

Division of Research
Graduate School of Business Administration
The University of Michigan

August 1975

AN ANALYSIS OF TRANSFORMATIONS
FOR TIME SERIES

Working Paper No. 117

by

C.F. Ansley, W.A. Spivey, W.J. Wroblewski
The University of Michigan

© The University of Michigan, 1975

FOR DISCUSSION PURPOSES ONLY

None of this material is to be quoted or
reproduced without the express permission
of the Division of Research.

Introduction

To apply the Box-Jenkins methodology to the analysis and forecasting of nonstationary time series, a preliminary transformation of the data is oftentimes required. Evidence suggests that the choice of this preliminary transformation is of critical importance. Chatfield and Prothero [4], for example, were unable to obtain a satisfactory forecasting model using Box-Jenkins methods on their transformed data, and several discussants of this paper, [2], [6], and [8], suggested that this result was attributable to an improper choice of initial transformation. It was also indicated by discussants that a more flexible parametric family of transformations, first suggested by Box and Cox [1] for the fixed effects analysis of variance model, could be used to select an initial transformation. This family is given by

$$(1) \quad y_{\lambda}(t) = \frac{(y(t))^{\lambda} - 1}{\lambda} \quad \lambda \neq 0$$
$$y_0(t) = \ln y(t).$$

where for our purposes $y(t)$, assumed to be positive, denotes a nonstationary time series and λ is the transformation parameter.

Estimates of λ for the time series of Chatfield and Prothero were presented in [2], [5], and [8]. Apparently these estimates were developed by extending to time series

models the likelihood methods associated with the Box-Cox transformation (1). To the best of our knowledge, however, no formal development of such an extension has been published. This paper presents such a development, together with a numerical algorithm for estimating λ , and illustrates the use of the algorithm in the analysis of several time series.

Development of the Likelihood Function

We have called (1) a preliminary transformation because one may wish to employ it before entering the identification phase of the Box-Jenkins methodology. Moreover, the reduction of a nonstationary time series to a stationary ARMA process by successive backward differencing is a central feature of the Box-Jenkins methodology. Thus to develop the likelihood function for the observed (transformed) time series data we must consider the effects of two systems of transformations, the preliminary system and the differencing system.

Suppose we have observations from an underlying time series $y(t)$ and that for some value λ of the transformation parameter the transformed observations $y_\lambda(t)$ in (1) follow an ARIMA(p,d,q) process having autoregressive parameters $\underline{\phi} = (\phi_1, \dots, \phi_p)$ and moving average parameters $\underline{\theta} = (\theta_1, \dots, \theta_q)$. Suppose further that the random shocks $a(t)$ associated with the process are normally distributed, independent random variables with mean zero and variance σ^2 . The constant

term in the model is denoted by δ .

Since the $y_\lambda(t)$ form an ARIMA(p,d,q) process, the d^{th} order backward differences,

$$(2) \quad w_\lambda(t) = \nabla^d y_\lambda(t)$$

form an ARMA(p,q) process. The joint probability density of n such differences $w_\lambda(1), \dots, w_\lambda(n)$ is given by Box and Jenkins [3; p. 273]:

$$(3) \quad f_n(\underline{w}_\lambda | \underline{\phi}, \underline{\theta}, \delta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} |M_n|^{1/2} \exp\left\{-\frac{S}{2\sigma^2}\right\}$$

where

$$(4) \quad \underline{w}_\lambda = (w_\lambda(n), \dots, w_\lambda(1)),$$

$$(5) \quad S = S(\underline{\phi}, \underline{\theta}, \delta) = \sum_{t=-\infty}^n \{E[a(t) | \underline{w}_\lambda; \underline{\phi}, \underline{\theta}, \delta]\}^2,$$

and where $M_n = M_n(\underline{\phi}, \underline{\theta})$ is the inverse of the n by n correlation matrix of the differenced series $w_\lambda(t)$.

Because it is impossible to reconstruct the joint probability density of the n+d transformed observations using only the joint probability density of the n differences $w_\lambda(t)$, we will assume that the initial d observations $y_\lambda(0), y_\lambda(-1), \dots, y_\lambda(-d+1)$ are fixed. We now seek the joint probability density of the n most recent observations $y_\lambda(n), \dots, y_\lambda(1)$ of the time series, regarding the initial d values as constants. Accordingly we substitute (2) for $w_\lambda(t)$ in (3) and multiply the resulting

function by the Jacobian of the transformation for the d^{th} backward differences. This Jacobian is equal to 1 because it is the determinant of a triangular matrix whose principal diagonal elements are each equal to 1. Thus the joint density of the $y_\lambda(t)$, regarding the initial d values $y(0), y(-1), \dots, y(-d+1)$ as fixed, is

$$(6) \quad g_n(\underline{y}_\lambda | y_\lambda(0), \dots, y_\lambda(-d+1); \underline{\phi}, \underline{\theta}, \delta, \sigma^2) \\ = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} |M_n|^{1/2} \exp\left\{-\frac{S}{2\sigma^2}\right\}$$

where

$$(7) \quad \underline{y}_\lambda = (y_\lambda(n), \dots, y_\lambda(1)),$$

$$(8) \quad S = S(\underline{\phi}, \underline{\theta}, \delta) = \sum_{t=-\infty}^n \{E[a(t) | \nabla^d y_\lambda(n), \dots, \nabla^d y_\lambda(1); \underline{\phi}, \underline{\theta}, \delta]\}^2.$$

Finally, we can obtain the joint density of the original (untransformed) time series $y(1), \dots, y(n)$, assuming $y(0), y(-1), \dots, y(-d+1)$ to be fixed, by substituting for $y_\lambda(t)$ in (6) and multiplying the result by the Jacobian of the transformation (1),

$$(9) \quad h_n(\underline{y} | y(0), \dots, y(-d+1); \underline{\phi}, \underline{\theta}, \delta, \sigma^2, \lambda) \\ = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} |M_n|^{1/2} \exp\left\{-\frac{S}{2\sigma^2}\right\} J$$

where

$$(10) \quad \underline{y} = (y(n), \dots, y(1))$$

$$(11) \quad S = S(\underline{\phi}, \underline{\theta}, \delta, \lambda) = \sum_{t=-\infty}^n \{E[a(t) | \sqrt[n]{d} Y_{\lambda}(n), \dots, \sqrt[n]{d} Y_{\lambda}(1); \underline{\phi}, \underline{\theta}, \delta, \lambda]\}^2$$

from (8), where S is now regarded as a function of λ as well as of the other parameters, and where

$$(12) \quad J = J_n(\lambda, \underline{y}) = \prod_{t=1}^n \left| \frac{\partial y_{\lambda}(t)}{\partial Y(t)} \right| = \prod_{i=1}^n \{y(t)\}^{\lambda-1}.$$

Interpreting (9) as a likelihood and taking the logarithm gives

$$(13) \quad L = \text{const} - \frac{n}{2} \ln \sigma^2 + \frac{1}{2} \ln |M_n| - \frac{S}{2\sigma^2} + \ln J.$$

Maximum likelihood estimates of $\underline{\phi}$, $\underline{\theta}$, δ , σ^2 and λ can now be obtained by solving the corresponding simultaneous system of nonlinear normal equations.

Numerical Algorithms for Parameter Estimation

Following Box and Jenkins [3; p. 213] we will assume that for moderate and large values of n , the term $\frac{1}{2} \ln |M_n|$ in (13) is dominated by $S/2\sigma^2$. Discarding this term we approximate (13) by

$$(14) \quad L^* = \text{const} - \frac{n}{2} \ln \sigma^2 - \frac{S}{2\sigma^2} + \ln J,$$

and note that the solution to the normal equation obtained from (14) for σ^2 is

$$(15) \quad \hat{\sigma}^2 = \frac{S}{n}.$$

For fixed values of the remaining parameters we can maximize L^* over σ^2 by substituting (15) into (14), obtaining

$$(16) \quad L_{\max}^* = L_{\max}^*(\underline{\phi}, \underline{\theta}, \delta, \lambda) = \text{const} - \frac{n}{2} \ln\left(\frac{S}{n}\right) + \ln J.$$

Maximization of L_{\max}^* remains complicated. It is possible, however, to reexpress (16) as a monotonic function of a sum of squares, enabling the use of a standard, nonlinear, least-squares algorithm without major revision of existing computer programs. The first step in this process is to examine the sum of squares function S in (14). From (11) we have

$$(17) \quad S = \sum_{t=-\infty}^n [a(t)]^2.$$

We simplify the notation here and below by abbreviating $E[a(t) | \underline{w}_\lambda; \underline{\phi}, \underline{\theta}, \delta, \lambda]$ to $[a(t)]$. Numerical values for $[a(t)]$ are obtained from the difference equations

$$(18) \quad \phi(B) [w_\lambda(t)] = \delta + \theta(B) [a(t)]$$

where

$$(19) \quad [w_\lambda(t)] = E[w_\lambda(t) | \underline{w}_\lambda; \underline{\phi}, \underline{\theta}, \delta, \lambda],$$

$$(20) \quad \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p,$$

$$(21) \quad \theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q,$$

and where B is the backshift operator defined by Box and Jenkins [3; p. 8].

