AD-A056 540 STATE UNIV OF NEW YORK AT BUFFALO AMHERST STATISTICA--ETC F/G 12/1
FORECASTING AND WHITENING FILTER ESTIMATION.(U)
JUN 77 E PARZEN
UNCLASSIFIED TR-54 N00014-72-C-0508
NL

State University of New York at Buffalo
Department of Computer Science

LEVEL II

12

(6) FORECASTING AND WHITENING FILTER ESTIMATION*,

(9) Technical rept.,

by

(10) Emanuel/Parzen

Statistical Science Division
State University of New York at Buffalo

(14) TR-54

(15) N0014-72-C-0508

TECHNICAL REPORT NO. 54

(11) June 1977

(12) 38 p.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

* Research supported in part by the Office of Naval Research under Contract N00014-72-C-0508 (NR 042-234).

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 54	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Forecasting and Whitening Filter Estimation		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) N00014-72-C-0508
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistical Science Division State University of New York at Buffalo Amherst, New York 14226		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-234
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE June 1977 13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Forecasting
Time Series Analysis
Autoregressive schemes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An approach to empirical time series analysis is described in which the identification stage is not accomplished chiefly by graphical inspection of the time series and of computed auxiliary sample functions such as the autocorrelation function, partial autocorrelation function, and spectrum. Rather the transfer function g_∞ of the whitening filter is directly estimated and parsimoniously parametrized.

(Continued on attached sheet.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

78 07 12 013

20. ABSTRACT (cont'd)

A criterion for choosing a regression model for forecasting is described in Section 1. A model identification procedure for a stationary time series is described in Section 2. Model identification for a non-stationary time series is discussed in Section 3. Our approach is illustrated by an example in Section 4.

FORECASTING AND WHITENING FILTER ESTIMATION*

by

Emanuel Parzen

Statistical Science Division
State University of New York at Buffalo
Amherst, New York 14226

Abstract

An approach to empirical time series analysis is described in which the identification stage is not accomplished chiefly by graphical inspection of the time series and of computed auxiliary sample functions such as the autocorrelation function, partial autocorrelation function, and spectrum. Rather the transfer function g_∞ of the whitening filter is directly estimated and parsimoniously parametrized.

A criterion for choosing a regression model for forecasting is described in Section 1. A model identification procedure for a stationary time series is described in Section 2. Model identification for a non-stationary time series is discussed in Section 3. Our approach is illustrated by an example in Section 4.

*Research supported by the Office of Naval Research under contract N00014-72-C-0508 (NR 042-234).

1. Choosing a Regression Model for Forecasting

Let Y be a variable (representing hourly electricity demand, daily births, weekly money supply, monthly retail sales of department stores, quarterly nonform inventory investment, annual gross national product, and so on) whose value denoted $Y(t)$, at time t one desires to forecast (predict) or explain. Either aim is accomplished by means of a decomposition of the value $Y(t)$ into the sum of two components as follows:

$$Y(t) = Y^{\mu}(t) + Y^{\nu}(t)$$

$Y^{\mu}(t)$ is the explained or predictable part of $Y(t)$

$Y^{\nu}(t)$ is the error, or unexplained, or unpredictable part of $Y(t)$

The explained part $Y^{\mu}(t)$ is usually a <u>linear</u> function of "explanatory" variables $X_1(t), X_2(t), \ldots$ with coefficients denoted $\beta_1, \beta_2, \ldots$; explicitly,

$$Y^{\mu}(t) = \beta_1 X_1(t) + \ldots + \beta_j X_j(t) + \ldots + \beta_k X_k(t) + \ldots$$

We write the sum as possibly infinite series, because in theory there is an infinite number of possible explanatory variables. However, only a finite number of explanatory variables X_j are expected to have coefficients β_j which are different enough from zero that the benefit (in mean square error terms) of estimating β_j is to be preferred to the cost of considering β_j to be equal to 0, and thus omitting X_j from the model.

The error term $Y^{\nu}(t)$ is best regarded as the residual $Y(t) - Y^{\mu}(t)$ after constructing $Y^{\mu}(t)$ to explain as much as possible of the value of $Y(t)$. We call $Y^{\nu}(t)$ the innovation at time t.

We use the Greek letter nu as a superscript to indicate that $Y^{\nu}(t)$ represents what's "new" (or "transitory") in $Y(t)$ after explaining as much of its value as possible by the best available explanatory variables $X_1(t),\ldots,X_k(t)$. We use the Greek letter μ as a superscript because it connotes a "mean," and $Y^{\mu}(t)$ connotes an average or smooth value about which $Y(t)$ fluctuates.

It is customary to denote the error term by $\epsilon(t)$, and write the model for $Y(t)$ as a regression model

$$Y(t) = \beta_1 X_1(t) + .. + \beta_k X_k(t) + \epsilon(t), \quad t = 1,2,..,n$$

The "ordinary" or "naive" regression model assumes that the errors $\epsilon(t)$ are independent identically distributed random variables, each with mean zero and variance σ^2. The best estimators of $\beta_1,\ldots,\beta_k$ (according to such criteria as least squares, maximum likelihood, or minimum variance unbiased), denoted $\hat{\beta}_1,..,\hat{\beta}_k$, in the model (now written in matrix form)

$$Y = X\beta + \epsilon$$

are the solutions of the normal equations

$$\hat{\beta} = (X'X)^{-1} X'Y$$

where

$$\beta = \begin{bmatrix} \beta_1 \\ \dots \\ \beta_k \end{bmatrix} , \quad X = \begin{bmatrix} X_1(1) \dots X_k(1) \\ \dots \quad \dots \quad \dots \\ X_n(1) \dots X_n(k) \end{bmatrix} , \quad Y = \begin{bmatrix} Y(1) \\ \dots \\ Y(n) \end{bmatrix} , \quad \epsilon = \begin{bmatrix} \epsilon(1) \\ \dots \\ \epsilon(n) \end{bmatrix}$$

Let Y^{μ} and Y^{ν} now denote the vectors of smoothed and residual values:

$$Y^{\mu} = \begin{bmatrix} Y^{\mu}(1) \\ \dots \\ Y^{\mu}(n) \end{bmatrix} , \quad Y^{\nu} = \begin{bmatrix} Y^{\nu}(1) \\ \dots \\ Y^{\nu}(n) \end{bmatrix}$$

Representations of Y^{μ} and Y^{ν} are

$$Y^{\mu} = AY , \quad Y^{\nu} = (I - A) Y$$

where

$$A = X(X'X)^{-1} X' .$$

A is called (see Hoaglin and Welch (1976)) the hat matrix; I call I - A the whitening matrix. The residual sum of squares can be represented

$$\sum_{t=1}^{n} |Y^{\nu}(t)|^2 = \|Y^{\nu}\|^2 = \|(I - A)\ Y\|^2$$

The sample multiple correlation coefficient R_X of Y given X is defined by $1 - R_X^2 = \|(I - A)\ Y\|^2 + \|Y\|^2$.

Given the regression model $Y = X\beta + \epsilon$ one may distinguish three problems: estimate

(i) $X\beta = Y^{\mu}$, smoothing or forecasting;

(ii) β , parameter estimation,

(iii) X , model identification.

We have described the solution to the parameter estimation problem when the model matrix X is fixed, but this solution is regarded as unsatisfactory on various criteria; Dempster, Schatzoff, and Wermuth (1977) compare alternative solutions, including <u>subset regression</u> (choose an optimal submatrix $X_{(p)} = [X_{j_1} : X_{j_2} : \ldots : X_{j_p}]$ of variables on which to regress Y) and ridge regression (estimate β by

$$\hat{\beta} = (X'X + \lambda I)^{-1} X'Y$$

for a suitable choose of ridge parameter λ). Which procedure to use in practice is best determined <u>adaptively</u> from the data rather than a priori on the basis of theoretical considerations. I believe recent research by Wahba (1976) provides insight into a criterion which can be used to choose the optimal regression model and parameter estimator. To each model and estimator one

can associate: (1) a hat matrix A [for subset regression it is $A_p = X_{(p)}\left(X'_{(p)}X_{(p)}\right)^{-1} X'_{(p)}$ while for ridge regression it is $A_\lambda = X(X'X + \lambda I)^{-1} X'$] and (2) a "cross-validation" criterion

$$CV(A) = \frac{\|(I - A)\, Y\|^2}{\{\text{Trace } (I - A)\}^2}$$

The optimum smoother $Y^\mu = \hat{A}Y$ corresponds to the hat matrix $\hat{A}$, defined as the matrix minimizing CV(A) over the family of hat matrices A one is considering.

The justification for this assertion is partly its successful application in practice and partly a variety of theoretical properties. Its justification is not the aim of this paper; rather we seek only to note the existence of criteria for selecting a regression model in order to motivate our approach toward finding criteria for selecting a time series model.

2. Identifying a Stationary Time Series Model

Let $Y(t)$, $t = 0, \pm 1, \ldots$ be a normal stationary time series. Then the explanatory variables to be used to explain or predict $Y(t)$ are past values $X_1(t) = Y(t-1), X_2(t) = Y(t-2), \ldots$ of which in theory there are an infinite number. In the decomposition $Y(t) = Y^\mu(t) + Y^\nu(t)$, $Y^\mu(t)$ is defined to be the conditional expectation of $Y(t)$ given the infinite past $Y(t-1), Y(t-2), \ldots$:

$$Y^\mu(t) = E[Y(t) \,|\, Y(t-1), Y(t-2), \ldots] ,$$

called the infinite memory one step ahead predictor. The model identification problem then corresponds to finding the memory m such that the finite memory predictor

$$Y^{\mu,m}(t) = E[Y(t)|Y(t-1),\ldots,Y(t-m]$$

performs as well as the infinite memory predictor, or more precisely $Y^{\mu}(t)$ and $Y^{\mu,m}(t)$ <u>almost</u> coincide.

