**76.50**	**4.95**	0.46	0.02
		QR-AE-NET	1.38(2.44)	0.50	2.59	0.20	0.01	0.01(2.21)	0.00	2.89	0.26	0.01
	$d=1,\sigma=1$	WQR-LASSO	1.57(2.51)	21.50	4.85	1.65	0.04	0.03(2.21)	29.50	4.73	1.14	0.04
		WQR-E-NET	1.61(2.60)	4.00	4.35	1.55	0.04	0.02(2.30)	9.00	4.11	1.00	0.04
		WQR-ALASSO	1.68(2.63)	11.00	4.96	1.84	0.00	0.01(2.30)	20.50	4.95	1.38	0.00
		WQR-AE-NET	1.70(2.68)	8.50	4.86	1.74	0.00	0.03(2.42)	16.00	4.76	1.27	0.00
	$d=1,\sigma=0.5$	WQR-LASSO	**0.81(1.41)**	**46.50**	**4.74**	0.82	0.04	**0.01(1.15)**	**51.50**	**4.67**	0.51	0.04
		WQR-E-NET	0.89(1.56)	5.00	3.77	0.74	0.04	0.01(1.25)	7.00	3.51	0.46	0.04
		WQR-ALASSO	**0.94(1.60)**	**43.50**	**4.98**	1.01	0.01	**0.02(1.26)**	**64.50**	**5.00**	0.64	0.01
		WQR-AE-NET	1.02(1.65)	39.50	4.84	0.87	0.01	0.03(1.34)	51.00	4.74	0.56	0.02
$D2-t_d$	$d=6,\sigma=1$	QR-LASSO	0.82(1.30)	14.00	3.40	0.01	0.02	-0.01(1.27)	11.50	3.31	0.00	0.02
		QR-E-NET	1.18(1.80)	0.00	1.90	0.00	0.02	-0.02(1.60)	0.00	1.89	0.00	0.02
		QR-ALASSO	**0.81(1.29)**	**88.50**	**4.88**	0.00	0.01	**-0.02(1.26)**	**90.00**	**4.90**	0.00	0.01
		QR-AE-NET	0.96(1.52)	0.00	2.43	0.00	0.02	-0.02(1.47)	0.00	2.40	0.00	0.02
	$d=6,\sigma=0.5$	QR-LASSO	0.40(0.64)	16.00	3.51	0.01	0.02	-0.02(0.61)	10.50	3.32	0.00	0.03
		QR-E-NET	0.62(1.02)	0.00	1.90	0.00	0.02	0.00(0.84)	0.00	1.94	0.00	0.02
		QR-ALASSO	**0.40(0.64)**	**92.00**	**4.92**	0.00	0.01	**-0.01(0.62)**	**93.00**	**4.93**	0.00	0.02
		QR-AE-NET	0.51(0.78)	0.00	2.38	0.00	0.02	-0.02(0.72)	0.00	2.33	0.00	0.02
	$d=6,\sigma=1$	WQR-LASSO	0.50(0.97)	42.00	4.10	0.00	0.04	-0.03(0.83)	48.00	4.29	0.00	0.04
		WQR-E-NET	0.51(1.00)	1.50	2.72	0.00	0.04	-0.03(0.86)	3.00	2.69	0.00	0.04
		WQR-ALASSO	**0.47(0.93)**	**93.50**	**4.94**	0.00	0.02	**-0.04(0.84)**	**97.50**	**4.98**	0.00	0.03
		WQR-AE-NET	**0.52(0.99)**	**68.50**	**4.66**	0.00	0.03	**-0.03(0.87)**	**70.00**	**4.67**	0.00	0.04
	$d=6,\sigma=0.5$	WQR-LASSO	0.26(0.49)	31.00	3.92	0.00	0.04	0.00(0.41)	39.50	4.16	0.00	0.04
		WQR-E-NET	0.27(0.51)	2.50	2.62	0.00	0.04	0.00(0.43)	3.00	2.78	0.00	0.04
		WQR-ALASSO	**0.27(0.50)**	**95.50**	**4.95**	0.00	0.02	**0.00(0.41)**	**98.00**	**4.98**	0.00	0.02
		WQR-AE-NET	**0.31(0.54)**	**78.50**	**4.76**	0.00	0.03	**-0.01(0.44)**	**83.50**	**4.83**	0.00	0.04
$D3-t_d$	$d=1,\sigma=1$	QR-LASSO	1.32(2.39)	17.00	3.84	0.43	0.03	-0.02(2.39)	28.50	4.30	0.46	0.06
		QR-E-NET	2.07(3.49)	0.00	1.77	0.30	0.02	-0.07(3.56)	1.00	1.52	0.25	0.02
		QR-ALASSO	1.35(2.38)	49.50	4.68	0.46	0.00	**-0.03(2.28)**	**65.00**	**4.91**	0.50	0.01
		QR-AE-NET	1.93(3.30)	0.50	2.12	0.31	0.00	-0.05(3.35)	0.50	2.02	0.26	0.00
	$d=1,\sigma=0.5$	QR-LASSO	0.77(1.28)	56.00	4.55	0.25	0.03	0.01(1.23)	63.00	4.71	0.31	0.04
		QR-E-NET	1.75(2.65)	0.00	2.17	0.20	0.02	0.00(2.15)	0.00	2.08	0.21	0.02
		QR-ALASSO	0.85(1.43)	8.00	4.18	0.26	0.00	0.02(1.22)	29.50	4.45	0.32	0.00
		QR-AE-NET	1.65(2.44)	0.00	2.25	0.18	0.00	0.03(2.05)	0.00	2.35	0.22	0.00
	$d=1,\sigma=1$	WQR-LASSO	1.50(2.62)	37.00	4.96	1.51	0.04	-0.04(2.30)	51.00	4.96	1.12	0.05
		WQR-E-NET	1.61(2.78)	17.00	4.45	1.36	0.04	-0.03(2.48)	21.00	4.21	0.96	0.05
		WQR-ALASSO	1.49(2.59)	30.00	4.98	1.59	0.00	-0.04(2.29)	38.50	4.99	1.29	0.01
		WQR-AE-NET	1.64(2.79)	36.00	4.75	1.37	0.00	-0.04(2.45)	46.00	4.70	0.98	0.01
	$d=1,\sigma=0.5$	WQR-LASSO	**0.67(1.30)**	**67.50**	**4.87**	0.54	0.04	**-0.01(1.17)**	**69.50**	**4.88**	0.49	0.04
		WQR-E-NET	0.77(1.47)	24.00	3.91	0.47	0.04	-0.01(1.29)	10.50	3.36	0.40	0.04
		WQR-ALASSO	**0.69(1.28)**	**69.00**	**4.99**	0.66	0.02	**0.00(1.14)**	**74.50**	**5.00**	0.60	0.03
		WQR-AE-NET	**0.77(1.43)**	**65.50**	**4.82**	0.49	0.03	**0.00(1.29)**	**65.50**	**4.83**	0.47	0.04
$D3-t_d$	$d=6,\sigma=1$	QR-LASSO	0.82(1.29)	3.50	2.91	0.02	0.01	0.02(1.26)	2.50	2.67	0.01	0.02
		QR-E-NET	1.57(2.23)	0.00	0.17	0.00	0.01	0.00(2.48)	0.00	0.02	0.02	0.02
		QR-ALASSO	**0.82(1.28)**	**94.50**	**4.97**	0.02	0.01	**0.04(1.20)**	**94.50**	**4.95**	0.00	0.01
		QR-AE-NET	1.03(1.56)	0.00	2.52	0.00	0.01	0.04(1.63)	0.00	2.27	0.00	0.02
	$d=6,\sigma=0.5$	QR-LASSO	0.37(0.66)	4.00	2.84	0.00	0.01	-0.02(0.64)	1.50	2.62	0.00	0.01
		QR-E-NET	1.04(1.66)	0.00	0.14	0.00	0.01	-0.07(2.03)	0.00	0.01	0.00	0.02
		QR-ALASSO	**0.38(0.64)**	**92.00**	**4.92**	0.00	0.01	**-0.03(0.61)**	**89.50**	**4.88**	0.00	0.02
		QR-AE-NET	0.53(0.90)	0.00	1.40	0.00	0.01	-0.06(1.08)	0.00	1.06	0.00	0.02
	$d=6,\sigma=1$	WQR-LASSO	0.40(0.76)	49.00	4.28	0.00	0.04	-0.01(0.69)	53.50	4.40	0.00	0.05
		WQR-E-NET	0.42(0.77)	4.50	2.89	0.00	0.04	-0.01(0.73)	7.00	2.84	0.00	0.04
		WQR-ALASSO	**0.40(0.76)**	**96.00**	**4.96**	0.01	0.03	**-0.01(0.68)**	**97.50**	**4.98**	0.00	0.03
		WQR-AE-NET	**0.43(0.75)**	**77.00**	**4.74**	0.00	0.04	**0.00(0.69)**	**70.50**	**4.66**	0.00	0.05
	$d=6,\sigma=0.5$	WQR-LASSO	0.19(0.36)	43.00	4.25	0.00	0.05	0.00(0.33)	47.50	4.26	0.00	0.05
		WQR-E-NET	0.22(0.38)	3.50	2.77	0.00	0.04	0.00(0.36)	4.00	2.84	0.00	0.05
		WQR-ALASSO	**0.19(0.36)**	**97.00**	**4.97**	0.00	0.03	**0.00(0.32)**	**97.00**	**4.97**	0.00	0.03
		WQR-AE-NET	**0.21(0.38)**	**70.00**	**4.67**	0.00	0.04	**-0.01(0.33)**	**62.00**	**4.58**	0.00	0.05

t_d denotes t-distribution with d degrees of freedom.

Table 5.13 shows the results of variable/model selection and prediction performance in the presence of collinearity hiding points $D3$ under the t-distribution. The adaptive penalized QR and WQR versions performed better than the non-adaptive penalized versions with respect to all the metrics.

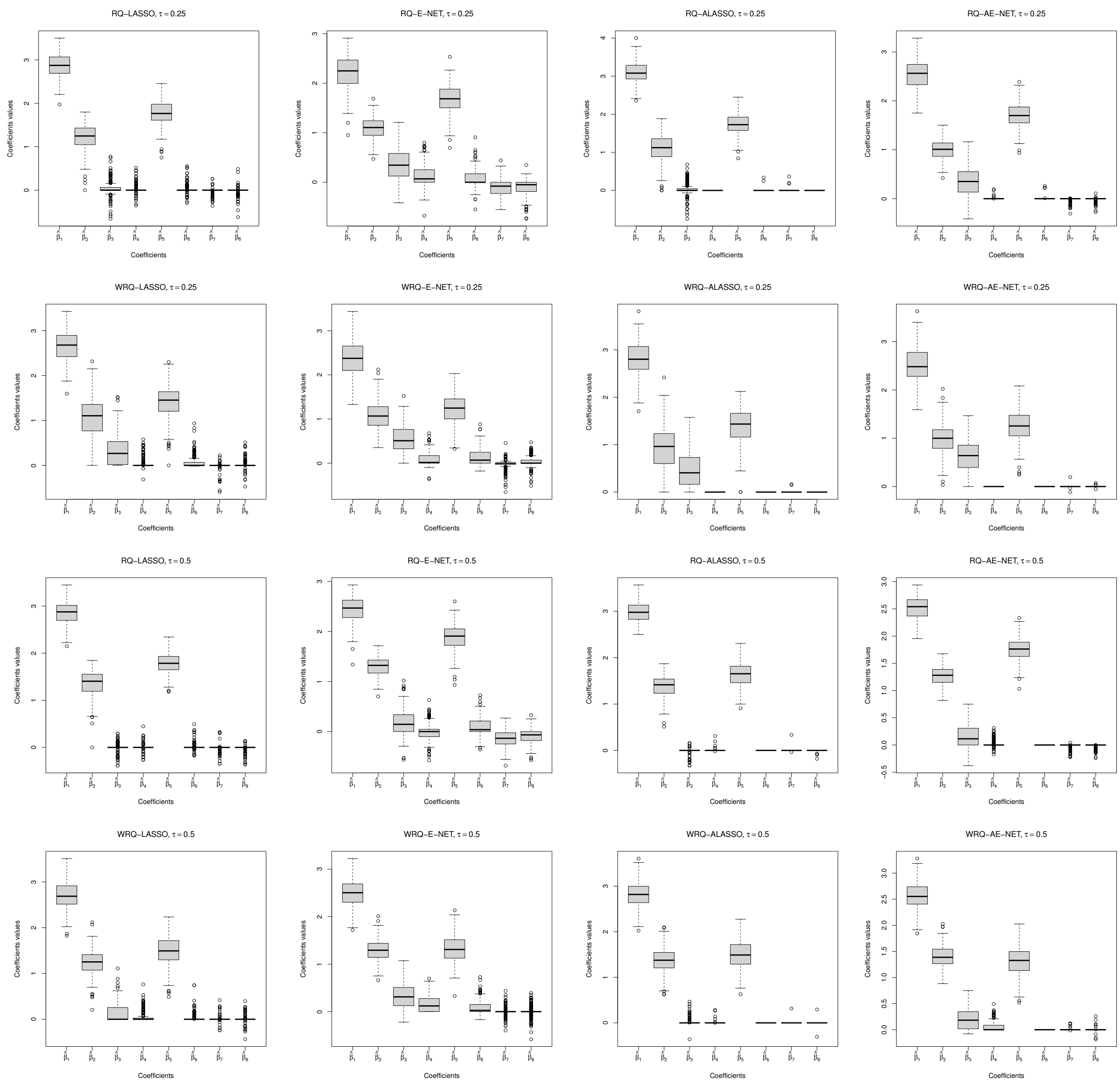


Figure 5.20: Box plots at $\tau = 0.25$ and $\tau = 0.50$ quantile level for $D2$ scenario under the t-distribution with $\sigma = 1$. Horizontal panels 1 and 2 are at $\tau = 0.25$ and horizontal panels 3 and 4 are at $\tau = 0.50$ RQ level (weighted and unweighted). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

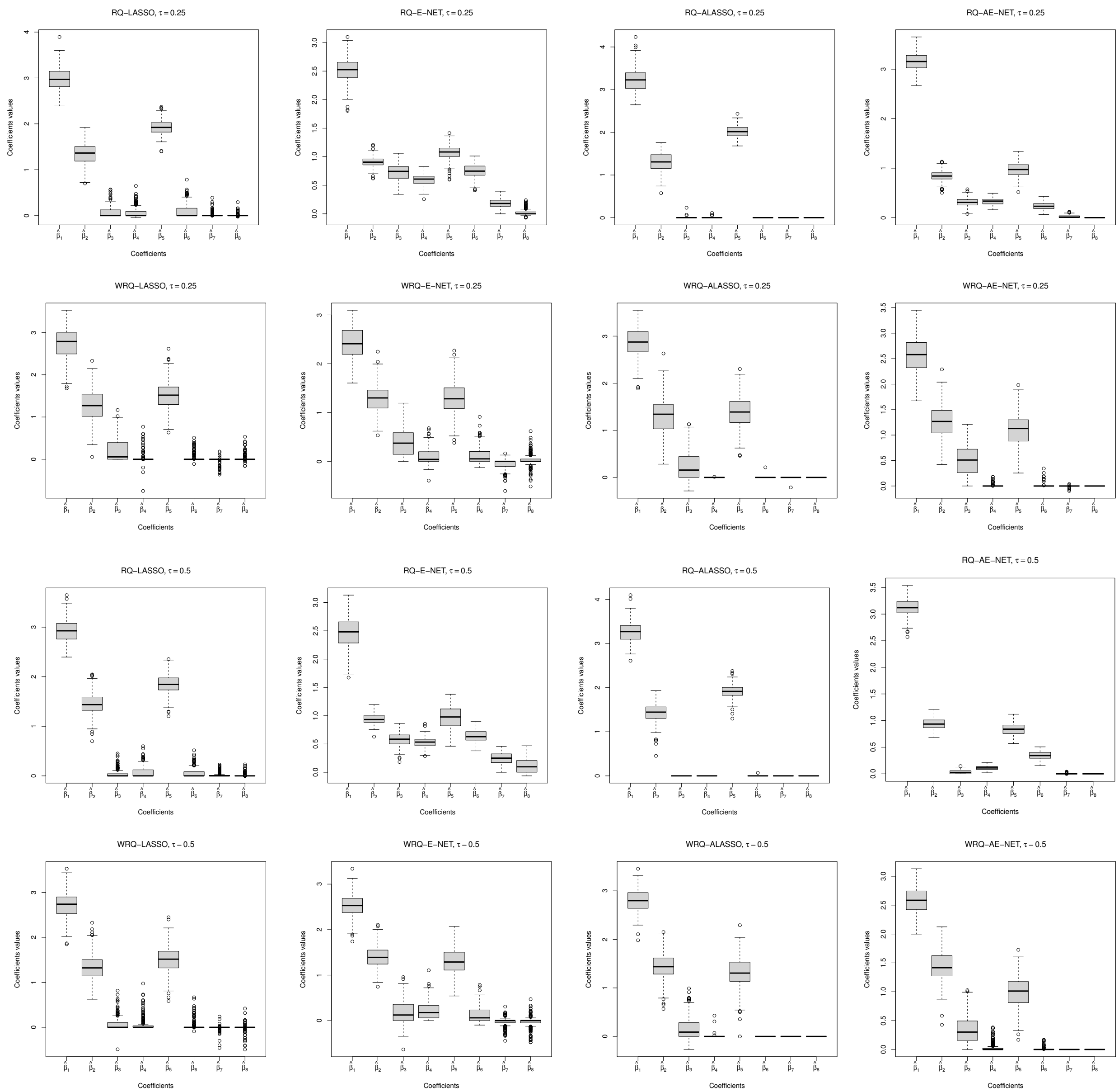


Figure 5.21: Box plots at $\tau = 0.25$ and $\tau = 0.50$ quantile level for $D3$ scenario under the t-distribution with $\sigma = 1$. Horizontal panels 1 and 2 are at $\tau = 0.25$ and horizontal panels 3 and 4 are at $\tau = 0.50$ *RQ* level (weighted and unweighted). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

The pairwise comparisons show the *QR-ALASSO* and *QR-AE-NET* procedures dominating the non-adaptive versions. The *QR-ALASSO* and *QR-AE-NET* procedures outperform the non-adaptive versions with respect to prediction, that is, 63% and 75% of the time, thus correctly fitting the models 100% and 38% of the time, respectively and correctly shrinking zero coefficients (both 100% of the time). However, these penalized *QR* procedures perform equally at correctly fitting the models 62% of the time.

