



Least product relative error estimation



Kani Chen^a, Yuanyuan Lin^b, Zhanfeng Wang^{c,*}, Zhiliang Ying^d

^a Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China

^b Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T, Hong Kong, China

^c Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China

^d Department of Statistics, Columbia University, New York, NY 10027, United States

ARTICLE INFO

Article history:

Received 9 October 2014

Available online 18 November 2015

AMS 2000 subject classifications:

Primary 62F12

62F03

Keywords:

Multiplicative regression model

Product form

Relative error

Scale invariance

Variance estimation

ABSTRACT

A least product relative error criterion is proposed for multiplicative regression models. It is invariant under scale transformation of the outcome and covariates. In addition, the objective function is smooth and convex, resulting in a simple and uniquely defined estimator of the regression parameter. It is shown that the estimator is asymptotically normal and that the simple plug-in variance estimation is valid. Simulation results confirm that the proposed method performs well. An application to body fat calculation is presented to illustrate the new method.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

In regression analysis, the least squares (LS) and least absolute deviation (LAD) are the most commonly used criteria based on absolute errors [10,8]. In some situations, however, criteria based on relative errors that are scale invariant and less sensitive to outliers are more desirable [5,4,3,13,6,1,14]. Consider the following multiplicative regression model

$$Y_i = \exp(X_i^\top \beta) \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where Y_i is the response variable, X_i is the p -vector of explanatory variables with the first component being 1 (intercept), β is the corresponding p -vector of regression parameters with the first component being the intercept and ϵ_i is the error term, which is strictly positive. An additional constraint on ϵ needs to be imposed so that the first component of β (intercept) becomes identifiable. Model (1) is also known as the accelerated failure time (AFT) model in the survival analysis literature.

For the multiplicative regression model (1), Chen et al. [1] give a convincing argument that a proper criterion should take into account both types of relative errors: one relative to the response and the other relative to the predictor of the response. A criterion with only one type of relative errors often leads to biased estimation. They introduced the least absolute relative error (LARE) estimation for model (1) by minimizing

$$LARE_n(\beta) \equiv \sum_{i=1}^n \left\{ \left| \frac{Y_i - \exp(X_i^\top \beta)}{Y_i} \right| + \left| \frac{Y_i - \exp(X_i^\top \beta)}{\exp(X_i^\top \beta)} \right| \right\}, \quad (2)$$

* Corresponding author.

E-mail address: zfw@ustc.edu.cn (Z. Wang).

the sum of the two types of the relative errors. The LARE estimation enjoys the robustness and scale-free property. However, like the LAD, the LARE criterion function is nonsmooth, and, as a result, the limiting variance of the corresponding estimator involves the density of the error. Furthermore, its computation is slightly more complicated than linear programming.

It would be desirable to develop a criterion function which not only incorporates the relative error terms, but also is smooth and convex. The latter would ensure the numerical uniqueness of the resulting estimator and the consistency of the usual plug-in sandwich-type variance estimation. The main purpose of this paper is to introduce a simple, smooth, convex and interpretable criterion function and to develop a related inference procedure.

The rest of the paper is organized as follows. Section 2 introduces the least product relative error (LPRE) criterion and extension of the LPRE to a general class of relative error criteria, along with simple inference procedures, including point and variance estimation, hypothesis testing and related large sample properties. Sections 3 and 4 contain simulation results and a real example. Some discussion and concluding remarks are given in Section 5.

2. Method

The least absolute relative error (LARE) criterion (2) of Chen et al. [1] is the result of adding together the two relative error terms. In this paper, we consider multiplying the two relative error terms and propose the following least product relative error (LPRE) criterion

$$LPRE_n(\beta) \equiv \sum_{i=1}^n \left\{ \left| \frac{Y_i - \exp(X_i^\top \beta)}{Y_i} \right| \times \left| \frac{Y_i - \exp(X_i^\top \beta)}{\exp(X_i^\top \beta)} \right| \right\}. \quad (3)$$

Note that the summand can be written as $\{Y_i - \exp(X_i^\top \beta)\}^2 / \{Y_i \exp(X_i^\top \beta)\}$. Thus, it may be viewed as a symmetrized version of the squared relative errors [6].

A simple algebraic manipulation leads to the following alternative expression

$$LPRE_n(\beta) \equiv \sum_{i=1}^n \{Y_i \exp(-X_i^\top \beta) + Y_i^{-1} \exp(X_i^\top \beta) - 2\}, \quad (4)$$

from which we can see major advantages. First, the criterion function is infinitely differentiable. Second, it is strictly convex since the exponential function is strictly convex. As a result, finding the minimizer is equivalent to finding the root of its first derivative. The usual asymptotic properties can therefore be derived by a local quadratic expansion and standard inference methods for M-estimation are applicable.

