Published in final edited form as:
J Am Stat Assoc. 2010 ; 105(491): 1104–1112. doi:10.1198/jasa.2010.tm09307.

Least Absolute Relative Error Estimation

Kani CHEN [Professor],
Department of Mathematics, HKUST, Kowloon, Hong Kong, China (makchen@ust.hk)

Shaojun GUO [Assistant Professor],
Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R.China (guoshaoj@amss.ac.cn)

Yuanyuan LIN [Ph.D. Candidate], and
Department of Mathematics, HKUST, Kowloon, Hong Kong, China (linyy@ust.hk)

Zhiliang YING [Professor]
Department of Statistics, 618 Mathematics, Columbia University, New York NY 10027 (zying@stat.columbia.edu)

Abstract

Multiplicative regression model or accelerated failure time model, which becomes linear regression model after logarithmic transformation, is useful in analyzing data with positive responses, such as stock prices or life times, that are particularly common in economic/financial or biomedical studies. Least squares or least absolute deviation are among the most widely used criterions in statistical estimation for linear regression model. However, in many practical applications, especially in treating, for example, stock price data, the size of relative error, rather than that of error itself, is the central concern of the practitioners. This paper offers an alternative to the traditional estimation methods by considering minimizing the least absolute relative errors for multiplicative regression models. We prove consistency and asymptotic normality and provide an inference approach via random weighting. We also specify the error distribution, with which the proposed least absolute relative errors estimation is efficient. Supportive evidence is shown in simulation studies. Application is illustrated in an analysis of stock returns in Hong Kong Stock Exchange.

Keywords

Multiplicative regression model; Logarithm transformation; Relative error; Random weighting

1. INTRODUCTION

Linear regression model is one of the most fundamental statistical models. And the most popular method of estimation, which dates back to Gauss, is the method of least squares (LS); see Gauss (1809) and Stigler (1981). Specifically, consider

$$Y_i^* = \boldsymbol{X}_i^\top \beta + \varepsilon_i^*, \qquad i = 1, \cdots, n, \quad (1)$$

where Y_i^* and $\boldsymbol{X}_i$ are, respectively, the response variable and observable p-vector of covariates, β is the p-vector of regression coefficients including an intercept and ε_i^* is the unobservable error term independent of $\boldsymbol{X}_i$. The least squares criterion is to minimize the sum of squares of the errors: $\sum_{i=1}^{n} \left(Y_i^* - \boldsymbol{X}_i^\top \beta\right)^2$. The resulting LS estimator enjoys some important optimality, such as best linear unbiased estimator. It is efficient when the errors follow normal distribution. An important alternative to the least squares method is the least

absolute deviation (LAD) method, which is to minimize the sum of absolute values of the errors: $\sum_{i=1}^{n}|Y_i^* - \boldsymbol{X}_i^{\intercal}\beta|$. The LAD estimator is more robust than the LS estimator, and its computation and inference procedure is now rather straightforward with the help of linear program and random weighting. A comprehensive discussion may be found in Portnoy and Koenker (1997). We note that the LS method requires finite second moment of the errors while the LAD requires positivity of the density of the errors at 0.

The above LS and LAD criterions are based on absolute errors. In many practical applications, however, the relative errors, rather than the absolute errors, are more of concern. Narula and Wellington (1977) presented an estimation method based on minimizing the sum of absolute relative errors for linear model. Makridakis *et al* (1984) used relative error as a model selection criterion in time series modeling. Khoshgoftaar *et al* (1992) gave sufficient conditions to ensure the strong consistency of the estimators minimizing the sum of squared relative errors: $\sum_{i=1}^{n}[\{Y_i - f(\boldsymbol{X}_i,\beta)\}/Y_i]^2$ (RLS for relative least squares) and minimizing the sum of the absolute relative errors: $\sum_{i=1}^{n}|\{Y_i - f(\boldsymbol{X}_i,\beta)\}/Y_i|$ (MRE for minimum relative errors) for nonlinear regression model $Y_i = f(\boldsymbol{X}_i,\beta)+\varepsilon_i$, where $f(x,\beta)$ is the regression function and Y_i, $\boldsymbol{X}_i$, β, ε_i are given in model (1). Park and Stefanski (1998) derived a closed form expression for the best mean squared relative error predictor of Y given $\boldsymbol{X}$, where Y is the response variable and $\boldsymbol{X}$ is the predictor variable. These approaches are conceptually appealing and quite easy to implement. Under certain restrictive, such as parametric, modeling assumptions, Park and Stefanski (1998) and Khoshgoftaar *et al* (1992) reported some elegant results. However, the theoretical justifications of the RLS and MRE methods are in general quite challenging. The consistency and asymptotic normality of RLS and MRE estimators for linear or nonlinear models are not established under general regularity conditions. Moreover, in all these studies, the relative error is defined as the spread between the target value and the predictor divided by the target value, i.e., the ratio of the error relative to the target. Such a relative error can be quite inadequate when, in particular, the unknown target value is large and the predictor is relatively small. On the other hand, the ratio of the error relative to the predictor can very well be an alternative representation of the relative error. More discussions on the choice of criterion of relative errors are given in Section 2. A similar consideration is seen in an accounting model in Ye (2007).

In the next section, we propose the least absolute relative errors criterion (LARE) for multiplicative models, by using both types of relative errors. Since the responses are usually positive when relative error is of concern, the multiplicative model or accelerated failure time (AFT) model naturally handles positive responses. In section 3, a large sample theory including consistency and asymptotic normality is presented along with an inference procedure with random weighting. Conditions, especially on the error terms, are also specified. In addition, the error distribution with which the LARE is efficient is given. Section 4 contains results of simulation studies. An illustration with a real example is given in Section 5. All proofs are deferred to the Appendix.

2. THE MODEL AND THE LARE CRITERION

Consider the following multiplicative model or accelerated failure time model:

$$Y_i = \exp\left(\boldsymbol{X}_i^{\intercal}\beta\right)\varepsilon_i, \qquad i=1,\cdots,n, \quad (2)$$

which, by taking logarithmic transformation, is model (1) with $Y_i^* = \log(Y_i)$ and $\varepsilon_i^* = \log(\varepsilon_i)$. Such logarithmic transformation is a reasonable choice in some cases due to its theoretical

simplicity. However, a linear relationship in the transformed model is not linear in the original one. And one need to transform the analysis results back to the original measurement scale.

Observe that the predictor of Y_i with covariate $\boldsymbol{X}_i$ is $\exp\left(\boldsymbol{X}_i^\tau\beta\right)$. It is intuitively appealing and interpretable to consider the relative error

$$\left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)}{Y_i}\right| \quad \text{or} \quad \left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)}{\exp\left(\boldsymbol{X}_i^\tau\beta\right)}\right|.$$

We note that $|\log(Y_i) - \boldsymbol{X}_i^\tau\beta$ is approximately equal to $|Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)|/Y_i$ or $|Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)|/\exp\left(\boldsymbol{X}_i^\tau\beta\right)$ only when the relative error is very small.

Remark 1

A measurement of relative error in terms of the ratio of the error relative to the target value can be inappropriate. Consider, for example, Y_i being large, say, 100, and the predictor $\exp\left(\boldsymbol{X}_i^\tau\beta\right)$ being small, say 10. The relative error so defined, $|Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)|/Y_i$, returns a value 0.9, whilst the alternative $|Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)|/\exp\left(\boldsymbol{X}_i^\tau\beta\right)$ returns 9. The latter, in this case, more properly reflects the inaccuracy of the predictor. The criteria RLS and MRE which use the former as the relative error are thus inadequate in this case. Conversely, only using the latter as relative error can be equally inappropriate when the predictor is large but the response is small. The criterion LARE that we propose below takes into consideration both types of relative errors. We note that the criteria RLS and MRE, if using both types of relative errors, are increasingly difficult to analyze. In particular, the closed form expression of the best mean squared relative error predictor of Y given $\boldsymbol{X}$ shall not be available anymore.