The *WQR-ALASSO* and *WQR-AE-NET* procedures outperform the unweighted adaptive penalized versions overall, though marginally with respect to correctly fitting models and correctly shrinking zero coefficients (see also *Figure* 5.20 and 5.21). The *WQR-ALASSO* procedure outperforms all other penalized procedures with respect to all metrics, though marginally with respect to the last two metrics (percentage of correctly fitting models and the average number correctly shrunk zero coefficients).

Table 5.14: Weighted and unweighted quantile regression at *D*6 under heavy tailed t-distribution for $n = 50$ ($\tau = 0.25$ and $\tau = 0.50$); bold text indicate better performance.

			$\tau = 0.25$					$\tau = 0.50$					
			median(MAD)	Correctly	No of Zeros			median(MAD)	Correctly	No of Zeros			
Distribution	Parameter	Method	Test Error	Fitted	c.zero	inc.zero	median(λ)	Test Error	Fitted	c.zero	inc.zero	median(λ)	
$D6-t_d$	$d=1,\sigma=1$	QR-LASSO	2.54(4.43)	22.50	4.71	1.34	0.04	-0.06(3.48)	37.00	4.69	0.90	0.04	
		QR-E-NET	2.79(4.69)	3.00	3.06	0.96	0.04	-0.06(3.72)	1.50	2.67	0.76	0.04	
		QR-ALASSO	2.69(4.59)	21.00	4.95	1.64	0.02	**-0.05(3.45)**	**42.50**	**4.98**	1.14	0.02	
		QR-AE-NET	0.73(4.67)	10.00	4.19	0.99	0.03	-0.07(3.72)	11.50	4.02	0.77	0.03	
	$d=1,\sigma=0.5$	QR-LASSO	1.10(1.95)	25.50	4.15	0.52	0.03	-0.04(1.62)	33.00	4.30	0.36	0.04	
		QR-E-NET	1.34(2.37)	0.00	2.36	0.44	0.03	0.00(1.83)	0.00	2.45	0.33	0.03	
		QR-ALASSO	**1.11(1.96)**	**74.00**	**4.98**	0.56	0.02	**-0.02(1.60)**	**81.50**	**5.00**	0.43	0.02	
		QR-AE-NET	1.34(2.37)	22.50	4.22	0.48	0.03	-0.02(1.87)	12.00	4.01	0.36	0.04	
	$d=1,\sigma=1$	WQR-LASSO	1.63(3.05)	14.00	4.66	1.52	0.04	0.04(2.47)	26.00	4.65	1.09	0.04	
		WQR-E-NET	1.82(3.22)	1.00	3.61	1.07	0.04	0.05(2.74)	2.50	3.57	0.81	0.04	
		WQR-ALASSO	1.67(3.11)	13.00	4.79	1.66	0.02	0.06(2.46)	23.50	4.75	1.24	0.02	
		WQR-AE-NET	1.84(3.28)	2.00	4.25	1.27	0.02	0.04(2.71)	0.00	4.05	0.98	0.03	
	$d=1,\sigma=0.5$	WQR-LASSO	**1.08(1.99)**	**67.00**	**4.80**	0.38	0.04	**0.02(1.52)**	**58.50**	**4.67**	0.33	0.04	
		WQR-E-NET	1.27(2.23)	1.00	2.45	0.32	0.03	0.03(1.78)	0.00	2.02	0.29	0.04	1
		WQR-ALASSO	**1.05(1.98)**	**79.00**	**5.00**	0.46	0.02	**0.02(1.58)**	**83.00**	**4.99**	0.36	0.01	
		WQR-AE-NET	1.25(2.24)	25.00	4.10	0.32	0.01	0.02(1.76)	8.00	3.59	0.29	0.02	
$D6-t_d$	$d=6,\sigma=1$	QR-LASSO	0.82(1.25)	53.50	4.26	0.00	0.04	0.05(1.26)	51.00	4.35	0.00	0.05	
		QR-E-NET	0.86(1.35)	3.00	2.44	0.00	0.03	0.03(1.30)	1.50	2.29	0.00	0.04	
		QR-ALASSO	**0.83(1.24)**	**96.00**	**4.96**	0.00	0.02	**0.05(1.23)**	**95.50**	**4.95**	0.00	0.03	
		QR-AE-NET	0.89(1.31)	30.50	4.03	0.00	0.03	0.04(1.29)	29.00	4.05	0.00	0.04	
	$d=6,\sigma=0.5$	QR-LASSO	**0.40(0.66)**	**63.50**	**4.51**	0.00	0.04	**0.00(0.64)**	**72.50**	**4.65**	0.00	0.05	
		QR-E-NET	0.41(0.72)	2.50	2.77	0.00	0.04	0.01(0.67)	5.00	3.00	0.00	0.04	
		QR-ALASSO	**0.40(0.65)**	**88.00**	**4.88**	0.00	0.03	**0.01(0.63)**	**93.50**	**4.93**	0.00	0.03	
		QR-AE-NET	0.42(0.70)	24.50	4.08	0.00	0.03	0.00(0.67)	25.50	4.01	0.00	0.04	
	$d=6,\sigma=1$	WQR-LASSO	0.54(0.92)	43.00	4.23	0.00	0.04	0.01(0.87)	55.50	4.45	0.00	0.04	
		WQR-E-NET	0.55(0.96)	2.00	3.07	0.00	0.04	-0.01(0.89)	6.00	3.23	0.00	0.04	
		WQR-ALASSO	**0.55(0.91)**	**69.50**	**4.69**	0.01	0.02	**0.00(0.85)**	**83.00**	**4.82**	0.00	0.02	
		WQR-AE-NET	0.60(0.98)	25.50	4.21	0.00	0.03	-0.01(0.93)	37.50	4.32	0.00	0.04	
	$d=6,\sigma=0.5$	WQR-LASSO	0.26(0.47)	35.50	4.06	0.00	0.04	-0.01(0.44)	44.00	4.19	0.00	0.04	
		WQR-E-NET	0.25(0.51)	5.00	3.16	0.00	0.03	0.00(0.47)	5.00	3.13	0.00	0.04	
		WQR-ALASSO	**0.26(0.46)**	**72.50**	**4.72**	0.00	0.02	**-0.01(0.42)**	**85.00**	**4.85**	0.00	0.02	
		WQR-AE-NET	0.26(0.48)	14.00	4.06	0.00	0.03	0.00(0.44)	31.00	4.26	0.00	0.03	

t_d denotes t-distribution with d degrees of freedom.

Heavy-tailed distribution scenario (D6)

Table 5.14 shows simulation results of variable/model selection and prediction performance at the design matrix $D6$ with collinearity introduced by the exponential decay $0.50^{|j-i|}$ under the

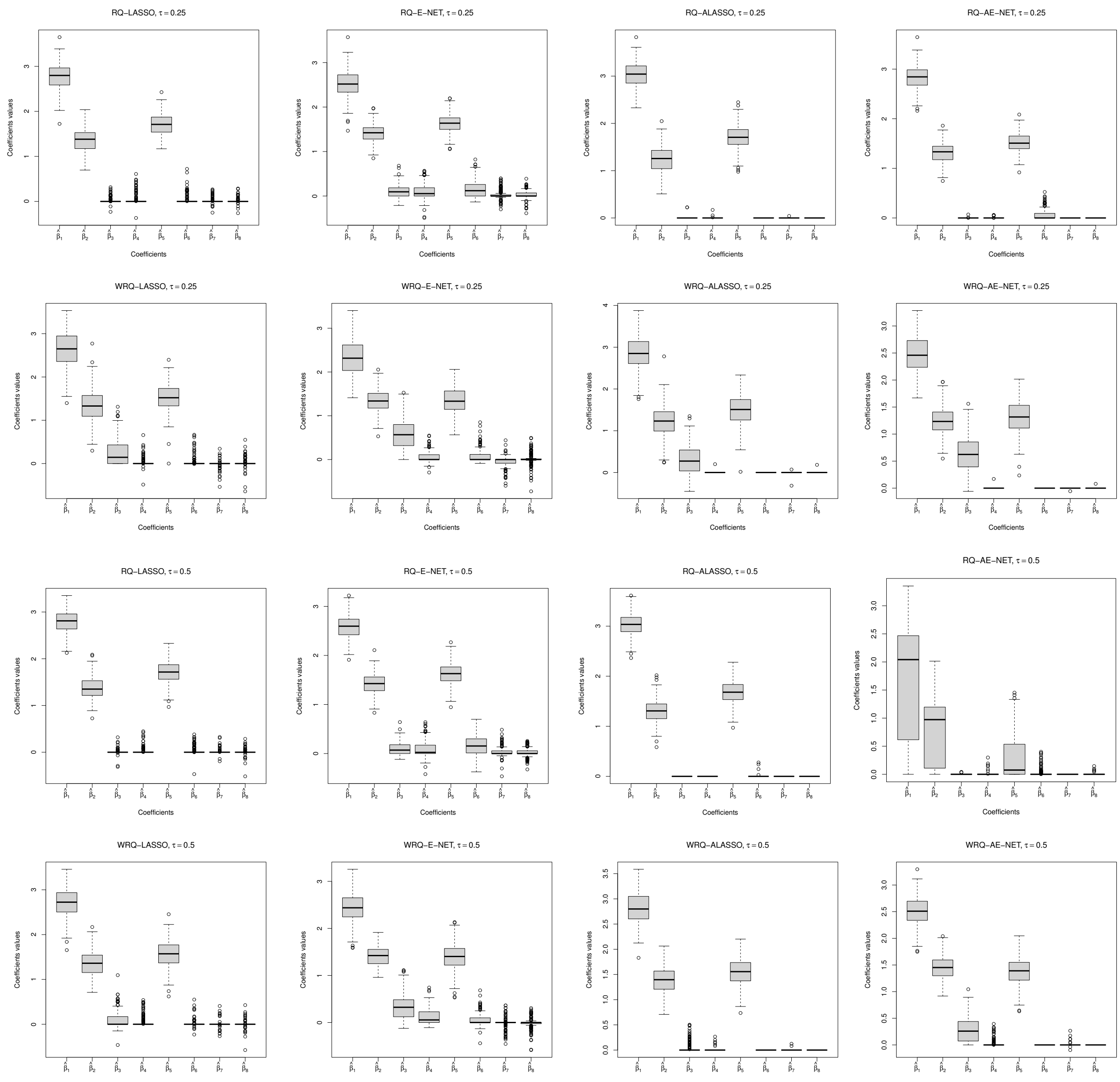


Figure 5.22: Box plots at $\tau = 0.25$ and $\tau = 0.50$ quantile level for $D6$ scenario under the t-distribution with $\sigma = 1$. Horizontal panels 1 and 2 are at $\tau = 0.25$ and horizontal panels 3 and 4 are at $\tau = 0.50$ *RQ* level (weighted and unweighted). The non-zero coefficients and zero coefficients are $(\beta_1, \beta_2, \beta_5)' = (3, 1.5, 2)'$ and $(\beta_3, \beta_4, \beta_6, \beta_7, \beta_8)' = (0, 0, 0, 0, 0)'$, respectively.

t-distribution with 1 and 6 degrees of freedom. In the unweighted scenarios, adaptive penalized procedures (*QR-ALASSO* and *QR-AE-NET*) outperform the non-adaptive versions with respect to all metrics. The *QR-ALASSO* and *QR-AE-NET* procedures outperform the non-adaptive versions with respect to prediction 75% and 62% of the time, and 100% and 88% of the time with respect to correctly fitting models, respectively. The adaptive procedures dominate the non-adaptive versions with respect to the average correctly shrunk zero coefficients 100% of the time in both cases.

The adaptive penalized *WQR* procedures exhibit the same pattern as the unweighted scenarios, with the *WQR-ALASSO* and *WQR-AE-NET* procedures dominating the weighted non-adaptive penalized versions (*WQR-LASSO* and *WQR-E-NET*) with respect to prediction, correctly fitted models and correctly shrunk zero coefficients. Both the *WQR-ALASSO* and *WQR-AE-NET* procedures outperform the non-adaptive weighted ones in prediction, 100% and correctly fit models 88% of the time. The adaptive penalized ones also correctly shrink zero coefficients better than the non-adaptive ones (see also *Figure* 5.22). The pairwise comparisons show the *WQR-ALASSO* and *WQR-AE-NET* procedures dominating the unweighted versions in prediction, though all procedures perform more or less similar with respect to correctly fitted models and correctly shrunk zero coefficients.

5.5 Applications of Adaptive Penalized Procedures to Well-known Data Sets from the Literature

The Jet-Turbine Engine (Montgomery et al. 2009) and Gunst and Mason Gunst & Mason (1980) data sets are used in this section to demonstrate the efficacy of adaptive weights $\breve{\omega}_j$ based on

$\hat{\beta}_j^{WR}(\tau)$ estimates in regularized QR and WQR procedures. In the literature, these data sets are used to illustrate the effectiveness of some robust methodologies to mitigate against adverse effects of collinearity and collinearity influential points. The Jet-Turbine Engine data set is very popular with Engineers and has high collinearity reducing points (see Montgomery et al. 2009, Bagheri & Midi 2012). The Gunst and Mason data set, unlike the Jet-Turbine Engine data set, is known to contain collinearity-inducing points (see also Gunst & Mason 1980). We measure the performance of adaptive procedures using biases of the estimated coefficients and how zero coefficients are correctly penalized to zero.

