UC Berkeley
Other Recent Work

Title
The Sampling Distribution of the Least Absolute Residuals Regression Estimates

Permalink
https://escholarship.org/uc/item/8nv4g3mj

Authors
Rosenberg, Barr
Carlson, Daryl

Publication Date
1971-11-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

THE COMMITTEE ON ECONOMETRICS AND MATHEMATICAL ECONOMICS

Working Paper No. IP-164

THE SAMPLING DISTRIBUTION OF LEAST ABSOLUTE RESIDUALS REGRESSION ESTIMATES

by

Barr Rosenberg and Daryl Carlson

November 1971
Revised

Note: The authors are Assistant Professor and graduate student at the School of Business Administration, University of California, Berkeley. Much of this work was completed while the senior author was supported by an NSF graduate fellowship at Harvard University. Later stages of the work were supported by NSF Grants GS-2102 and 3306 through the Institute of Business and Economic Research, University of California, Berkeley. The authors gratefully acknowledge the helpful comments of Professors L. D. Taylor, H. S. Houthakker, R. Dobell, J. B. Ramsey, C. Kelso, and M. Stefanski. An earlier version of this article was presented at the Econometric Society Meetings, December, 1970.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
CENTER FOR RESEARCH IN MANAGEMENT SCIENCE
University of California, Berkeley

THE SAMPLING DISTRIBUTION OF LEAST ABSOLUTE RESIDUALS REGRESSION ESTIMATES

by

Barr Rosenberg and Daryl Carlson

Regression by minimization of the sum of the absolute values of the residuals (LAR) *is shown to be preferable to least squares regression* (LS) *when the disturbance distribution has massive tails. For the special case of a single regressor, the exact sampling distribution of the* LAR *estimation error is derived. For multivariate regression with a symmetric disturbance distribution, the* LAR *estimation error is approximately multivariate normally distributed with mean zero and variance matrix* $\lambda(X'X)^{-1}$, *where* T *is the sample size and* λ/T *is the variance of the median of a sample of size* T *from the disturbance distribution. The approximate sampling theory is validated by extensive Monte Carlo studies.*

I. INTRODUCTION

This article is concerned with the familiar linear regression model:

$$(1.1) \qquad y_t = \sum_{i=1}^{K} x_{it}\beta_i + u_t = \underset{\sim}{x}_t\underset{\sim}{\beta} + u_t \qquad t=1,\ldots,T$$

where $\underset{\sim}{x}_t$ is the row vector of explanatory variables for observation t, and $\underset{\sim}{\beta}$ is the column vector of parameters, or in matrix form:

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{u},$$

where the disturbances are independently distributed according to some probability distribution yet to be specified and are independent of the explanatory variable. The Least Squares (LS) estimate of the regression coefficient vector is the vector, $\hat{\underset{\sim}{b}}_{LS}$, which minimizes the sum of squared residuals $\sum_{t=1}^{T} r_t^2$, where $r_t = y_t - \underset{\sim}{x}_t\hat{\underset{\sim}{b}}_{LS}$, t=1,...,T. A general "least-alpha" estimate may be defined analogously as that estimate $\hat{\underset{\sim}{b}}$ which minimizes

$$(1.2) \qquad S_\alpha = \sum_{t=1}^{T} |r_t|^\alpha, \qquad \text{for} \qquad \alpha > 0.$$

The estimator for $\alpha = 1$ ——→ ——→ has been variously termed the "Minimum Absolute Deviations" (MAD) [1], "Minimum Sum of Absolute Errors" (MSAE) [19,20,25], "Minimum Deviations" (MD) [21], "Least Deviations" (LD) [16], "Least Absolute" (LA) [14], "L_1" (since the least-alpha estimator is the outcome of projection of the dependent on the explanatory variables with Euclidean norm L_α) [2], and

"Least Absolute Residuals" (LAR) [23] Estimator. The last term, LAR, being perhaps the clearest and most analogous to the accepted terminology for Least Squares estimation, will be used here.

In this paper several theorems concerning the sampling distribution of the LAR estimator are derived and an approximate sampling theory, which is closely analogous to the LS sampling theory, is proposed. In multivariate regression with a disturbance distribution having mean zero and variance σ^2, the familiar sampling distribution for the LS estimators is

$$A_{LS}: \quad \hat{\underset{\sim}{\beta}}_{LS} \sim \eta(\underset{\sim}{\beta}, \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}) \ .$$

This is the exact distribution when the disturbances are normally distributed, and the large-sample approximation when the disturbances have any finite-variance distribution. Provided that the disturbance distribution is symmetric, the proposed approximate sampling theory for LAR is

$$A_{LAR}: \quad \hat{\underset{\sim}{\beta}}_{LAR} \sim \eta(\underset{\sim}{\beta}, \lambda(F,T)(\underset{\sim}{X}'\underset{\sim}{X})^{-1}),$$

where $\lambda(F,T) = T\sigma^2_{MED}$, and σ^2_{MED} is the variance of the median of a sample of size T from the disturbance distribution F. Since $\sigma^2 = T\sigma^2_{MEAN}$, the LAR approximate distribution theory differs from LS only in that the variance of the sample median replaces the variance of the sample mean. A similar approach can be applied to the sampling distributions of the remaining least-alpha estimators, but we omit the details since the computational difficulties associated with these estimators appear to render their use impractical.

In the balance of this section, some background on the "least-alpha" family of estimators is provided. In Section II, the LAR parameter estimates are expressed as a function of the explanatory variables and the unobserved disturbances, and several Theorems concerning the basic properties of the error distribution are deduced. Then, for the special case of a single regressor, a nearly complete sampling theory is derived in Section III. When the disturbance distribution is symmetric, the existence of moments for the estimation error distribution is simply related to the existence of the moments of the disturbance distribution and to the sample size. The LAR estimator is highly robust and, in particular, LAR will have finite variance for common sample sizes when the disturbances have the infinite-variance stable Paretian distribution. Moreover, the LAR estimates are consistent for a wide class of disturbance distributions. In Section IV, several easily computed approximations to the exact sampling theory are evaluated by Monte Carlo studies. Section V introduces the difficult case of several regressors, and in Section VI A_{LAR} is validated by extensive Monte Carlo studies. Some remarks concerning the use of LAR estimates complete the article.

1.1 The members of the least-alpha family of estimators can be compared on the basis of three criteria: computational economy, the precision of the estimators, and the richness of the distribution theory available for the estimators.

With regard to computational economy, the LS estimator is far superior, being the only estimator that is linear in the observations. In general, a least-alpha estimator minimizes

$$S_\alpha = \sum_{t=1}^{T} |r_t|^\alpha$$

subject to the constraints

$$\begin{aligned}
x_{11} b_1 + \ldots + x_{K1} b_K + r_1 &= y_1 \\
x_{12} b_1 + \ldots + x_{K2} b_K + r_2 &= y_2 \\
\vdots \qquad\qquad \vdots \qquad\qquad \ddots \qquad &\vdots \\
x_{1T} b_1 + \ldots + x_{KT} b_K + r_T &= y_T \,.
\end{aligned}$$

When alpha equals 1, the objective function is linear, so the LAR estimate can be computed by the familiar simplex algorithm for linear programming.[1] Several important simplifications in the dual problem result from the simple structure of the constraint tableau [28,7,23]. A program prepared by the present authors may be obtained on request. Computation time for a seven-variable regression with eighty observations on the CDC 6400 is 2.866 seconds, compared to 0.348 seconds for an LS regression program which computed the variance-covariance matrix of estimation errors. Another computational approach that is claimed to be still more efficient [26,27] has come into use recently [14].

Since the LAR estimate is computed in a linear programming framework, inequality constraints on the coefficients and differing weights for

---

[1] Edgeworth, who suggested LAR as an alternative to LS in 1887 [10,11], proposed a geometric algorithm which was improved by Rhodes [21] and Singleton [24]. Wagner [28] developed the linear programming solution.

positive and negative residuals in the objective function (appropriate for asymmetric loss functions) are introduced at negligible cost. However, the use of quadratic loss functions to introduce prior estimates necessitates quadratic programming. For LS, in contrast, quadratic loss functions (the degenerate case of which is an exact linear constraint) can easily be used for prior information on the parameters, but inequality constraints and asymmetry require quadratic programming. In sum, LAR is computationally more difficult than LS, but not prohibitively so.

The remaining estimators in the least-alpha family require nonlinear programming and do not yield the same simplified dual problem. Accordingly, the required computation time may be as much as several orders of magnitude greater than for LAR.[2] Thus, LAR stands next after LS in computational simplicity.

1.2 With regard to the accuracy of the estimators, LS is ideal in the presence of normally distributed disturbances, where it is the maximum-likelihood and minimum-variance unbiased estimator from the classical viewpoint and the mean for the posterior distribution of the parameters from the Bayesian standpoint. However, this superiority of LS does not extend to other disturbance distributions. The least-alpha estimator is the maximum-likelihood estimator for the disturbance distribution:[3]

$$f(u) \sim \exp\left[-\left(\frac{|u|}{S}\right)^{\alpha}\right]. \tag{1.3}$$

---

[2] One exception is the case alpha = ∞ (the Chebychev estimator), where the maximal absolute residual is minimized [3]. However, this estimator has the very worst sampling characteristics for disturbance distributions with massive tails, and hence does not compete with LAR in those applications where LAR is shown below to be superior to LS.

[3] For a more complete discussion of this point, see the recent paper of Zeckhauser and Thompson [30].

For alpha equal to two, this is the normal distribution, and LS is the associated ML estimator. For alpha equal to one, this is the Laplace distribution, and LAR is the ML estimator. As $\alpha \to 0$, the distribution becomes increasingly peaked, with a small cusp at $u = 0$ and with extremely flat tails extending from this cusp to infinity; in response, the objective function becomes sensitive only to those residuals with smallest absolute value, and the estimator approaches the mode. Conversely, as $\alpha \to \infty$, the distribution approaches the uniform distribution with a flat plateau between $u = -S$ and $u = +S$, and a near-zero level outside of this interval; in response, the objective function becomes sensitive only to those residuals with largest magnitude, and the estimator approaches the mid-range or Chebychev estimator. Moving along the continuum from $\alpha = \infty$ toward $\alpha = 0$, the tails of the distributions become relatively more massive, and the weight accorded to large residuals by the least-alpha estimator declines. LAR, being nearer the massive-tailed case, can be expected to outperform LS when there is a high probability of large residuals or outliers in a data series [2].

When the only explanatory variable in the regression is a constant--that is, where the central tendency of the population of dependent variables is to be estimated--LS is the sample average and LAR is the sample median. Thus, the four cases, $\alpha \to 0$, $\alpha = 1$, $\alpha = 2$, and $\alpha \to \infty$, correspond to the mode, median, mean, and mid-range estimators.

For a continuous disturbance distribution, where the mode is not well defined, the median is the preferred estimator when the tails of the distribution are massive. For instance, in a recent Monte Carlo study, Fama and Roll [12] found that for nonnormal members of the stable Paretian distribution, the error variance of the median is considerably smaller. Since these distributions have been suggested as appropriate to describe the probability distribution of price changes in speculative markets [5: 297-332], the Monte Carlo results have been interpreted as justification for the use of LAR in this context.

Several empirical studies over the past few years have illustrated the superiority of LAR estimates over LS estimates in forecasting applications. For example, in a study by Meyer and Glauber, various quarterly investment models were estimated by both LAR and LS [19]. These estimates were then evaluated over a seven-quarter forecast period, and the LAR estimates were superior for five out of six of the investment models, both on a sum-of-squared-forecast-errors and a sum-of-absolute-forecast-errors criterion. Similarly, in a study by Richard Oveson, with the Houthakker-Taylor consumption equations, the LAR estimates were superior over the forecast-evaluation period [20]. In a slightly different application,

John Wiginton used both LAR and LS to estimate the parameters of the Sharpe diagonal "portfolio-selection" model [29]. His results indicated that the model utilizing the LAR estimates yielded portfolios that performed better, with respect to a return-versus-risk criterion over a forecast period, than the model using the LS estimates. It is probable that in at least some of these studies the models were poorly specified, and that the superior forecasting performance of LAR was due to the resulting unstable disturbance distribution. However, the fact remains that LAR was more robust in the presence of this misspecification and, in many areas of applied statistics, data limitations render misspecification unavoidable.

1.3 Finally we come to the question of available sampling theory for the estimates. Here, again, LS has had the dominant advantage. Since the estimation error is linear in the disturbance terms, the distribution of the estimates is most simply related to the distribution of disturbances. In contrast, the sampling theory for the other least-alpha estimates has been virtually nil. Several Monte Carlo studies [1,4,13,14] have reached rather tentative conclusions about the relative error moments of LAR and LS. In the special case of the median, the distribution has been derived directly [17, Vol. I, Ch. 14; 15].

In this article, the sampling distribution of the LAR estimates is described fairly satisfactorily for the case where the disturbance distribution is symmetric. The results of the Monte Carlo studies suggest that the approximate sampling theory will be satisfactory for many applications.

II. BASIC PROPERTIES OF THE LAR ESTIMATOR

The first step in developing a distribution theory for LAR estimators is to express the estimates as a function of the observed explanatory variables and the unobserved disturbances. This function maps each possible vector of disturbances onto a vector of LAR parameter estimates. Any probability distribution of the disturbances determines, through this mapping, a probability distribution for the parameter estimates. In the case of least squares, the function relating estimates to disturbances has the explicit form:

$$(2.1) \qquad \hat{\underset{\sim}{b}}_{LS} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim}{y} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'(\underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{u}) = \underset{\sim}{\beta} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim}{u}.$$

For LAR estimation, there is no explicit expression of this kind. The LAR parameter estimates are defined implicitly as the vector $\underset{\sim}{b}$ which minimizes

$$S(\underset{\sim}{b}) = \sum_{t=1}^{T} |y_t - \underset{\sim}{x}_t\underset{\sim}{b}|.$$

A problem of nonuniqueness can arise, since $S(\underset{\sim}{b})$ may achieve its minimum, not at a single point, $\underset{\sim}{b}$, but on a subset of any dimension up to K of B, the space of all estimates, $\underset{\sim}{b}$. For the moment, we will consider any vector $\underset{\sim}{b}$ to be an LAR estimator if it minimizes $S(\underset{\sim}{b})$, not requiring the estimator to be unique.

$S(\underset{\sim}{b})$ can be rewritten in terms of $\underset{\sim}{X}$ and $\underset{\sim}{u}$ as:

$$(2.2) \qquad S(\underset{\sim}{b}) = \sum_{t=1}^{T} |(\underset{\sim t\sim}{x}\beta + u_t) - \underset{\sim t\sim}{x}\underset{\sim}{b}| = \sum_{t=1}^{T} |\underset{\sim t}{x}(\underset{\sim}{b}-\beta) - u_t|$$

$$= \sum_{t=1}^{T} |\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t| \equiv \sum_{t=1}^{T} O_t(\underset{\sim}{e}) \equiv O(\underset{\sim}{e})$$

where $\underset{\sim}{e}$ is the vector of errors in the parameter estimates and $O_t(\underset{\sim}{e})$ is the absolute value of the t^{th} residual. Since each individual function O_t is convex from below, so is the summation $O(\underset{\sim}{e})$ and, equivalently, the surface $S(\underset{\sim}{b})$. This yields:

LEMMA 1: The surface S is convex from below.

Therefore, any local minimum for S will be a global minimum and, hence, an LAR estimate.