2.1. Estimation

We now deal with parameter estimation and develop the corresponding theory. Our estimator for β will be denoted by $\hat{\beta}_n$ and defined as the minimizer of (3) or, equivalently, (4). The strict convexity of (4) entails that the minimizer, if it exists, must be unique. Assume the design matrix $\sum_{i=1}^n X_i X_i^\top$ is nonsingular. This is a minimum condition for the purpose of identifiability. Then, $LPRE_n(\beta)$ is strictly convex, and, as $\|\beta\| \rightarrow \infty$, $\sum_{i=1}^n (X_i^\top \beta)^2 \rightarrow \infty$, implying $\max\{|X_i^\top \beta| : i = 1, \dots, n\} \rightarrow \infty$. It follows that $LPRE_n(\beta) \rightarrow \infty$ as $\|\beta\| \rightarrow \infty$. And the following theorem holds.

Theorem 1. *If $\sum_{i=1}^n E(X_i X_i^\top)$ is nonsingular, then $\hat{\beta}_n$ exists and is unique.*

Remark 1. The nonsingularity of $\sum_{i=1}^n E(X_i X_i^\top)$ is also a necessary and sufficient condition for the least squares estimator to be unique.

We next establish asymptotic properties for $\hat{\beta}_n$ under suitable regularity conditions. For notational simplicity, we assume that $(X^\top, Y)^\top, (X_i^\top, Y_i)^\top, i = 1, \dots, n$ are independent and identically distributed. It allows for heteroskedasticity in that it does not require the error term ϵ to be independent of the explanatory variable X . We will use the following conditions for the development of the asymptotic theory.

Condition C1. There exists $\delta > 0$ such that $E\{(\epsilon + 1/\epsilon) \exp(\delta \|X\|)\} < \infty$.

Condition C1*. There exists $\delta > 0$ such that $E\{(\epsilon + 1/\epsilon)^2 \exp(\delta \|X\|)\} < \infty$.

Condition C2. The expected design matrix, $E(XX^\top)$, is positive definite.

Condition C3. The error terms satisfy $E(\epsilon|X) = E(1/\epsilon|X)$.

Condition C1 is almost minimal for the criterion function (4) to have a finite expectation in a neighborhood of the true parameter β_0 . It also ensures that the limit of (4) is twice differentiable with respect to β and that the differentiation and expectation is interchangeable. Condition C2 ensures that the design matrix is nonsingular, a minimal requirement for the regression parameter to be identifiable. Under C1 and C2, the limiting criterion function is strictly convex in a neighborhood of β_0 . Condition C3 is equivalent to that the derivative of the criterion function at β_0 has mean 0, again a minimal condition for the resulting estimator to be asymptotically unbiased. The strict convexity and the asymptotic unbiasedness ensure that the estimator is consistent. Condition C1* is simply a stronger version of C1 for the asymptotic normality to hold.

Theorem 2. Under Conditions C1, C2 and C3, $\hat{\beta}_n$ is strongly consistent.

Proof. Under C1, C2 and C3, one can show that $LPRE_n(\beta)/n$ converges to $E\{LPRE_n(\beta)\}/n$ in a small neighborhood of β_0 and that both are convex. Thus, by Theorem 10.8 in [9], $\hat{\beta}_n$, the minimizer of $LPRE_n(\beta)$, converges to β_0 , the minimizer of $E\{LPRE_n(\beta)\}$.

The next theorem establishes the asymptotic normality and the validity of the plug-in variance estimation. Let $D = E\{XX^\top(\epsilon + 1/\epsilon)\}$ and $V = E\{XX^\top(\epsilon - 1/\epsilon)^2\}$. Define their plug-in estimators $\hat{D} = (1/n) \sum_{i=1}^n X_i X_i^\top \{ \exp(X_i^\top \hat{\beta}_n)/Y_i + Y_i/\exp(X_i^\top \hat{\beta}_n) \}$ and $\hat{V} = (1/n) \sum_{i=1}^n X_i X_i^\top \{ Y_i/\exp(X_i^\top \hat{\beta}_n) - \exp(X_i^\top \hat{\beta}_n)/Y_i \}^2$.

Theorem 3. Under Conditions C1*, C2 and C3, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is asymptotically normal with mean 0 and covariance matrix $D^{-1}VD^{-1}$, which is consistently estimated by $\hat{D}^{-1}\hat{V}\hat{D}^{-1}$.

Proof. Since $\hat{\beta}_n$ is consistent, by the Taylor expansion,

$$\hat{\beta}_n - \beta_0 = \hat{D}_*^{-1} \frac{1}{n} \sum_{i=1}^n X_i \left\{ \frac{Y_i}{\exp(X_i^\top \beta_0)} - \frac{\exp(X_i^\top \beta_0)}{Y_i} \right\},$$

where $\hat{D}_* = (1/n) \sum_{i=1}^n X_i X_i^\top \{ \exp(X_i^\top \beta^*)/Y_i + Y_i/\exp(X_i^\top \beta^*) \}$ and β^* lies in between the true parameter β_0 and the LPRE estimate $\hat{\beta}_n$. The desired results follow from the law of large numbers and the central limit theorem.

It can be shown that when the error ϵ has density

$$f(x) = c \exp(-x - 1/x - \log x + 2)I(x > 0), \tag{5}$$

where c is the normalizing constant, $D = V$ becomes the Fisher information. It then follows that $\hat{\beta}_n$ is asymptotically efficient.