5.5.1 The Jet-Turbine Engine Data Set

This section compares the performance of adaptive versus non-adaptive penalized QR and WQR procedures using the Jet-Turbine Engine data set (Montgomery et al. 2009). The 40 observation Jet-Turbine Engine data set consists of variables namely, (i) the response variable "thrust of a jet-turbine engine" denoted by Y and (ii) predictor variables, primary speed of rotation denoted by X_1, secondary speed of rotation denoted by X_2, fuel flow rate denoted by X_3, pressure denoted by X_4, exhaust temperature denoted by X_5 and ambient temperature at time of test denoted by X_6. The observations 6 and 20 are high leverage collinearity reducing points (see Bagheri & Midi 2012). The data is standardized, and the predictor variable generated by the equation: $Y_1 = \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \varepsilon_1$, $\varepsilon_1 \sim t_6$ for 38 observations, excluding high leverage points 6 and 20 with $\boldsymbol{X}_2$ and $\boldsymbol{Y}_2$ comprising cases (observations) 6 and 20, such that $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)'$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')'$, where $\boldsymbol{\beta}_1 = (50, 0, 0, 10, 15, 0)'$.

Table 5.15 shows parameter estimates (βs) and their respective biases from the true coefficients/true βs ($\boldsymbol{\beta}_1 = (50,0,0,10,15,0)'$). In the majority of cases, the zero coefficients (β_2, β_3, β_6) are correctly shrunk to zero (see *Table* 5.15). The adaptive regularized and non-adaptive regularized *QR* and *WQR* procedures are compared in terms of their biases from the true coefficients ($\boldsymbol{\beta}_1 = (50,0,0,10,15,0)'$) and the penalization to zero of zero coefficients. Zero coefficients are correctly penalized to zero/near zero as expected in all versions of the regularized procedures. The adaptive penalized versions are more than 50% better than the non-adaptive penalized versions (2 out of 4 zero coefficients are exact zeros). The penalized *WQR* procedures outperform the unweighted versions in the collinearity reducing scenario at the $\tau = 0.50$ *QR* levels, and the converse is true at the $\tau = 0.25$ *RQ* levels. The adaptive penalized *QR* and *WQR* versions outperform the non-adaptive penalized procedures at both $\tau = 0.50$ and $\tau = 0.25$ for *WQR-ALASSO*. The best performing procedure is the *WQR-ALASSO* at $\tau = 0.50$. In most cases, the optimal λ is close to zero ($\lambda < 0.01$).

5.5.2 The Gunst and Mason Data Set

We assess the performance of the adaptive regularized *QR* and *WQR* procedures in this section using estimated coefficient bias and how zero coefficients are correctly penalized to zero in the Gunst and Mason data set (Gunst & Mason 1980). The 49 observation data set (data set of 49 countries), where the response variable "gross national product" is denoted by $Y = GNP$ and predictor variables are as follows: (i) infant death rate denoted by $X_1 = TNFD$, (ii) physician to population ratio denoted by $X_2 = PHYS$, (iii) population density denoted by $X_3 = DENS$, (iv) density of agricultural land denoted by $X_4 = AGDS$, (v) measure of literacy $X_5 = LIT$ and (vi) higher education index denoted by $X_6 = HIED$. In the literature, predictor variables *DENS* and *AGDS* are strongly

Table 5.15: Regularized QR and WQR parameter estimation results for the Jet-Turbine Engine data set. (Montgomery et al. 2009); bold text indicate better performance.

		β	NON-BIASED $\beta(Bias)$	QR-LASSO $\beta(Bias)$	QR-E-NET $\beta(Bias)$	QR-ALASSO $\beta(Bias)$	QR-AE-NET $\beta(Bias)$	WQR-LASSO $\beta(Bias)$	WQR-E-NET $\beta(Bias)$	WQR-ALASSO $\beta(Bias)$	WQR-AE-NET $\beta(Bias)$
$\tau = 0.25$	λ			0.01	0.01	0.07	0.02	0.01	0.01	0.00	0.00
	intercept	-0.72	-1.11(0.39)	-0.60(-0.12)	-0.69(-0.03)	-0.72(0.00)	-0.67(-0.05)	-0.10(-0.62)	-0.08(-0.64)	-0.07(-0.65)	-0.23(-0.49)
	X_1	50.00	15.15(-34.85)	33.20(-16.80)	21.63(-28.37)	17.55(-32.45)	24.53(-25.47)	35.34(-14.66)	26.98(-23.02)	44.01(-5.99)	51.84(1.84)
	X_2	0.00	84.78(84.78)	10.34(10.34)	14.14(14.14)	26.40(26.40)	27.17(27.17)	0.00(0.00)	0.00(0.00)	0.00(0.00)	-15.12(-15.12)
	X_3	0.00	-89.20(-89.20)	0.00(0.00)	0.00(0.00)	-20.53(-20.53)	-26.10(-26.10)	0.00(0.00)	0.25(0.25)	0.00(0.00)	-14.00(-14.00)
	X_4	10.00	28.94(18.94)	14.18(4.18)	19.48(9.48)	27.98(17.98)	25.27(15.27)	16.13(6.13)	22.52(12.52)	10.23(0.23)	20.28(10.28)
	X_5	15.00	30.54(15.54)	16.51(1.51)	18.57(3.57)	22.19(7.19)	22.54(7.54)	25.85(10.85)	27.79(12.79)	23.02(8.02)	35.00(20.00)
	X_6	0.00	10.21(10.21)	-4.28(-4.28)	-4.75(-4.75)	-0.78(-0.78)	0.57(0.57)	-9.56(-9.56)	-10.65(-10.65)	-8.81(-8.81)	-7.73(-7.73)
$\tau = 0.50$	λ			0.08	0.03	0.32	0.10	0.04	0.00	0.97	0.42
	intercept	0.00	0.40(0.40)	0.44(0.44)	0.39(0.39)	0.57(0.57)	0.53(0.53)	**0.01(0.01)**	0.02(0.02)	**0.01(0.01)**	0.01(0.01)
	X_1	50.00	38.47(-11.53)	9.68(-40.32)	16.28(-33.72)	27.98(-22.02)	16.09(-33.91)	**49.19(-0.81)**	41.15(-8.85)	**54.16(4.16)**	20.42(-29.58)
	X_2	0.00	41.62(41.62)	33.30(33.30)	17.72(17.72)	35.97(35.97)	21.51(21.51)	**0.00(0.00)**	0.00(0.00)	**0.00(0.00)**	10.97(10.97)
	X_3	0.00	-37.77(-37.77)	0.00(0.00)	8.14(8.14)	0.00(0.00)	0.00(0.00)	**0.00(0.00)**	-5.85(-5.85)	**0.00(0.00)**	6.34(6.34)
	X_4	10.00	12.73(2.73)	21.57(11.57)	18.00(8.00)	5.43(-4.57)	21.60(11.60)	**0.00(-10.00)**	18.47(8.47)	**0.80(-9.20)**	17.20(7.20)
	X_5	15.00	19.06(4.06)	6.02(-8.98)	12.42(-2.58)	1.56(-13.44)	13.52(-1.48)	**14.63(-0.37)**	21.20(6.20)	**18.96(3.96)**	18.26(3.26)
	X_6	0.00	4.21(4.21)	0.00(0.00)	-3.06(-3.06)	0.00(0.00)	-0.90(-0.90)	**0.00(0.00)**	4.58(4.58)	**0.00(0.00)**	0.00(0.00)

[1] The intercept $F^{-1}(\tau)+\boldsymbol{\beta}_0$, translate to 0 and -0.72 under the t_6 error term distribution, at quantile levels $\tau = 0.50$ and $\tau = 0.25$, respectively. The biases are calculated by the difference between the true and estimated parameters ($\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$). They show the ability of our regularization and variable selection procedures in QR procedures, to shrink some zero coefficients to zero.

correlated (high collinearity) with few high leverage points. These outlying and influential points are namely, 7 (Canada), 13 (El Salvador), 17 (Hong Kong), 20 (India), 39 (Singapore) and 46 (United States of America), with observations 17 and 39 being collineairity inducing points. Data are standardized and the response variable is generated by the equation $Y_1 = \boldsymbol{X}_1'\boldsymbol{\beta}_1 + \varepsilon_1$, $\varepsilon_1 \sim t_6$ for the first 43 observations, with $\boldsymbol{X}_2$ and $\boldsymbol{Y}_2$ comprising observations 7, 13, 17, 20, 39 and 46 such that $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1', \boldsymbol{Y}_2')'$, where $\boldsymbol{\beta}_1 = (0, 8, -13, 0, 0, 6)'$.

We summarize the results in *Table* 5.16, by showing estimated βs and their corresponding biases. The zero coefficients (β_1, β_4, β_5) are correctly penalized to zero and show low estimation biases. The non-zero coefficients are very close to the true β values $(\beta_2, \beta_3, \beta_6)' = (8, -13, 6)'$ at the optimal λ both at the $\tau = 0.25$ and $\tau = 0.50$ *RQ* levels. The adaptive versions outperform the non-adaptive versions in prediction and correct penalization most of the time. The adaptive versions have fewer biases 67% of the time. The *WQR-ALASSO* outperforms the rest of the procedures. The *WQR-ALASSO* correctly penalizes zero coefficients, that is, 100% and 50% of the time more than the rest of the procedures at $\tau = 0.50$ and $\tau = 0.25$ *QR* levels, respectively. Adaptive penalized procedures are better than non-adaptive penalized procedures 67% of the time. The *ALASSO* penalized procedures are marginally better than the rest of the procedures at $\tau = 0.50$. Most unweighted models show an affinity to be optimal when $\lambda = 0$.

5.6 Concluding Remark

In this chapter, we performed simulation studies by considering six predictor space design matrices with data aberrations comprising high leverage points, collinearity influential points and collinearity under different distribution scenarios (Gaussian and t-distributions). We investigated

Table 5.16: Regularized QR and WQR parameter estimation results for the Gunst and Mason data set (Gunst & Mason 1980); bold text indicate better performance.

		β	NON-BIASED $\beta(Bias)$	QR-LASSO $\beta(Bias)$	QR-E-NET $\beta(Bias)$	QR-ALASSO $\beta(Bias)$	QR-AE-NET $\beta(Bias)$	WQR-LASSO $\beta(Bias)$	WQR-E-NET $\beta(Bias)$	WQR-ALASSO $\beta(Bias)$	WQR-AE-NET $\beta(Bias)$
$\tau = 0.25$	λ			0.00	0.00	0.01	0.00	0.04	0.04	0.05	0.06
	intercept	-0.72	0.38(-1.10)	-0.85(0.13)	-0.85(0.13)	-0.98(0.26)	-0.69(0.03)	-0.16(-0.56)	-0.10(-0.62)	**-0.11(-0.61)**	**-0.11(-0.61)**
	X_1	0.00	-0.14(0.14)	-0.02(0.02)	0.00(0.00)	0.00(0.00)	-0.19(0.19)	-2.31(2.31)	-2.44(2.44)	**0.00(0.00)**	**0.00(0.00)**
	X_2	8.00	9.29(-1.29)	4.37(3.63)	4.89(3.11)	2.97(5.03)	7.73(0.27)	8.33(-0.33)	9.01(-1.01)	**6.43(1.57)**	**6.44(1.56)**
	X_3	-13.00	-10.97(-2.03)	0.00(-13.00)	-2.00(-11.00)	-6.35(-6.65)	-9.59(-3.41)	0.00(-13.00)	0.00(-13.00)	**-0.50(-12.50)**	**-0.52(-12.48)**
	X_4	0.00	11.54(-11.54)	0.87(-0.87)	3.01(-3.01)	6.68(-6.68)	10.09(-10.09)	0.00(0.00)	0.00(0.00)	**0.00(0.00)**	**0.00(0.00)**
	X_5	0.00	6.21(-6.21)	1.62(-1.62)	2.09(-2.09)	0.00(0.00)	4.90(-4.90)	0.00(0.00)	0.00(0.00)	**0.00(0.00)**	**0.00(0.00)**
	X_6	6.00	1.61(4.39)	2.22(3.78)	2.19(3.81)	3.30(2.70)	1.74(4.26)	9.03(-3.03)	9.33(-3.33)	**8.42(-2.42)**	**8.42(-2.42)**
$\tau = 0.50$	λ			0.01	0.02	0.01	0.01	0.04	0.04	0.01	0.03
	intercept	0.00	0.44(-0.44)	0.03(-0.03)	0.03(-0.03)	**0.14(-0.14)**	0.07(-0.07)	-0.01(0.01)	0.00(0.00)	**0.00(0.00)**	**-0.01(0.01)**
	X_1	0.00	0.72(-0.72)	0.40(-0.40)	0.41(-0.41)	**0.00(0.00)**	0.00(0.00)	0.00(0.00)	0.00(0.00)	**0.00(0.00)**	**0.00(0.00)**
	X_2	8.00	9.87(-1.87)	10.13(-2.13)	10.13(-2.13)	**7.83(0.17)**	8.21(-0.21)	7.48(0.52)	6.27(1.73)	**8.37(-0.37)**	**8.44(-0.44)**
	X_3	-13.00	-5.15(-7.85)	-4.80(-8.20)	-4.80(-8.20)	**-0.98(-12.02)**	-8.41(-4.59)	-5.75(-7.25)	-5.99(-7.01)	**-9.29(-3.71)**	**-9.67(-3.33)**
	X_4	0.00	4.68(-4.68)	4.38(-4.38)	4.38(-4.38)	**0.00(0.00)**	6.89(-6.89)	0.00(0.00)	0.00(0.00)	**0.00(0.00)**	**0.00(0.00)**
	X_5	0.00	3.64(-3.64)	3.71(-3.71)	3.71(-3.71)	**0.00(0.00)**	0.81(-0.81)	-0.18(0.18)	-1.19(1.19)	**0.00(0.00)**	**0.00(0.00)**
	X_6	6.00	4.94(1.06)	4.92(1.08)	4.91(1.09)	**6.24(-0.24)**	4.31(1.69)	10.68(-4.68)	11.21(-5.21)	**11.38(-5.38)**	**11.32(-5.32)**

[1] The intercept $F^{-1}(\tau) + \boldsymbol{\beta}_0$, translate to 0 and -0.72 under the t_6 error term distribution, at quantile levels $\tau = 0.50$ and $\tau = 0.25$, respectively. The biases are calculated by the difference between the true and estimated parameters ($\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}$). They show the ability of our regularization and variable selection procedures in QR procedures, to shrink some zero coefficients to zero.

the finite sample performance of regularized QR and WQR procedures. In Section 5.1, we described the design of the simulation studies. In Section 5.2, we presented results for the penalized WQR procedures and in Section 5.3, we applied the penalized WQR procedures to well-known data sets in the literature. We consider the Hawkins, Bradu and Kass data set in Subsection 5.3.1 and the Hocking and Pendleton data set in Subsection 5.3.2. In Section 5.4, we presented results for the adaptive penalized QR and WQR procedures. Applications of the adaptive penalized QR and WQR procedures to well-known data sets in the literature were presented in Section 5.5. In Subsections, 5.5.1 and 5.5.2, we presented applications to the Jet-Turbine Engine data set and the Gunst and Mason data set, respectively. In Sections 5.2 and 5.4, MAD of test errors, percentages of correctly fitted models and average number of correctly/incorrectly fitted zero coefficients were used to measure the performance of the penalized procedures. In addition to these metrics, box plots and stacked bar graphs were also used. The suggested penalized procedures performed better than existing ones in the majority of cases (see Chapter 6 (Contributions and Recommendations Chapter) for a more detailed discussion). The discussion of all the results and recommendations for further study are presented in Chapter 6.

The penalized WQR procedures used in this thesis to analyze low-dimensional problems can also be used to analyze high-dimensional problems, making them appropriate for examining complicated big data sets found in a variety of domains. The extended penalized QR framework will hopefully successfully combat the problem of dimensionality by encouraging sparsity and offering more precise estimates of the conditional quantiles in high-dimensional scenarios. The extension would increase the options for statistical methods for high-dimensional data analysis by giving researchers a powerful tool for examining collinearity among predictors in multidimensional datasets.

Chapter 6

Contributions and Recommendations for Further Study

In this thesis, we suggested several variable selection and regularization techniques in a *QR* setting. These regularization techniques are categorized broadly into penalized *WQR* and adaptive penalized *QR* and *WQR* techniques. The penalized *WQR* and adaptive (both weighted and unweighted) procedures were suggested as solutions to predictor space data aberrations' (collinearity, high leverage points and collinearity influential points) in mitigating against their adverse effects on *QR*. These high leverage points that induce or reduce collinearity are collinearity influential points. The suggested regularized methods are attractive because of their robustness in the presence of these data aberrations and heavy-tailed distributions. The first part of this study is premised on extending the *WLAD* procedure (Arslan 2012) to the penalized *WQR* framework with *LASSO*, *RIDGE* and *E-NET* penalties. The second part is premised on the proposed adaptive weights based *QR* procedures (both weighted and unweighted).