By definition, a vector $\underset{\sim}{b}$ minimizes $S(\underset{\sim}{b})$ *locally* if $S(\underset{\sim}{b}+\underset{\sim}{d}) \geq S(\underset{\sim}{b})$ for all vectors $\underset{\sim}{d}$ whose length, $||\underset{\sim}{d}||$ (read "the norm of d"), is sufficiently small. The minimum is unique if there is a strict inequality $S(\underset{\sim}{b}+\underset{\sim}{d}) > S(\underset{\sim}{b})$. It will be convenient to select the norm $||\underset{\sim}{d}|| = \sum_{i=1}^{K} |d_i|$. We now derive:

LEMMA 2: The necessary and sufficient condition that $\underset{\sim}{b}$ be an LAR estimate of $\underset{\sim}{\beta}$, expressed in terms of the error, $\underset{\sim}{e} = \underset{\sim}{b} - \underset{\sim}{\beta}$, is: For all $\underset{\sim}{d}$ such that (2.3) is satisfied, (2.4) holds.

$$(2.3) \qquad 0 < ||\underset{\sim}{d}|| < \delta = \frac{\min_t \{|\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t| \;||\; (\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t) \neq 0\}}{\max_t \left\{ \sum_{i=1}^{K} |x_{it}| \right\}}$$

$$(2.4) \qquad \sum_{t=1}^{T} (\text{sign } [\underset{\sim t}{x}(\underset{\sim}{e} + \underset{\sim}{d}) - u_t])\; \underset{\sim t\sim}{x}\underset{\sim}{d} \geq 0.$$

The estimate is unique iff (2.4) holds as a strict inequality.

Proof:

$$\text{From (2.2)} \quad S(\underset{\sim}{b}+\underset{\sim}{d}) - S(\underset{\sim}{b}) = \sum_{t=1}^{T} (|\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t| - |\underset{\sim t}{x}\underset{\sim}{e} - u_t|)$$

When $|\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t| = 0$,

$$|\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t| - |\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t| = \text{sign } [\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t]\; \underset{\sim t\sim}{x}\underset{\sim}{d}.$$

When $|\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t| \neq 0$, and $\underset{\sim}{d}$ satisfies (2.3)

$$|\underset{\sim t\sim}{x}\underset{\sim}{d}| \leq \sum_{i=1}^{K} |x_{it}d_i| \leq \left(\sum_{i=1}^{K} |x_{it}| \right)\left(\sum_{i=1}^{K} |d_i| \right) < \delta \sum_{i=1}^{K} |x_{it}| \leq |x_t\underset{\sim}{e} - u_t|,$$

and, therefore, $\text{sign } [\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t] = \text{sign } [\underset{\sim t}{x}\underset{\sim}{e} - u_t]$ and, again,

$$|\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t| - |\underset{\sim t\sim}{x}\underset{\sim}{e} - u_t| = \text{sign } [\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t]\; \underset{\sim t}{x}\underset{\sim}{d}.$$

Hence, for those $\underset{\sim}{d}$ which satisfy (2.3), $O(\underset{\sim}{e}+\underset{\sim}{d}) - O(\underset{\sim}{e}) = \sum_{t=1}^{T}$ (sign $[\underset{\sim}{x}_t(\underset{\sim}{e}+\underset{\sim}{d}) - u_t]\ \underset{\sim}{x}_t\underset{\sim}{d}$) and (2.4) is equivalent to the condition that $\underset{\sim}{e}$ minimize O locally.

Equation 2.4 is the implicit equation relating $\underset{\sim}{e}$, $\underset{\sim}{X}$ and $\underset{\sim}{u}$ which will be used in the balance of this paper. Notice that the estimation error does not depend on the true value of the parameters:

THEOREM 1: In the regression model (1.1), the estimation error is independent of $\underset{\sim}{\beta}$.

The symmetry of (2.4) yields:

THEOREM 2: In regression model (1.1), if either of the following holds:

(i) u_t is symmetrically distributed for all t, or

(ii) the explanatory variable vectors $\underset{\sim}{x}_t$ are symmetrically distributed about zero and the disturbances u_t are identically distributed,

the estimation error vector, $\underset{\sim}{e}$, will be symmetrically distributed about zero.

Proof:

Equation (2.4) is equivalent to the condition

$$(2.5) \qquad -\underset{\sim}{d} \sum_{t=1}^{T} \text{sign}\,[(-\underset{\sim}{e}-\underset{\sim}{d})\,\underset{\sim}{x}_t - (-u_t)]\,\underset{\sim}{x}_t \geq 0$$

and also to the condition

$$(2.6) \qquad -\underset{\sim}{d} \sum_{t=1}^{T} \text{sign}\,[(-\underset{\sim}{e}-\underset{\sim}{d})(-\underset{\sim}{x}_t) - u_t]\,(-\underset{\sim}{x}_t) \geq 0.$$

Since $\underset{\sim}{d}$ takes both positive and negative signs, substituting $\underset{\sim}{d}^* = -\underset{\sim}{d}$ does not change the nature of (2.5) and (2.6), and they can be rewritten as:

$$(2.7) \qquad \underset{\sim}{d}^* \sum_{t=1}^{T} \text{sign}\,[(-\underset{\sim}{e}+\underset{\sim}{d}^*)\,\underset{\sim}{x}_t - (-\underset{\sim}{u}_t)]\,\underset{\sim}{x}_t \geq 0$$

$$(2.8) \qquad \underset{\sim}{d}^* \sum_{t=1}^{T} \text{sign}\,[(-\underset{\sim}{e}+\underset{\sim}{d}^*)(-\underset{\sim}{x}_t) - \underset{\sim}{u}_t]\,(-\underset{\sim}{x}_t) \geq 0.$$

Since (2.4) and (2.7) are equivalent, the vector of disturbances $\underset{\sim}{u} = (u_1,\ldots,u_t)$ will result in error $\underset{\sim}{e}$ in the LAR estimator, iff the vector $-\underset{\sim}{u}$ will result in error $-\underset{\sim}{e}$. Thus, if $\underset{\sim}{u}$ is symmetrically distributed about zero, so is $\underset{\sim}{e}$.

Similarly, (2.8) is the condition that $-\underset{\sim}{e}$ is the error in the LAR estimator with the signs of the explanatory variables reversed. If the vectors $\underset{\sim}{x}_t$ are symmetrically distributed around zero, a change of sign to $-\underset{\sim}{x}_t$ will only permute the order of appearance of the $\underset{\sim}{x}_t$, and if u_t is identically distributed for all t, the order of appearance is immaterial. Thus, the likelihood of $\underset{\sim}{e}$ given $\underset{\sim}{X}$ (equal to the likelihood of $-\underset{\sim}{e}$ given $-\underset{\sim}{X}$) is equal to the likelihood of $-\underset{\sim}{e}$ given $\underset{\sim}{X}$, and $\underset{\sim}{e}$ is symmetrically distributed about zero.

THEOREM 3: Let $F^*(\underset{\sim}{e}|\underset{\sim}{X}, F(\underset{\sim}{u}))$ be the cumulative probability distribution for the LAR estimation error in (1.1), as a function of the explanatory variables $\underset{\sim}{X}$ and the disturbance distribution $F(u)$. Then for $\delta > 0$ and $\underset{\sim}{\Lambda}$ any nonsingular KxK matrix,

$$F^*(\underset{\sim}{e}|\underset{\sim}{X},F(\underset{\sim}{u})) = F^*(\delta\underset{\sim}{\Lambda}^{-1}\underset{\sim}{e}|\underset{\sim\sim}{X\Lambda},F(\delta\underset{\sim}{u})).$$

That is, the scale of the estimation error distribution varies proportionately with the scale of the disturbance distribution, and any linear transformation of the X variables results in the inverse transformation of the estimation error.

Proof:

The theorem follows immediately from Lemma 2. Since

$$\sum_{t=1}^{T} \text{sign}[\underset{\sim t}{x}(\underset{\sim}{e}+\underset{\sim}{d}) - u_t]\underset{\sim t}{x}\underset{\sim}{d} = \frac{1}{\delta}\sum_{t=1}^{T} \text{sign}[\underset{\sim t}{x}\underset{\sim}{\Lambda}(\delta\underset{\sim}{\Lambda}^{-1}\underset{\sim}{e} + \delta\underset{\sim}{\Lambda}^{-1}\underset{\sim}{d})-\delta u_t](\underset{\sim t}{x}\underset{\sim}{\Lambda})(\delta\underset{\sim}{\Lambda}^{-1}\underset{\sim}{d}),$$

the pair $\underset{\sim}{X}$ and $\underset{\sim}{u}$ result in estimation error $\underset{\sim}{e}$ if and only if the pair $\underset{\sim\sim}{X\Lambda}$ and $\delta\underset{\sim}{u}$ result in estimation error $\delta\underset{\sim}{\Lambda}^{-1}\underset{\sim}{e}$.

III. THE SAMPLING THEORY FOR LAR REGRESSION WITH ONE EXPLANATORY VARIABLE

Let $y_t = \beta x_t + u_t$, $t=1,\ldots,T$, where x_t is a single explanatory variable. Assume that the disturbances are independent of x and are distributed independently and identically according to some probability distribution.

The sum of absolute residuals $S(b) = 0(e)$ is now a function of a single variable. The function is convex from below, and accordingly will reach its minimum at a point e greater than c if the function 0 is downward sloping at c, i.e. $\left.\frac{d0(e)}{d(e)}\right|_c < 0$. The converse of this also holds, with the exception that when the slope is zero in an interval containing c, the LAR estimator is ill-defined over the interval. From a practical standpoint, this difficulty is resolved by defining the LAR estimator as the midpoint of the interval, but the estimation error cannot be bounded by the slope of the objective function. For the purposes of this section, assume that the occurrence of a zero slope has zero probability. (As will be apparent below, this assumption will hold when no summation of the form $\sum_{t=1}^{T} (\pm x_t)$ equals zero.) Then the cumulative distribution $F^*(\cdot)$ of the estimation error relates simply to the slope of the objective function:[4]

---

[4]It may illuminate the coming sections to note that the approach to be taken toward the LAR sampling theory, in which the estimation error is bounded by the slope of the objective function, can also be used to derive the sampling distribution of the LS estimators. For LS, the objective function $S_2 = \sum_{t=1}^{T} (x_t e - u_t)^2$ is again convex, so that $P[e \geq c] = P\left[\left.\frac{dS_2}{de}\right|_c < 0\right]$ and $\left.\frac{dS_2}{de}\right|_c = 2 \sum_{t=1}^{T} x_t(x_t c - u_t)$. This summation has mean $2c \sum_{t=1}^{T} x_t^2$ and variance $4\sigma^2 \sum_{t=1}^{T} x_t^2$ when σ^2 is finite. Approximating the probability distribution of the summation by a normal distribution $P[e \geq c] \sim P\left[\mathcal{N}\left(2c \sum_{t=1}^{T} x_t^2,\ 4\sigma^2 \sum_{t=1}^{T} x_t^2\right) > 0\right] = P\left[\mathcal{N}\left(0,\ \frac{\sigma^2}{\sum_{t=1}^{T} x_t^2}\right) > c\right]$. Thus, the

$$(3.1) \qquad 1 - F^*(c) = P[e > c] = P\left[\left.\frac{dO(e)}{de}\right|_c < 0\right].$$

The right-hand derivative of the objective function is given by

$$(3.2) \qquad V(c) \equiv \left.\frac{dO(e)}{de}\right|_c = \sum_{t=1}^{T} \left.\frac{dO_t(e)}{de}\right|_c \equiv \sum_{t=1}^{T} V_t(c),$$

where

$$(3.3) \qquad V_t(c) = \left\{\begin{array}{l} |x_t| \\ \text{sign } (c_t - u_t)\, x_t \end{array}\right\} \quad \text{if} \quad \left\{\begin{array}{l} cx_t - u_t = 0 \\ cx_t - u_t \neq 0 \end{array}\right.$$

$$(3.4) \qquad = \left\{\begin{array}{l} |x_t| \\ -|x_t| \end{array}\right\} \quad \text{if} \quad \left\{\begin{array}{l} u_t \leq cx_t \text{ and } x_t > 0 \text{ or } u_t \geq cx_t \text{ and } x_t < 0 \\ u_t > cx_t \text{ and } x_t > 0 \text{ or } u_t < cx_t \text{ and } x_t < 0 \end{array}\right.$$

For the balance of this section, the disturbance is assumed to be symmetrically distributed. Then, $P[u_t \geq cx_t] = P[-u_t \leq c(-x_t)] = P[u_t \leq c(-x_t)]$, and we find

---

approximation yields the result that the error is normally distributed with mean zero and variance $\sigma^2 \Big/ \left(\sum_{t=1}^{T} x_t^2\right)$, which is the exact result when disturbances are normally distributed and the large sample approximation in general.

$$
\begin{aligned}
&P[V_t(c) = \ \ |x_t|] = P[u_t \leq c|x_t|] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad , \ t = 1,\ldots,T \\
&P[V_t(c) = -|x_t|] = P[u_t > c|x_t|]
\end{aligned}
\tag{3.5}
$$

Let $G(\cdot)$ be the cumulative distribution function of the absolute value of u, i.e. $G(w) = P[|u| \leq w] = 2F(|w|) - 1$. Then, for $c \geq 0$,

$$
V_t(c) = \left\{ \begin{array}{c} |x_t| \\ \\ -|x_t| \end{array} \right\} \text{ with probability } \left\{ \begin{array}{l} 1/2 + \dfrac{G(c|x_t|)}{2} \\ \\ 1/2 - \dfrac{G(c|x_t|)}{2} \end{array} \right. \tag{3.6}
$$

The exact cumulative distribution function for e can now be represented as follows: Let Q be the set of 2^T vectors comprising all possible ordered combinations of T elements equal to ± 1. Let $|\underset{\sim}{X}| = (|x_1|,\ldots,|x_T|)'$. Then each possible combination of values $V_1(c),\ldots,V_T(c)$ can be written as $(q_1 x_1,\ldots,q_T x_T)$, for some $\underset{\sim}{q} \in Q$, and the slope of the objective function will be given by $\sum_{t=1}^{T} V_T(c) = \underset{\sim}{q}'|\underset{\sim}{X}|$. Therefore, for $c \geq 0$,

$$
1 - F^*(c) = P[e \geq c] = P\left[\sum_{t=1}^{T} V_T(c) < 0\right] = \sum_{\underset{\sim}{q} \cdot \exists \cdot \underset{\sim}{q}'|\underset{\sim}{X}| < 0} \left(\prod_{t=1}^{T} \left(\frac{1 - q_t G(c|x_t|)}{2} \right) \right). \tag{3.7}
$$

By Theorem 2, the estimation error distribution will be symmetric since the disturbance distribution is symmetric. Hence, the distribution for $c < 0$ will be the reflection of the distribution given here, i. e. $F^*(-c) = 1 - F^*(c)$.

For the special case $x_t = 1$, $t=1,\ldots,T$, where the LAR estimator reduces to the sample median, the sample size must be odd for the median to be defined. When $T = 2r + 1$, the above formula reduces to:

$$(3.8) \qquad 1 - F^*(c) = \sum_{j=r+1}^{T} \frac{T!}{j!(T-j)!} F(c)^{T-j}(1-F(c))^j,$$

and the probability differential simplifies further [17, Vol. 1, Ch. 14] to

$$(3.9) \qquad dF^*(c) = \frac{T!}{(r!)^2} dF(c)F(c)^r(1-F(c))^r.$$

In the general case, (3.7) is useless for computational purposes. However, it does yield a strong existence theorem for the moments of the LAR estimation error. Let the asymptotic <u>order</u> of a function $f(z)$ be defined as the power γ such that $\lim_{z\to\infty} \frac{f(z)}{z^\gamma} = \eta$, where η is some finite constant [9]. This will be written "$f(z)$ is $O(z^\gamma)$," or if $\eta = 0$, "$f(z)$ is $o(z^\gamma)$." The following Lemma will be needed:

LEMMA 3: Let $H(z)$ be any cumulative distribution function for z. If $\int_{-\infty}^{\infty} z^\kappa dH(z)$, the moment of order κ, exists, then

$$(1 - G(w)) = P[|z| \geq w] = o(w^{-\kappa}).$$

Conversely, if $(1 - G(w)) = O(w^{-(\kappa + \varepsilon)})$ for any $\varepsilon > 0$, then z has moments of order up to and including κ.