In the theory of time series analysis it is useful to define an ideal time series model by the condition $Y^{\mu}(t)$ and $Y^{\mu,m}(t)$ exactly coincide; we define the time series $Y(\cdot)$ to be an autoregressive scheme of order m if

$$Y^{\mu}(t) = Y^{\mu,m}(t)$$

or equivalently

$$Y^{\nu,m}(t) = Y(t) - Y^{\mu,m}(t)$$

is white noise in the sense that

$$E\left(Y^{\nu,m}(s)\, Y^{\nu,m}(t)\right) = 0 \qquad \text{for } s \neq t$$

The coefficients in the representation of $Y^{\mu,m}(t)$ as a linear combination of $Y(t-j)$, $j = 1,..,m$ are denoted by $-\alpha_{j,m}$, where α symbolizes "autoregressive." Thus we write

$$-Y^{\mu,m}(t) = \alpha_{1,m}Y(t-1) + .. + \alpha_{m,m}Y(t-m) \quad .$$

Define the backward shift operator L by

$$LY(t) = Y(t-1)$$

(which means the same as B in the notation of Box and Jenkins). Finally define the polynomial

$$g_m(z) = 1 + \alpha_{1,m}z + .. + \alpha_{m,m}z^m$$

Then

$$Y^{\nu,m}(t) = Y(t) - Y^{\mu,m}(t)$$

$$= g_m(L)\, Y(t)$$

Similarly we define infinite memory (i) predictions, (ii) autoregressive transfer function, and (iii) innovations:

(i) $$-Y^{\mu}(t) = \alpha_{1,\infty}Y(t-1) + .. + \alpha_{m,\infty}Y(t-m) + \ldots$$

(ii) $$g_\infty(z) = 1 + \alpha_{1,\infty}z + \ldots + \alpha_{m,\infty}z^m + \ldots$$

(iii) $$Y^{\nu}(t) = g_\infty(L)\, Y(t)$$

For future reference define the mean square prediction errors

$$\sigma_m^2 = E[|Y^{\nu,m}(t)|^2] \ , \ \sigma_\infty^2 = E[|Y^{\nu}(t)|^2] \ .$$

We call $g_\infty(z)$ the autoregressive transfer function (ARTF) of the stationary time series $Y(\cdot)$ since one can write symbolically

time series $Y(\cdot)$ — g_∞ — $\epsilon(\cdot)$ white noise .

In words, $g_\infty(L)$ is the whitening filter.

<u>The time series (model identification) problem can be defined to be the estimation of</u> g_∞ ; this is equivalent to defining the regression modeling problem as estimation of the "optimal" hat matrix A .

We will attempt to clarify the role of finite parameter schemes, such as AR (autoregressive) schemes of order p :

$$g_p(L)\ Y(t) = \epsilon(t) \ ;$$

MA (moving average) schemes of order q :

$$Y(t) = h_q(L)\ \epsilon(t) \ , \ h_q(z) = 1 + \beta_1 z + .. + \beta_q z^q \ ;$$

ARMA (autoregressive-moving average) schemes of order (p,q) :

$$g_p(L)\, Y(t) = h_q(L)\, \epsilon(t) \quad .$$

To interpret the polynomials $g_p(z)$ and $h_q(z)$ parametrizing an ARMA scheme, one uses them to form the ARTF:

$$g_\infty(z) = h_q^{-1}(z)\, g_p(z)$$

The assumption of a parsimonious ARMA scheme is adopted to provide a parsimonious finite parametric representation of g_∞ in order to enable it to be estimated. However, one can estimate g_∞ non-parametrically using approximating AR schemes. It is my view that in modeling <u>stationary</u> time series only AR schemes are needed. However, it will be shown in the next section that ARMA schemes are useful for modeling non-stationary time series.

To form estimators $\hat{\sigma}_m^2, \hat{\alpha}_{1,m}, \ldots, \hat{\alpha}_{m,m}$ of the parameters of an AR scheme of order m we use the Yule-Walker equations: for $j = 1,\ldots,m$

$$\hat{\rho}(-v) + \hat{\alpha}_{1,m}\hat{\rho}(1 - v) + \ldots + \hat{\alpha}_{m,m}\hat{\rho}(m - v) = 0$$

where

$$\hat{\rho}(v) = \sum_{t=1}^{T-v} Y(t)\, Y(t + v) \div \sum_{t=1}^{T} Y^2(t)$$

is an estimator of $\rho(v) = \mathrm{corr}\Big(Y(t), Y(t+v)\Big)$. Further

$$\hat{\sigma}_m = \hat{\rho}(0) + \hat{\alpha}_{1,m}\hat{\rho}(1) + \ldots + \hat{\alpha}_{m,m}\hat{\rho}(m) \quad .$$

It seems plausible that there is a value of m, denoted $\hat{m}$, such that

$$\hat{g}_m(z) = 1 + \hat{\alpha}_{1,m} z + \ldots + \hat{\alpha}_{m,m} z^m$$

for $m = \hat{m}$ is an "optimal" estimator of g_∞; in symbols,

$$\hat{g}_\infty(z) = \hat{g}_{\hat{m}}(z) \quad .$$

One can show that approximately the overall mean square percentage error of any $\hat{g}_m$ as an estimator of g_∞ satisfies

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left| \frac{\hat{\sigma}_m^{-2}\hat{g}_m(e^{i\omega}) - \sigma_\infty^{-2} g_\infty(e^{i\omega})}{\sigma_\infty^{-2} g_\infty(e^{i\omega})} \right|^2 d\omega$$

$$= \frac{1}{T}\sum_{j=1}^{m} \sigma_j^{-2} + \sigma_\infty^{-2} - \sigma_m^{-2}$$

where T is the sample size. Therefore in practice we choose $\hat{m}$ to minimize a criterion function $CAT(m)$, defined as follows:

$$CAT(0) = -(1 + \frac{1}{T})$$

$$CAT(m) = \frac{1}{T}\sum_{j=1}^{m} \hat{\sigma}_j^{-2} - \hat{\sigma}_m^{-2}$$

where $\hat{\hat{\sigma}}^2_m$ is an unbiased estimator of σ^2_m defined by

$$\hat{\hat{\sigma}}^2_m = \left(1 - \frac{m}{T}\right)^{-1} \hat{\sigma}^2_m \quad .$$

When $\hat{m} = 0$ we accept the hypothesis that the observed time series is white noise.

Having determined the maximum order $\hat{m}$ of the approximating autoregressive transfer function $\hat{g}_{\hat{m}}(z)$ we next use stepwise regression techniques to determine the significantly non-zero autoregressive coefficients in the transfer function. As an example, on monthly data if one determined that $\hat{m} = 13$, it would be of interest to determine whether $\hat{g}_{13}(z)$ were approximately of the form

$$\hat{g}_{13}(z) = (1 - \theta_1 z)(1 - \theta_{12} z^{12}) \quad .$$

Stepwise, or subset, autoregression is discussed by McClave (1975).

3. Identifying a Non-Stationary Time Series Model

For a non-stationary time series, the modeling problem is not only to find the whitening filter (which transforms $\{Y(t)\}$ to $\{Y^{\nu}(t)\}$, but interpret it as several filters in series:

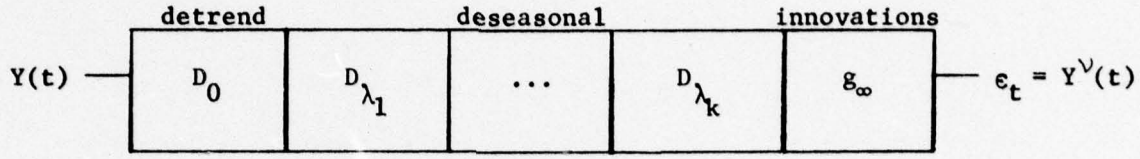
D_0 : a detrending filter which in the spectral domain eliminates the low frequency components corresponding to trend,

D_λ : a de-seasonal filter which in the spectral domain eliminates the components corresponding to a periodic component with

period λ , or to the harmonics with frequencies which are multiples of $\frac{2\pi}{\lambda}$,

g_∞ or Π : an innovations filter which transforms to white noise the series $Y^{(stat)}(t) = D_0 D_\lambda Y(t)$ representing a transformation of $Y(t)$ to a stationary series.

The time series modeling problem is thus to find the filter representation



where we admit the possibility of several different periods $\lambda_1,\ldots,\lambda_k$ (for example in monthly data λ values are often 12 and 3 , in daily data λ values are often 7 and 365 , in hourly data λ values are often 24 and 168).

Given the above decomposition, one can form various derived series:

$Y^{(0)}(t) = D_0 Y(t)$ the detrended series,

$Y^{(\lambda)}(t) = D_\lambda Y(t)$ the seasonally adjusted series,

$Y^{(0,\lambda)}(t) = Y^{(stat)}(t) = D_0 D_\lambda Y(t) = D_\lambda D_0 Y(t)$, the detrended seasonally adjusted series,

$Y^\nu(t) = Y^{(white)}(t) = g_\infty D_0 D_\lambda Y(t)$ the innovations series.

Such decompositions seem to be crucial to the study of the relations between time series $Y_1(\cdot)$ and $Y_2(\cdot)$. To study their relations it seems

clear that if one relates $Y_1(\cdot)$ and $Y_2(\cdot)$ without filtering one will often find "spurious" relationships. It has been suggested therefore that one attempt to relate $Y_1^{\nu}(\cdot)$ and $Y_2^{\nu}(\cdot)$, the individual innovations of each series. What remains to be examined is the insight to be derived from relating $Y_1^{(\lambda)}(t)$ and $Y_2^{(\lambda)}(t)$, the seasonally-adjusted series, or $Y_1^{(stat)}(t)$ and $Y_2^{(stat)}(t)$, the detrended and seasonally adjusted series.

The question remains of how to find in practice the detrending and deseasonal filters. To seasonally adjust for a period λ in data, several possibilities are available which may be interpreted as seasonal adjustment filters.

A filter with the same zeroes in the frequency domain as some usual procedures, which is recursive (acts only on past values), and yields a variety of filter shapes (in the frequency domain) between a square wave and a sinusoid is the one-parameter family of filters

$$D_{\lambda}(\theta) = \frac{I - L^{\lambda}}{I - \theta L^{\lambda}} ,$$

where the parameter θ is chosen (usually by an estimation procedure) between 0 and 1. When $\theta = 0$ the filter is denoted ∇_{λ} and called λ-th difference.

To understand the role of the filter $D_{\lambda}(\theta)$, denote it for brevity by D and rewrite it as follows; writing $I - L^{\lambda} = I - \theta L^{\lambda} - (1 - \theta) L^{\lambda}$ we obtain

$$D = I - \frac{(1 - \theta) L^{\lambda}}{I - \theta L^{\lambda}} = I - \{1 - \theta)(L^{\lambda} + \theta L^{2\lambda} + \theta^2 L^{3\lambda} + \ldots)\} .$$

Then the output $Y^{(D)}(t) = DY(t)$ of a filter D with input $Y(t)$ can be written

$$Y^{(D)}(t) = Y(t) - (1 - \theta)\{Y(t - 2\lambda) + \dots\} \quad .$$

In words, $Y^{(D)}(t)$ is the result of subtracting from $Y(t)$ the exponentially weighted average of $Y(t - \lambda), Y(t - 2\lambda), \dots$.

