An Exponential Extrapolator*

R. J. Duffin

Department of Mathematics, Carnegie Institute of Technology
Pittsburgh 13, Pennsylvania

AND

Phillips Whidden

Aluminum Company of America, Pittsburgh 10, Pennsylvania

A linear combination of \( m \) exponential functions (exponential) is fitted to a time series, such as daily observations. The fitting is carried out over all past time by weighted least squares with an exponential weight factor. The resulting minimizing function could be continued into the future. In particular tomorrow's predicted value is defined by this continuation. To obtain an explicit solution of the problem a formula is constructed which gives the predicted value as a linear combination of the last \( m \) observed values and the last \( m \) predicted values. The \( 2m \) coefficients of this formula are expressed as explicit rational functions of the \( m \) exponential bases. The extrapolating functions available with this method include polynomials, trigonometric polynomials, damped waves, etc. The particular class of extrapolating functions to be used for a given problem depends on the genesis of the data.

I. Introduction

This note describes a method of extrapolation of data by fitting with functions of the form

\[
\hat{p}(x) = \sum_{1}^{m} a_i \beta_i^x
\]  

(1)

where the $\beta_j$ are fixed complex numbers, assumed distinct and nonzero. Such exponential polynomials are here termed \textit{exponomials}.

The data are supposed to be observations given at a regular sequence of values of $x$, say $x = 1, 2, 3, \ldots$. The observations for these values of $x$ form a sequence of complex numbers denoted by $y_1, y_2, y_3, \ldots$. Then the central problem here is the extrapolation of this sequence of observations to obtain a predicted value for $x = 0$. This predicted value is denoted by $y_0^*$ and is defined as

$$y_0^* = \sum_{1}^{m} a_j.$$  \hspace{1cm} (2)

The method of discounted least squares is used to fit the exponomial to the data. This means that the coefficients $a_1, \ldots, a_m$ are chosen to minimize the expression

$$E = \sum_{1}^{\infty} \theta^n |y_n - \hat{p}(n)|^2.$$  \hspace{1cm} (3)

Here $\theta$ is a positive constant termed the discount factor. It is assumed to satisfy the following condition

$$\theta |\beta_j| < 1; \hspace{0.5cm} j = 1, 2, \ldots, m.$$  \hspace{1cm} (4)

In a previous paper Duffin and Schmidt [1] treated extrapolation by discounted least squares but they used ordinary polynomials,

$$\hat{p}(x) = \sum_{0}^{m-1} a_j x^j$$  \hspace{1cm} (5)

instead of exponomials. The present treatment is aimed at obtaining the same type of results for exponomials. Actually the previous results can be obtained as a limiting case of this treatment. To understand this suppose that $\beta_j = e^{ij}$ for $j = 1, \ldots, m$. The Taylor series of the $\beta_j x$ show that if $\varepsilon$ is small and $x$ is on a finite interval then the exponomials (1) uniformly approximate the polynomials (5).

When the minimization of (3) is carried out relation (2) leads to the so called "long formula",

$$y_0^* = \sum_{1}^{\infty} Q_n y_n.$$  \hspace{1cm} (6)
Here the coefficients $Q_n$ do not depend on the sequence $y_n$. Moreover it is found that these coefficients decrease exponentially so (6) is a rapidly converging series. Thus in practice one may terminate the long formula at a finite value of $n$.

Let $y^*$ be defined as

$$y^*_x = \sum_{n=1}^{\infty} Q_n y_{n+x}.$$  \hfill (7)

Then $y^*_x$ is the predicted value of $y_x$ based on the “previous values” $y_{x+1}, y_{x+2}, \ldots$. The discrepancy $\delta_x$ is defined as

$$\delta_x = y^*_x - y_x.$$  \hfill (8)

In the paper on extrapolation with polynomials it was found that there is an identity giving $y^*_x$ as a linear combination of the $m$ previous observations and the $m$ previous discrepancies. A similar “short formula” is here found for exponentials. Thus

$$y^*_x = - \sum_{k=1}^{m} g_k y_{k+x} - \sum_{k=1}^{m} f_k \theta^k \delta_{k+x}.$$  \hfill (9)

Here the $g_k$ and $f_k$ are constants which do not depend on the sequence $y_n$ or the discount factor $\theta$. The short formula has several advantages over the long formula; this is discussed in [1].

