ROBUST GENERALIZED LIKELIHOOD RATIO TEST BASED ON PENALIZATION

by

Meijiao Zhang

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2018

Approved by:

____
Dr. Jiancheng Jiang

____
Dr. Zhiyi Zhang

____
Dr. Weihua Zhou

____
Dr. Yongge Wang

ABSTRACT

MEIJIAO ZHANG. Robust Generalized Likelihood Ratio Test Based On Penalization. (Under the direction of DR. JIANCHENG JIANG)

The Least absolute deviation combined with the Least absolute shrinkage and selection operator (LAD-LASSO) estimator can do regression shrinkage and selection and is also resistant to outliers or heavy-tailed errors which is proposed in Wang et al. (2007). Generalized likelihood ratio (GLR) test motivated by the likelihood principle, which does not require knowing the underlying distribution family and also shares the Wilks property, has wide applications and nice interpretations [cf. Fan et al. (2001) and Fan and Jiang (2005)]. In this dissertation, we propose a GLR test based on LAD-LASSO estimators in order to combine their advantages together. We obtain the asymptotic distributions of the test statistic by applying the Bahadur representation to the LAD-LASSO estimators. Furthermore, we show that the test has oracle property and can detect alternatives nearing the null hypothesis at a maximum rate of root-n. Simulations are conducted to compare test statistics under different procedures for a variety of error distributions including standard normal, t_3 and mixed normal. A real data example is used to illustrate the performance of the testing approach.

ACKNOWLEDGMENTS

Upon the completion of this dissertation I would sincerely express my gratitude thanks to many people. First of all, I would like to thank my advisor Dr. Jiancheng Jiang for his excellent guidance, caring and patience during my graduate studies at UNC Charlotte. Besides providing me with an effective guidance on my research, his enthusiasm inspired me to push myself to always be better. This dissertation would not be possible without his guidance and encouragement. My appreciation also extends to Dr. Zhiyi Zhang, Dr. Weihua Zhou, and Dr. Yongge Wang for serving as very valuable members in my advisory committee who gave great suggestions on my research. I would also like to thank our Mathematics and Statistics Department. I enjoy the environment here where people are always willing to help each other, especially to Dr. Yuanan Diao, Dr. Mohammad Kazemi, Dr. Shaozhong Deng, Dr. Joel Avrin, Dr. Stanislav Molchanov, Dr. Yanqing Sun, Dr. Hae-Soo Oh, Dr. Isaac Sonin, Dr. Mingxin Xu, Dr. Yang Li, Dr. Jaya Bishwal, Dr. Wafaa Shaban, Dr. Elizabeth Eagle and Mark Hamrick. I would like to thank Dr. Weidong Tian, Dr. Hwan Lin, Dr. Judson Russell and Dr. Xiuli He in the Belk College of Business. In addition, I owe my thanks to my classmates and friends. They had given me very concrete guidance. Last but not least, I would like to deeply thank to my family who have provided endless support and inspiration whenever I was facing difficulties during my PhD life.

TABLE OF CONTENTS

Notation

H_0	Null hypothesis
H_1	Alternative hypothesis
$\xrightarrow{a.s}$	Convergence almost surely
$\xrightarrow{p}$	Convergence in probability
$\xrightarrow{d}$	Convergence in distribution
$o_p(1)$	Converges to zero in probability
$O_p(1)$	Bounded in probability
$a \perp b$	a and b are orthogonal
E	Expectation
Var	Variance
Cov	Covariance
x'	Transpose of x
$N_k(\mu, \Gamma)$	k-variate normal distribution with mean vector μ and covariance matrix Γ
LAD	Least absolute deviation
LASSO	Least absolute shrinkage and selection operator
GLR	Generalized likelihood ratio
CLT	Central limit theorem
LLN	Law of large numbers
LSE	Least-squares estimator
MLE	Maximum likelihood estimator
CDF	Cumulative distribution function
PDF	Probability density function

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Motivation

Consider the linear model

$$y_i = x_i'\beta + z_i'\gamma + \varepsilon_i, \quad i = 1, ..., n, \tag{1}$$

where ε_i are identically independently distributed (i.i.d.) random errors with PDF $f(x)$, median 0 and $E|\varepsilon_i| = \sigma > 0$. Let β and γ are unknown parameters where $p = dim(\beta)$ and $q = dim(\gamma)$.

To obtain an estimator to be robust against outliers and error distributions and also enjoy a sparse representation, Wang et al. (2007) proposed a robust LASSO-type estimator, minimizing from the following LAD-LASSO criterion:

$$\text{LAD-LASSO} = Q(\beta) = \sum_{i=1}^{n} |y_i - x_i'\beta| + n \sum_{j=1}^{p} \lambda_j |\beta_j|$$

In the current study we propose a robust GLR test based on L_1 regression to improve likelihood ratio test. The idea is applicable to some parametric, semiparametric, and nonparametric models.

Our interest here lies on the following testing problem

$$H_0 : \gamma = \gamma_0 \quad \text{versus} \quad H_1 : \gamma \neq \gamma_0$$

regarding β as nuisance parameters.

Let $Y = (y_1, ..., y_n)'$, $X = (x_1, ..., x_n)'$, $Z = (z_1, ..., z_n)'$, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$ in model (1). The reduced model is

$$Y = X\beta + Z\gamma_0 + \varepsilon,$$

and the full model is

$$Y = X\beta + Z\gamma + \varepsilon.$$

When $\{\varepsilon_i\}_{i=1}^n$ are normal, it is known that the LR test is equivalent to the F-test statistic

$$F_n = \frac{(RSS_0 - RSS_1)/q}{RSS_1/(n-p-q)},$$

where RSS_0 and RSS_1 are the residual sum of squares under H_0 and H_1, respectively, based on the least squares estimation. Under the null hypothesis, F_n follows the $\mathcal{F}_{q,n-p-q}$ distribution. Under the alternative F_n has a non-central $\mathcal{F}_{q,n-p-q}(\nu^2)$ distribution with non-centrality parameter

$$\nu^2 = \sigma^{-2} \parallel (I_n - P_1)Z(\gamma - \gamma_0) \parallel^2,$$

where $P_1 = X(X'X)^{-1}X'$, I_n is the $n \times n$ identity matrix, and $\parallel \cdot \parallel$ denotes the L_2 norm of a vector. In general, ν^2 depends on the sample correlations between the variables in X and those in Z.

According to the previous argument, under H_1, we consider the penalized least absolute deviation estimator minimizing

$$Q(\beta, \gamma) = \sum_{i=1}^{n} |y_i - x_i'\beta - z_i'\gamma| + n \sum_{j=1}^{p} \lambda_j |\beta_j|, \quad (2)$$

over β and γ. Let $\hat{\beta}$ and $\hat{\gamma}$ be the resulting estimators. Then the residual sum of absolute deviations under H_1 is

$$RSS_1^\star = \sum_{i=1}^{n} |y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}|$$

Under H_0, we minimize

$$Q(\beta) = \sum_{i=1}^{n} |y_i - x_i'\beta - z_i'\gamma_0| + n\sum_{j=1}^{p} \lambda_j|\beta_j|, \tag{3}$$

over β and get the minimizer $\hat{\beta}_0$. Then the residual sum of absolute deviations under H_0 is

$$RSS_0^\star = \sum_{i=1}^{n} |y_i - x_i'\hat{\beta}_0 - z_i'\gamma_0|$$

Since the error distribution is not specified, the LR test is not available here. Intuitively, we can compare the residual sum of squares from the null and alternative models. Following the idea in Fan et al. (2001) and Fan and Jiang (2007), we define the GLR statistic

$$T_n = \frac{n}{2}\log(RSS_0^\star/RSS_1^\star) \approx \frac{n}{2}\frac{RSS_0^\star - RSS_1^\star}{RSS_1^\star} \tag{4}$$

Large values of T_n suggest rejection of H_0. It is worth pointing out that the GLR test of Fan et al. (2001) is different from the GLR test proposed here, since their GLR test did not use regularization.

In the above estimation we have penalized the nuisance parameters but not the parameters of interest which is different from the common penalized estimation for variable selection where all parameters are penalized. Firstly, it improves the power of the GLR test by penalizing the nuisance parameters. Secondly, if the true values of parameters of interest γ are zero and all parameters are penalized, then asymptotically there is no difference between the penalized estimators of parameters under the null and the alternative hypotheses or T_n can be very small. Now it is hard to reject the null hypothesis. So the size of the test is very small and the power could not be improved.

1.2 Outline

The rest of this dissertation is organized as follows. In Chapter 2, we begin to discuss the model and the theoretical results. We proposed the new test statistic, termed as the GLR test, to test if the parameters of interest under the high dimensional multiple linear regression model is constant or not. The test statistic is constructed based on the comparison of the residual sum of absolute deviations under the null and the alternative hypotheses respectively. The asymptotic distribution of the test statistic has been derived and the detailed proofs are provided in the Appendix. In Chapter 3, we use the simulation results to show the performance of our test statistics and compare our working procedure with the oracle procedure to illustrate the oracle properties of our test statistics. In Chapter 4, a real data example has been applied to show the significance of the testing procedure. In Chapter 5, we conclude the dissertation and discuss some possible directions for future work.

CHAPTER 2: ROBUST GLR TEST BASED ON PENALIZATION

2.1 Notations and Assumptions

For convenience, we define the regression coefficient as $(\beta', \gamma')' = (\beta_a', \beta_b', \gamma')'$, where $\beta_a = (\beta_1, ..., \beta_{p_0})'$, $\beta_b = (\beta_{p_0+1}, ..., \beta_p)'$ and $\gamma = (\gamma_1, ..., \gamma_q)'$. Moreover, assume that $\beta_j \neq 0$ for $j \leq p_0$ and $\beta_j = 0$ for $j > p_0$ for some $p_0 \geq 0$ or $\beta_b = 0$. Thus the correct model has p_0 significant and $(p - p_0)$ insignificant regression variables of nuisance parameter β. Under H_0, its corresponding LAD-LASSO estimator is denoted by $\hat{\beta}_0 = (\hat{\beta}_{0a}', \hat{\beta}_{0b}')'$. Under H_1, its corresponding LAD-LASSO estimator is denoted by $(\hat{\beta}', \hat{\gamma}')' = (\hat{\beta}_a', \hat{\beta}_b', \hat{\gamma}')'$. In addition, we also decompose the covariate $x_i = (x_{ia}', x_{ib}')'$ with $x_{ia} = (x_{i1}, ..., x_{ip_0})'$ and $x_{ib} = (x_{i(p_0+1)}, ..., x_{ip})'$ and define $w_i = (x_i', z_i')' = (w_{i1}, ..., w_{il})'$ where $z_i = (z_{i1}, ..., z_{iq})'$ and $l = p + q$.

