Linear Regression with Non-Normal Error Terms
Author(s): Richard Zeckhauser and Mark Thompson
Source: *The Review of Economics and Statistics*, Vol. 52, No. 3 (Aug., 1970), pp. 280-286
Published by: The MIT Press
Stable URL: https://www.jstor.org/stable/1926296

# LINEAR REGRESSION WITH NON-NORMAL ERROR TERMS

Richard Zeckhauser and Mark Thompson *

## Background and Introduction

THIS paper considers the linear regression technique when the error term is not assumed to have a normal distribution. The weaker assumption is made that errors are distributed by a power distribution of the class made popular by such papers as Dianada [4], Turner [10], and Box and Tiao [2 and 3]. For these distributions the density at $z$ depends on three parameters, $\mu$, $\sigma$, $\theta$, and is given by

$$f(z;\,\mu,\sigma,\theta) = k(\sigma,\theta)\exp\left\{ -\left|\frac{z-\mu}{\sigma}\right|^{\theta}\right\}, \tag{1a}$$

where

$$k(\sigma,\theta) = [2\sigma\Gamma(1+1/\theta)]^{-1}. \tag{1b}$$

The ranges of the variables are

$$-\infty < z < \infty,$$
$$-\infty < \mu < \infty,$$
$$\sigma > 0 \text{ and } \theta > 0. \tag{1c}$$

The density functions are centered about the location parameter, $\mu$; $\sigma$ is the scale parameter. The parameter $\theta$ measures the degree of peakedness (kurtosis) of the distribution. The multiplicative constant, $k(\sigma,\theta)$, which incorporates the well-tabulated gamma function, serves as a normalizing factor to insure that the area under the density curve is one.[1] For the normal distribution $\theta = 2$; $\theta = 1$ gives the double exponential distribution; where $\theta$ tends to $\infty$, the distribution tends to the rectangular.

We shall investigate the model

$$y_i = a + bx_i + u_i, \qquad i = 1,\ldots,n,$$

where the manifest observations are the $(x_i, y_i)$ pairs, and where the error random variables, the $u_i$'s, are independent, identically distributed according to (1) with $\mu = 0$. The parameters of the model [2] are $a$, $b$, $\sigma$, and $\theta$.

Given the observations, the likelihood of the parameters is found by multiplying together the likelihoods for the individual error random variables. The likelihood of the parameters given the entire sample is

$$L(a,b,\sigma,\theta) = \prod_{i=1}^{n} f(y_i|a+bx_i,\sigma,\theta) \tag{2}$$

$$= k^n(\sigma,\theta)\exp\left\{-\frac{1}{\sigma^\theta}S\right\}, \tag{3a}$$

where

$$S = \sum_{i=1}^{n} |y_i - a - bx_i|^\theta; \tag{3b}$$

and where the vectors of observations in the likelihood function are suppressed for convenience. For this general family of distributions, $\sigma$ has no effect on the maximum likelihood parameters of the regression line.[3] Throughout this discussion we will look at the generalized likelihood function of $a$, $b$, $\theta$ defined by

$$L^*(a,b,\theta) = \max_{\sigma} L(a,b,\theta,\sigma).$$

It may be seen from the form of (3) that with $\theta = 2$ the maximum likelihood estimates of $a$ and $b$ are the least squares estimates, and that with $\theta = 1$ the maximum likelihood estimates of $a$ and $b$ are those that minimize the sum of absolute deviations.

### *Least Squares Regression*

The least squares technique has traditionally been justified by two assumptive arguments: (1) that it provides the maximal likelihood regression coefficients; and (2) that of all unbiased linear estimators, least squares has minimal variance about the regression line. Both of these properties have at times been adduced to call least squares the "best" of regression techniques. Because least squares possesses in addition the attribute of computational facility, this method long has reigned as the foremost tool of the social scientist in reducing data to mathematically descriptive relationships.