6.1 Contributions

Six predictor design matrix scenarios ($D1$-$D6$) were considered in the simulation study to investigate the efficacy of the suggested penalized WQR and adaptive penalized procedures in mitigating the adverse effects of collinearity, high leverage points and collinearity influential points. These design matrix scenarios range from the orthogonal case $D1$, the collinearity inducing points in $D2/D4$, the collinearity reducing case in $D3/D5$ to the exponential decay $0.5^{|j-i|}$ induced collinearity case coupled with high leverage points. All these design scenarios contain high leverage points, except $D1$. The design scenarios $D1$-$D5$ were considered under the Gaussian and t-distribution with $D6$ considered under the t-distributions only. The procedures were also applied to real data sets from the literature.

Firstly, the thesis suggested penalized WQR procedures using robust weights based on the computationally intensive high breakdown MCD method rather than the well-known classical Mahalanobis distance. We used the MCD-based weights ω_i that Arslan (2012) successfully implemented in the $WLAD$-$LASSO$. Though the implementation was in a $WLAD$ scenario, we extended it to the WQR scenarios with the accrued benefits of robustness in the predictor space. These robust weights are used in the formulation of the suggested robust penalized WQR methods (see Chapter 4). The $RIDGE$, $LASSO$ and E-NET penalized WQR procedures are the WQR-$RIDGE$, WQR-$LASSO$ and the WQR-E-NET procedures, respectively. These penalized WQR procedures are robust in the presence of high leverage points and collinearity influential points. The weights ω_i based on MCD down-weighs extreme points, thereby reducing their undue influence. These regularized WQR procedures, unlike the LS estimators, are also robust in the response space, since RQs influence functions are bounded in the Y-space.

We used a simulation study and applications to well known data sets from the literature to investigate the efficacy of the suggested WQR procedures in mitigating the adverse effects of high leverage points and collinearity influential points. The results show that generally penalized WQR procedures outperform the unweighted QR versions, with the WQR-$LASSO$ outperforming all others, albeit marginally to the WQR-E-NET in some case, in the $D2/D4$ design scenario under the Gaussian distribution. Although the WQR-$LASSO$ procedure dominates at $D2/D4$ design scenario under the Gaussian distribution, in few cases, it is dominated by the WQR-E-NET procedure in prediction. As explained by Zou & Hastie (2005), this occasional dominance of the WQR-E-NET version over the $LASSO$ penalized WQR version is to be expected . When the penalized procedures are applied to well-known data sets from the literature, the application of the MCD based robust weight is adequate (without penalty) in some cases, with the non-penalized versions outperforming the penalized versions, i.e., the optimal tuning parameter will be $\lambda = 0$. The WQR-$LASSO$ and WQR-E-NET procedures are too "greedy" at $\tau = 0.25$, shrinking all parameters to zero, and the best performance is mostly at the RQ level $\tau = 0.25$. Simulations show an improvement in variable selection due to the robust weighting formulation.

Simulations also show that WQR-$RIDGE$, WQR-$LASSO$ and the WQR-E-NET procedures perform better than QR variable selection procedures QR-$RIDGE$, QR-$LASSO$ and QR-E-NET, respectively. In conclusion, the simulation study and applications to the (Hawkins et al. 1984) and Hocking & Pendleton (1983) data sets show an improvement in variable selection and regularization due to the MCD based robust weighting formulation.

Secondly, this thesis suggested $ALASSO$ and AE-NET penalized WQR (WQR-$ALASSO$ and WQR-AE-NET) methodologies with a slope coefficient as the initial coefficient estimator used to compute adaptive weights. This study is premised on estimated local RQ estimator rather than the global conditional mean regression (*LS*) estimator, which motivated the choice of our adaptive weights. We used a *RIDGE* penalized *WQR* based coefficient estimate in the adaptive weights construction in which the weights are used to down-weigh extreme values (high leverage points) and are based on the minimum covariance determinant (*MCD*) estimator of Rousseeuw (1985). The *WQR-LASSO* and *WQR-E-NET* procedures (the initially proposed penalized *WQR* procedures) have been further extended to the *ALASSO* and *AE-NET* regularized *WQR* versions namely, the *WQR-ALASSO* and *WQR-AE-NET* procedures, as mitigation against both high leverage and collinearity influential observations. The adaptive regularized *QR*/*WQR* methodologies satisfy sparsity conditions, are asymptotically normally distributed and provide a balance between bias-variance trade-offs.

In this adaptive regularization framework, we carried out a comparative study of the models via simulation studies, as well as applications to well-known real data sets from the literature. With respect to correctly shrinking zero coefficients in the presence of collinearity-inducing or reducing observations, adaptive penalized procedures outperform non-adaptive versions in the vast majority of cases. The same pattern is exhibited in terms of percentage of correctly fitted models, and the same procedure is better with respect to prediction. Under both the normal and t-distribution scenarios, the regularized *WQR* (*WQR-ALASSO* and *WQR-AE-NET*) procedures outperform the non-adaptive penalized versions in prediction and percentage of correctly fitting models. As exhibited in the unweighted adaptive versions, we have the same pattern of results in the non-adaptive, where the unweighted ones are less effective than the weighted ones. The proposed adaptive penalized procedures were applied to the real-life data sets, including the Jet-Turbine Engine (Montgomery et al. 2009) used in the Engineering and the Gunst and Mason (Gunst & Mason 1980) data sets. Our adaptive penalized procedures showed that the proposed adaptive weights are effective. The real-life example depicts better performance by adaptive penalized procedures in comparison to the non-penalized versions (adaptive versions are better 67% the time), with the *WQR-ALASSO* outperforming all the procedures. Under the normal distribution scenarios (*D*1-*D*5), the adaptive versions (*QR-ALASSO*, *QR-AE-NET*, *WQR-ALASSO* and *WQR-AE-NET*) perform better than the non-adaptive versions when σ is small. In the presence of large errors, the *ALASSO* penalized *QR* is superior to other unweighted procedures, as in the Zou (2006) *LS* case. Under the heavy-tailed distribution scenario (see design matrices *D*2, *D*3 and *D*6 under the t-distribution), the same pattern of results is exhibited, as in the Gaussian cases, with the *ALASSO* penalized procedures being more effective variable/model selection and parameter estimation procedures than other versions. The suggested adaptive weights perform better than a constant one, and these adaptive weights have an advantage of having an adaptive weight at a particular *RQ* level. We conclude that the proposed adaptive variable selection and parameter estimation procedures effectively deal with collinearity and collinearity influential point adverse effects, particularly in cases of heavy-tailed distributions (t-distribution). In heavy-tailed distribution with higher degrees of freedom, the *AE-NET* penalty is recommended as an alternative to the *ALASSO* penalty.

Although the coefficient estimates from *LASSO* penalized *WQR* procedures are biased since they are shrunken towards zero, model simplicity is increased. The less biased, adaptive penalized *WQR* techniques offer an alternative.

6.2 Recommendations

In this section, we suggest two major areas of research. These areas include diagnosis of collinearity influential points in a *QR* setting and the extension of our work to include group penalties and high dimensional data sets.

6.2.1 Other Penalty Based Regularization Quantile Regression Procedures

This thesis considered penalized *QR* variable selection procedures with *RIDGE*, *LASSO*, *E-NET*, *ALASSO* and *AE-NET* penalties. New research can focus on other penalties, such as *SCAD*, *LARS*, *AR-LASSO* and group *LASSO* penalties, *etc*. The group *LASSO* penalty is given by $\lambda \Sigma_{\ell}^{m} \sqrt{p_\ell} |\beta^{(\ell)}|$, where $\beta^{(\ell)}$ is the coefficient vector of group ℓ, p_ℓ is the number of predictors in group ℓ from m different groups (Simon & Tibshirani 2012). The *LARS* and *SCAD* penalties are discussed in this thesis. The extension of these penalties to the *QR* procedures scenarios can be pursued in the future. We suggest these procedures to check if our methods can be further improved.

Our penalized *QR* procedures are most suitable when the main aim is variable selection in the presence of high leverage and collinearity influential points. However, the investigation of the applicability of our proposed penalized *WQR* procedures, specifically in the context of statistical inference, is beyond the scope of this research. We advise further research into how well our penalized procedures perform in various statistical inference settings, such as hypothesis testing and confidence interval estimates. Understanding how covariates are interpreted and used in these settings will be useful for applied researchers.

Further research could focus on developing and evaluating post-selection inference techniques specifically tailored for quantile regression, e.g., post-*LASSO* penalized *WQR* procedures and their post-selection inference. The oracle property of these post-selection procedures can also be investigated in the future, along with their strengths, providing valuable insights into the performance and limitations of the procedures and refining their applicability in practice. A comparative study can be conducted involving our methods and post-LASSO penalized WQR procedures by checking the accuracy, bias reduction, computational efficiency, and interpretability of the procedures.

6.2.2 Collinearity Influential Points Detection in a Quantile Regression Setting

For future research, we suggest collinearity influential point diagnostics in a *QR* setup using the diagnostic robust generalized potential *DRGP*, the *ER* approaches and residuals. One may use the single point deletion or group deletion criteria in a *QR* scenario in the formulation of these diagnostics procedures. The group deletion approach of diagnosing high leverage points (Imon 2002) partitions the design matrix into good and bad cases identified by the minimum volume ellipsoid (MVE) (Rousseeuw 1985) or the minimum covariance determinant (MCD) (Rousseeuw 1984). The group deletion criterion also known as generalized potential (*GP*) is an extension of the single deletion approach (see also Bagheri et al. 2012, Nurunnabi et al. 2016). One can also formulate diagnostic tools using the Hadi & Simonoff (1993) idea of an *ES*-based multiple outlier diagnostic procedures, or the Habshah et al. (2009) idea of the diagnostic robust generalized potential $(DRGP)$ for identifying multiple collinearity influential observations in a *QR* setting.

6.2.3 High Dimensional Quantile Regression

This section summarizes some QR aspects in high-dimensional that can be considered for further study. *QR* with p fixed asymptotics were extensively explored by Koenker & Bassett (1978). These *QR* models with p fixed asymptotics and traditional models pose problems in handling the increasing availability of data, and high-dimensional models proffer a solution through the improvement of accuracy and validity (Koenker et al. 2018). According to Koenker et al. (2018), two asymptotic regimes exist in high-dimensional models namely, p increases slowly relative to sample size n, and that p increases fast with n. Two asymptotic regimes exist in high-dimensional models namely, p increases slowly relative to sample size n, and that p increases fast with n. The QR estimator under necessary conditions achieves a uniform ℓ_2-rate of convergence of $\sqrt{p/n}$, and the Bahadur (1966) embedded representation aids the construction of uniform convergence bands.

Consider the outcome of interest Y and the conditioning variables $\boldsymbol{Z}$, then the conditional quantile function is given by $Q_{Y|Z}(\tau)$ for $\tau \in \mathscr{T} \subset (0,1)$ Koenker et al. (2018). To estimate the conditional quantile function, we let $\boldsymbol{X} = \boldsymbol{G}(\boldsymbol{Z})$, where $\boldsymbol{G}(\boldsymbol{Z})$ denotes a p-vector of known transformations of $\boldsymbol{Z}$, and the resultant $\boldsymbol{\beta}(\tau)$ vector is then given by

$$Q_{Y|Z}(\tau) = \boldsymbol{X}'\boldsymbol{\beta}(\tau) + r_\tau(\boldsymbol{Z}),$$

where $r_\tau(\boldsymbol{Z}) = Q_{Y|Z}(\tau) - \boldsymbol{X}'\boldsymbol{\beta}(\tau)$ denotes the approximate error function, and $p \gg n$. We have $\boldsymbol{Z} = \boldsymbol{X}$ and $r_\tau(\boldsymbol{Z}) = 0$ in the parametric case, and the non-parametric case coefficient process $\{\boldsymbol{\beta}(\tau) : \tau \in \mathscr{A}\}$ is taken to give a good approximate of the conditional quantile function. Regularization methods are used to explore high-dimensional data patterns in pursuit of high-dimensional models. In the literature, small sub-models $T_\tau \subset \{1,2,\dots,p\}$, with parameters $(|T_\tau| \geq s)$, yield good approximations. These approximate sparse models with $s \ll n$ non-zero coefficients enough to achieve a vanishing approximation error for each τ, $sup_{\tau\in\mathscr{T}}||\beta_\tau||_0 \leq s$ and $sup_{\tau\in\mathscr{T}}E_n[r_\tau(\boldsymbol{Z})^2] \leq Cs\frac{log(p\vee n)}{n}$, where $||.||_0$ is the number of non-zero components of a vector, and C denotes a fixed constant.

6.2.4 Inference

We recommend future research in statistical inference. In implementing our methods, we recommend reading Cahyani et al. (2016). It is important to consider data sets which have high leverage points in the implementation.

Bibliography

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE transactions on automatic control* **19**(6), 716–723.

Alkhamisi, M., Khalaf, G. & Shukur, G. (2006), 'Some modifications for choosing ridge parameters', *Communications in Statistics-Theory and Methods* **35**(11), 2005–2020.

Arslan, O. (2012), 'Weighted lad-lasso method for robust parameter estimation and variable selection in regression', *Computational Statistics & Data Analysis* **56**(6), 1952–1965.

Bagheri, A., Habshah, M. & Imon, R. H. M. R. (2012), 'A novel collinearity-influential observation diagnostic measure based on a group deletion approach', *Communications in Statistics - Simulation and Computation* **41**(8), 1379–1396.

Bagheri, A. & Midi, H. (2012), 'On the performance of the measure for diagnosing multiple high leverage collinearity-reducing observations', *Mathematical Problems in Engineering* **2012**.

Bahadur, R. R. (1966), 'A note on quantiles in large samples', *The Annals of Mathematical Statistics* **37**(3), 577–580.

Belloni, A. & Chernozhukov, V. (2011), 'ℓ 1 -penalized quantile regression in high-dimensional sparse models', *Ann. Statist.* **39**(1), 82–130.

Belloni, A. & Chernozhukov, V. (2013), 'Least squares after model selection in high-dimensional sparse models', *Bernoulli* **19**(2), 521–547.

Bickel, P. J., Doksum, K. & Hodges, J. L. (1982), *A Festschrift for Erich L. Lehmann*, CRC Press.

Breiman, L. (1995), 'Better subset regression using the nonnegative garrote', *Technometrics* **37**(4), 373–384.

Cahyani, T. B. N., Wigena, A. H. & Djuraidah, A. (2016), 'Quantile regression with elastic-net in statistical downscaling to predict extreme rainfall', *Global Journal of Pure and Applied Mathematics* **12**(4), 3517–3524.

Chatterjee, S. & Hadi, A. (1988*a*), *Sensitivity Analysis in Linear Regression*, Wiley Series in Probability and Statistics, Wiley.

Chatterjee, S. & Hadi, A. S. (1988*b*), 'Impact of simultaneous omission of a variable and an observation on a linear regression equation', *Computational Statistics & Data Analysis* **6**(2), 129–144.

Conniffe, D. & Stone, J. (1973), 'A critical view of ridge regression', *Journal of the Royal Statistical Society. Series D (The Statistician)* **22**(3), 181–187.

Dai, L., Chen, K. & Li, G. (2020), 'The broken adaptive ridge procedure and its applications', *Statistica Sinica* **30**(2), 1069–1094.

Dai, L., Chen, K., Sun, Z., Liu, Z. & Li, G. (2018), 'Broken adaptive ridge regression and its asymptotic properties', *Journal of multivariate analysis* **168**, 334–351.

Dicker, L., Huang, B. & Lin, X. (2013), 'Variable selection and estimation with the seamless-l 0 penalty', *Statistica Sinica* pp. 929–962.

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004), 'Least angle regression', *Ann. Statist.* **32**(2), 407–499.

Efroymson, M. A. (1960), 'Multiple regression analysis', *Mathematical Methods for Digital Computers* pp. 191–203.