To study the theoretical properties of our GLR test statistics, the following assumptions are necessary throughout:

Assumption 2.1. The error ε has continuous and positive density at the origin.

Assumption 2.2. $n^{-1/2} \max_{l \leq p+q, i \leq n} |w_{il}| = o_p(1)$.

Assumption 2.3. There exists positive definite Σ_{xz} such that

$$n^{-1}(W'W) \xrightarrow{p} \Sigma_{xz}, \quad as\ n \to \infty,$$

where $(w_{i1}, ..., w_{il}) = w_i'$ be the ith row of W.

Denote

$$\Sigma_{xz} = E\begin{pmatrix} x_{1a}x_{1a}' & x_{1a}x_{1b}' & x_{1a}z_1' \\ x_{1b}x_{1a}' & x_{1b}x_{1b}' & x_{1b}z_1' \\ z_1x_{1a}' & z_1x_{1b}' & z_1z_1' \end{pmatrix} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix},$$

so $\Sigma_{xx} \triangleq E(x_1x_1') = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ is positive definite and $\Sigma_{33} = E(z_1z_1')$ is also positive definite. Define $\Sigma \triangleq \begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}$, $\Sigma^{-1} \triangleq \begin{pmatrix} \Sigma^{11} & \Sigma^{13} \\ \Sigma^{31} & \Sigma^{33} \end{pmatrix}$. Then Σ and Σ^{-1} are positive definite.

Assumption 2.4. Let $a_n = max\{\lambda_j, 1 \leq j \leq p_0\}$ and $b_n = min\{\lambda_j, p_0 < j \leq p\}$. $\sqrt{n}a_n \to 0$ and $\sqrt{n}b_n \to \infty$ as $n \to \infty$.

Note that Assumption 2.1, 2.2 and 2.3 are typical assumptions and used extensively in literature for establishing the $\sqrt{n}$-consistency and the asymptotic normality of the unpenalized LAD estimator. Furthermore, the Assumption 2.4 appears in Wang et al. (2007) to build the oracle property of the penalized LAD-LASSO estimator.

2.2 Bahadur Representations of the LAD-LASSO Estimators

Under H_0, $\hat{\Delta}_{\beta_{0a}} \triangleq \sqrt{n}(\hat{\beta}_{0a} - \beta_a)$ and $\hat{\Delta}_{\beta_{0b}} \triangleq \sqrt{n}(\hat{\beta}_{0b} - \beta_b)$. Then we have the following theorem:

Theorem 2.1. Assume that the assumptions in Section 2.1 hold. The Bahadur representations for $\hat{\Delta}_{\beta_{0a}}$ and $\hat{\Delta}_{\beta_{0b}}$ are

$$\hat{\Delta}_{\beta_{0a}} = \frac{1}{2}f(0)^{-1}\Sigma_{11}^{-1}n^{-1/2}\sum_{i=1}^{n} x_{ia}\text{sgn}(\varepsilon_i) + o_p(1), \tag{5}$$

$$\hat{\Delta}_{\beta_{0b}} = o_p(1). \tag{6}$$

where the function

$$\operatorname{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

Under H_1, $\hat{\Delta}_{\beta_a} \triangleq \sqrt{n}(\hat{\beta}_a - \beta_a)$, $\hat{\Delta}_{\beta_b} \triangleq \sqrt{n}(\hat{\beta}_b - \beta_b)$ and $\hat{\Delta}_\gamma \triangleq \sqrt{n}(\hat{\gamma} - \gamma)$. Then we have the following theorem states as below.

Theorem 2.2. Assume that the assumptions in Section 2.1 hold. The Bahadur representations for $\hat{\Delta}_{\beta_a}$, $\hat{\Delta}_{\beta_b}$ and $\hat{\Delta}_\gamma$ are

$$\hat{\Delta}_{\beta_a} = \frac{1}{2} f(0)^{-1} n^{-1/2} (\Sigma^{11} \sum_{i=1}^{n} x_{ia} \operatorname{sgn}(\varepsilon_i) + \Sigma^{13} \sum_{i=1}^{n} z_i \operatorname{sgn}(\varepsilon_i)) + o_p(1), \tag{7}$$

$$\hat{\Delta}_\gamma = \frac{1}{2} f(0)^{-1} n^{-1/2} (\Sigma^{31} \sum_{i=1}^{n} x_{ia} \operatorname{sgn}(\varepsilon_i) + \Sigma^{33} \sum_{i=1}^{n} z_i \operatorname{sgn}(\varepsilon_i)) + o_p(1), \tag{8}$$

$$\hat{\Delta}_{\beta_b} = o_p(1). \tag{9}$$

Theorem 2.1 and Theorem 2.2 show that the Bahadur representation of the penalized estimator is the same as that of the unpenalized estimator [cf. Ruppert and Carroll (1980)], indicating that the penalized estimator has oracle property.

2.3 Asymptotic Theory of the GLR Test Statistics

Now let us consider the asymptotic properties of our GLR test statistics.

Theorem 2.3. Assume that the assumptions in Section 2.1 hold. Under H_0, $T_n \xrightarrow{d} \frac{1}{8f(0)\sigma}\chi_q^2$.

However, the distribution of T_n depends on nuisance parameters. So we define $\widetilde{T_n} \triangleq$

$8\hat{f}(0)\hat{\sigma}T_n$, where $\hat{f}(0) \triangleq \frac{1}{nh}\sum_{i=1}^{n} K(\frac{y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}}{h})$ and $\hat{\sigma} \equiv \frac{RSS_1^*}{n}$. In the definition of $\hat{f}(0)$, the kernel $K(x)$ is the normal density function and h is the bandwidth. It is well known that $\hat{f}(0)$ is a consistent estimator of $f(0)$. Applying Lemma 7 in Appendix B, $\hat{\sigma}$ is also a consistent estimator of σ. So we propose the following corollary.

Corollary 2.3.1. *Assume that the assumptions in Section 2.1 hold. Under H_0, $\widetilde{T_n} \xrightarrow{d} \chi_q^2$.*

This is an extension of the Wilks type of phenomenon, by which, we mean that the asymptotic null distribution of $\widetilde{T_n}$ is independent of the nuisance parameter σ and the nuisance design density function f.

To study the power of the proposed test, we consider the local (Pitman) alternatives of the form

$$H_{1n} : \gamma = \gamma_0 + n^{-r}\Delta_\gamma,$$

where $\|\Delta_\gamma\| \neq 0$.

Theorem 2.4. *Assume that the assumptions in Section 2.1 hold. For the testing problem $H_0 \leftrightarrow H_{1n}$ when$r < 1/2$, the test T_n can detect alternative H_{1n} asymptotically with probability one.*

Corollary 2.4.1. *Assume that the assumptions in Section 2.1 hold. For the testing problem $H_0 \leftrightarrow H_{1n}$ when $r < 1/2$, the test $\widetilde{T_n}$ can detect alternative H_{1n} asymptotically with probability one.*

We conclude this section by considering the limiting behavior of the test statistic under the local alternative H_{1n} with $r = 1/2$.

Theorem 2.5. Assume that the assumptions in Section 2.1 hold. Under H_{1n} with $r = 1/2$, $T_n \xrightarrow{d} \frac{1}{8f(0)\sigma}\chi_q^2(\rho^2) + C^2$,

where $\rho^2 = 4f(0)^2\Delta_\gamma'(\Sigma^{33})^{-1}\Delta_\gamma$ and $C^2 = \frac{f(0)}{2\sigma}\Delta_\gamma'\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Delta_\gamma$.

Corollary 2.5.1. Assume that the assumptions in Section 2.1 hold. Under H_{1n} with $r = 1/2$, $\widetilde{T_n} \xrightarrow{d} \chi_q^2(\rho^2) + D^2$,

where $\rho^2 = 4f(0)^2\Delta_\gamma'(\Sigma^{33})^{-1}\Delta_\gamma$ and $D^2 = 4f(0)^2\Delta_\gamma'\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Delta_\gamma$.

The above theorem and corollary show that both tests T_n and $\widetilde{T_n}$ can detect the local alternatives at a maximum rate of $\sqrt{n}$, the optimal rate in all regular parametric tests.

The power of both tests T_n and $\widetilde{T_n}$ can be calculated. Denote the PDF of $\chi_q^2(\rho^2)$ is $g(x)$. The power of T_n is

$$
\begin{aligned}
P(T_n > \frac{1}{8f(0)\sigma}\chi^2_{q,1-\alpha}|H_{1n}) &= \int_{\frac{1}{8f(0)\sigma}\chi^2_{q,1-\alpha}}^{+\infty} 8f(0)\sigma * g(8f(0)\sigma(t - C^2))dt \\
&= \int_{\chi^2_{q,1-\alpha}}^{+\infty} g(\tilde{t} - D^2))d\tilde{t} \\
&= 1 - G(\chi^2_{q,1-\alpha} - D^2) \\
&= Q_{\frac{q}{2}}\left(\sqrt{\rho^2}, \sqrt{\chi^2_{q,1-\alpha} - D^2}\right)
\end{aligned}
$$

where $G(x)$ is the CDF of $\chi_q^2(\rho^2)$ and $Q_M(a, b)$ is the Marcum Q-function.

The $Q_M(a, b)$ function increases as b decreases. It is obvious that non-zero C^2 or D^2 will improve the power.