Continuing with the problem of reducing (16) to a monotonic function of a sum of squares, we follow Box and Cox and define an additional transformation

$$(22) \quad z_{\lambda}(t) = \frac{w_{\lambda}(t)}{J^{1/n}} \quad t = 1, \dots, n;$$

outside the range $1, \dots, n$ we write

$$(23) \quad z_{\lambda}(t) = \frac{[w_{\lambda}(t)]}{J^{1/n}} .$$

Because $[w_{\lambda}(t)] = w_{\lambda}(t)$ for $t = 1, \dots, n$, (23) is a simple extension of (22).

If we now write

$$(24) \quad \hat{a}_z(t) = \frac{[a(t)]}{J^{1/n}} ,$$

we can divide (18) by $J^{1/n}$ and obtain

$$(25) \quad \phi(B)z_{\lambda}(t) = \delta_z + \theta(B)\hat{a}_z(t)$$

where

$$(26) \quad \delta_z = \frac{\delta}{J^{1/n}} .$$

Given $z_{\lambda}(1), \dots, z_{\lambda}(n)$, it is clear that values of $\hat{a}_z(t)$ can be obtained from (25) in exactly the same way that values of $[a(t)]$ are obtained from (18).

Using (24), the sum of squares S in (17) can be written

$$(27) \quad S = \sum_{t=-\infty}^n [a(t)]^2 = J^{2/n} \sum_{t=-\infty}^n (\hat{a}(t))^2 = J^{2/n} S_z$$

where

$$(28) \quad S_z = \sum_{t=-\infty}^n (\hat{a}_z(t))^2.$$

Finally, we can substitute (27) into (16) and obtain

$$(29) \quad L_{\max}^* = \text{const} - \frac{n}{2} \ln \left(J^{2/n} \frac{S_z}{n} \right) + \ln J \\ = \text{const} - \frac{n}{2} \ln \left(\frac{S_z}{n} \right).$$

The problem of maximizing (14) is now reduced to that of minimizing S_z in (29) with respect to $\underline{\phi}$, $\underline{\theta}$, δ_z , and λ .

From the estimates $\hat{\delta}_z$ and $\hat{\lambda}$ we then obtain

$$\hat{\delta} = J^{1/n}(\hat{\lambda}) \hat{\delta}_z.$$

We have found S_z to be well behaved and approximately quadratic in λ for a variety of time series, suggesting that standard nonlinear least squares algorithms may be used successfully to minimize S_z . In fact, we have adapted a least squares version of the Marquardt algorithm [7] to minimize S_z simultaneously over all the parameters including λ , and this has proven to be efficient over a wide range of ARIMA models.

Examples of Likelihood Functions and Parameter Estimation

We demonstrate the approach described above by examining the likelihood surfaces associated with four, simulated time series, denoted A, B, C, and D. Each of these is an ARIMA time series transformed by the inverse of (1) using a different value of λ . The four underlying models were chosen to reflect a variety of correlation structures commonly found in practice. Each series contains 100 observations and in each case the initial value $y_\lambda(1)$ was arbitrarily chosen to be 1000. The simulation process is described in detail for Series A only.

Series A was generated from an ARIMA (1,1,0) model with $\phi = 0.5$, $\delta = 2.5$, and $\sigma^2 = 1$, transformed by the inverse of (1) for $\lambda = 0.7$. The steps in the process were as follows:

- (i) A stream of independent $N(0,1)$ variates $a(t)$ was generated and from it an initial value, $w_\lambda(1)$ was determined.
- (ii) Subsequent values of the ARMA (1,0) series $w_\lambda(t)$ were then obtained from the equation
$$w_\lambda(t) = 2.5 + 0.5w_\lambda(t-1) + a(t).$$
- (iii) Using a starting value $y_\lambda(1) = y_{.7}(1) = 1000$, values $y_{.7}(2), \dots, y_{.7}(100)$ were determined using $y_{.7}(t) = y_{.7}(t-1) + w_\lambda(t)$.
- (iv) Values of $y(1), \dots, y(100)$ were obtained from the inversion of (1) for $\lambda = 0.7$ and
$$y(t) = (0.7y_{.7}(t) + 1)^{1/.7}.$$

Using the approximate maximum likelihood procedure described above, estimates were obtained as follows:

$$\begin{aligned}\hat{\phi} &= 0.345, \\ \hat{\delta} &= 4.19, \\ \hat{\lambda} &= 0.724, \\ \hat{\sigma}^2 &= 1.56.\end{aligned}$$

For each pair of (fixed) ϕ and λ the likelihood L^* in (14) was maximized over δ and σ^2 . The nature of the resulting likelihood is indicated by the contours in Figure 1. It should be noted that the differences between values of L^* on adjacent contours are not equal in Figure 1, nor in the other contour maps shown below.

It is clear that L^* is well behaved and unimodal in this case. Because we work with S_z in (28) rather than L^* in our computing algorithm, it is appropriate to examine the variation of S_z with respect to ϕ and λ . In Figure 2 S_z is minimized over the remaining parameters.

The similarity of Figure 1 and Figure 2 supports the use of (29) in obtaining maximum likelihood estimates. The approximately quadratic nature of L^* with respect to λ can be more easily seen by plotting L_{\max}^* (maximized over δ , σ^2 and ϕ) against λ as in Figure 3.

The underlying model for Series B was ARIMA(2,1,0) with $\phi_1 = 0.75$, $\phi_2 = -0.5$, $\delta = 3.75$, and $\sigma^2 = 1$. Series B was then obtained by an inversion of (1) for $\lambda = 0.6$. Maximum likelihood estimates were