The criterion we propose, called least absolute relative errors (LARE), is to minimize the sum of the absolute relative errors for model (2):

$$LARE_n(\beta) \equiv \sum_{i=1}^{n}\left\{\left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)}{Y_i}\right| + \left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)}{\exp\left(\boldsymbol{X}_i^\tau\beta\right)}\right|\right\}. \quad (3)$$

One advantage is that they are scale free or unit free. This is particularly important for applying LARE criterion to certain types of data. For example, in regression analysis of a number of stocks, comparison of share prices of different stocks is generally meaningless, especially because of possible share split or reverse split. In other words, different stocks have different units which are not well defined. The criterions based on absolute errors is not directly applicable here without accounting for the heterogeneity.

The proposed LARE criterion is based on the sum of the two types of the relative errors. There are also several different ways of combining the two types of errors. For example, one might consider the maximum of the two, as appeared in Ye (2007), in which case, a theory can be developed in an analogous fashion; see more discussion in Section 6. The computation of minimizing $LARE(\beta)$ can be carried out by the conventional numerical tools, such as the Newton-Raphson method, or by the programming similar to that of LAD regression which is now a standard practice.

3. ASYMPTOTIC PROPERTIES

Some notations are needed. Throughout the paper, $\|\cdot\|$ is the Euclidean norm and $I(\cdot)$ is the indicator function. For simplicity of presentation, we make a notion $(\mathbf{X}, \mathbf{Y}, \varepsilon)$ and assume $(\boldsymbol{X}_i, Y_i, \varepsilon_i)$, $i \geq 1$, are independent and identically distributed (i.i.d) copies of $(\boldsymbol{X}, Y, \varepsilon)$, where $\boldsymbol{X}_i$ and ε_i are independent. Let β_0 be the true value of β. The following assumptions are needed for the consistency and asymptotic normality of the LARE estimator.

Assumption 1

ε has a continuous density $f(\cdot)$ in a neighborhood of 1.

Assumption 2

$P(\varepsilon > 0) = 1$.

Assumption 3

$\boldsymbol{X}$ is bounded, i.e, $P(\|\boldsymbol{X}\| \leq K) = 1$ for some $0 < K < \infty$, and does not concentrate on any hyperplane of $p - 1$ dimension.

Assumption 4

$E(\varepsilon + \varepsilon^{-1}) < \infty$ and $E[(\varepsilon + \varepsilon^{-1})\mathrm{sgn}(\varepsilon - 1)] = 0$.

Assumption 5

$E\{(\varepsilon + \varepsilon^{-1})^2\} < \infty$.

Assumptions 1-3 are regularity conditions. In Assumption 4, the condition on the first moment $E(\varepsilon + \varepsilon^{-1}) < \infty$ is to ensure the weak consistency of the LARE estimator. The condition $E[(\varepsilon + \varepsilon^{-1})\mathrm{sgn}(\varepsilon - 1)] = 0$ is only an identifiability condition, which plays the same role as the assumptions of zero mean and zero median for the LS and LAD methods, respectively, for linear regression. In fact, as shown in Lemma 2 in Appendix, if ε is nondegenerate and satisfies $E(\varepsilon + \varepsilon^{-1}) < \infty$, then there exists a unique scale transformation $\varepsilon_a = a \cdot \varepsilon$ such that $E\left[\left(\varepsilon_a + \varepsilon_a^{-1}\right)\mathrm{sgn}(\varepsilon_a - 1)\right] = 0$. It implies that this condition ensures the identifiability of the intercept component of the parameter β in model (2). Assumption 5 is to ensure the asymptotic normality of the LARE estimator, similar to the finite second moment assumption for the LS estimator for linear regression.

Remark 2

The first moment condition $E(\varepsilon + \varepsilon^{-1}) < \infty$ ensures consistency and the second moment condition $E\{(\varepsilon + \varepsilon^{-1})^2\} < \infty$ ensures the asymptotic normality of the LARE estimator, while the RLS estimator in Park and Stefanski (1998) requires second moment condition $E(\varepsilon^{-2}) < \infty$ for consistency.

Remark 3

These technical conditions may not be the weakest possible ones. They are imposed to facilitate the proofs. Some conditions could be relaxed for general limit theory. Knight (1998) gave a general limit theory for LAD estimation. Correspondingly, we could follow those steps to construct more general limit theory. This leaves space for future research.

Assumption 3 implies that $\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^{\intercal}$ is positive definite almost surely. By Lemma 1 in the Appendix, $LARE_n(\beta)$ is strictly convex in β under Assumption 3. Therefore, the

minimizer of $LARE_n(\beta)$, denoted as $\widehat{\beta}_n$, exists and is unique almost surely. The following theorem establishes the consistency and asymptotic normality for $\widehat{\beta}_n$.

Theorem 1

Suppose Assumptions 1-4 hold. Then, $\widehat{\beta}_n$ *converges to* β_0 *in probability as* $n \to \infty$. If, in addition to Assumptions 1-4, Assumption 5 holds, then as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow{\mathscr{D}} N\left(0, \frac{1}{4}\{J+2f(1)\}^{-2} A \boldsymbol{V}^{-1}\right).$$

where $\xrightarrow{\mathscr{D}}$ presents 'convergence in distribution', $A = E\{(\varepsilon + \varepsilon^{-1})^2\}$, $J = E\{\varepsilon \mathrm{sgn}(\varepsilon - 1)\}$ and $\mathbf{V} = E(\mathbf{X}\mathbf{X}^{\mathrm{T}})$.

Remark 4

Note that

$$2E\{\varepsilon I(\varepsilon>1)\} > E\left\{\left(\varepsilon+\varepsilon^{-1}\right) I(\varepsilon>1)\right\} = E\left\{\left(\varepsilon+\varepsilon^{-1}\right) I(\varepsilon \le 1)\right\} > 2E\{\varepsilon I(\varepsilon \le 1)\}$$

under Assumptions 1 and 4, which ensures $J > 0$. So the positivity of the density of the error in a neighborhood of 1 is not required here. It is different from the LAD estimation for linear regression models, where the positivity of the density of the error in a neighborhood of zero is essential to ensure the asymptotic normality.

Unlike the least squares estimator, the asymptotic covariance matrix involves the density function of the error terms and cannot be properly estimated using the plug-in rules. To avoid density estimation, we propose a distributional approximation based on random weighting method by externally generating *i.i.d.* random variables. Let $w_1, \ldots, w_n$ be a sequence of *i.i.d.* nonnegative random variables, with mean and variance both equal to 1. For instance, the standard exponential distribution has mean and variance equal to 1. Define

$$LARE_n^{\star}(\beta) \equiv \sum_{i=1}^{n} w_i \left\{ \left| \frac{Y_i - \exp\left(\boldsymbol{X}_i^{\mathrm{T}}\beta\right)}{Y_i} \right| + \left| \frac{Y_i - \exp\left(\boldsymbol{X}_i^{\mathrm{T}}\beta\right)}{\exp\left(\boldsymbol{X}_i^{\mathrm{T}}\beta\right)} \right| \right\},$$

and $\widehat{\beta}_n^{\star} = \arg\min_{\beta \in \mathscr{B}_0} LARE_n^{\star}(\beta)$. The distribution of $\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right)$ can be approximated by the resampling distribution of $\sqrt{n}\left(\widehat{\beta}_n^{\star} - \widehat{\beta}_n\right)$. Let $\mathscr{L}^{\star}$ denote the conditional distribution given $\{(Y_i, \boldsymbol{X}_i), i = 1, \ldots, n\}$.