From the power derivation of T_n, it also indicates that $\widetilde{T_n}$ has the same asymptotic power as T_n. Power results of T_n and $\widetilde{T_n}$ under different error distributions are shown in Table 1-6 in Chapter 3.

CHAPTER 3: SIMULATION

Similarly to Section 2.1 of Wang et al. (2007), we can easily get the LAD-LASSO estimator by creating an augmented dataset including the penalized terms for nuisance parameters. In addition, we use the method in Section 2.3 of Wang et al. (2007) to get the tuning parameter estimate for each λ_j which makes the LAD-LASSO estimator enjoy the same asymptotic efficiency as the oracle estimator.

3.1 Density Estimations under H_0

Specifically, we set $p = 9$ and $\beta = (1, 0, 0, 0, 0, 0, 0, 0, 0)'$. In other words, the first $p_0 = 1$ regression variable is significant, while the other 8 are insignificant. We also set $q = 3$ and $\gamma_0 = (1, 2, 3)'$. For a given i, the covariates x_i and z_i are generated from $N_{12}(0, \Gamma)$,

$$\text{where } \Gamma = \begin{bmatrix} 1 & 0.8 & 0.8^2 & \dots & 0.8^{11} \\ 0.8 & 1 & 0.8 & \dots & 0.8^{10} \\ 0.8^2 & 0.8 & 1 & \dots & 0.8^9 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.8^{11} & 0.8^{10} & 0.8^9 & \dots & 1 \end{bmatrix}.$$

The sample size considered is given by $n = 500$. Furthermore, each response variable y_i is generated according to

$$y_i = x_i'\beta + z_i'\gamma_0 + \varepsilon_i$$

where ε_i is generated from $N(0,1)$.

According to Theorem 2.3 and its corresponding corollary, the distribution of T_n should be asymptotically $\frac{1}{8f(0)\sigma}\chi^2_3$-distributed and $\widetilde{T_n}$ should be asymptotically χ^2_3-distributed. To verify this empirically, we plot the sampling distribution of 1000 simulation statistics of T_n and $\widetilde{T_n}$ against their true distribution density respectively via the kernel density estimate as shown in Figure 1. The two plots depict the T_n and $\widetilde{T_n}$ closely following their true distributions, which is consistent with our asymptotic theory.

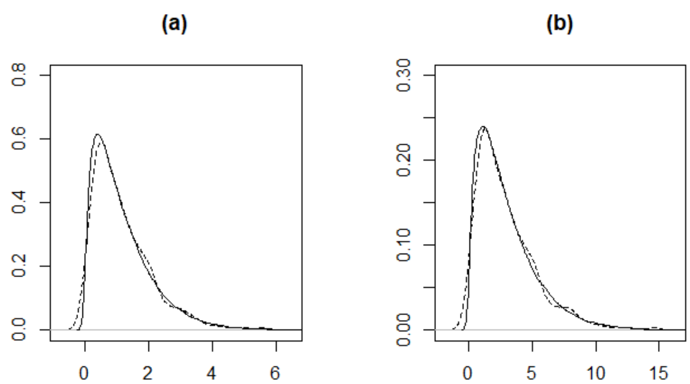

Figure 1: Estimated densities. $(a) : T_n; (b) : \widetilde{T_n}$. Solid: true; dashed: the simulation approximation.

3.2 Power Functions under H_1

We next investigated the power of our tests by considering the following alternative sequences indexed by $\theta = 0, 0.2, 0.4, 0.6, 0.8, 1.0$:

$$H_{1n} : \gamma = \gamma_0 + n^{-1/2}\theta\Delta^*_\gamma,$$

where $\Delta^*_\gamma = (-6, 0, 2)'$ and $\Delta^*_\gamma \perp \gamma_0$.

Furthermore, each response variable y_i is generated according to

$$y_i = x_i'\beta + z_i'\gamma + \varepsilon_i$$

where ε_i is generated from $N(0,1)$.

Note that when $\theta = 0$, the null and the alternative are the same. Therefore, we can expect that: 1) when $\theta = 0$, the power of the test should be close to the significance level; 2) the further is θ away from 0, the greater is the power. These are consistent with the plots as shown in Figure 2. Figure 2 illustrate the power functions of T_n and $\widetilde{T_n}$ against them of their oracle tests based on 1000 simulation iterations of sample size $n = 500$ at three different significance levels: 0.1, 0.05, and 0.01. We can tell from the figure that our tests perform closely to the oracle tests, so our tests have oracle property and should mimic the oracle tests.

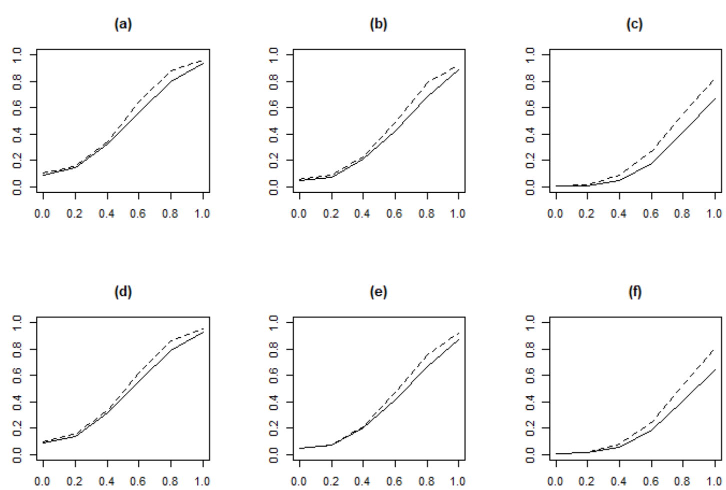

Figure 2: Power functions of T_n and $\widetilde{T_n}$. (a), (b) and (c): T_n; (d), (e) and (f): $\widetilde{T_n}$; From left to right, significance levels are $\alpha = 0.1[(a),(d)], 0.05[(b),(e)], 0.01[(c),(f)]$. Solid: Our test; dashed: Oracle test.

We next compare our test (GLR) with the oracle test, and the test with $\lambda_j = 0$ denoted by (GLR*) for T_n and $\widetilde{T_n}$ respectively in three error distributions: $N(0,1), t_3$ and mixed normal $(0.95N(0,1) + 0.05N(0,9))$. The results shown in tables below indicating that our test is robust against heavy-tailed errors and outliers due to the LAD and also has oracle property due to the LASSO.

Table 1: Power results of T_n for $N(0,1)$ error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	Oracle	0.104	0.154	0.340	0.643	0.877	0.961
		GLR	0.092	0.143	0.323	0.563	0.800	0.936
		GLR*	0.087	0.124	0.202	0.320	0.491	0.701
	0.05	Oracle	0.057	0.087	0.227	0.486	0.787	0.925
		GLR	0.049	0.071	0.209	0.421	0.681	0.888
		GLR*	0.044	0.063	0.115	0.204	0.363	0.589
	0.01	Oracle	0.009	0.019	0.088	0.269	0.555	0.823
		GLR	0.007	0.011	0.052	0.183	0.418	0.665
		GLR*	0.004	0.015	0.027	0.064	0.169	0.336
1000	0.1	Oracle	0.094	0.164	0.354	0.622	0.866	0.974
		GLR	0.084	0.160	0.316	0.574	0.802	0.933
		GLR*	0.100	0.125	0.217	0.348	0.507	0.706
	0.05	Oracle	0.046	0.101	0.245	0.488	0.776	0.940
		GLR	0.034	0.077	0.220	0.416	0.703	0.891
		GLR*	0.051	0.062	0.128	0.238	0.370	0.599
	0.01	Oracle	0.008	0.024	0.082	0.275	0.558	0.814
		GLR	0.005	0.019	0.069	0.184	0.432	0.716
		GLR*	0.012	0.015	0.038	0.090	0.176	0.338

Table 2: Power results of $\widetilde{T_n}$ for $N(0,1)$ error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	Oracle	0.098	0.152	0.333	0.622	0.868	0.955
		GLR	0.090	0.137	0.320	0.553	0.787	0.930
		GLR*	0.087	0.125	0.199	0.324	0.489	0.702
	0.05	Oracle	0.052	0.077	0.215	0.471	0.758	0.918
		GLR	0.048	0.072	0.202	0.418	0.669	0.875
		GLR*	0.043	0.062	0.114	0.201	0.368	0.584
	0.01	Oracle	0.008	0.017	0.084	0.245	0.527	0.804
		GLR	0.009	0.012	0.057	0.184	0.406	0.644
		GLR*	0.005	0.014	0.028	0.061	0.166	0.329
1000	0.1	Oracle	0.080	0.153	0.335	0.599	0.856	0.969
		GLR	0.084	0.149	0.313	0.560	0.788	0.928
		GLR*	0.094	0.121	0.212	0.345	0.493	0.694
	0.05	Oracle	0.038	0.094	0.224	0.467	0.759	0.929
		GLR	0.033	0.069	0.209	0.391	0.678	0.878
		GLR*	0.047	0.059	0.123	0.233	0.357	0.581
	0.01	Oracle	0.007	0.023	0.078	0.265	0.536	0.796
		GLR	0.005	0.019	0.063	0.177	0.411	0.693
		GLR*	0.011	0.015	0.040	0.087	0.172	0.325

Table 3: Power results of T_n for t_3 error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	Oracle	0.109	0.151	0.288	0.542	0.789	0.934
		GLR	0.107	0.131	0.269	0.493	0.709	0.879
		GLR*	0.110	0.116	0.205	0.310	0.439	0.622
	0.05	Oracle	0.051	0.081	0.181	0.420	0.687	0.891
		GLR	0.050	0.061	0.162	0.367	0.587	0.802
		GLR*	0.057	0.058	0.110	0.189	0.319	0.491
	0.01	Oracle	0.007	0.018	0.063	0.190	0.417	0.724
		GLR	0.008	0.010	0.049	0.149	0.313	0.560
		GLR*	0.009	0.012	0.028	0.067	0.135	0.267
1000	0.1	Oracle	0.090	0.143	0.289	0.529	0.808	0.921
		GLR	0.077	0.122	0.283	0.476	0.769	0.888
		GLR*	0.093	0.115	0.162	0.282	0.469	0.605
	0.05	Oracle	0.048	0.079	0.197	0.399	0.686	0.870
		GLR	0.034	0.065	0.162	0.341	0.638	0.805
		GLR*	0.043	0.062	0.088	0.197	0.352	0.483
	0.01	Oracle	0.008	0.021	0.076	0.176	0.444	0.704
		GLR	0.003	0.013	0.040	0.142	0.364	0.573
		GLR*	0.012	0.013	0.021	0.077	0.141	0.246

