Journal of Multivariate Analysis 144 (2016) 91–98

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

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.

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

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

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

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

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

http://dx.doi.org/10.1016/j.jmva.2015.10.017
0047-259X/

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

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

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

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

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 $\|\beta\| \to \infty$, $\sum_{i=1}^n (X_i^\top \beta)^2 \to \infty$, implying $\max\{|X_i^\top \beta| : i = 1, \ldots, n\} \to \infty$. It follows that $LPRE_n(\beta) \to \infty$ as $\|\beta\| \to \infty$. 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 $(X^\top, Y)^\top, (X_i^\top, Y_i)^\top, i = 1, \ldots, n$ 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.