2.2. Hypothesis testing

We now turn to hypothesis testing. Although the asymptotic theory developed in the preceding subsection can be used to construct Wald-type testing statistics, we will focus on an approach that is based directly on the LPRE criterion. For simplicity, we assume homogeneous errors, i.e. ϵ is independent of X and consider the following null hypothesis

$$H_0 : \beta \in \Omega_0 = \{b \in R^p : H^\top b = 0\}, \tag{6}$$

where $H = (h_1, \dots, h_q)$ and $h_j, j = 1, \dots, q$ are p -vectors that are linearly independent and lie in the linear space spanned by the column vectors of the design matrix X .

Let

$$M_n \equiv \min_{\beta \in \Omega_0} LPRE_n(\beta) - \min_{\beta \in R^p} LPRE_n(\beta). \tag{7}$$

Through the usual quadratic expansion, we can arrive at an asymptotic ANOVA-type decomposition. The asymptotic normality can then be applied to show that, under the null hypothesis, M_n converges in distribution to $K\chi_q^2$, where $K = 4E(\epsilon)/E\{(\epsilon - 1/\epsilon)^2\}$ and χ_q^2 is the central chi-squared distribution with q degree of freedom. The constant K can be estimated consistently by $\hat{K} = 4 \sum_{i=1}^n \{Y_i \exp(-X_i^\top \hat{\beta}_n)\} / \sum_{i=1}^n \{Y_i \exp(-X_i^\top \hat{\beta}_n) - 1/Y_i \exp(X_i^\top \hat{\beta}_n)\}^2$. Therefore we can use M_n/\hat{K} as the testing statistic with $\chi_{q,1-\alpha}^2$ as the cut-off point, where α is a given nominal significance level.

2.3. General relative error criteria

A general relative error (GRE) criterion can be constructed:

$$GRE_n(\beta) \equiv \sum_{i=1}^n g \left(\left| \frac{Y_i - \exp(X_i^\top \beta)}{Y_i} \right|, \left| \frac{Y_i - \exp(X_i^\top \beta)}{\exp(X_i^\top \beta)} \right| \right), \tag{8}$$

where $g(a, b)$ is a bivariate function satisfying certain regularity conditions. Taking $g(a, b) = a + b$, it becomes the LARE criterion function while $g(a, b) = ab$, it becomes the LPRE of the preceding section. One may also consider $g(a, b) = \max\{a, b\}$ [13]. Note that all three criteria here are symmetric functions. A possible non-symmetric one could be $g(a, b) = a + \exp(b)$, where we pay more attention to the relative error of b and more heavily penalize large value of b compared to a .

The derivative of GRE_n with respect to β is denoted as

$$S_n(\beta) = \sum_{i=1}^n \phi \left(\left| \frac{Y_i - \exp(X_i^\top \beta)}{Y_i} \right|, \left| \frac{Y_i - \exp(X_i^\top \beta)}{\exp(X_i^\top \beta)} \right| \right) X_i.$$

Its expectation with $\beta = \beta_0$ becomes 0 when

$$E \{ \phi (|1 - \epsilon^{-1}|, |\epsilon - 1|) | X \} = 0. \quad (9)$$

With the LARE and the LPRE as special cases, we have $\phi (|1 - \epsilon^{-1}|, |\epsilon - 1|) = (1/\epsilon + \epsilon) \{ I(\epsilon \leq 1) - I(\epsilon \geq 1) \}$ and $1/\epsilon - \epsilon$, respectively. Then, in the special cases of LARE and LPRE, the expression (9) takes the forms $E \{ [(1/\epsilon + \epsilon) \{ I(\epsilon \leq 1) - I(\epsilon \geq 1) \}] | X \} = 0$ and $E (1/\epsilon - \epsilon | X) = 0$ respectively, when $\log(\epsilon)$ has a density function symmetric around 0.

Let $\hat{\beta}_n$ be a minimizer of the criterion function (8). It follows that, under (9), $\hat{\beta}_n$ is asymptotically unbiased. In fact we have the following result concerning the limiting distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

Theorem 4. Under (9) and additional regularity conditions concerning the nonsingularity of the design matrix and finite moment condition on the error, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is asymptotically normal with mean 0 and covariance matrix $(1/a^2)JV^{-1}$ where $J = E \{ [\phi (|\epsilon - 1|, |1 - 1/\epsilon|)]^2 \}$, $V = E(XX^T)$ and constant a satisfies $E \{ \phi (|1 - \epsilon \exp(-c)|, |\epsilon^{-1} \exp(c) - 1|) \} = ac + o(|c|)$, as $|c| \rightarrow 0$.