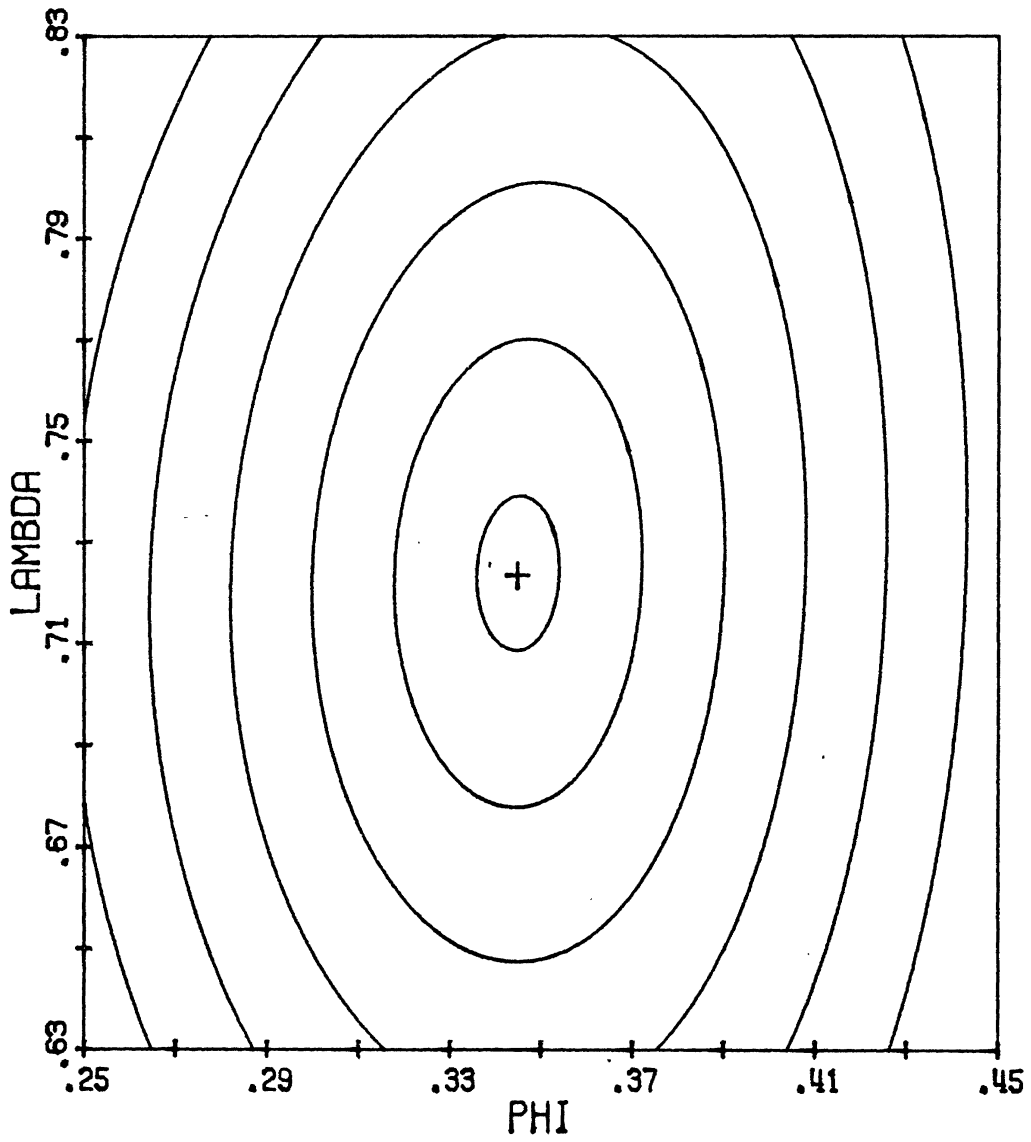


Fig. 1. Log Likelihood Contours,

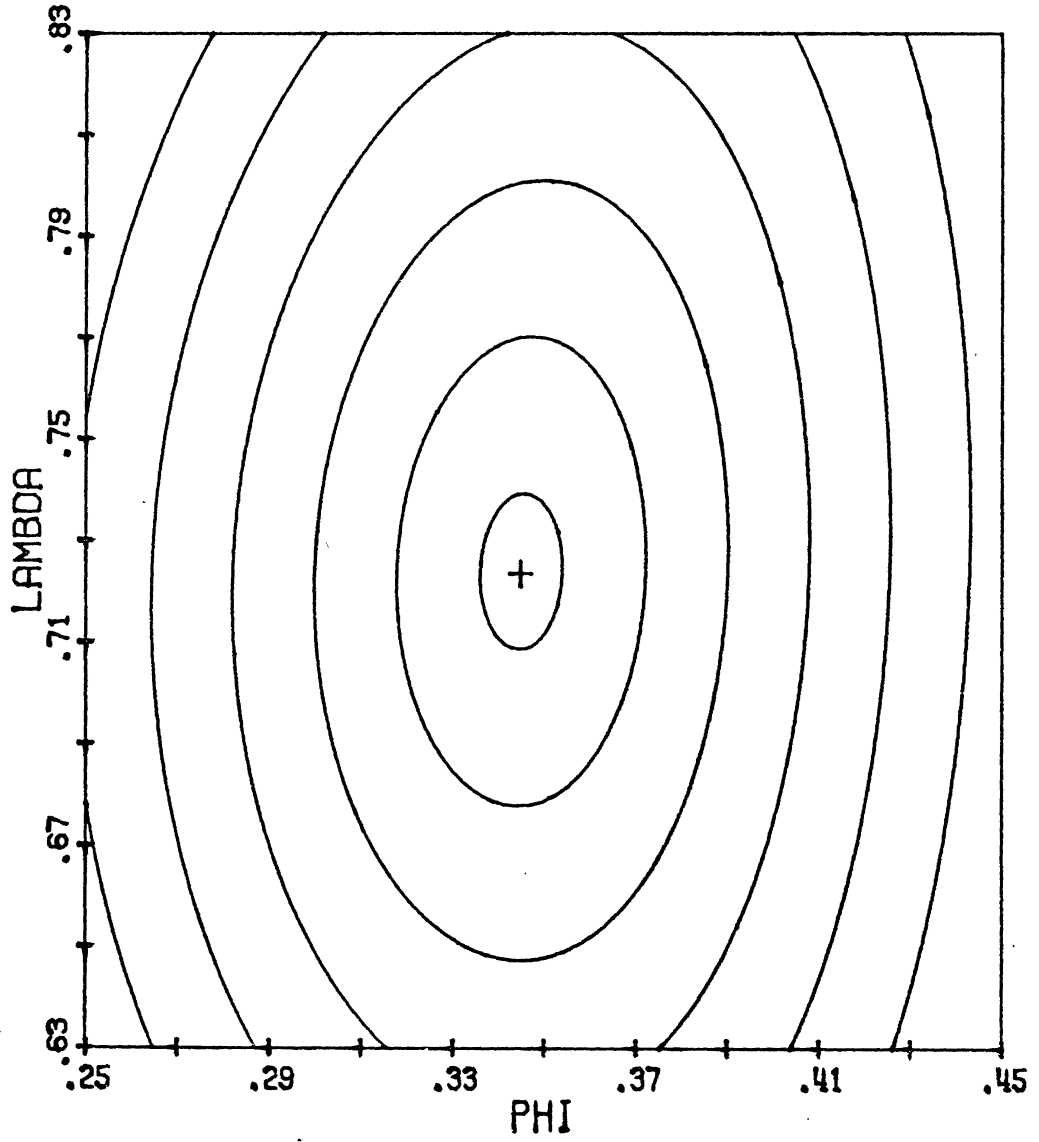


Fig. 2. Sums of Squares of Contours.

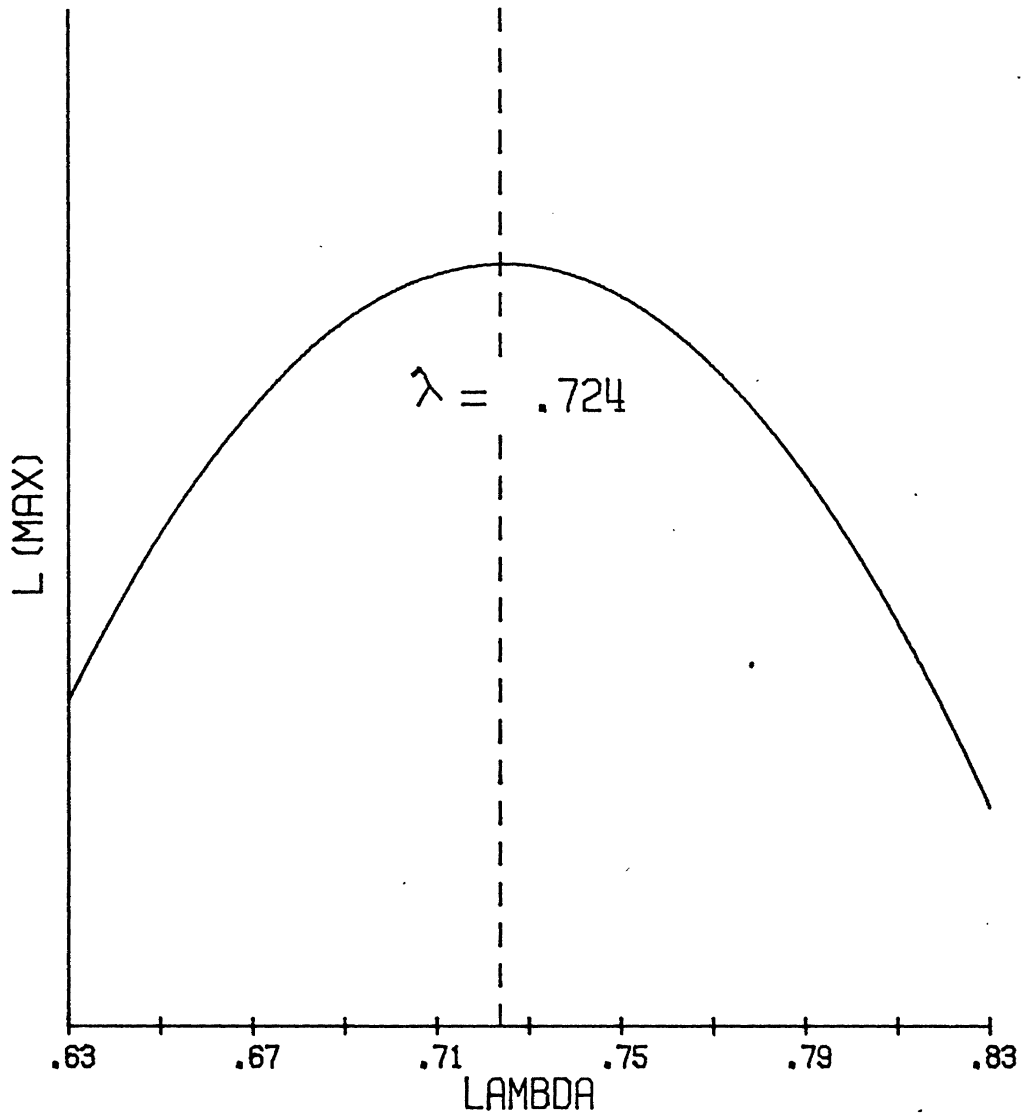


Fig. 3. Log Likelihood Series A.

$$\hat{\phi}_1 = 0.665,$$

$$\hat{\phi}_2 = -0.537,$$

$$\hat{\delta} = 3.67,$$

$$\hat{\sigma}^2 = 0.428,$$

$$\hat{\lambda} = 0.585.$$

The regularity of L^* is apparent from the surface contours for (ϕ_1, λ) and (ϕ_2, λ) in Figures 4 and 5. In each case, L^* is maximized over the remaining parameters.