Proposition 1

Suppose Assumptions 1-5 hold. Then as $n \to \infty$,

$$\mathscr{L}^{\star}\left(\sqrt{n}\left(\widehat{\beta}_n^{\star} - \widehat{\beta}_n\right)\right) \xrightarrow{\mathscr{D}} N\left(0, \frac{1}{4}\{J+2f(1)\}^{-2} A \boldsymbol{V}^{-1}\right),$$

which is the asymptotic distribution of $\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right)$, where J, A and **V** are given in Theorem 1.

The proof of Proposition 1 is similar to the proof of Theorem 1 in Chen *et al* (2008) and is omitted here. The inference procedure via resampling is as follows. First, nonnegative *i.i.d.* random weights $\{w_1, \ldots, w_n\}$ of mean one and variance one are generated M times, where M is a large number. Each time, $\widehat{\beta}_n^{\star}$ is computed. Denote them as $b_1, \ldots, b_M$. Then, the distribution of $\sqrt{n}\left(\widehat{\beta}_n^{\star} - \widehat{\beta}_n\right)$ is approximated by the empirical distribution of $\left\{\sqrt{n}\left(b_i - \widehat{\beta}_n\right), i=1, \ldots, M\right\}$.

It is known that the variance of an efficient estimator attains the Cramer-Rao lower bound. The least squares estimator and least absolute deviation estimator are efficient when the error terms follow normal distribution and double exponential distribution, respectively. In the following, we give the error distribution with which the LARE estimator is efficient.

Proposition 2

Suppose Assumption 3 holds. If the error ε has a density function as follows:

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

where c is a normalizing constant, then the estimator $\widehat{\beta}_n$ is efficient.

Remark 5

If a random variable X is distributed with density $f(x)$ in Proposition 2, then $1/X$ is equal in distribution to X.

4. SIMULATION STUDIES

Simulation studies are conducted to compare the finite sample efficiency of the least squares (LS), the least absolute deviation (LAD), the relative least squares (RLS) in which the predictor is the best mean squared relative error predictor of Y given X and our proposed least absolute relative errors (LARE) estimator. The studies are based on the model

$$Y_i = \exp\left(\beta_0 + \beta_1 X_{1_i} + \beta_2 X_{2_i}\right) \varepsilon_i, \qquad i = 1, \cdots, n, \quad (4)$$

where X_{1_i} and X_{2_i} are two independent random variables following the standard normal distribution N(0, 1), and β_0, β_1 and β_2 are the regression parameters. We consider three error distributions: ε follows the distribution with which the LARE estimator is efficient; log(ε) follows Uniform(−2, 2); and log(ε) follows $N(0, 1)$. The sample size n is 200. The variance inference is based on the random weighting and the resampling size N is 500. The simulation results are based on 1000 replications.

We get the LS and LAD estimators by minimizing $\sum_{i=1}^{n} (\log Y_i - \beta_0 - \beta_1 X_{1_i} - \beta_2 X_2)^2$ and $\sum_{i=1}^{n} |\log Y_i - \beta_0 - \beta_1 X_{1_i} - \beta_2 X_{2_i}|$ respectively. And we get the RLS estimators by minimizing $\sum_{i=1}^{n} \left[\left\{Y_i - g_i^*(X)\right\} / Y_i\right]^2$, where $g_i^*(X) = E\left(Y_i^{-1} | X_{1_i}, X_{2_i}\right) / E\left(Y_i^{-2} | X_{1_i}, X_{2_i}\right)$ for model (4) is the best mean squared relative error predictor proposed in Park and Stefanski (1998).

In the following Table 4-1, we present the average of the estimates $\widehat{\beta}_n$ the empirical standard error (SE), the average of the estimated standard errors (SEE) and coverage probabilities (CP) of 95% confidence intervals based on the resampling. Table 4-2 shows the asymptotic standard error for $\widehat{\beta}_n$.

The main findings can be summarized as follows:

- For ε follows the efficient distribution, LARE is slightly better than the LS and LAD and much better than the RLS in terms of accuracy and stability of the estimation of the regression parameters.
- For $\log(\varepsilon)$ follows uniform distribution, LARE performs considerably better than the LS, LAD and RLS.
- For $\log(\varepsilon)$ follows normal distribution, LS is efficient theoretically for linear regression models. It is seen from Tables 4-1 and 4-2 that, LARE does well with comparable results to the LS.
- For the error distributions considered in our simulation, Tables 4-1 and 4-2 show that, the SE, SEE and the asymptotic standard error of LARE estimator are generally close.

Further simulation shows that LARE is not reliable when $\log(\varepsilon)$ follows double exponential distribution. This result is not strange because Assumption 4 is not satisfied in this case. Indeed, our proposed method performs well in practical settings.

5. APPLICATIONS

The dataset to be analyzed is obtained by the Reuters 3000 Xtra which is a major tool used by financial and investment analysts worldwide. The dataset contains the monthly close stock prices for 408 firms from 2007 to 2008 and their corresponding Book Value Per Share (BVPS) and Earning Per Share (EPS) in Hong Kong Stock Exchange. The *P/B* ratio is the price-to-book ratio which is a financial ratio to compare book value of a company to its current market price. And the *P/E* ratio is the price-to-earning ratio which is also a financial ratio to measure the price paid for a share relative to the annual income or profit per share earned by the firm.

Let P_{C_i} and P_{N_i} be the current price and the price for a fixed period of time later for $i = 1,\ldots, n$, respectively. The sample size n here is 408. We consider the following model:

$$P_{N_i} = P_{C_i} \exp(\beta_0 + \beta_1 PE_i + \beta_2 PB_i)\,\varepsilon_i, \qquad i = 1, \cdots, n, \quad (5)$$

where PE_i and PB_i are the *P/E* ratio and *P/B* ratio corresponding to the current price P_{C_i}.

The purpose of this study is to analyze the stock returns by using LARE and LS to estimate $\beta = (\beta_0, \beta_1, \beta_2)$ in model (5). Table 5-1 presents the estimator $\widehat{\beta}$ for β where P_{C_i} are the monthly close prices of 2007 and P_{N_i} are the corresponding monthly close prices one year later in model (5). Table 5-2 shows summary statistics of $\widehat{\beta}_0, \widehat{\beta}_1$ and $\widehat{\beta}_2$.

The results show that, LARE and LS give similar estimates which are statistically stable. The predictor based on LARE are financially meaningful and could give better estimates for the intrinsic value of a firm. Moreover, it can be seen that the proposed estimates for β_1 which is the coefficient of *P/E* ratio in model (5) are substantially more stable than that of *P/B* ratio.

6. CONCLUDING REMARKS

This paper proposes the least absolute relative errors estimation for multiplicative model. The main point of the paper is to advocating such a criterion, which may have broader applications in financial/economic data analysis, as shown in the real example of this paper and Ye (2007), survival analysis or categorical analysis. Heuristically, in survival analysis, less accuracy in terms of absolute error may be required for predicting longer life times; and, in categorical data analysis, a category with larger percentage of observations may require more accuracy of prediction in terms of absolute error. Such consideration bears the same rationale of using relative error rather than absolute error. Our future work shall consider further extension of the method to censored data and categorical data.

