Received: 10 October 2019 | Revised: 3 September 2020 | Accepted: 5 September 2020

DOI: 10.1002/sim.8765

RESEARCH ARTICLE

Statistics in Medicine WILEY

Kernel density-based likelihood ratio tests for linear regression models

Feifei Yan[1] | **Qing-Song Xu**[1] | **Man-Lai Tang**[2] | **Ziqi Chen**[3]

[1]School of Mathematics and Statistics, Central South University, Changsha, China

[2]Department of Mathematics, Statistics and Insurance, Hang Seng University of Hong Kong, Hong Kong, China

[3]School of Statistics, KLATASDS-MOE, East China Normal University, Shanghai, China

Correspondence
Ziqi Chen, School of Statistics, KLATASDS-MOE, East China Normal University, Shanghai, 200062, China.
Email: zqchen@fem.ecnu.edu.cn

Funding information
Innovation Program of Central South University, Grant/Award Number: 2018zzts102; Research Grant Council of the Hong Kong Special Administrative Region, Grant/Award Numbers: UGC/FDS14/P01/16, UGC/FDS14/P02/18; National Natural Science Foundation of China, Grant/Award Numbers: 11871124, 11871477

In this article, we develop a so-called profile likelihood ratio test (PLRT) based on the estimated error density for the multiple linear regression model. Unlike the existing likelihood ratio test (LRT), our proposed PLRT does not require any specification on the error distribution. The asymptotic properties are developed and the Wilks phenomenon is studied. Simulation studies are conducted to examine the performance of the PLRT. It is observed that our proposed PLRT generally outperforms the existing LRT, empirical likelihood ratio test and the weighted profile likelihood ratio test in sense that (i) its type I error rates are closer to the prespecified nominal level; (ii) it generally has higher powers; (iii) it performs satisfactorily when moments of the error do not exist (eg, Cauchy distribution); and (iv) it has higher probability of correctly selecting the correct model in the multiple testing problem. A mammalian eye gene expression dataset and a concrete compressive strength dataset are analyzed to illustrate our methodologies.

KEYWORDS
likelihood ratio test, profile likelihood ratio test, semiparametric approach, Wilks phenomenon

1 | INTRODUCTION

The likelihood ratio test (LRT) has been widely adopted for assessing the goodness of fit of two competing statistical models based on the ratio of their likelihoods and its large sample distribution enjoys the Wilks theorem.[1,2] It is noticed that the LRT for multiple linear regression model assumes that the common density function of the error terms is known (eg, normally distributed), which is, however, seldom satisfied in practical applications. To relax the distributional assumption of the LRT, Peng et al[3] proposed a test based on an empirical likelihood objective function for the hypothesis testing problem of parameters. However, their method requires to split the sample into two parts and assumes the existence of the second or higher order moments of the error terms.

The profile likelihood approach is another widely used method for estimating regression parameters in semiparametric models.[4-13] Its basic idea is to replace the unknown density function by its nonparametric (kernel) estimate for the given parametric components. That is, we treat the error density as an unknown nonparametric function and estimate it via kernel smoothing. Based on the estimated error density, we propose the profile likelihood ratio test (PLRT) for the hypothesis testing problem of parameters of interest in the multiple linear regression models. Our proposed PLRT performs well without the specification of the density function of the error term. It performs satisfactorily even when the second or higher order moments of the error term do not exist (eg, Cauchy distribution).

The rest of this article is organized as follows. In Section 2, we introduce the PLRT. The Wilks theorems for simple and composite null hypotheses will be shown and the power functions will be examined. We conduct simulation studies to assess the performance of the proposed PLRT in Section 3. We illustrate the PLRT via two real datasets about mammalian eye gene expression and concrete strength in Section 4. Some concluding remarks will be presented in Section 5. All proofs and some simulation results are deferred to Supplementary Materials.

2 | THE PLRT

2.1 | The simple null hypotheses

Define $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^T$. Suppose $(y_i, \mathbf{X}_i)_{i=1}^n$ are n independent samples from the following linear regression models:

$$y_i = \mathbf{X}_i^T \beta + \epsilon_i, \quad i = 1, \dots, n,$$

where $\beta = (\beta_1, \dots, \beta_p)^T$. We assume $0 < Var(\mathbf{X}_1) < \infty$ and $\mathbf{X}_i$ is independent of $\epsilon_i (i = 1, \dots, n)$. Denote $\epsilon_i(\beta) = y_i - \sum_{j=1}^p x_{ij}\beta_j$. Let $f_{\epsilon(\beta)}$ and f_0 be the probability density function of $\epsilon_1(\beta)$ and the true density of ϵ_1, respectively. We have $f_{\epsilon(\beta^*)} = f_0$.

Consider the following two-sided hypotheses:

$$H_0 : \beta = \beta^* \quad \text{vs} \quad H_1 : \beta \neq \beta^*, \tag{1}$$

where $\beta^* = (\beta_1^*, \dots, \beta_p^*)^T$. Let $\mathbf{y} = (y_1, \dots, y_n)$ and $\tilde{\mathbf{X}} = \{\mathbf{X}_i, i = 1, \dots, n\}$. Ideally, if we know the true density function of ϵ (ie, $f_0(\epsilon)$), we can use the following LRT for testing the above hypotheses

$$\lambda(\mathbf{y}, \tilde{\mathbf{X}}) = \frac{\prod_{i=1}^n f_0\{\epsilon_i(\hat{\beta})\}}{\prod_{i=1}^n f_0\{\epsilon_i(\beta^*)\}},$$

where $\hat{\beta}$ is the maximum likelihood estimate of β based on the true density function. Under H_0, $2\log\lambda(\mathbf{y}, \tilde{\mathbf{X}})$ converges in distribution to the chi-square distribution with degrees of freedom p as $n \to \infty$, which is the well-known Wilks phenomenon/theorem. At significance level α with $0 < \alpha < 1$, we reject H_0 if $2\log\lambda(\mathbf{y}, \tilde{\mathbf{X}}) \geq \chi^2_{1-\alpha}(p)$.

In practice, ones seldom know the true density of ϵ. Given β, the density function of $\epsilon(\beta)$ can be readily estimated by

$$\hat{f}_{\epsilon(\beta)(-i)}(u) = \frac{1}{nh}\sum_{l \neq i} K\left\{\frac{\epsilon_l(\beta) - u}{h}\right\},$$

where $K(\cdot)$ is a scalar kernel function. The bandwidth parameter h can be selected by the following maximum-estimated-likelihood cross-validation method

$$\hat{h} = \arg\max_{h>0} \sum_{i=1}^n \log\left\{\frac{1}{nh}\sum_{k \neq i} K\left(\frac{\epsilon_k(\hat{\beta}) - \epsilon_i(\hat{\beta})}{h}\right)\right\},$$

where $\hat{\beta}$ is any consistent estimate of β.