It seems to me open to investigation whether the filter D (of mixed autoregressive-moving average type) is superior to the approximately equivalent autoregressive filter

$$D' = (I - L^{\lambda})(I + \theta L^{\lambda}) = I - (1 - \theta)\, L^{\lambda} - \theta L^{2\lambda}$$

whose output $Y^{(D')}(t) = D'Y(t)$ can be written

$$Y^{(D')}(t) = Y(t) - (1 - \theta)\, Y(t - \lambda) - \theta Y(t - 2\lambda) \quad .$$

It appears to me that <u>the role of moving averages in Box-Jenkins ARIMA models is to build filters of the type</u> $D_{\lambda}(\theta)$. Thus the ARIMA model

$$(I - L)(I - L^{12})\, Y(t) = (I - \theta_1 L)(I - \theta_{12} L^{12})\, \epsilon_t$$

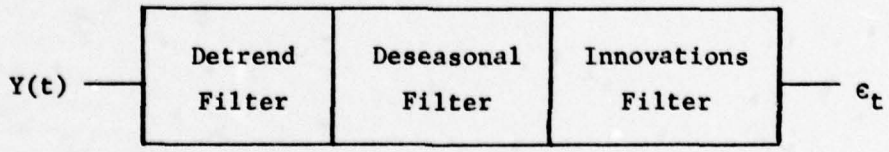
should be viewed as the whitening filter

$$D_1(\theta_1)\, D_{12}(\theta_{12})\, Y(t) = \left(\frac{I - L}{I - \theta_1 L}\right)\left(\frac{I - L^{12}}{I - \theta_{12} L^{12}}\right) Y(t) = \epsilon_t$$

I should like to emphasize that the output of the filter $D_1(\theta_1)\, D_{12}(\theta_{12})$ is often not white noise but is only a stationary time series. For purposes of one-step ahead prediction, it is often not important to differentiate between the case that $D_{12}D_1Y(t)$ is white noise or not, since most of the predictability is obtained by finding a suitable transformation to stationarity of the form D_1D_{12} (below we discuss naive prediction as a transformation to stationarity).

The moral to be drawn from the foregoing considerations is as follows. To find a transformation of a non-stationary time series to stationarity it may suffice to apply pure differencing operators such as $I - L$ and $I - L^{12}$. However, the transformation of the residuals to the innovations series should be expressed if possible in terms of factors corresponding to the filters $I - \theta_1 L$ and $I - \theta_{12}L^{12}$ since such factors enable us to interpret the overall whitening filter

as a series of filters



which can be interpreted as helping to provide solutions to the seasonal adjustment problem.

Naive Prediction and Transformations to Stationarity: To predict a time series $Y(t)$ one can often suggest a "naive" predictor of the form

$$Y^{naive}(t) = Y(t - \lambda_1) + Y(t - \lambda_2) - Y(t - \lambda_1 - \lambda_2) .$$

The prediction error of this predictor is given by

$$\tilde{Y}(t) = Y(t) - Y^{naive}(t) = Y(t) - Y(t - \lambda_1) - Y(t - \lambda_2) + Y(t - \lambda_1 - \lambda_2) = (I - L^{\lambda_1})(I - L^{\lambda_2}) Y(t) .$$

In words, taking λ_1-th and λ_2-th differences is equivalent to forming the naive prediction errors.

A criterion that $Y(t)$ be non-stationary is that it be predictable (in the sense that the ratio of the average square of $\tilde{Y}(t)$ to the average square of $Y(t)$ is of the order of $1/T$). When $\tilde{Y}(t)$ is stationary (non-predictable) one models it by an approximate autoregressive scheme,

$$\hat{g}_{m}(L)\, \tilde{Y}(t) = \varepsilon_t$$

which can be used to form $\tilde{Y}^{\mu}(t)$, the best one-step ahead predictor of $\tilde{Y}(t)$.

The best one-step ahead predictor of $Y(t)$ is given by

$$Y^{\mu}(t) = Y^{naive}(t) + \tilde{Y}^{\mu}(t) \quad ;$$

to prove this, note the identity

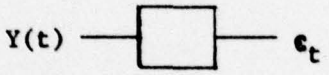
$$Y(t) = Y^{naive}(t) + \tilde{Y}(t)$$

and form the conditional expectation of both sides of this identity with respect to $Y(t-1), Y(t-2), \ldots$.

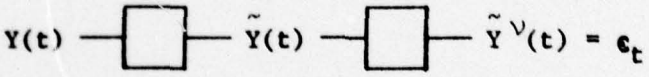
A remarkable fact is the equality of the prediction errors of $Y(t)$ and $\tilde{Y}(t)$:

$$Y^{\nu}(t) = Y(t) - Y^{\mu}(t) = \tilde{Y}(t) - \tilde{Y}^{\mu}(t) = \tilde{Y}^{\nu}(t) \ .$$

It follows that to find the whitening filter



for a non-stationary time series $Y(t)$ (which includes almost all time series with seasonal components) it suffices to apply any one-sided filter (in practice either suggested by an ad hoc deseasonalizing procedure or found by applying the CAT method and discovering that $\hat{\sigma}^2_{\hat{m}}$ is less than $8/T$) whose output $\tilde{Y}(t)$ is stationary. The series filter



then yields the whitening filter. While the filter leading to $\tilde{Y}(t)$ is not unique, the overall filter leading to ϵ_t is unique.

The final seasonal-adjustment procedure is a filter D_λ which comes from interpreting the overall whitening filter as a series of filters in series which can be interpreted as detrending and deseasonalizing filters.

4. Illustrative Example: International Airline Passengers

To illustrate our approach to time series modeling of seasonal data, let us consider a series (used as an illustrative example by Box and Jenkins (1970), p. 305): monthly passenger totals (measured in thousands) in international air travel 1949-1960 which we denote by $Y(t)$ and its logarithms which we denote by $Z(t)$. The series length is 144.

The model fitted by Box and Jenkins to the airline data is

$$\nabla_{12}\nabla_1 Z(t) = (1 - \theta_1 L)(1 - \theta_{12} L^{12})\, \varepsilon(t) ;$$

in words, take first and twelfth differences to transform to a stationary time series which is modeled as a multiplicative moving average. The parameters of this model are estimated by Box and Jenkins to be $\theta_1 = .4$, $\theta_{12} = .6$, $\sigma_\varepsilon^2 = .0013$. The model fitted to the airline data by Box and Jenkins can be written in our notation

$$D_1(.4)\, D_{12}(.6)\, Z(t) = \varepsilon(t), \quad \sigma_\varepsilon^2 = .0013 .$$

In our approach, one has a choice of first steps.

Choice 1: Take first and twelfth differences as <u>obvious</u> "naive" prediction errors; then analyze

$$\tilde{Z}(t) = D_{12} D_1 Z(t)$$

which is a time series of length 131.

Choice II: To determine suitable transformations to transform from a non-stationary time series Z to a stationary time series $\tilde{Z}$, examine a best approximating autoregressive scheme whose order $\hat{m}$ is determined by CAT and whose parsimonious form is determined by subset autoregression. In this way one might be led (without directly examining sample autocorrelations) to try

$$\tilde{Z}(t) = \nabla_1 Z(t) \ , \quad \text{first differences,}$$

or

$$\tilde{Z}(t) = \nabla_{12} Z(t) \ , \quad \text{twelfth differences}$$

as possible transformations to stationarity.

If we adopt Choice I, we note first that $\tilde{Z}(\cdot)$ has variance $R_{\tilde{Z}}(0) = .002$ which is about 1% of the variance of $Z(\cdot)$ which is .194. This indicates that $Z(\cdot)$ is non-stationary. Fitting $\tilde{Z}(\cdot)$ by a suitably long autoregressive scheme with order $\hat{m}$ determined by CAT, one finds $\hat{m} = 12$ with autoregressive coefficients

$\alpha_1 = .36$	$\alpha_7 = .01$
$\alpha_2 = .05$	$\alpha_8 = -.03$
$\alpha_3 = .15$	$\alpha_9 = -.16$
$\alpha_4 = .11$	$\alpha_{10} = -.03$
$\alpha_5 = -.05$	$\alpha_{11} = .08$
$\alpha_6 = -.09$	$\alpha_{12} = .34$

and residual variance $\sigma_\epsilon^2 = .0014$. The best stepwise AR scheme fitted to $\tilde{Z}(t)$ is found to have only lags 1 and 12 :

$$\tilde{Z}(t) + .32\,\tilde{Z}(t-1) + .37\,\tilde{Z}(t-12) = \epsilon(t) \ , \quad \sigma_\epsilon^2 = .0015$$

The residual variance σ_ϵ^2 of a model is actually computed in our program as a proportion (here .75) of $R_{\tilde{Z}}(0)$.

Next one might examine the form of ARMA and MA schemes that fit $\tilde{Z}(\cdot)$. A best fitting ARMA scheme is

$$\tilde{Z}(t) + .34\,\tilde{Z}(t-1) = \epsilon(t) - .39\,\epsilon(t-12)$$

with residual variance $\sigma_\epsilon^2 = .73\ R_{\tilde{Z}}(0) = .0015$. We would regard the ARMA scheme as identical to the AR scheme. If one forces the model for $\tilde{Z}(t)$ to be of the MA form

$$\tilde{Z}(t) = \epsilon(t) + \beta_1\epsilon(t-1) + \beta_{12}\epsilon(t-12) + \beta_{13}\epsilon(t-13)$$

one finds that (using for computational speed a statistically inefficient linear algorithm) $\beta_1 = -.27$, $\beta_{12} = -.38$, $\beta_{13} = .13$ with residual variance $\sigma_\epsilon^2 = .76\ R_{\tilde{Z}}(0) = .0015$. Note that $\beta_1\beta_{12} = .10$ which is close enough to .13 that one could conclude a multiplicative model

$$\tilde{Z}(t) = (1 - .3L)(1 - .4L)\ \epsilon(t) \ ;$$

this multiplicative model for $\tilde{Z}(t)$ enables one to write, in agreement with Box-Jenkins, that

$$D_1(.3)\, D_{12}(.4)\, Z(t) = \varepsilon(t) \quad .$$

If we adopt Choice II, we find that the best approximating autoregressive scheme to $Z(t)$ with order $\hat{m}$ determined by CAT has $\hat{m} = 13$, with coefficients

$$\alpha_1 = -1.00 \qquad \alpha_7 = .02$$
$$\alpha_2 = .09 \qquad \alpha_8 = .08$$
$$\alpha_3 = -.03 \qquad \alpha_9 = -.11$$
$$\alpha_4 = .03 \qquad \alpha_{10} = .02$$
$$\alpha_5 = -.02 \qquad \alpha_{11} = .07$$
$$\alpha_6 = 1.00 \qquad \alpha_{12} = -.45$$
$$\alpha_{13} = .48$$

The residual variance is .06 (of the original variance .1935); it just about equals the threshold $8/T$ (here $T = 131$ for the residual series, and $8/T = .06$) below which we consider a residual variance indicates predictability and non-stationarity. Having determined $\hat{m}$ one then uses stepwise autoregression to determine a more parsimonious AR scheme of order $\hat{m}$ which includes only autoregressive coefficients significantly different from zero. In the present case, one would find only lag 1 significant with coefficient .95; therefore as a $\tilde{Z}$ series representing a transformation of Z from non-stationarity to stationarity we might choose first differences: $\tilde{Z}(t) = Z(t) - Z(t-1)$, which has residual variance .01.