Using $\lambda = 0.5$, Series C was obtained by inverting (1) for an ARIMA(1,1,1) process with $\phi = 0.8$, $\theta = -0.5$, $\delta = 1.0$, and $\sigma^2 = 1$. Maximum likelihood estimates obtained were

$$\hat{\phi} = 0.752,$$

$$\hat{\theta} = -0.383,$$

$$\hat{\delta} = 0.195,$$

$$\hat{\sigma}^2 = 0.0235,$$

$$\hat{\lambda} = 0.362.$$

Figures 6 and 7 show the nature of L^* (maximized over the remaining parameters) with respect to (ϕ, λ) and (θ, λ) .

Series D was generated from an ARIMA(0,1,1) process by inverting (1) for $\lambda = 0.4$ with $\theta = -0.4$, $\delta = 5.0$, and $\sigma^2 = 1$. Maximum likelihood estimates were

$$\hat{\theta} = -0.444,$$

$$\hat{\delta} = 3.04,$$

$$\hat{\sigma}^2 = 0.388,$$

$$\hat{\lambda} = 0.366.$$

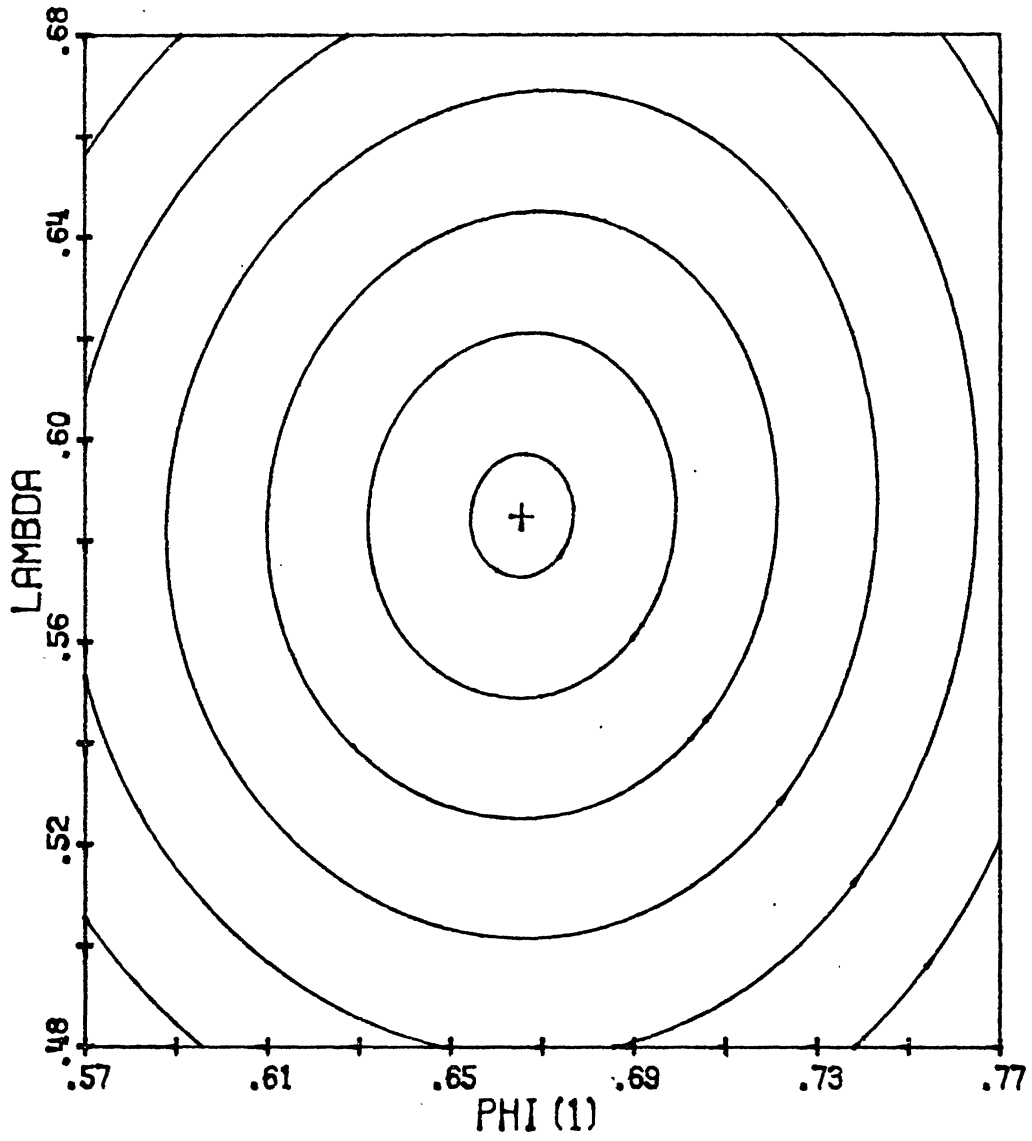


Fig. 4. Log Likelihood Contours,

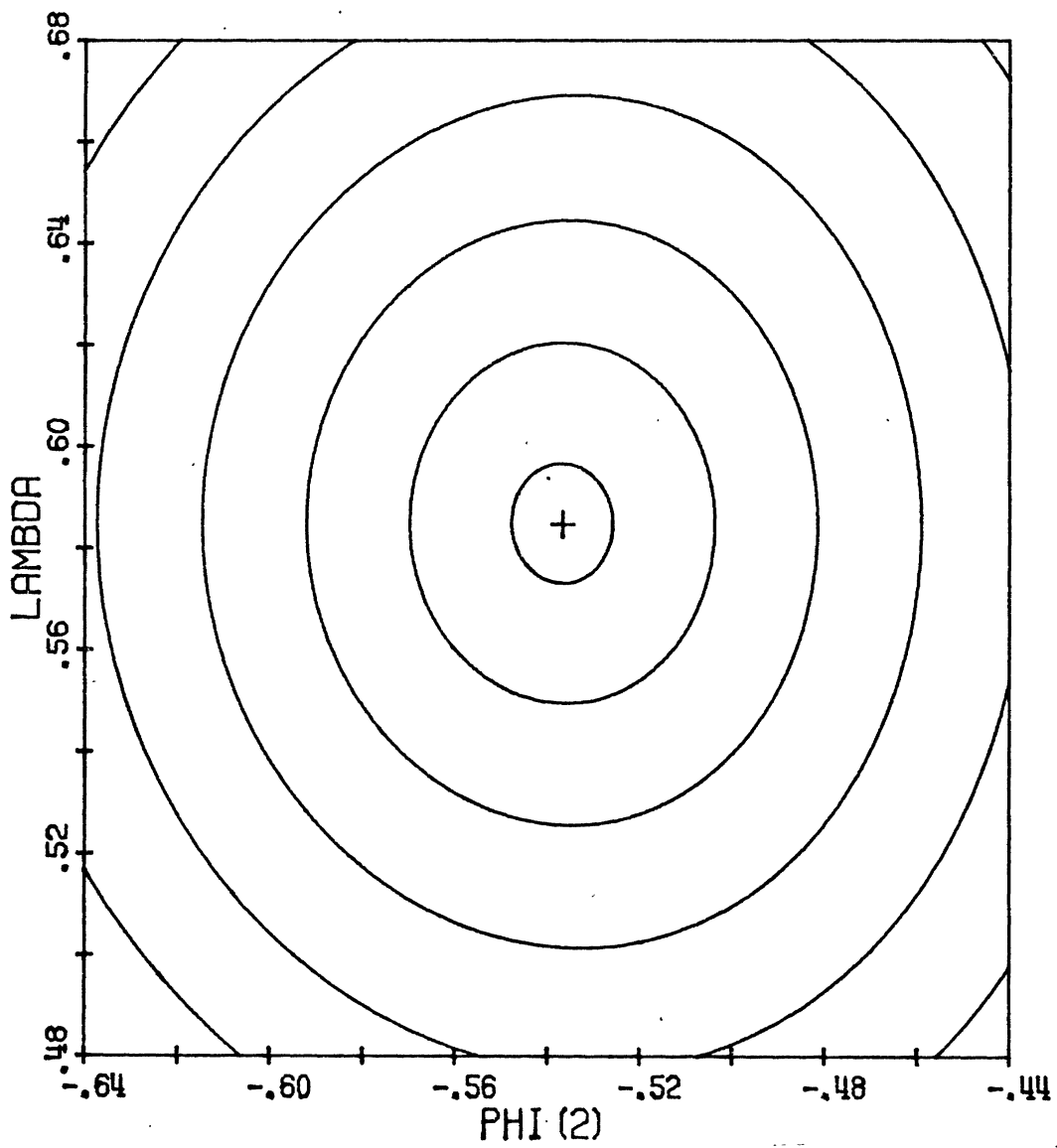


Fig. 5. Log Likelihood Contours.