The PLRT statistic for hypotheses (1) can then be defined by

$$\lambda_P(\mathbf{y}, \tilde{\mathbf{X}}) = \frac{\prod_{i=1}^n \hat{f}_{\epsilon(\hat{\beta}^{\text{MPL}})(-i)}\{\epsilon_i(\hat{\beta}^{\text{MPL}})\}}{\prod_{i=1}^n \hat{f}_{\epsilon(\beta^*)(-i)}\{\epsilon_i(\beta^*)\}}, \tag{2}$$

where

$$\hat{\beta}^{\text{MPL}} = \arg\max_{\beta} \frac{1}{n}\sum_{i=1}^n \log \hat{f}_{\epsilon(\beta)(-i)}\{\epsilon_i(\beta)\}.$$

Remark 1. Let a be any constant. Since

$$\hat{f}_{\epsilon(\hat{\beta}^{\mathrm{MPL}})(-i)}(\epsilon_i(\hat{\beta}^{\mathrm{MPL}})) = \frac{1}{nh}\sum_{l\neq i} K\left\{\frac{\epsilon_l(\hat{\beta}^{\mathrm{MPL}}) - \epsilon_i(\hat{\beta}^{\mathrm{MPL}})}{h}\right\}$$
$$= \frac{1}{nh}\sum_{l\neq i} K\left\{\frac{\{\epsilon_l(\hat{\beta}^{\mathrm{MPL}}) + a\} - \{\epsilon_i(\hat{\beta}^{\mathrm{MPL}}) + a\}}{h}\right\}$$

and

$$\hat{f}_{\epsilon(\beta^*)(-i)}(\epsilon_i(\beta^*)) = \frac{1}{nh}\sum_{l\neq i} K\left\{\frac{\epsilon_l(\beta^*) - \epsilon_i(\beta^*)}{h}\right\}$$
$$= \frac{1}{nh}\sum_{l\neq i} K\left\{\frac{\{\epsilon_l(\beta^*) + a\} - \{\epsilon_i(\beta^*) + a\}}{h}\right\},$$

location shift of the error term has no effect on the PLRT. That is, $E(\epsilon|X) = 0$ is not required for our PLRT. The PLRT performs based on the estimated density function of the error and thus $D(\epsilon|X) < \infty$ is not required for our proposed PLRT. It is expected that the PLRT works well for, for example, Cauchy error.

Remark 2. The simulation studies in Supplementary Materials show that our proposed PLRT is robust to the choice of bandwidth. In practice, we can use the maximum-estimated-likelihood cross-validation method to get the bandwidth parameter.

Let $\chi^2(p)$ be the chi-square distribution with p degrees of freedom. The following theorem reports the Wilks phenomenon of the PLRT.

Theorem 1. *Suppose Conditions (1) to (7) in Appendix hold. Under H_0, we have*

$$2\log\lambda_P(\mathbf{y}, \tilde{\mathbf{X}}) \xrightarrow{L} \chi^2(p).$$

Given any significance level α with $0 < \alpha < 1$, we can therefore reject H_0 if $2\log\lambda_P(\mathbf{y}, \tilde{\mathbf{X}}) \geq \chi^2_{1-\alpha}(p)$. For PLRT in (2), we use the MPL estimator $\hat{\beta}^{\mathrm{MPL}}$. The usage of the MPL estimator is crucial in order to enjoy the Wilks phenomenon for PLRT. For example, this phenomenon may not hold when the ordinary least squares estimator is used. The kernel function used does not affect the Wilks phenomenon of PLRT. Moreover, the simulation studies in Supplementary Materials show that the performance of our PLRT is robust to the choice of kernel function. In practice, the kernel function may be selected by the leave-one-out cross-validation method. For details, see Supplementary Materials.

We next consider the power function of the proposed PLRT. By Theorem 1 above and Theorem 2 in Chen et al,[14] the power function of PLRT (denoted as $\phi_P(\beta; n)$) can be shown to be

$$P_\beta\{2\log\lambda_P(\mathbf{y}, \tilde{\mathbf{X}}) \geq \chi^2_{1-\alpha}(p)\} = P_\beta\left\{2\log\frac{\prod_{i=1}^n \hat{f}_{\epsilon(\hat{\beta}^{\mathrm{MPL}})(-i)}\{\epsilon_i(\hat{\beta}^{\mathrm{MPL}})\}}{\prod_{i=1}^n \hat{f}_{\epsilon(\beta^*)(-i)}\{\epsilon_i(\beta^*)\}} \geq \chi^2_{1-\alpha}(p)\right\}$$
$$= P_\beta\left[2\sum_{i=1}^n \log\hat{f}_{\epsilon(\hat{\beta}^{\mathrm{MPL}})(-i)}\{\epsilon_i(\hat{\beta}^{\mathrm{MPL}})\} - 2\sum_{i=1}^n \log\hat{f}_{\epsilon(\beta)(-i)}\{\epsilon_i(\beta)\} \geq \chi^2_{1-\alpha}(p) - 2nM(\beta, \beta^*) + O_p(n^{\frac{1}{2}})\right]$$
$$= 1 - F(\chi^2_{1-\alpha}(p) - 2nM(\beta, \beta^*); p)(1 + o_p(1)),$$

where $\chi^2_{1-\alpha}(p)$ is the $1-\alpha$ quantile of $\chi^2(p)$, $F(x; p)$ is the cumulative distribution function of $\chi^2(p)$ and $M(\beta, \beta^*) := \int f_{\epsilon(\beta)}(v)\log f_{\epsilon(\beta)}(v)dv - \int f_{\epsilon(\beta^*)}(v)\log f_{\epsilon(\beta^*)}(v)dv$. Since $M(\beta, \beta^*)$ is positive when the true value of the regression coefficient is β, by Lemma 1 of Chen et al,[14] we have

$$\phi_P(\beta; n) \to 1 \text{ as } n \to \infty,$$

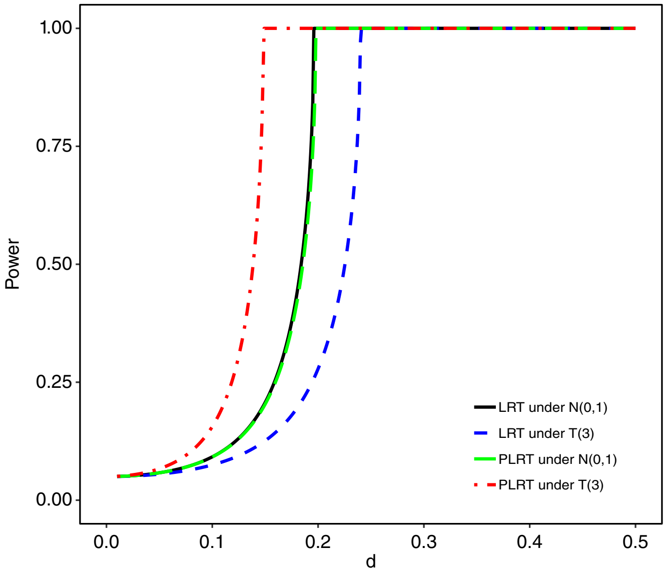

FIGURE 1 The power as functions of $d = |\beta - \beta^*|$ for PLRT and LRT. LRT, likelihood ratio test; PLRT, profile likelihood ratio test [Colour figure can be viewed at wileyonlinelibrary.com]

for $\beta \neq \beta^*$. This result indicates that the PLRT can achieve high power provided that the sample size is large enough.