Proof:

By the definition of existence for the moment of order κ, $\int_0^\infty z^\kappa dh(z)$ and $\int_{-\infty}^0 z^\kappa dH(z)$ are finite and, therefore, the absolute moment of order κ, $\int_{-\infty}^\infty |z|^\kappa dH(z)$, is finite. For any $w > 0$,

$$\int_{-\infty}^\infty |z|^\kappa dH(z) = \lim_{w \to \infty} \left(\int_{-\infty}^{-w} |z|^\kappa dH(z) + \int_{-w}^{w} |z|^\kappa dH(z) + \int_{w}^{\infty} |z|^\kappa dH(z) \right)$$

$$\geq \lim_{w \to \infty} \left(w^\kappa \left(\int_{-\infty}^{w} dH(z) + \int_{w}^{\infty} dH(z) \right) + \int_{-w}^{w} |z|^\kappa dH(z) \right)$$

$$= \lim_{w \to \infty} \left(w^\kappa (1-G(w)) + \int_{-w}^{w} |z|^\kappa dH(z) \right)$$

$$= \lim_{w \to \infty} \left(w^\kappa (1-G(w)) \right) + \int_{-\infty}^{\infty} |z|^\kappa dH(z). \tag{3.10}$$

Thus, $\lim_{w \to \infty} w^\kappa (1-G(w)) = 0$, and therefore $(1-G(w))$ is $o(w^{-\kappa})$.

Conversely, all moments of order up to and including κ will exist when the κ^{th} absolute moment is finite. This moment can be expressed in terms of the function G as follows:

$$\int_{-\infty}^\infty |z|^\kappa dH(z) \leq \sum_{w=1}^{\infty} w^\kappa \cdot p[w - 1 \leq |z| < w]$$

$$= \sum_{w=1}^{\infty} w^\kappa \left((1-G(w-1)) - (1-G(w)) \right)$$

$$(3.11) \qquad = \sum_{z=0}^{\infty} \left(1-G(z)\right)\left((z+1)^{K} - z^{K}\right).$$

But $\lim_{z \to \infty} \frac{(z+1)^{K} - z^{K}}{z^{K-1}} = K$ by L'Hospital's rule. Thus, the right-hand terms in the summation are $O(z^{K-1})$. Therefore, when $(1-G(z))$ is $O(z^{-(K+\varepsilon)})$, this is a summation of terms which are $O(z^{-(1+\varepsilon)})$ and, hence, the summation converges to a finite limit. Under these conditions, then, $\int_{-\infty}^{\infty} |z|^{K} dH(z)$ is finite and the proof is complete.

We now derive an existence theorem for the moments of the estimation error.

THEOREM 4: Assume that the disturbances in the regression model (1.1) are identically and symmetrically distributed. Let h be the smallest number, such that for some combination of the indices $1,2,\ldots,T$, denoted by $t_1,\ldots,t_T$,

$$(3.12) \qquad \sum_{j=1}^{h} |x_{t_j}| > \sum_{j=h+1}^{T} |x_{t_j}|.$$

Then the LAR estimate b will have moments up to and including K, if u has moments up to $(K/h + \delta)$ for some $\delta > 0$.

Proof:

We first show that

$$(3.13) \qquad \underset{c \to \infty}{\text{ORDER}} \{P[e > c]\} = h \underset{c \to \infty}{\text{ORDER}} \{(1-G(c))\}.$$

From (3.7), since the order of a finite sum is the maximum of the orders of its elements, $P[e \geq c]$ has order equal to

$$(3.14)\quad \underset{c \to \infty}{\text{ORDER}} \left\{ P[e \geq c] \right\} = \max_{q_j \cdot \varepsilon \cdot q_j' |X| < 0} \left(\underset{c \to \infty}{\text{ORDER}} \left\{ \prod_{t=1}^{T} \left(\frac{1 - q_j G(c|x_t|)}{2} \right) \right\} \right).$$

Now, as $c \to \infty$, $G(c|x_t|)$ approaches 1 for all $x_t \neq 0$ and remains constant for $x_t = 0$. Hence, all terms where $q_j = -1$ approach a positive constant and do not affect the order of $P[e > c]$. Where $q_j = 1$ and $x_t \neq 0$, the terms approach zero with order equal to that of $(1-G(w))$.

The order of a product is the sum of orders of its terms, so the above equals:

$$(3.15)\quad \max_{q \cdot \exists \cdot q' |X| < 0} \left(\sum_{\substack{t \cdot \exists \cdot q_t = +1 \\ \text{and } x_t \neq 0}} \text{ORDER } \{(1-G(c))\} \right)$$

$$= \max_{q \cdot \exists \cdot q' |X| < 0} \left(h(q) \text{ ORDER } (1-G(c)) \right)$$

where $h(q)$ is the number of positive entries in q corresponding to nonzero values of x.

The order of $(1-G(c))$ is nonpositive, since $G(\cdot)$ is a decreasing function. Therefore, the summation will be maximized if $h(q)$ is minimized. Since we are considering only those sign combinations q for which $q'|x| < 0$, the minimum is the smallest number h, such that if the terms $|x_t|$ are arranged in order from largest to smallest with subscripts

$t_1,\ldots,t_T$, the condition $-|x_{t_1}|- \ldots -|x_{t_h}| + |x_{t_{h+1}}| + \ldots + |x_{t_T}| < 0$ holds. This is the desired result.

Next, since the estimation error distribution is symmetric,

$\underset{c \to \infty}{\text{ORDER}} \{P[e > c]\} = \underset{c \to \infty}{\text{ORDER}} \{1-G^*(c)\}$, where $G^*(c) \equiv P[|e| < c]$.

We now apply Lemma 3: e will have moments of order up to K if $(1-G^*(c))$ is $O(c^{-(K + \varepsilon)})$, $\varepsilon > 0$; which will hold if $(1-G(c))$ is $O(c^{-(K/h + \delta)})$, $\delta > 0$, by (3.13); and this, in turn, will hold if the disturbance has moments of order $\frac{K}{h} + \delta$.

This is a very strong result. It implies, in particular, that if $h \geq 3$, the estimate $\hat{b}$ must have a finite variance when the disturbance u has a finite mean. Since h will be as large as 3 for all usual samples, and since all the disturbance distributions that have been suggested (including the Pareto with location parameter, generally denoted "α", between 1 and 2) have finite means, this implies that the LAR estimate will have a finite variance in general. In contrast, LS has a finite variance only when the disturbance u has a finite variance.

Turning to the asymptotic properties of LAR as the sample size T increases, let the estimator b be termed <u>consistent</u> if for all $\varepsilon > 0$, $\lim_{T\to\infty} P[|b-\beta| > \varepsilon] = 0$. Then

THEOREM 5: For the regression model in Theorem 4, the following are sufficient conditions for the LAR estimator to be consistent:

(i) $\lim_{T \to \infty} \left(\frac{\sum_{t=1}^{T} x_t^2}{T^2} \right) = 0$

(ii) as $T \to \infty$ some positive proportion, ρ, of the x_t have absolute value greater than some positive constant, γ.

(iii) for every $\varepsilon > 0$, $P[|u_t| < \varepsilon] = (1-G(\varepsilon)) > 0$. That is, the disturbance falls in every interval containing zero with positive probability.

Proof:

Let $V(c,T) = \sum_{t=1}^{T} v_t(c)$ be the slope of the objective function, with mean $M(c,T)$ and variance $\sigma^2(c,T)$. For any $c > 0$,

$$(3.16) \quad M(c,T) = \left(\sum_{t=1}^{T} |x_t| G(c|x_t|) \right) \geq \sum_{\substack{t \leq T \\ \cdot 3 \cdot |x_t| > \gamma}} \gamma G(c|x_t|) \geq \sum_{\substack{t \leq T \\ \cdot 3 \cdot |x_t| > \gamma}} \gamma G(c\gamma)$$

Therefore, by conditions (ii) and (iii)

$$(3.17) \quad \lim_{T \to \infty} \frac{M(c,T)}{T} \geq \frac{\rho T \gamma G(c\gamma)}{T} > 0.$$

Also, by condition (i),

$$(3.18) \quad \lim_{T \to \infty} \frac{\sigma^2(c,T)}{T^2} = \lim_{T \to \infty} \left(\sum_{t=1}^{T} \frac{x_t^2 (1-G(c|x_t|)^2)}{T^2} \right) < \lim_{T \to \infty} \left(\frac{\sum_{t=1}^{T} x_t^2}{T^2} \right) = 0.$$

Hence, by Chebychev's inequality,

$$(3.19) \quad \lim_{T \to \infty} P[c > c] = \lim_{T \to \infty} P[V(c,T) < 0] \leq \lim_{T \to \infty} \left(\frac{\sigma^2(c,T)}{M(c,T)^2} \right) \leq \lim_{T \to \infty} \left(\frac{o(T^2)}{O(T^2)} \right) = 0.$$

By the symmetry of the disturbance distribution, the probability that the error is negative behaves identically, and we have, for any $c > 0$, $\lim_{T \to \infty} P[|b-\beta| > c] = 0$.

The robustness of the LAR estimator in the face of massive tailed distributions is shown by the fact that consistency does not require any restriction on the tails of the disturbance distribution. The conditions stated here are not necessary for consistency, and alternative sets of sufficient conditions can be formulated.

To summarize the conclusions of this section, the univariate LAR estimator will be unbiased, will possess finite variance, and will be consistent for virtually all regressions with symmetric disturbance distributions which can be imagined in practice. When the disturbance distribution is asymmetric, the proofs of Theorems 4 and 5 can be generalized to show that the moments of the LAR estimator will still exist, and the LAR estimator will still converge to its mean, but the estimator will no longer be unbiased unless the explanatory variable is symmetrically distributed. Since the bias is, in general, a cumbersome function of the disturbance distribution and the explanatory variables, it will be difficult to correct for, and may be regarded as a defect of the estimator.

The results of this section can be extended to cover the case of heteroscedasticity. The proofs of Theorems 4 and 5 are possible with minor changes when the disturbance distributions are drawn from some family, with the range of the dispersion parameters bounded above and below. Efficient estimation of $\hat{b}$ will then require weighting the observations by a function of the changing parameters.

IV. AN APPROXIMATE SAMPLING THEORY FOR LAR REGRESSION WITH ONE EXPLANATORY VARIABLE AND SYMMETRICALLY DISTRIBUTED DISTURBANCES

When the disturbance distribution is symmetric, from (3.7)

$$1 - F^*(c) = P\left[\sum_{t=1}^{T} V_t(c) < 0\right].$$

Since each V_t depends only on the corresponding u_t, and the disturbances are independently distributed, the V_t are independent of one another. Therefore, the slope V is the sum of T independent variables and will be approximately normally distributed. From (3.6), for $c \geq 0$ the moments of the V_t are

$$\mu'_{1t} = E[V_t(c)] = |x_t|\, G(c|x_t|) \tag{4.1}$$

$$\mu_{2t} = E[(V_t(c) - \mu'_{1t})^2] = |x_t|^2\,(1 - G(c|x_t|)^2) \tag{4.2}$$

$$\mu_{3t} = E[(V_t(c) - \mu'_{1t})^3] = |x_t|^3\,(2G(c|x_t|)^3 - 2G(c|x_t|)) \tag{4.3}$$

$$\mu_{4t} = E[(V_t(c) - \mu'_{1t})^4] = |x_t|^4\,(-3G(c|x_t|)^4 + 2G(c|x_t|)^2 + 1) \tag{4.4}$$

The slope of the objective function, being the sum of the T independent random variables V_t, therefore has moments:

$$\mu'_1 = E[V(c)] = \sum_{t=1}^{T} \mu'_{1t} \qquad \mu_2 = E[(V(c)-\mu'_1)^2] = \sum_{t=1}^{T} \mu_{2t}$$

$$\mu_3 = E[(V(c)-\mu'_1)^3] = \sum_{t=1}^{T} \mu_{3t} \qquad \mu_4 = E[(V(c)-\mu'_1)^4] = 3\mu_2^2 + \sum_{t=1}^{T} (\mu_{4t} - 3\mu_{2t}^2) \tag{4.5}$$

Thus, for $c \geq 0$,

$$1 - F^*(c) \sim P\left[\mathcal{N}\left(\sum_{t=1}^{T} |x_t| G(c|x_t|),\ \sum_{t=1}^{T} |x_t|^2 (1-G(c|x_t|)^2)\right) \leq 0\right].$$

$$(4.6) \qquad = P\left[\mathcal{N}(0,1) \geq \frac{\sum_{t=1}^{T} |x_t| G(c|x_t|)}{\sqrt{\sum_{t=1}^{T} x_t^2 (1-G(c|x_t|))^2}}\right] = H(c).$$

Thus the cumulative distribution of the LAR estimation error may be approximated by $H(\cdot)$. The third and fourth moments of V can be used to improve this approximation somewhat (by Edgeworth's expansion [6:227-230]), but in all cases these moments have been so close to the corresponding moments of the normal distribution that the improvement has been insignificant. The approximate cumulative distribution is

$$(4.7) \qquad A_1: \quad F(c) = P[e \leq c] = \begin{cases} H(|c|) & \text{if} \quad c < 0 \\ 1 - H(c) & \text{if} \quad c \geq 0 \end{cases}.$$

The LAR estimation error can be expected to be approximately normally distributed, since the error is the compound effect of a large number of independent disturbances, and is insensitive to the value of any one disturbance. Accordingly, a second approximation is

$$A_2: \quad e_{LAR} \sim \mathcal{N}(0, \sigma^2_{A_1}),$$

where $\sigma^2_{A_1}$, the variance of the approximate distribution A_1, is computed by numerical integration.

In searching for an easily computed approximation, we conjectured that the variance of the estimation error for LAR might be related to the second moment of the explanatory variable as in the case of LS, i.e.,

that $\mathrm{VAR}(e_{LAR}) \propto \left(\sum_{t=1}^{T} x_t^2 \right)^{-1}$. Then, the estimation error variance can be computed as $\mathrm{VAR}(e_{LAR}) = \lambda(F,T) \left(\sum_{t=1}^{T} x_t^2 \right)^{-1}$, where $\lambda(F,T)/T$ is the variance of the LAR estimation in the simplest possible regression with a sample of size T from the disturbance distribution F, where $x_t = 1$, $t=1,\ldots,T$, and the LAR estimator reduces to the median of the disturbances. Computations are then greatly reduced, for when $x_t \equiv 1$, expression (4.6) simplifies to

$$P[e \geq c] = p\left[\eta(0,1) \geq \frac{\sqrt{T}\, G(c)}{\sqrt{1-G(c)^2}}\right] = H(c). \tag{4.8}$$

Moreover, the exact distribution of the median is known, so that λ can be computed by numerical integration of the distribution (3.9), thus avoiding approximation (4.6) entirely. Finally, and perhaps most importantly, the computations need only be done once for any given disturbance distribution $F(\cdot)$, for a variety of sample sizes T, and then the computed values $\lambda(F,T)$ can be substituted, whenever the distribution is encountered, into the approximation

$$A_{LAR}: \quad e_{LAR} \sim \eta\left(0, \lambda(F,T) \left(\sum_{t=1}^{T} x_t^2 \right)^{-1}\right)$$

To evaluate these approximations, we resorted to Monte Carlo studies. This approach, which has been aptly criticized as capital-intensive, is the only viable approach in the present case, since evaluation of the exact distribution (3.7) is prohibitively costly for large T. The one exception is the case of the median, where the

exact distribution given by (3.9) can be computed and compared to the approximations. In this case approximation A_1, computed via (4.8), yields a good approximation to the true standard error. In Table 4.1, ratios of the standard error of the median to that of the mean ($\sigma/\sqrt{T}$) are given for the normal distribution and for two massive-tailed distributions which are used in the Monte Carlo studies reported below. These "contaminated normal" distributions are defined by the parameters (p,V) as follows: the disturbance is selected randomly from two normal distributions, being drawn from a standard normal with probability (1-p), and from a normal with mean 0 and variance V with probability p. The resulting distribution has variance (pV + (1-p)) and kurtosis $3(pV^2 + (1-p))/(pV + (1-p))^2$. As is apparent from (4.8), the approximated standard error decreases proportionately with $1/\sqrt{T}$, and this is not accurate as seen in Table 4.1. Although the approximation is quite satisfactory for odd values of T between 31 and 119, the most common range in econometrics, the imperfections in the approximation A_1 underscore the potential importance of A_{LAR}, where the exact distribution (3.9) may be used.