The least absolute relative error criterion that we adopt in (3) is not necessarily the unique choice. There are variations such as

$$LARE_n'(\beta) \equiv \sum_{i=1}^{n} \max\left\{\left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^{\intercal}\beta\right)}{Y_i}\right|, \left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^{\intercal}\beta\right)}{\exp\left(\boldsymbol{X}_i^{\intercal}\beta\right)}\right|\right\}, \quad (6)$$

as also considered in Ye (2007). For such variations, the asymptotic theories analogous to Theorem 1 and Propositions 1 and 2 can be established without further difficulty. In this paper, we choose to present a typical one of the criterions.

For completion, we give the main results for the estimator of such variations here without proof as a note. The assumptions parallel Assumptions 1-5 in Section 3. Similar to Lemma 1 in the Appendix, one can prove that $LARE_n'(\beta)$ is strictly convex in β under Assumption 3. Therefore, there exists a unique β_n' which minimizes $LARE_n'(\beta)$ almost surely. Other than Assumptions 1-3, the following assumptions are needed for consistency and asymptotic normality for $\widetilde{\beta}_n'$ the minimizer of $LARE_n'(\beta)$.

Assumption 6

$E(\varepsilon + \varepsilon^{-1}) < \infty$ and $E\{\varepsilon^{-1}I(\varepsilon \le 1) - \varepsilon I(\varepsilon > 1)\} = 0$.

Assumption 7

$E\{\varepsilon^2 \mathrm{I}(\varepsilon > 1) + \varepsilon^{-2} I(\varepsilon \le 1)\} < \infty$.

Assumptions 6-7 play the same role as Assumptions 4-5 in Section 3. $E\{\varepsilon^{-1}I(\varepsilon \le 1) - \varepsilon I(\varepsilon > 1)\} = 0$ shares similar property as $E[(\varepsilon + \varepsilon^{-1})\mathrm{sgn}(\varepsilon - 1)] = 0$ in Section 3, which is only an identifiability condition.

Proposition 3

Suppose Assumptions 1-3 and Assumption 6 hold. Then, $\widetilde{\beta}_n'$ *converges to* β_0 *in probability as* $n \to \infty$. If, in addition to Assumptions 1-3 and Assumption 6, Assumption 7 holds, then as $n \to \infty$,

$$\sqrt{n}\left(\widetilde{\beta}_n' - \beta_0\right) \xrightarrow{\mathscr{D}} N\left(0, \frac{1}{4}\{K + f(1)\}^{-2} B V^{-1}\right),$$

where $B = E\{\varepsilon^2 \mathrm{I}(\varepsilon > 1) + \varepsilon^{-2}\mathrm{I}(\varepsilon \le 1)\}$, $K = E\{\varepsilon \mathrm{I}(\varepsilon > 1)\}$ and $\mathbf{V} = E(\mathbf{X}\mathbf{X}^{\mathrm{T}})$.

Acknowledgments

The authors are grateful to two anonymous referees, the Associate Editor and the Editor for comments and suggestions that lead to substantial improvements in the paper.

APPENDIX: PROOFS

We state two lemmas that will be used later.

Lemma 1

Let $\psi(x, a) = |1 - a^{-1}e^{x}| + |1 - ae^{-x}|$ for a > 0 and $x \in R$. Then, for fixed $a > 0$, $\psi(x, a)$ is a strictly convex function in $x \in R$.

The proof is omitted.

Lemma 2

Suppose that ξ^* is nondegenerate and $E\{\exp(\xi^*) + \exp(-\xi^*)\} < \infty$. Let $\phi(a) = E\{\exp(\xi^* - a) + \exp(a - \xi^*)\}\operatorname{sgn}(\xi^* - a)]$ and $a^* = \max\{a : \phi(a) \geq 0\}$. If $\phi(a)$ is continuous at a^*, then there exists a unique constant $a \in R$ such that $\phi(a) = 0$.

Proof

Observe the following inequality

$$\begin{aligned} &\{\exp(x-b) \quad + \exp(b-x)\}\operatorname{sgn}(x-b) \\ &\qquad - \{\exp(x-a) + \exp(a-x)\}\operatorname{sgn}(x-a) \\ &\leq \int_a^b \{-\exp \quad (x-y) + \exp(-x+y)\}\operatorname{sgn}(x-y)\,dy, \end{aligned} \tag{A.1}$$

for any x, a and $b \in R$ with $a < b$. Then,

$$\begin{aligned} \phi(b) - \phi(a) \quad &\leq E\left[\int_a^b \{-\exp(\xi^* - y) + \exp(-\xi^* + y)\}\operatorname{sgn}(\xi^* - y)\,dy\right] \\ &= \int_a^b E\left[\{-\exp(\xi^* - y) + \exp(-\xi^* + y)\}\operatorname{sgn}(\xi^* - y)\right] dy. \end{aligned} \tag{A.2}$$

It is easy to show that $\{\exp(-x + y) - \exp(x - y)\}\operatorname{sgn}(x - y) < 0$ for $x \neq y$. It follows that

$$E\left[\{-\exp(\xi^* - y) + \exp(-\xi^* + y)\}\operatorname{sgn}(\xi^* - y)\right] < 0,$$

which implies that $\phi(b) - \phi(a) < 0$. Thus, $\phi(\cdot)$ is strictly decreasing. On the other hand, it is seen from the expression $\phi(\cdot)$ that

$$\phi(a) \to -\infty \quad \text{as} \quad a \to \infty \quad \text{and} \quad \phi(a) \to \infty \quad \text{as} \quad a \to -\infty.$$

Together with the continuity of $\phi(\cdot)$ at a^*, there exists a unique solution to $\phi(a) = 0$. The proof is complete.

A.1. Proof of Theorem 1

The proof will be done in several steps.

Step 1

To prove consistency, denote

$$\psi_n(\beta) \equiv \sum_{i=1}^{n} \left[\left| 1 - \varepsilon_i^{-1} \exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} \right| + \left| 1 - \varepsilon_i \exp\left\{ -\boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} \right| \right].$$

It follows from the Convexity Lemma in Pollard (1991, p. 187) and the convexity of $\psi_n(\beta)$ by Lemma 1 that, for any compact set $\mathscr{B}$,

$$\sup_{\beta \in \mathscr{B}} \frac{1}{n} \left| \psi_n(\beta) - E\{\psi_n(\beta)\} \right| \to 0 \quad \text{(A.3)}$$

in probability as $n \to \infty$. Then,

$$\begin{aligned} & E\{\psi_n(\beta) - \psi_n(\beta_0)\} \\ = & \sum_{i=1}^{n} E\left[\left| 1 - \varepsilon_i^{-1} \exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} \right| + \left| 1 - \varepsilon_i \exp\left\{ -\boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} \right| - \left| 1 - \varepsilon_i^{-1} \right| - \left| 1 - \varepsilon_i \right| \right] \\ = & \sum_{i=1}^{n} E\left(\left(\varepsilon_i + \varepsilon_i^{-1} \right) \operatorname{sgn}(1 - \varepsilon_i) \left[\exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} - 1 \right] \right) \\ & + \sum_{i=1}^{n} E\left(\varepsilon_i \operatorname{sgn}(\varepsilon_i - 1) \left[\exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} + \exp\left\{ -\boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} - 2 \right] \right) \\ & + 2 \sum_{i=1}^{n} E\left(\left\{ I\left(\varepsilon_i \le \exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} \right) - I(\varepsilon_i \le 1) \right\} \right. \\ & \qquad \left. \left[\varepsilon_i^{-1} \exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} - \varepsilon_i \exp\left\{ -\boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} \right] \right). \end{aligned} \quad \text{(A.4)}$$