On the other hand, the power function of the LRT (denoted as $\phi(\beta; n)$) is given by

$$P_\beta\{2\log\lambda(\mathbf{y}, \tilde{\mathbf{X}}) \geq \chi^2_{1-\alpha}(p)\} = 1 - F(\chi^2_{1-\alpha}(p) - 2nM_1(\beta, \beta^*); p)(1 + o_p(1)),$$

where $M_1(\beta, \beta^*) := \int f_{\epsilon(\beta)}(v)\log f_{\epsilon(\beta^*)}(v)dv - \int f_{\epsilon(\beta^*)}(v)\log f_{\epsilon(\beta^*)}(v)dv$. We first compare the power functions of the PLRT and LRT when the error terms follow the normal distribution. Without loss of generality, we assume that β is one-dimensional, $X_1 \sim N(0, \sigma_1^2)$ and $\epsilon_1 \sim N(0, \sigma_2^2)$. The power functions of the PLRT and LRT can be, respectively, shown to be

$$\phi_P(\beta) = 1 - F\left(\chi^2_{1-\alpha}(1) - n\log\left\{1 + \frac{\sigma_1^2(\beta - \beta^*)^2}{\sigma_2^2}\right\}; 1\right)(1 + o_p(1)), \quad \text{and}$$
$$\phi(\beta) = 1 - F\left(\chi^2_{1-\alpha}(1) - \frac{n\sigma_1^2(\beta - \beta^*)^2}{\sigma_2^2}; 1\right)(1 + o_p(1)).$$

Since $\log\left[\{\sigma_1^2(\beta - \beta^*)^2\}/\sigma_2^2 + 1\right] - \{\sigma_1^2(\beta - \beta^*)^2\}/\sigma_2^2 < 0$, the LRT is always more powerful than the PLRT for $\beta \neq \beta^*$ and the error following the normal distribution. When $|\beta - \beta^*| < \epsilon$ with ϵ being small, the LRT and PLRT have similar powers. To illustrate the effect of $d = |\beta - \beta^*|$ on the power functions of the LRT and PLRT, we plot their corresponding power functions in Figure 1 with $n = 100$ and $\alpha = 0.05$ when both X_1 and ϵ_1 follow (a) $N(0, 1)$; and (b) Student t distribution with 3 degrees of freedom. Under normal distribution, the power functions of the LRT and PLRT perform similarly. However, the PLRT obviously outperforms the LRT (based on normal distribution) when the error actually follows nonnormal distribution (T-distribution). Detailed calculations can be found in Supplementary Materials.

2.2 | The composite null hypotheses

Let β^* be the true parameter of β $(p \times 1)$ and $\beta^* \in \text{Interior}(\Theta)$ with Θ being a compact subset of R^p. Consider the following hypothesis testing problem:

$$H_0^C : \beta \in \Theta_0 \quad \text{vs} \quad H_1^C : \beta \in \Theta \setminus \Theta_0. \tag{3}$$

We assume that there exists a bijection $\mathbf{g} = (g_1, \ldots, g_p)$ between Ψ_0 and Θ_0 such that $\Theta_0 = \{\beta : \beta = \mathbf{g}(\varphi)$ with $\varphi \in \Psi_0\}$, where Ψ_0 is a r-dimensional subspace of R^r $(r < p)$ and $g_i(i = 1, \ldots, p)$ is second differentiable with support Ψ_0.

Under H_0^C, the profile likelihood estimate of φ can be obtained by

$$\hat{\varphi}_0^{\mathrm{MPL}} = \arg\max_{\varphi \in \Psi_0} \frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_{\epsilon\{\mathbf{g}(\varphi)\}(-i)}[\epsilon_i\{\mathbf{g}(\varphi)\}],$$

where $\epsilon_i\{\mathbf{g}(\varphi)\} = y_i - \mathbf{X}_i^T \mathbf{g}(\varphi)$. The profile likelihood estimate of β can be written as $\hat{\beta}_0^{\mathrm{MPL}} = \mathbf{g}(\hat{\varphi}_0^{\mathrm{MPL}})$ under H_0^C. We can then define the PLRT statistic for (3) as

$$\lambda_{P_C}(\mathbf{y}, \tilde{\mathbf{X}}) = \frac{\prod_{i=1}^{n} \hat{f}_{\epsilon(\hat{\beta}^{\mathrm{MPL}})(-i)}\{\epsilon_i(\hat{\beta}^{\mathrm{MPL}})\}}{\prod_{i=1}^{n} \hat{f}_{\epsilon(\hat{\beta}_0^{\mathrm{MPL}})(-i)}\{\epsilon_i(\hat{\beta}_0^{\mathrm{MPL}})\}}.$$

The following theorem shows that the PLRT also enjoys the Wilks phenomenon under the composite null hypotheses.

Theorem 2. *Suppose Conditions (1) to (7) in Appendix hold. Under H_0^C, we have*

$$2 \log \lambda_{P_C}(\mathbf{y}, \tilde{\mathbf{X}}) \xrightarrow{L} \chi^2(p - r).$$

3 | NUMERICAL STUDIES

In this section, we conduct several simulation studies to examine the type I error rate and power performance of our proposed PLRT, the LRT with error following $N(0, \sigma^2)$, the empirical likelihood ratio test (ELRT),[15] and the weighted profile likelihood ratio test (WPLRT) proposed by Huang.[16] Our proposed PLRT and the WPLRT are implemented based on the Gaussian kernel function (ie, $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$).

3.1 | Study 1

We first consider the following simple linear regression model

$$y_i = x_i \beta + \epsilon_i, \quad i = 1, \dots, n.$$

Here, we assume $x_i \sim_{i.i.d.} N(0, 4^2)$ and the error (ie, ϵ_i) follows (a) $N(0, 3)$; (b) the mixture of normal variates: $0.9N(0,1) + 0.1N(0, 10^2)$; (c) the Student t distribution with 3 degrees of freedom; (d) the chi-squared distribution with 3 degrees of freedom; and (e) the standard Cauchy distribution with $n = 100$ or 200.