The first argument above assumes a normal distribution of the error terms. We argue that

* Professor Howard Raiffa generously gave us his expert advice on this paper. A referee made suggestions which helped us clarify our argument. The research was supported by the Society of Fellows, the Harvard Institute of Economic Research, and the Public Policy Program of the Kennedy School of Government.

[1] This may be seen from the derivation,

$$\int_0^\infty e^{-x^\theta}dx = (1/\theta)\int_0^\infty e^{-t}t^{1/\theta-1}dt = \frac{1}{\theta}\Gamma(1/\theta) = \Gamma(1+1/\theta).$$

[2] For ease of computation and exposition we discuss only two-variable regressions. Multivariate regressions introduce no new conceptual problems.

[3] The maximum likelihood $\sigma$ is given by $(\theta S/n)^{1/\theta}$.

this supposition is often unwarranted and shall show that significant gains in likelihood may be achieved when the regression technique allows for the more general class of error distribution defined in (1).

Various conditions may render inoperative the Central Limit Theorem and thus vitiate the a priori hypothesis of a normal error distribution. The number of errors making significant contributions to the error term is quite often small and these factors treated as random variables may be dependent. Unless these few factors are themselves normally distributed, they will not aggregate to produce a normal distribution of the overall error term. Furthermore, the presence of non-random aspects of human error, as well as the fact that few studied relationships are truly linear, will further weaken any argument that errors are normally distributed.

The second assumptive argument, the well-known Gauss-Markov result, makes the implicit assertion that variance is a valuable measure in gauging the efficiency of the fit of a regression line. We argue that the value and significance of this measure diminish as the underlying error distribution diverges from the normal. Note, for example, that for error distributions of the class in the preceding section defined by $\theta \leq 1$, the variance of the errors about the true regression line is infinite.[4] This fact means that in many cases of interest any variance measure will be utterly impractical and misleading as a measure of the efficiency of a regression.

Whenever the precise best fit of a regression line is sought and the non-normality of the error terms cannot justify the use of least squares, we assert that it is desirable that the regression technique allow for the general class of error distributions considered in this paper. We support this contention below with examples drawn from two classic papers. There, we compare coefficient estimates and likelihood values derived using least squares and using the more general regression technique proposed here. We will find that the estimated coefficients will differ for the two methods, and that the advantage in terms of likelihood of the proposed method is significant.

Least squares has one unassailable advantage, its simplicity. The computational process used to prepare this paper takes several times longer than least squares, but it still turned out to consume little computer time for the kinds of examples considered here. For regressions involving a great number of coefficients, computer cost might be more of a consideration. Even for these cases we believe that the advent of rapid new computing machinery together with experience with relevant algorithms might soon make it worthwhile to employ the regression procedures discussed here whenever the assumptions motivating the least squares procedure are not a reasonable specification.

*Maximum Likelihood Estimates for $a$, $b$, and $\theta$ for Real Data*

In what follows we will apply maximum likelihood methods to some well-known sets of data in order to compare results derived with least squares with those that are achieved when in addition to $a$ and $b$, $\theta$ also is estimated.

With the aid of these empirical examples, we were able to discern the general shapes of likelihood functions for different values of $\theta$. Imagine $a$ and $b$ on the axes of the horizontal plane, with the likelihood of the regression parameters conditional upon the sample given by the height above the plane. In general, the likelihood surface will have a ridge running obliquely across the plane. The highest point on this ridge will correspond to the $(a, b)$ combination that gives the maximum likelihood regression line for the particular value of $\theta$.

