International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Conditions for the Optimality of Exponential Smoothing Forecast Procedures

Ledolter, J. and Box, G.E.P.

IIASA Professional Paper
March 1976

PP-76-4

CONDITIONS FOR THE OPTIMALITY OF EXPONENTIAL SMOOTHING FORECAST PROCEDURES

Johannes Ledolter

March 1976

Conditions for the Optimality of Exponential Smoothing Forecast Procedures

Johannes Ledolter* and George E.P. Box**

PREFACE

The use at time t of available observations from a time series to forecast its value at some future time t+ provides an important basis for planning and control. The obtaining of good forecasts is an important part of model building at IIASA.

*currently at IIASA, Laxenburg, Austria

**University of Wisconsin-Madison, USA.

Conditions for the Optimality of Exponential Smoothing Forecast Procedures

Johannes Ledolter and George E.P. Box

Abstract

Exponential smoothing procedures, in particular those recommended by Brown [3] are used extensively in many areas of economics, business and engineering. It is shown in this paper that:

i) Brown's forecasting procedures are optimal in terms of achieving minimum mean square error forecasts only if the underlying stochastic process is included in a limited subclass of ARIMA (p,d,q) processes. Hence, it is shown what assumptions are made when using these procedures.

ii) The implication of point (i) is that the users of Brown's procedures tacitly assume that the stochastic processes which occur in the real world are from the particular restricted subclass of ARIMA (p,d,q) processes. No reason can be found why these particular models should occur more frequently than others.

iii) It is further shown that even if a stochastic process which would lead to Brown's model occurred, the actual methods used for making the forecasts are clumsy and much simpler procedures can be employed.

1. The class of autoregressive integrated moving average processes and their minimum mean square error forecasts

An approach to the modelling and forecasting of stationary and nonstationary processes, such as commonly occur in business, economics and engineering, is discussed by Box and Jenkins [2]. Utilizing earlier work by Kolmogorov [7,8], Wold [12], Yaglom [13], Yule [14], it uses a three stage iterative model building procedure of identification, estimation and diagnostic checking.

The class of autoregressive integrated moving average (ARIMA) models of order (p,d,q) which is discussed in [2] can be written

$$\varphi_{p+d}(B)z_t = \phi_p(B)(1-B)^d z_t = \theta_q(B)a_t \qquad (1.1)$$

where

i) z_t is a discrete stochastic process

ii) $\phi_p(B) = 1-\phi_1 B-\ldots-\phi_p B^p$

$\theta_q(B) = 1-\theta_1 B-\ldots-\theta_q B^q$

$\varphi_{p+d}(B) = \phi_p(B)(1-B)^d = 1-\varphi_1 B-\ldots-\varphi_{p+d}B^{p+d}$

and B is the backshift operator: $B^m z_t = z_{t-m}$

iii) $\{a_t\}$ is a white noise sequence

$$Ea_t = 0 \quad \text{and} \quad Ea_t a_{t-k} = \begin{cases} 0 & k \neq 0 \\ \sigma_a^2 & k=0 \end{cases}$$

The roots of $\phi_p(B) = 0$ and $\theta_q(B) = 0$ are assumed to lie outside the unit circle.

ARIMA (p,d,q) processes provide a class of models capable of representing time series which, although not necessarily stationary, are homogeneous and in statistical equilibrium.

The stochastic process in (1.1) can equivalently be written in terms of current and previous shocks a_t

$$z_t = a_t + \sum_{j\geq 1} \psi_j a_{t-j} \qquad (1.2)$$

where

$$\psi(B) = 1 + \sum_{j\geq 1} \psi_j B^j = \varphi_{p+d}^{-1}(B)\theta_q(B)$$

or in terms of a weighted sum of previous values of the stochastic process and the current shock a_t.

$$z_t = \sum_{j\geq 1} \pi_j z_{t-j} + a_t \qquad (1.3)$$

where

$$\pi(B) = 1 - \sum_{j\geq 1} \pi_j B^j = \varphi_{p+d}(B)\theta_q^{-1}(B)$$

Forecasts of ARIMA (p,d,q) processes:

Minimum mean square error forecasts for linear stochastic processes are given by the conditional expectation of future observations

$$\hat{z}_t(\ell) = [z_{t+\ell}] = E\{z_{t+\ell}|z_t, z_{t-1}, \ldots\} \qquad (1.4)$$

Forecasts are calculated using the difference equation form of the model

$$\hat{z}_t(\ell) = [z_{t+\ell}] = \varphi_1[z_{t+\ell-1}] + \ldots + \varphi_{p+d}[z_{t+\ell-(p+d)}] + [a_{t+\ell}] - \theta_1[a_{t+\ell-1}] - \ldots - \theta_q[a_{t+\ell-q}] \qquad (1.5)$$

where

$$[z_{t+j}] = \begin{cases} z_{t+j} & \text{for } j \leq 0 \\ \hat{z}_t(j) & \text{for } j > 0 \end{cases}$$

$$[a_{t+j}] = \begin{cases} a_{t+j} & \text{for } j \leq 0 \\ 0 & \text{for } j > 0 \end{cases}$$

Forecasts can equivalently be expressed as a linear function of previous observations

$$[z_{t+\ell}] = \sum_{j\geq 1} \pi_j [z_{t+\ell-j}] + [a_{t+\ell}] \quad . \qquad (1.6)$$

In particular, for $\ell = 1$

$$\hat{z}_t(1) = \sum_{j\geq 1} \pi_j z_{t+1-j}$$

Forecasts can be updated from one time origin to the other by

$$\hat{z}_{t+1}(\ell) = \hat{z}_t(\ell+1) + \psi_\ell a_{t+1} \quad . \tag{1.7}$$

Although forecasts are calculated and updated most conveniently from the difference equation form (1.5), from the point of studying the nature of the forecasts it is profitable to consider the explicit form of the forecast function. The eventual forecast function is the solution of the difference equation

$$\hat{z}_t(\ell) - \varphi_1\hat{z}_t(\ell-1)-\ldots-\varphi_{p+d}\hat{z}_t(\ell-p-d) = 0 \quad \text{for} \quad \ell > q$$

and is given by

$$\hat{z}_t(\ell) = b_1^*(t)f_1^*(\ell)+\ldots+b_{p+d}^*(t)f_{p+d}^*(\ell) \quad \text{for } \ell>q-p-d \tag{1.8}$$

$f_1^*(\ell),\ldots,f_{p+d}^*(\ell)$ are functions of the lead time ℓ and depend only on the autoregressive part of the model $\varphi_{p+d}(B)$. In general, these functions can be polynomials, exponentials, sines, cosines or combinations of these functions.

For a given forecast origin t, the coefficients $\underset{\sim}{b}^*(t) = [b_1^*(t),\ldots,b_{p+d}^*(t)]'$ are constants and are the same for all lead times ℓ; however they change from one forecast origin to the next and as shown by Box and Jenkins [2] they can be updated by

$$\underset{\sim}{b}^*(t) = L^{*'}\underset{\sim}{b}^*(t-1) + \underset{\sim}{g}[z_t - \hat{z}_{t-1}(1)] \tag{1.9}$$

where

$$L^{*'} = F_\ell^{*-1}F_{\ell+1}^*$$

$$F^*_\ell = \begin{bmatrix} f^*_1(\ell) & \cdots & f^*_{p+d}(\ell) \\ \vdots & & \vdots \\ f^*_1(\ell+p+d-1) & \cdots & f^*_{p+d}(\ell+p+d-1) \end{bmatrix}$$

and

$$\underset{\sim}{g} = F^{*-1}_\ell \underset{\sim}{\psi}_\ell \qquad \text{with } \underset{\sim}{\psi}_\ell = [\psi_\ell, \psi_{\ell+1}, \ldots, \psi_{\ell+p+d-1}]'$$

for any $\ell > q-p-d$.

2. Exponential smoothing forecast procedures

Exponential smoothing techniques have received broad attention in the existing literature, especially in the area of management science. These procedures are fully automatic which means that once a computer program has been written, forecasts for any time series can be derived without manual intervention. The fact that they are automatic has been put forward as an advantage of the scheme. However it can equally well be argued that this is a great disadvantage since it discourages the use of the human mind in circumstances where this instrument could be used with profit.