The developments to follow in this note give explicit evaluation of the coefficients $Q_n$ of the long formula and the coefficients $g_k$ and $f_k$ of the short formula. They are expressed as rational function of $\beta_1, \ldots, \beta_m$.

II. Discounted Least Squares

The sequences $y_n$ to be considered are assumed to give $E$ defined by (3) a finite value. Let $y_n$ and $v_n$ be two such sequences. Then a bilinear form, associated with the quadratic form $E$, is denoted by $[y, v]$ and is defined as

$$[y, v] = \sum_{1}^{\infty} \theta^n y_n v_n.$$  

Then (3) may be written as

$$E(y - \phi) = [\bar{y} - \bar{\phi}, y - \phi]$$

where the bar denotes the complex conjugate.
If the coefficients \( a_i \) of \( p(x) \) are chosen to minimize \( E(y - p) \) then it is seen that \( y - p \) is orthogonal to an arbitrary exponomial \( P \) of the class (1). So

\[
[P, y - p] = 0. \tag{10}
\]

This leads to a system of \( m \) linear equations for the \( m \) coefficients \( a_i \). The solution of this system involves inverting the matrix \( \beta_{ij} \). Instead of pursuing this direct approach we shall now follow a synthetic operational method.

### III. An Operational Method

Let \( X \) denote the translation operator such that if \( f \) is a function of \( x \) then \( Xf(x) = f(x + 1) \). Then an operator \( G \) is defined as

\[
G(X) = \prod_{1}^{m} (1 - \beta_{ij}^{-1} X). \tag{11}
\]

It is clearly seen that \( G \) is the anihilator of exponomials of the form (1); thus

\[
G(X)p(x) = 0. \tag{12}
\]

An operator \( F \), related to \( G \), is defined as

\[
F(X) = \prod_{1}^{m} (1 - \alpha_{ij} X) \tag{13}
\]

where \( \alpha_{ij} \) denotes \( \beta_{ij} \). Finally an operator \( Q \) is defined as

\[
Q(X) = 1 - \frac{G(X)}{F(\theta X)}. \tag{14}
\]

To interpret \( Q \) the right side is expanded in a power series\(^1\) in \( X \). Thus

\[
Q(X) = \sum_{1}^{\infty} Q_n X^n. \tag{15}
\]

\(^1\) The manipulations performed in this paper on operators expressed as power series are easily justified. The proof depends on the fact that the operands are such that the resulting series are absolutely convergent.
It is now to be shown that the coefficients $Q_n$ in (15) are actually the same as the coefficients in the long formula. Then relation (7) can be written as

$$ y_x^* = Q y_x. \quad (16) $$

To prove relation (16) we first write

$$ X^{-1} Q = \frac{X^{-1}(F - G)}{F}. $$

The denominator of the fraction on the right is a polynomial of degree $m$ in $X$. The numerator is a polynomial of degree $m - 1$. Thus the Lagrange interpolation formula may be applied to give

$$ Q = X \sum_{1}^{m} \frac{-G(r_j)r_j^{-1}}{\theta F'(\theta r_j)(X - r_j)} $$

where $r_1, \ldots, r_m$ are the roots of $F(\theta X)$. Expanding the denominators by the binomial theorem gives

$$ Q = \sum_{1}^{m} \sum_{1}^{\infty} \frac{G(r_j)}{\theta r_j F'(\theta r_j)} \left(\frac{X}{r_j}\right)^n $$

Of course $r_j = \theta^{-1} \alpha_j^{-1}$ so

$$ Q = \sum_{1}^{m} \left( \sum_{1}^{\infty} \frac{\alpha_j G(\theta^{-1} \alpha_j^{-1})}{F'(\alpha_j^{-1})} \alpha_j^n \right) \theta^n X^n. \quad (17) $$

This shows that $Q_n = q(n) \theta^n$ where

$$ q(x) = \sum_{1}^{m} \frac{a_j G(\theta^{-1} \alpha_j^{-1})}{F'(\alpha_j^{-1})} \alpha_j^{x}. \quad (18) $$

Thus $q$ is an exponential of form (1).