Fan, J., Fan, Y. & Barut, E. (2014), 'Adaptive robust variable selection', *Ann. Statist.* **42**(1), 324–351.

Fan, J. & Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.

Fan, J. & Peng, H. (2004), 'Nonconcave penalized likelihood with a diverging number of parameters', *Ann. Statist.* **32**(3), 928–961.

Farrar, D. E. & Glauber, R. R. (1967), 'Multicollinearity in regression analysis: the problem revisited', *The Review of Economic and Statistics* pp. 92–107.

Fox, J. & Monette, G. (1992), 'Generalized collinearity diagnostics', *Journal of the American Statistical Association* **87**(417), 178–183.

Frommlet, F. & Nuel, G. (2016), 'An adaptive ridge procedure for l 0 regularization', *PloS one* **11**(2), e0148620.

Fu, W. J. (1998), 'Penalized regressions: the bridge versus the lasso', *Journal of computational and graphical statistics* **7**(3), 397–416.

Furnival, G. & Wilson, R. (1974), 'Regression by leaps and bounds-a program for finding the best subset of regressions [computer program]', *New Haven, CT: Yale University and US Forest Service* .

Galvao, A. F., Gu, J. & Volgushev, S. (2020), 'On the unbiased asymptotic normality of quantile regression with fixed effects', *Journal of Econometrics* **218**(1), 178–215.

Ghorbani, H. (2019), 'Mahalanobis distance and its application for detecting multivariate outliers', *Facta Univ Ser Math Inform* **34**, 583–95.

Giloni, A., Simonoff, J. S. & Sengupta, B. (2006), 'Robust weighted lad regression', *Computational Statistics & Data Analysis* **50**(11), 3124–3140.

Gunst, R. & Mason, R. (1980), 'Regression analysis and its applications. new york: Marce dekker'.

Habshah, M., Norazan, M. R. & Imon, A. H. R. (2009), 'The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression', *Journal of Applied Statistics* **36**(5), 507–520.

Hadi, A. S. & Simonoff, J. S. (1993), 'Procedures for the identification of multiple outliers in linear models', *Journal of the American statistical association* **88**(424), 1264–1272.

Hawkins, D. M. (1993), 'Regression adjustment for variables in multivariate quality control', *Journal of Quality Technology* **25**(3), 170–182.

Hawkins, D. M., Bradu, D. & Kass, G. V. (1984), 'Location of several outliers in multiple-regression data using elemental sets', *Technometrics* **26**(3), 197–208.

Hawkins, D. M. & Olive, D. J. (1999), 'Improved feasible solution algorithms for high breakdown estimation', *Computational Statistics and Data Analysis* **30**(1), 1–11.

Hawkins, D. M. & Olive, D. J. (2002), 'Inconsistency of resampling algorithms for high-breakdown regression estimators and a new algorithm', *Journal of the American Statistical Association* **97**, 136–159.

Heinze, G., Wallisch, C. & Dunkler, D. (2018), 'Variable selection–a review and recommendations for the practicing statistician', *Biometrical journal* **60**(3), 431–449.

Hesterberg, T., Choi, N. H., Meier, L. & Fraley, C. (2008), 'Least angle and ℓ_1 penalized regression. a review', *Statistics Surveys* **2**, 61–93.

Hocking, R. & Pendleton, O. (1983), 'The regression dilemma', *Communications in Statistics - Theory and Methods* **12**(5), 497–527.

Hocking, R. R., Speed, F. & Lynn, M. (1976), 'A class of biased estimators in linear regression', *Technometrics* **18**(4), 425–437.

Hoerl, A. E., Kannard, R. W. & Baldwin, K. F. (1975), 'Ridge regression:some simulations', *Communications in Statistics* **4**(2), 105–123.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Hosmer Jr, D. W., Lemeshow, S. & Sturdivant, R. X. (2013), *Applied logistic regression*, Vol. 398, John Wiley & Sons.

Hubert, M. & Rousseeuw, P. J. (1997), 'Robust regression with both continuous and binary regressors', *Journal of Statistical Planning and Inference* **57**(1), 153 – 163. Robust Statistics and Data Analysis, Part I.

Imon, A. R. (2002), ‘On deletion residuals’, *Calcutta Statistical Association Bulletin* **53**(1-2), 65–80.

James, W. & Stein, C. (1961), Estimation with quadratic loss, *in* ‘Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics’, University of California Press, Berkeley, Calif., pp. 361–379.

Jiang, X., Jiang, J. & Song, X. (2012), ‘Oracle model selection for nonlinear models based on weighted composite quantile regression’, *Statistica Sinica* **22**(4), 1479–1506.

Jiang, Y., Wang, Y., Zhang, J., Xie, B., Liao, J. & Liao, W. (2021), ‘Outlier detection and robust variable selection via the penalized weighted lad-lasso method’, *Journal of Applied Statistics* **48**(2), 234–246.

Jureckova, J. & Procházka, B. (1994), ‘Regression quantiles and trimmed least squares estimator in nonlinear regression model’, *Journal of Nonparametric Statistics* **3**.

Karunamuni, R. J., Kong, L. & Tu, W. (2019), ‘Efficient robust doubly adaptive regularized regression with applications’, *Statistical methods in medical research* **28**(7), 2210–2226.

Kato, K., Galvao Jr, A. F. & Montes-Rojas, G. V. (2012), ‘Asymptotics for panel quantile regression models with individual effects’, *Journal of Econometrics* **170**(1), 76–91.

Khalaf, G. & Shukur, G. (2005), ‘Choosing ridge parameter for regression problems’, *Communications in Statistics - Theory and Methods* **34**(5), 1177–1182.

Kibria, B. G. (2003), ‘Performance of some new ridge regression estimators’, *Communications in Statistics-Simulation and Computation* **32**(2), 419–435.

Knight, K. (1999), ‘Asymptotics for l1-estimators of regression parameters under heteroscedasticityy’, *Canadian Journal of Statistics* **27**(3), 497–507.

Koenker, R. (2005), *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

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

Koenker, R., Chernozhukov, V., He, X. & Peng, L. (2018), *Handbook of Quantile Regression*, CHAPMAN & HALL/CRC HANDBOOKS O, CRC Press, Taylor & Francis Group.

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

Koenker, R. W. & D’Orey, V. (1987), ‘Algorithm as 229: Computing regression quantiles’, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **36**(3), 383–393.

Lawless, J. F. & Wang, P. (1976), ‘A simulation study of ridge and other regression estimators’, *Communications in Statistics - Theory and Methods* **5**(4), 307–323.

Li, Y. & Zhu, J. (2008), 'L1-norm quantile regression', *Journal of Computational and Graphical Statistics* **17**(1), 163–185.

Lin, W. & Lv, J. (2013), 'High-dimensional sparse additive hazards regression', *Journal of the American Statistical Association* **108**(501), 247–264.

Lin, Y. & Zhang, H. H. (2006), 'Component selection and smoothing in multivariate nonparametric regression', *Ann. Statist.* **34**(5), 2272–2297.

Mahalanobis, P. C. (1936), On the generalized distance in statistics, National Institute of Science of India.

Mantel, N. (1970), 'Why stepdown procedures in variable selection', *Technometrics* **12**(3), 621–625.

Mason, R. L. & Gunst, R. F. (1985), 'Outlier-induced collinearities', *Technometrics* **27**(4), 401–407.

McDonald, G. C. (1980), 'Some algebraic properties of ridge coefficients', *Journal of the Royal Statistical Society. Series B (Methodological)* **42**(1), 31–34.

Meier, L. & Bühlmann, P. (2008), 'Smoothing l1-penalized estimators for high-dimensional time-course data', *Electronic Journal of Statistics* **1**.

Midi, H. & Bagheri, A. (2013), 'Robust multicollinearity diagnostic measures based on minimum covariance determination approach', *Economic Computation & Economic Cybernetics Studies & Research* **47**(4), 1–15.

Montgomery, D. C., Runger, G. C. & Hubele, N. F. (2009), *Engineering statistics*, John Wiley & Sons.

Montgomery, D., Peck, E. & Vining, G. (2012), *Introduction to Linear Regression Analysis*, Wiley Series in Probability and Statistics, Wiley.

Mudhombo, I. & Ranganai, E. (2022), 'Robust variable selection and regularization in quantile regression based on adaptive-lasso and adaptive e-net', *Computation* **10**(11), 203.

Muniz, G. & Kibria, B. G. (2009), 'On some ridge regression estimators: An empirical comparisons', *Communications in Statistics—Simulation and Computation®* **38**(3), 621–630.

Norouzirad, M., Hossain, S. & Arashi, M. (2018), 'Shrinkage and penalized estimators in weighted least absolute deviations regression models', *Journal of Statistical Computation and Simulation* **88**(8), 1557–1575.

Nurunnabi, A., Nasser, M. & Imon, A. (2016), 'Identification and classification of multiple outliers, high leverage points and influential observations in linear regression', *Journal of Applied Statistics* **43**(3), 509–525.

O'brien, R. M. (2007), 'A caution regarding rules of thumb for variance inflation factors', *Quality & Quantity* **41**(5), 673–690.

Ogundimu, E. O. (2022), 'On lasso and adaptive lasso for non-random sample in credit scoring', *Statistical Modelling* p. 1471082X221092181.

Park, M. & Hastie, T. (2006), 'Regularization path algorithms for detecting gene interactions', *IEEE Transactions on Reliability - TR* .

Pollard, D. (1991), 'Asymptotics for least absolute deviation regression estimators', *Econometric Theory* **7**(2), 186–199.

Ranganai, E. (2007), Aspects of model development using regression quantiles and elemental regressions, PhD thesis, Stellenbosch: Stellenbosch University.

Ranganai, E. (2016*a*), 'On studentized residuals in the quantile regression framework', *SpringerPlus* **5**(1), 1231.

Ranganai, E. (2016*b*), 'Quality of fit measurement in regression quantiles: An elemental set method approach', *Statistics and Probability Letters* **111**(C), 18–25.

Ranganai, E. & Mudhombo, I. (2021), 'Variable selection and regularization in quantile regression via minimum covariance determinant based weights', *Entropy* **23**(1), 33.

Ranganai, E., Vuuren, J. O. V. & Wet, T. D. (2014), 'Multiple case high leverage diagnosis in regression quantiles', *Communications in Statistics - Theory and Methods* **43**(16), 3343–3370.

Rousseeuw, P. (1985), 'Multivariate estimation with high breakdown point', *Mathematical Statistics and Applications* p. 283–297.

Rousseeuw, P. J. (1984), 'Least median of squares regression', *Journal of the American Statistical Association* **79**(388), 871–880.

Rousseeuw, P. J. & Croux, C. (1993), 'Alternatives to the median absolute deviation', *Journal of the American Statistical association* **88**(424), 1273–1283.

Rousseeuw, P. J. & Driessen, K. V. (1999), 'A fast algorithm for the minimum covariance determinant estimator', *Technometrics* **41**(3), 212–223.

Rousseeuw, P. J. & Hubert, M. (2018), 'Anomaly detection by robust statistics', *WIREs Data Mining and Knowledge Discovery* **8**(2), e1236.

Rousseeuw, P. J. & Leroy, A. (1987), *Robust Regression and Outlier Detection*, Wiley Series in Probability and Statistics, Wiley.

Rousseeuw, P. J. & van Zomeren, B. C. (1990), 'Unmasking multivariate outliers and leverage points', *Journal of the American Statistical Association* **85**(411), 633–639.

Salibián-Barrera, M. & Wei, Y. (2008), 'Weighted quantile regression with nonelliptically structured covariates', *Canadian Journal of Statistics* **36**(4), 595–611.

Schaefer, S., McPhail, T. & Warren, J. (2006), 'Image deformation using moving least squares', *ACM Trans. Graph.* **25**(3), 533–540.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The annals of statistics* pp. 461–464.

Sengupta, D. & Bhimasankaram, P. (1997), 'On the roles of observations in collinearity in the linear model', *Journal of the American Statistical Association* **92**(439), 1024–1032.

Shao, J. (1993), 'Linear model selection by cross-validation', *Journal of the American Statistical Association* **88**(422), 486–494.

Siegel, A. F. (1982), 'Robust regression using repeated medians', *Biometrika* **69**(1), 242–244.

Silvey, S. (1969), 'Multicollinearity and imprecise estimation', *Journal of the Royal Statistical Society: Series B (Methodological)* **31**(3), 539–552.

Simon, N. & Tibshirani, R. (2012), 'Standardization and the group lasso penalty', *Statistica Sinica* **22**(3), 983.

Smyth, G. K. & Hawkins, D. M. (2000), 'Robust frequency estimation using elemental sets', *Journal of Computational and Graphical Statistics* **9**(1), 196–214.

Strang, G. (2006), *Linear algebra and its applications*, Thomson, Brooks/Cole, Belmont, CA.

Su, M. & Wang, W. (2021), 'Elastic net penalized quantile regression model', *Journal of Computational and Applied Mathematics* **392**, 113462.

Suhail, M., Chand, S. & Kibria, B. G. (2019), 'Quantile-based robust ridge m-estimator for linear regression model in presence of multicollinearity and outliers', *Communications in Statistics-Simulation and Computation* pp. 1–13.

Suhail, M., Chand, S. & Kibria, B. G. (2020), 'Quantile based estimation of biasing parameters in ridge regression model', *Communications in Statistics-Simulation and Computation* **49**(10), 2732–2744.

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

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. (2005), 'Sparsity and smoothness via the fused lasso', *Journal of the Royal Statistical Society Series B* **67**(1), 91–108.

Tsamardinos, I., Greasidou, E. & Borboudakis, G. (2018), 'Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation', *Machine learning* **107**, 1895–1922.

Turlach, B., Venables, W. & Wright, S. (2005), 'Simultaneous variable selection', *Technometrics* **47**(3), 349–363.

Wang, H., Li, G. & Jiang, G. (2007), 'Robust regression shrinkage and consistent variable selection through the lad-lasso', *Journal of Business & Economic Statistics* **25**(3), 347–355.

Wang, L., Wu, Y. & Li, R. (2012), 'Quantile regression for analyzing heterogeneity in ultra-high dimension', *Journal of the American Statistical Association* **107**(497), 214–222. PMID: 23082036.

Weisberg, S. (2014), *Applied Linear Regression*, Wiley Series in Probability and Statistics, Wiley.

Willmott, C. J. & Matsuura, K. (2005), 'Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance', *Climate research* **30**(1), 79–82.

Wu, T. T. & Lange, K. (2008), 'Coordinate descent algorithms for lasso penalized regression', *The Annals of Applied Statistics* **2**(1), 224–244.

Wu, Y. & Liu, Y. (2009), 'Variable selection in quantile regression', *Statistics Sinica* pp. 801–817.

Yi, C. (2017), 'hqreg: Regularization paths for lasso or elastic-net penalized huber loss regression and quantile regression, r package version 1.4'.

Yu, Y. & Feng, Y. (2014), 'Modified cross-validation for penalized high-dimensional linear regression models', *Journal of Computational and Graphical Statistics* **23**(4), 1009–1027.

Yuan, M. & Lin, Y. (2006), 'Model selection and estimation in regression with grouped variables', *Journal of the Royal Statistical Society Series B* **68**(1), 49–67.

Zhang, C.-H. et al. (2010), 'Nearly unbiased variable selection under minimax concave penalty', *The Annals of statistics* **38**(2), 894–942.

Zhao, P., Rocha, G. & Yu, B. (2009), 'The composite absolute penalties family for grouped and hierarchical variable selection', *Ann. Statist.* **37**(6A), 3468–3497.

Zhao, Q. (2000), 'Restricted regression quantiles', *Journal of Multivariate Analysis* **72**(1), 78–99.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**(476), 1418–1429.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society, Series B* **67**, 301–320.

Zou, H. & Li, R. (2008), 'One-step sparse estimates in nonconcave penalized likelihood models', *Ann. Statist.* **36**(4), 1509–1533.

Zou, H. & Yuan, M. (2008), 'Composite quantile regression and the oracle model selection theory', *Ann. Statist.* **36**(3), 1108–1126.

Zou, H. & Zhang, H. H. (2009), 'On the adaptive elastic-net with a diverging number of parameters', *Annals of statistics* **37**(4), 1733.