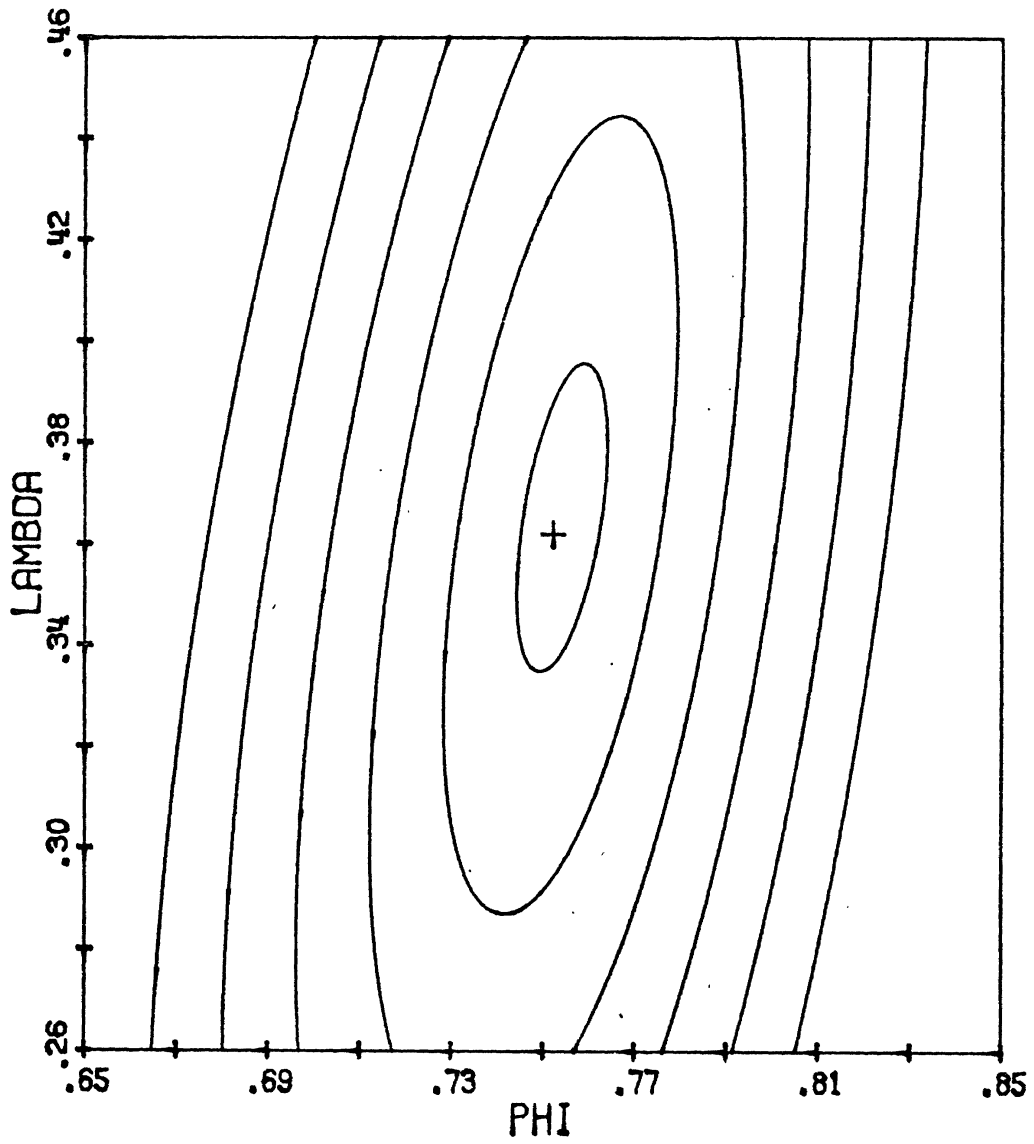


Fig. 6. Log Likelihood Contours.

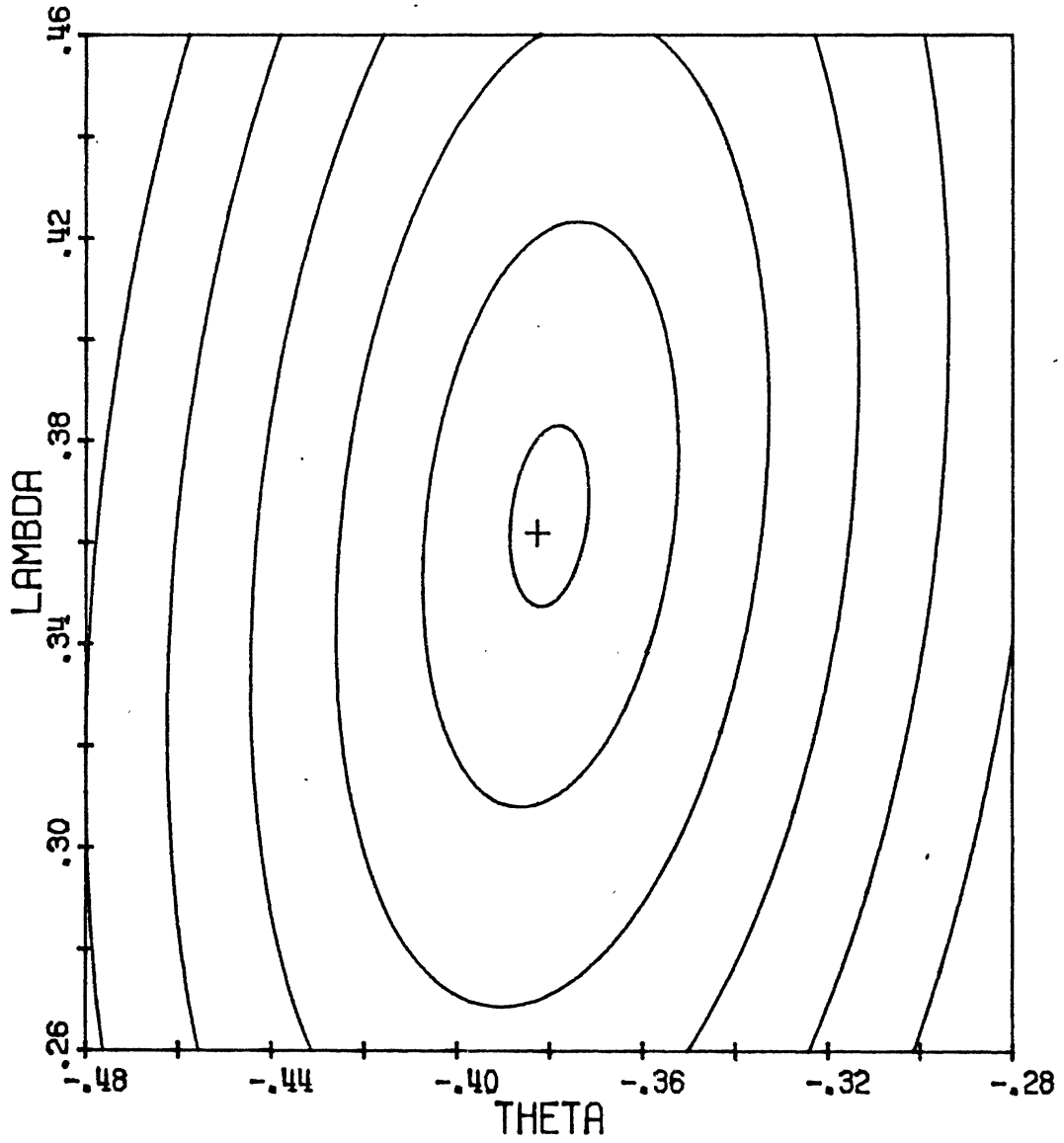


Fig. 7. Log Likelihood Contours.

The variation of L^* over (θ, λ) is shown in Figure 8 in the same manner as for the previous cases.