We consider the following simple hypothesis testing problem

$$H_0^{S1} : \beta = 3 \quad \text{vs} \quad H_1^{S1} : \beta \neq 3$$

at the significance level of α being 0.05.

Data are generated with $\beta = 2.5 + 0.01d$, $d = 0, \dots, 100$. For all β's, the number of replications is 1000 for each sample size and each error distribution. Figure 2 reports the empirical powers for β's based on 1000 replications for the four tests.

3.2 | Study 2

We consider the following multiple linear regression model

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \epsilon_i, \quad i = 1, \dots, n.$$

Here, (x_1, x_2, x_3) follows the multivariate normal distribution $N(0, \Sigma_x)$ with $(\Sigma_x)_{i,j} = 0.5^{|i-j|}$ for $1 \leq i, j \leq 3$ and the error (ie, ϵ_i) follows those distributions considered in Study 1 with $n = 100$ or 200.

We consider the following simple hypothesis testing problem

$$H_0^{S2} : \beta_1 = 3 \quad \text{vs} \quad H_1^{S2} : \beta_1 \neq 3$$

at the significance level of α being 0.05.

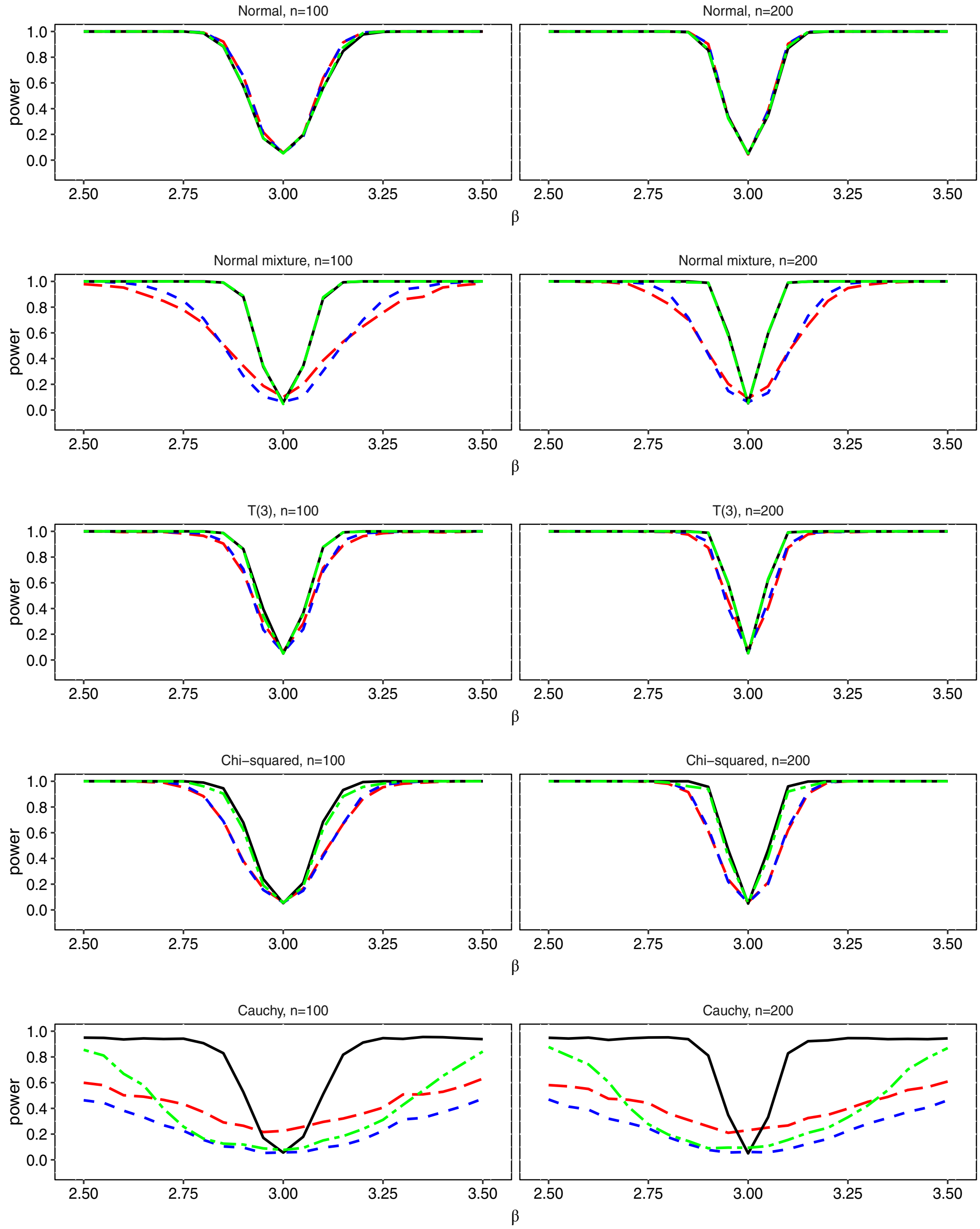

FIGURE 2 Empirical power functions for different errors in Study 1. ......: LRT based on normal distribution ; - - -: ELRT;[15] -.-.-: WPLRT;[16] —: PLRT. ELRT, empirical likelihood ratio test; LRT, likelihood ratio test; PLRT, profile likelihood ratio test; WPLRT, weighted profile likelihood ratio test [Colour figure can be viewed at wileyonlinelibrary.com]

Data are generated with $\beta_1 = 1 + 0.01d, d = 0, \dots, 400$ and $(\beta_2, \beta_3) = (1.5, 2)$. For all β_1's, the number of replications is 1000 for each sample size and each error distribution. Figure 3 reports the empirical powers for β_1's based on 1000 replications for the four tests.

3.3 | Study 3

We adopt the model in Study 2 and consider the following hypothesis testing problem

$$H_0^{S3} : \beta_1 = 3, \beta_2 = 1.5 \quad \text{vs} \quad H_1^{S3} : \beta_1 \neq 3, \beta_2 \neq 1.5$$

at the significance level of α being 0.05.