The shape of the likelihood surface was investigated analytically and with the aid of an on-line computer. For $\theta = 1$, the likelihood surface is composed of many planar segments. For $\theta > 1$, the ridge is round-topped, and the slope near the maximum is rather gentle. The first derivatives of the likelihood with respect to $a$ and $b$ are necessarily continuous, and the sections of the likelihood surface parallel to the axes are unimodal. With $\theta < 1$, the sections of the likelihood surface parallel to the axes are no longer unimodal; the ridge top is sharply peaked and non-differentiable. There are in addition several minor ridges, each correspond-

[4] The unboundedness of a variable is implicit in the consideration of any distribution, such as the normal, which has infinite tails. Distributions which also have infinite variances imply nothing further in this regard.

ing to a set of coordinates $(a, b)$ that takes the regression line through a data point. The second derivatives in the neighborhood of a maximum become positive, thereby, incidentally, rendering all quadratic optimization techniques infeasible.

The maximum likelihood values of $a$ and $b$ will depend significantly on the $\theta$ for which they are estimated. Thus, even if we are primarily interested in the likelihood function for $a$ and $b$, our results will be strongly influenced by any assumptions about the value of $\theta$.

We can best illustrate with an empirical example. The paper by Arrow et al. [1] on the estimation of CES production functions gives cross-nation data for the textiles spinning and weaving industry.[5] The regression was of the logarithm of value added per man year on the logarithm of the wage rate, each measured in dollars.

Table 1 gives the maximum likelihood estimates of $a$ and $b$ for each of four values of $\theta$.

TABLE 1. — TEXTILES SPINNING AND WEAVING MAXIMUM LIKELIHOOD PARAMETERS OF REGRESSION LINE FOR DIFFERENT VALUES OF $\theta$

| | Value of $\theta$ | | | |
|---|---|---|---|---|
| | 0.5 | 1.0 | 2.0 | 3.0 |
| $a$ | 2.543 | 2.119 | 2.023 | 1.733 |
| $b$ | 0.73 | 0.79 | 0.81 | 0.85 |

Quite obviously, if the true value of $\theta$ were 0.5 it would not be very satisfactory to use maximum likelihood estimates that were derived on the assumption that $\theta$ equals 2. One way around this difficulty is to derive maximum likelihood estimates with $\theta$ as well as $a$ and $b$ being estimated. That is, with $\theta$ set equal to $\hat{\theta}$, its maximum likelihood estimate.

In each of four examples we found the maximum likelihood estimates for $a$ and $b$ with $\theta$ specified equal to 2. These were the least squares estimates. These estimates were compared with the maximum likelihood estimates for the regression parameters when $\theta$ was allowed to take its maximum likelihood value, $\hat{\theta}$. The examples were:

I) The cross-section regression to estimate the textiles spinning and weaving production function.

II) A cross-section regression of 1965 United States foreign aid to Near East and South Asian countries on their populations.

III) A time-series regression of consumption on the supply of money for the years 1897–1958. The data and original regression appear in a classic article by Milton Friedman [5].

IV) A time-series regression of national income on money supply for the 1897–1958 period that also is taken from Friedman's article.[6]

Table 2 shows the results of these calculations.

The values of the maximum likelihood regression parameters vary greatly when $\theta$ is set equal to $\hat{\theta}$ rather than assuming that $\theta = 2$. The intercepts were changed from 15.0 to 57.1 per cent. In the four examples, the slopes of the regression lines changed less on a percentage basis than did the intercepts; they changed from 1.8 to 9.3 per cent. In the second example, the positive slope and intercept both decreased.[7]

Unfortunately, without much further statistical work, there will be no easy formulas which give the distributions of the maximum likelihood estimators of the regression line when $\theta$ is allowed to take on values different from 2. This makes it impossible to employ the usual classical tests for the significance of regression coefficients. It would of course be possible to employ the conditional sampling distributions of the estimators (given the population parameters) to derive reference tables of boundary values that indicate different levels of significance. Such tables are not presented in this paper. An alternate procedure, more in the spirit of Bayesian investigations, is available. It compares aggregate likelihoods with and without the inclusion of a particular regression coefficient. The framework of the likelihood

[5] This was the first mentioned of the two industries for which they presented the greatest number of observations.

[6] The data for the second example is taken from *The World Almanac* [12]. The units are aid in millions of dollars and population in millions. The units for the third and fourth examples are millions of current dollars.