The proof of Theorem 4 is similar to that of Theorem 1 in [1] and is omitted. In general, the asymptotic variance of $\hat{\beta}_n$, the minimizer of $GRE_n(\beta)$, may involve the density function of the error. To avoid density estimation, a distributional approximation based on random weighting can be applied. Let w_1, w_2, \dots, w_n be a sequence of i.i.d. nonnegative random variables independent of $\{(X_i^T, Y_i)\}$ with $E(w_1) = \text{Var}(w_1) = 1$. Define

$$GRE_n^*(\beta) \equiv \sum_{i=1}^n w_i g \left(\left| \frac{Y_i - \exp(X_i^T \beta)}{Y_i} \right|, \left| \frac{Y_i - \exp(X_i^T \beta)}{\exp(X_i^T \beta)} \right| \right), \quad (10)$$

and let $\hat{\beta}_n^*$ be a minimizer of (10). Similar to Proposition 1 in [1], it can be shown that, given the data $\{(X_i^T, Y_i) : i = 1, \dots, n\}$, the asymptotic conditional distribution of $\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n)$ is $N(0, (1/a^2)JV^{-1})$, which is exactly the asymptotic distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$. Thence, the numerical procedure for the random weighting distributional approximation is as follows: fix the samples $\{(X_i^T, Y_i) : i = 1, \dots, n\}$,

- (1) Generate random weights w_i , $i = 1, \dots, n$, independently from a preset weight distribution, for example, the exponential distribution with mean 1;
- (2) Obtain estimate $\hat{\beta}_n^*$ by minimizing $GRE_n^*(\beta)$ over β ;
- (3) Repeat steps (1) and (2) for N times, and obtain N estimators of β , saying $\{\hat{\beta}_{nk}^*, k = 1, \dots, N\}$.
- (4) Empirical distribution and empirical variance of $\{\hat{\beta}_{nk}^*, k = 1, \dots, N\}$ are used to approximate the distribution and variance of $\hat{\beta}_n$, respectively.

If the error ϵ has a density function as follows:

$$f(x) = c \exp \{ -g(|1 - x|, |1 - x^{-1}|) - \log x \} I(x > 0), \quad (11)$$

where c is a normalizing constant, then the estimator $\hat{\beta}_n$ is asymptotically efficient. Density $f(x)$ in (11) belongs to a class of inverse transformation invariant densities, meaning that if a random variable ϵ is distributed with density $f(x)$, then $1/\epsilon$ is equal in distribution to ϵ . Fig. 1 shows densities of some particular choices of function g . One can see that the error distribution with which the product criterion is efficient has heavier tails than others, indicating that the product criterion is more robust in practical application.

Based on general relative error criterion (8), a general test statistic to test hypothesis (6) can be constructed as

$$M_n \equiv \min_{\beta \in \Omega_0} GRE_n(\beta) - \min_{\beta \in RP} GRE_n(\beta). \quad (12)$$

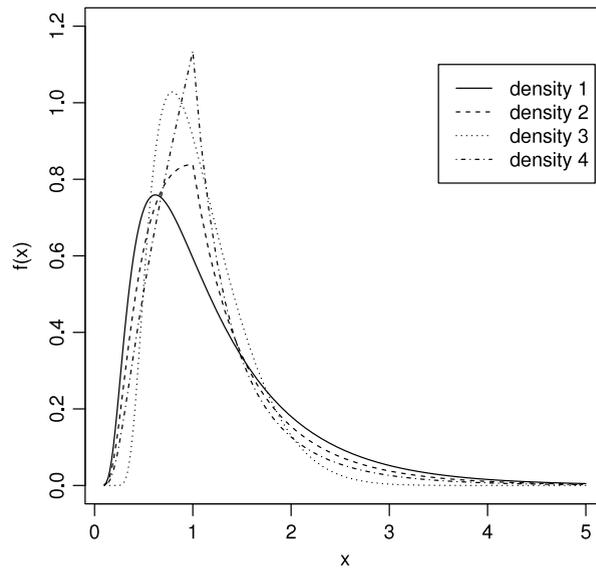
Especially, when the error terms follow the distribution described in (11), M_n is identical to the log-likelihood ratio test statistic. The following theorem demonstrates the asymptotic distribution of M_n .

Theorem 5. Under (6) and additional regularity conditions,

$$M_n \rightarrow \frac{J}{2a} \chi_q^2 \quad \text{in distribution,}$$

as $n \rightarrow \infty$, where χ_q^2 refers to the chi-square distribution with q degrees of freedom.

The proof of Theorem 5 is similar to that of Theorem 1 in [2] and is also omitted. In general, the asymptotic distribution of M_n may involve the density of the errors. The plug-in method involving density estimation can be inaccurate and computationally troublesome. In this case, a random weighting method, as used by Chen et al. [2], Wang et al. [11], Xiao



density1: $f_1(x) = c_1 \exp(-x - x^{-1} - \log x + 2)I(x > 0)$;
 density2: $f_2(x) = c_2 \exp(-\max\{|1 - x|, |1 - x^{-1}|\} - \log x)I(x > 0)$;
 density3: $f_3(x) = c_3 \exp\{-(1 - x)^2 - (1 - x^{-1})^2 - \log x\}I(x > 0)$;
 density4: $f_4(x) = c_4 \exp(-|1 - x| - |1 - x^{-1}| - \log x)I(x > 0)$.