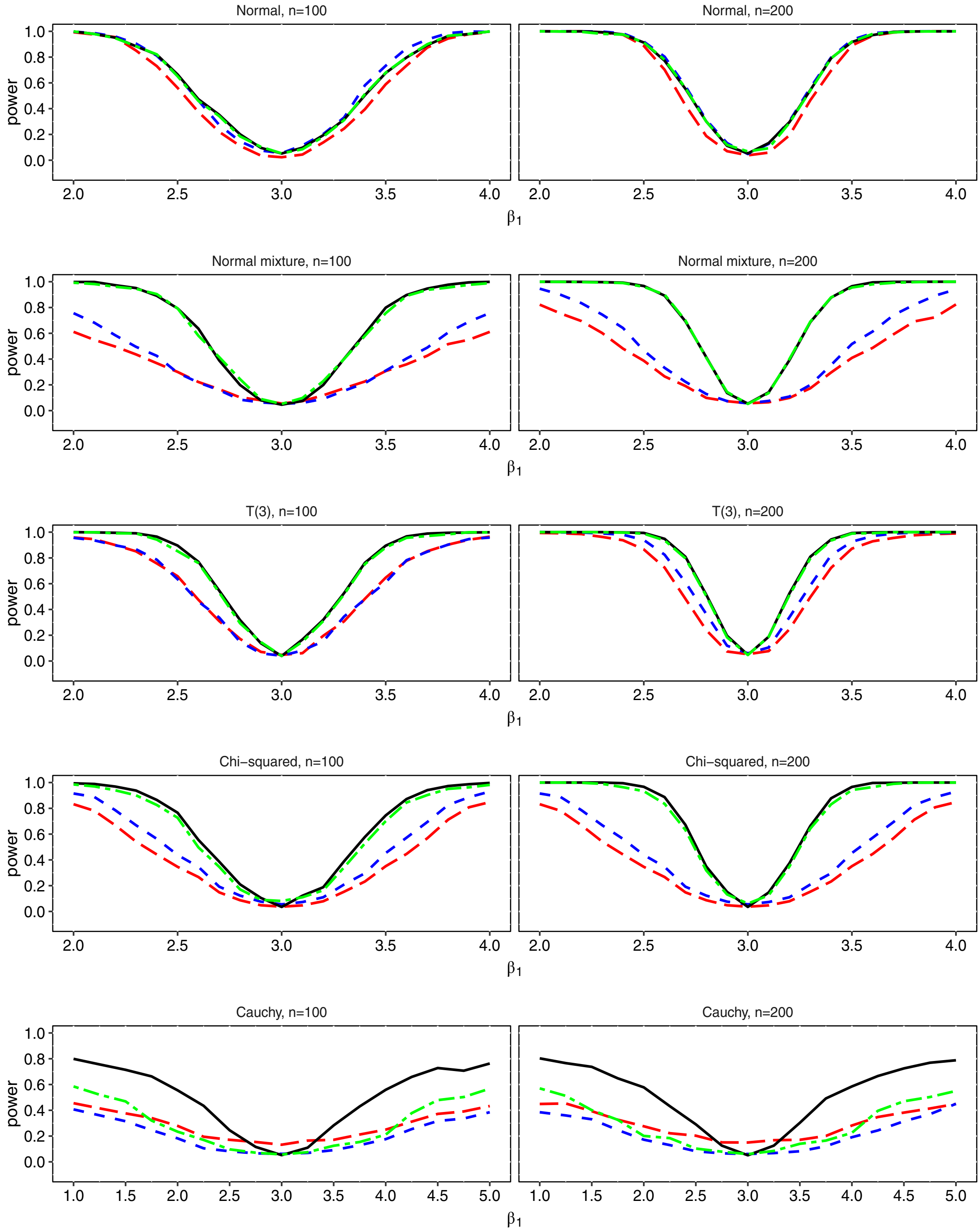

FIGURE 3 Empirical power functions for different errors in Study 2. ------: LRT based on normal distribution ; ----: ELRT;[15] -·-·-: WPLRT;[16] —: PLRT. ELRT, empirical likelihood ratio test; LRT, likelihood ratio test; PLRT, profile likelihood ratio test; WPLRT, weighted profile likelihood ratio test [Colour figure can be viewed at wileyonlinelibrary.com]

To examine the type I error rate behavior, we generate data with $\beta_1 = 3$, $\beta_2 = 1.5$ and $\beta_3 = 2$. The number of replications is 1000 for each sample size and each error distribution. To examine the power performance, we generate randomly five hundred (β_1, β_2)'s with $\beta_1 \sim U(2, 4)$ and $\beta_2 \sim U(0.5, 2.5)$. We similarly calculate the rejection probability (empirical power) at each of the five hundred (β_1, β_2)'s, and then take the average. Table 1 reports the empirical type I error rates and powers based on 1000 replications for all the four tests.

Method	Type I error n = 100	Type I error n = 200	Power n = 100	Power n = 200
$N(0, 3)$				
PLRT	0.051	0.052	0.796	0.904
WPLRT	0.052	0.055	0.799	0.901
ELRT	0.044	0.040	0.791	0.887
LRT	0.048	0.051	0.818	0.904
$0.9N(0, 1) + 0.1N(0, 10^2)$				
PLRT	0.049	0.050	0.850	0.938
WPLRT	0.051	0.047	0.844	0.934
ELRT	0.141	0.099	0.536	0.646
LRT	0.058	0.054	0.478	0.689
T distribution $df = 3$				
PLRT	0.050	0.054	0.860	0.940
WPLRT	0.051	0.054	0.859	0.943
ELRT	0.063	0.056	0.787	0.894
LRT	0.061	0.056	0.803	0.911
Chi-squared $df = 3$				
PLRT	0.052	0.047	0.817	0.936
WPLRT	0.057	0.055	0.781	0.918
ELRT	0.062	0.054	0.598	0.767
LRT	0.061	0.063	0.632	0.803
Standard Cauchy				
PLRT	0.052	0.053	0.593	0.640
WPLRT	0.064	0.061	0.328	0.335
ELRT	0.292	0.306	0.394	0.395
LRT	0.067	0.070	0.132	0.126

Abbreviations: ELRT, empirical likelihood ratio test; LRT, likelihood ratio test; PLRT, profile likelihood ratio test; WPLRT, weighted profile likelihood ratio test.

TABLE 1 The empirical type I errors and powers in Study 3

3.4 | Study 4

We adopt the model in Study 2 and consider the following hypothesis testing problem

$$H_0^{S4} : \beta_1 = 3, \beta_2 = 1.5, \beta_3 = 2 \quad \text{vs} \quad H_1^{S4} : \beta_1 \neq 3, \beta_2 \neq 1.5, \beta_3 \neq 2$$

at the significance level of α being 0.05.

To examine the type I error rate behavior, we generate data with $\beta_1 = 3$, $\beta_2 = 1.5$, and $\beta_3 = 2$. The number of replications is 1000 for each sample size and each error distribution. To examine the power performance, we generate randomly five hundred $(\beta_1, \beta_2, \beta_3)$'s with $\beta_1 \sim U(2, 4)$, $\beta_2 \sim U(0.5, 2.5)$, and $\beta_3 \sim U(1, 3)$. We calculate the empirical rejection probability (empirical power) at each of the five hundred $(\beta_1, \beta_2, \beta_3)$'s, and then take the average. Table 2 reports the empirical type I error rates and powers based on 1000 replications for the four tests.