[7] The calculations of percentage changes were made before rounding. Percentage changes in the intercept will depend upon units of measurement if logs of variables enter the regression. The change in $a$ in example I therefore has little meaning.

TABLE 2. — MAXIMUM LIKELIHOOD ESTIMATES OF THE REGRESSION PARAMETERS $a$ AND $b$ FOR FOUR EXAMPLES WHEN $\theta = 2$ AND WHEN $\theta = \hat{\theta}$, ITS MLE VALUE

| Specification | Examples | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | I | | II | | III | | IV | |
| | $a$ | $b$ | $a$ | $b$ | $a$ | $b$ | $a$ | $b$ |
| With $\theta = 2$, least squares | 2.023 | .81 | 23.24 | 1.81 | 7813 | 1.31 | 9075 | 1.47 |
| With $\theta = \hat{\theta}$, the MLE value | 2.574 | .73 | 9.98 | 1.78 | 6615 | 1.34 | 7715 | 1.50 |
| The MLE value, $\hat{\theta}$ | .45 | | .60 | | .50 | | .675 | |

comparison procedure is set forth in the material that follows.

*Likelihood Gains and Their Significance*

Table 3 shows the gains in likelihood that were achieved by estimating rather than assuming a value for $\theta$.

*Maximum Likelihood Values of $\theta$*

It is somewhat startling to note that the maximum likelihood $\theta$ in each of our four examples was significantly less than one. This means that in each case the distribution of the error term that gives the greatest likelihood is one with more density in the tails in comparison to the central hump than is consistent with any normal distribution. For such a distribution, to estimate the regression parameters using least squares is highly inefficient.

TABLE 3. — MAXIMUM VALUES OF THE LIKELIHOOD FUNCTION FOR FOUR EXAMPLES WHEN $\theta=2$ AND WHEN $\theta=\hat{\theta}$, ITS MLE VALUE

| | Examples | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| Log $L^*$ with $\theta = 2$, least squares | 2.137 | $-69.06$ | $-676.9$ | $-678.1$ |
| Log $L^*$ with $\theta = \hat{\theta}$, the MLE value | 5.133 | $-66.94$ | $-657.0$ | $-669.5$ |
| Ratio $\dfrac{L^* \text{ with } \theta = \hat{\theta}}{L^* \text{ with } \theta = 2}$ | 19.298 | 8.331 | $4.498 \times 10^8$ | 5431 |
| Number of observations = $n$ | 18 | 13 | 62 | 62 |
| The $n^{th}$ root of ratio | 1.181 | 1.177 | 1.378 | 1.149 |

Table 3 shows that the gain in likelihood per data point ranges from 14.9 per cent to 37.8 per cent. This appears to be a significant gain, but it is difficult to get a feeling for what is a large and what is a small gain when an additional parameter is estimated.

To obtain a basis for comparison, we calculated the likelihood losses incurred when values of $a$ and $b$ are pre-specified in the textiles spinning and weaving example. We assumed $\theta$ equal to 2, pre-specified the value of $a$ and $b$ away from its maximum likelihood value, and then calculated the maximum likelihood allowing the other regression parameter and $\sigma$ to vary. Table 4 gives the results.

The last row of table 4 shows that the gains in likelihood per data point derived by freeing $a$ or $b$ from fairly restrictive pre-specifications were relatively small, ranging from 0.4 to 7.4 per cent.

Least-squares procedures weight large errors relatively much more heavily than does the maximum likelihood regression procedure for a $\theta$ significantly less than 2. If the true $\theta$ is equal to, let us say, 0.7, but we use least squares estimation techniques, we will be giving much too much weight to eccentric points. This will exert a randomizing effect on the estimated regression line. If the true $\theta$ is significantly greater than 2, the converse will be true. A random factor will be introduced because least squares underweights eccentric points. By failing to use the true, or approximately true, value of $\theta$ when $\theta$ is far from 2, least squares incurs a significant loss of efficiency.