However, it seems intuitively more meaningful to take twelfth differences: $\tilde{Z}(t) = Z(t) - Z(t-12)$ which has residual variance .0038 . The innovation series $Z^{\nu}(t)$ of $Z(\cdot)$ is found by finding the innovation series $\tilde{Z}^{\nu}(\cdot)$ of $\tilde{Z}(\cdot)$ by fitting to $\tilde{Z}(\cdot)$ an autoregressive scheme whose order we now denote by $\hat{\tilde{m}}$. One obtains $\hat{\tilde{m}} = 13$ (using CAT) with residual variance .00127 (comparable to .00134 obtained by Box and Jenkins). Using stepwise autoregression on $\tilde{Z}$ one discovers lags 1 , 12 , 13 have significant coefficients; we therefore form the residuals (with residual variance .0015)

$$\tilde{Z}^{\nu}(t) = \tilde{Z}(t) - .74\,\tilde{Z}(t-1) + .38\,\tilde{Z}(t-12) - .31\,\tilde{Z}(t-13)$$

$$\doteq (I - .74\,L)(I + .38\,L^{12})\,\tilde{Z}(t)$$

The model we obtain by fitting a parsimonious autoregressive scheme to twelfth differences is

$$(I - .74\,L)(I + .38\,L^{12})(I - L^{12})\,Z(t) = \epsilon(t) \quad .$$

It can be written approximately

$$\frac{I - L}{I - .26\,L}\;\frac{I - L^{12}}{I - .38\,L}\;Z(t) = \epsilon(t)$$

which is similar to the model found by Box and Jenkins. We prefer to write our model

$$D_1(.26)\,D_{12}(.38)\,Z(t) = \epsilon(t) \quad .$$

We have asserted in Section 3 a theorem that the ideal whitening filter transforming a non-stationary time series $Y(\cdot)$ to its innovations $Y^{\nu}(\cdot)$ is unique but its decompositions for purposes of interpretation are not unique. In practice, one may appear to be able to identify several "different" approximate whitening filters; additional experience in case studies is required to determine how to measure differences and equivalences of whitening filters.

References

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden Day: San Francisco.

Dempster, A. P., Schatzoff, M., and Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares, Jour. Amer. Stat. Assoc., 72 (March, 1977). With discussion.

Hoaglin, D. and Welch, P. (1976). The Hat Matrix in Regression and ANOVA NBER Cambridge Computer Center Technical Report 341.

Makridakis, S. (1976). A Survey of Time Series, International Statistical Review 44, 29-70.

McClave, J. (1975). Subset Autoregression, Technometrics, 17, 213-220.

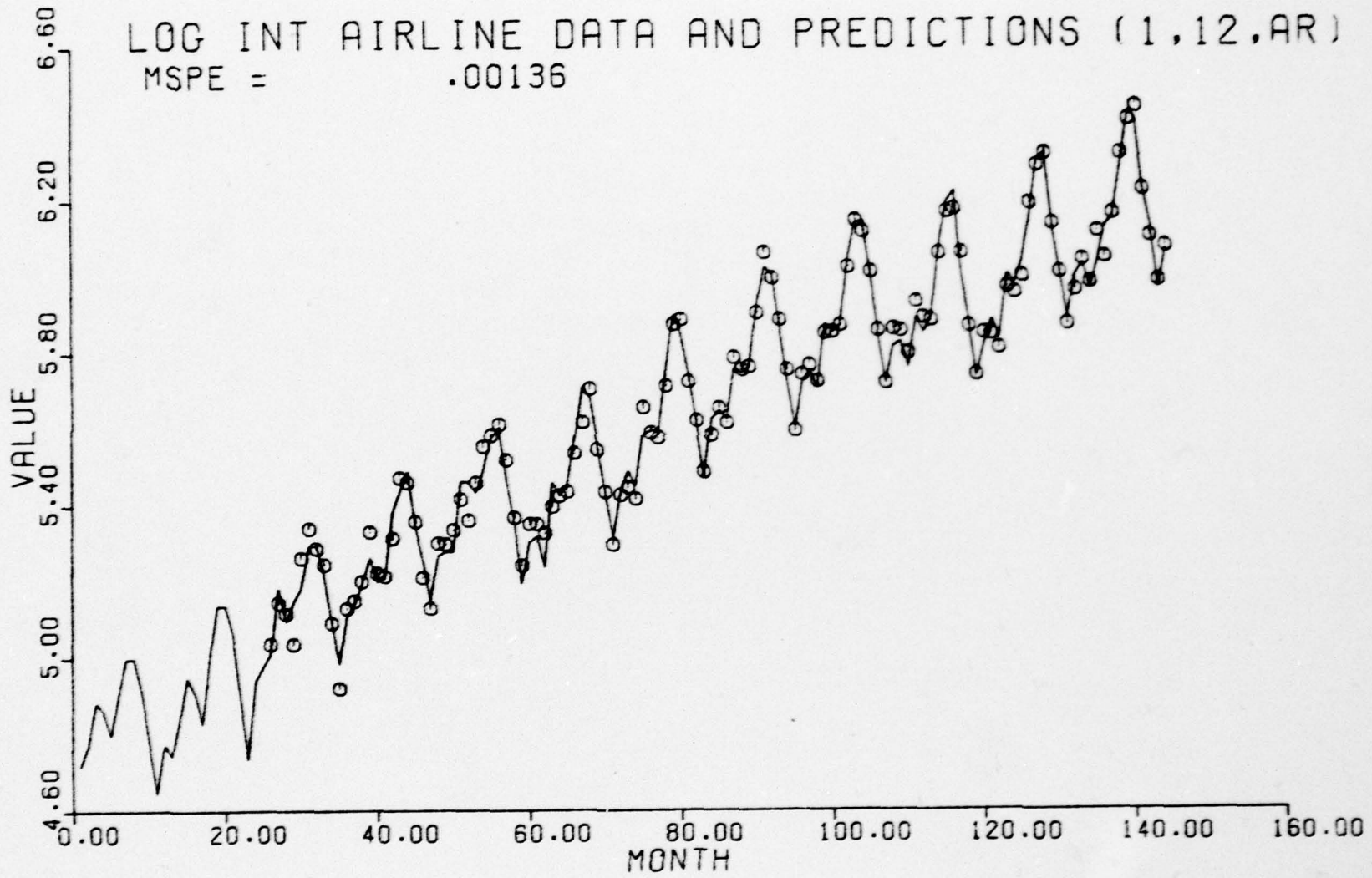
Parzen, E. (1974). Some Recent Advances in Time Series Modeling, IEEE Transactions on Automatic Control, AC-19, 723-730.

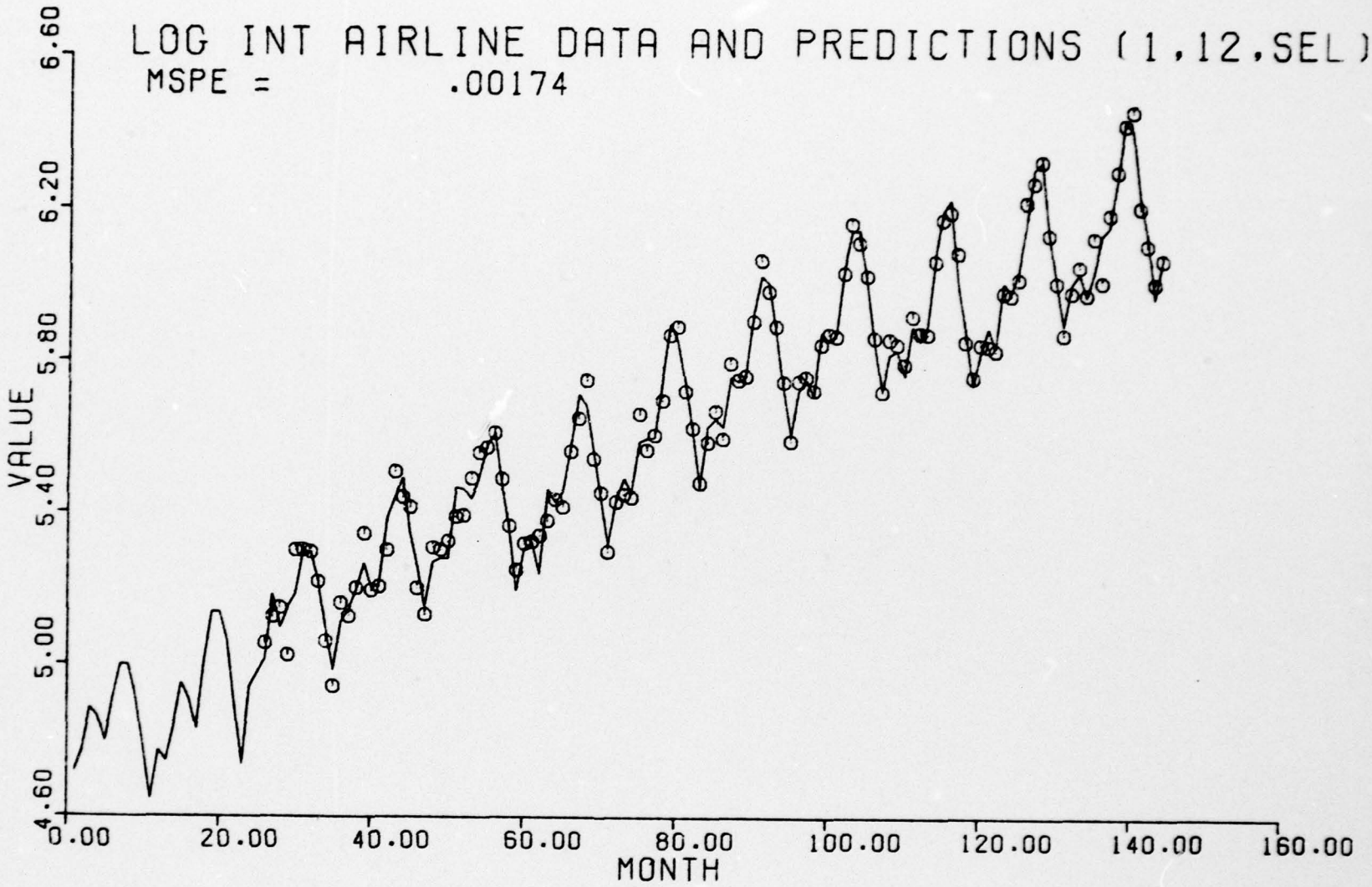
Parzen, E. (1975). Some Solutions to the Time Series Modeling and Prediction Problem, The Search for Oil, edited by D. Owens. Marcel Dekker: New York, 1-16.