APPENDIX A:TABLES AND FIGURES

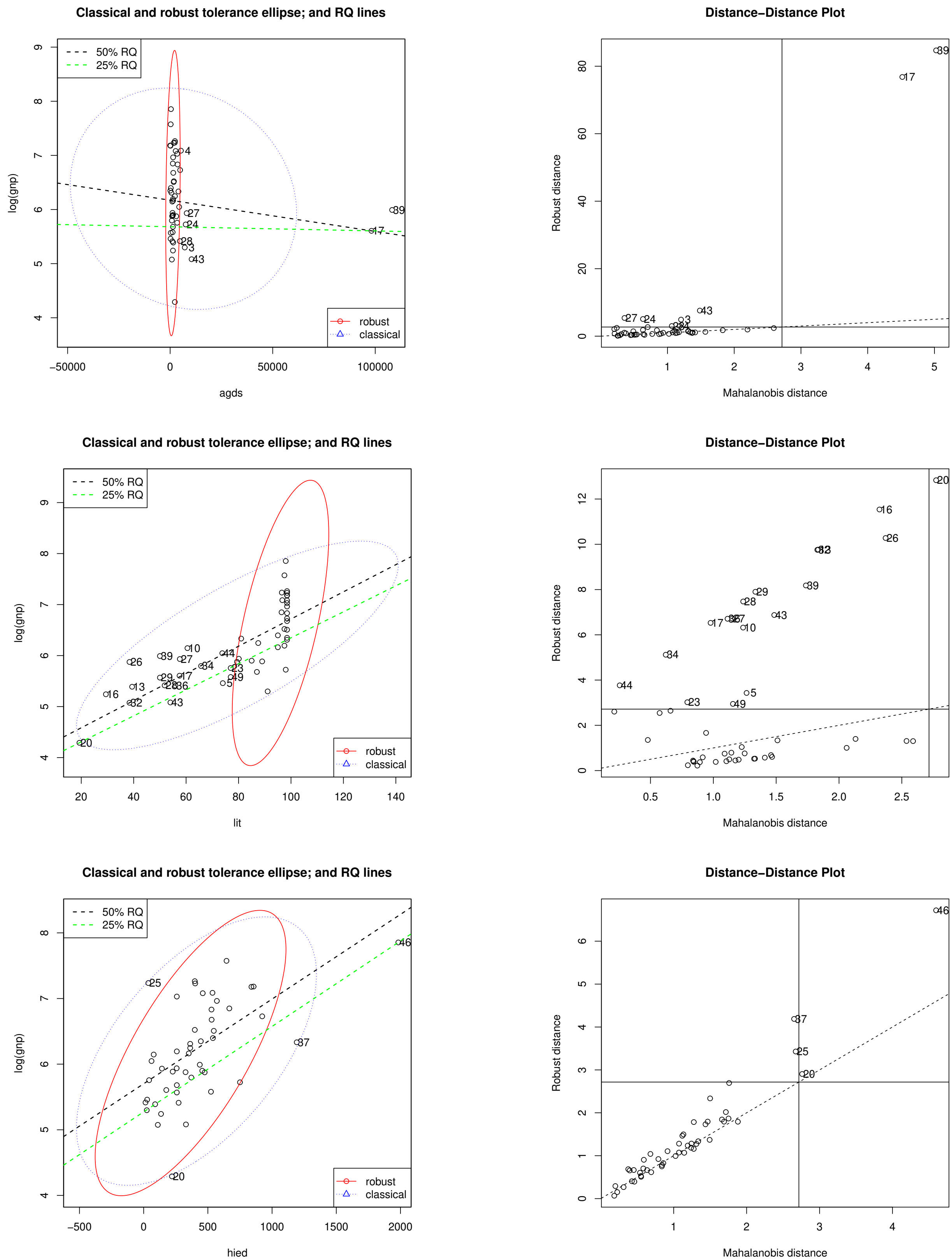


Figure 6.1: Tolerance Ellipses with RQ lines and Distance-Distance Plots, Gunst-Mason Data

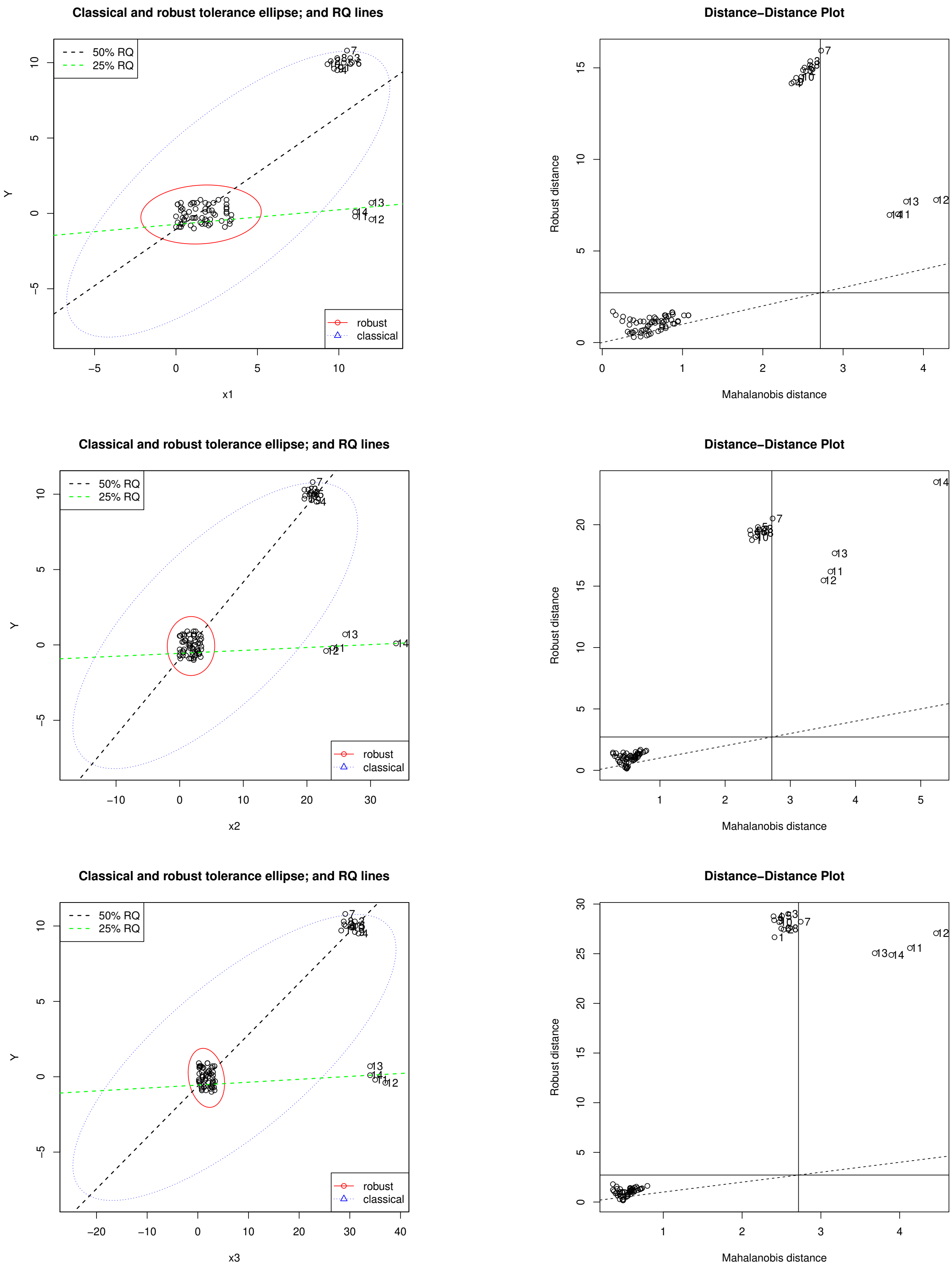


Figure 6.2: Tolerance Ellipses with RQ lines and Distance-Distance Plots, Hawkins Data

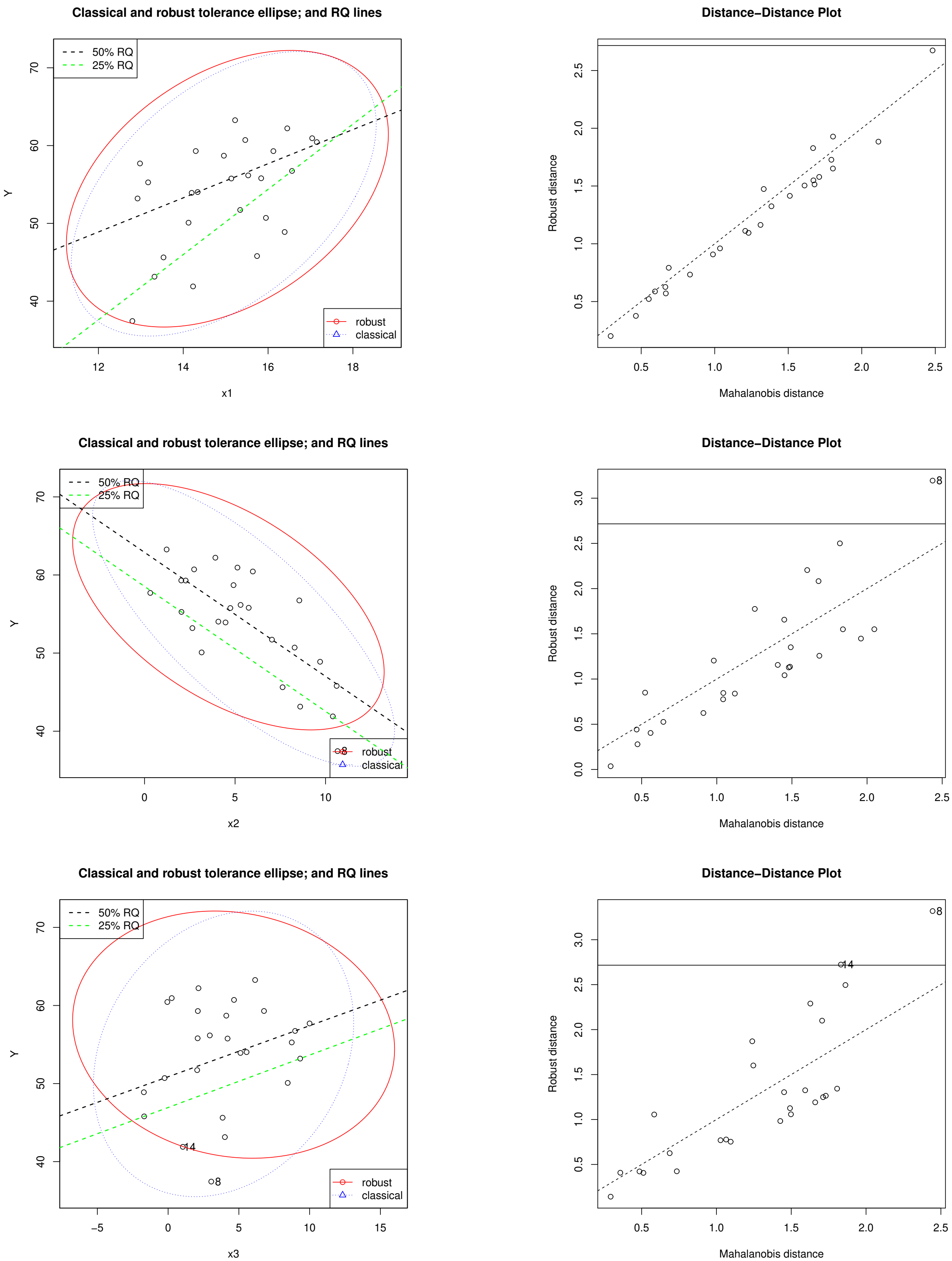


Figure 6.3: Tolerance Ellipses with RQ lines and Distance-Distance Plots, Hocking Data

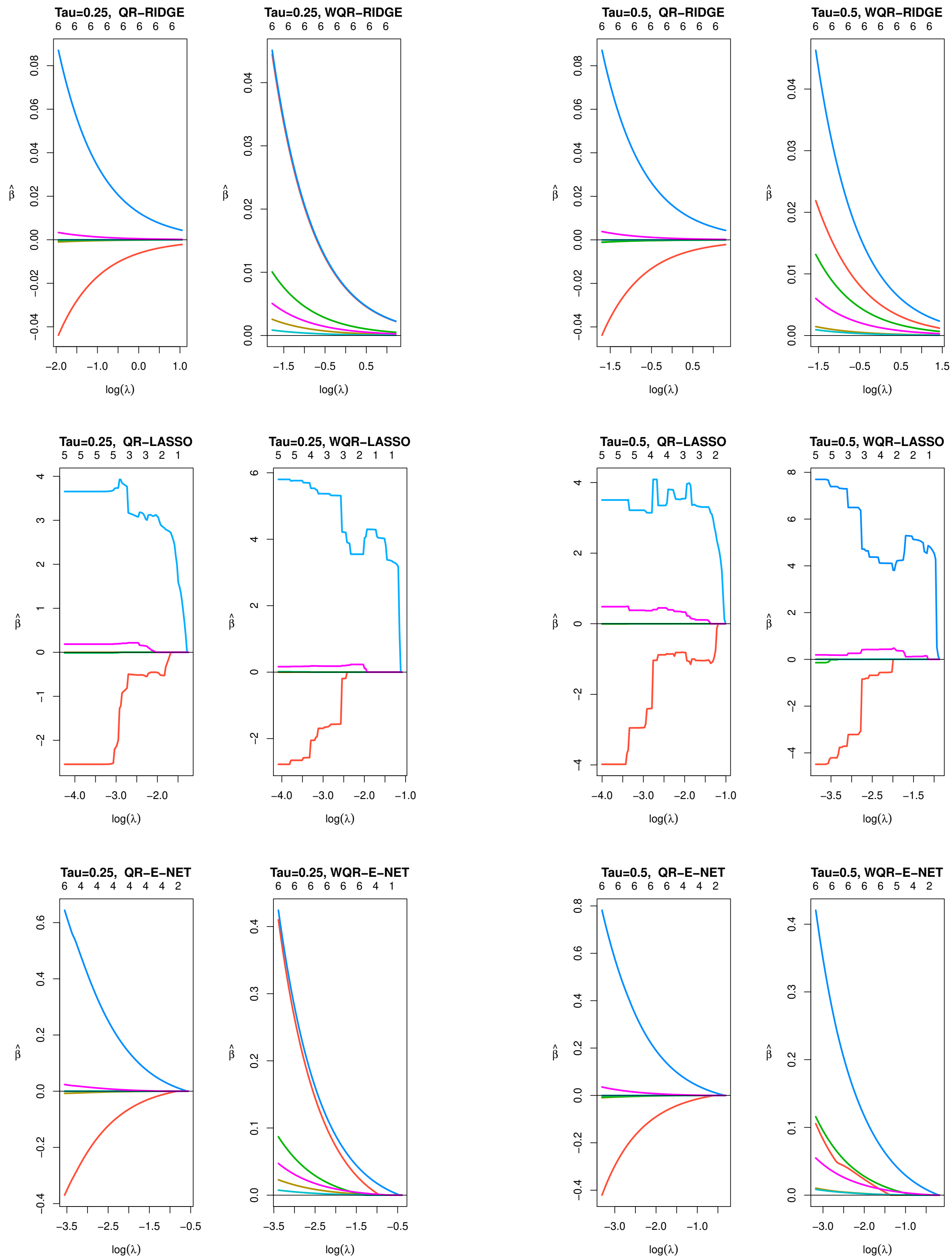


Figure 6.4: :Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for the Gunst-Mason data set.