The basic exponential smoothing equation replaces an observed series z_t by a smoothed series $\bar{z}_t$, an exponentially weighted average of current and past values of z.

$$\bar{z}_t = \alpha \sum_{j \geq 0} (1-\alpha)^j z_{t-j} \quad . \tag{2.1}$$

The latest available smoothed value is used to forecast all future observations

$$\hat{z}_t(\ell) = \bar{z}_t \quad . \tag{2.2}$$

This basic exponential smoothing procedure by Holt [6], Winters [11], Brown [3] was, and still is, used frequently to derive forecasts of economic and business data. Muth [9] investigated the conditions under which this procedure provides minimum mean square error forecasts. He showed that the underlying process has to be given by the ARIMA (0,1,1) process

$$(1-B)z_t = (1-(1-\alpha)B)a_t \quad . \tag{2.3}$$

Generalizations of exponential smoothing procedures have been considered by Brown [3], Brown and Meyer [4]. They select fitting functions $\underset{\sim}{f}(\ell) = [f_1(\ell),\ldots,f_m(\ell)]'$ from the class of functions

$$\underset{\sim}{f}(\ell+1) = L\underset{\sim}{f}(\ell) \quad . \tag{2.4}$$

L is a $(m \times m)$ non singular transition matrix and $\underset{\sim}{f}(0)$ is specified. The coefficients $\underset{\sim}{b}(t) = [b_1(t),\ldots,b_m(t)]'$ of the forecast function

$$\hat{z}_t(\ell) = \hat{p}(t,\ell) = \underset{\sim}{b}'(t)\underset{\sim}{f}(\ell) \tag{2.5}$$

are fitted by discounted least squares minimizing

$$\sum_{j\geq 0} \beta^j [z_{t-j} - \hat{p}(t,-j)]^2 \quad .$$

The fitting functions are chosen by visual inspection and the smoothing constant $\beta (0<\beta<1)$ is assumed to be known. Brown suggests picking β between .7 and .9.

The coefficients b(t) are updated from one time origin to the other by

$$\underset{\sim}{b}(t) = L'\underset{\sim}{b}(t-1) + \underset{\sim}{h}[z_t - \hat{p}(t-1,1)] \tag{2.6}$$

where

$$\underset{\sim}{h} = F^{-1}\underset{\sim}{f}(0)$$

and

$$F = \sum_{j\geq 0} \beta^j \underset{\sim}{f}(-j)\underset{\sim}{f}'(-j) \quad .$$

3. Equivalence theorem for forecasts derived from exponential smoothing forecast procedures and forecasts from ARIMA models.

Summary of the equivalence theorem:

In the appendix we prove the following theorem.

Brown's exponential smoothing forecast procedures with specified fitting functions $\underset{\sim}{f}(\ell)$ and smoothing parameter β will provide minimum mean square error forecasts if and only if the underlying stochastic process follows the ARIMA model

$$\varphi(B)z_t = \varphi(\beta B)a_t \quad . \tag{3.1}$$

The roots of $\varphi(B) = 0$ lie on the unit circle and the eventual forecast function of (3.1) is given by the forecast function of Brown's model.

Some examples:

1. Constant model

$f(\ell) = 1 \quad \forall \ell$ and smoothing parameter β

$\hat{p}(t,\ell) = b_1(t) \quad .$

The equivalent ARIMA model is given by

$(1-B)z_t = (1-\beta B)a_t \quad .$

2. Linear model

$$\underset{\sim}{f}(\ell) = \begin{Bmatrix} 1 \\ \ell \end{Bmatrix} \text{ and smoothing parameter } \beta$$

$$\hat{p}(t,\ell) = b_1(t) + b_2(t)\ell \quad .$$

The equivalent ARIMA model is given by

$$(1-B)^2 z_t = (1-\beta B)^2 a_t \quad .$$

3. 12-point sinusoidal model

$$\underset{\sim}{f}(\ell) = \begin{Bmatrix} 1 \\ \sin \frac{2\pi}{12}\ell \\ \cos \frac{2\pi}{12}\ell \end{Bmatrix} \text{ and smoothing parameter } \beta.$$