TABLE 4.1--Approximated and True Standard Errors of the the Median, as Multiples of the Standard Error of the Mean

Distribution	Approximation A_1	True Values for Sample Sizes 15	31	59	119
Normal	1.24	1.2351	1.2446	1.2488	1.2511
Contaminated Normal (0.15,16)	.777	0.7819	0.7824	0.7828	0.7831
Contaminated Normal (0.15,25)	.658	0.6642	0.6639	0.6639	0.6639

The several approximations were then compared by Monte Carlo simulation over a wide range of symmetric disturbance distributions, sample sizes, and distributions of the explanatory variable. In all cases the LAR estimation error distribution did not differ significantly from the normal distribution, so that A_1 was dropped from consideration. The normality of $\hat{b}_{LAR}$, which is preserved through all our Monte Carlo studies, even for infinite variance disturbance distributions, is the consequence of the estimator's insensitivity to extreme values of the disturbances. A second and unexpected result was that the approximation A_{LAR} was generally equivalent to or better than A_2. Apparently, despite its closer adherence to the form of the exact distribution, the approximation (4.6) introduces greater errors when the explanatory variable is widely dispersed than does approximation A_{LAR}. This result allows the most easily computed and most readily generalized approximation, A_{LAR}, to be accepted as the approximation of choice.

Extensive Monte Carlo studies of the validity of A_{LAR} were conducted next. We anticipated that the validity of the approximation would increase with T, would decrease with the kurtosis of the disturbance distribution, and would decrease with the kurtosis of the explanatory variable. Accordingly, the experiments are varied along each of these three factors: (i) sample sizes are T = 31 and T = 59; (ii) disturbance distributions are normal (kurtosis = 3), contaminated normal (0.15, 16.0) with kurtosis $\mu_4/\sigma^4 = 11.8$, and contaminated normal (0.15, 25.0) with $\mu_4/\sigma^4 = 13.4$; (iii) the explanatory variables are constant, normally distributed, and widely dispersed with very

high kurtosis. Thus, there are two sample sizes, three disturbance distributions, and three explanatory variable distributions, or eighteen cases in all. For each case, two to four experiments were conducted, each consisting of 1000 replications with identical values of the explanatory variable but different realizations from the disturbance distribution. The results for the sixty-two experiments, involving 62,000 regressions in all, are given in Tables 4.2, 4.3 and 4.4

At the left of the tables, the type of explanatory variable, and the actual coefficient of absolute variation ($\phi = E[(|x|-E[|x|])^2]/E[|x|]^2$) and kurtosis ($\mu_4/\sigma^4$) of the explanatory variable are given. Then the average sample variance and average sample kurtosis from the 1000 samples of T pseudo-random disturbances in the experiment are given.

Then, for each estimation method, the performance of the approximate theoretical distribution for the estimation errors (normal, with variance equal to $\sigma_u^2/(\Sigma x^2)$ in the LS case and with variance equal to $\lambda/(\Sigma x^2)$ in the LAR case, with λ computed by numerical integration of (3.9)), is summarized by four statistics. The first of these is the ratio of the actual root mean square estimation error to the standard deviation of estimation error implied by the theoretical distribution (S/σ).. Since the estimation errors are independently and approximately normally distributed, this ratio is distributed similarly to $r\sqrt{\chi^2_{1000}/1000}$, where r is the ratio of the standard error of the true estimation error distribution to the approximated standard error. Under the null hypothesis that the approximated distribution is correct, S/σ is approximately normally distributed with mean value

TABLE 4.2--Normal Disturbance Distribution

X Distribution		Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
Φ	μ_4/σ^4	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31														
Constant														
--	--	0.992	2.82	0.965	4.4	0.8	4.9	1.005	4.8	1.1	3.7	1.23	1.27	0.80
--	--	0.994	2.77	1.016	5.3	1.2	3.1	1.047	5.8	1.4	3.3	1.23	1.26	0.83
Normal														
.777	2.90	1.003	2.80	0.978	4.1	0.9	2.3	1.020	6.3	1.1	2.0	1.23	1.28	0.78
.442	2.62	0.993	2.79	0.988	4.2	0.9	2.5	1.030	6.0	1.1	2.8	1.23	1.28	0.80
.448	2.16	0.999	2.82	0.981	4.2	0.9	2.4	1.051	6.1	1.2	3.3	1.23	1.32	0.81
Dispersed														
1.93	6.83	0.985	2.78	0.991	4.1	0.8	2.3	1.004	5.2	1.1	2.6	1.23	1.24	0.81
1.66	9.46	1.009	2.80	1.013	5.1	1.2	1.7	1.000	4.9	1.1	2.7	1.23	1.21	0.81
1.79	9.26	0.999	2.82	0.990	4.4	0.8	2.9	0.990	4.8	0.9	2.6	1.23	1.22	0.81
Sample Size T = 59														
Constant														
--	--	0.997	2.91	1.000	4.9	0.9	1.5	1.023	4.4	0.9	3.8	1.24	1.27	0.80
--	--	0.993	2.89	1.003	4.9	0.8	2.8	1.010	4.4	1.4	1.7	1.24	1.25	0.80
--	--	0.993	2.90	1.009	5.5	1.1	2.6	1.008	5.6	1.2	1.7	1.24	1.24	0.80
Normal														
.717	2.85	1.002	2.87	0.978	4.1	0.7	2.8	0.984	4.5	0.9	1.7	1.24	1.24	0.77
.781	3.90	0.990	2.87	0.965	4.9	0.6	3.3	1.006	6.0	1.0	2.5	1.24	1.29	0.79
.403	2.22	0.993	2.90	1.003	5.6	0.7	2.3	0.987	4.8	1.1	2.0	1.24	1.22	0.79
Dispersed														
1.32	5.82	0.999	2.85	1.029	5.9	1.3	2.5	1.007	5.1	1.1	2.3	1.24	1.21	0.81
1.23	7.68	0.993	2.86	1.004	5.2	1.3	2.6	1.008	5.5	1.3	2.8	1.24	1.24	0.80
1.39	6.39	0.993	2.90	0.998	5.1	1.3	1.7	1.000	4.8	0.9	1.7	1.24	1.24	0.81

TABLE 4.3--Contaminated (0.15, 16.0) Normal Disturbance Distribution

X Distribution		Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
ϕ	μ_4/σ^4	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31														
Constant														
--	--	3.26	6.96	1.000	5.2	1.5	2.6	0.982	6.1	1.0	2.9	0.77	0.75	0.65
--	--	3.35	7.04	1.023	5.9	2.0	2.7	0.989	4.9	1.3	1.6	0.77	0.74	0.64
--	--	3.18	6.82	1.014	4.9	1.2	2.7	1.022	5.4	1.5	1.7	0.77	0.77	0.67
Normal														
.628	2.83	3.17	6.82	1.012	5.7	1.8	5.6	1.034	6.1	2.1	2.6	0.77	0.79	0.69
.631	2.65	3.25	6.86	1.004	5.8	2.1	3.5	1.017	6.2	1.3	1.9	0.77	0.78	0.67
.546	2.43	3.28	7.02	1.003	5.4	1.8	4.6	1.014	6.0	2.0	2.3	0.77	0.78	0.66
--	2.65	3.19	6.97	1.005	5.3	1.2	5.2	1.006	5.9	1.2	3.2	0.77	0.77	
Dispersed														
1.52	7.25	3.16	6.97	0.961	5.3	1.7	7.5	1.039	6.0	1.9	1.6	0.77	0.84	0.67
1.22	3.84	3.26	6.86	1.036	7.0	1.9	2.8	1.094	7.0	2.3	4.2	0.77	0.81	0.68
1.32	5.28	3.32	6.92	1.004	5.9	1.8	4.7	1.071	7.1	1.9	3.0	0.77	0.82	0.66
1.79	9.26	3.24	7.03	0.975	5.5	2.3	10.7	1.093	6.7	2.5	2.4	0.77	0.87	0.72
Sample Size T = 59														
Constant														
--	--	3.27	8.47	1.020	5.6	1.4	1.2	1.012	5.8	0.9	1.8	0.77	0.77	0.65
--	--	3.24	8.25	0.971	4.2	1.1	3.0	0.989	4.5	0.6	2.0	0.77	0.79	0.65
--	--	3.26	8.63	1.054	6.8	1.7	2.8	1.065	6.7	2.1	3.7	0.77	0.78	0.67
--	--	3.22	8.57	1.007	5.5	1.0	2.6	1.003	5.7	1.0	1.3	0.77	0.77	0.66
Normal														
.883	4.17	3.27	8.43	0.993	5.7	1.2	3.3	0.999	4.1	1.1	1.7	0.77	0.78	0.66
.549	2.67	3.19	8.46	1.027	5.9	1.6	2.9	1.042	5.2	1.7	2.9	0.77	0.79	0.67
.602	2.37	3.17	8.49	0.937	3.8	1.1	6.1	1.008	5.5	1.1	1.7	0.77	0.84	0.64
.687	3.02	3.25	8.57	1.008	5.0	1.3	3.4	1.002	5.1	0.9	2.7	0.77	0.77	0.66
Dispersed														
1.70	11.03	3.20	8.48	0.969	5.4	2.0	7.2	0.993	5.1	1.7	2.9	0.77	0.79	0.67
2.65	16.62	3.30	8.32	1.013	5.7	2.5	8.7	1.114	6.8	3.1	2.3	0.77	0.85	0.71
2.17	12.17	3.23	8.52	0.989	6.1	2.8	7.0	1.114	7.3	2.7	3.3	0.77	0.85	0.70
1.23	7.29	3.25	8.57	0.967	5.8	1.3	7.1	0.992	5.7	0.8	1.7	0.77	0.80	0.65

TABLE 4.4--Contaminated (0.15, 25.0) Normal Disturbance Distribution

X Distribution		Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
ϕ	μ_4/σ^4	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31														
Constant														
--	--	4.57	8.07	1.000	5.5	1.5	4.6	0.989	4.8	0.9	2.4	0.66	0.65	0.59
--	--	4.66	8.40	1.020	6.1	1.8	2.7	1.004	5.3	0.8	1.8	0.66	0.65	0.58
--	--	4.39	8.16	0.997	5.2	1.1	3.3	1.029	6.0	1.0	2.7	0.66	0.68	0.59
Normal														
.831	2.62	4.56	8.28	0.955	5.2	1.1	6.1	1.028	5.7	1.3	2.2	0.66	0.71	0.62
1.13	3.67	4.57	8.12	0.986	5.1	1.2	6.2	1.060	6.4	2.2	3.2	0.66	0.70	0.64
.465	2.39	4.59	8.09	1.030	6.4	2.0	4.3	1.017	5.5	1.6	1.8	0.66	0.65	0.62
.448	2.16	4.58	8.38	1.017	6.4	2.0	5.4	1.042	6.2	1.4	3.1	0.66	0.67	0.63
Dispersed														
2.14	11.27	4.55	8.31	0.997	5.7	2.7	10.0	1.188	7.5	3.5	2.8	0.66	0.77	0.68
.482	7.07	4.76	8.22	0.964	5.3	2.2	10.9	1.081	7.1	2.7	2.5	0.66	0.73	0.60
2.04	8.03	4.54	8.03	1.003	6.7	3.3	11.7	1.119	7.4	2.5	3.3	0.66	0.73	0.66
1.79	9.26	4.58	8.38	0.972	6.2	2.4	13.4	1.125	7.3	2.8	2.8	0.66	0.75	0.69
Sample Size T = 59														
Constant														
--	--	4.69	10.32	1.028	6.0	1.6	2.1	1.022	5.6	1.4	1.5	0.66	0.66	0.61
--	--	4.63	10.15	0.988	4.9	1.1	2.4	1.035	5.7	1.7	2.2	0.66	0.69	0.60
--	--	4.43	10.40	1.008	5.2	0.8	2.8	1.029	6.2	1.2	3.5	0.66	0.67	0.63
Normal														
.357	2.15	4.60	10.58	0.991	4.7	1.1	2.8	1.027	6.0	1.4	1.8	0.66	0.69	0.59
.440	2.26	4.64	10.56	1.051	7.1	1.4	2.4	1.023	6.0	1.5	1.9	0.66	0.64	0.61
.770	4.00	4.56	10.33	1.017	5.4	1.6	2.7	1.036	6.4	1.4	2.0	0.66	0.67	0.61
.403	2.22	4.55	10.36	1.004	5.4	1.1	2.4	1.008	4.9	0.8	2.2	0.66	0.66	0.60
Dispersed														
1.44	6.75	4.62	10.32	1.016	6.2	2.4	5.6	1.056	5.9	1.5	3.4	0.66	0.69	0.61
1.65	9.34	4.54	10.45	0.998	5.5	2.1	7.6	1.060	5.8	1.4	3.6	0.66	0.70	0.63
2.86	14.09	4.61	10.03	1.044	7.4	3.2	10.4	1.137	6.6	3.2	2.5	0.66	0.71	0.65
1.39	6.39	4.55	10.36	1.015	6.5	2.0	4.1	1.036	5.7	1.3	1.9	0.66	0.67	0.65