TABLE 4. — TEXTILES SPINNING AND WEAVING.
THE EFFECT OF PRESETTING THE VALUE OF ONE REGRESSION PARAMETER ON THE MAXIMUM VALUE OF THE LIKELIHOOD FUNCTION AND THE MAXIMUM LIKELIHOOD ESTIMATE OF THE OTHER PARAMETER

| | Pre-Specification | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | None | $a=2.5$ | $a=1.5$ | $b=.7$ | $b=.78$ | $b=.083$ | $b=0.9$ |
| $a$ | 2.023 | 2.5 | 1.5 | 2.723 | 2.186 | 1.864 | 1.396 |
| $b$ | 0.805 | 0.734 | 0.884 | 0.7 | 0.78 | 0.83 | 0.9 |
| Log $L^*$ | 2.137 | 1.510 | 1.388 | 0.835 | 2.055 | 2.059 | 1.079 |
| Ratio of $L^*$ without restrictions to $L^*$ with a parameter pre-specified | | 1.871 | 2.115 | 3.743 | 1.085 | 1.081 | 2.881 |
| The $n^{\text{th}}$ root of ratio ($n$ = 18) | | 1.035 | 1.042 | 1.074 | 1.005 | 1.004 | 1.061 |

*Effect of Estimating $\theta$*

It might be objected that the likelihood improvements achieved using our procedures were primarily attributable to the loss of one degree of freedom when $\theta$ was made variable. For example, even if the distribution of the error term were normal, we would expect random aspects of small samples to lead to fluctuating values for the maximum likelihood estimate of $\theta$.

To test this conjecture we ran four simulations with a pre-determined regression equation and error terms drawn from a table of random normal deviates. There were forty data points in each simulation. The values of the independent variable were assigned at unit intervals. For each of the simulations we calculated the maximum likelihood with $a$ and $b$ variable, but $\theta$ set equal to its known true value of 2. Then we calculated the maximum likelihood for each sample allowing $\theta$ to vary as well. The results are shown in table 5.

In these four simulations, we note with some surprise that the maximum likelihood $\theta$'s are far from two. What is much more startling is that the likelihood gains are so small given these sizable divergences. The gains from estimating $\theta$ rather than using its known true value ranged from 0.6 per cent to 2.8 per cent. This provides an interesting extension of a previous observation. For $\theta$ equal to 2, the likelihood function has a rather gentle slope in the neighborhood of the maximum not only for changes in $a$ and $b$, but for changes in $\theta$ as well.

TABLE 5. — LIKELIHOOD GAINS FROM SETTING $\theta$ EQUAL TO ITS MLE VALUE RATHER THAN ITS KNOWN VALUE
FOUR SIMULATIONS WITH NORMAL ERROR RANDOM VARIABLES, WITH TRUE $\theta=2$

| | | | | |
|---|---|---|---|---|
| Log $L^*$ with $\theta=2$, its true value | $-60.377$ | $-55.569$ | $-50.477$ | $-57.146$ |
| Log $L^*$ with $\theta=\hat{\theta}$, the MLE value | $-59.418$ | $-54.482$ | $-49.469$ | $-56.945$ |
| The MLE value, $\hat{\theta}$ | 3.25 | 1.25 | 3.928 | 2.825 |
| Ratio $\dfrac{L^* \text{ with } \theta=\hat{\theta}}{L^* \text{ with } \theta=2}$ | 2.609 | 2.965 | 2.740 | 1.260 |
| The $n^{\text{th}}$ root of ratio ($n=40$) | 1.024 | 1.028 | 1.026 | 1.006 |

This observation has a practical corollary. In assessing the applicability of least squares we thought first to use the optimal likelihood $\theta$ as a measure of normality. Although this in a rough way is feasible, it now appears preferable to use the likelihood gain per data point from estimating $\theta$ as the yardstick. Thus, while Example I is by either standard the least amenable to least squares, we conclude on grounds of the likelihood gains that Example II and not Example III — as one might have expected from its lower $\hat{\theta}$ — includes the next most non-normal data.