By Assumption 4, the first term in the summand is 0. It follows from Assumptions 1 and 4 that,

$$2E\{\varepsilon I(\varepsilon > 1)\} > E\left\{ \left(\varepsilon + \varepsilon^{-1} \right) I(\varepsilon > 1) \right\} = E\left\{ \left(\varepsilon + \varepsilon^{-1} \right) I(\varepsilon \le 1) \right\} > 2E\{\varepsilon I(\varepsilon \le 1)\}, \quad \text{(A.5)}$$

which implies $J = E\{\varepsilon \operatorname{sgn}(\varepsilon - 1)\} > 0$. This result leads to the fact that the second term in (A.4) is nonnegative. It is easy to check that the third term in (A.4) is also nonnegative. Hence, $E\{\psi_n(\beta) - \psi_n(\beta_0)\} \ge 0$ for all β. Furthermore, $E\{\psi_n(\beta) - \psi_n(\beta_0)\} = 0$ ensures

$$\sum_{i=1}^{n} E\left(\varepsilon_i \operatorname{sgn}(\varepsilon_i - 1) \left[\exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} + \exp\left\{ -\boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} - 2 \right] \right) = 0.$$

As $\beta = \beta_0$ is the unique minimizer of $\exp\left\{ \boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\} + \exp\left\{ -\boldsymbol{X}_i^{\tau} (\beta - \beta_0) \right\}$, it follows from Assumption 3 and $E\{\varepsilon \operatorname{sgn}(\varepsilon - 1)\} > 0$ that $\beta = \beta_0$ is the unique minimizer of $E\{\psi_n(\beta) - \psi_n(\beta_0)\}$. Denote $\psi(\beta) = n^{-1} E\{\psi_n(\beta)\}$. Then, for every $\delta > 0$, there exists $\eta > 0$ such that $\psi(\beta) > \psi(\beta_0) + \eta$ for $\|\beta - \beta_0\| \ge \delta$. For any constant δ and C, let $\widehat{\beta}_n^*$ be the minimizer of $\psi_n(\beta)$ over $\delta \le \|\beta - \beta_0\| \le C$. Then by (A.3), $\psi_n(\beta_n^*) \to \psi(\beta_n^*)$ in probability as $n \to \infty$ and $\psi(\beta_n^*) < \psi(\beta_0) + \eta$ for some $\eta > 0$. On the other hand, for any constant δ,

$$\inf_{\|\beta - \beta_0\| \le \delta} \psi_n(\beta) \le \psi_n(\beta_0) \to \psi(\beta_0)$$

in probability by (A.3). Therefore, with probability going to 1, the minimum of $\psi_n(\beta)$ in $\|\beta - \beta_0\| \le C$ is achieved inside $\|\beta - \beta_0\| \le \delta$. Since $\psi_n(\beta)$ is strictly convex, the local

minimizer inside $\|\beta-\beta_0\| \le \delta$ is the unique global minimizer. By the definition of $\widehat{\beta}_n$, $P\left(\widehat{\beta}_n \in \{\beta:\|\beta-\beta_0\| \le \delta\}\right) \to 1$ as $n \to \infty$. Thus, the weak consistency of $\widehat{\beta}_n$ is proved by letting $\delta \to 0$.

Step 2

To prove asymptotic normality, we approximate $E\{\psi_n(\beta) - \psi_n(\beta_0)\}$ for every fixed β in a neighborhood of β_0 first. Observe that $\exp(x) + \exp(x) - 2 = x^2 + O(|x|^3)$ if x closes to zero. By the Taylor expansion,

$$\begin{aligned} & \tfrac{1}{n}E\{\psi_n(\beta)-\psi_n(\beta_0)\} \\ = \; & J\cdot\tfrac{1}{n}\sum_{i=1}^{n}E\left[\exp\left\{-\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}+\exp\left\{\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}-2\right] \\ & +2f(1)\tfrac{1}{n}\sum_{i=1}^{n}E\left\{(\beta-\beta_0)^{\tau}\boldsymbol{X}_i\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}+O\left(\|\beta-\beta_0\|^3\right) \qquad \text{(A.6)} \\ = \; & J(\beta-\beta)^{\tau}\boldsymbol{V}(\beta-\beta_0)+2f(1)(\beta-\beta_0)^{\tau}\boldsymbol{V}(\beta-\beta_0)+O\left(\|\beta-\beta_0\|^3\right) \\ = \; & \{J+2f(1)\}(\beta-\beta_0)^{\tau}\boldsymbol{V}(\beta-\beta_0)+O\left(\|\beta-\beta\|^3\right), \end{aligned}$$

where $J = E\{\varepsilon\operatorname{sgn}(\varepsilon-1)\}$ and $\boldsymbol{V} = E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})$.

Step 3

Write $\boldsymbol{W}_n = \sum_{i=1}^{n}\left(\varepsilon_i+\varepsilon_i^{-1}\right)\operatorname{sgn}(\varepsilon_i-1)\boldsymbol{X}_i$. We are now in position to show

$$\sup_{\|\beta-\beta\|\le Cn^{-1/2}}|\psi_n(\beta)-\psi_n(\beta_0)+\boldsymbol{W}_n^{\tau}(\beta-\beta_0)-E\{\psi_n(\beta_0)-\psi_n(\beta_0)\}| \to 0 \qquad \text{(A.7)}$$

in probability as $n \to \infty$, for each positive constant C. To this end, let $\theta = \sqrt{n}(\beta-\beta_0)$, it is equivalent to show

$$\sup_{\|\theta\|\le C}\left|\psi_n\left(\beta_0+\frac{\theta}{\sqrt{n}}\right)-\psi_n(\beta_0)+\frac{1}{\sqrt{n}}\boldsymbol{W}_n^{\tau}\theta-E\left\{\psi_n\left(\beta_0+\frac{\beta}{\sqrt{n}}\right)-\psi_n(\beta_0)\right\}\right| \to 0 \qquad \text{(A.8)}$$

in probability as $n \to \infty$. In order to establish (A.8), we shall first show that, for each fixed θ,

$$\psi_n\left(\beta_0+\frac{\theta}{\sqrt{n}}\right)-\psi_n(\beta_0)+\frac{1}{\sqrt{n}}\boldsymbol{W}_n^{\tau}\theta-E\left\{\psi_n\left(\beta_0+\frac{\theta}{\sqrt{n}}\right)-\psi_n(\beta_0)\right\} \to 0 \qquad \text{(A.9)}$$

in probability as $n \to \infty$. Analogous to (A.4), denote

$$G_i(\beta) \equiv \varepsilon_i\operatorname{sgn}(\varepsilon_i-1)\left[\exp\left\{\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}+\exp\left\{-\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}-2\right]$$

and

$$\begin{aligned} R_i(\beta) \equiv \Big[& I\left(\varepsilon_i>\exp\left\{\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}\right) \quad -I(\varepsilon_i>1)] \\ & \left[\varepsilon_i\exp\left\{-\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}-\varepsilon_i^{-1}\exp\left\{\boldsymbol{X}_i^{\tau}(\beta-\beta_0)\right\}\right]. \end{aligned}$$