Table 4: Power results of $\widetilde{T_n}$ for t_3 error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	Oracle	0.064	0.103	0.214	0.458	0.732	0.906
		GLR	0.070	0.087	0.205	0.408	0.643	0.834
		GLR*	0.075	0.078	0.143	0.229	0.367	0.538
	0.05	Oracle	0.031	0.049	0.128	0.339	0.595	0.846
		GLR	0.033	0.033	0.113	0.291	0.500	0.743
		GLR*	0.033	0.042	0.074	0.143	0.252	0.411
	0.01	Oracle	0.003	0.009	0.043	0.113	0.343	0.617
		GLR	0.003	0.005	0.026	0.098	0.220	0.445
		GLR*	0.005	0.006	0.014	0.038	0.083	0.182
1000	0.1	Oracle	0.065	0.105	0.248	0.453	0.751	0.895
		GLR	0.056	0.096	0.224	0.420	0.710	0.848
		GLR*	0.062	0.091	0.119	0.242	0.414	0.547
	0.05	Oracle	0.035	0.056	0.160	0.330	0.613	0.829
		GLR	0.020	0.049	0.114	0.283	0.572	0.742
		GLR*	0.029	0.046	0.065	0.155	0.284	0.422
	0.01	Oracle	0.005	0.010	0.051	0.133	0.376	0.629
		GLR	0.002	0.009	0.019	0.113	0.289	0.489
		GLR*	0.007	0.007	0.014	0.047	0.110	0.193

Table 5: Power results of T_n for mixed normal error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	Oracle	0.101	0.145	0.289	0.582	0.833	0.941
		GLR	0.108	0.130	0.309	0.526	0.787	0.908
		GLR*	0.115	0.120	0.200	0.314	0.522	0.675
	0.05	Oracle	0.051	0.077	0.200	0.451	0.744	0.907
		GLR	0.054	0.074	0.191	0.375	0.670	0.846
		GLR*	0.065	0.070	0.120	0.202	0.375	0.540
	0.01	Oracle	0.007	0.014	0.067	0.243	0.524	0.743
		GLR	0.008	0.013	0.049	0.161	0.381	0.628
		GLR*	0.007	0.011	0.034	0.081	0.171	0.307
1000	0.1	Oracle	0.089	0.148	0.347	0.631	0.835	0.959
		GLR	0.103	0.162	0.302	0.551	0.800	0.932
		GLR*	0.097	0.148	0.192	0.327	0.511	0.661
	0.05	Oracle	0.047	0.085	0.231	0.482	0.731	0.916
		GLR	0.057	0.085	0.205	0.425	0.680	0.869
		GLR*	0.042	0.078	0.112	0.217	0.382	0.535
	0.01	Oracle	0.009	0.020	0.088	0.229	0.501	0.797
		GLR	0.008	0.015	0.067	0.172	0.405	0.665
		GLR*	0.007	0.011	0.035	0.077	0.167	0.310

Table 6: Power results of $\widetilde{T_n}$ for mixed normal error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	Oracle	0.092	0.128	0.274	0.556	0.812	0.934
		GLR	0.100	0.122	0.286	0.503	0.767	0.900
		GLR*	0.106	0.106	0.188	0.292	0.497	0.652
	0.05	Oracle	0.042	0.068	0.169	0.420	0.720	0.890
		GLR	0.046	0.062	0.174	0.351	0.640	0.830
		GLR*	0.055	0.064	0.115	0.190	0.356	0.524
	0.01	Oracle	0.005	0.011	0.056	0.223	0.487	0.715
		GLR	0.008	0.010	0.043	0.146	0.351	0.599
		GLR*	0.007	0.009	0.032	0.074	0.158	0.281
1000	0.1	Oracle	0.081	0.135	0.329	0.599	0.817	0.950
		GLR	0.090	0.151	0.288	0.524	0.785	0.920
		GLR*	0.089	0.136	0.182	0.306	0.483	0.641
	0.05	Oracle	0.038	0.079	0.212	0.455	0.711	0.910
		GLR	0.053	0.081	0.191	0.409	0.660	0.853
		GLR*	0.038	0.070	0.101	0.209	0.365	0.522
	0.01	Oracle	0.005	0.017	0.080	0.204	0.462	0.762
		GLR	0.008	0.014	0.059	0.146	0.370	0.624
		GLR*	0.006	0.010	0.030	0.063	0.152	0.287

In addition, we only change one condition of the true parameters in Section 3.1. We set $\beta = (1, 1, 1, 1, 1, 1, 1, 1, 1)'$. We compare our test (GLR) with the test with $\lambda_j = 0$ denoted by (GLR*) for T_n in three error distributions: $N(0, 1), t_3$ and mixed normal $(0.95N(0, 1) + 0.05N(0, 9))$. In this case, the GLR* is the oracle test since all nuisance parameters are significant. The results shown in tables below indicating that the power of our GLR test is lower than the GLR* test when there are no insignificant nuisance parameters.

Table 7: Power comparisons of T_n for $N(0,1)$ error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	GLR*	0.097	0.133	0.208	0.341	0.537	0.670
		GLR	0.070	0.094	0.157	0.283	0.465	0.618
	0.05	GLR*	0.046	0.067	0.119	0.228	0.413	0.553
		GLR	0.028	0.044	0.086	0.163	0.330	0.491
	0.01	GLR*	0.006	0.013	0.035	0.077	0.167	0.328
		GLR	0.003	0.005	0.018	0.051	0.115	0.261
1000	0.1	GLR*	0.094	0.099	0.196	0.329	0.534	0.686
		GLR	0.085	0.077	0.165	0.306	0.496	0.659
	0.05	GLR*	0.044	0.047	0.108	0.227	0.412	0.572
		GLR	0.036	0.040	0.091	0.199	0.369	0.523
	0.01	GLR*	0.011	0.008	0.036	0.077	0.184	0.326
		GLR	0.008	0.005	0.028	0.057	0.149	0.275

Table 8: Power comparisons of T_n for t_3 error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	GLR*	0.099	0.153	0.172	0.310	0.484	0.599
		GLR	0.077	0.114	0.150	0.271	0.443	0.553
	0.05	GLR*	0.046	0.085	0.110	0.209	0.348	0.483
		GLR	0.030	0.059	0.078	0.178	0.298	0.438
	0.01	GLR*	0.006	0.016	0.030	0.079	0.157	0.246
		GLR	0.006	0.009	0.022	0.056	0.120	0.192
1000	0.1	GLR*	0.095	0.140	0.196	0.298	0.455	0.607
		GLR	0.083	0.128	0.184	0.277	0.432	0.578
	0.05	GLR*	0.047	0.072	0.113	0.187	0.320	0.463
		GLR	0.042	0.061	0.098	0.171	0.295	0.434
	0.01	GLR*	0.009	0.012	0.027	0.063	0.148	0.230
		GLR	0.008	0.009	0.021	0.053	0.120	0.217

Table 9: Power comparisons of T_n for mixed normal error

			θ					
n	α	**Test**	**0.0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
500	0.1	GLR*	0.092	0.130	0.194	0.329	0.481	0.678
		GLR	0.058	0.095	0.151	0.267	0.427	0.623
	0.05	GLR*	0.037	0.068	0.109	0.203	0.357	0.546
		GLR	0.024	0.045	0.081	0.165	0.286	0.472
	0.01	GLR*	0.010	0.016	0.038	0.063	0.155	0.302
		GLR	0.005	0.008	0.025	0.035	0.113	0.228
1000	0.1	GLR*	0.103	0.122	0.201	0.312	0.497	0.645
		GLR	0.088	0.103	0.185	0.287	0.471	0.616
	0.05	GLR*	0.051	0.067	0.137	0.204	0.369	0.525
		GLR	0.038	0.049	0.114	0.175	0.337	0.483
	0.01	GLR*	0.010	0.022	0.027	0.069	0.174	0.269
		GLR	0.008	0.015	0.019	0.050	0.152	0.237

CHAPTER 4: REAL DATA EXAMPLE

We use the data set in Jiang et al. (2012). It consists a random sample of 113 hospitals and for each hospital there are 11 variables.

- Infection risk (y): Average estimated probability of acquiring an infection in the hospital.
- Age (x_1): Average age of patients (in years).
- Length of stay (x_2): Average length of stay of all patients in the hospital (in days).
- Routine culturing ratio (x_3): Ratios of number of cultures performed to number of patients without signs or symptoms of hospital-acquired infection, times 100.
- Routine chest X-ray ratio (x_4): Ratio of number of X-rays performed to numbers of patients without signs or symptoms of pneumonia, times 100.
- Number of beds (x_5): Average number of beds in the hospital during the study period.
- Average daily census (x_6): Average number of patients in the hospital per day during the study period.
- Number of nurses (x_7): Average number of full-time equivalent registered and

licensed practical nurses during the study period (number full time plus one half the number part time).

- Available facilities and services (x_8): Percent of 35 potential facilities and services that are provided by the hospital.

- Medical school affiliation (x_9): 1=Yes, 2=No.

- Region (x_{10}, x_{11}, x_{12}): 1=NE, 2=NC, 3=S, 4=W.