The large discrepancies between estimates and true values of δ and σ^2 , respectively, are an interesting feature of the estimates determined from the simulation results. The transformation (1) is in effect a choice of a new metric for the analysis. Because the units of measurement of $\hat{\delta}$ are units of the new metric and the units of $\hat{\sigma}^2$ are (units)² of this metric, these estimates are very sensitive to the choice of λ . An investigation of the joint properties of the maximum likelihood estimates of λ and the structural parameters will be the subject of a future paper by the authors.

Application to the Analysis of U.S. Money Supply, M2

We illustrate the results of the preceding analysis by applying them to the U.S. money supply (M2), seasonally adjusted. In Table 1 the data are shown by month for the period January, 1970 through April, 1975, and a plot of the data is given in Figure 9.

An examination of Figure 9 suggests that there may be an exponential trend in the data so that an initial value of λ less than 1 may be appropriate. In Figure 10 we have plotted the series after transformation for $\lambda = 0.75, 0.50, 0.25$, and 0 (the logarithm).

A slight overtransformation is suggested for $\lambda = 0.5$ which becomes more pronounced for $\lambda = 0.25$ and $\lambda = 0$. The

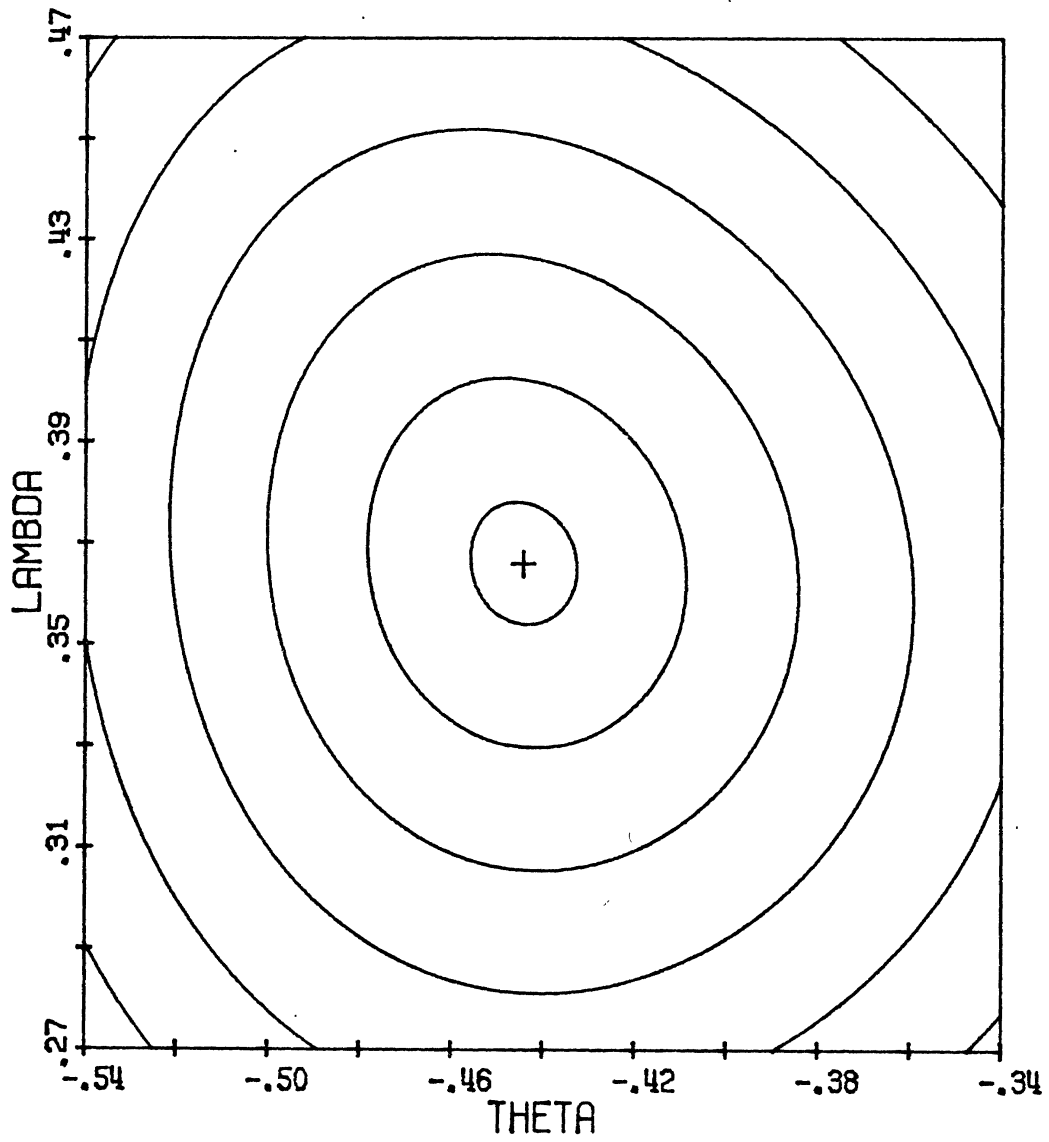


Fig. 8. Log Likelihood Contours.

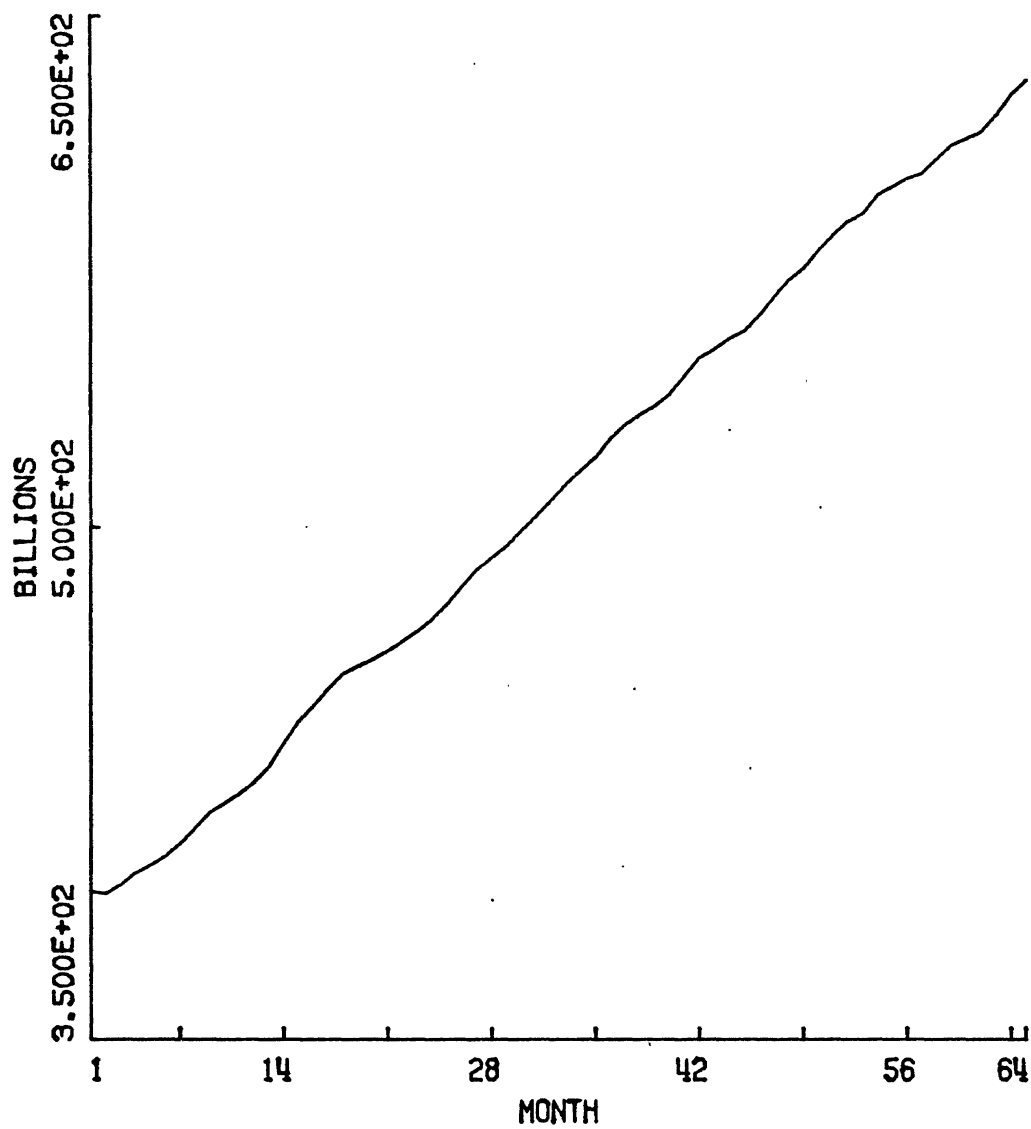


Fig. 9. U.S. Money Supply (M2).