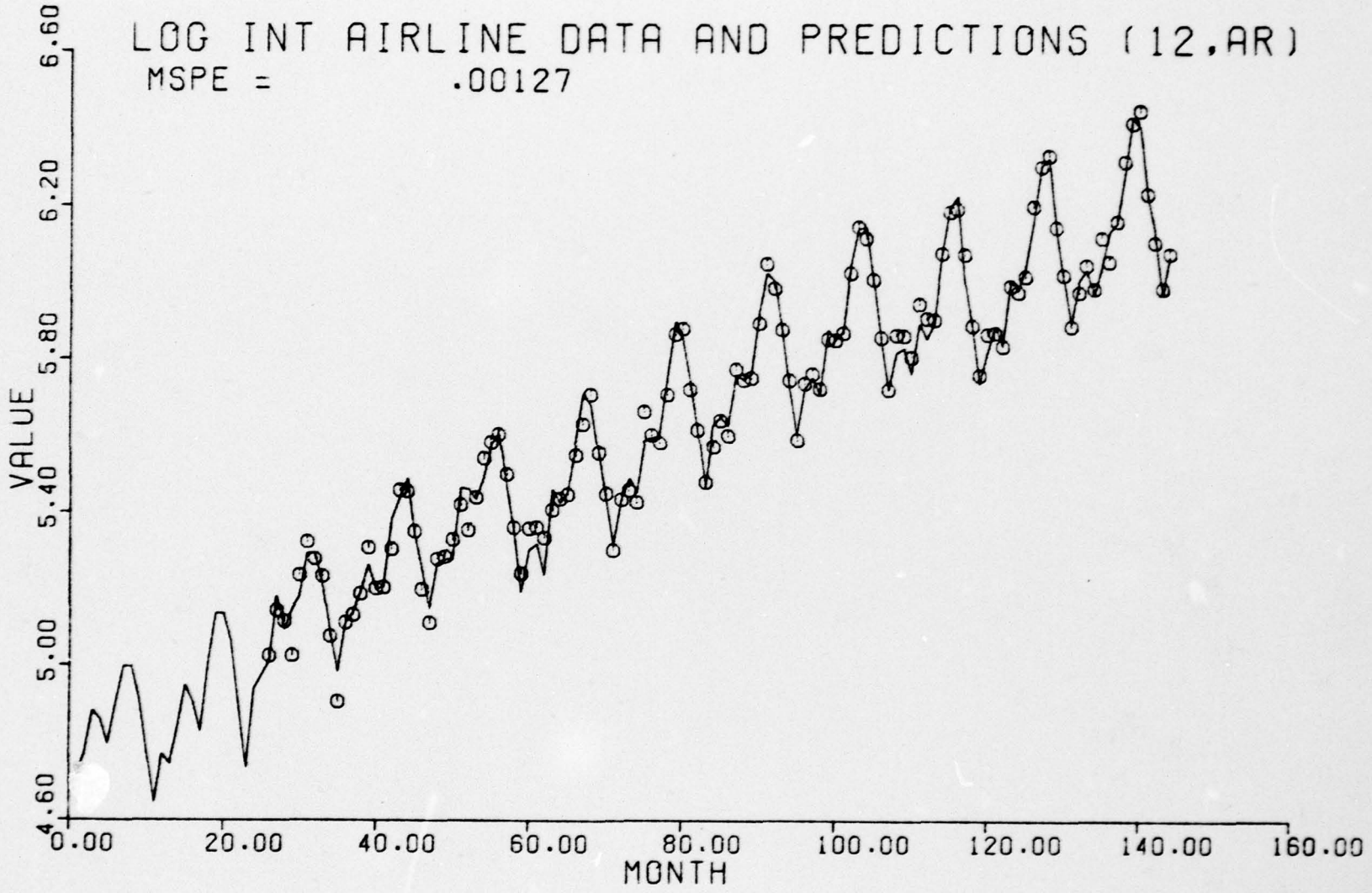
Parzen, E. (1976). An Approach to Time Series Modeling and Forecasting Illustrated by Hourly Electricity Demands, Statistical Science Division, SUNY at Buffalo, Technical Report.

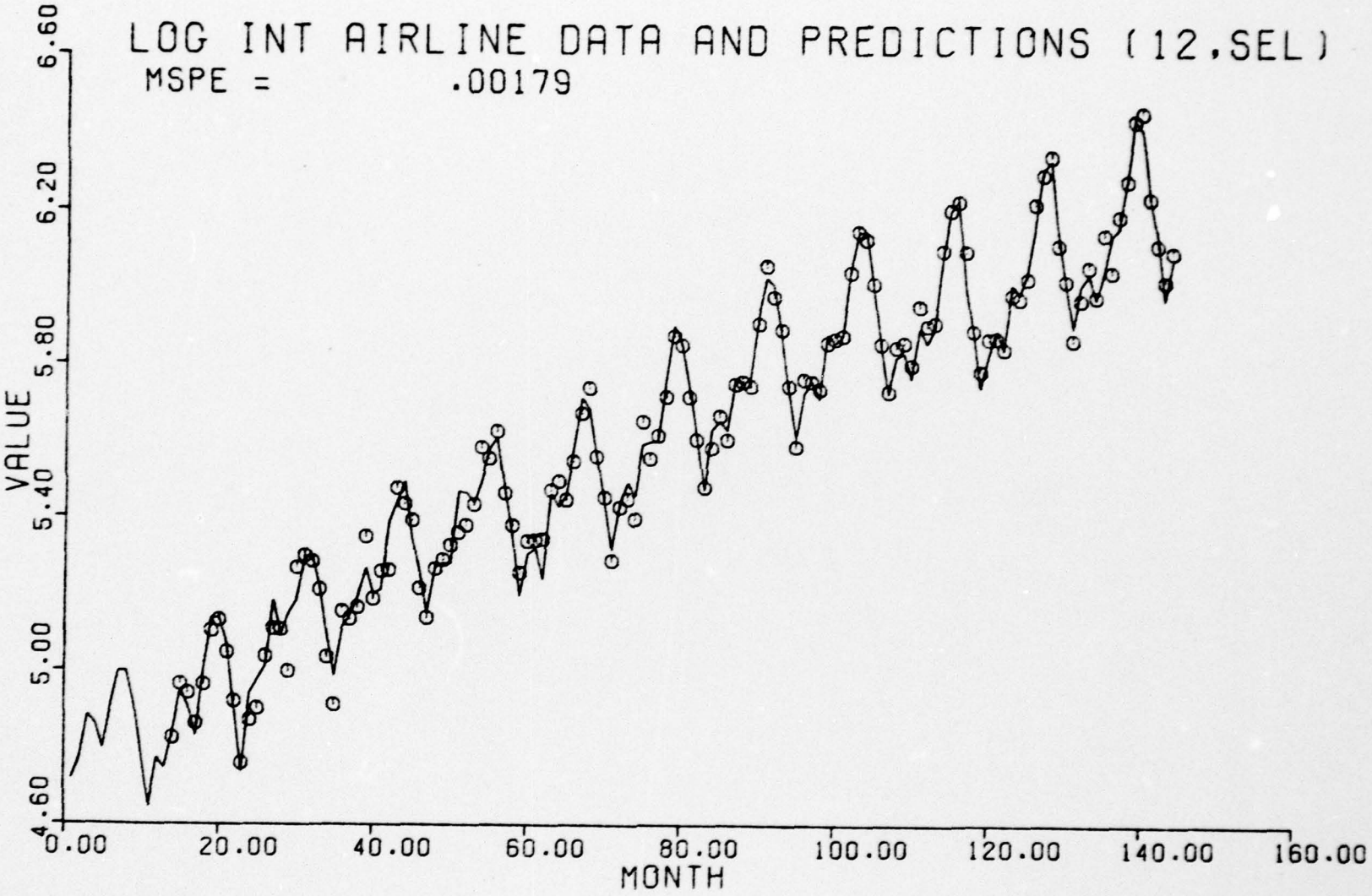
Parzen, E. (1977). Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes, Multivariate Analysis - IV, edited by P. Krishnaiah, Academic Press: New York.