TABLE 4.5--Summary of Univariate Monte Carlo Results

		Disturbance Distribution															
		Normal						Contaminated Normal (0.15, 16.0)						Contaminated Normal (0.15, 25.0)			
		LS			LAR			LS			LAR			LS			LAR		
		S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%
Regression with X Constant																			
T = 31	MAX	1.016	5.3	1.2	1.047	5.8	1.4	1.023	5.9	2.0	1.022	6.1	1.5	1.020	6.1	1.8	1.029	6.0	1.0
	MEAN	0.991	4.9	1.0	1.026	5.3	1.2	1.012	5.3	1.6	0.998	5.5	1.3	1.006	5.6	1.5	1.007	5.2	0.9
	MIN	0.965	4.4	0.8	1.005	4.8	1.1	1.000	4.9	1.2	0.982	4.9	1.0	0.997	5.2	1.1	0.989	4.8	0.8
T = 59	MAX	1.009	5.5	1.1	1.023	5.6	1.4	1.054	6.8	1.7	1.065	6.7	2.1	1.028	6.0	1.6	1.035	6.2	1.7
	MEAN	1.004	5.1	0.9	1.014	4.8	1.2	1.013	5.5	1.3	1.017	5.7	1.2	1.008	5.2	1.2	1.029	5.8	1.4
	MIN	1.000	4.9	0.8	1.008	4.4	0.9	0.971	4.2	1.0	0.989	4.5	0.6	0.988	4.9	0.8	1.022	5.6	1.2
Regression with X Normally Distributed																			
T = 31	MAX	0.988	4.2	0.9	1.051	6.3	1.2	1.012	5.8	2.1	1.034	6.2	2.1	1.030	6.4	2.0	1.060	6.4	2.2
	MEAN	0.982	4.2	0.9	1.034	6.1	1.1	1.006	5.6	1.7	1.018	6.0	1.6	0.997	5.8	2.1	1.039	5.8	1.6
	MIN	0.978	4.1	0.9	1.020	6.0	1.1	1.003	5.3	1.2	1.006	5.9	1.2	0.955	5.1	1.1	1.017	5.5	1.3
T = 59	MAX	1.003	5.6	0.7	1.006	6.0	1.1	1.027	5.9	1.6	1.042	5.5	1.7	1.051	7.1	1.6	1.036	6.4	1.5
	MEAN	0.982	4.9	0.7	0.992	5.1	1.0	0.991	5.1	1.3	1.013	5.0	1.2	1.016	5.6	1.3	1.023	5.8	1.3
	MIN	0.965	4.1	0.6	0.984	4.5	0.9	0.937	3.8	1.1	0.999	4.1	0.9	0.991	4.7	1.1	1.008	4.9	0.8
Regression with X Highly Dispersed																			
T = 31	MAX	1.013	5.1	1.2	1.004	5.2	1.1	1.036	7.0	2.3	1.094	7.1	2.5	1.003	6.7	3.3	1.118	7.5	3.5
	MEAN	0.998	4.5	1.0	0.998	5.0	1.0	0.994	5.9	1.9	1.074	6.7	2.1	0.984	6.0	2.6	1.128	7.3	2.9
	MIN	0.990	4.1	0.8	0.990	4.8	0.9	0.961	5.3	1.7	1.039	6.0	1.9	0.964	5.3	2.2	1.081	7.1	2.5
T = 59	MAX	1.029	5.9	1.3	1.008	5.5	1.3	1.013	6.1	2.8	1.114	7.3	3.1	1.044	7.4	3.2	1.137	6.6	3.2
	MEAN	1.010	5.4	1.3	1.005	5.1	1.1	0.985	5.6	2.1	1.053	6.2	2.1	1.018	6.4	2.4	1.072	6.0	1.8
	MIN	0.998	5.1	1.3	1.000	4.8	0.9	0.967	5.4	1.3	0.992	5.1	0.8	0.998	5.5	2.0	1.036	5.7	1.3

unity and standard deviation of $\sqrt{1/2000}$ [17,I:371-374], and the 99 percent confidence region for the ratio is (.942, 1.058). The second and third statistics are the percentages of the actual estimation errors which fall outside of the 95 and 99 percent confidence regions of the approximated distribution. Under the null hypothesis that the approximated distribution is correct, the mean values of these binomial variables are 5 and 1 and the exact 99% confidence regions are (3.4, 6.9) and (.3, 1.9). The final statistic is the Kolmogorov test statistic (d) for the distribution as a whole, equal to the maximum, over n = 1,..., 1000 of $|P(e_n) - F(e_n)|$, where $P(e_n)$ is the percentage of errors in the sample of 1000 which are smaller than e_n, and $F(e_n)$ is the value of the approximated cumulative distribution function at e_n, expressed as a percentage. Under the null hypothesis, the 99 percent confidence region for d is (0, 5.14) [17:II: 457]. At the right-hand side of the table, the theoretical (σ/σ) and actual (S/S) ratios of the LAR to LS root mean square errors are given, and then the sample correlation between LS and LAR estimation errors. The maxima, means and minima of the statistic S/σ, % >95, and % >99 are given in Table 4.5 for the normal disturbances and for the two contaminated normal distributions taken together. The latter can be discussed jointly, since the results for the two are similar.

The pseudo-random disturbances satisfy all tests of randomness.[6] The mean is never significantly different from zero and the variance

---

[6]To avoid the serial dependence of the pseudo-random numbers generated by most algorithms, several stages of random sampling were

is always appropriate for the frequency distribution from which the disturbances are drawn. The average sample kurtosis is generally lower than for the frequency distribution, but this is due to the downward bias of the estimated kurtosis in a small sample. The true LS sampling theory is never rejected at the 99 percent confidence level in the sixty-two experiments: the theoretical LS standard error is never rejected, and for the seventeen experiments with normally distributed disturbances, where the hypothesized normal distribution for the LS estimation error is also the true distribution, the normal distribution is never rejected at the 99 percent level by any of the three applicable statistics (% >95, % >99, and d). It is, therefore, unlikely that the results of the Monte Carlo experiments are significantly biased due to deviations from randomness in the pseudo-random disturbances.

Turning to the evaluation of A_{LAR}, the results for the normally distributed disturbances are excellent (Tables 4.2 and 4.5). The null hypothesis that the estimation error is normally distributed with variance $\lambda(\eta,T)(\sum_{t=1}^{T} x_t^2)^{-1}$ is nowhere rejected at the 99% confidence level by any of the statistics d, S/σ, %>95, %>99, and in fact

---

coupled in generating the disturbances. The integers from 1 to 1000 were first pseudo-randomly reordered by pseudo-random sampling without replacement, and then thoroughly scrambled by 10,000 pseudo-random exchanges of position, to form a pool of 1000 pseudo-random integers. "Uniformly distributed random numbers" were then constructed by pseudo-random sampling with replacement of three integers, I_1, I_2, I_3, from this pool by the formula $R = (I_1-1)\ 10^{-3} + (I_2-1)\ 10^{-6} + (I_3-1)\ 10^{-9}$. For any desired disturbance distribution other than the normal, the random disturbances are constructed through the inverse of the cumulative probability distribution, $u = F^{-1}(R)$. Random normal variates are constructed as the sum of sixteen independent uniformly distributed variables.

the divergences from the mean values of these statistics are only marginally worse than for the LS sampling theory.

The results for the two high-kurtosis disturbance distributions may be discussed jointly (Tables 4.3, 4.4 and 4.5). Note that in these experiments, the LS estimation error is no longer exactly normally distributed, so that the normal approximation to the LS sampling theory is susceptible to error. Where x is constant, the performance of A_{LAR} continues to be satisfactory. The approximation is never rejected at the 99% level of confidence for incorrect standard error (any rejection would indicate that λ was computed incorrectly), and the coincidence to the normal distribution of the LAR estimation error is as good as for the LS estimation error. Both the LS and LAR approximations are rejected once at the 99% level, in both cases because the percentage of errors beyond the 99% confidence region is excessively large.

In the experiments where x is not constant and the u kurtosis is high, the LAR estimation error is distributed more closely to the normal: the null hypothesis that the estimation error is normally distributed is rejected at the 99% level by the d-statistic 19 times for LS, never for LAR.

Where x is normally distributed, the performance of A_{LAR} is satisfactory. Insofar as the standard error is concerned, there appears to be a downward bias of about 1 percent, but the approximated standard error is never rejected at the 99% confidence level. Insofar as the confidence regions are concerned, A_{LAR} is slightly superior to A_{LS}. A_{LS} is rejected 3 times because of excessive errors beyond

the 99% confidence region while A_{LAR} is rejected only once, and in addition A_{LS} is rejected once because of excessive errors beyond the 95% confidence region. Both approximations underestimate the tails of the error distribution on average, with A_{LS} the worse offender, but the degree of underestimation will probably not be regarded as a serious problem in hypothesis testing.

Where both the x distribution and the u distribution have high kurtosis, A_{LAR} clearly underestimates the standard error of the LAR estimate. The average downward bias is 8.6% for sample size 31, and 5.4% for sample size 59. For each sample size, the bias is almost monotonically related to the kurtosis of the explanatory variables, and the extreme biases occur where the x kurtosis is extraordinarily high (the kurtosis of the 31 observations of x leading to the extreme downward bias is 11.27, and the largest downward biases with 59 observations correspond to x kurtoses of 16.62, 14.09 and 12.17). For $T = 31$, the A_{LS} confidence regions are superior, with a smaller average error, and with 6 rejections versus 9 for A_{LAR} at the 99% level of confidence. For $T = 59$, the A_{LAR} confidence regions are superior, with the smaller average error, and with 4 rejections versus 8 for A_{LS}. The degree of underestimation of the tails of the error distribution is now substantial, and warrants attention in hypothesis testing.

To summarize the results of this section, A_{LAR} is not significantly inferior to the familiar LS normal approximation A_{LS}, with the exception that when both the x distribution and the u distribution

have high kurtosis, A_{LAR} understates the standard deviation of the estimation error.

V. The Case of Several Explanatory Variables

In multivariate regression with the explanatory variable matrix $\underset{\sim}{X}$, let $\underset{\sim}{M}_{LAR}$ be the variance matrix of estimation errors for the multivariate LAR estimator. Let $\underset{\sim}{D}\underset{\sim}{D}' = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}$. Then the transformed variables $w_1,\ldots,w_K$ defined by $\underset{\sim}{W} = \underset{\sim}{X}\underset{\sim}{D}$ are orthonormal, that is, $\underset{\sim}{W}'\underset{\sim}{W} = \underset{\sim}{I}$. It follows from Theorem 3 that $\underset{\sim}{M}_{LAR} = \underset{\sim}{D}\underset{\sim}{J}_{LAR}\underset{\sim}{D}'$, where $\underset{\sim}{J}_{LAR}$ is the variance matrix of the LAR estimation errors in the transformed regression with explanatory variables w. Thus, the variance matrix of the LAR estimation error can always be related to the variance matrix in a multivariate regression with orthonormal explanatory variables.

We know that for the LS estimator, $\underset{\sim}{J}_{LS} = \sigma^2\underset{\sim}{I}$, and hence $\underset{\sim}{M}_{LS} = \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$. From a geometrical point of view, this holds because LS is a projection, with quadratic norm L_2, of the vector $\underset{\sim}{y}$ onto the subspace χ spanned by $\underset{\sim}{X}$ within Euclidean T-dimensional space, and this projection decomposes into the sum of projections onto any orthogonal basis for the subspace χ. When the explanatory variables are transformed into the variables w, the estimates of the regression coefficients for these transformed variables are distinct orthogonal linear functions of the dependent variables, yielding estimation errors that are uncorrelated when disturbances have finite variance, are independent, and ——————→ ——————→ are normally distributed. In deriving the sampling distribution for each of these transformed regression coefficients, either

the familiar univariate approach or the indirect approach used in Section III (see fn. 4, p. 15) can be followed. In either case, the results yield the estimation error variance matrix $\sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ in the original coordinate system.

The least-alpha estimators are projections in L_α and the difficulties in developing multivariate sampling theory can be associated with the properties of these spaces. For L_α spaces other than L_2, there is no analog to orthogonality and, indeed, distance cannot be expressed in terms of an inner product over any basis. For this reason, it is impossible to decompose the multivariate estimator into a set of unrelated univariate components. Thus, although the w are orthonormal with respect to a quadratic norm, there is no assurance that their least-alpha parameter estimates will be uncorrelated, and it is not possible to construct analytically a basis for χ which will possess this property.

When the multivariate estimator cannot be decomposed into univariate elements, the slope of the objective function cannot be used to bound the estimation error--it is quite possible for the slope of a multivariate objective function to be positive with respect to a parameter at some point, although the minimum of the objective function is reached at a greater value of that parameter. Hence, the approach of Section III cannot be generalized exactly. Moreover, the complexity of the LAR projection makes it unlikely that any easily computable function of the explanatory variables will exist which is analogous to the information matrix $\underset{\sim}{X}'\underset{\sim}{X}$ in least squares.

Thus, the outlook is pessimistic, insofar as generalizing the exact distribution is concerned, and it appears certain that the computational difficulty will be at least as great as for the univariate case, where it is prohibitive. Accordingly, an approximate sampling theory appears to be required for any applications. It seems reasonable to expect that if the variables w are orthonormal, the corresponding LAR parameter estimates will be virtually uncorrelated and that their marginal distributions will follow the univariate approximation developed in Section IV. In this case $\underset{\sim}{J}_{LAR} \sim \lambda(F,T)\underset{\sim}{I}$, and $\underset{\sim}{M}_{LAR} \sim \lambda(F,T)(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$. The Monte Carlo results in Section VI below confirm this approximation, in that the multivariate results are entirely consistent with the univariate results already reported. The reader who is primarily interested in applications can skip to Section VI.

One possible approach to the exact probability distribution of the multivariate LAR estimates is as follows: examining the LAR regression as a linear programming problem, it is easy to establish that when the explanatory variables are not linearly dependent, the regression hyperplane $\hat{\underset{\sim}{y}} = \underset{\sim}{X}\underset{\sim}{\beta}$ will pass through K observations of the dependent variables. (In rare cases where the optimal regression estimates are not uniquely determined, some optimal estimate will satisfy this condition; and in cases where more than K of the observation vectors $(y_t : \underset{\sim}{x}_t)$ are linearly dependent, the regression hyperplane may pass through more than K observations.) Thus, in general, K or more of the residuals in the final estimated regression will be zero, and the corresponding observations will determine the hyperplane.

We can speak of the K explanatory variable vectors as a basis. It can be shown that as long as a set of K linearly independent explanatory variable vectors exists, there will always be a linearly independent basis that yields the minimal value of the objective function and, for a continuous disturbance distribution, there is zero probability that a linearly dependent basis can also yield the minimal value. One possible approach to the sampling distribution is to partition the problem by deducing necessary and sufficient conditions for each set of K regression observations to be selected as those determining the regression hyperplane.

From the linear programming viewpoint, the condition that the estimate is optimal is that introduction of no other regression observation into the basis can lead to a reduction in the objective function. For ease of exposition, suppose that the observations are renumbered so as to have the first K be those in the optimal basis. Since these vectors $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_K$ will be linearly independent, let each of the remaining T-K explanatory variable vectors be expressed in terms of the basis vectors as $\underset{\sim}{x}_t' = (\underset{\sim}{x}_1' \ \ldots \ \underset{\sim}{x}_K')\underset{\sim}{a}_t = \sum_{i=1}^{K} a_{ti}\underset{\sim}{x}_i'$. Then it is readily verified that the erroneous component of the regression hyperplane at the t^{th} explanatory variable vector is $\underset{\sim}{x}_t(\underset{\sim}{b}-\underset{\sim}{\beta}) = \underset{\sim}{a}_t'\underset{\sim}{u}^*$ where $\underset{\sim}{u}^*$ is the vector of disturbances in the K basis observations. Hence, the fitting error in the regression hyperplane is given for the t^{th} observation by $\hat{y}_t - y_t = \underset{\sim}{a}_t'\underset{\sim}{u}^* - u_t$.

If we now consider rotating the regression hyperplane through the shortest arc from some basis observation $(y_i : \underset{\sim}{x}_i)$ to some excluded observation $(y_j : \underset{\sim}{x}_j)$, keeping the remaining basis observations unchanged, it can be shown that for a sufficiently small rotation, the shift in the objective function is proportional to

$$(5.1) \qquad 1 - \text{sign}[a_{ji}(\underset{\sim}{a}_j'\underset{\sim}{u}^* - u_j)] \left\{ \sum_{t=K+1}^{N} a_{ti} \, \text{sign}[\underset{\sim}{a}_t'\underset{\sim}{u}^* - u_t] \right\}$$

and that the sign of this expression gives the sign of the change in the objective function when the hyperplane is rotated until the nearest observation is touched and enters the basis. (To make the condition completely general, note that when one of the excluded observations also lies on the hyperplane by chance, so that $\underset{\sim}{a}_t'\underset{\sim}{u}^* - u_t = 0$, the expression "$\text{sign}[\underset{\sim}{a}_t'\underset{\sim}{u}^* - u_t]$" should be assigned whichever value will lead to an increase in the objective function, i.e. $-\text{sign}[a_{ji}(\underset{\sim}{a}_j'\underset{\sim}{u}^* - u_j)\, a_{ti}]$.