Then,

$$\begin{aligned} & \psi_n(\beta) - \psi_n(\beta_0) - E\{\psi_n(\beta) - \psi_n(\beta_0)\} \\ = & -\sum_{i=1}^{n} \left(\varepsilon_i + \varepsilon_i^{-1}\right) \operatorname{sgn}(\varepsilon_i - 1) \left[\exp\left\{\boldsymbol{X}_i^{\tau}(\beta - \beta_0)\right\} - 1\right] \\ & + \sum_{i=1}^{n} \{G_i(\beta) - EG_i(\beta)\} + 2\sum_{i=1}^{n} \{R_i(\beta) - ER_i(\beta)\}. \end{aligned}$$

For each fixed θ,

$$\begin{aligned} & \sum_{i=1}^{n} E\left[G_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right) - E\left\{G_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right)\right\}\right]^2 \\ \leq & \sum_{i=1}^{n} E\{\varepsilon_i \operatorname{sgn}(\varepsilon_i - 1)\}^2 E\left\{\exp\left(-\frac{1}{\sqrt{n}}\boldsymbol{X}_i^{\tau}\theta\right) + \exp\left(\frac{1}{\sqrt{n}}\boldsymbol{X}_i^{\tau}\theta\right) - 2\right\}^2 \quad \text{(A.10)} \\ = & \sum_{i=1}^{n} E\{\varepsilon_i \operatorname{sgn}(\varepsilon_i - 1)\}^2 E\left(\frac{1}{n}\theta^{\tau}, \boldsymbol{X}_i\boldsymbol{X}_i^{\tau}\theta + a_i\right)^2, \text{ say} \\ \to & 0 \end{aligned}$$

as $n \to \infty$, where $P(\|a_i\| \leq cn^{-3/2}) = 1$ for some constant c and $i = 1, \ldots, n$. It then follows that

$$\sum_{i=1}^{n} \left[G_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right) - E\left\{G_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right)\right\}\right] \to 0 \quad (11)$$

in probability as $n \to \infty$. On the other hand, by the Taylor expansion, for each fixed θ,

$$\begin{aligned} & E\left\{\varepsilon \exp\left(-\frac{1}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta\right) - \varepsilon^{-1}\exp\left(\frac{1}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta\right)\right\}^2 \\ = & E\left(-\varepsilon\frac{1}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta - \varepsilon^{-1}\frac{1}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta + \varepsilon - \varepsilon^{-1} + b\right)^2 \\ = & E\left\{-(\varepsilon - 1)\frac{1}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta - \left(\varepsilon^{-1} - 1\right)\frac{1}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta - \frac{2}{\sqrt{n}}\boldsymbol{X}^{\tau}\theta + (\varepsilon - 1) - \left(\varepsilon^{-1} - 1\right) + b\right\}^2 \\ \leq & E\left[2\left\{(\varepsilon - 1)^2 + \left(\varepsilon^{-1} - 1\right)^2 + 4\right\}\frac{1}{n}\theta^{\tau}\boldsymbol{X}\boldsymbol{X}^{\tau}\theta + 2(\varepsilon - 1)^2 + 2\left(\varepsilon^{-1} - 1\right)^2 + b^2\right], \text{ say} \end{aligned}$$

where $P(\|b\| \leq cn^{-1}) = 1$ for some constant c. Hence, an argument similar to (A.10) leads to

$$\begin{aligned} & \sum_{i=1}^{n} E\left[R_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right) - E\left\{R_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right)\right\}\right]^2 \\ \leq & \sum_{i=1}^{n} E\left\{\left\{I\left(\frac{1}{\sqrt{n}}\boldsymbol{X}_i^{\tau}\theta > 0\right) I\left(0 < \log\varepsilon_i \leq \frac{1}{\sqrt{n}}\boldsymbol{X}_i^{\tau}\theta\right)\right.\right. \\ & \left.\left. + I\left(\frac{1}{\sqrt{n}}\boldsymbol{X}_i^{\tau}\theta \leq 0\right) I\left(0 \geq \log\varepsilon_i > \frac{1}{\sqrt{n}}\boldsymbol{X}_i^{\tau}\theta\right)\right\}\right. \\ & \left.\left[2\left\{(\varepsilon_i - 1)^2 + \left(\varepsilon_i^{-1} - 1\right)^2 + 4\right\}\frac{1}{n}\theta^{\tau}\boldsymbol{X}_i\boldsymbol{X}_i^{\tau}\theta + 2(\varepsilon_i - 1)^2 + 2\left(\varepsilon_i^{-1} - 1\right)^2 + b_i^2\right], \text{ say} \right. \\ \to & 0 \end{aligned}$$

as $n \to \infty$, where $P(\|b_i\| \leq cn^{-1}) = 1$ for some constant c and $i = 1, \ldots, n$. Thus, for each fixed θ,

$$\sum_{i=1}^{n} \left[R_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right) - E\left\{R_i\left(\beta_0 + \frac{\theta}{\sqrt{n}}\right)\right\}\right] \to 0 \quad \text{(A.12)}$$

in probability as $n \to \infty$. Combining (A.11) and (A.12), together with Assumption 4, we have shown (A.9).

Next, $\psi_n\left(\beta_0+\theta/\sqrt{n}\right)-\psi_n(\beta_0)+\boldsymbol{W}_n^{\top}\theta/\sqrt{n}$ is convex by Lemma 1. It follows from (A.9) and the Convexity Lemma in Pollard (1991, p. 187) that, for each constant $C>0$,

$$\sup_{\|\theta\|\le C}\left|\psi_n\left(\beta_0+\frac{\theta}{\sqrt{n}}\right)-\psi_n(\beta_0)+\frac{1}{\sqrt{n}}\boldsymbol{W}_n^{\top}\theta-E\left\{\psi_n\left(\beta_0+\frac{\theta}{\sqrt{n}}\right)-\psi_n(\beta_0)\right\}\right|\to 0$$

in probability. Then (A.7) is proved.

Step 4

Let $\xi_n(\beta)=\psi_n(\beta)-\psi_n(\beta_0)-n\{J+2f(1)\}(\beta-\beta_0)^{\top}\boldsymbol{V}(\beta-\beta_0)+\boldsymbol{W}_n^{\top}(\beta-\beta_0)$. Combining step 2 and step 3, we have

$$\sup_{\|\beta-\beta_0\|\le Cn^{-1/2}}|\xi_n(\beta)|\to 0 \qquad \text{(A.13)}$$

in probability as $n\to\infty$ for each constant $C>0$. Let $\widehat{\beta}_n^*$ be the minimizer of $n\{J+2f(1)\}(\beta-\beta_0)^{\top}\boldsymbol{V}(\beta-\beta_0)-\boldsymbol{W}_n^{\top}(\beta-\beta_0)$. Clearly $\widehat{\beta}_n^*-\beta_0=\{J+2f(1)\}^{-1}\boldsymbol{V}^{-1}\boldsymbol{W}_n/(2n)$. By the definition of $\boldsymbol{W}_n$, for every $\delta>0$, there exist some constants $K\delta>0$ and N_δ, such that $P\left(\|\widehat{\beta}_n^*-\beta\|>K_\delta n^{-1/2}\right)\le\delta/2$ for any $n\ge N_\delta$. In view of (A.13), for every $\eta>0$, there exists some constant N_η such that, for any $n\ge N_\eta$,

$$P\left(\sup_{\|\beta-\beta_0\|\le K_\delta n^{-1/2}}|\xi_n(\beta)|>\eta\right)\le\delta/2.$$