The likelihood gains from freeing $\theta$ in our simulations are quite small in comparison to those that are achieved by freeing $\theta$ in our real world examples. This is to be expected. Even though the maximum likelihood estimate for a parameter may be far from its true value, the likelihood gain to be derived by using the former rather than the latter should not be expected to be great. Given the magnitude of the likelihood gains that we achieved in our four empirical examples, we would conclude that the observed maximum likelihood values were not in fact chance occurrences with the true $\theta$ being close to 2.[8]

[8] To deal with cases that might be less clear-cut, we

### Summary and Conclusions

This paper presents a procedure for estimating linear regression lines when errors are distributed by one of a class of power distributions. The degree of peakedness of these distributions is indicated by the parameter $\theta$; it equals 2 in the case of the normal distribution.

The maximum likelihood parameters of a regression line can be significantly affected by the value of $\theta$ that defines the distribution with which they are estimated. If the true $\theta$ is far from 2, least squares regression will give inefficient estimates of $a$ and $b$.

The four empirical examples showed that there can be sizeable gains in likelihood if $\theta$ is estimated rather than pre-specified equal to 2.

devised a simple procedure to test whether a $\theta$ for a power distribution, either assumed or estimated from a sample, was consistent with the observed sample. The procedure computes the ratio of the width of the interval that contains the third of the observations with the smallest absolute values, to the width of the interval that contains the next third. The greater the true value of $\theta$, the smaller is the expected value of this ratio. As a complement to this test we ran some Monte Carlo studies to get the distribution of this ratio for the normal distribution. The results of these tests supported, but did not provide overwhelming evidence for, our conclusions about the values of $\theta$. The tests and results, too lengthy to detail here, are available on request.

We noted with some surprise that the maximum likelihood values for $\theta$ in the four examples were significantly less than one, that the maximum likelihood distributions of the error term were relatively flat with long tails. Future studies will investigate whether this is a general phenomenon for errors about a regression line. This will give us some insight into the question of what values of $\theta$ should be expected.

A logical extension of this analysis would be to adopt a fully Bayesian approach: (1) place a joint prior distribution on $\theta$, $a$, $b$, $\sigma$; (2) update this prior by multiplying it by the likelihood function; (3) integrate out the nuisance parameters $\sigma$ and $\theta$ to get a joint posterior on $a$ and $b$. This would obviate the need to have any procedure to test the null hypothesis that $\theta = 2$.

All of the evidence we uncovered leads us to the conclusion that if accurate estimation of a linear regression line is important, it will usually be desirable to estimate not only the coefficients of the regression line, but also the parameters of the power distribution that generated the errors about the regression line. The effect on the estimates of regression coefficients may not be small.

## APPENDIX
## The Computation of Maximum Likelihood Parameters

We start with $\theta = 2$ and find maximum likelihood estimates for $a$ and $b$ which are the least squares estimates. Next, using these estimates of $a$ and $b$ we find the $\sigma$ and $\theta$ that maximize the likelihood function. Then we find the maximum likelihood $a$ and $b$ for this particular $\theta$ value; $\sigma$ has no influence. This procedure is then iterated with some creative modification to be described. To find the maximizing values it is useful to know the derivatives of the likelihood function with respect to $a$, $b$, and $\theta$. They are,

$$\frac{\partial \log L}{\partial a} = \frac{\theta}{\sigma^\theta} \sum_{i=1}^{n} |y_i - a - bx_i|^{\theta-1} \cdot x_i \cdot sgn(y_i - a - bx_i),$$

$$\frac{\partial \log L}{\partial b} = \frac{\theta}{\sigma^\theta} \sum_{i=1}^{n} |y_i - a - bx_i|^{\theta-1} sgn(y_i - a - bx_i),$$