We have the following observations based on the four studies:

(i) For normally distributed errors, the resultant empirical type I errors, and powers of the four tests are very close;

TABLE 2 The empirical type I errors and powers in Study 4

Method	Type I error n = 100	Type I error n = 200	Power n = 100	Power n = 200
$N(0,3)$				
PLRT	0.050	0.054	0.916	0.952
WPLRT	0.051	0.055	0.917	0.950
ELRT	0.074	0.071	0.915	0.961
LRT	0.050	0.051	0.921	0.961
$0.9N(0,1)+0.1N(0,10^2)$				
PLRT	0.050	0.055	0.951	0.981
WPLRT	0.052	0.056	0.950	0.971
ELRT	0.205	0.156	0.760	0.825
LRT	0.056	0.058	0.627	0.810
T distribution $df=3$				
PLRT	0.049	0.048	0.960	0.978
WPLRT	0.048	0.052	0.950	0.975
ELRT	0.107	0.080	0.938	0.968
LRT	0.059	0.058	0.924	0.964
Chi-squared $df=3$				
PLRT	0.049	0.048	0.935	0.980
WPLRT	0.058	0.056	0.911	0.968
ELRT	0.110	0.086	0.820	0.932
LRT	0.070	0.065	0.799	0.935
Standard Cauchy				
PLRT	0.054	0.052	0.587	0.611
WPLRT	0.061	0.062	0.462	0.475
ELRT	0.259	0.327	0.549	0.587
LRT	0.060	0.060	0.160	0.157

Abbreviations: ELRT, empirical likelihood ratio test; LRT, likelihood ratio test; PLRT, profile likelihood ratio test; WPLRT, weighted profile likelihood ratio test.

(ii) For nonnormally distributed errors, the PLRT generally controls its type I error rates better than the other methods in the sense that the PLRT yields empirical type I error rates closer to the preassigned nominal level (ie, 0.05) than the other methods;

(iii) The PLRT is generally more powerful than the other methods, especially when the errors deviate from the normal distribution (eg, the mixture of normal variates: $0.9N(0,1)+0.1N(0,10^2)$), the Student t distribution with 3 degrees of freedom, the chi-squared distribution with 3 degrees of freedom and the standard Cauchy distribution). When the error density is wrongly specified, the maximum likelihood estimators based on normal error may be inefficient or inconsistent. Thus, the resultant LRT based on normal distribution can lose power.[17] PLRT is based on the estimated error density while ELRT is based on mean of the error. In other words, PLRT adopts more information and should be more powerful than ELRT. Moreover, the WPLRT relies on mean of the error but our proposed PLRT is invariant to location shift of errors, thus the PLRT performs better than WPLRT for the chi-squared error;

(iv) Since LRT, ELRT, and WPLRT all assume the existence of the error mean, the three tests can lose power, for example, for Cauchy error. The PLRT is based on the estimated error density and thus performs well even when moments (eg, mean and variance) do not exist (eg, Cauchy distribution).

In conclusion, our proposed PLRT has better performance than the other three methods.

3.5 | Study 5

We consider the multiple testing problem in this study. Consider the following linear regression model:

$$y_i = \sum_{j=1}^{m} x_{ij}\beta_j + \epsilon_i, \quad i = 1, \dots, n.$$

We have simultaneous null hypotheses H_j : $\beta_j = 0$ $(j = 1, \dots, m)$ to be tested and control the false discovery rate (FDR) at level α. The Benjamini-Hochberg procedure[18] based on PLRT can be summarized as follows:

Step 1. For H_j, we construct the PRLT at the significance level of α and compute the p-value as p_j for $j=1, \dots, m$.

Step 2. We arrange these p-values in ascending order and denote them as $p_{(1)}, \dots, p_{(m)}$. Their corresponding null hypotheses are $H_{(1)}, \dots, H_{(m)}$.

Step 3. Find the largest k such that $p_{(k)} \leq \frac{k}{m}\alpha$.

Step 4. Reject the null hypotheses for all $H_{(j)}$ for $j=1, \dots, k$.

Set $m=5$. Let $(x_1, x_2, x_3, x_4, x_5)^T$ follow the multivariate normal distribution $N(0, \Sigma_x)$ with $(\Sigma_x)_{i,j} = 0.5^{|i-j|}$ for $1 \leq i, j \leq 5$ and the errors follow the distributions considered in Study 1 with $n=100$ or 200. Data are generated with $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = (3, 1.5, 2, 2.5, 3.5)$. We consider the following simultaneous multiple hypothesis testing problem

$$H_0^{S5} : \beta_i = 0 \quad \text{vs} \quad H_1^{S5} : \beta_i \neq 0, \quad i = 1, \dots, 5.$$

with FDR at level $\alpha = 0.05$. The number of replications is 1000 for each sample size and each error distribution. The Benjamini-Hochberg procedure based on LRT, ELRT, and WPLRT can be conducted similarly.

Table 3 reports the average number of nonzero coefficients by the four tests based on 1000 replications. It shows that our proposed PLRT has higher probability of correctly selecting the correct model than the other three methods for the multiple testing problem under consideration. It is noted that LRT, ELRT, and WPLRT break down for error with infinite variance (eg, Cauchy distribution).

4 | REAL DATA ANALYSIS

In this section, we apply our proposed methodology to analyze two datasets, namely, the mammalian eye gene expression dataset and concrete compressive strength dataset. For each dataset, we have more than one hypotheses simultaneously and thus use the controlling FDR method[18] in multiple testing to control the type I error rates for our PLRT and the other methods.

4.1 | The mammalian eye gene expression dataset

In this subsection, we apply our proposed methodologies to analyze an example regarding the gene expression and regulation in the mammalian eye.[19] Briefly, Bardet-Biedl syndrome (BBS) is a genetically heterogeneous disease characterized primarily by retinal dystrophy, obesity, polydactyly, renal malformations and learning disabilities.[20-22] The current prevalence of BBS in Newfoundland is approximately 1 in 18 000.[23] Beales et al[20] found that the progress of BBS is rapid, with a mean of 7 years from diagnosis to blindness. BBS expression varies and its diagnosis is often difficult. TRIM32 is recently found to cause BBS, that is, the mutation (P130S) in TRIM32 gives rise to BBS. For this reason, TRIM32 is also sometimes called BBS11.[22] The regulatory mechanisms shared by related genes would likely cause their expression to respond to biological variations in a coordinated fashion.[22] Hence, we are interested in finding those genes whose expression are correlated with TRIM32.