Hence, for every $\delta,\ \eta>0$, there exists $N=\max\{N_\delta, N_\eta\}$ such that, for any $n\ge N$,

$$\begin{aligned}P\left(|\xi_n(\widehat{\beta}_n^*)|>\eta\right) &= P\left(|\xi_n(\widehat{\beta}_n^*)|>\eta, \|\widehat{\beta}_n^*-\beta_0\|>K_\delta n^{-1/2}\right)\\ &\quad +P\left(|\xi_n(\widehat{\beta}_n^*)|>\eta, \|\widehat{\beta}_n^*-\beta_0\|\le K_\delta n^{-1/2}\right)\\ &\le P\left(\|\widehat{\beta}_n^*-\beta_0\|>K_\delta n^{-1/2}\right)+P\left(\sup_{\|\beta-\beta_0\|\le K_\delta n^{-1/2}}|\xi_n(\beta)|>\eta\right)\\ &\le \delta,\end{aligned}$$

which implies $\xi_n(\widehat{\beta}_n^*)=o_p(1)$. Similar arguments also lead to

$$\sup_{\|\beta-\widehat{\beta}_n^*\|\le Cn^{-1/2}}|\xi_n(\beta)|=o_p(1)$$

for each constant $C>0$.

Observe that

$$\psi_n(\beta) - \psi_n(\beta_0) = n\{J+2f(1)\}\left(\beta - \widehat{\beta}_n^*\right)^\tau \boldsymbol{V}\left(\beta - \widehat{\beta}_n^*\right) - \tfrac{1}{4n}\{J+2f(1)\}^{-1}\boldsymbol{W}_n^\tau \boldsymbol{V}^{-1}\boldsymbol{W}_n + \xi_n(\beta).$$

For any constants c and C with $0 < c < C < \infty$,

$$\begin{aligned} & \inf_{cn^{-1/2} \le \|\beta - \widehat{\beta}_n^*\| \le Cn^{-1/2}} \{\psi_n(\beta) - \psi_n(\beta_0)\} \\ \le & \inf_{cn^{-1/2} \le \|\beta - \widehat{\beta}_n^*\| \le Cn^{-1/2}} \left[n\{J+2f(1)\}\left(\beta - \widehat{\beta}_n^*\right)^\tau \boldsymbol{V}\left(\beta - \widehat{\beta}_n^*\right) - \tfrac{1}{4n}\{J+2f(1)\}^{-1}\boldsymbol{W}_n^\tau \boldsymbol{V}^{-1}\boldsymbol{W}_n\right] - \sup_{cn^{-1/2} \le \|\beta - \widehat{\beta}_n^*\| \le Cn^{-1/2}} |\xi_n(\beta)| \\ \ge & \{J+2f(1)\}c^2\lambda - \tfrac{1}{4n}\{J+2f(1)\}^{-1}\boldsymbol{W}_n^\tau \boldsymbol{V}^{-1}\boldsymbol{W}_n + o_p(1), \end{aligned} \tag{A.14}$$

where λ is the smallest eigenvalue of $\boldsymbol{V}$. On the other hand, for any constant c,

$$\begin{aligned} \inf_{\|\beta - \widehat{\beta}_n^*\| \le cn^{-1/2}} \{\psi_n(\beta) - \psi_n(\beta_0)\} & \le \psi_n\left(\widehat{\beta}_n^*\right) - \psi_n(\beta_0) \\ & = -\tfrac{1}{4n}\{J+2f(1)\}^{-1}\boldsymbol{W}_n^\tau \boldsymbol{V}^{-1}\boldsymbol{W}_n + \xi_n\left(\widehat{\beta}_n^*\right) \\ & = -\tfrac{1}{4n}\{J+2f(1)\}^{-1}\boldsymbol{W}_n^\tau \boldsymbol{V}^{-1}\boldsymbol{W}_n + o_p(1). \end{aligned} \tag{A.15}$$

Both (A.14) and (A.15) together imply that, with probability going to 1, the minimum of $\psi_n(\beta) - \psi_n(\beta_0)$ in $\|\beta - \widehat{\beta}_n^*\| \le Cn^{-1/2}$ is achieved inside $\|\beta - \widehat{\beta}_n^*\| \le cn^{-1/2}$. Since $\psi_n(\beta) - \psi_n(\beta_0)$ is convex, the local minimizer inside $\|\beta - \widehat{\beta}_n^*\| \le cn^{-1/2}$ is the global minimizer. Thus,

$$\begin{aligned} \widehat{\beta}_n - \beta_0 & = \widehat{\beta}_n^* - \beta_0 + o_p\left(n^{-1/2}\right) \\ & = \tfrac{1}{2\sqrt{n}}\{J+2f(1)\}^{-1}\boldsymbol{V}^{-1}\tfrac{1}{\sqrt{n}}\boldsymbol{W}_n + o_p\left(n^{-1/2}\right). \end{aligned}$$

Hence, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{4}\{J+2f(1)\}^{-2}A\boldsymbol{V}^{-1}\right).$$

A.2. Proof of Proposition 2

For the given density of ε, the density of Y_i given $\boldsymbol{X}_i$ is

$$f_{Y_i|\boldsymbol{X}_i}(y) = c\exp\left\{-\left|\frac{\exp\left(\boldsymbol{X}_i^\tau\beta\right) - y}{\exp\left(\boldsymbol{X}_i^\tau\beta\right)}\right| - \left|\frac{y - \exp\left(\boldsymbol{X}_i^\tau\beta\right)}{y}\right| - \log y\right\}.$$

Then, the likelihood function of Y is

$$L(\beta) = c^n \exp\left[-\sum_{i=1}^n \left\{\left|\frac{\exp\left(\boldsymbol{X}_i^\tau\beta\right) - Y_i}{\exp\left(\boldsymbol{X}_i^\tau\beta\right)}\right| - \left|\frac{Y_i - \exp\left(\boldsymbol{X}_i^\tau\beta\right)}{Y_i}\right| - \log Y_i\right\}\right].$$

Maximizing this likelihood function is equivalent to minimizing our proposed LARE criterion

$$\sum_{i=1}^{n}\left\{\left|\frac{\exp\left(\boldsymbol{X}_i^{\top}\beta\right)-Y_i}{\exp\left(\boldsymbol{X}_i^{\top}\beta\right)}\right|+\left|\frac{Y_i-\exp\left(\boldsymbol{X}_i^{\top}\beta\right)}{Y_i}\right|\right\}.$$

Therefore the estimator $\widehat{\beta}_n$, which minimizes $LARE_n(\beta)$, is efficient when $\varepsilon \sim \mathrm{f}(\cdot) = c \exp(-|1-x| - |1-x^{-1}| - \log x)I(x > 0)$. The proof is complete.

References

Chen K, Ying Z, Zhang H, Zhao L. Analysis of least absolute deviation. Biometrika. 2008; 95:107–122.

Gauss, CF. Theoria Motus Corporum Coelestium. Perthes; Hamburg: 1809. Theory of the Motions of the Heavenly Bodies Moving about the Sun in Conic Sections. Dover; New York: 1963. Translation reprinted as

Khoshgoftaar TM, Bhattacharyya BB, Richardson GD. Predicting software errors, during development, using nonlinear regression models: a comparative study. IEEE Transactions on Reliability. 1992; 41:390–395.

Knight K. Limiting distribution for L_1 regression estimators under general conditions. the Annals of Statistics. 1998; 26:755–770.