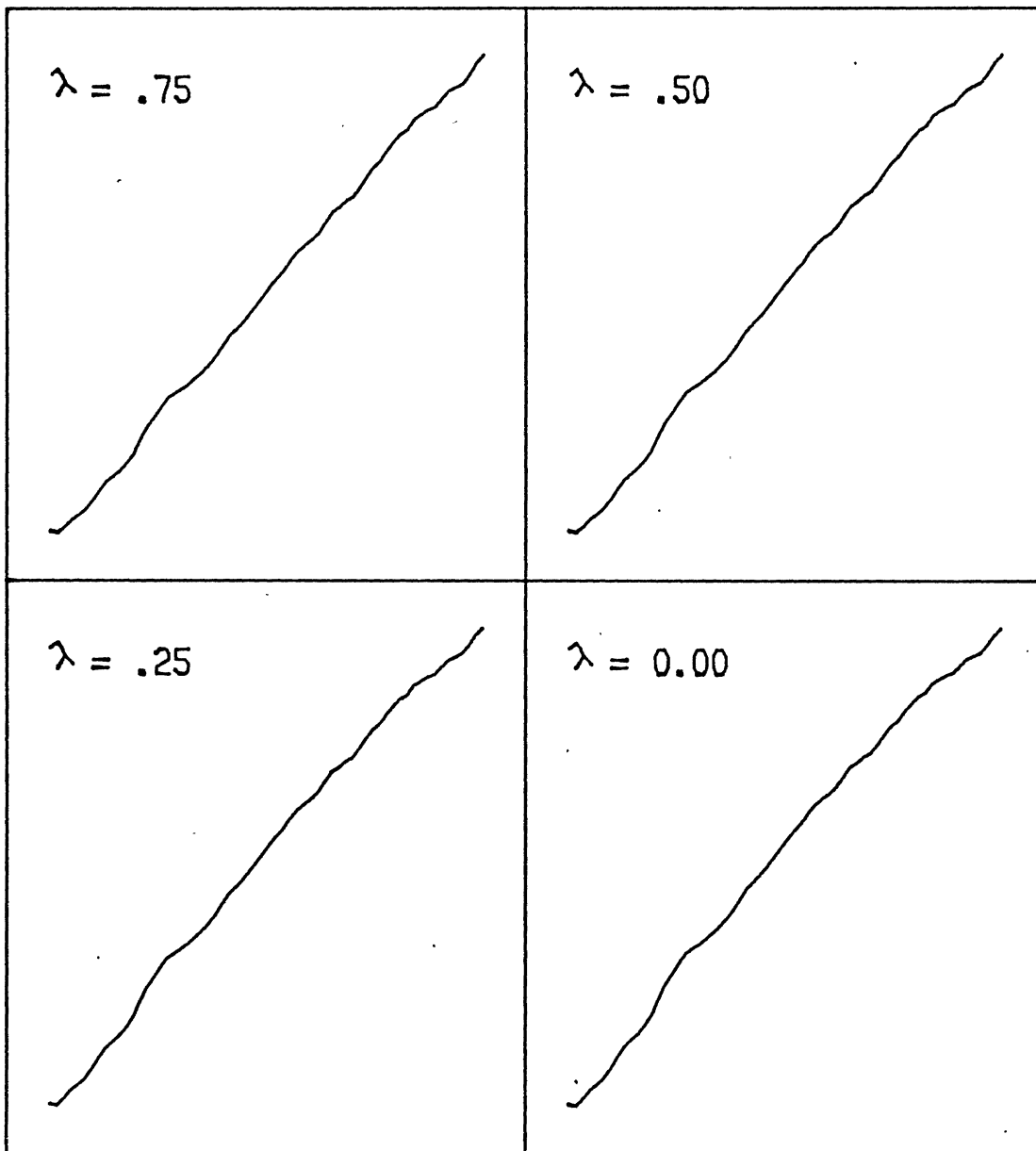


Fig. 10. U.S. Money Supply (M2) S.A.

Table 1

U.S. MONEY SUPPLY (M2), SEASONALLY ADJUSTED
(billions of dollars)

	1970	1971	1972	1973	1974	1975
J	393.3	429.9	477.5	529.8	575.5	615.5
F	392.6	436.7	482.7	532.9	580.9	620.3
M	395.3	443.1	487.5	535.3	585.5	626.4
A	398.8	447.6	490.9	538.8	589.4	630.4
M	401.0	452.6	494.3	544.2	591.6	
J	403.6	457.1	498.7	549.5	597.1	
J	407.4	459.3	503.2	551.9	599.6	
A	411.9	461.2	507.8	555.1	601.9	
S	416.4	463.7	512.3	557.2	603.4	
O	419.2	466.6	516.5	561.6	607.6	
N	421.9	469.6	520.3	567.2	611.6	
D	425.3	473.1	525.7	572.2	613.5	

Source: Various issues of the Federal Reserve Bulletin.

series transformed by $\lambda = 0.75$, however, shows a trend that appears to be approximately linear; thus we could expect that the first differences would be approximately stationary. The identification stage of the Box-Jenkins procedure was therefore applied to the series transformed with $\lambda = 0.75$. First differences of the transformed series appeared to be stationary, and the differenced series was identified as an ARMA(2,0) process. Initial estimates obtained from the identification phase were:

$$\phi_1^{(0)} = 0.44,$$

$$\phi_2^{(0)} = -0.19,$$

$$\delta(0) = 0.60,$$

$$\lambda(0) = 0.75.$$

Maximum likelihood estimates obtained by our procedure were:

$$\hat{\phi}_1 = 0.561,$$

$$\hat{\phi}_2 = 0.266,$$

$$\hat{\delta} = 0.588,$$

$$\hat{\lambda} = 0.759,$$

$$\hat{\sigma}^2 = 0.718.$$

The variation of the likelihood function L^* with respect to λ for this series is shown in Figure 11. Finally, we show in Figure 12 the effect on the linearity of the trend when $\lambda = 0.759$ is used in the transformation (1).