To continue the proof it is seen from (12) that

$$ Q \hat{p} = \hat{p} - F^{-1} G \hat{p} = \hat{p} $$

for any exponential $\hat{p}$. Then making use of (17) with $x = 0$ gives

$$ \hat{p}(0) = \sum_{1}^{\infty} \theta^n q(n) \hat{p}(n) = [q, \hat{p}]. \quad (19) $$

In particular (19) holds for a polynomial $\hat{p}$ which minimizes $E(y - \hat{p})$. 

Next the orthogonality relation (10) is applied with $P = \tilde{q}$ giving 

$$[q, \tilde{p}] = [q, y].$$

This relation together with (19) gives 

$$\tilde{p}(0) = [q, y].$$  \hfill (20)

Since $\gamma_0^*$ is defined to be $\tilde{p}(0)$ this proves the long formula (6). Consequently the coefficients of the long formula are explicitly given by (18).

**IV. The Short Formula**

Carrying out the indicated multiplication in the formula (11) for $G$ gives 

$$G(X) = \sum_{0}^{m} g_k X^k$$  \hfill (21)

where the $g_k$ are constants. Likewise formula (13) for $F$ becomes 

$$F(X) = \sum_{0}^{m} f_k X^k.$$  \hfill (22)

It is seen from (11) and (13) that 

$$F(X) = X^m \frac{G(X^{-1})/g_m.} \hfill (23)$$

Thus the coefficients $f_k$ and $g_k$ have the following relationship 

$$f_k = g_{m-k}/g_m.$$  \hfill (24)

Multiplying formula (14) by $F(\theta X)$ gives 

$$F(\theta X)Q(X) = F(\theta X) - G(X).$$

The expressions (21) and (22) are substituted in this relation. Since $g_0 = f_0 = 1$ this gives 

$$Q = \sum_{1}^{m} g_k X^k - \sum_{1}^{m} f_k \theta^k (Q - 1) X^k.$$ 

Operating on the sequence $\gamma_x$ proves that 

$$\gamma_x^* = -\sum_{1}^{m} g_k \gamma_{x+k} - \sum_{1}^{m} f_k \theta^k \delta_{x+k}. $$
This is the short formula (9). A further simplification results by using (24) to eliminate \( f_k \). Thus

\[
y_x^* = - \sum_{1}^{m} g_k y_{k+x} - \bar{g}_m^{-1} \sum_{1}^{m} \bar{g}_{m-k} \bar{g}_k \delta_{k+x}.
\]  

(25)

Here the \( g_k \) are given by

\[
g_1 = - \sum_{1}^{m} \beta_k^{-1},
\]

\[
g_2 = \sum_{1}^{m} \sum_{1}^{m} \beta_j^{-1} \beta_k^{-1} \quad j < k
\]

\[
\vdots
\]

\[
g_m = (-1)^m \beta_1^{-1} \beta_2^{-1} \ldots \beta_m^{-1}.
\]

Formula (25) is the central result of this note.

V. The Standard Deviation of the Extrapolation

Suppose that the observation \( y_n \) have independent random errors with the same variance. Then it follows from the long formula (6) that the variance in \( y_0^* \) is reduced by a factor \( \Sigma Q_n^2 \). In this connection it is of interest to evaluate the more general series

\[
S = \sum_{1}^{\infty} |Q_n|^2 c^n.
\]  

(26)

Here \( c \) is a positive constant such that \( c \beta^2 \beta < 1 \) where \( \beta = \max |\beta_j| \). This inequality insures that series (26) converges.