We study whether the infection risk depends on the possible influential factors. Since the medical school affiliation and region are categorical, we introduced a dummy variable x_9 for the medical school affiliation and three dummy variables (x_{10}, x_{11}, x_{12}) for the region as covariates. The model is linear,

$$y_i = \sum_{i=1}^{12} \beta_i x_i + \varepsilon_i, \ i = 1, ..., 113.$$

Table 10: Estimates and standard errors

Method	**LAD**	**LAD-LASSO**
x_1	0.12593(0.01547)	0.11584(0.00784)
x_2	0.29449(0.13214)	0.21798(0.06344)
x_3	0.02575(0.01589)	0.01493(0.00616)
x_4	0.01973(0.00726)	0.02056(0.00406)
x_5	-0.00351(0.00373)	0.00000(0.00086)
x_6	0.01129(0.00458)	0.00405(0.00136)
x_7	-0.00277(0.00232)	0.00000(0.00070)
x_8	-0.01541(0.01401)	0.00000(0.00395)
x_9	-0.08286(0.44532)	0.00000(0.02239)
x_{10}	-0.19322(0.33639)	0.00000(0.05143)
x_{11}	-0.91443(0.34513)	-0.40235(0.18229)
x_{12}	-1.49347(0.43851)	-1.41223(0.26854)

We applied the LAD and LAD-LASSO to get the coefficient estimates for each covariate. The results of variable estimation and selection are presented in Table 10. From Table 10, we can see that all coefficients are nonzero in LAD since we did not apply penalty method. The LAD-LASSO selected six variables: age (x_1), length of stay (x_2), routine culturing ratio (x_3), routine chest X-ray ratio (x_4), average daily census (x_6) and the categorical variable region.

Since the estimated coefficients were positive for x_1, x_2, x_3, x_4, x_6 and negative for x_{11}, x_{12}, which indicates that, during the study period, infection risk (y) increases with the average age of patients (x_1), average length of stay of all patients in the hospital (x_2), the routine culturing ratio (x_3), the routine chest X-ray ratio (x_4), average number of patients in the hospital per day (x_6) and decreases with the region corresponding to x_{11} and x_{12}. This is reasonable, since elderly patients tend to have a weak resistance to infection, and larger x_2 and x_6 increase the chance of cross-infection among patients. In addition, routine cultures and chest X-ray may do harm to the body, and patients without signs or symptoms of hospital-acquired infection or pneumonia should receive it as little as possible. There may be also a region effect to the infection. People from South or West area have stronger resistance to infection comparing with those from other areas.

To check the significance of the selected variables, we performed the hypothesis testing problem:

$$H_0 : \beta_3 = \beta_6 = \beta_8 = 0 \text{ versus } H_1 : \text{at least one of them is not zero.}$$

So our interesting parameters are $\beta_3,\ \beta_6,\ \beta_8$ while the nuisance parameters are the rest. We perform both the GLR test and the GLR* test of $\widetilde{T_n}$ with significance level $\alpha = 0.05$. Their asymptotic null distributions are χ^2_3 which does not depend on nuisance parameters. The realized value of the GLR is calculated as 8.819581 and corresponding p-value is 0.03178837. The realized value of the GLR* is 7.493779 and corresponding p-value is 0.05771851. We reject the null hypothesis in the GLR test while we fail to reject the null hypothesis in the GLR* test. According to Table 10, x_3 is significant in either LAD or LAD-LASSO estimate indicating that we need to reject the null hypothesis. So the GLR test performs better than the GLR* test in this case.

CHAPTER 5: CONCLUSION

In summary, under both the null and the alternative hypothesis we have proposed, the penalized estimators enjoy the oracle property of estimation. The resulting test statistics imitate the oracle test statistics in the sense that those unknown insignificant nuisance parameters were known in advance. Hence:

The GLR test should mimic the oracle GLR test.

This is very useful when there are insignificant nuisance parameters especially in high-dimensional regression models or in classical multiple linear regression models.

In future work, we would like to allow the parameter dimensions depend on the sample size, apply to ARIMA models which the errors are not i.i.d distributed, and expand to various penalty functions such as SCAD in Fan and Li (2001). In addition, we could also apply our procedure to semiparametric and nonparametric models.

REFERENCES

[1] Bickel, P. J. (1975). One-step Huber estimates in linear models. Journal of the American Statistics Association 70, 428-433.

[2] Bickel, P. J. and Ghosh, J. K. (1990). A Decomposition for the Likelihood Ratio Statistic and the Bartlett Correction - A Bayesian Argument. Ann. Statist. 3, 1070-1090.

[3] Cai, Z. (2002). Regression quantile for time series. Econometric Theory 18, 169-192.

[4] Cai, Z., Das, M., Xiong, H., and Wu, X. (2006). Functional coefficient instrumental variables models. Journal of Econometrics. 133, 207-241.

[5] Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. J. Amer. Statist. Assoc. Vol. 95, No. 451, pp. 888-902.

[6] Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941-956.

[7] Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. Journal of the American Statistical Association. 103, 1595-1608.

[8] Chaudhuri, P. (1991). Nonparametric quantile estimates of regression quantiles and their local Bahadur representation. The Annals of Statistics. 19, 760-777.

[9] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1991). Local regression models. Statistical Models in S (Chambers, J. M. and Hastie, T. J., eds), 309C376. Wadsworth & Brooks, Pacific Grove.

[10] El Bantli, F. and Hallin, M. (1999). L_1 Estimation in linear models with heterogeneous white noise, Statistics and Probability Letters, 45, 305-315.

[11] Fan, J. and Gijbels, I.(1996). Local Polynomial Modeling and Its Applications. Chapman and Hall, London.

[12] Fan, J. and Jiang, J. (2005). Nonparametric Inferences for Additive Models. Journal of the American Statistical Association, Vol. 100, No. 471.

[13] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, Vol. 96, No. 456.

[14] Fan, J., Yao, Q., and Tong, H. (1996). Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems. Biometrika, 83, 189-206.

[15] Fan, J., Zhang, C. and Zhang, J. (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon. The Annuals of Statistics. 29, 153-193.

[16] Gourieroux, C., Monfort, A. and Renault, E. (1987). Consistent M-Estimators in a Semi- Parametric Model. CEPREMAP Working Papers (Couverture Orange) 8720, CEPREMAP.

[17] Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. J. Roy. Statist. Soc. B. 55 757C796.

[18] He, X., Ng, P. and Portnoy, S. (1998). Bivariate Quantile Smoothing Splines. Journal of the Royal Statistical Society. Ser. B, 60. 537-550.

[19] He, X. and Shi,P. (1996). Bivariate tensor-product B-splines in a partly linear model. Journal of Multivariate Analysis 58, 162-181.

[20] He, X. and Zhu, L.-X. (2003). A lack-of-fit test for quantile regression. J. Amer. Statist. Assoc. 98 1013C1022. MR2041489.

[21] Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. J. Amer. Statist. Assoc. 87 58C68.

[22] Honda,T. (2004). Quantlie Regression in Varying Coedfficient Models. Journal of Statistical Planning and Inference, 121, 113-125.

[23] Horowitz, J. L. and Lee, S. (2005). Nonparametric Estimation of an Additive Quantile Regression Model.Journal of the American Statistical Association, 100, 1238-1249.

[24] Horowitz, J. L. and Spokoiny, V. G. (2002). An adaptive, rate-optimal test of linearity for median regression models. J. Amer. Statist. Assoc. 97 822C835. MR1941412.

[25] Jiang, J. and Jiang, X. (2011). Inference for partly linear additive Cox models. Statistica Sinica 21, 901-921.

[26] Jiang, X., Jiang, J. and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. Statistica Sinica 22, 1479-1506.

[27] Koenker, R. (2004). Quantreg: An R Package for Quantile Regression and Related Methods. http://cran.r-project.org.

[28] Koenker, R. (2005). Quantile Regression. Econometric Society Monograph Series, Cambridge University Press, New York.

[29] Koenker, R., Bassett, G. (1978). Regression Quantiles. Econometrica 46, 33-50.

[30] Koenker, R. and Bassett, G. (1982a). Robust Tests for Heteroscedasticity Based on Regression Quantiles. Econometrica, 50, 43-61.

[31] Koenker, R. and Bassett, G. (1982b). Test of Linear Hypotheses and L_1 Estimation. Econometrica, 50, 1577-1584.

[32] Koenker, R. and Machado, J. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. Journal of the Royal Statistical Society, Series B,66, 145-163.

[33] Koenker, R., and Xiao, Z. (2004). Unit Root Quantile Autoregression Inference. Journal of the American Statistical Association, 99, 775-787.

[34] Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. Journal of Econometrics 128, 137-164.

[35] Lehmann, E. L. and Romano, J. P. (2005). Testing Statistical Hypotheses. Springer.

[36] Ravikumar, P., Lafferty, J., Liu, H. and Wasserman, L. (2009). Sparse additive models. J. R. Statist. Soc. B 71, 1009-1030.

[37] Roncheti, E. (1985). Robust Model Selection in Regression. Statistics & Probability Letters, 3, 21-23.

[38] Ruppert, D. and Carroll, R. J. (1980). Trimmed Least Squares Estimation in the Linear Model. Journal of the American Statistical Association, Vol. 75, No. 372, pp. 828-838.

[39] Wang, H., Li, G. and Jiang, G. (2007). Robust Regression Shrinkage and Consistent Variable Selection through the LAD-Lasso. Journal of Business & Economic Statistics, Vol. 25, No. 3, pp. 347-355.

[40] Wu, T., Yu, K. and Yu,Y. (2010). Single-index quantile regression. Journal of Multivariate Analysis. 101, 1607-1621.

APPENDIX A: PROOFS OF THEOREMS IN SECTION 2.2

Lemma .1. Under H_0, $\sqrt{n}(\hat{\beta}_{0a} - \beta_a) = O_p(1)$ and $\hat{\beta}_{0b} = 0$ with probability tending to 1.