The dataset includes 120 subjects with 18 976 genes being observed for each subject. We conduct an initial screening for the dataset, that is, we select 200 genes with top values of gene-expression variances.[24] We then use our PLRT to test whether each of the 200 gene has a significant impact on TRIM32. The null and alternative hypotheses are $\beta_i = 0$ and $\beta_i \neq 0$, respectively, for $i=1, \dots, 200$. Our method suggests that 81 genes are correlated to TRIM32 at the FDR of 0.00001.[18,25] It is noteworthy that the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, https://string-db.org/) suggests that 73 of the 81 selected genes are correlated to TRIM32, and our test further suggests

TABLE 3 The average number of nonzero coefficients in Study 5

Method	Number of nonzero n = 100	n = 200
$N(0, 3)$		
PLRT	5.000	5.000
WPLRT	5.000	5.000
ELRT	5.000	5.000
LRT	5.000	5.000
$0.9N(0, 1) + 0.1N(0, 10^2)$		
PLRT	5.000	5.000
WPLRT	4.999	5.000
ELRT	4.511	4.895
LRT	4.904	4.993
T distribution $df = 3$		
PLRT	5.000	5.000
WPLRT	5.000	5.000
ELRT	4.951	4.989
LRT	4.986	4.995
Chi-squared $df = 3$		
PLRT	4.999	5.000
WPLRT	4.998	4.999
ELRT	4.912	4.996
LRT	4.994	4.998
Standard Cauchy		
PLRT	4.999	5.000
WPLRT	3.876	3.889
ELRT	1.921	2.005
LRT	1.936	1.229

Abbreviations: ELRT, empirical likelihood ratio test; LRT, likelihood ratio test; PLRT, profile likelihood ratio test; WPLRT, weighted profile likelihood ratio test.

that genes "LOC102546420," "Cnpy1," "Gosr1," "Krtap14l," "Fabp12," "Serpine3," "Fam120c," "Ramac" may also be correlated with gene TRIM32.

4.2 | The concrete compressive strength dataset

This dataset contains 1030 observations. We are interested in the relationships between the concrete compressive strength (Y) and the following covariates: Cement (X_1), Blast Furnace Slag (X_2), Fly Ash (X_3), Water (X_4), Superplasticizer (X_5), Coarse Aggregate (X_6), Fine Aggregate (X_7), Age (X_8). We use the PLRT, LRT with error following $N(0, \sigma^2)$, ELRT and WPLRT to test whether each covariate has a significant impact on concrete compressive strength. Specifically, the null and alternative hypotheses are $\beta_i = 0$ and $\beta_i \neq 0$, respectively, for $i = 1, \dots, 8$.

Table 4 reports the testing results at the FDR 0.05. We also use the adaptively penalized maximum profile likelihood (AMPL) method[14] to estimate the coefficients. The ELRT[15] suggests that water is not significant for the concrete compressive strength. However, Swanmy[26] and Yeh[27] suggested that water is very important. Both LRT and WPLRT conclude that all covariates are significant at FDR being 0.05, while our proposed PLRT suggests that Superplasticizer,

TABLE 4 The results for analyzing the concrete compressive strength dataset

Covariate	PLRT	WPLRT	ELRT	LRT	Coeff
Cement	+	+	+	+	0.113
Blast Furnace Slag	+	+	+	+	0.096
Fly Ash	+	+	+	+	0.079
Water	+	+	−	+	−0.245
Superplasticizer	−	+	−	+	0.000
Coarse Aggregate	−	+	−	+	0.000
Fine Aggregate	−	+	−	+	0.000
Age	+	+	+	+	0.114

Note: "+" means this covariate has a significant impact on concrete compressive strength, otherwise "−." "coeff" represents the estimated coefficient by AMPL.
Abbreviations: AMPL, adaptively penalized maximum profile likelihood; ELRT, empirical likelihood ratio test; LRT, likelihood ratio test; PLRT, profile likelihood ratio test; WPLRT, weighted profile likelihood ratio test.

Coarse Aggregate and Fine Aggregate are not significant, which are consistent with the results by the AMPL method. Moreover, the conclusions made by PLRT are consistent with those of Swanmy[26] and Yeh.[27]

5 | DISCUSSION

We develop a so-called PLRT for testing the coefficients in the multiple linear regression model. Unlike the existing LRT, our proposed PLRT does not need to specify the error distribution. Our simulation results support that it performs better than the LRT in terms of power when the error distribution for LRT is wrongly specified, and also outperforms the ELRT and WPLRT. The PLRT can be applied to the situations that involve the simultaneous testing of multiple hypotheses, see the simulation studies and real data analysis. We observed in the simulation studies that the PLRT performs much better than the WPLRT for Cauchy error. However, the WPLRT can be substantially improved by replacing the l_2 constraint with the l_1 constraint for heavy-tailed errors (eg, Cauchy distribution), which is worth future investigation. Moreover, it is of great interest to pursue the parameter hypothesis testing problem in the framework of profile likelihood for longitudinal data[11] in the future.

The simulation studies in the Supplementary Materials show that our proposed PLRT performs very well when the error is no longer additive. We are interested in investigating the theoretical properties and the comprehensive numerical performances of the PLRT under the nonadditive model framework in the future.

ACKNOWLEDGEMENTS

The authors thank the Editor, Associate Editor, and the three anonymous reviewers for helpful comments and suggestions, which have led to great improvements of this article. The work of Feifei Yan was partially supported through the Innovation Program of Central South University 2018zzts102; the work of Man-Lai Tang was partially supported through grants from the Research Grant Council of the Hong Kong Special Administrative Region (UGC/FDS14/P01/16, UGC/FDS14/P02/18 and the Research Matching Grant Scheme (RMGS)) and a grant from the National Natural Science Foundation of China 11871124; The computing facilities/software were supported from SAS Viya and the Big Data Intelligence Centre at Hang Seng University of Hong Kong; and the work of Ziqi Chen was partially supported through NSFC grant 11871477.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The mammalian eye gene expression dataset is available via the link http://myweb.uiowa.edu/pbreheny/data/Scheetz2006.html;[19] The concrete compressive strength dataset is available from the UCI Machine Learning Repository via the link http://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength.[28]

ORCID

Man-Lai Tang https://orcid.org/0000-0003-3934-2676
Ziqi Chen https://orcid.org/0000-0002-4128-2986