The formal relation (14) may be converted to a numerical identity by replacing the operator \( X \) by a complex variable \( z \). Thus

\[
\sum_{0}^{\infty} Q_n z^n = - \frac{G(z)}{F(\theta z)}
\]  

(27)

As a notational convenience \( Q_0 \) is here defined to be \(-1\). Likewise by replacing \( X \) by \( c/z \) in the complex conjugate of (14) gives

\[
\sum_{0}^{\infty} Q_k \frac{c^k}{z^k} = - \frac{\bar{G}(c/z)}{\bar{F}(\theta c/z)}.
\]  

(28)
Both series (25) and (26) will converge if
\[ c \theta \beta < |z| < \theta^{-1} \beta^{-1}. \]  
(29)

If \( z \) satisfies this inequality then (27) and (28) may be multiplied to yield a Laurent series \( L(z) \). Thus
\[
L = \sum_0^\infty \sum_0^\infty Q_n \bar{Q}_k c^k \bar{c}^{k-n} - \frac{G(z)}{F(\theta z)} \frac{\bar{G}(\bar{c}/z)}{\bar{F}(\bar{c} z)}. \]

Now relation (23) is used to convert numerator and denominator here to polynomials in ascending powers of \( z \). Thus
\[
L(z) = \frac{G(z)}{F(\theta z)} \frac{F(z/c)}{G(z/c)} \frac{g_m^2}{\theta^m}. \]

It is seen that the constant term of \( L(z) \) is precisely \( S + 1 \) where \( S \) is defined by (24). On the other hand the constant term of a Laurent series is given by the contour integral formula
\[
S + 1 = \frac{1}{2\pi i} \int \frac{L(z)}{z} dz. \]

Here the contour could be a circle of radius \( r \) with \( c \theta \beta < r < \theta^{-1} \beta^{-1} \). The contour integral may be evaluated by residues. The only singularities are simple poles where \( zG(z/c) = 0 \). Thus
\[
S + 1 = \frac{g_m^2}{\theta^m} + \frac{g_m^2 \theta c}{\theta^m} \sum_1^m \frac{G(c \theta \beta_i)}{F(c \theta^2 \beta_i)} \frac{F(\theta \beta_i)}{G'(\beta_i)}. \]  
(30)

If the errors in the observations are independent and have the same variance then the variance in the extrapolation is reduced by a factor \( S \) obtained by putting \( c = 1 \) in (30). Another case of interest is \( c = \theta^{-1} \) because then (30) simplifies to
\[
S = -1 + \theta^{-m} |g_m|^2. \]  
(31)

This corresponds to a situation in which the errors vary with "distance," obeying a law of the form
\[
\sigma^2(y_n) = K \theta^{-n}. \]  
(32)

Here \( K \) is a constant and \( \sigma(y_n) \) denotes the standard deviation of \( y_n \). Then since \( \sigma^2(y_n^*) = KS \) we obtain from (31) and (32)
\[
\sigma^2(y_n^*) = \sigma^2(y_m) (|g_m|^2 - \theta^m). \]  
(33)
In connection with a law of errors of the form (32) the following
hypothetical situation is not without interest. Suppose that:
(a) The $y_n$ are actually observations of an exponential of the form (1).
(b) The $y_n$ have independent Gaussian errors.
(c) The standard deviation varies with distance according to an
exponential law.

Under these hypotheses it may be shown that Fisher's principle of
maximum likelihood leads to the determination of the exponential by the
method of discounted least squares.

VI. RADIOACTIVE DECAY

An example of exponential extrapolation is furnished by radioactive
decay. Thus suppose that a mixture contains three radioactive elements
with half lives known to be $T_1$, $T_2$, and $T_3$ days. Let $\beta_1 = \exp\left(-1/T_1\right)$ etc. Thus the amount of the first element left after $x$ days is proportional
to $\beta_1^x$. Then a Geiger counter in the vicinity of the mixture would register
a counting rate
\[ p(x) = a_1 \beta_1^x + a_2 \beta_2^x + a_3 \beta_3^x. \]

The coefficients $a_1$, $a_2$, and $a_3$ depend on the initial amounts of the
elements present. It is supposed that these amounts are not given.