Proof. See Lemma 1 and Lemma 2 of Wang et al. (2007) for reference. □

Lemma .2. The sequence of solutions $\hat{\beta}_0$ of (3) satisfies

$$n^{-1/2} \sum_{i=1}^{n} x_{ia} \mathrm{sgn}(y_i - x_i' \hat{\beta}_0) \xrightarrow{a.s} 0. \tag{10}$$

Proof. Let $\{e_j\}_{j=1}^{p}$ be the standard basis of R^p. Define

$$G_j(a) = \sum_{i=1}^{n} |y_i - x_i'(\hat{\beta}_0 + ae_j)| + n\lambda_j |\hat{\beta}_{0j} + a|,$$

and let $H_j(a)$ be the derivative of $G_j(a)$, so that

$$H_j(a) = -\sum_{i=1}^{n} x_{ij} \mathrm{sgn}(y_i - x_i'(\hat{\beta}_0 + ae_j)) + n\lambda_j \mathrm{sgn}(\hat{\beta}_{0j} + a).$$

Using the method of Ruppert and Carroll (1980, proof of Lemma A.2), we can show that

$$n^{-1/2}|H_j(0)| \leq 2n^{-1/2} \sum_{i=1}^{n} |x_{ij}| I(y_i - x_i'\hat{\beta}_0 = 0) + 2\sqrt{n}\lambda_j I(\hat{\beta}_{0j} = 0).$$

Applying Lemma A.1 of Ruppert and Carroll (1980), it can be shown that

$$\sum_{i=1}^{n} |x_{ij}| I(y_i - x_i'\hat{\beta}_0 = 0) \xrightarrow{a.s} 0.$$

By applying $\sqrt{n}\lambda_j \rightarrow 0$ when $1 \leq j \leq p_0$ of Assumption 2.4, it is clear that

$n^{-1/2}H_j(0) \xrightarrow{a.s} 0$ when $1 \leq j \leq p_0$. In addition, due to

$$n^{-1/2}H_j(0) = -n^{-1/2}\sum_{i=1}^{n} x_{ij}\text{sgn}(y_i - x_i'\hat{\beta}_0) + \sqrt{n}\lambda_j\text{sgn}(\hat{\beta}_{0j}),$$

using $\sqrt{n}\lambda_j \to 0$ when $1 \leq j \leq p_0$ of Assumption 2.4 again, the result (10) is proved. □

Proof of Theorem 2.1 :

Proof. For $\Delta \in R^p$, define

$$M(\Delta) = n^{-1/2}\sum_{i=1}^{n} x_i\psi_\tau(\varepsilon_i - n^{-1/2}x_i'\Delta),$$

where $\psi_\tau(x) = \tau - I(x < 0)$.

Let $\hat{\Delta}_0 \triangleq (\hat{\Delta}_{\beta_{0a}}', \hat{\Delta}_{\beta_{0b}}')'$. From Lemma 1, we know that $\hat{\Delta}_0 = O_p(1)$. Applying Lemma A.3 of Ruppert and Carroll (1980), we have

$$M(\hat{\Delta}_0) - M(0) + f(0)\Sigma_{xx}\hat{\Delta}_0 = o_p(1).$$

Plug in the definition of $M(\Delta)$, we have

$$n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} x_{ia} \\ x_{ib} \end{pmatrix}\psi_\tau(\varepsilon_i - n^{-1/2}x_i'\hat{\Delta}_0) - n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} x_{ia} \\ x_{ib} \end{pmatrix}\psi_\tau(\varepsilon_i) + f(0)\Sigma_{xx}\begin{pmatrix} \hat{\Delta}_{\beta_{0a}} \\ \hat{\Delta}_{\beta_{0b}} \end{pmatrix}$$

which is equal to $o_p(1)$. This leads to

$$n^{-1/2}\sum_{i=1}^{n} x_{ia}\psi_\tau(\varepsilon_i - n^{-1/2}x_i'\hat{\Delta}_0) - n^{-1/2}\sum_{i=1}^{n} x_{ia}\psi_\tau(\varepsilon_i) + f(0)(\Sigma_{11}\hat{\Delta}_{\beta_{0a}} + \Sigma_{12}\hat{\Delta}_{\beta_{0b}}) = o_p(1).$$

Applying Lemma 1, it is obvious that $\hat{\Delta}_{\beta_{0b}} = o_p(1)$. Using Lemma 2, it shows that

$n^{-1/2}\sum_{i=1}^{n} x_{ia}\psi_\tau(\varepsilon_i - n^{-1/2}x_i'\hat{\Delta}_0) \xrightarrow{a.s} 0$, then we have

$$f(0)\Sigma_{11}\hat{\Delta}_{\beta_{0a}} = n^{-1/2}\sum_{i=1}^{n} x_{ia}\psi_\tau(\varepsilon_i) + o_p(1).$$

In L_1-norm regression, $\tau = \frac{1}{2}$. So $\psi_\tau(\varepsilon_i) = \frac{1}{2}\text{sgn}(\varepsilon_i)$ when $\varepsilon_i \neq 0$. Thus, the Bahadur representation of $\hat{\Delta}_0$ in Theorem 2.1 has been proved. □

Lemma .3. Under H_1, $\sqrt{n}(\hat{\beta}_a - \beta_a) = O_p(1)$, $\sqrt{n}(\hat{\gamma} - \gamma) = O_p(1)$ and $\hat{\beta}_b = 0$ with probability tending to 1.

Proof. Since we don't use penalty on our interest parameter γ, so the tuning parameter for γ is zero, which satisfies the Assumption 2.4. The result follows from Lemma 1 and Lemma 2 of Wang et al. (2007). □

Lemma .4. The sequence of solutions $(\hat{\beta}, \hat{\gamma})$ of (2) satisfies

$$n^{-1/2}\sum_{i=1}^{n} x_{ia}\text{sgn}(y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}) \xrightarrow{a.s} 0. \quad (11)$$

and

$$n^{-1/2}\sum_{i=1}^{n} z_i\text{sgn}(y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}) \xrightarrow{a.s} 0. \quad (12)$$

Proof. Let $\{e_j\}_{j=1}^{p}$ be the standard basis of R^p. Define

$$L_j(a) = \sum_{i=1}^{n} |y_i - x_i'(\hat{\beta} + ae_j) - z_i'\hat{\gamma}| + n\lambda_j|\hat{\beta}_j + a|,$$

and let $N_j(a)$ be the derivative of $L_j(a)$, so that

$$N_j(a) = -\sum_{i=1}^{n} x_{ij}\text{sgn}(y_i - x_i'(\hat{\beta} + ae_j) - z_i'\hat{\gamma}) + n\lambda_j\text{sgn}(\hat{\beta}_j + a).$$

Then result (11) follows from Lemma 2.

Let $\{u_k\}_{k=1}^q$ be the standard basis of R^q. Define

$$L_k^*(a) = \sum_{i=1}^n |y_i - x_i'\hat{\beta} - z_i'(\hat{\gamma} + au_k)|$$

and let $N_k^*(a)$ be the derivative of $L_k^*(a)$, so that

$$N_k^*(a) = -\sum_{i=1}^n z_{ik}\text{sgn}(y_i - x_i'\hat{\beta} - z_i'(\hat{\gamma} + au_k)).$$

The result (12) follows from Ruppert and Carroll (1980, proof of Lemma A.2). □

Proof of Theorem 2.2 :

Proof. For $\Delta \in R^{p+q}$, define

$$M(\Delta) = n^{-1/2}\sum_{i=1}^n w_i\psi_\tau(\varepsilon_i - n^{-1/2}w_i'\Delta).$$

Let $\hat{\Delta} \triangleq (\hat{\Delta}_{\beta_a}', \hat{\Delta}_{\beta_b}', \hat{\Delta}_\gamma')'$. From Lemma 3, we know that $\hat{\Delta} = O_p(1)$. Applying Lemma A.3 of Ruppert and Carroll (1980), we have

$$M(\hat{\Delta}) - M(0) + f(0)\Sigma_{xz}\hat{\Delta} = o_p(1).$$

Plug in the definition of $M(\Delta)$ and simple algebra, we have both

$$n^{-1/2}\sum_{i=1}^n x_{ia}\psi_\tau(\varepsilon_i - n^{-1/2}w_i'\hat{\Delta}) - n^{-1/2}\sum_{i=1}^n x_{ia}\psi_\tau(\varepsilon_i) + f(0)(\Sigma_{11}\hat{\Delta}_{\beta_a} + \Sigma_{12}\hat{\Delta}_{\beta_b} + \Sigma_{13}\hat{\Delta}_\gamma)$$

and

$$n^{-1/2}\sum_{i=1}^n z_i\psi_\tau(\varepsilon_i - n^{-1/2}w_i'\hat{\Delta}) - n^{-1/2}\sum_{i=1}^n z_i\psi_\tau(\varepsilon_i) + f(0)(\Sigma_{31}\hat{\Delta}_{\beta_a} + \Sigma_{32}\hat{\Delta}_{\beta_b} + \Sigma_{33}\hat{\Delta}_\gamma)$$

equal to $o_p(1)$.

By applying Lemma 3, it is obvious that $\hat{\Delta}_{\beta_b} = o_p(1)$. Using Lemma 4, we can see that $n^{-1/2}\sum_{i=1}^{n} x_{ia}\psi_\tau(\varepsilon_i - n^{-1/2}w_i'\hat{\Delta}) \xrightarrow{a.s} 0$ and $n^{-1/2}\sum_{i=1}^{n} z_i\psi_\tau(\varepsilon_i - n^{-1/2}w_i'\hat{\Delta}) \xrightarrow{a.s} 0$. So the above results can be simplified as

$$f(0)\Sigma \begin{pmatrix} \hat{\Delta}_{\beta_a} \\ \\ \hat{\Delta}_\gamma \end{pmatrix} = n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} x_{ia} \\ \\ z_i \end{pmatrix} \psi_\tau(\varepsilon_i) + o_p(1),$$

which is equivalent as

$$\begin{pmatrix} \hat{\Delta}_{\beta_a} \\ \\ \hat{\Delta}_\gamma \end{pmatrix} = f(0)^{-1}\Sigma^{-1}n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} x_{ia} \\ \\ z_i \end{pmatrix} \psi_\tau(\varepsilon_i) + o_p(1).$$

Using the definition of Σ^{-1} and $\psi_\tau(\varepsilon_i)$ when $\tau = \frac{1}{2}$, the Bahadur representation of $\hat{\Delta}$ in Theorem 2.2 has been proved. □

APPENDIX B: PROOFS OF THEOREMS IN SECTION 2.3

Lemma .5. According to the notations and assumptions in Section 2.1,

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{13}B^{-1}\Sigma_{31}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{13}B^{-1} \\ -B^{-1}\Sigma_{31}\Sigma_{11}^{-1} & B^{-1} \end{pmatrix},$$

where $B = \Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}$, and provided Σ_{11}^{-1} and B^{-1} exist.