Fig. 1. Plot of four densities.

et al. [12] can be applied. For illustration, let $\hat{\beta}_c$ and $\hat{\beta}_c^*$ be the minimizers of $GRE_n(\beta)$ and $GRE_n^*(\beta)$ under H_0 defined in (6) respectively. We define the following testing statistic

$$M_n^* = \left[GRE_n^*(\hat{\beta}_c^*) - GRE_n^*(\hat{\beta}_n^*) \right] - \left[GRE_n^*(\hat{\beta}_c) - GRE_n^*(\hat{\beta}_n) \right]. \tag{13}$$

Following the lines of the proof of Theorem 4 in [12], we can show that, conditional on the data $\{(X_i^\top, Y_i) : i = 1, \dots, n\}$, the asymptotic distribution of M_n^* is $(J/2a)\chi_q^2$, which allows us to use M_n^* to approximate the distribution of M_n under the null and hence avoid the estimation of a .

3. Simulation

Simulation studies are conducted to compare the finite sample performance of the proposed LPRE, the LARE, the LS and the LAD. The data are generated from the model

$$Y = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)\epsilon, \tag{14}$$

where X_1 and X_2 are two covariates following the standard normal distribution $N(0, 1)$ with correlation coefficient $\rho = 0$ or 0.6 . We take $(\beta_0, \beta_1, \beta_2)^\top = (1, 1, 1)^\top$. We consider five error distributions: the distribution with which the LARE estimator is efficient; the distribution with which the LPRE estimator is efficient; the exponential of the uniform distribution on $(-2, 2)$; the log-standard normal distribution; and the uniform distribution on $(0.5, a)$ with a chosen such that $E(\epsilon) = E(1/\epsilon)$. Note that the first four error distributions are such that $1/\epsilon$ is distributed same as ϵ . The sample size n is 200. The variance estimation for the LARE and the LAD is based on random weighting with resampling size $N = 500$, while the variance estimation for the LPRE and the LS is based on the plug-in rule. The LS and LAD estimates are obtained by minimizing $\sum_{i=1}^n (\log Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})^2$ and $\sum_{i=1}^n |\log Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}|$, respectively. The simulation results are based on 1000 replications.

It is seen from Table 1 that the LPRE performs considerably better than the LARE, the LS and the LAD when $\log(\epsilon)$ is uniformly distributed on $(-2, 2)$. With log-normal error distribution, the LPRE performs as well as the LARE and is comparable to the LS. With the uniform error distribution, the LPRE performs slightly better than the LS, and much better than the LARE and the LAD. The variance estimation of the LPRE gives accurate coverage probability in the study.

The performance of the proposed test statistic M_n is evaluated with the product relative error criterion. We consider two null hypotheses $H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$. Tables 2 and 3 present the empirical significance levels and powers with $n = 200$ when ϵ follows the distributions with which the LPRE and the LARE are respectively efficient, the log-uniform distribution on $(-2, 2)$ and the log-standard normal distribution. It is seen that the empirical significance levels are close to the nominal levels, suggesting that M_n is adequate. Under $H_0 : \beta_2 = 0$ and nominal level 0.05, the power increases from

Table 1
Comparison among various approaches with $\beta = (1, 1, 1)^T$.