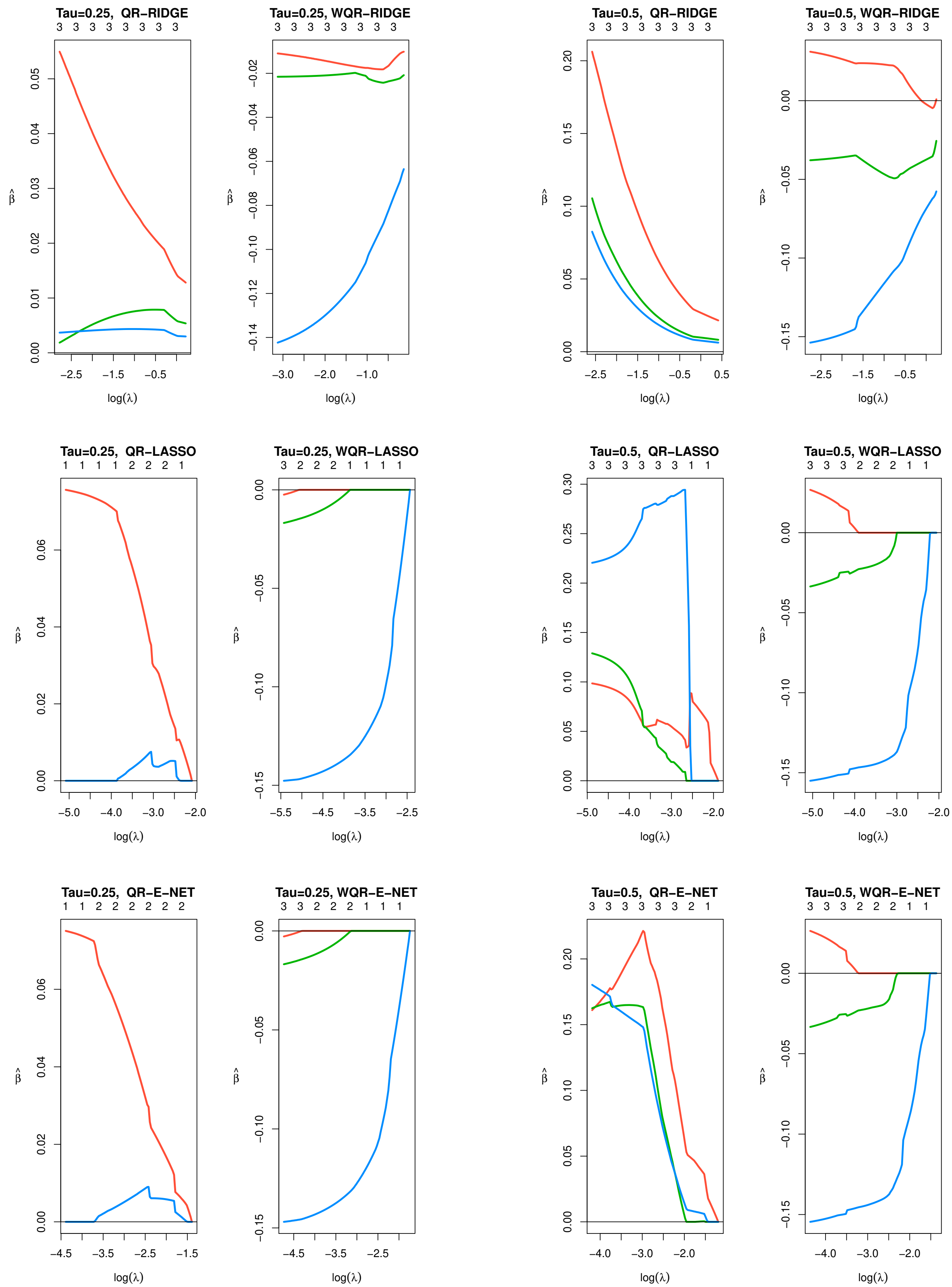


Figure 6.5: :Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for the Hawkins data set.

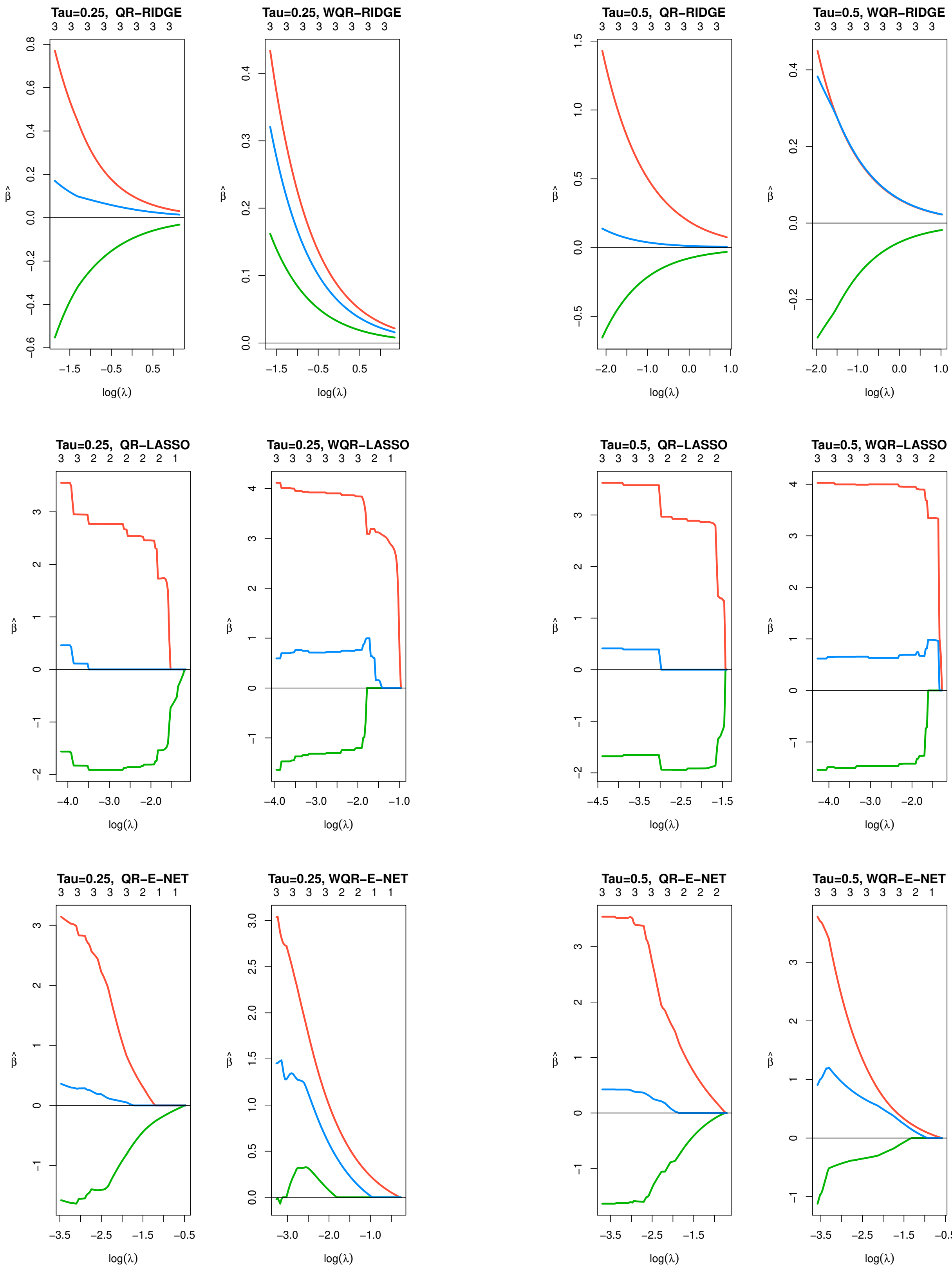


Figure 6.6: :Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for the Hocking dataset.

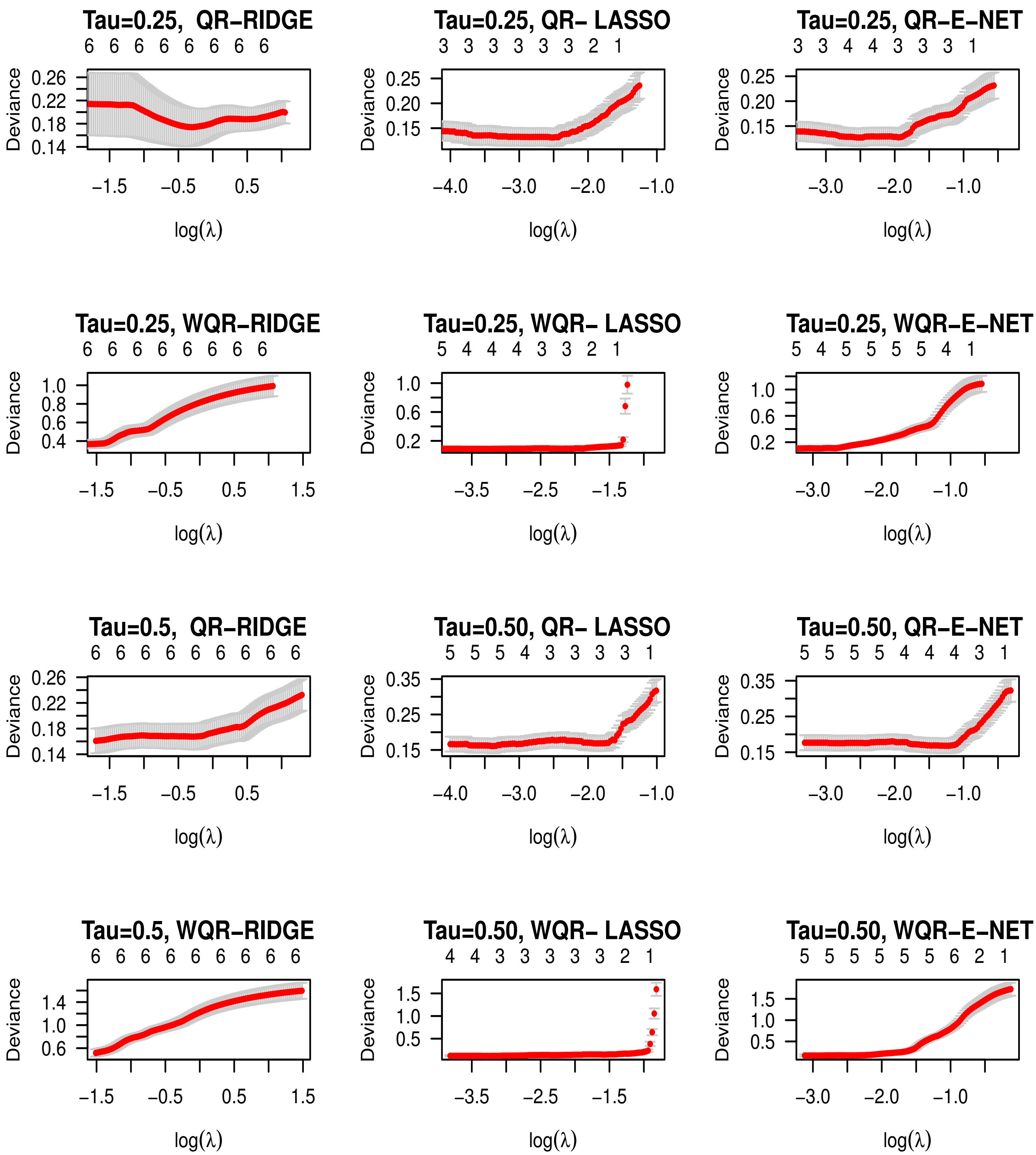


Figure 6.7: Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for Gunst and Mason dataset.

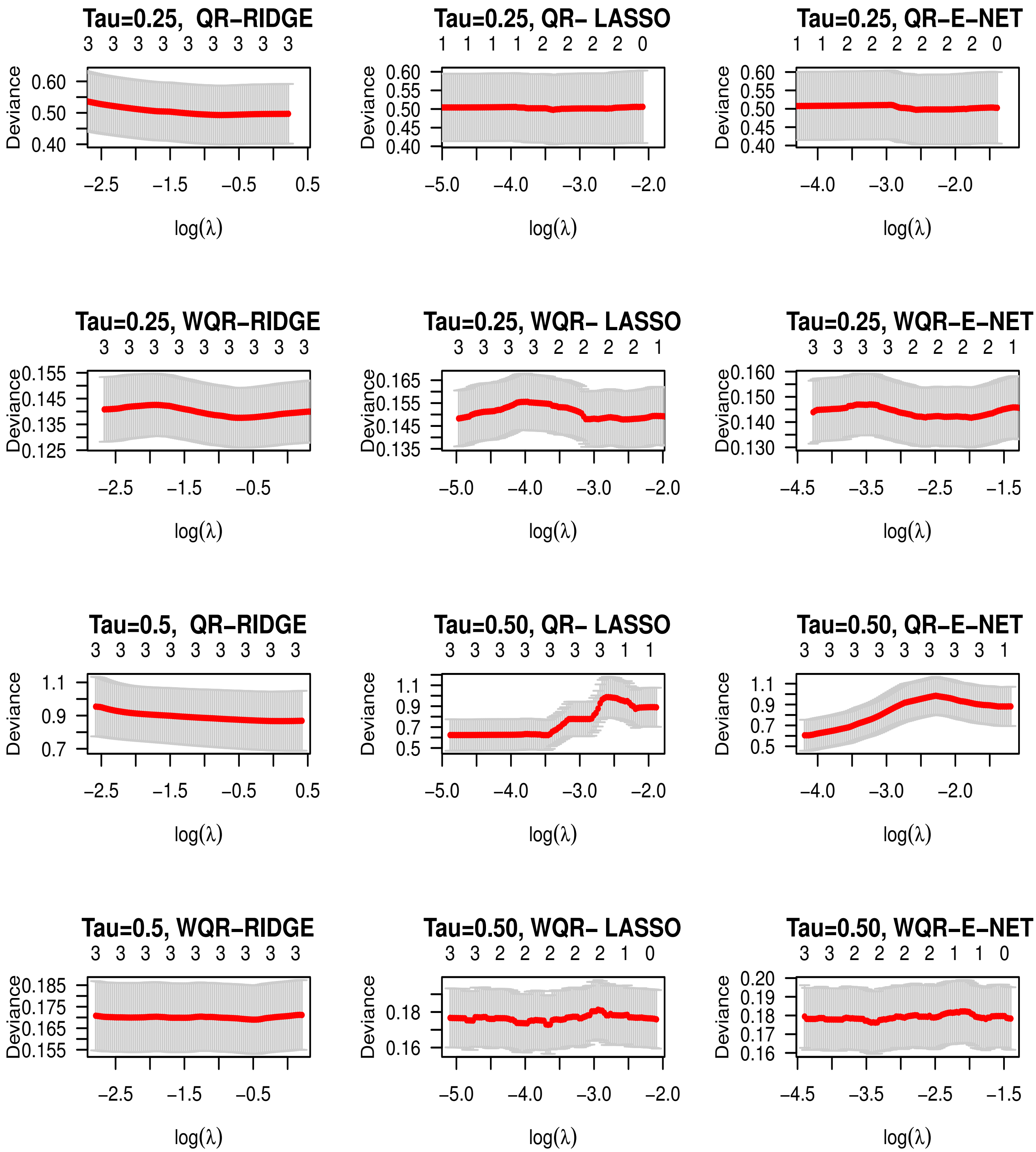


Figure 6.8: Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for Hawkins dataset.