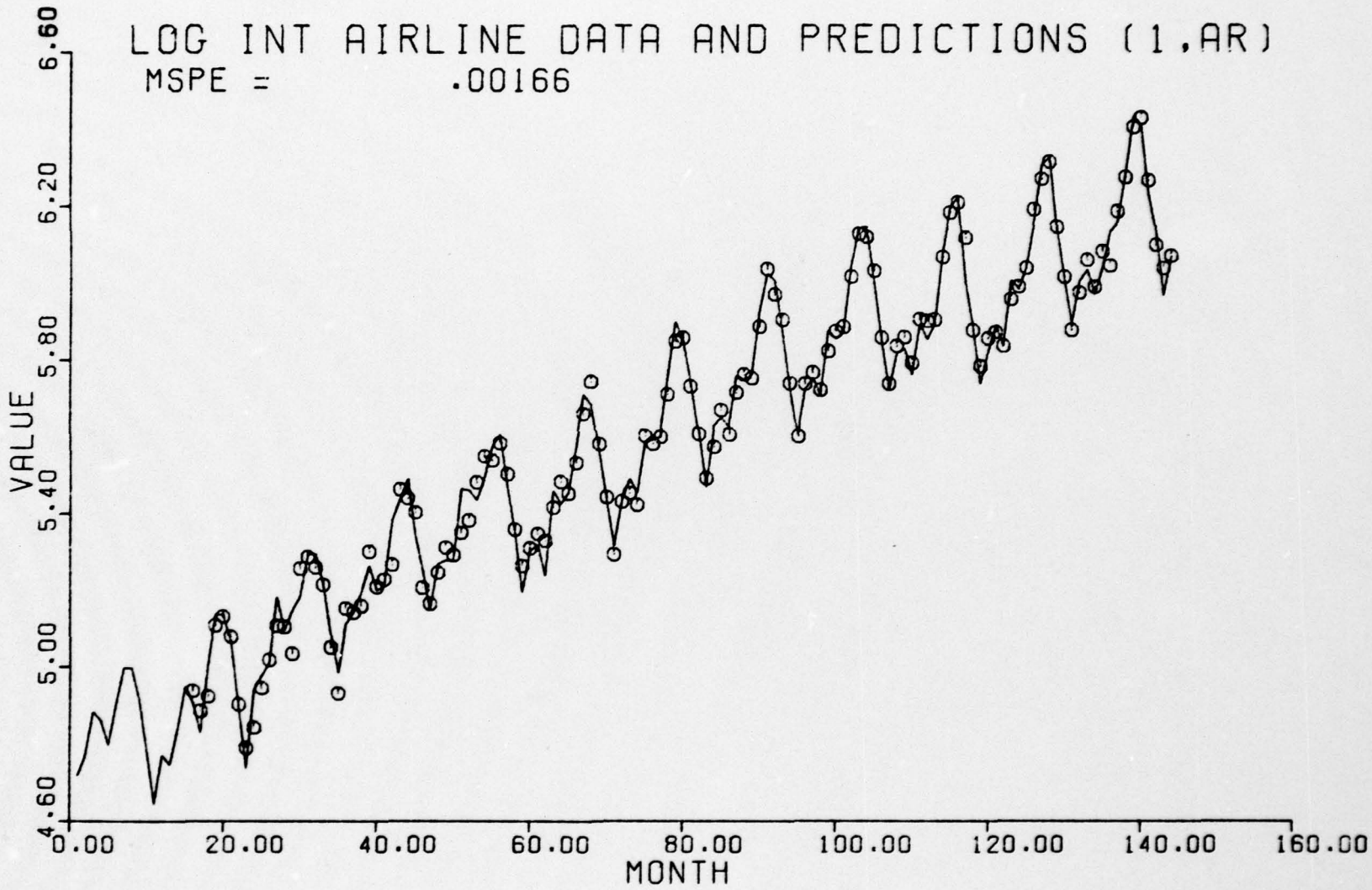
Wahba, G. (1976). A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them, Symposium on Applications of Statistics. To be edited by P. Krishnaiah.

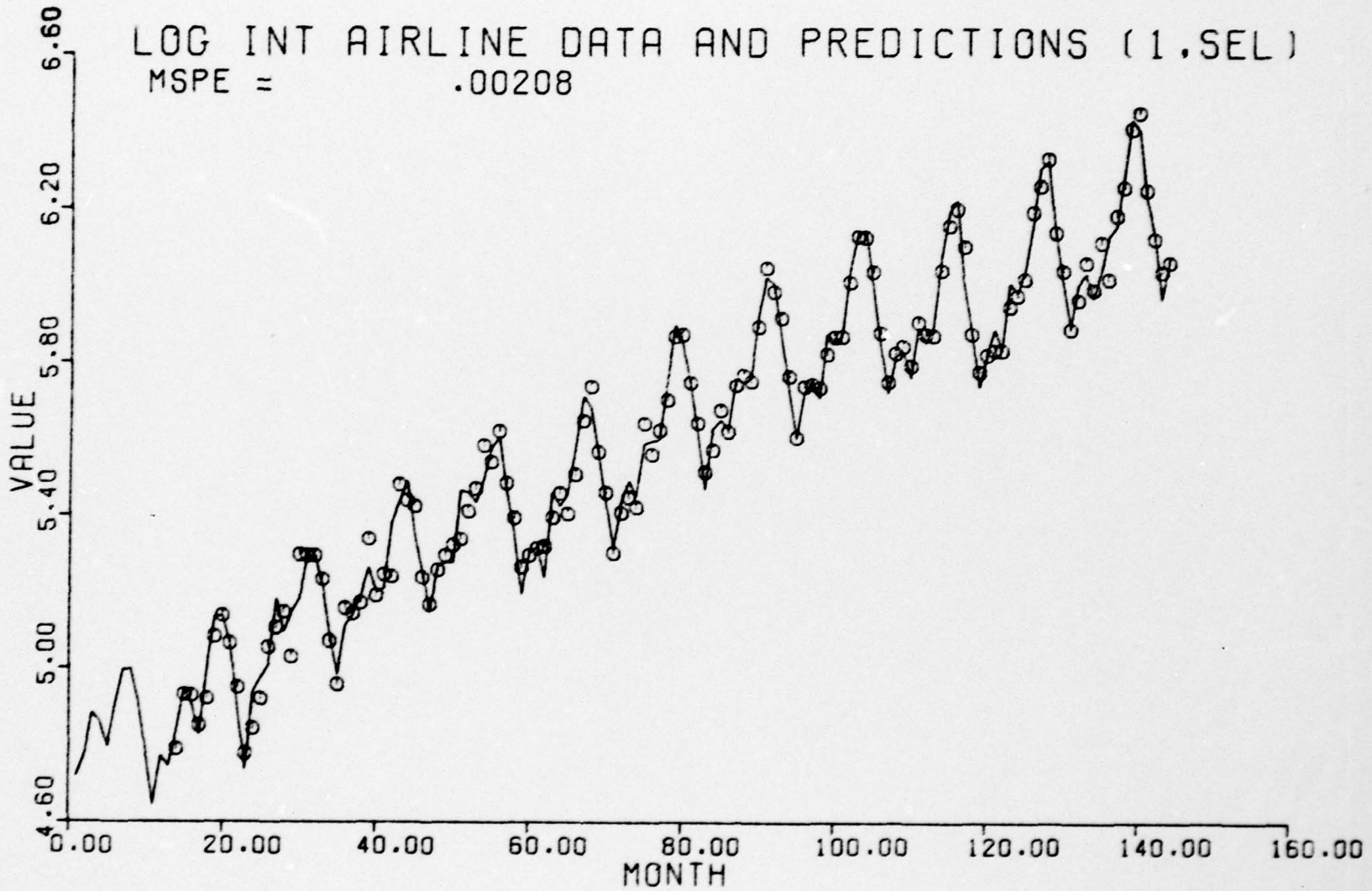

LOG INT AIRLINE DATA AND PREDICTIONS (1,12,AR)
MSPE = .00136
VALUE
6.60
6.20
5.80
5.40
5.00
4.60
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
MONTH

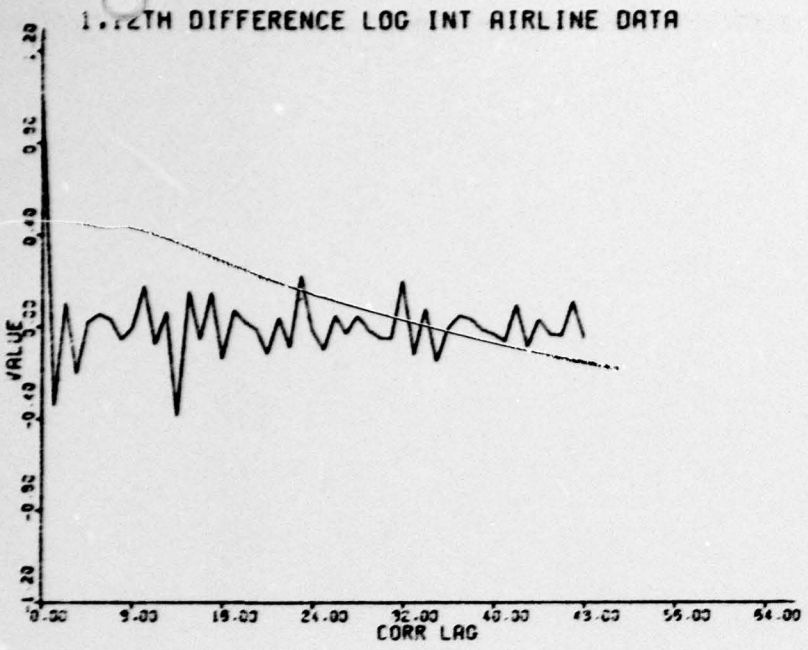

LOG INT AIRLINE DATA AND PREDICTIONS (1,12,SEL)
MSPE = .00174
VALUE
6.60
6.20
5.80
5.40
5.00
4.60
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
MONTH

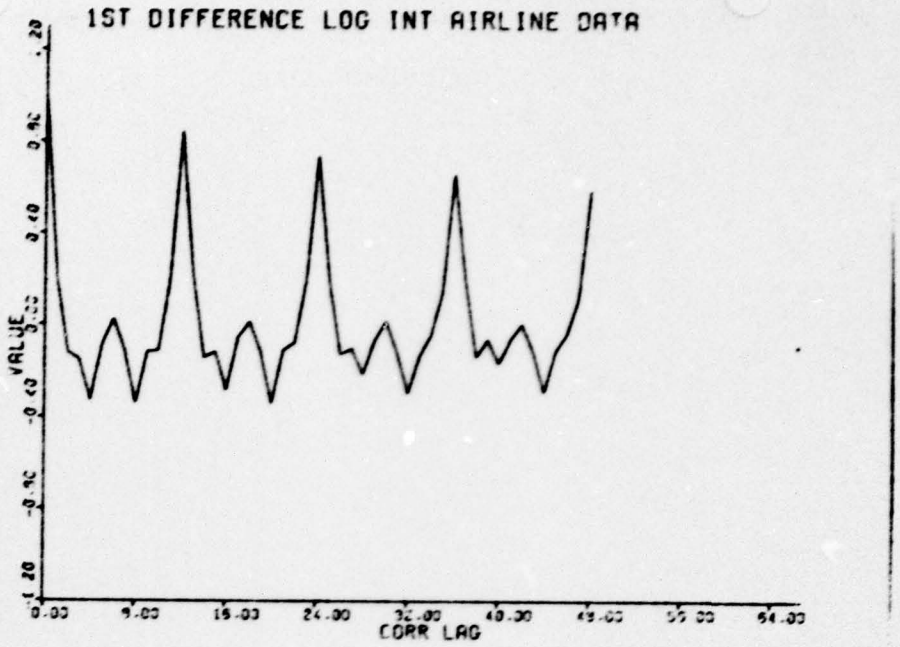

LOG INT AIRLINE DATA AND PREDICTIONS (12,AR)
MSPE = .00127
VALUE
6.60
6.20
5.80
5.40
5.00
4.60
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
MONTH