Makridakis, S.; Andersen, A.; Carbone, R.; Fildes, R.; Hibon, M.; Lewandowski, R.; Newton, J.; Parzen, E.; Winkler, R. The Forecasting Accuracy of Major Time Series Methods. Wiley; New York: 1984.

Narula SC, Wellington JF. Prediction, linear regression and the minimum sum of relative errors. Technometrics. 1977; 19:185–190.

Park H, Stefanski LA. Relative-error prediction. Statistics and Probability Letters. 1998; 40:227–236.

Pollard D. Asymptotics for least absolute deviations regression estimators. Econometric Theory. 1991; 7:186–199.

Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators (with discussion). Statistical Science. 1997; 12:279–300.

Stigler SM. Gauss and the Invention of Least Squares. the Annals of Statistics. 1981; 9:465–474.

Ye J. Price models and the value relevance of accounting information. Technical report. 2007

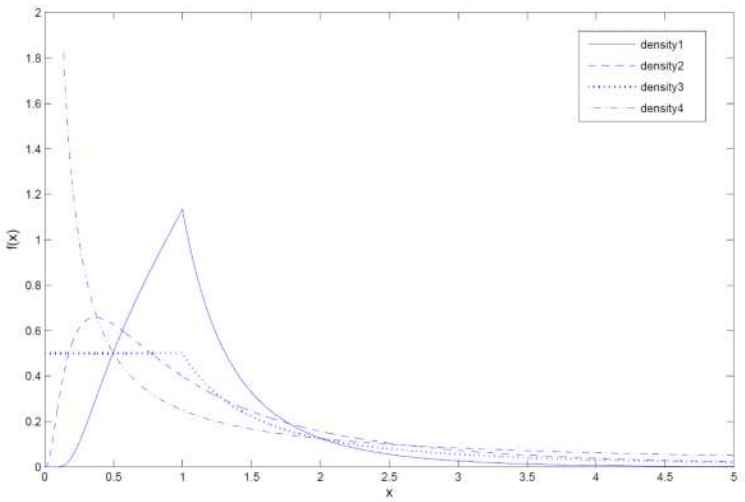

Figure 1.
Plot of four densities.
density1: $f(x) = c\exp(-|1 - x| - |1 - x^{-1}| - \log x)I(x > 0)$.
density2: the density of ε where $\log(\varepsilon) \sim N(0, 1)$.
density3: the density of ε where $\log(\varepsilon) \sim$ Double Exponential(0, 1).
density4: the density of ε where $\log(\varepsilon) \sim$ Uniform(−2, 2).

Table 4-1

Comparison among various approaches with $\beta = (1, 1, 1)^{\mathrm{T}}$

		$\varepsilon \sim f(\cdot)^{\dagger}$			$\log(\varepsilon) \sim Unif(-2,2)$			$\log(\varepsilon)\ \varepsilon\ N(0, 1)$		
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
LARE	BIAS	0.001	0.002	0.001	0.001	0.002	0.000	0.004	0.004	0.002
	SE	0.032	0.033	0.034	0.077	0.075	0.073	0.076	0.073	0.076
	SEE	0.033	0.034	0.034	0.075	0.075	0.075	0.073	0.072	0.072
	CP	0.945	0.944	0.951	0.944	0.943	0.959	0.926	0.928	0.931
LS	BIAS	0.001	0.002	0.001	0.001	0.002	0.000	0.004	0.003	0.002
	SE	0.035	0.035	0.037	0.083	0.081	0.078	0.071	0.069	0.072
	SEE	0.035	0.035	0.035	0.081	0.080	0.080	0.070	0.069	0.070
	CP	0.945	0.952	0.926	0.948	0.937	0.951	0.950	0.939	0.935
LAD	BIAS	0.001	0.002	0.001	0.001	0.004	0.001	0.004	0.003	0.001
	SE	0.033	0.034	0.034	0.143	0.140	0.135	0.090	0.085	0.090
	SEE	0.036	0.038	0.038	0.145	0.144	0.144	0.093	0.094	0.094
	CP	0.938	0.915	0.921	0.897	0.868	0.888	0.917	0.907	0.906
RLS	BIAS	0.145	0.010	0.003	0.268	0.004	0.002	0.071	0.001	0.003
	SE	1.231	0.269	0.180	1.663	0.253	0.243	1.414	0.286	0.283
	SEE	0.692	0.144	0.143	0.818	0.173	0.167	0.822	0.216	0.215
	CP	0.854	0.889	0.872	0.925	0.921	0.925	0.660	0.708	0.751

†Note: $f(x) = c \exp(-|1 - x| - |1 - x^{-1}| - \log x) I(x > 0)$.

J Am Stat Assoc. Author manuscript; available in PMC 2013 September 04.

Table 4-2

Asymptotic standard errors for estimators of β

	$\varepsilon \sim f(\cdot)^{\dagger}$			$\log(\varepsilon) \sim Unif(-2,2)$			$\log(\varepsilon) \sim N(0, 1)$		
LARE	0.030	0.030	0.030	0.074	0.074	0.074	0.075	0.075	0.075
LS	0.035	0.035	0.035	0.082	0.082	0.082	0.071	0.071	0.071
LAD	0.031	0.031	0.031	0.141	0.141	0.141	0.089	0.089	0.089

†Note: $f(x) = c \exp(-|1 - x| - |1 - x^{-1}| - \log x) I(x > 0)$.

Table 5-1

Comparison of regression coefficients: LARE vs LS

	LARE			LS		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
JAN	1.0549	−0.0002	−0.0171	1.0840	0.0001	−0.0210
FEB	1.1976	0.0004	−0.0222	1.2244	0.0002	−0.0241
MAR	1.3198	−0.0009	−0.0168	1.3243	−0.0005	−0.0227
APR	0.9157	−0.0012	−0.0085	0.8961	−0.0007	−0.0092
MAY	0.6336	−0.0009	−0.0067	0.6291	−0.0006	−0.0069
JUN	0.6461	−0.0007	−0.0078	0.6330	−0.0004	−0.0071
JUL	0.4676	−0.0007	−0.0061	0.4478	−0.0005	−0.0056
AUG	0.2313	−0.0003	−0.0053	0.2838	−0.0004	−0.0048
SEP	0.0623	−0.0002	−0.0039	0.0844	−0.0002	−0.0031
OCT	0.0106	−0.0000	−0.0040	0.0373	−0.0001	−0.0038
NOV	−0.1079	−0.0000	−0.0034	−0.1060	−0.0002	−0.0035
DEC	−0.1429	−0.0003	−0.0014	−0.1442	−0.0003	−0.0033

Table 5-2

Summary statistics: LARE vs LS

		Min	Max	Mean	Stdev	Median	10th Percentile	90th Percentile
LARE	$\hat{\beta}_0$	−0.1429	1.3198	0.5241	0.5186	0.5506	−0.1079	1.1976
	$\hat{\beta}_1$	−0.0012	0.0004	−0.0004	0.0005	−0.0003	−0.0009	0.0000
	$\hat{\beta}_2$	−0.0222	−0.0014	−0.0086	0.0065	−0.0064	−0.0171	−0.0034
LS	$\hat{\beta}_0$	−0.1442	1.3243	0.5328	0.5172	0.5384	−0.1060	1.2244
	$\hat{\beta}_1$	−0.0007	0.0002	−0.0003	0.0003	−0.0004	−0.0006	0.0001
	$\hat{\beta}_2$	−0.0241	−0.0031	−0.0096	0.0081	−0.0063	−0.0227	−0.0033