The equivalent ARIMA model is given by

$$(1-B)(1-\sqrt{3}\,B + B^2)\, z_t = (1-\beta B)(1-\sqrt{3}\,\beta B + \beta^2 B^2)\, a_t \quad .$$

Interpretation and shortcomings of Brown's exponential smoothing method:

In the light of the above theorem the shortcomings of Brown's forecasting procedure are threefold.

i) The class of fitting functions decides the form of the left hand side of the difference equation model (3.1) (autoregressive operator). In Brown's method the fitting functions are chosen by visual inspection of the time series itself. As is shown in Box and Jenkins [2] visual inspection of the time series alone can be quite misleading and more reliable identification tools such as sample autocorrelation and sample partial autocorrelation function have to be considered.

ii) The exponentially discounted least squares procedure then forces the right hand side of the difference equation (moving average operator) to be of the form $\varphi(\beta B)$. It is thus automatically determined by the autoregressive part on the left hand side of model (3.1) and is a function of the smoothing constant only.

iii) The smoothing constant β is assumed to be known. Brown states that the smoothing constant should be picked between .7 and .9. Actual study of time series, however, gives no empirical support to this assertion and no theoretical reasons seem to be available for discussion. The supposition that β ought to be picked in this range appears strange.

The π-weights implied by Brown's model

The π-weights for the ARIMA model (3.1) are given by

$$\pi_1 = (1-\beta)\varphi_1$$

$$\pi_j = \beta\varphi_1\pi_{j-1}+\beta^2\varphi_2\pi_{j-2}+\ldots+\beta^{j-1}\varphi_{j-1}\pi_1+(1-\beta^j)\varphi_j \qquad 2\leq j\leq n$$

$$\pi_j = \beta\varphi_1\pi_{j-1}+\beta^2\varphi_2\pi_{j-2}+\ldots+\beta^n\varphi_n\pi_{j-n} \qquad j\geq n+1$$

(3.2)

It is instructive to look at the π-weights since they show how past observations are discounted to derive one step ahead minimum mean square error forecasts

$$\hat{z}_t(1) = \sum_{j\geq 1} \pi_j z_{t+1-j} \quad .$$

One must ask the question: "Is there any reason to believe that the world behaves according to this class of weight functions given in (3.2)?" Pandit [8] has tried to find some theoretical reasons why business, economic, and quality control systems can be predicted by exponential smoothing methods, giving it a "spring-dashpoint" interpretation. The analogy

seems strange and is contradicted by many time series, which have been modelled by the three stage iterative Box-Jenkins method.

The data themselves should determine the form of the model and the value of its parameters. The π-weights should depend on the underlying process which has to be identified properly.

Computation of the forecasts

Brown's forecasting procedures are claimed to be computationally efficient. It is easily seen, however, that the forecasts can be derived more readily directly from the difference equation of the equivalent ARIMA model (3.1). Thus, even if one believed in the adequacy of Brown's implied model, one should not use his method to calculate and update the forecasts.

These points are best brought out by consideration of specific examples.

4. Illustrative examples

Example 1: Daily IBM common stock closing prices. The data is given in Box and Jenkins [2]. After inspection of the series, Brown [3] argued that short pieces of the data could be represented by quadratic curves and that one, therefore, ought to consider quadratic fitting functions given by

$$\hat{p}(t,\ell) = b_0(t)+b_1(t)\ell+b_2(t)\ell^2 \quad . \tag{4.1}$$

He updates the coefficients of the forecast function by discounted least squares and chooses a smoothing constant of $\beta = .9$.

The forecasts for this form of Brown's model are shown

for several time origins and for lead times $\ell = 1,2,3$ in the first column of Table 1.

If Brown's model were to be used then it would be much more convenient to use the theory developed in this paper and to calculate the forecasts directly from the equivalent difference equation given below

$$(1 - B)^3 z_t = (1 - .9B)^3 a_t \quad . \tag{4.2}$$

The forecasts are given in the second column of Table 1.