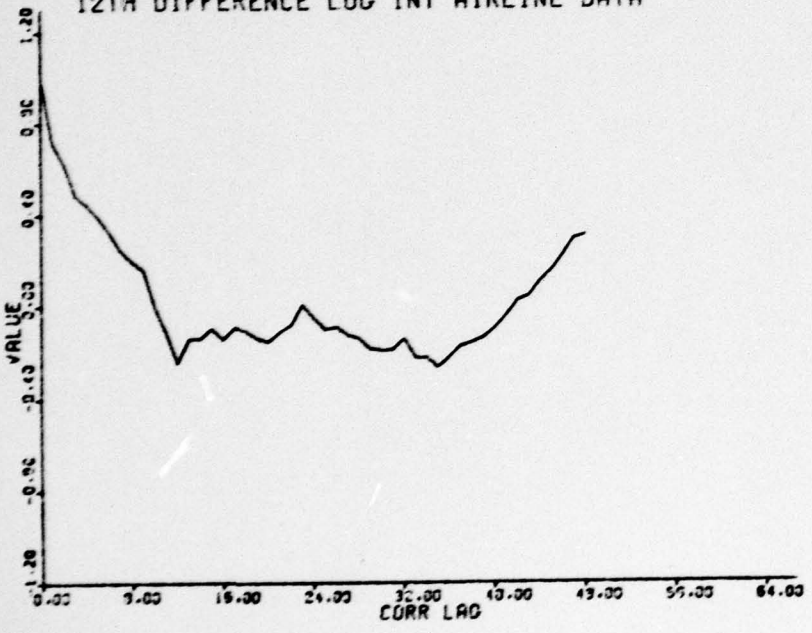

LOG INT AIRLINE DATA AND PREDICTIONS (12,SEL)
MSPE = .00179
VALUE
6.60
6.20
5.80
5.40
5.00
4.60
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
MONTH

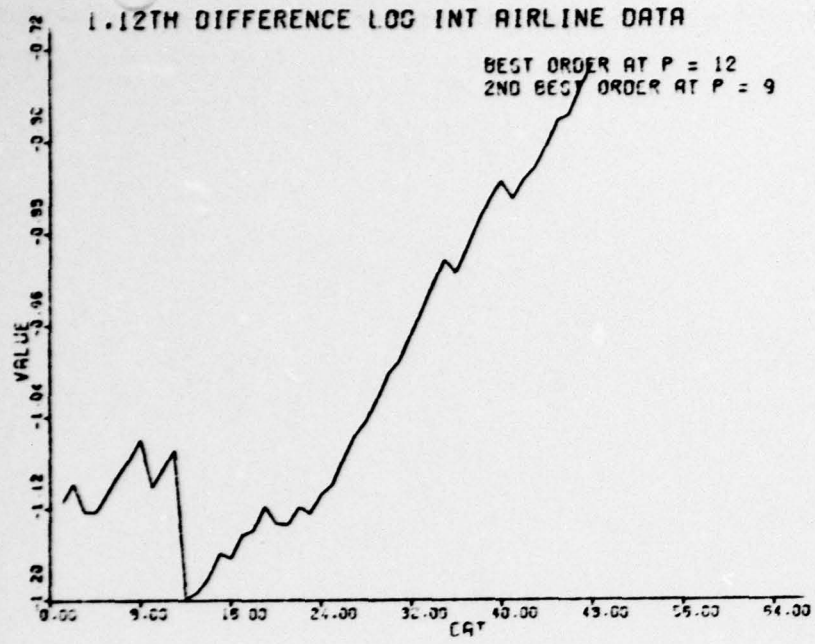

LOG INT AIRLINE DATA AND PREDICTIONS (1,AR)
MSPE = .00166
VALUE
6.60
6.20
5.80
5.40
5.00
4.60
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
MONTH

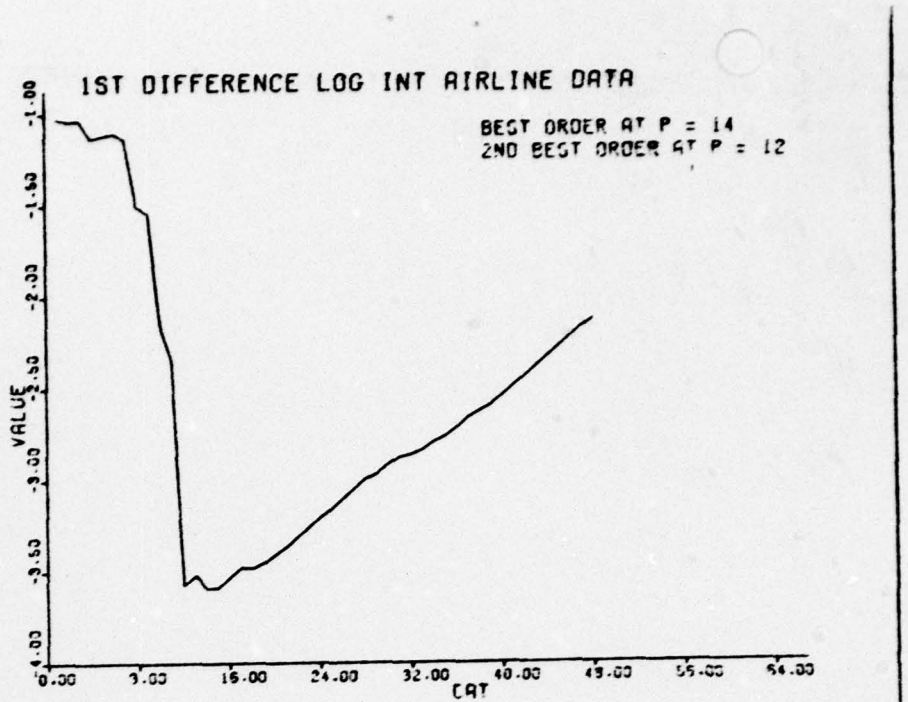

LOG INT AIRLINE DATA AND PREDICTIONS (1,SEL)
MSPE = .00208
VALUE
6.60
6.20
5.80
5.40
5.00
4.60
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
MONTH