ρ	$\epsilon \sim f_1(\cdot)$			$\log(\epsilon) \sim Unif(-2, 2)$			$\log(\epsilon) \sim N(0, 1)$			$\epsilon \sim f_2(\cdot)$			$\epsilon \sim Unif(0.5, a)$				
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$		
0	LPRE	BIAS	0.007	0.004	0.009	0.001	0.001	0.004	0.002	0.001	0.001	0.004	0.005	0.004	0.002	0.000	0.001
		SE	0.037	0.037	0.037	0.067	0.067	0.067	0.074	0.075	0.076	0.045	0.045	0.045	0.023	0.023	0.023
		SEE	0.036	0.036	0.036	0.067	0.067	0.067	0.075	0.075	0.075	0.045	0.045	0.045	0.023	0.023	0.023
		CP	0.943	0.942	0.955	0.952	0.952	0.952	0.948	0.949	0.945	0.945	0.956	0.947	0.948	0.950	0.950
	LARE	BIAS	0.001	0.002	0.001	0.001	0.002	0.000	0.004	0.004	0.002	0.001	0.001	0.001	0.044	0.000	0.001
		SE	0.032	0.033	0.034	0.077	0.075	0.073	0.076	0.073	0.076	0.047	0.048	0.047	0.035	0.034	0.034
		SEE	0.033	0.034	0.034	0.075	0.075	0.075	0.073	0.072	0.072	0.047	0.047	0.047	0.032	0.032	0.033
		CP	0.945	0.944	0.951	0.944	0.943	0.959	0.926	0.928	0.931	0.936	0.927	0.934	0.942	0.943	0.943
	LS	BIAS	0.001	0.002	0.001	0.001	0.002	0.000	0.004	0.003	0.002	0.004	0.005	0.005	0.004	0.005	0.001
		SE	0.035	0.035	0.037	0.083	0.081	0.078	0.071	0.069	0.072	0.047	0.047	0.047	0.025	0.025	0.025
		SEE	0.035	0.035	0.035	0.081	0.080	0.080	0.070	0.069	0.070	0.045	0.045	0.045	0.025	0.026	0.026
		CP	0.945	0.952	0.926	0.948	0.937	0.951	0.950	0.939	0.935	0.939	0.941	0.950	0.945	0.946	0.947
LAD	BIAS	0.001	0.002	0.001	0.001	0.004	0.001	0.004	0.003	0.001	0.009	0.008	0.010	0.053	0.001	0.002	
	SE	0.033	0.034	0.034	0.143	0.140	0.135	0.090	0.085	0.090	0.061	0.060	0.058	0.037	0.037	0.036	
	SEE	0.036	0.038	0.038	0.145	0.144	0.144	0.093	0.094	0.094	0.063	0.062	0.062	0.040	0.040	0.039	
	CP	0.938	0.915	0.921	0.897	0.868	0.888	0.917	0.907	0.906	0.899	0.887	0.906	0.900	0.902	0.901	
0.6	LPRE	BIAS	0.002	0.002	0.001	0.001	0.001	0.002	0.003	0.001	0.000	0.001	0.002	0.003	0.001	0.001	0.001
		SE	0.037	0.068	0.069	0.065	0.123	0.125	0.077	0.142	0.142	0.046	0.083	0.083	0.023	0.041	0.041
		SEE	0.036	0.066	0.066	0.067	0.122	0.122	0.074	0.136	0.136	0.045	0.082	0.082	0.023	0.041	0.041
		CP	0.938	0.950	0.946	0.960	0.941	0.947	0.936	0.941	0.940	0.942	0.945	0.947	0.940	0.955	0.959
	LARE	BIAS	0.002	0.000	0.001	0.002	0.000	0.000	0.002	0.002	0.000	0.001	0.002	0.002	0.042	0.002	0.001
		SE	0.033	0.061	0.061	0.073	0.137	0.139	0.076	0.140	0.139	0.050	0.091	0.092	0.035	0.061	0.061
		SEE	0.034	0.063	0.063	0.075	0.136	0.136	0.072	0.130	0.129	0.050	0.093	0.093	0.034	0.061	0.061
		CP	0.962	0.961	0.950	0.953	0.934	0.942	0.933	0.921	0.929	0.935	0.947	0.946	0.739	0.950	0.950
	LS	BIAS	0.002	0.002	0.000	0.002	0.000	0.001	0.002	0.001	0.000	0.001	0.002	0.003	0.003	0.001	0.001
		SE	0.036	0.066	0.066	0.079	0.149	0.152	0.073	0.133	0.132	0.046	0.083	0.084	0.023	0.041	0.041
		SEE	0.035	0.065	0.065	0.082	0.150	0.150	0.071	0.130	0.130	0.046	0.084	0.083	0.023	0.042	0.042
		CP	0.941	0.948	0.948	0.963	0.948	0.942	0.942	0.948	0.948	0.940	0.948	0.950	0.936	0.957	0.961
LAD	BIAS	0.002	0.001	0.000	0.002	0.001	0.001	0.001	0.000	0.001	0.001	0.002	0.003	0.010	0.001	0.001	
	SE	0.035	0.063	0.064	0.082	0.154	0.158	0.074	0.136	0.134	0.048	0.086	0.087	0.026	0.043	0.043	
	SEE	0.034	0.063	0.063	0.085	0.155	0.155	0.073	0.133	0.133	0.048	0.087	0.087	0.024	0.044	0.044	
	CP	0.946	0.952	0.953	0.961	0.948	0.940	0.943	0.940	0.950	0.949	0.953	0.947	0.901	0.955	0.962	

$f_1(x) = c_1 \exp(-|1 - x| - |1 - x^{-1}| - \log x)I(x > 0)$;
 $f_2(x) = c_2 \exp(-x - x^{-1} - \log x + 2)I(x > 0)$.

Table 2
Type I error and power with the null hypothesis $(\beta_0, \beta_1, \beta_2) = (1, 1, 0)$.

ρ	β	$\log(\epsilon) \sim Unif(-2, 2)$		$\log(\epsilon) \sim N(0, 1)$		$\epsilon \sim f_1(\cdot)$		$\epsilon \sim f_2(\cdot)$	
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
0	(1.0, 1.0, 0.0)	0.053	0.015	0.049	0.009	0.057	0.013	0.053	0.006
	(1.0, 1.0, 0.1)	0.338	0.157	0.280	0.120	0.593	0.350	0.772	0.569
	(1.0, 1.0, 0.2)	0.823	0.637	0.745	0.555	0.987	0.965	0.999	0.997
	(1.0, 1.0, 0.3)	0.984	0.955	0.980	0.921	1.000	1.000	1.000	1.000
	(1.0, 1.0, 0.4)	1.000	0.999	0.998	0.991	1.000	1.000	1.000	1.000
0.6	(1.0, 1.0, 0.0)	0.056	0.013	0.063	0.013	0.056	0.013	0.061	0.013
	(1.0, 1.0, 0.1)	0.130	0.042	0.116	0.032	0.202	0.086	0.346	0.130
	(1.0, 1.0, 0.2)	0.376	0.180	0.336	0.151	0.670	0.407	0.854	0.665
	(1.0, 1.0, 0.3)	0.711	0.471	0.627	0.377	0.942	0.838	0.991	0.967
	(1.0, 1.0, 0.4)	0.900	0.764	0.836	0.635	0.999	0.990	1.000	1.000

α represents the nominal significance level.