and

$$\frac{\partial \log L}{\partial \theta} = \frac{-D}{\sigma^\theta} + \frac{S \log \sigma}{\sigma^\theta} + \frac{n\Gamma'(1+1/\theta)}{\sigma^2 \Gamma(1+1/\theta)},$$

where

$$D = \sum_{i=1}^{n} |y_i - a - bx_i|^\theta \cdot \log |y_i - a - bx_i|$$

and $S$ is as indicated in (3b) in the text. The notation "$sgn$" denotes the sign function which gives the values of one for positive operands and the minus one for negative operands. The combined factor $\Gamma'(1+1/\theta)/\Gamma(1+1/\theta)$ from the third term of the derivative with respect to $\theta$ is the digamma function. If only a finite set of values of $\theta$ is used, the well-tabulated values of this factor may be held in memory. For continuously varying $\theta$, Euler's equation

$$\Gamma'(x)/\Gamma(x) = -c - \frac{1}{x} + \sum_{i=1}^{\infty} \left( \frac{1}{i} - \frac{1}{x+i} \right).$$

where $c$ is Euler's constant, can serve as the basis of a simple subroutine to determine the value of this factor.

In the conventional method of computing, the search follows the gradient in the $ab$-plane to the likelihood maximum. The major problem is determination of step size. Because of the erratic behavior of the second derivative with $\theta < 1$, the step determination formula for quadratic maxima could not be used. The most efficient method found involved presetting the step length and altering it as the subsequent searches showed the best length to be either longer or shorter. $\theta$ was periodically checked and varied to keep the general likelihood-maximizing $\theta$ for the new combination of $a$, $b$. Because of the nature of the likelihood topology of a curving ridge, the optimization procedure of Rosenbrock, designed for such terrains, was also found efficient when modified slightly.

Problems of this type are also uniquely suited to the on-line method of programming. With the programmer at the console, errors are found immediately and the programmer acquires a much better grasp of the nature of his object function than can be obtained from printed output. This intuition acts along a feedforward principle in the search for a maximum and achieves efficiency that cannot be preprogrammed. For computation ease at the console, the program was modified so that for a given $b$ and $\theta$ the likelihood-maximizing $a$ was found automatically. The program operator could consider his problem one of two variable optimization.

Copies of the program are available on request.

## REFERENCES

[1] Arrow, K., H. B. Chenery, B. Minhas, and R. Solow, "Substitution and Economic Efficiency," this REVIEW, XLII (Aug. 1961), 225–250.

[2] Box, G. E. P., and G. C. Tiao, "A Further Look at Robustness via Bayes's Theorem," *Biometrika*, 49 (1962), 419–432.

[3] ———, "A Bayesian Approach to the Importance of Assumptions Applied to the Comparison of Variances," *Biometrika*, 51 (1964), 153–167.

[4] Dianada, P. H., "Note on Some Properties of Maximum Likelihood Estimates," *Proceedings, Cambridge Philosophical Society* 45 (1949), 536–544.

[5] Friedman, M., "The Lag in the Effect of Monetary Policy," *Journal of Political Economy*, 60 (1961), 447–466.

[6] Johnston, J., *Econometric Methods*. (New York: McGraw-Hill, 1960.)

[7] Mandelbrot, B., "New Methods in Statistical Economics," *Journal of Political Economy*, 71 (1963), 421–440.

[8] RAND Corporation, *A Million Random Digits and 100,000 Normal Deviates*. (Glencoe, Ill.: Free Press, 1955.)

[9] Rosenbrock, H. H., and C. Story, *Computational Techniques for Chemical Engineers*. (London: Oxford, 1966.)

[10] Turner, M. C., "On Heuristic Estimation Methods," *Biometrics*, 16 (1960), 299–301.

[11] Wilson, R. B., "Two Notes on Estimating Shape Parameters." RAND Corporation Memorandum, RM-4459-PR.

[12] *The World Almanac*. (New York: Newspaper Enterprise Association, 1967.)