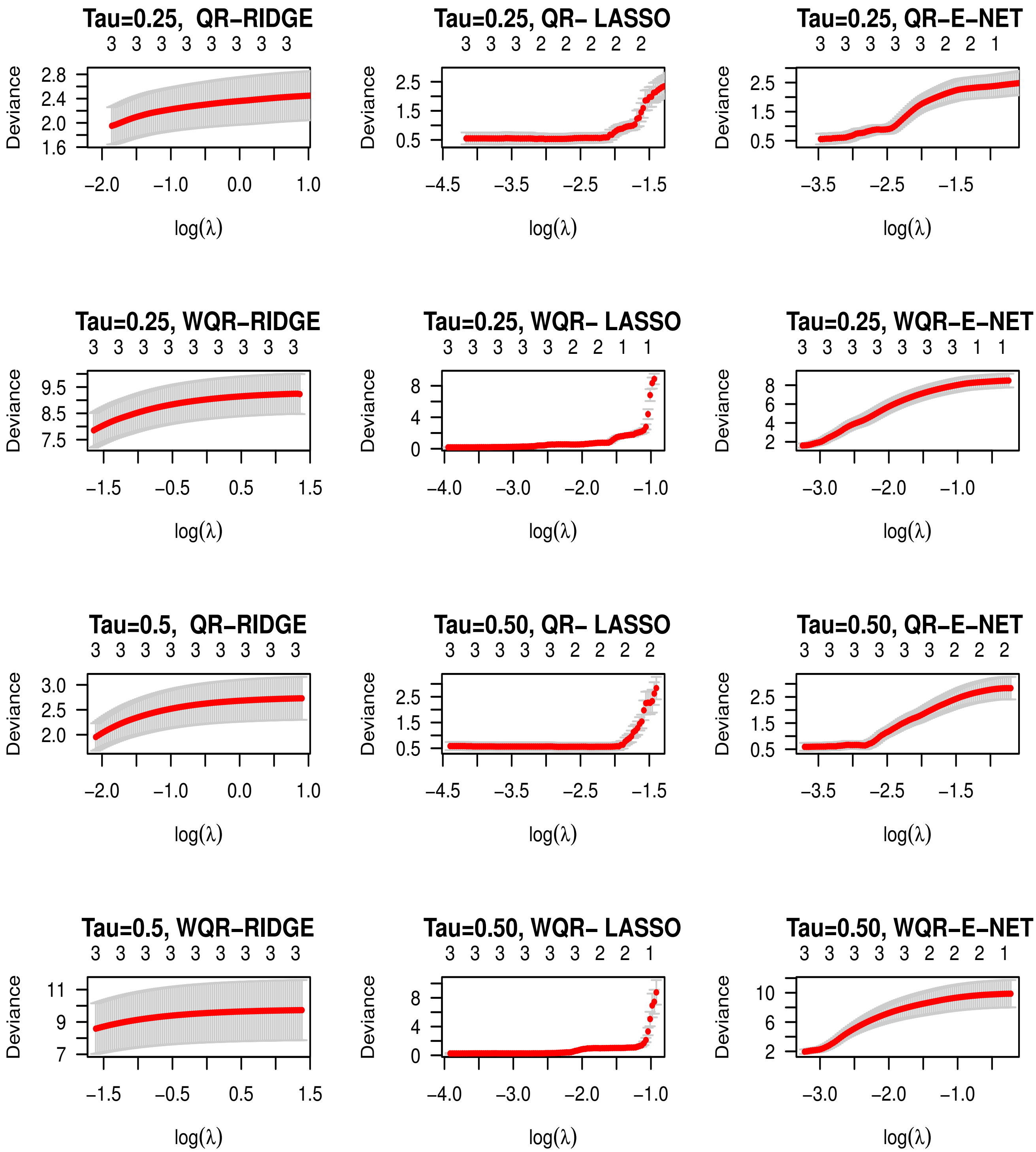


Figure 6.9: Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for Hocking dataset.

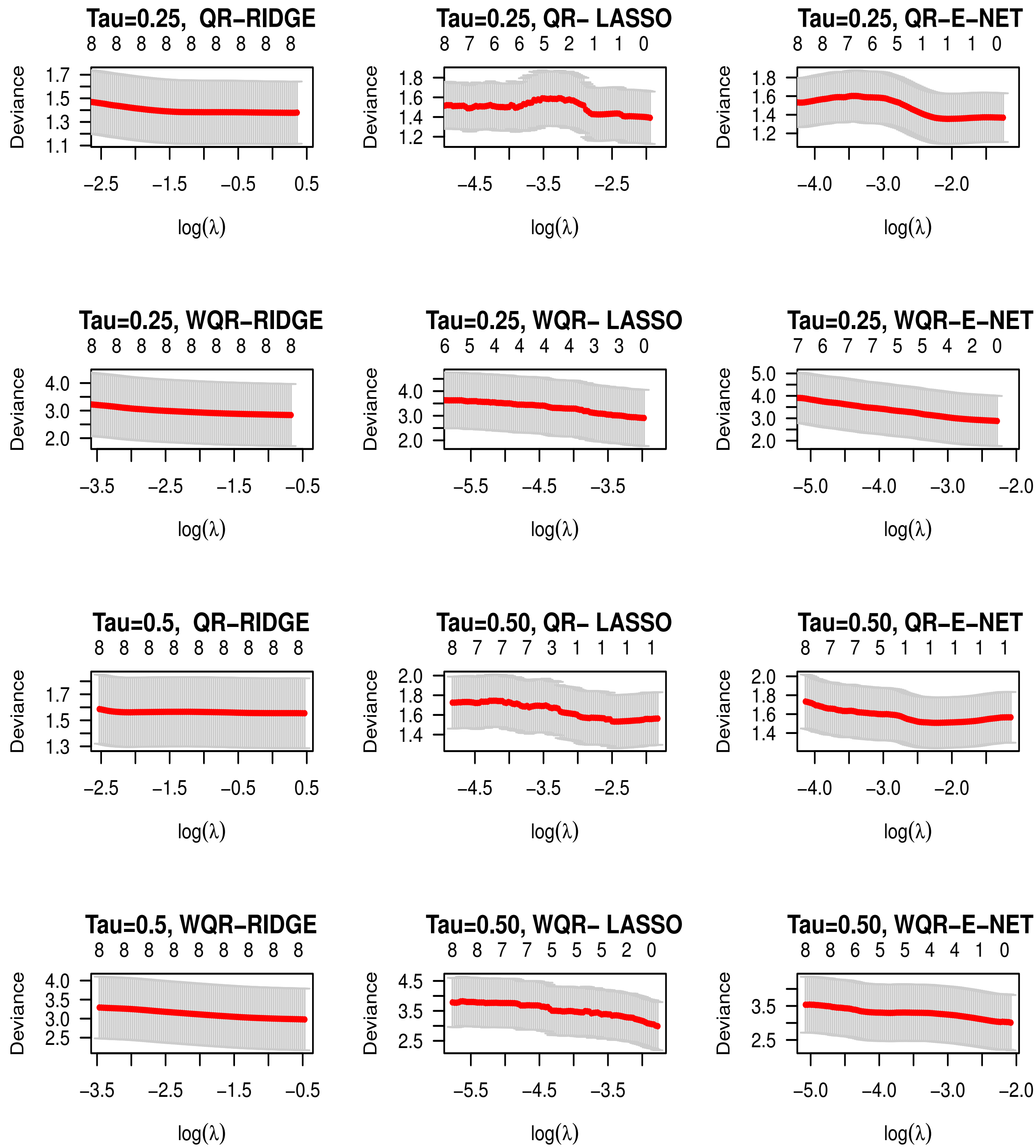


Figure 6.10: Profiles for *RIDGE*-type, *LASSO*-type and *E-NET*-type estimators for D6 dataset.

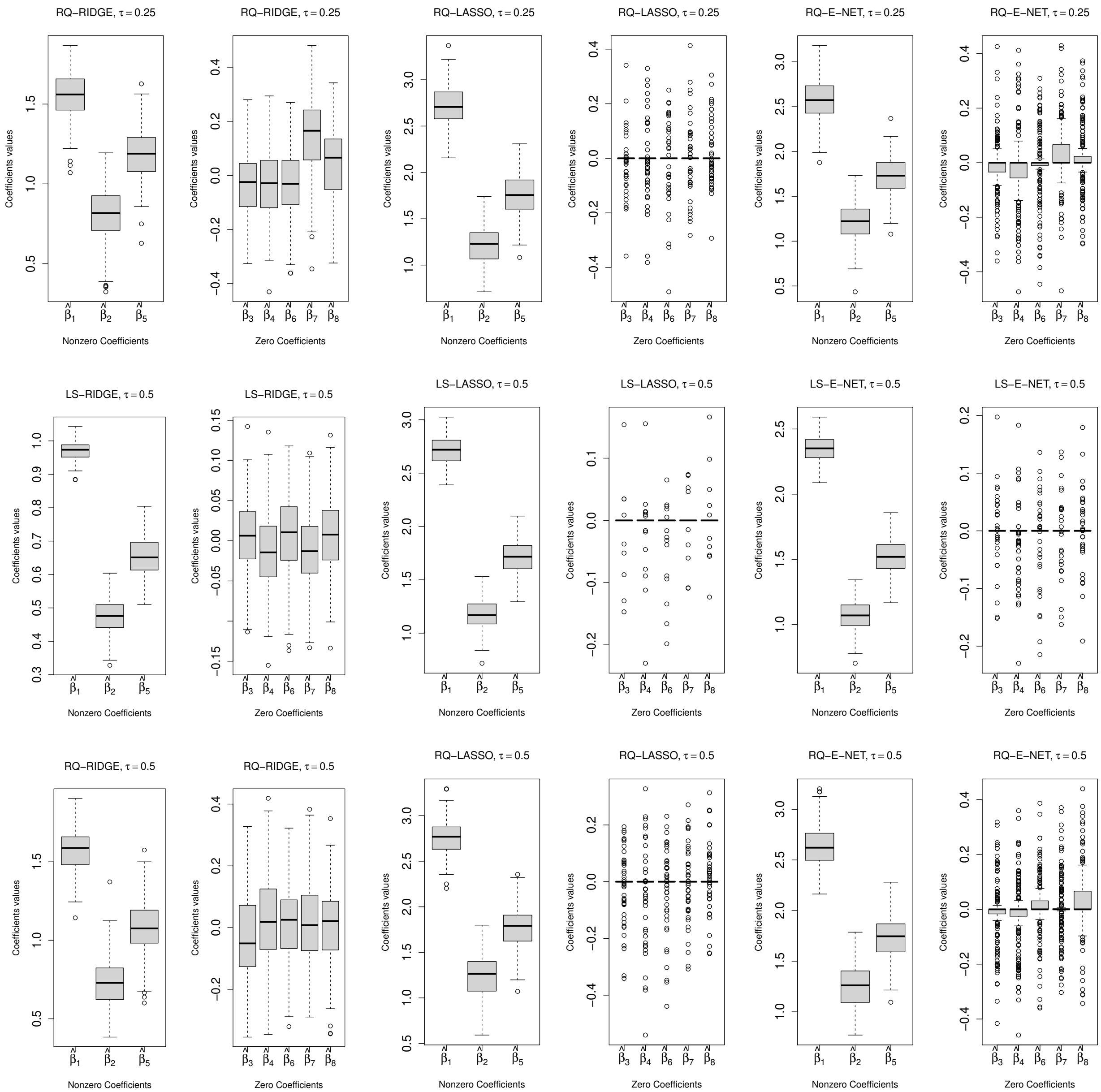


Figure 6.11: Box Plots for $D1$ under normal distribution with $\sigma = 1$ at $\tau = 0.25$ and $\tau = 0.5$ quantile level.

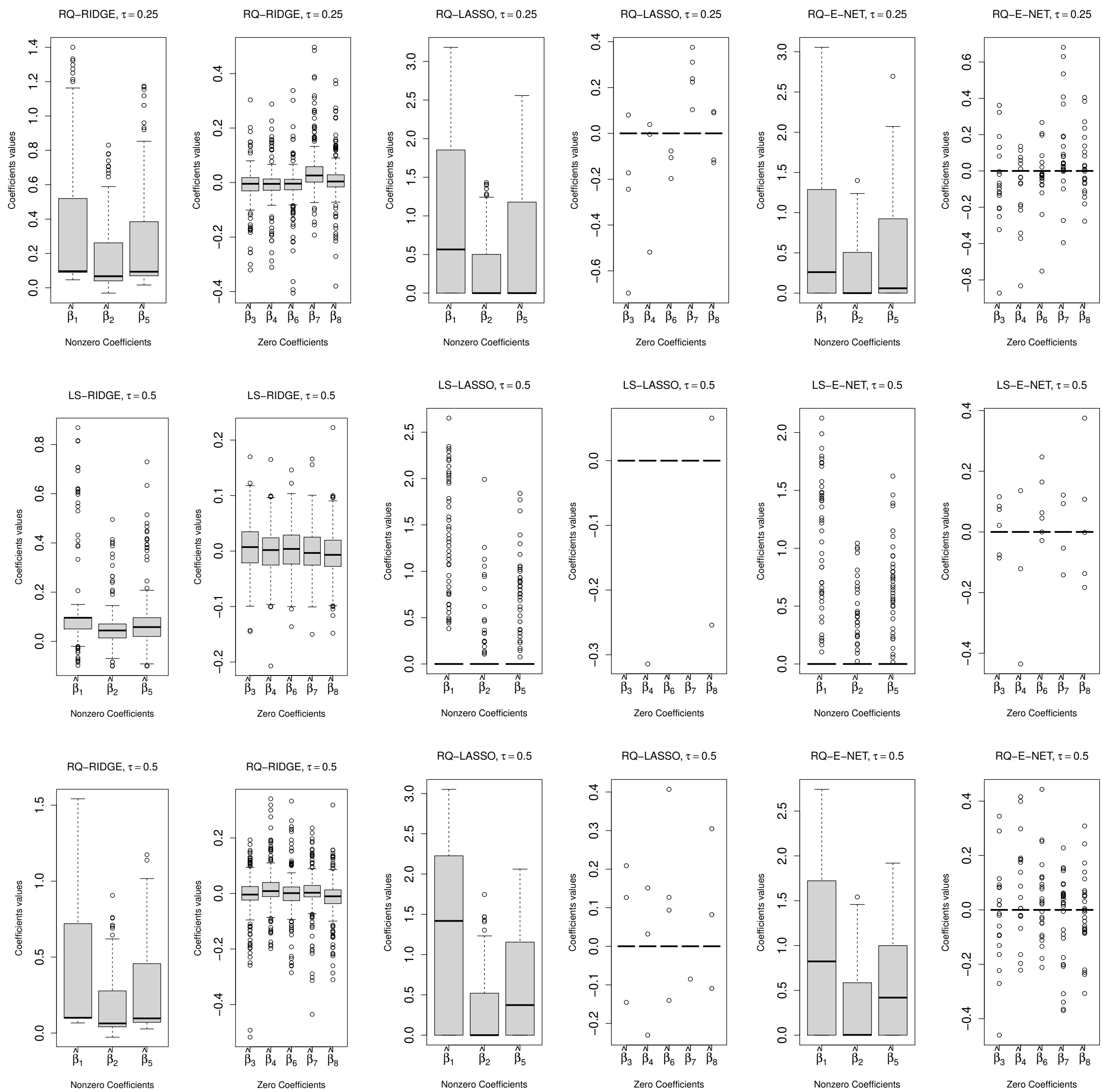


Figure 6.12: Box Plots for $D1$ under t-distribution with $\sigma = 1$, $d = 1$ at $\tau = 0.25$ and $\tau = 0.5$ quantile level.

APPENDIX B: ALGORITHMS

In case a reader wants to use bootstrapping in the simulations as suggested by one Examiner, we present an algorithm as an extra (see also Ogundimu 2022).

Algorithm : ***Regularized Quantile Regression***
Input :
Data set $\{\boldsymbol{x}_i^{*'}, \boldsymbol{y}^*\}$ such that $\boldsymbol{x}_i^{*'} = \omega_i \boldsymbol{x}_i'$, the input i^{th} row
of a $n \times p$ matrix $\boldsymbol{X}^*$, and p is the number of predictor variables and n is the sample size.
$\boldsymbol{y}^* = \omega_i \boldsymbol{y}$: # The weighted response vector of size n.
λ: # The regularization parameter.
ω_i: # The MCD based weight vector of size n.
$\varepsilon = 10^{-6}$ # Set convergence threshold.
Output: Penalized WQR Model
(i) Fit the penalized WQR models and MAD_{app} of test errors
(ii) Generate K data sets of the same sample size (n) using bootstrap samples with replacement.
(iii) Compute $MAD_{boot}(k)$. # For $k \in [1:K]$, apply the penalized WQR procedures, making sure optimal
λ is selected for each sample to provide models having predictors coefficients
based on the regularization parameter. # The bootstrap performance.
(iv) Compute $MAD_{orig}(k)$. # MAD from the original data set for each of the K models.
(v) Compute $Opt(k) = MAD_{boot}(k) - MAD_{orig}(k)$ # Optimal in model fit for each of the bootstrap sample.
(vi) Compute $Opt = \frac{1}{K}\Sigma_{k=1}^{K} Opt(k)$ # The average optimism to measure the test performance.
(vii) Compute $MAD_{adj} = MAD_{app} - Opt$. # Optimism adjusted measure for the original model.
(viii) ***Return*** $\langle Opt(k), Opt, MAD_{adj} \rangle$