Consider the problem of estimating the counting rate at time $x = 0$
from the counting rates $y_1, y_2, \ldots$ observed on the 1, 2, \ldots days. This
problem is solved by setting up the long formula. It is possible to take
$\theta = 1$ because there is no convergence problem.

For certain measurement techniques it would be reasonable to assume
that the $y_n$ have independent Gaussian errors and all with the same
standard deviation $\sigma$. Then if $\theta = 1$ the determination of $y_0^*$ is in accord
with Fisher's principle of maximum likelihood. Moreover by formula (33)
\[ \sigma^2(y_0^*) = (\beta_1^{-2} \beta_2^{-2} \beta_3^{-2} - 1)\sigma^2. \]

It is seen from this formula that "smoothing" takes place if and only
if the $\beta_j$ are close to unity.

To write the short formula we expand $G(X) = (1 - X/\beta_1)(1 - X/\beta_2)$
$(1 - X/\beta_3)$. Then substituting the coefficients $g_k$ in (25) gives
\[ y_0^* = Ay_{x+1} - By_{x+2} + Cy_{x+3} + C^{-1}(B\theta \delta_{x+1} - A\theta^2 \delta_{x+2} + \theta^3 \delta_{x+3}). \]

Here $A = \beta_1^{-1} + \beta_2^{-1} + \beta_3^{-1}$, $B = \beta_1^{-1} \beta_2^{-1} + \beta_2^{-1} \beta_3^{-1} + \beta_3^{-1} \beta_1^{-1}$, and $C = \beta_1^{-1} \beta_2^{-1} \beta_3^{-1}$. The short formula can be used to compute
$y_0^*$ by techniques discussed in [1].
The problem just discussed is the backward extrapolation of a time series. Forward extrapolation presents a similar problem. Thus let $x$ denote days elapsed. Then the short formula (36) still holds provided $\beta_i = \exp(T_i^{-1})$ and provided $\theta$ satisfies inequality (4).

VII. Extrapolation with Trigonomials

An exponomial in which the bases have unit absolute value, $|\beta_k| = 1$, is a linear combination of sines and cosines and so may be termed a trigonomial. In particular we shall be concerned with trigonomials of the form

$$p(x) = a_0 + \sum_{k=1}^{M} (a_k \cos \rho_k x + b_k \sin \rho_k x)$$

(37)

where the $\rho_k$ are distinct positive constants. This is an exponomial such that if $\beta$ is a base so also is $\tilde{\beta}$ and $\beta \tilde{\beta} = 1$.

For example if $M = 1$ then

$$p(x) = a_0 + a_1 \cos \rho_1 x + b_1 \sin \rho_1 x$$

and

$$G(X) = (1 - X)(1 - 2 \cos \rho_1 X + X^2) = f(X).$$

Thus the short formula is

$$y_0^* = A(y_1 + \theta \delta_4) - A(y_2 + \theta^2 \delta_2) + (y_3 + \theta^3 \delta_3)$$

(38)

where $A = 1 + 2 \cos \rho_1$. Again, if $M = 2$, then

$$p(x) = a_0 + a_1 \cos \rho_1 x + b_1 \sin \rho_1 x + a_2 \cos \rho_2 x + b_2 \sin \rho_2 x$$

and the short formula is

$$y_0^* = A(y_1 + \theta \delta_4) - B(y_2 + \theta^2 \delta_2) + B(y_3 + \theta^3 \delta_3) - A(y_4 + \theta^4 \delta_4) + (y_5 + \theta^5 \delta_5)$$

(39)

where

$$A = 1 + 2 \cos \rho_1 + 2 \cos \rho_2,$$

$$B = 2 + 2 \cos \rho_1 + 2 \cos \rho_2 + 4 \cos \rho_1 \cos \rho_2.$$
The studies in this note and in [1] have been exclusively concerned with least squares using an exponential weight factor. In another place Duffin and Schmidt will present a study of certain other weight factors.

Reference