Since the present basis will be at an optimum if and only if the shift in the objective function is nonnegative for all such rotations, the necessary and sufficient condition for the present basis to be optimal is that, for all pairs of included and excluded observations,

$$(5.2) \qquad 1 \geq \text{sign}[a_{ji}[\underset{\sim}{a}_j'\underset{\sim}{u}^* - u_j]] \left\{ \sum_{t=K+1}^{N} a_{ti} \, \text{sign}[\underset{\sim}{a}_t'\underset{\sim}{u}^* - u_t] \right\} .$$

The term multiplying the summation can be chosen from among all the excluded observations and, therefore, a term can always be found that

has the same sign as the summation. Hence, this condition is equivalent to the condition that, for i=1,...,K,

$$\left| \sum_{t=K+1}^{N} a_{ti} \operatorname{sign}[a_{\sim t\sim}' u^* - u_t] \right| \leq 1.$$

Thus, we have:

LEMMA 4. When the explanatory variables are of full rank K, a necessary and sufficient condition for a set of K observations with indices $(i_1,\ldots,i_K \in I)$ to be chosen as uniquely determining the LAR parameter estimates, $\hat{b} = \Lambda'^{-1} y^*$ where $y^* = \begin{pmatrix} y_{i_1} \\ \vdots \\ y_{i_K} \end{pmatrix}$ and $\Lambda = (x_{\sim i_1}' \ \ldots \ x_{\sim i_K}')$, is that the vectors $x_{\sim i_1}, \ldots, x_{\sim i_K}$ be linearly independent and that, for i=1,...,K,

$$\left| \sum_{t \notin I} a_{ti} \operatorname{sign}[a_{\sim t\sim}' u^* - u_t] \right| < 1, \tag{5.3}$$

where for $t \notin I$, $a_{\sim t} = \Lambda^{-1} x_{\sim t}'$, and where $u^* = \begin{pmatrix} u_{i_1} \\ \vdots \\ u_{i_K} \end{pmatrix}$. When there is no unique estimate, a linearly independent set will exist which yields the minimal sum of the objective function and which satisfy condition (5.3) with $\leq$ rather than $<$.

We may use this lemma to express the probability distribution of the estimation error vector. For any finite sample size T, there

will be $\frac{T!}{(T-K)!\ K!}$ possible bases, each comprising K vectors. We may eliminate those bases which contain a set of linearly dependent vectors. Let $\mathcal{A}$ be the set of all linearly independent bases. The probability element for a given estimation error vector becomes:

$$(5.4)\qquad P[\underset{\sim}{e}=\underset{\sim}{\eta}] = \sum_{\underset{\sim}{\Lambda}\in\mathcal{A}} \left(P[\underset{\sim}{u}=\underset{\sim}{\Lambda}'\underset{\sim}{\eta}] \cdot P\left[\left| \sum_{t\notin I} a_{ti}\,\mathrm{sign}(\underset{\sim t\sim}{a}'\underset{\sim}{u}-u_t) \right| \leq 1 \text{ for } i\in I \right] \right).$$

The first term is just $\prod_{i\in I} [f(\underset{\sim i\sim}{x}\underset{\sim}{\eta})]$, where f is the density function of $\underset{\sim}{u}$. The second term may be simplified by noting that $\underset{\sim t\sim}{a}'\underset{\sim}{u} = \underset{\sim t\sim}{x}'\underset{\sim}{\Lambda}'^{-1}\underset{\sim}{u} = \underset{\sim t\sim}{x}\underset{\sim}{\eta}$ and that $a_{ti} = \underset{\sim}{\Lambda}^{i}\underset{\sim t}{x}'$, where $\underset{\sim}{\Lambda}^{i}$ is the i^{th} row of $\underset{\sim}{\Lambda}^{-1}$. Hence,

$$(5.5)\qquad P[\underset{\sim}{e}=\underset{\sim}{\eta}] = \sum_{\underset{\sim}{\Lambda}\in\mathcal{A}} \left(\left(\prod_{i\in I} f(\underset{\sim i\sim}{x}\underset{\sim}{\eta}) \right) P\left[\left| \underset{\sim}{\Lambda}^{j} \sum_{t\notin I} \underset{\sim t}{x}' \,\mathrm{sign}\,(\underset{\sim t\sim}{x}\underset{\sim}{\eta}-u_t) \right| \leq 1 \right.\right.$$
$$\left.\left. \text{for } j=1,\ldots,K \right] \right).$$

This condition is somewhat sharper than (2.7) in that the only directions of movement which are considered for each basis are the K rows of $\underset{\sim}{\Lambda}^{-1}$, i.e. $\underset{\sim}{d} = \underset{\sim}{\Lambda}^{1},\ldots,\underset{\sim}{\Lambda}^{K}$. However, it is less manageable since the probability has been expressed as the sum over all possible bases. The condition reduces to condition (2.7) in the univariate case. It may yield multivariate analogues to Theorems 2 and 3. L. D. Taylor has conducted some Monte Carlo studies of the probability of occurrence of bases [19]. However, the appropriate direction for future research on the exact multivariate sampling distribution does not emerge clearly.

VI. An Approximate Multivariate Sampling Theory

In Tables 6.1 through 6.6, results of Monte Carlo experiments identical in format to the univariate experiments are given for regressions with two explanatory variables and a constant. The previous summary statistics are given for each of the three parameter estimates. The distributions of these test statistics remain as in Section 4. In addition, the statistic

$$w = -2\log(((e/1000)^{K}\,\mathrm{DET}(\underset{\sim}{B}\underset{\sim}{M}^{-1}))^{1000/2}\, e^{-\frac{1}{2}\mathrm{TR}(\underset{\sim}{B}\underset{\sim}{M}^{-1})})$$

where

$$\underset{\sim}{e}_i = \underset{\sim}{\hat{\beta}}_i - \beta, \qquad \underset{\sim}{\bar{e}} = \frac{\sum_{i=1}^{1000} \underset{\sim}{e}_i}{1000}$$

$$\underset{\sim}{M} = \begin{cases} \lambda(\underset{\sim}{X}'\underset{\sim}{X})^{-1} & \text{for LAR} \\ \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1} & \text{for LS} \end{cases},$$

$$\underset{\sim}{B} = \sum_{i=1}^{1000} (\underset{\sim}{e}_i - \underset{\sim}{\bar{e}})(\underset{\sim}{e}_i - \underset{\sim}{\bar{e}})',$$

is given as a test of the coincidence of the dispersion of the estimation error vector with the approximate normal distribution A. Under the null hypothesis that the approximated multivariate distribution is valid, the statistic w is asymptotically distributed as $\chi^2_{\frac{K(K+1)}{2}}$. In these regressions, where K = 3, w is approximately distributed with mean equal to 6, and the 99% confidence region for rejection of excessive variance is (0,16.81).

In Tables 6.7 and 6.8, Monte Carlo experiments with six explanatory variables and a constant are reported. The format of the

experiments is the same as before, but only the standard errors of the parameter estimates have been computed. Here the statistic w has mean value equal to 28 under the null hypothesis, and the 99% confidence region for rejection of excessive variance is (0,48.28). Finally, in Table 6.9 the results of the multivariate Monte Carlo experiments are summarized as the univariate results were summarized in Table 4.5. There are 42 multivariate regressions, with 150 estimated parameters in all.

The explanatory variables are constructed so as to be orthogonal to one another and to the constant. As a precautionary measure, the variances of the explanatory variables differ by factors of as much as 1000. This was done in order to confirm that LAR is not sensitive to the numerical difficulties which arise from multicollinearity; the results of the previous section have already shown that the accuracy of A_{LAR} is not affected by the second moments of the explanatory variables. Following the reasoning of the previous section, we hoped that the marginal distributions of the estimated parameters would approximate the distributions already encountered in the univariate regressions; that is, the constant term would be distributed as the median, the regression coefficients for normally distributed variables would be distributed as in univariate regression on a normal explanatory variable, etc. We also hoped that the different parameter estimates would be approximately uncorrelated, since the explanatory variables are orthogonal. The results are entirely consistent with these expectations.

TABLE 6.1--Normal Disturbance Distribution

Regression with a Constant and 2 Normally Distributed Variables

X	Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
			w				w						
	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31													
Experiment 1	1.004	2.99		1.62				6.00					
Constant			1.011	5.3	1.4	1.8	1.030	5.6	1.5	3.9	1.24	1.27	.80
X1			0.999	5.2	1.0	3.8	1.033	5.6	1.1	4.6	1.24	1.29	.80
X2			0.984	5.1	0.8	2.8	0.999	4.9	1.3	1.7	1.24	1.26	.78
Sample Size T = 59													
Experiment 2	0.994	2.87		0.803				2.53					
Constant			0.995	6.0	1.7	2.8	0.984	4.2	1.1	1.9	1.24	1.24	.80
X1			0.998	5.4	0.6	2.0	0.974	4.4	0.4	2.1	1.24	1.22	.79
X2			0.988	4.3	1.3	2.1	0.990	4.9	0.7	2.0	1.24	1.25	.80
Experiment 3	0.998	2.86		1.436				2.37					
Constant			1.009	5.5	1.0	2.6	1.025	5.2	1.3	3.4	1.24	1.27	.79
X1			1.013	5.6	1.2	1.9	1.007	5.1	1.3	2.4	1.24	1.24	.80
X2			1.003	4.8	0.7	1.9	0.986	4.1	0.8	2.2	1.24	1.23	.80
Experiment 4	1.00	2.91		5.815				2.52					
Constant			1.003	5.1	0.9	2.0	1.014	4.9	0.8	5.2	1.24	1.26	.80
X1			0.988	4.3	1.0	2.6	0.995	5.1	0.5	2.7	1.24	1.26	.80
X2			0.983	5.0	0.8	2.3	0.990	5.1	0.8	2.7	1.24	1.26	.80
Experiment 5	0.995	2.88		5.973				7.02					
Constant			1.027	4.9	0.9	2.5	0.992	4.8	0.8	2.2	1.24	1.22	.82
X1			1.039	6.0	1.2	3.3	1.045	6.2	1.7	3.3	1.24	1.26	.82
X2			0.977	5.0	1.3	2.2	0.962	4.2	1.4	3.6	1.24	1.22	.78
Experiment 6	1.00	2.89		3.988				4.31					
Constant			1.021	4.7	1.3	3.0	1.006	5.4	1.1	2.5	1.24	1.24	.80
X1			0.967	4.5	1.6	5.0	1.007	6.5	1.3	3.5	1.24	1.30	.80
X2			1.017	5.9	1.0	1.5	1.036	5.5	1.2	2.3	1.24	1.27	.81

TABLE 6.2—Contaminated (0.15, 16.0) Normal Disturbance Distribution

Regression with a Constant and 2 Normally Distributed Variables

X	Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
				w				w					
	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31													
Experiment 1	3.24	7.08		6.73				9.46					
Constant			.974	3.8	1.0	2.9	1.044	5.5	0.9	3.7	.782	.837	.65
X1			.979	4.3	1.1	2.8	1.022	4.8	1.0	2.8	.782	.816	.67
X2			1.022	6.1	2.1	3.5	1.037	6.7	1.7	2.3	.782	.794	.68
Sample Size T = 59													
Experiment 2	3.22	8.54		15.450				11.83					
Constant			1.011	4.9	1.2	3.4	0.976	5.3	1.0	3.0	.783	.756	.61
X1			1.052	7.0	2.2	2.8	1.034	6.5	1.8	2.2	.783	.772	.67
X2			0.942	4.9	0.7	6.9	0.977	4.6	1.1	3.4	.783	.811	.64
Experiment 3	3.25	8.56		2.898				4.57					
Constant			0.982	5.1	1.2	2.7	0.996	5.0	1.1	3.3	.783	.793	.65
X1			0.982	5.0	1.0	3.5	0.974	4.4	1.2	2.9	.783	.778	.63
X2			1.000	5.3	1.5	1.9	1.019	5.3	1.6	1.6	.783	.798	.62
Experiment 4	3.28	8.66		6.856				11.61					
Constant			1.022	6.0	1.5	1.6	1.017	4.5	1.1	2.2	.783	.777	.62
X1			1.045	6.4	1.4	3.3	1.012	6.7	1.3	5.6	.783	.802	.68
X2			1.004	6.4	1.2	3.2	1.015	5.5	1.6	1.9	.783	.790	.66
Experiment 5	3.21	8.56		10.500				8.94					
Constant			0.970	4.6	0.8	3.0	0.964	4.2	0.7	2.9	.783	.778	.62
X1			0.956	4.4	0.9	6.5	0.974	4.5	1.1	3.8	.783	.793	.66
X2			0.964	4.0	0.9	4.5	0.964	3.4	0.9	2.5	.783	.787	.63
Experiment 6	3.27	8.75		2.418				6.40					
Constant			0.994	5.8	1.0	4.6	1.019	5.3	1.0	4.5	.783	.802	.67
X1			0.993	6.0	1.1	4.2	1.020	5.0	1.1	2.7	.783	.804	.66
X2			0.972	4.7	1.0	3.9	0.995	5.7	1.3	3.6	.783	.799	.62

TABLE 6.3--Contaminated (0.15, 25.0) Normal Disturbance Distribution

Regression with a Constant and 2 Normally Distributed Variables

X	Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
			w				w						
	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31													
Experiment 1	4.60	8.29		6.87				11.80					
Constant			1.002	5.2	1.2	3.0	1.019	5.2	1.7	1.9	.664	.675	.60
X1			0.975	5.4	1.3	6.5	1.044	5.1	1.8	2.1	.664	.710	.64
X2			0.981	5.5	1.6	7.8	1.035	5.2	1.7	3.6	.664	.701	.64
Sample Size T = 59													
Experiment 2	4.53	10.16		.775				4.77					
Constant			0.995	4.8	0.9	1.6	1.027	6.3	1.1	2.3	.664	.685	.63
X1			1.007	5.4	1.7	3.9	1.023	5.3	1.5	2.0	.664	.674	.62
X2			1.003	5.5	1.0	3.0	1.011	6.1	1.0	1.7	.664	.669	.62
Experiment 3	4.52	10.47		1.125				2.00					
Constant			0.988	5.0	1.2	3.8	1.004	5.4	1.7	2.0	.664	.676	.56
X1			0.988	5.5	1.2	5.7	1.018	5.6	1.7	1.5	.664	.684	.62
X2			1.014	5.6	1.2	2.1	1.014	5.3	1.1	2.1	.664	.663	.62
Experiment 4	4.52	10.25		1.051				10.48					
Constant			0.992	4.6	2.1	3.7	1.015	4.9	1.8	1.8	.664	.676	.62
X1			0.996	4.7	1.2	2.5	0.985	4.1	1.2	2.2	.664	.657	.59
X2			0.996	4.8	1.8	3.4	1.012	5.5	1.1	2.1	.664	.676	.62
Experiment 5	4.64	10.41		6.192				4.78					
Constant			0.980	4.7	1.3	4.4	0.985	5.9	1.2	4.5	.664	.668	.57
X1			1.025	5.5	1.6	2.1	1.044	5.2	1.2	4.7	.664	.676	.58
X2			1.037	5.6	1.4	3.7	1.006	5.5	1.1	1.5	.664	.644	.62
Experiment 6	4.64	10.26		3.688				7.06					
Constant			1.002	5.2	1.5	3.2	0.984	4.9	1.0	2.7	.664	.652	.61
X1			0.990	5.1	1.4	6.3	1.010	5.2	1.0	2.7	.664	.678	.61
X2			1.012	5.2	1.5	2.4	1.033	6.2	2.0	1.6	.664	.678	.58