0.05 to 1.0 as β_2 varies from 0.0 to 0.4. In other words, the power increases as the parameters move away from the null hypothesis, a common phenomenon in hypothesis testing.

4. Real example

We apply the proposed method for the analysis of the body fat data. The data contain various body measurement indices related with percentage of body fat for 252 men, which are available at <http://lib.stat.cmu.edu/datasets/bodyfat>; [7]. We select 12 explanatory variables: age (X_1), height⁴/weight²(X_2) and 10 other body circumference indexes (neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm and wrist, denoted by $X_i, i = 3, \dots, 12$). We note that X_2 is a transform of the well-known body mass index (BMI) defined as the ratio of weight to height². The sample size n is 251. The response

Table 3
Type I error and power with the null hypothesis $(\beta_0, \beta_1, \beta_2) = (1, 0, 0)$.

ρ	β	$\log(\epsilon) \sim Unif(-2, 2)$		$\log(\epsilon) \sim N(0, 1)$		$\epsilon \sim f_1(\cdot)$		$\epsilon \sim f_2(\cdot)$	
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
0	(1.0, 0.0, 0.0)	0.045	0.013	0.057	0.015	0.048	0.008	0.058	0.008
	(1.0, 0.1, 0.0)	0.270	0.103	0.222	0.084	0.524	0.281	0.683	0.463
	(1.0, 0.1, 0.1)	0.462	0.258	0.391	0.214	0.809	0.607	0.941	0.828
	(1.0, 0.2, 0.0)	0.773	0.568	0.663	0.455	0.984	0.933	0.997	0.990
0.6	(1.0, 0.2, 0.2)	0.965	0.900	0.914	0.810	1.000	1.000	1.000	1.000
	(1.0, 0.0, 0.0)	0.045	0.012	0.061	0.010	0.048	0.012	0.052	0.015
	(1.0, 0.1, 0.0)	0.315	0.136	0.297	0.138	0.601	0.391	0.827	0.638
	(1.0, 0.1, 0.1)	0.860	0.707	0.770	0.574	0.998	0.976	1.000	0.998
	(1.0, 0.2, 0.0)	0.902	0.746	0.810	0.623	0.999	0.988	1.000	0.999
	(1.0, 0.2, 0.2)	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000

Table 4
Analysis of the body fat data with LPRE, LARE, LS and LAD.

	LPRE		LARE		LS		LAD	
	Est	p-value	Est	p-value	Est	p-value	Est	p-value
β_0	2.823 (0.026)	0.000	2.851 (0.027)	0.000	2.835 (0.022)	0.000	2.883 (0.029)	0.000
β_1	0.085 (0.038)	0.013	0.052 (0.027)	0.027	0.072 (0.031)	0.011	0.055 (0.038)	0.074
β_2	-0.155 (0.068)	0.011	-0.205 (0.073)	0.002	-0.156 (0.056)	0.003	-0.211 (0.088)	0.008
β_3	-0.103 (0.052)	0.024	-0.064 (0.044)	0.073	-0.102 (0.043)	0.009	-0.064 (0.047)	0.087
β_4	-0.167 (0.076)	0.014	-0.168 (0.063)	0.004	-0.134 (0.063)	0.017	-0.093 (0.081)	0.125
β_5	0.582 (0.091)	0.000	0.547 (0.084)	0.000	0.558 (0.075)	0.000	0.501 (0.105)	0.000
β_6	-0.231 (0.085)	0.003	-0.200 (0.069)	0.002	-0.217 (0.07)	0.001	-0.235 (0.079)	0.001
β_7	0.105 (0.077)	0.086	0.047 (0.055)	0.196	0.090 (0.063)	0.076	0.064 (0.073)	0.190
β_8	0.026 (0.054)	0.315	0.003 (0.038)	0.469	0.020 (0.044)	0.327	-0.011 (0.048)	0.409
β_9	-0.009 (0.036)	0.401	-0.018 (0.037)	0.313	-0.005 (0.029)	0.437	-0.002 (0.032)	0.475
β_{10}	0.081 (0.049)	0.049	0.034 (0.049)	0.244	0.051 (0.041)	0.106	0.008 (0.056)	0.443
β_{11}	0.033 (0.040)	0.205	0.035 (0.034)	0.152	0.033 (0.033)	0.157	0.028 (0.042)	0.252
β_{12}	-0.088 (0.047)	0.031	-0.084 (0.036)	0.010	-0.088 (0.039)	0.012	-0.095 (0.044)	0.015

Est: parameter estimate. The estimated standard deviations are given in the parentheses.

Table 5
Comparisons of median prediction errors with LPRE, LARE, LS and LAD.

	LPRE	LARE	LS	LAD
MPE	3.679	3.957	3.861	3.933
MPPE	0.039	0.046	0.043	0.041
MPAE	0.401	0.433	0.418	0.410
MSPE	13.537	15.660	14.907	15.468

variable Y is the percentage of body fat. We delete one observation with $Y = 0$ and fit the model

$$Y_i = \exp\left(\beta_0 + \sum_{j=1}^{12} \beta_j Z_j\right) \epsilon_i, \quad i = 1, \dots, n, \tag{15}$$

where $Z_j, j = 1, \dots, 12$, denote the normalized explanatory variables.