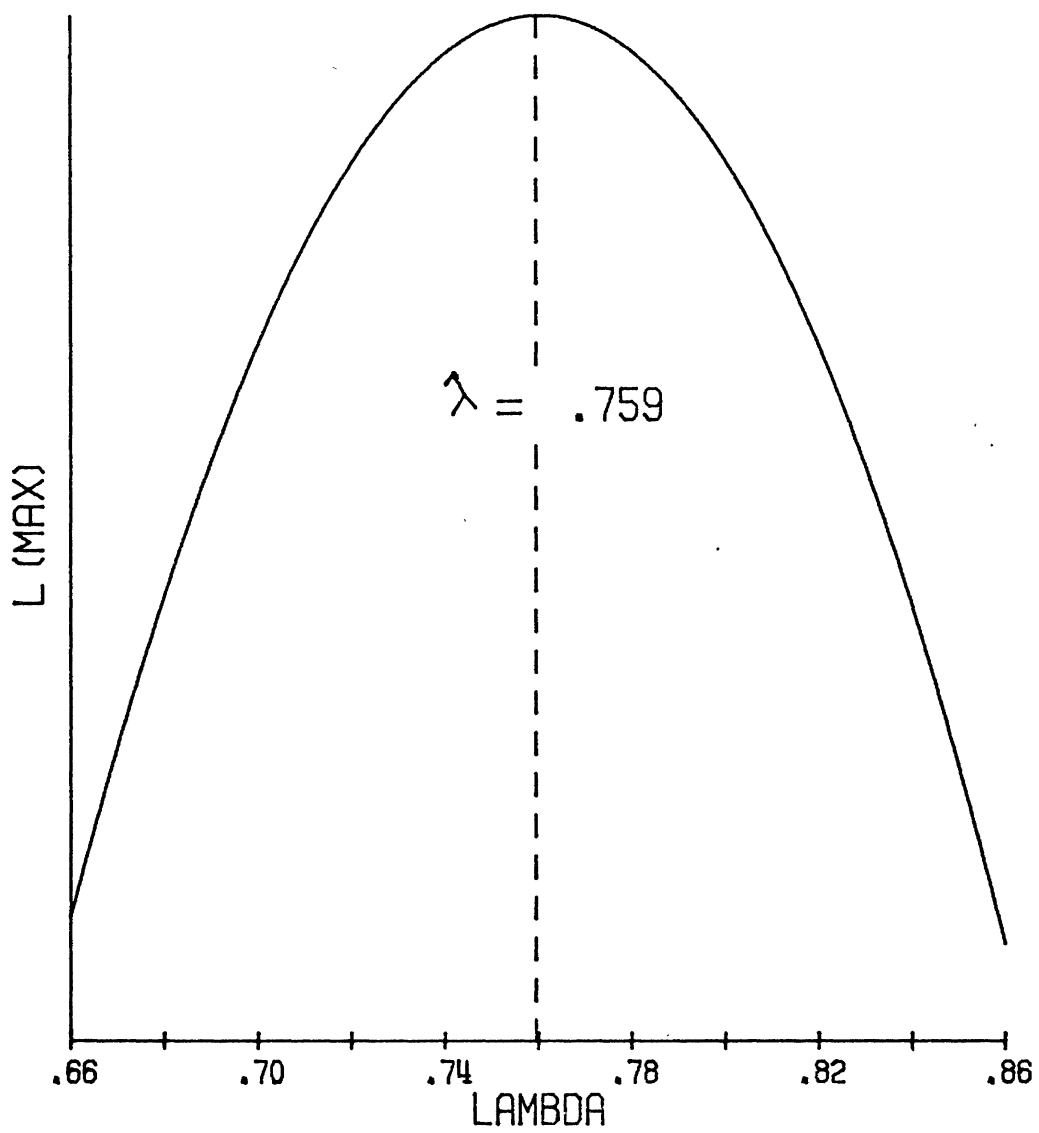


Fig. 11. Log Likelihood (M2).

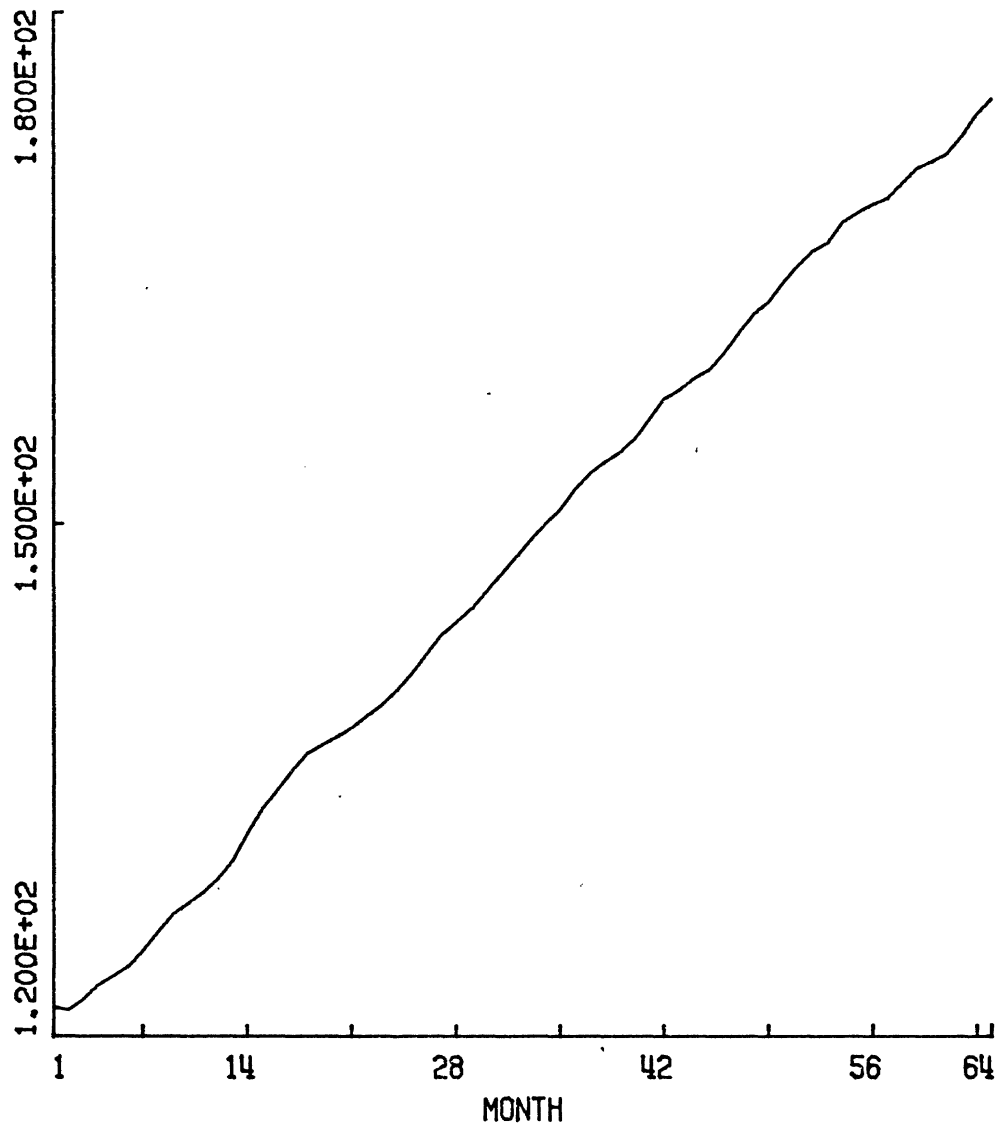


Fig. 12. Transformed M2 (Lambda = .759).

BIBLIOGRAPHY

1. Box, G.E.P., and Cox, D.R. "An Analysis of Transformations." Journal of the Royal Statistical Society, B, 26 (1964): 211-43.
2. Box, G.E.P., and Jenkins, G.M. "Some Comments on a Paper by Chatfield and Prothero and on a Review by Kendall." Journal of the Royal Statistical Society, A, 135 (1973): 337-45.
3. _____. Time Series Analysis, Forecasting and Control. San Francisco, Calif.: Holden-Day, 1970.
4. Chatfield, C., and Prothero, D.L. "Box-Jenkins Seasonal Forecasting: Problems in a Case-study." Journal of the Royal Statistical Society, A, 136 (1973): 295-315.
5. _____. "A Reply by Dr. Chatfield and Dr. Prothero." Journal of the Royal Statistical Society, A, 136 (1973): 345-52.
6. Harrison, P.J. "Discussion of a Paper by Dr. Chatfield and Dr. Prothero." Journal of the Royal Statistical Society, A, 136 (1973): 319-24.
7. Marquardt, D.W. Journal of the Society of Industrial and Applied Mathematics, 11 (1963): 431-41.
8. Tunnicliffe Wilson, G. "Discussion of a Paper by Dr. Chatfield and Dr. Prothero." Journal of the Royal Statistical Society, A, 136 (1973): 315-19.

REQUISITION

UNIVERSITY OF MICHIGAN

CHARGE TO

PRINTING ADDRESSING

Fund/Dept. Expend. Restr.

Acct. No. 307680 Acct. Res., Publ. & Serv.

FILL IN EXACT NAME OF FUND OR DEPARTMENT AND BUDGET ACCOUNT

SEND ORIGINAL AND
DUPLICATE TO ABOVE
NAMED DEPARTMENT

DATE 8-19-75

QUANTITY	DESCRIPTION	REF. NO. LEAVE BLANK
100 copies	One-sided from bond typescript	
	Working Paper 117 (Spivey and Wroblewski)	
	28 pp.	
	(Our project 911)	

DO NOT WRITE BEYOND THIS LINE

INSTRUCTIONS

Services of Printing Office or Addressing Service may be ordered direct on this requisition form. Check the appropriate square at top of form. Do not intermix items for the different services but use separate requisitions instead. Send white and yellow copies of requisition. Yellow copy will be returned to you with reference number added. When calling, refer to reference number.

ADDRESSING — Services include preparation of metal addressing plates and addressograph printing from plates on labels, envelopes, self mailing pieces, publications, file cards, etc. Care and upkeep of plates, deletions, new additions, changes, coding done on order of list owner. Use of other than own list must have authorization by owning department. Diamond labels kept in stock for immediate use. High-speed mechanical application of labels is available.

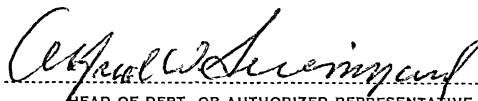
PRINTING—Give quantity and description by title or name of item. If similar to a previous order give old reference number and form number if previously used. Provide as many details as possible. Copy, sample, sketch, photocopy or other material to identify or describe finished product must accompany requisition.

Proofs desired
Yes No Phone Campus Mail Via 164 LS&A

Contact person Alice Preketes
Address 412A Bus. Ad. Phone 4-1366

Deliver to

Will call for at: North Campus Office

Signed 
HEAD OF DEPT. OR AUTHORIZED REPRESENTATIVE