TABLE 6.4--Normal Disturbance Distribution

Regression with a Constant and 2 Widely Dispersed X Variables

X Distribution		Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
				w				w						
σ^2	μ_4/σ^4	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31														
Experiment 1		.999	2.84	8.72				11.08						
Constant				1.018	5.5	1.1	1.7	1.024	5.2	0.6	3.7	1.245	1.272	.80
2.15	3.16			0.995	4.4	1.1	1.6	0.975	6.0	1.3	1.9	1.245	1.236	.80
3.76	9.26			0.982	4.9	1.1	2.2	1.015	4.5	1.3	1.9	1.245	1.251	.80
Experiment 2		.996	2.79	7.92				3.59						
Constant				0.960	3.9	0.8	2.3	0.990	4.8	1.1	3.0	1.245	1.210	.80
3.38	9.41			1.009	5.3	1.2	2.0	0.989	5.0	1.0	1.6	1.245	1.286	.81
3.36	4.75			1.006	4.7	1.2	2.1	0.978	4.1	0.3	1.9	1.245	1.214	.81
Experiment 3		.996	2.82	3.19				5.93						
Constant				1.000	5.2	0.7	2.2	0.971	5.0	1.1	2.0	1.245	1.210	.80
3.09	6.29			0.985	4.7	0.8	1.6	0.995	3.3	0.5	2.5	1.245	1.251	.82
3.14	12.50			0.975	4.1	0.4	2.3	1.000	3.9	0.7	3.0	1.245	1.226	.80
Sample Size T = 59														
Experiment 4		.997	2.86	14.25				11.24						
Constant				1.002	4.4	0.7	2.8	1.014	5.7	1.0	2.0	1.249	1.263	.82
3.30	9.59			1.070	7.0	1.6	4.4	1.042	6.0	1.5	4.6	1.249	1.219	.81
3.90	7.41			1.024	6.5	1.2	1.8	1.034	6.0	1.3	3.3	1.249	1.261	.80
Experiment 5		.997	2.90	17.06				11.24						
Constant				0.969	5.0	1.2	5.5	0.997	4.4	1.1	2.6	1.249	1.280	.79
3.10	7.97			1.032	6.4	1.6	3.5	1.011	5.7	1.0	3.2	1.249	1.222	.81
7.58	8.64			0.969	4.5	0.9	3.1	0.941	3.9	0.7	4.4	1.249	1.212	.80
Experiment 6		1.005	2.88	2.35				6.03						
Constant				0.995	5.0	0.9	1.3	0.969	4.4	0.5	1.7	1.249	1.219	.80
3.70	6.93			0.998	5.3	0.4	3.0	0.992	5.1	1.1	4.1	1.249	1.241	.79
2.08	6.70			1.002	5.6	1.2	3.0	0.986	4.9	1.6	3.8	1.249	1.226	.81

TABLE 6.5--Contaminated (0.15, 16.0) Normal Disturbance Distribution

X Distribution		Actual U Distribution		Performance of LS Approximation w				Performance of LAR Approximation w				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
σ^2	μ_4/σ^4	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31														
Experiment 1		3.30	6.93		10.31				126.18					
Constant				1.021	5.8	1.3	2.2	1.039	5.9	1.2	4.5	.771	.795	.67
5.26	7.20			1.018	7.0	2.5	9.5	1.100	6.9	2.5	3.2	.771	.847	.71
3.35	16.12			1.060	5.7	3.8	10.5	1.233	8.7	4.9	4.4	.771	.908	.80
Experiment 2		3.11	6.74		3.83				23.96					
Constant				0.977	4.4	1.1	2.5	1.065	7.1	1.6	5.2	.771	.850	.67
2.06	5.43			0.994	4.9	1.8	6.8	1.054	6.7	2.3	3.6	.771	.832	.69
2.97	8.16			0.971	5.1	2.2	7.6	1.075	5.2	2.7	1.9	.771	.864	.73
Experiment 3		3.25	6.57		8.04				42.51					
Constant				0.997	5.9	1.5	5.8	1.035	5.5	1.1	4.3	.771	.812	.64
4.14	15.03			0.961	4.9	1.8	10.7	1.052	5.7	2.8	4.6	.771	.858	.74
9.63	7.73			1.048	6.4	3.1	8.5	1.130	7.3	3.3	3.8	.771	.844	.73
Sample Size T = 59														
Experiment 4		3.19	8.47		5.57				5.54					
Constant				0.966	3.9	1.3	2.7	0.992	4.8	1.0	2.5	.783	.803	.62
3.56	5.56			1.012	5.6	1.7	4.2	1.004	5.2	1.0	1.8	.783	.777	.69
3.52	7.20			0.995	5.3	1.0	3.7	1.024	4.9	1.3	1.7	.783	.807	.66
Experiment 5		3.24	8.50		1.35				36.80					
Constant				1.001	4.7	0.8	2.7	0.996	4.8	0.9	1.9	.783	.767	.66
3.13	11.89			1.003	5.9	1.8	5.2	1.050	6.8	2.1	2.6	.783	.815	.72
6.00	18.11			1.023	7.0	2.8	8.4	1.135	8.4	2.6	3.9	.783	.862	.71
Experiment 6		3.28	8.54		10.71				38.66					
Constant				1.035	5.6	1.7	2.7	1.034	5.8	1.8	2.8	.783	.782	.66
3.21	9.35			1.045	6.0	2.3	2.6	1.104	7.3	1.8	5.7	.783	.826	.70
3.42	6.49			1.020	5.7	2.1	4.2	1.090	7.6	2.4	3.8	.783	.833	.67

TABLE 6.6—Contaminated (0.15, 25.0) Normal Disturbance Distribution

X Distribution		Actual U Distribution		Performance of LS Approximation				Performance of LAR Approximation				$\frac{\sigma_{LAR}}{\sigma_{LS}}$	$\frac{S_{LAR}}{S_{LS}}$	Corr.
				w				w						
σ^2	μ_4/σ^4	σ^2	μ_4/σ^4	S/σ	>95	>99	d	S/σ	>95	>99	d			
Sample Size T = 31														
Experiment 1		4.65	8.23		15.04				35.36					
Constant				0.972	3.8	1.3	4.8	0.980	4.8	0.7	3.9	.664	.668	.59
1.98	2.35			1.068	7.2	2.7	5.8	1.118	7.6	2.1	4.1	.664	.695	.67
3.29	5.09			1.013	5.5	2.0	3.4	1.055	6.2	1.3	4.2	.664	.690	.62
Experiment 2		4.55	8.16		7.57				108.39					
Constant				1.020	5.4	1.9	2.3	1.018	5.5	1.3	3.1	.664	.663	.61
3.42	9.18			0.956	5.0	2.3	15.9	1.137	8.5	3.0	4.2	.664	.789	.68
12.01	11.50			0.983	5.7	3.0	15.7	1.195	7.4	3.2	4.0	.664	.804	.66
Experiment 3		4.64	8.32		7.52				40.00					
Constant				1.034	5.9	1.6	2.8	1.005	5.1	1.1	6.0	.664	.636	.60
1.42	3.92			1.020	6.2	2.7	6.7	1.100	6.3	1.7	3.7	.664	.705	.65
1.62	3.94			0.967	5.1	1.9	9.4	1.080	6.3	1.8	4.7	.664	.731	.61
Sample Size T = 59														
Experiment 4		4.63	10.56		11.84				70.49					
Constant				1.009	5.5	1.4	2.5	0.984	4.3	0.6	1.9	.664	.641	.60
6.71	11.03			1.008	6.1	2.1	7.3	1.082	6.9	2.3	2.7	.664	.707	.63
4.19	20.77			0.985	5.6	2.6	10.6	1.185	8.4	4.0	4.0	.664	.792	.74
Experiment 5		4.71	10.40		5.59				15.45					
Constant				1.022	5.8	1.7	3.2	1.016	5.9	1.7	1.5	.664	.659	.60
5.65	7.72			0.988	5.6	2.1	8.4	1.040	5.6	1.5	3.0	.664	.697	.60
3.74	14.26			1.013	6.7	1.4	8.3	1.078	6.0	2.2	2.2	.664	.705	.64
Experiment 6		4.68	10.50		6.79				3.91					
Constant				0.980	5.7	1.0	4.3	0.996	4.9	1.1	2.1	.664	.674	.58
2.11	6.23			0.994	5.0	2.0	5.5	1.014	5.9	1.0	1.8	.664	.679	.60
5.35	7.48			1.033	6.4	2.0	5.0	0.988	4.6	1.1	2.3	.664	.636	.63

TABLE 6.7--Regression on a Constant and 6 Normally Distributed Variables, T = 59

Disturbance Distribution	Normal $\sigma^2 = 0.994$ $\mu_4/\sigma^4 = 2.90$			Contaminated (0.15,16.0) $\sigma^2 = 3.22$ $\mu_4/\sigma^4 = 8.52$			Contaminated (0.15,25.0) $\sigma^2 = 4.56$ $\mu_4/\sigma^4 = 10.3$		
w_{LS}	20.5			37.7			40.9		
w_{LA}	28.5			31.6			34.1		
	$\frac{S_{LS}}{\sigma_{LS}}$	$\frac{S_{LAR}}{\sigma_{LAR}}$	$\frac{S_{LAR}}{S_{LS}}$	$\frac{S_{LS}}{\sigma_{LS}}$	$\frac{S_{LAR}}{\sigma_{LAR}}$	$\frac{S_{LAR}}{S_{LS}}$	$\frac{S_{LS}}{\sigma_{LS}}$	$\frac{S_{LAR}}{\sigma_{LAR}}$	$\frac{S_{LAR}}{S_{LS}}$
Constant	1.003	1.000	1.24	1.004	1.001	.780	1.001	1.013	0.666
X1	0.974	0.960	1.22	0.948	0.972	.798	0.950	0.976	0.679
X2	1.000	1.010	1.25	0.968	1.021	.828	0.963	1.040	0.711
X3	1.025	1.014	1.23	1.027	1.049	.798	1.027	1.060	0.678
X4	1.020	1.036	1.26	1.000	1.037	.813	0.998	1.050	0.694
X5	0.988	1.000	1.26	0.974	0.992	.799	0.971	1.010	0.686
X6	0.974	0.999	1.27	0.968	1.006	.817	0.966	1.012	0.692

TABLE 6.8--Regression on a Constant and 6 Highly Dispersed Variables, T = 59

Disturbance Distribution	Normal $\sigma^2 = 1.004$ $\mu_4/\sigma^4 = 2.88$				Contaminated (0.15,16.0) $\sigma^2 = 3.28$ $\mu_4/\sigma^4 = 8.65$				Contaminated (0.15,25.0) $\sigma^2 = 4.64$ $\mu_4/\sigma^4 = 10.40$			
w_{LS}		29.66				37.65				62.01		
w_{LA}		23.40				60.02				127.11		
	X-kurt.	$\frac{S_{LS}}{\sigma_{LS}}$	$\frac{S_{LAR}}{\sigma_{LAR}}$	$\frac{S_{LAR}}{S_{LS}}$	X-kurt.	$\frac{S_{LS}}{\sigma_{LS}}$	$\frac{S_{LAR}}{\sigma_{LAR}}$	$\frac{S_{LAR}}{S_{LS}}$	X-kurt.	$\frac{S_{LS}}{\sigma_{LS}}$	$\frac{S_{LAR}}{\sigma_{LAR}}$	$\frac{S_{LAR}}{S_{LS}}$
Constant		0.977	0.942	1.206		1.018	1.049	0.806		0.999	1.022	0.679
X1	4.81	0.988	0.989	1.251	3.90	1.029	1.098	0.835	9.59	1.035	1.060	0.680
X2	13.17	1.006	1.002	1.245	6.68	0.964	1.013	0.822	7.41	1.005	1.103	0.732
X3	7.61	1.037	1.032	1.244	7.84	0.999	1.030	0.808	11.74	1.047	1.104	0.701
X4	12.47	1.020	0.992	1.216	9.45	1.027	1.037	0.790	6.25	1.010	1.054	0.692
X5	14.54	1.021	0.982	1.203	7.20	1.005	1.018	0.796	12.91	1.030	1.088	0.702
X6	7.02	0.988	0.988	1.251	6.18	0.984	1.027	0.817	7.76	1.101	1.130	0.683

TABLE 6.9---Summary of Multivariate Monte Carlo Results

Sample Size and Variable (number of cases in parentheses)		Disturbance Distribution: Normal						Contaminated Normal (0.15, 16.0)						Contaminated Normal (0.15, 25.0)					
		LS			LAR			LS			LAR			LS			LAR		
		S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%	S/σ	>95%	>99%
Regression with a Constant Term and Several Normally Distributed X																			
T = 31	MAX	1.011	5.3	1.4	1.030	5.6	1.5	0.974	3.8	1.0	1.044	5.5	0.9	1.002	5.2	1.2	1.019	5.2	1.7
Constant	MEAN	1.011	5.3	1.4	1.030	5.6	1.5	0.974	3.8	1.0	1.044	5.5	0.9	1.002	5.2	1.2	1.019	5.2	1.7
Term (1)	MIN	1.011	5.3	1.4	1.030	5.6	1.5	0.974	3.8	1.0	1.044	5.5	0.9	1.002	5.2	1.2	1.019	5.2	1.7
T = 31	MAX	0.999	5.2	1.0	1.033	5.6	1.3	1.022	6.1	2.1	1.037	6.7	1.7	0.981	5.5	1.6	1.044	5.2	1.8
Normal X	MEAN	0.992	5.2	0.9	1.017	5.2	1.2	1.000	5.2	1.6	1.030	5.8	1.4	0.978	5.4	1.4	1.040	5.2	1.8
(2)	MIN	0.984	5.1	0.8	0.999	4.9	1.1	0.979	4.3	1.1	1.022	4.8	1.0	0.975	5.4	1.3	1.035	5.1	1.7
T = 59	MAX	1.027	6.0	1.7	1.025	5.4	1.3	1.022	6.0	1.5	1.019	5.3	1.1	1.002	5.2	2.1	1.027	6.3	1.8
Constant	MEAN	1.010	5.2	1.2	1.003	4.9	1.0	0.997	5.3	1.1	0.996	4.9	1.0	0.992	4.9	1.4	1.005	5.5	1.4
Term (6)	MIN	0.995	4.7	0.9	0.984	4.2	0.8	0.970	4.6	0.8	0.964	4.2	0.7	0.980	4.6	0.9	0.984	4.9	1.0
T = 59	MAX	1.039	6.0	1.6	1.045	6.5	1.7	1.052	7.0	2.2	1.072	6.7	1.8	1.037	5.6	1.8	1.060	6.2	2.0
Normal X	MEAN	0.997	5.1	1.1	1.001	5.1	1.0	0.987	5.4	1.2	0.996	5.2	1.3	0.996	5.3	1.4	1.019	5.4	1.3
(17)	MIN	0.967	4.3	0.6	0.960	4.1	0.4	0.942	4.0	0.7	0.964	3.4	0.9	0.950	4.7	1.0	0.976	4.1	1.0
Regression with a Constant Term and Several Highly Dispersed X																			
T = 31	MAX	1.018	5.5	1.1	1.024	5.2	1.1	1.021	5.9	1.5	1.065	7.1	1.6	1.034	5.9	1.9	1.018	5.5	1.3
Constant	MEAN	0.992	4.9	0.9	0.995	5.0	0.9	0.998	5.4	1.3	1.046	6.2	1.3	1.009	5.0	1.6	1.001	5.1	1.0
Term (3)	MIN	0.960	3.9	0.7	0.971	4.8	0.6	0.977	4.4	1.1	1.035	5.5	1.1	0.972	3.8	1.3	0.980	4.8	0.7
T = 31	MAX	1.009	5.3	1.2	1.015	6.0	1.3	1.060	7.0	3.8	1.233	8.7	4.9	1.068	7.2	2.7	1.195	8.5	3.2
Dispersed X	MEAN	0.992	4.7	1.0	0.992	4.6	0.8	1.009	5.7	2.5	1.108	6.8	3.1	1.001	5.8	2.4	1.114	7.0	2.2
(6)	MIN	0.982	4.1	0.4	0.975	3.3	0.3	0.961	4.9	1.8	1.052	5.2	2.3	0.956	5.0	1.9	1.055	6.2	1.3
T = 59	MAX	1.002	5.0	1.2	1.014	5.7	1.1	1.035	5.6	1.7	1.049	5.8	1.8	1.022	5.5	1.7	1.022	5.9	1.7
Constant	MEAN	0.986	4.8	0.9	0.981	4.8	0.9	1.005	4.7	1.3	1.018	5.1	1.2	1.003	5.7	1.4	1.005	5.0	1.1
Term (4)	MIN	0.969	4.4	0.7	0.942	4.4	0.5	0.966	3.9	0.8	0.992	4.8	0.9	0.980	5.8	1.0	0.984	4.3	0.6
T = 59	MAX	1.070	7.0	1.6	1.042	6.0	1.5	1.045	7.0	1.7	1.135	8.4	2.6	1.101	6.7	2.6	1.185	8.4	4.0
Dispersed X	MEAN	1.013	5.9	1.2	0.999	5.3	1.2	1.009	5.9	2.0	1.053	6.7	1.9	1.021	5.9	2.0	1.077	6.2	2.0
(13)	MIN	0.969	4.5	0.4	0.941	3.9	0.7	0.964	5.3	1.0	1.004	4.9	1.0	0.985	5.0	1.4	0.988	4.6	1.0