REFERENCES

1. Wilks S. The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Stat*. 1938;9(1):60-62.
2. Edwards A. The history of likelihood. *Int Stat Rev*. 1974;42(1):9-15.
3. Peng L, Qi Y, Wang R. Empirical likelihood test for high dimensional linear models. *Statist Probab Lett*. 2014;86(1):85-90.
4. Severini T, Wong W. Profile likelihood and conditionally parametric models. *Ann Stat*. 1992;20(4):1768-1802.
5. Fan J, Huang T. Profile likelihood inferences on semiparametric varying coefficient partially linear models. *Bernoulli*. 2005;11(6):1031-1057.
6. Linton O, Sperlich S, Keilegom I. Estimation of a semiparametric transformation model. *Ann Stat*. 2008;36(2):686-718.
7. Lombardia M, Sperlich S. Semiparametric inferencein generalized mixed effects models. *J R Stat Soc Ser B Stat Methodol*. 2008;70(5):913-930.
8. Lee A, Hirose Y. Semi-parametric efficiency bounds for regression models under response-selective sampling: the profile likelihood approach. *Ann Inst Stat Math*. 2010;62(6):1023-1052.
9. Lin L, Liu L, Cui X. A simple and efficient profile likelihood for semiparametric exponential family; 2017. https://arxiv.org/abs/1711.11426. Submitted November 30, 2017.
10. Rice J, Tsodikov A. Semiparametric profile likelihood estimation for continuous outcomes with excess zeros in a random-threshold damage-resistance model. *Stat Med*. 2017;36(12):1924-1935.
11. Chen Z, Tang M, Gao W. A profile likelihood approach for longitudinal data analysis. *Biometrics*. 2018;74(1):220-228.
12. Hirose Y, Liu I. Statistical generalized derivative applied to the profile likelihood estimation in a mixture of semiparametric models; 2018. https://arxiv.org/abs/1807.07670. Submitted July 20, 2018.
13. Sang H, Morikawa K. A profile likelihood approach to semiparametric estimation with nonignorable nonresponse; 2018. https://arxiv.org/abs/1809.03645. Submitted September 11, 2018.
14. Chen Z, Tang M, Gao W, Shi N. New robust variable selection methods for linear regression models. *Scand J Stat*. 2014;41(3):725-741.
15. Owen A. Empirical likelihood for linear models. *Ann Stat*. 1991;19(4):1725-1747.
16. Huang A. Density estimation and nonparametric inferences using maximum likelihood weighted kernels. *J Nonparametr Stat*. 2013;25(3):561-571.
17. Foutz R, Srivastava R. The performance of the likelihood ratio test when the model is incorrect. *Ann Stat*. 1977;5(6):1183-1194.
18. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B Stat Methodol*. 1995;57(1):289-300.
19. Scheetz T, Kim K, Swiderski R, et al. Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proc Natl Acad Sci*. 2006;103(39):14429-14434.
20. Beales P, Elcioglu N, Woolf A, Parker D, Flinter F. New criteria for improved diagnosis of Bardet–Biedl syndrome: results of a population survey. *J Med Genet*. 1999;36(6):437-446.
21. Ansley S, Badano J, Blacque O, et al. Basal body dysfunction is a likely cause of pleiotropic Bardet–Biedl syndrome. *Nature*. 2003;425(6958):628-633.
22. Chiang A, Beck J, Yen H, et al. Homozygosity mapping with SNP arrays identifies TRIM32 , an E3 ubiquitin ligase, as a Bardet–Biedl syndrome gene (BBS11). *Proc Natl Acad Sci*. 2006;103(16):6287-6292.
23. Moore S, Green J, Fan Y, et al. Clinical and genetic epidemiology of Bardet–Biedl syndrome in newfoundland: a 22-year prospective, population-based, cohort study. *J Med Genet Part A*. 2005;132(4):352-360.
24. Huang J, Ma S, Li H, Zhang C. The sparse Laplacian shrinkage estimator for high-dimensional regression. *Ann Stat*. 2005;39(4):2021-2046.
25. Storey J, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci*. 2003;100(16):9440-9445.
26. Swanmy R. Design for durability and strength through the use of fly ash and slag in concrete. Paper presented at: Proceedings of the 3rd CANMET/ACI International Conference CANMET/ACI 1–72; August 24-27, 1997; Auckland, New Zealand.
27. Yeh IC. Analysis of strength of concrete using design of experiments and neural networks. *J Mater Civ Eng*. 2006;18(4):597-604.
28. Yeh IC. Modeling of strength of high performance concrete using artificial neural networks. *Cem Concr Res*. 1998;28(12):1797-1808.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Yan F, Xu Q-S, Tang M-L, Chen Z. Kernel density-based likelihood ratio tests for linear regression models. *Statistics in Medicine*. 2021;40:119–132. https://doi.org/10.1002/sim.8765

APPENDIX

In order to investigate the Theorems in this article, we give the following regular conditions.

(1): $f_{\epsilon(\beta)}(\cdot)$ has a continuous second-order derivative with respect to each u and each $\beta \in \mathcal{N}$, where $\mathcal{N}$ is a neighborhood of β^*. $\beta^* \in$ interior(Θ) with Θ being a compact subset of R^p.

(2): $f_{\epsilon(\beta^*)}(\cdot)$ has support R and is uniformly continuous.

(3): $K(\cdot)$ is uniformly continuous with modulus of continuous W_k, twice continuously differentiable, of bounded variation $V(K)$ and absolutely integrable. $K(\cdot)$ is symmetric, $\int K(u)du = 1$, $\int uK(u)du = 0$ and $\int u^3K(u)du = 0$.

(4): $\int |u \log |u||^{\frac{1}{2}} |dK(u)| < \infty$, $\sup |K(u)| < \infty$, $\sup |K'(u)| < \infty$ and $\int K^2(u)du < \infty$. $\lim_{u\to\infty} |K(u)u| = 0$, $\lim_{u\to\infty} |K(u)u^2| = 0$, $\lim_{u\to\infty} |K'(u)u| = 0$ and $\lim_{u\to\infty} |K'(u)u^2| = 0$.

(5): $E[\sup_{\beta\in\Theta} |\log f_{\epsilon(\beta)}\{\epsilon_1(\beta)\}|] < \infty$, $E[\sup_{\beta\in N} \|\partial[\log f_{\epsilon(\beta)}\{\epsilon_1(\beta)\}]/\partial\beta\|] < \infty$, $E[\sup_{\beta\in N} \|\partial^2[\log f_{\epsilon(\beta)}\{\epsilon_1(\beta)\}]/\partial\beta\partial\beta^T\|] < \infty$.

(6): The density of $\mathbf{X}$, that is, $f_{\mathbf{X}}(\cdot)$ is bounded away from 0 and ∞ and is Lipschitz continuous on its compact support.

(7): $nh^5/(\log n)^2 \to \infty$, $nh^8 \to 0$, and $n^{1-b}h \to \infty$ for some $b > 0$.