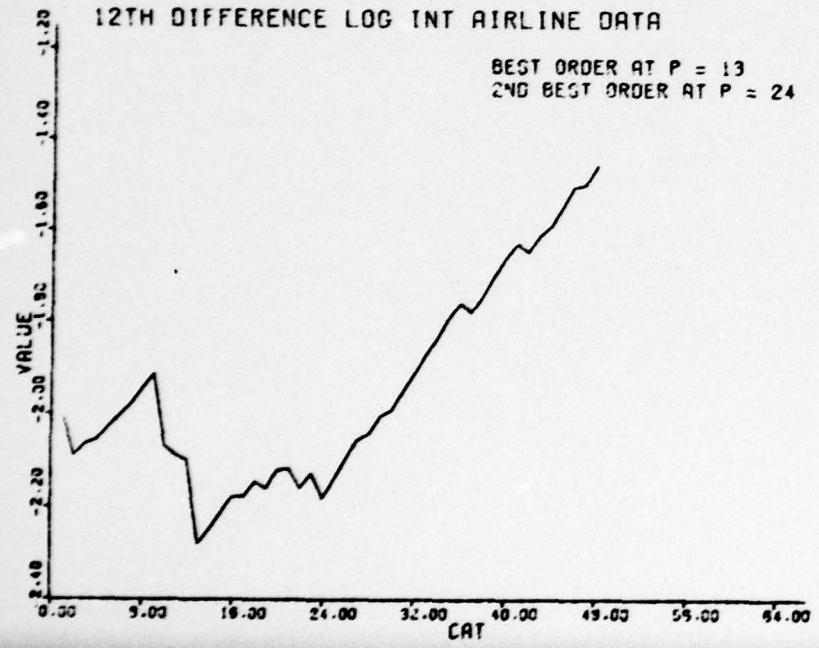

1.12TH DIFFERENCE LOG INT AIRLINE DATA
VALUE
CORR LAG

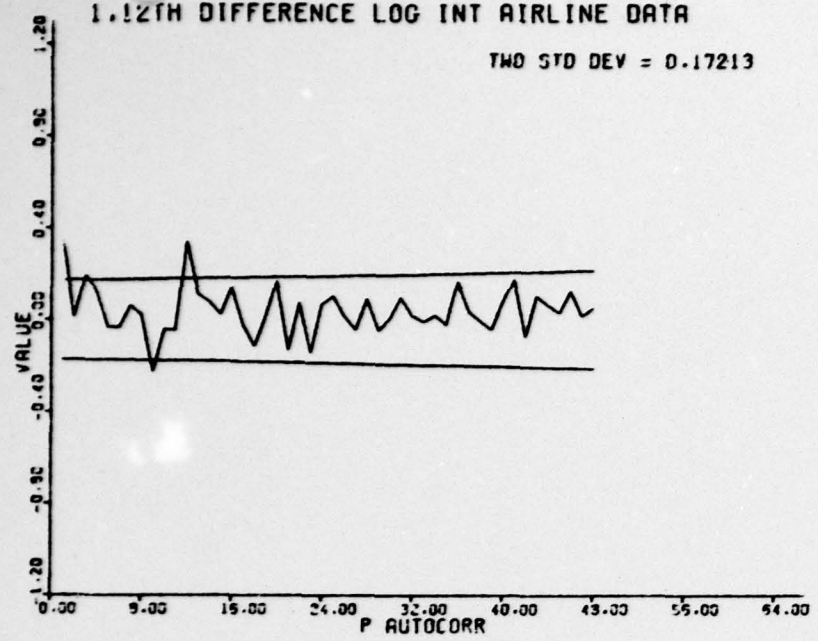

1ST DIFFERENCE LOG INT AIRLINE DATA
VALUE
CORR LAG

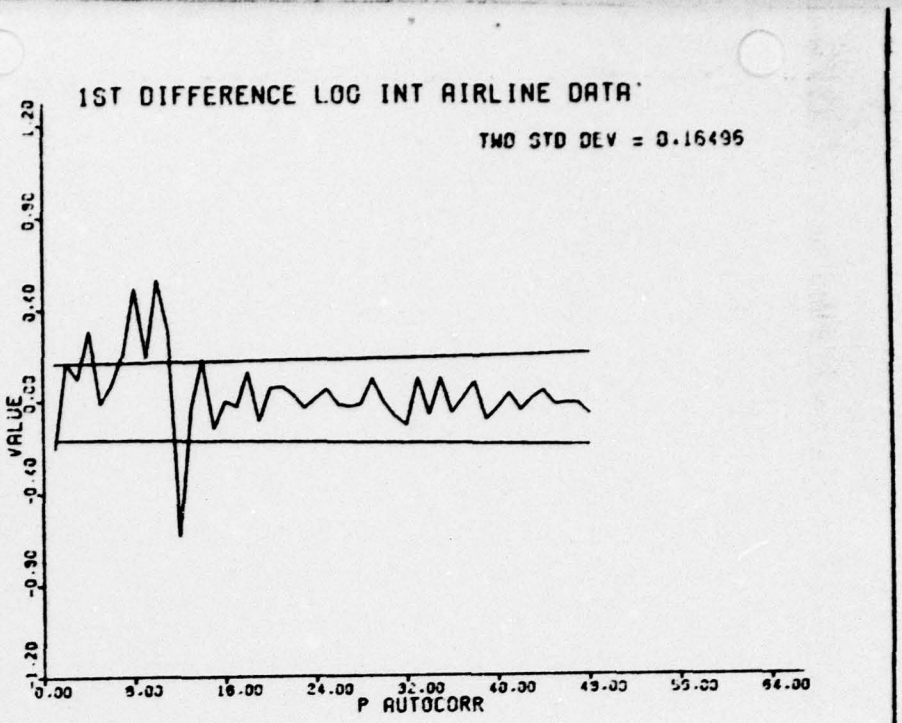

12TH DIFFERENCE LOG INT AIRLINE DATA
VALUE
CORR LAG

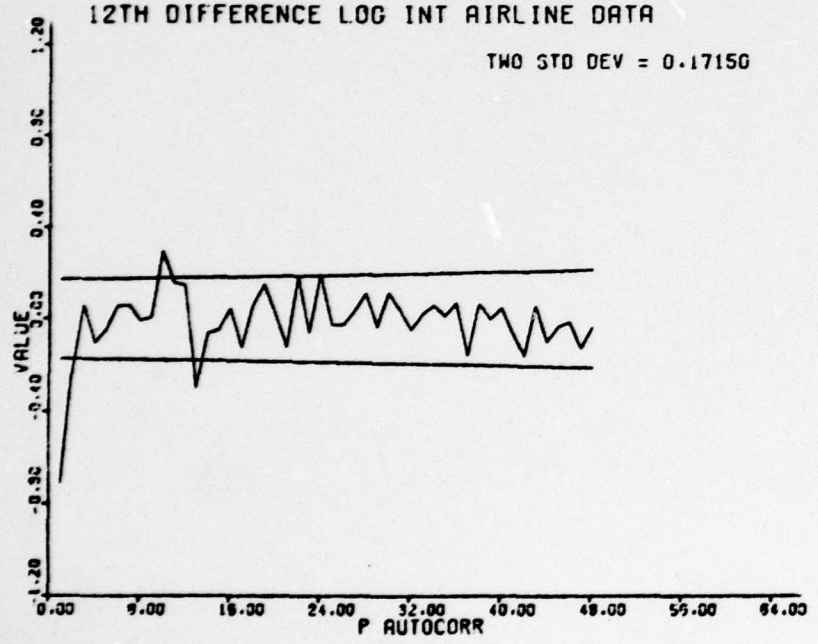

1.12TH DIFFERENCE LOG INT AIRLINE DATA
BEST ORDER AT P = 12
2ND BEST ORDER AT P = 9
VALUE
CAT
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00

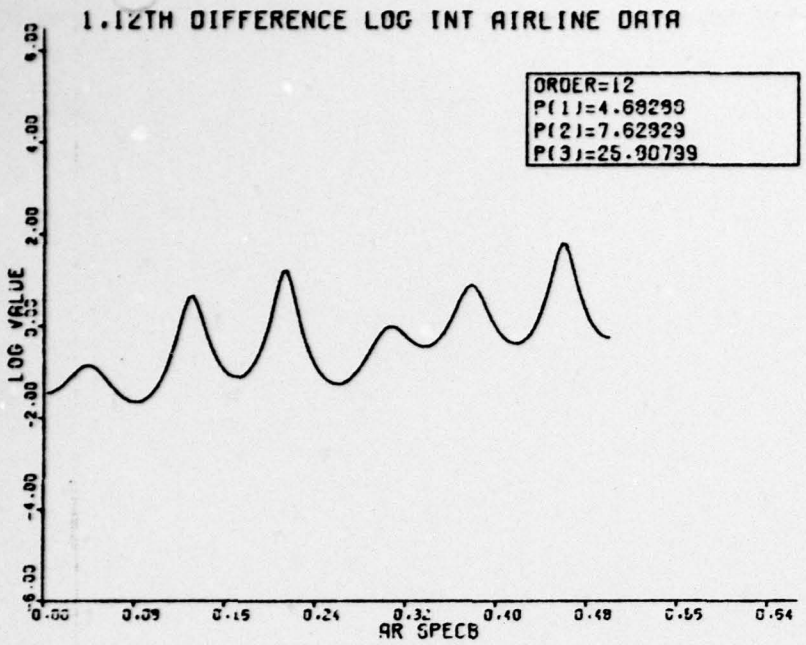

1ST DIFFERENCE LOG INT AIRLINE DATA
BEST ORDER AT P = 14
2ND BEST ORDER AT P = 12
VALUE
CAT
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00

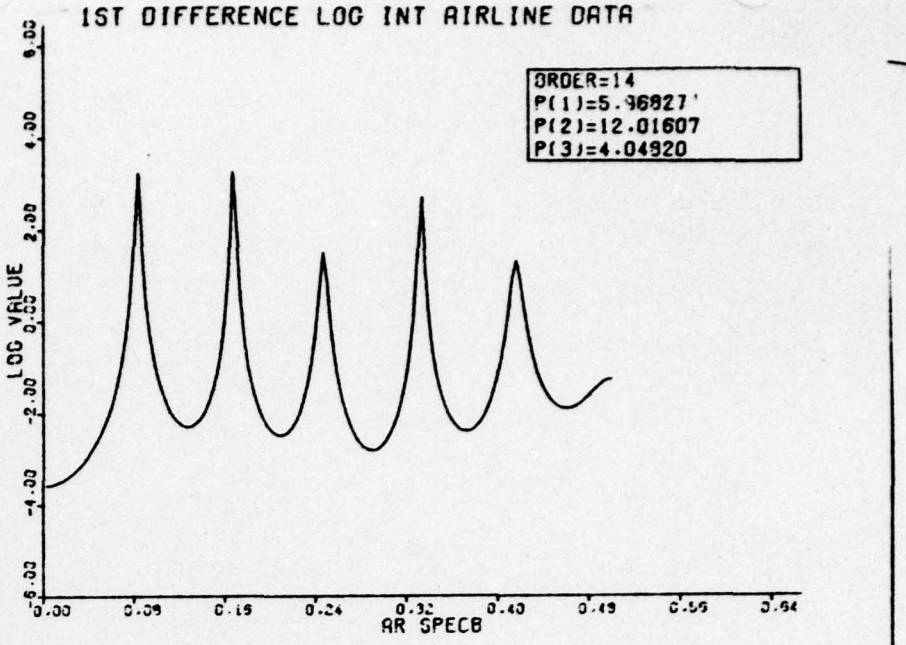

12TH DIFFERENCE LOG INT AIRLINE DATA
BEST ORDER AT P = 13
2ND BEST ORDER AT P = 24
VALUE
CAT
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00

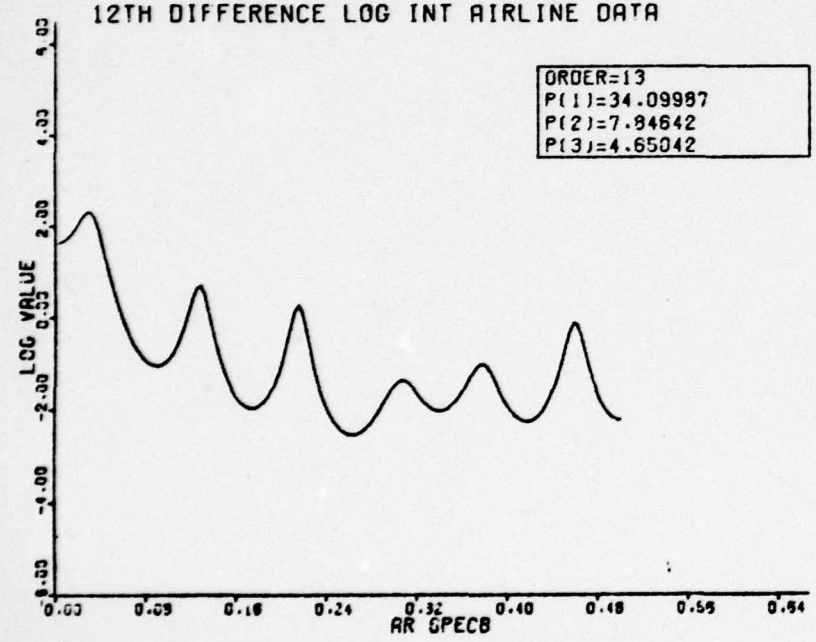

1,12TH DIFFERENCE LOG INT AIRLINE DATA
TWO STD DEV = 0.17213
VALUE
1.20
0.80
0.40
0.00
-0.40
-0.80
-1.20
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00
P AUTOCORR


1ST DIFFERENCE LOG INT AIRLINE DATA
TWO STD DEV = 0.16496
VALUE
1.20
0.80
0.40
0.00
-0.40
-0.80
-1.20
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00
P AUTOCORR


12TH DIFFERENCE LOG INT AIRLINE DATA
TWO STD DEV = 0.17150
VALUE
1.20
0.80
0.40
0.00
-0.40
-0.80
-1.20
0.00
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00
P AUTOCORR


1.12TH DIFFERENCE LOG INT AIRLINE DATA
ORDER=12
P(1)=4.68299
P(2)=7.62929
P(3)=25.80799
LOG VALUE
AR SPECB


1ST DIFFERENCE LOG INT AIRLINE DATA
ORDER=14
P(1)=5.96827
P(2)=12.01607
P(3)=4.04920
LOG VALUE
AR SPECB


12TH DIFFERENCE LOG INT AIRLINE DATA
ORDER=13
P(1)=34.09987
P(2)=7.84642
P(3)=4.65042
LOG VALUE
AR SPECB