For the regressions with normally distributed explanatory variables, the performance of A_{LAR} is quite satisfactory. Only one experiment was run with T=31 for each disturbance distribution; the number of cases is too small to draw firm conclusions, although A_{LS} appears to be slightly superior to A_{LAR}. A_{LAR} is never rejected at the 99% level by any of the statistics, while A_{LS} is rejected twice by the d-statistic. The w-statistics are near the mean value, indicating that the multivariate distribution of estimation errors does not differ significantly from the approximated multivariate normal distribution. One 7-variable and five 3-variable experiments were run with T=59 for each disturbance distribution. A_{LAR} is rejected twice for inappropriate standard error and once for excessive errors beyond the 99% confidence region, while A_{LS} is rejected once for excessive errors beyond the 95% confidence region, twice for excessive errors beyond the 99% confidence region, and three times by the d-statistic. For the constant term in the regression, the performance of A_{LAR} is virtually perfect, with small differences between the average values of the various statistics and their mean values under the null hypothesis. These results are consistent with the results for univariate regression on a constant. For the normally distributed explanatory variables, the A_{LAR} confidence regions perform comparably to the A_{LS} confidence regions, with both slightly understating the tails of the estimation error distribution, and the A_{LAR} estimated standard error has an average downward bias of .8% for the two high kurtosis distributions. These results are again entirely consistent with the univariate results, for normally distributed X. The A_{LAR}

w-statistics are well behaved, clustering around their mean value but with a slight upward bias which is consistent with the understatement of the standard error. In summary, the multivariate A_{LAR} performs satisfactorily for applications with normally distributed X.

For regression with highly dispersed X and normally distributed disturbances, the performance of A_{LAR} continues to be satisfactory. The average values of all statistics are near their mean values under the null hypothesis. A_{LAR} is rejected only once at the 99% level (this rejection, in contrast to all others, is for too high a standard error), while A_{LS} is rejected once for inappropriate standard error, once for excessive errors beyond the 95% confidence region, once by the w-statistic, and once by the d-statistic. (All these rejections of A_{LS} occur in two regressions; since the sampling theory is exact, these rejections should not be regarded as disconfirming the null hypothesis, but rather as random samples from the tail of the sampling distributions of the statistics!)

For regression with high-kurtosis U and high-kurtosis X, A_{LAR} shows the same underestimation of the standard error of estimates as before. The average downward bias for the coefficients of high-kurtosis X is 11.1% for T=31 and 6.5% for T=59, and the bias increases with the kurtosis of the explanatory variable. As a result, the confidence intervals for A_{LAR} show 14 rejections for T=31 versus 8 for A_{LS}, and 10 rejections for T=59 versus 8 for A_{LS}. As before A_{LAR} underestimates the tails of the estimation error distribution more seriously than A_{LS} for T=31, and similarly to A_{LS} for T=59. Again as before, A_{LS} is

rejected far more often (7 times) by the d-statistic than is A_{LAR} (1 time). The underestimation of the standard errors leads to consistently high values for the LAR w-statistic, but the correlations between errors remain near zero.

In these regressions with highly dispersed X, A_{LAR} again performs as in the univariate case. The marginal distribution of the estimation error in the constant term is consistent with the distribution of the median, and the marginal distributions of the coefficients of the dispersed X are consistent with the univariate Monte Carlo results with dispersed X. The presence of an explanatory variable with very high kurtosis in a regression appears to increase the standard errors of the other coefficients slightly, as is to be expected if the estimated parameters are not entirely independent, but the spillover effect is too small to be conclusively demonstrated.

VII. Summary and Conclusion

Three characteristics which are relevant in evaluating an estimator are (i) computational difficulty, (ii) sampling characteristics (the precision of the estimator), and (iii) the availability of a sampling distribution theory. This paper has been concerned with the comparison of the LS and LAR estimators. A first conclusion is that LAR should not be rejected on the basis of computational difficulty alone. An LAR regression program developed by the present authors requires roughly 7 times as many computations as a standard LS regression program, and this difference will be insignificant in most applications.

The authors have attempted to provide a sampling theory for the LAR estimator. The important consequence of Theorem 1 is that the LAR estimator is symmetrically distributed about the true value of the parameter vector when the disturbances are symmetrically distributed. Theorem 3 demonstrates that any linear transformation of the explanatory variables results in the inverse transformation of the estimation errors, so that the problem of multivariate regression can be reduced to that of regression on orthonormal explanatory variables. The exact distribution of the LAR estimator as a function of the disturbance distribution is exhibited for univariate regression, and the estimator is shown to be extremely robust in the presence of massive tails in the disturbance distribution. The univariate LAR estimator is consistent (Theorem 5) and achieves finite variance in small samples (Theorem 4), even in cases where the tails of the disturbance distribution are so massive that the LS estimator is neither consistent nor ever achieves finite variance.

The exact distribution could not be generalized to the multivariate case, and even in the univariate case evaluation of the distribution is prohibitively costly. Accordingly, the following approximate sampling theory was conjectured and evaluated by Monte Carlo studies: in a regression with T observations and disturbance distribution $F(\cdot)$, $\hat{\underset{\sim}{b}}_{LAR} \sim \eta(\underset{\sim}{b}, \lambda(F,T)(\underset{\sim}{X}'\underset{\sim}{X})^{-1})$, where $\frac{\lambda(F,T)}{T}$ is the variance of the median of a sample of size T from the distribution $F(\cdot)$. This approximation has the great virtue of computational simplicity, since tables of $\lambda(F,T)$ can be computed by numerical integration of the exact

distribution for the median (3.9). Moreover, since the approximate distribution differs from the normal approximation to the LS sampling theory only in the scale factor, all of the familiar hypothesis tests and inference procedures of LS sampling theory can be applied as an approximation to LAR, with only this scale factor changed.

The ratios of LAR to LS standard error predicted by the approximation and actually observed in the Monte Carlo studies are given in Table 7.1 for three disturbance distributions and two sample sizes.

TABLE 7.1--Relative Performance of LS and LAR for 3 Disturbance Distributions

Normal or Contaminated Normal Distribution	Kurtosis			Ratios of LAR/LS standard errors							
	Actual	Sample Averages		A_{LAR}		Experimental Averages					
						X constant		X normal		X dispersed	
	--	T=31	T=59	T=31	T=59	T=31	T=59	T=31	T=59	T=31	T=59
Normal	3.0	2.8	2.9	1.245	1.249	1.248	1.246	1.286	1.250	1.238	1.231
(0.15,16.0)	11.8	6.9	8.5	.782	.783	.794	.784	.788	.798	.849	.819
(0.15,25.0)	13.4	8.2	10.4	.664	.664	.661	.669	.690	.675	.739	.698

The results are also divided according to the kurtosis of the explanatory variable, with the "dispersed" X having been constructed with very high kurtosis. The approximated standard errors appear to be entirely satisfactory when the kurtosis of the explanatory variable is not greater than that of the normal distribution. In those cases where both the explanatory variables and the disturbances have high kurtosis, the approximation does understate the standard error of the estimate. The

understatement is severe enough to warrant attention and in a subsequent paper we will discuss a correction to the approximated standard error which is a function of the kurtosis of the explanatory variable. However, the understatement is never large enough to threaten the predicted superiority of LAR over LS for the disturbance distributions with higher kurtosis.

The approximation was also evaluated on the basis of the accuracy of the 95% and 99% confidence intervals for the parameters. The performance is as satisfactory here as in the case of the predicted standard errors, and when compared with the normal approximation to the LS sampling theory, the performance appears to be better still. The LAR estimates are almost exactly normally distributed in the presence of high-kurtosis disturbances, whereas the distribution of the LS estimates has a more massive tail than the normal distribution. (The greater coincidence with normality of the LAR estimator is due to the ability of LAR to ignore extreme values of the disturbances.) As a result, the normal approximation to the LS sampling theory understates the probability of large errors, and this bias is comparable to the LAR understatement resulting from too low a predicted standard error.

Another important result of the Monte Carlo studies is that the multivariate distribution of the LAR estimation error is entirely consistent, according to all statistics which we have measured, with the hypothesis that the multivariate distribution is a direct generalization of the univariate distribution, as proposed in the approximate

sampling theory. The generalization is known to be inexact. Nevertheless, the Monte Carlo results suggest that the Theorems about the existence of moments and consistency of the univariate estimator do generalize in much the same form to the multivariate case.

Many previous articles have suggested that the LAR estimator may be preferable to the LS estimator in applications where the disturbance distribution has massive tails, and our results confirm this. The average measured kurtosis of the two high-kurtosis distributions studied here is given in Table 7.1 for sample sizes of 31 and 59. Kurtosis of this magnitude is frequently encountered in at least two areas of statistical applications. In speculative markets, the distribution of prices often has high kurtosis, and the infinite variance distributions have been proposed as an approximation to the actual price distribution. Also, in disaggregated data of all kinds, extreme values, due both to legitimate random events and to undetected data errors, are common. In these applications LAR should be seriously considered where the problem is sufficiently important to warrant the computational effort. In a sequel to this article, Tables of $\lambda(F,T)$ will be published and some other questions of importance in applications will be considered. In particular, (i) Should the approximated standard error be adjusted in response to the kurtosis of the explanatory variable? (ii) How should the distribution of the disturbances be estimated from the regression residuals? (iii) To what extent are the estimated parameters independent of the estimated $\lambda(F,T)$, and is the application of F and t tests appropriate?

In conclusion, two notes of caution are appropriate. First, the LAR estimator is unbiased only if either the disturbance distribution or the explanatory variable distribution is symmetric. Second, the LAR estimator has been shown to be superior to the LS estimator in the presence of massive-tailed distributions as a result of its ability to minimize the effect of extreme observations, but the LAR estimator has not been shown to be optimal. A nonlinear estimation procedure consisting of a preliminary screen to reject extreme observations followed by the application of least squares to the surviving data may be superior to both LS and LAR.

REFERENCES

[1] Asher, V. G. and T. D. Wallace. "A Sampling Study of Minimum Absolute Deviation Estimates," *Operations Research*, 11 (September-October 1963), 747-758.

[2] Barrodale, I. "L_1 Approximation and the Analysis of Data," *Applied Statistics (JRSS Ser. C)*, 17 (1968), 51-57.

[3] Barrodale, I. and A. Young. "Algorithms for Best L_1 and L_∞ Linear Approximations on a Discrete Set," *Numer. Math.*, 8 (1966), 295-306.

[4] Blattberg, R. and T. Sargent. "Regression with Non-Gaussian Disturbance: Some Sampling Results," *Econometrica* , 39 (May 1971), 501-510.

[5] Cootner, P. (ed.). *The Random Character of Stock Market Prices*. Cambridge, Mass.: M.I.T. Press, 1964.

[6] Cramér, H. *Mathematical Methods of Statistics*. Princeton, N.J.: Princeton University Press, 1963.

[7] Dantzig, G. B. *Linear Programming and Extensions*. Princeton, N.J.: Princeton University Press, 1963.

[8] Darboux, G. (ed.). *Oeuvres de Fourier*. Paris: Gauthier-Villars, 1888-1890, Vol. II, pp. 324-325.

[9] de Bruijn, N. G. *Asymptotic Methods in Analysis*. 2nd ed. Amsterdam: North Holland Publishing Company, 1961.

[10] Edgeworth, F. Y. "A New Method of Reducing Observations Relating to Several Quantities," *Philosophical Magazine*,, 5th Series, 24 (1887), 222-223.

[11] _____. "On a New Method of Reducing Observations Relating to Several Quantities," *Philosophical Magazine*, 5th Series, 25 (1888), 185-191.

[12] Fama, E. and R. Roll, "Some Properties of Symmetric Stable Distributions," *Journal of the American Statistical Association* (September 1968), 817-836.

[13] Fisher, W. D. "A Note on Curve Fitting with Minimum Deviations by Linear Programming," *Journal of the American Statistical Association*, 56 (1961), 359-362.

[14] Glahe, F. R. and J. G. Hunt. "The Small Sample Properties of Simultaneous Equation Least Absolute Estimators vis-à-vis Least Squares Estimators, " *Econometrica*, 38 (September 1970), 742-753.

[15] Hojo, T. "Distribution of the Median, Quartile, and Interquartile Distance in Samples from a Normal Population," *Biometrika* (1931), 316-360.

[16] Karst, O. J. "Linear Curve Fitting Using Least Deviations," *Journal of the American Statistical Association*, 53 (1958), 118-132.

[17] Kendall, M. G. and A. Stuart. *The Advanced Theory of Statistics*. 3 Vols. London: Charles Griffin & Co., 1961-1966.

[18] Ladd, G. W. "Effects of Shocks and Errors in Estimation: An Empirical Comparison," *Journal of Farm Economics*, 38 (1956), 485-495.

[19] Meyer, J. R. and R. R. Glauber. *Investment Decisions, Economic Forecasting, and Public Policy*. Boston, Mass.: Harvard Business School Press, 1964.

[20] Oveson, R. M. "An Analysis of Forecasting Accuracy Using MSAE," unpublished paper, Harvard University, 1966.

[21] Rhodes, E. C. "Reducing Observations by the Method of Minimum Deviations," *Philosophical Magazine*, 7th Series, 9 (1930), 974-992.

[22] Rice, J. R. and J. S. White. "Norms for Smoothing and Estimation," *SIAM Review*, 6 (1964), 243-256.

[23] Rosenberg, B. and D. Carlson. "Least Absolute Residuals Regression Routine," Working Paper No. IP-162. Berkeley: Center for Research in Management Science, University of California, 1970.

[24] Singleton, R. R. "A Method of Minimizing the Sum of the Absolute Values of Deviations," *Annals of Mathematical Statistics*, 11 (1940), 301-310.

[25] Taylor, L. D. "On Estimation by Minimizing the Sum of Absolute Errors," unpublished manuscript, University of Michigan, Ann Arbor, Michigan (December 1970).

[26] Usow, K. H. "On L_1 Approximation: Computation for Continuous Functions and Continuous Dependence," *SIAM Journal of Numerical Analysis*, 4 (1967), 70-88.

[27] _____. "On L_1 Approximation: Computation for Discrete Functions and Discretization Effects," *SIAM Journal of Numerican Analysis*, 4 (1967), 233-244.

[28] Wagner, H. M. "Linear Programming Techniques for Regression Analysis," *Journal of the American Statistical Association*, 54 (March 1959), 206-212.

[29] Wiginton, J. C. "Portfolio Analysis Under the Stable Paretian Hypothesis: An Empirical Evaluation," unpublished manuscript, Capital Markets Research Programme, York University, Toronto, Canada (April 1970).

[30] Zeckhauser, Richard and Mark Thompson. "Linear Regression with Non-Normal Error Terms," *Review of Economics and Statistics*, 52 (August 1970), 280-287.