In fact, however, as was shown originally by Box and Jenkins [2], Brown's model seems to be totally inadequate. This is seen for example by the much larger mean square error of the one, two, and three steps ahead forecasts given in Table 1. Identification using Box-Jenkins methods leads to consider an ARIMA (0,1,1) model with the moving average parameter estimated close to zero

$$(1 - B) z_t = (1 + .087B) a_t \quad . \tag{4.3}$$

It was noted that the model in (4.3) is very nearly a random walk as originally suggested by Bachelier [1]. This model implies that the best forecasts of future observations are very nearly the current value of the stock price. This is very different from Brown's model which implies that information about the next value is not only contained in the current observation but also in the observations before.

The π-weights for the models (4.2) and (4.3) are shown in the diagram in Table 2.

The autocorrelation of the one step ahead forecast errors for models (4.2) and (4.3) are given in Table 1. It can be seen that there is significant autocorrelation among the one step ahead forecast errors for the difference equation model

(4.2) which is implied by Brown's forecasting procedure in (4.1). For the model (4.3) the autocorrelations are essentially zero.

Calculating the forecasts

It is worth emphasizing again that if forecasts were to be derived from Brown's model, one should not use Brown's method of calculating and updating, which is extremely laborious. It is much easier to calculate the forecasts directly from the equivalent difference equation. This will give the same result, except for rounding errors, as it is shown in Table 1.

The same point can be made in terms of a further example. Example 2: Warmdot filter sales. This series is given in Brown [3].

Brown chooses the simple 12-point sinusoidal model

$$\hat{p}(t,\ell) = b_0(t)+b_1(t)\sin\frac{2\pi}{12}\ell+ b_2(t)\cos\frac{2\pi}{12}\ell \qquad (4.4)$$

and updates the coefficients of the forecast function by discounted least squares with smoothing constant $\beta = .9$. The forecasts for this form of Brown's model are shown in the first column of Table 3.

There is no point in going through the Box-Jenkins three stage iterative method for this series because this appears to be an artificial series which has been manufactured from the model (4.4).

However if at the moment we suppose that forecasts of this type are needed, the theory of this paper supplies a much simpler method of obtaining the forecasts. The stochastic difference equation which would provide minimum mean square error forecasts for the procedure (4.4) is given by

$$(1-B)(1-\sqrt{3}B+B^2)z_t = (1-.9B)(1-\sqrt{3}(.9)B+(.9)^2B^2)a_t \quad . \qquad (4.5)$$

Table 3 shows that except for rounding errors the forecasts using the stochastic difference equation coincide with the forecasts derived by Brown.

time origin	lead time	Brown's model $\hat{p}(t,\ell)=b_0(t)+b_1(t)\ell+b(t)\ell^2$ $\beta = .9$ (4.1)	Difference equation form equivalent to Brown's model (4.2)	Difference equation form of the ARIMA (0,1,1) suggested by Box-Jenkins (4.3)
		F O R E C A S T S		
300	1	382.59	382.50	376.63
	2	385.56	385.56	376.63
	3	388.66	388.66	376.63
310	1	385.93	385.92	375.96
	2	387.49	387.48	375.96
	3	389.09	389.09	375.96
320	1	409.33	409.32	408.82
	2	411.98	411.98	408.82
	3	414.71	414.70	408.82
330	1	387.31	387.31	384.22
	2	386.72	386.71	384.22
	3	386.07	386.07	384.22
340	1	375.63	375.62	362.92
	2	374.16	374.15	362.92
	3	372.63	372.62	362.92
350	1	346.60	346.60	359.99
	2	344.15	344.14	359.99
	3	341.63	341.62	359.99
360	1	348.21	348.21	342.62
	2	347.06	347.05	342.62
	3	345.89	345.89	342.62
369	1	345.88	345.88	357.38
	2	345.65	345.65	357.38
	3	345.45	345.44	357.38

lead time	Observed mean square error of forecasts obtained at various lead times ℓ	
1	186.81	92.00
2	279.78	188.00
3	378.17	291.34
Correlation among the one step ahead forecast errors	$r_1 = .71$	$r_1 = .00$
	$r_2 = .44$	$r_2 = .00$