Proof. It is well known known result by using simple matrix algebra. □

Lemma .6.

$$\hat{\Delta}_{\beta_{0a}} = \hat{\Delta}_{\beta_a} + \Sigma_{11}^{-1}\Sigma_{13}\hat{\Delta}_{\gamma}$$

Proof. Define $\xi_n = n^{-1/2}\sum_{i=1}^{n} x_{ia}\text{sgn}(\varepsilon_i)$, $\xi_n^* = n^{-1/2}\sum_{i=1}^{n} z_i\text{sgn}(\varepsilon_i)$ and $\eta_n = (\xi_n', \xi_n^{*\prime})' = n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} x_{ia} \\ z_i \end{pmatrix}\text{sgn}(\varepsilon_i)$. It is obvious to see that $\xi_n \xrightarrow{d} N(0, \Sigma_{11})$, $\xi_n^* \xrightarrow{d} N(0, \Sigma_{33})$ and $\eta_n \xrightarrow{d} N(0, \Sigma)$. Applying the Bahadur representation of the LAD-LASSO estimators in Theorem 2.1 and Theorem 2.2, then we have

$$\hat{\Delta}_{\beta_{0a}} = \frac{1}{2}f(0)^{-1}\Sigma_{11}^{-1}\xi_n + o_p(1), \tag{13}$$

$$\hat{\Delta}_{\beta_a} = \frac{1}{2}f(0)^{-1}(\Sigma^{11}\xi_n + \Sigma^{13}\xi_n^*) + o_p(1), \tag{14}$$

$$\hat{\Delta}_{\gamma} = \frac{1}{2}f(0)^{-1}(\Sigma^{31}\xi_n + \Sigma^{33}\xi_n^*) + o_p(1). \tag{15}$$

We can plug in the above equations to $\hat{\Delta}_{\beta_a} + \Sigma_{11}^{-1}\Sigma_{13}\hat{\Delta}_{\gamma}$, so

$$\begin{aligned}\hat{\Delta}_{\beta_a} + \Sigma_{11}^{-1}\Sigma_{13}\hat{\Delta}_{\gamma} &= \frac{1}{2}f(0)^{-1}[\Sigma^{11}\xi_n + \Sigma^{13}\xi_n^* + \Sigma_{11}^{-1}\Sigma_{13}(\Sigma^{31}\xi_n + \Sigma^{33}\xi_n^*)] + o_p(1) \\ &= \frac{1}{2}f(0)^{-1}[(\Sigma^{11} + \Sigma_{11}^{-1}\Sigma_{13}\Sigma^{31})\xi_n + (\Sigma^{13} + \Sigma_{11}^{-1}\Sigma_{13}\Sigma^{33})\xi_n^*] + o_p(1).\end{aligned}$$

Consider the two matrix before ξ_n and ξ_n^*:

$$\Sigma^{11} + \Sigma_{11}^{-1}\Sigma_{13}\Sigma^{31} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{13}B^{-1}\Sigma_{31}\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{13}(-B^{-1}\Sigma_{31}\Sigma_{11}^{-1}) = \Sigma_{11}^{-1},$$

$$\Sigma^{13} + \Sigma_{11}^{-1}\Sigma_{13}\Sigma^{33} = -\Sigma_{11}^{-1}\Sigma_{13}B^{-1} + \Sigma_{11}^{-1}\Sigma_{13}B^{-1} = 0.$$

Plug them back in, so we have $\hat{\Delta}_{\beta_a} + \Sigma_{11}^{-1}\Sigma_{13}\hat{\Delta}_\gamma = \frac{1}{2}f(0)^{-1}\Sigma_{11}^{-1}\xi_n + o_p(1)$, or the Lemma is proved. □

Lemma .7. $RSS_1^\star/n = \sigma + o_p(1)$

Proof. Plug in the definition of $RSS_1^\star$, we can see that

$$\begin{aligned}
RSS_1^\star/n &= \frac{1}{n}\sum_{i=1}^n |y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}| \\
&= \frac{1}{n}\sum_{i=1}^n |y_i - x_{ia}'\hat{\beta}_a - x_{ib}'\hat{\beta}_b - z_i'\hat{\gamma}| \\
&= \frac{1}{n}\sum_{i=1}^n |y_i - x_{ia}'\hat{\beta}_a - z_i'\hat{\gamma}| + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^n |\varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z_i'\hat{\Delta}_\gamma| + o_p(1),
\end{aligned}$$

since $P(\hat{\beta}_b = 0) \to 1$ as $n \to \infty$.

Define $\hat{\varepsilon}_i \triangleq \varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z_i'\hat{\Delta}_\gamma$, and according to the LLN theorem,

$$RSS_1^\star/n \xrightarrow{p} E|\hat{\varepsilon}_i|.$$

Let $I \triangleq \sum_{i=1}^n(|\varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z_i'\hat{\Delta}_\gamma| - |\varepsilon_i|)$. According to Wang et al. (2007), it holds that

$$I = -\eta_n' \begin{pmatrix} \hat{\Delta}_{\beta_a} \\ \hat{\Delta}_\gamma \end{pmatrix} + f(0)(\hat{\Delta}_{\beta_a}', \hat{\Delta}_\gamma')\Sigma \begin{pmatrix} \hat{\Delta}_{\beta_a} \\ \hat{\Delta}_\gamma \end{pmatrix} + o_p(1). \tag{16}$$

From Equation (14) and (15), we have

$$\begin{pmatrix} \hat{\Delta}_{\beta_a} \\ \\ \hat{\Delta}_{\gamma} \end{pmatrix} = \frac{1}{2} f(0)^{-1} \Sigma^{-1} \eta_n + o_p(1). \tag{17}$$

Plug in (17) back to (16), we can get $I = -\frac{1}{4f(0)} \eta_n' \Sigma^{-1} \eta_n + o_p(1)$. Since $\eta_n \xrightarrow{d} N(0, \Sigma)$, then $\eta_n' \Sigma^{-1} \eta_n \xrightarrow{d} \chi^2_{p_0+q}$. Thus, $I \xrightarrow{d} -\frac{1}{4f(0)} \chi^2_{p_0+q} \Rightarrow I = O_p(1) \Rightarrow I/n = o_p(1)$.

Using the definition of I and the LLN theorem, $I/n \xrightarrow{p} E|\hat{\varepsilon}_i| - E|\varepsilon_i|$. Thus, $E|\hat{\varepsilon}_i| = E|\varepsilon_i| + o_p(1) \Rightarrow RSS_1^\star/n = E|\varepsilon_i| + o_p(1) = \sigma + o_p(1)$. □

Lemma .8. $\hat{\Delta}_\gamma \xrightarrow{d} N(0, \frac{1}{4f(0)^2} B^{-1})$

Proof. According to Equation (15), $\hat{\Delta}_\gamma = \frac{1}{2} f(0)^{-1} (\Sigma^{31}, \Sigma^{33}) \eta_n + o_p(1)$ where $\eta_n \xrightarrow{d} N(0, \Sigma)$. In order to calculate the variance of $\hat{\Delta}_\gamma$, we need to calculate the matrix $(\Sigma^{31}, \Sigma^{33}) \Sigma (\Sigma^{31}, \Sigma^{33})'$. According to Lemma 5, it can be shown that the matrix $(\Sigma^{31}, \Sigma^{33}) \Sigma (\Sigma^{31}, \Sigma^{33})' = B^{-1}$. Thus $\hat{\Delta}_\gamma \xrightarrow{d} N(0, \frac{1}{4f(0)^2} B^{-1})$. □

Proof of Theorem 2.3

Proof. Consider $\frac{RSS_0^\star - RSS_1^\star}{2}$ under H_0.

$$\begin{aligned} \frac{RSS_0^\star - RSS_1^\star}{2} &= \frac{1}{2} \sum_{i=1}^{n} (|y_i - x_i' \hat{\beta}_0 - z_i' \gamma_0| - |y_i - x_i' \hat{\beta} - z_i' \hat{\gamma}|) \\ &= \frac{1}{2} \sum_{i=1}^{n} (|y_i - x_{ia}' \hat{\beta}_{0a} - x_{ib}' \hat{\beta}_{0b} - z_i' \gamma_0| - |y_i - x_{ia}' \hat{\beta}_a - x_{ib}' \hat{\beta}_b - z_i' \hat{\gamma}|) \\ &= \frac{1}{2} \sum_{i=1}^{n} (|y_i - x_{ia}' \hat{\beta}_{0a} - z_i' \gamma_0| - |y_i - x_{ia}' \hat{\beta}_a - z_i' \hat{\gamma}|) + o_p(1), \end{aligned}$$

since $P(\hat{\beta}_{0b} = 0) \to 1$ and $P(\hat{\beta}_b = 0) \to 1$ as $n \to \infty$.