To evaluate the performance of different methods, the dataset is partitioned into two parts. The first part with 200 observations is used to fit model (15), and the rest 51 observations are used to evaluate the prediction power. The results are shown in Tables 4 and 5.

The p -value is calculated by $1 - \Phi(|\hat{\beta}_j/\hat{s}_j|)$, where $\hat{\beta}_j$ is the estimate of β_j, \hat{s}_j is the estimated standard deviation for $\hat{\beta}_j$, and $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. The variance estimation of the LPRE and the LS are obtained by the plug-in rule, while that for LARE and LAD are obtained by random weighting resampling. Table 4 shows that the four methods identify some common variables (with p -value < 0.05), such as age, 1/BMI and abdomen circumference. It makes sense that the percentage of body fat increases as age, BMI and abdomen circumference become large. However, the biceps circumference is identified only by the LPRE, indicating the percentage of body fat increases as the biceps circumference becomes larger.

The prediction accuracy based on the four methods estimation is measured by four different median indices: median of absolute prediction errors $\{|Y_i - \hat{Y}_i|\}$ (MPE), median of product relative prediction errors $\{|Y_i - \hat{Y}_i|^2 / (Y_i \hat{Y}_i)\}$ (MPPE), median of additive relative prediction errors $\{|Y_i - \hat{Y}_i|/Y_i + |Y_i - \hat{Y}_i|/\hat{Y}_i\}$ (MAPE) and median of squared prediction errors $\{(Y_i - \hat{Y}_i)^2\}$ (MSPE), where $\hat{Y}_i = \exp(\hat{\beta}_0 + \sum_{j=1}^{12} \hat{\beta}_j Z_j), i = 201, \dots, 251$. Table 5 shows that the LPRE has the smallest MPE, MPPE, MAPE and MSPE among the LPRE, the LARE, the LS and the LAD.

5. Conclusion

In our view, in the realm of criteria based on relative errors, the LPRE proposed in this paper has the best potential to be the basic and primary choice, just like the least squares in the realm of criteria based on absolute errors. The proposed LPRE represents a substantial improvement over that of Chen et al. [1] both theoretically and computationally, particularly in terms of the simplicity of inference. Extensions can be made to cover analysis of censored data and high dimensional data. Moreover, we present a more general GRE method and tests of linear hypotheses based on relative errors, which is not studied in [1] and other relevant literature. The LPRE criterion may have broad applications in financial and survival data analysis.

Acknowledgments

The authors are grateful to the Editor, the Associate Editor, and the anonymous referees for comments and suggestions that lead to improvements in the paper. Chen's work is supported by Hong Kong RGC General Research Funds (Nos. 601011, 600612, 600813 and 16300714). Lin's work is supported by the Hong Kong Research Grants Council (No. 509413) and a Direct Grant for Research at the Chinese University of Hong Kong (No. 4053136). Wang's work is supported by funds of the State Key Program of National Natural Science Foundation of China (No. 11231010) and National Natural Science Foundation of China (No. 11471302).

References

- [1] K. Chen, S. Guo, Y. Lin, Z. Ying, Least absolute relative error estimation, *J. Amer. Statist. Assoc.* 105 (2010) 1104–1112.
- [2] K. Chen, Z. Ying, H. Zhang, L. Zhao, Analysis of least absolute deviation, *Biometrika* 95 (2008) 107–122.
- [3] T. Khoshgoftaar, B. Bhattacharyya, G. Richardson, Predicting software errors, during development, using nonlinear regression models: a comparative study, *IEEE Trans. Reliab.* 41 (1992) 390–395.
- [4] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, R. Winkler, *The Forecasting Accuracy of Major Time Series Methods*, Wiley, New York, 1984.
- [5] S. Narula, J. Wellington, Prediction, linear regression and the minimum sum of relative errors, *Technometrics* 19 (1977) 185–190.
- [6] H. Park, L. Stefanski, Relative-error prediction, *Statist. Probab. Lett.* 40 (1998) 227–236.
- [7] K. Penrose, A. Nelson, A. Fisher, Generalized body composition prediction equation for men using simple measurement techniques (abstract), *Med. Sci. Sports Exerc.* 17 (1985) 189.
- [8] S. Portnoy, R. Koenker, The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators (with discussion), *Statist. Sci.* 12 (1997) 279–300.
- [9] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [10] S. Stigler, Gauss and the invention of least squares, *Ann. Statist.* 9 (1981) 465–474.
- [11] Z. Wang, Y. Wu, L. Zhao, Approximation by randomly weighting method for linear hypothesis testing in censored regression model, *Sci. China, Ser. A Math.* 52 (2009) 561–576.
- [12] L. Xiao, B. Hou, Z. Wang, Y. Wu, Random weighting approximation for tobit regression models with longitudinal data, *Comput. Statist. Data Anal.* 79 (2014) 235–247.
- [13] J. Ye, *Price models and the value relevance of accounting information*, Technical report, 2007.
- [14] Q. Zhang, Q. Wang, Local least absolute relative error estimating approach for partially linear multiplicative model, *Statist. Sinica* 23 (2013) 1091–1116.