Continue the derivation, we have

$$\frac{RSS_0^\star - RSS_1^\star}{2} = \frac{1}{2}\sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_{0a}}| - |\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z'_i\hat{\Delta}_\gamma|)$$
$$= \frac{1}{2}\sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_{0a}}| - |\varepsilon_i|) - \frac{1}{2}\sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z'_i\hat{\Delta}_\gamma| - |\varepsilon_i|)$$
$$= \frac{I_1 - I_2}{2}.$$

According to Wang et al. (2007), I_1 can be written as

$$I_1 = \sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_{0a}}| - |\varepsilon_i|)$$
$$= -\xi'_n\hat{\Delta}_{\beta_{0a}} + f(0)\hat{\Delta}'_{\beta_{0a}}\Sigma_{11}\hat{\Delta}_{\beta_{0a}} + o_p(1). \tag{18}$$

I_2 can be written as

$$I_2 = \sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z'_i\hat{\Delta}_\gamma| - |\varepsilon_i|)$$
$$= -\eta'_n(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)' + f(0)(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)\Sigma(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)' + o_p(1). \tag{19}$$

From Equation (13), we can get

$$\xi_n = 2f(0)\Sigma_{11}\hat{\Delta}_{\beta_{0a}} + o_p(1) \tag{20}$$

From Equation (17), we can get

$$\eta_n = 2f(0)\Sigma(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)' + o_p(1) \tag{21}$$

Plug Equation (20) and (21) back to (18) and (19), then we have

$I_1 = -f(0)\hat{\Delta}'_{\beta_{0a}}\Sigma_{11}\hat{\Delta}_{\beta_{0a}} + o_p(1)$, and $I_2 = -f(0)(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_{\gamma})\Sigma(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_{\gamma})' + o_p(1)$.

$$I_1 - I_2 = f(0)[(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_{\gamma})\Sigma(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_{\gamma})' - \hat{\Delta}'_{\beta_{0a}}\Sigma_{11}\hat{\Delta}_{\beta_{0a}}] + o_p(1)$$
$$= f(0)(\hat{\Delta}'_{\beta_a}\Sigma_{11}\hat{\Delta}_{\beta_a} + \hat{\Delta}'_{\gamma}\Sigma_{31}\hat{\Delta}_{\beta_a} + \hat{\Delta}'_{\beta_a}\Sigma_{13}\hat{\Delta}_{\gamma} + \hat{\Delta}'_{\gamma}\Sigma_{33}\hat{\Delta}_{\gamma} - \hat{\Delta}'_{\beta_{0a}}\Sigma_{11}\hat{\Delta}_{\beta_{0a}}) + o_p(1).$$

Applying Lemma 6, we can get

$$\hat{\Delta}'_{\beta_{0a}}\Sigma_{11}\hat{\Delta}_{\beta_{0a}} = \hat{\Delta}'_{\beta_a}\Sigma_{11}\hat{\Delta}_{\beta_a} + \hat{\Delta}'_{\gamma}\Sigma_{31}\hat{\Delta}_{\beta_a} + \hat{\Delta}'_{\beta_a}\Sigma_{13}\hat{\Delta}_{\gamma} + \hat{\Delta}'_{\gamma}\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\hat{\Delta}_{\gamma},$$

$$I_1 - I_2 = f(0)\hat{\Delta}'_{\gamma}(\Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13})\hat{\Delta}_{\gamma} + o_p(1) = f(0)\hat{\Delta}'_{\gamma}B\hat{\Delta}_{\gamma} + o_p(1).$$

By applying Lemma 7, we can get $T_n = \frac{I_1 - I_2}{2\sigma} + o_p(1)$ under H_0. In addition, using Lemma 8, we can get $I_1 - I_2 \xrightarrow{d} \frac{1}{4f(0)}\chi^2_q$. Thus, $T_n \xrightarrow{d} \frac{1}{8f(0)\sigma}\chi^2_q$ under H_0. □

Proof of Theorem 2.4

Proof. Consider $\frac{RSS_0^\star - RSS_1^\star}{2}$ under H_{1n}.

$$\begin{aligned}\frac{RSS_0^\star - RSS_1^\star}{2} &= \frac{1}{2}\sum_{i=1}^{n}(|y_i - x_i'\hat{\beta}_0 - z_i'\gamma_0| - |y_i - x_i'\hat{\beta} - z_i'\hat{\gamma}|)\\ &= \frac{1}{2}\sum_{i=1}^{n}(|y_i - x_{ia}'\hat{\beta}_{0a} - x_{ib}'\hat{\beta}_{0b} - z_i'\gamma_0| - |y_i - x_{ia}'\hat{\beta}_a - x_{ib}'\hat{\beta}_b - z_i'\hat{\gamma}|)\\ &= \frac{1}{2}\sum_{i=1}^{n}(|y_i - x_{ia}'\hat{\beta}_{0a} - z_i'\gamma_0| - |y_i - x_{ia}'\hat{\beta}_a - z_i'\hat{\gamma}|) + o_p(1),\end{aligned}$$

since $P(\hat{\beta}_{0b} = 0) \to 1$ and $P(\hat{\beta}_b = 0) \to 1$ as $n \to \infty$.

Continue the derivation, we have

$$\frac{RSS_0^\star - RSS_1^\star}{2} = \frac{1}{2}\sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_{0a}} + n^{-r}z_i'\Delta\gamma| - |\varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z_i'\hat{\Delta}_{\gamma}|)$$
$$= \frac{1}{2}\sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_{0a}} + n^{-r}z_i'\Delta\gamma| - |\varepsilon_i|) - \frac{1}{2}\sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x_{ia}'\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z_i'\hat{\Delta}_{\gamma}|$$
$$- |\varepsilon_i|) = \frac{I_3 - I_4}{2}.$$

According to Wang et al. (2007),

$$I_3 = \sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_{0a}} + n^{-r}z'_i\Delta\gamma| - |\varepsilon_i|)$$
$$= -\eta'_n(\hat{\Delta}'_{\beta_{0a}}, -n^{1/2-r}\Delta'_\gamma)' + f(0)(\hat{\Delta}'_{\beta_{0a}}, -n^{1/2-r}\Delta'_\gamma)\Sigma(\hat{\Delta}'_{\beta_{0a}}, -n^{1/2-r}\Delta'_\gamma)' + o_p(1),$$

and

$$I_4 = \sum_{i=1}^{n}(|\varepsilon_i - n^{-\frac{1}{2}}x'_{ia}\hat{\Delta}_{\beta_a} - n^{-\frac{1}{2}}z'_i\hat{\Delta}_\gamma| - |\varepsilon_i|)$$
$$= -\eta'_n(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)' + f(0)(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)\Sigma(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)' + o_p(1).$$

By using Equation (21), we can get

$$I_3 = -2f(0)(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)\Sigma(\hat{\Delta}'_{\beta_{0a}}, -n^{1/2-r}\Delta'_\gamma)'$$
$$+ f(0)(\hat{\Delta}'_{\beta_{0a}}, -n^{1/2-r}\Delta'_\gamma)\Sigma(\hat{\Delta}'_{\beta_{0a}}, -n^{1/2-r}\Delta'_\gamma)' + o_p(1),$$

and

$$I_4 = -f(0)(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)\Sigma(\hat{\Delta}'_{\beta_a}, \hat{\Delta}'_\gamma)' + o_p(1).$$

Replacing $\hat{\Delta}_{\beta_{0a}}$ with $\hat{\Delta}_{\beta_a}$ and $\hat{\Delta}_\gamma$ by using Lemma 6, we can get

$$I_3 - I_4 = f(0)\hat{\Delta}'_\gamma B\hat{\Delta}_\gamma + 2f(0)n^{1/2-r}\Delta'_\gamma B\hat{\Delta}_\gamma + f(0)n^{1-2r}\Delta'_\gamma\Sigma_{33}\Delta_\gamma + o_p(1)$$
$$= f(0)(\hat{\Delta}_\gamma + n^{1/2-r}\Delta_\gamma)'B(\hat{\Delta}_\gamma + n^{1/2-r}\Delta_\gamma) + f(0)n^{1-2r}\Delta'_\gamma\Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13}\Delta_\gamma + o_p(1)$$
$$= I_3^* + I_4^* + o_p(1).$$

From Lemma 8 we can see that $\hat{\Delta}_\gamma + n^{1/2-r}\Delta_\gamma \xrightarrow{d} N(n^{1/2-r}\Delta_\gamma, \frac{1}{4f(0)^2}B^{-1})$ which leads to $I_3^* \xrightarrow{d} \frac{1}{4f(0)}\chi_q^2(\rho^2)$ where the non-centrality parameter $\rho^2 = 4f(0)^2n^{1-2r}\Delta'_\gamma B\Delta_\gamma$. $I_4^* \to \infty$ as $n \to \infty$ when $r < 1/2$. By Slutsky's Theorem and Lemma 7, we can get $T_n \xrightarrow{d} \frac{I_3^*+I_4^*}{2\sigma}$ under H_1. $P(T_n > \frac{1}{8f(0)\sigma}\chi^2_{q,1-\alpha}|H_{1n}) \to 1$ where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)$th quantile of χ_q^2 if $r < 1/2$. □

Proof of Theorem 2.5

Proof. According to the proof of Theorem 2.4, if $r = 1/2$, we have

$$
\begin{aligned}
I_3 - I_4 &= f(0)\hat{\Delta}_\gamma' B \hat{\Delta}_\gamma + 2f(0)\Delta_\gamma' B \hat{\Delta}_\gamma + f(0)\Delta_\gamma' \Sigma_{33} \Delta_\gamma + o_p(1) \\
&= f(0)(\hat{\Delta}_\gamma + \Delta_\gamma)' B (\hat{\Delta}_\gamma + \Delta_\gamma) + f(0)\Delta_\gamma' \Sigma_{31} \Sigma_{11}^{-1} \Sigma_{13} \Delta_\gamma + o_p(1) \\
&= I_3^* + I_4^* + o_p(1).
\end{aligned}
$$

Then $I_3^* \xrightarrow{d} \frac{1}{4f(0)} \chi_q^2(\rho^2)$ where the non-centrality parameter $\rho^2 = 4f(0)^2 \Delta_\gamma' B \Delta_\gamma$.

Thus, by Slutsky's Theorem and Lemma 7, $T_n \xrightarrow{d} \frac{1}{8f(0)\sigma} \chi_q^2(\rho^2) + C^2$ where

$C^2 = \frac{f(0)}{2\sigma} \Delta_\gamma' \Sigma_{31} \Sigma_{11}^{-1} \Sigma_{13} \Delta_\gamma$ under H_{1n}. □