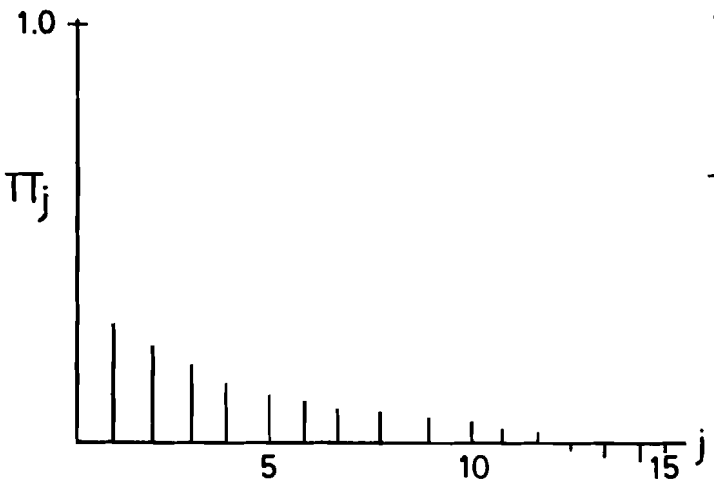
Table 1: Forecasts, observed mean square error of the forecasts and auto-correlation among one step ahead forecast errors for Brown's model (4.1), using both Brown's method of calculating the forecasts and the equivalent difference equation form (4.2), and for model (4.3) suggested by Box-Jenkins procedures.

π-weights implied by Brown's model

j	π_j	j	π_j
1	.300	11	.008
2	.240	12	.000
3	.190	13	-.001
4	.146	14	-.001
5	.114	15	-.012
6	.086	16	-.015
7	.063	17	-.016
8	.044	18	-.016
9	.030	19	-.017
10	.018	20	-.016

π-weights of the fitted ARIMA (0,1,1) model (4.3) suggested by Box and Jenkins

j	π_j
1	1.087
2	- .094
3	.008
4	- .001
$j \geq 5$	.000



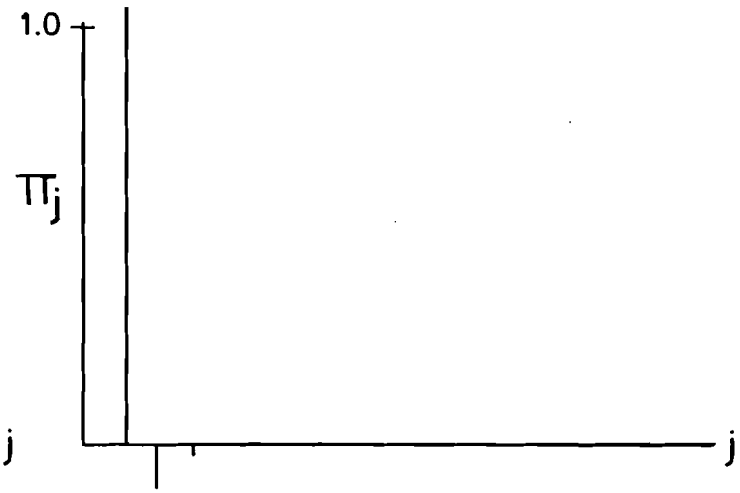


Table 2: The π-weights implied by Brown's model (4.2) and for the fitted ARIMA (0,1,1) model (4.3) suggested by Box and Jenkins.

time origin	lead time	Forecasts for the model (4.4) using Brown's method of forecasting with smoothing constant $\beta = .9$	Forecasts for the model (4.4) using the equivalent difference equation (4.5) for the calculation of the forecasts
80	1	130.01	130.12
	2	151.16	151.27
	3	169.64	169.73
90	1	101.48	101.52
	2	114.61	114.67
	3	135.23	135.28
100	1	105.71	105.72
	2	98.66	98.68
	3	102.36	102.37
110	1	132.17	132.17
	2	102.98	102.98
	3	81.52	81.52
120	1	163.80	163.80
	2	149.53	149.53
	3	133.03	133.03

Table 3: Forecasts for the Warmdot filter sales using Brown's model (4.4): Comparison of Brown's method of calculating the forecasts and the forecasts derived from the equivalent difference equation (4.5).

APPENDIX

Proof of equivalence theorems for forecasts derived from exponential smoothing forecast procedures and forecasts from ARIMA models.

Theorem 1.

Model A: Consider the Brown model with fitting functions

$$\underset{\sim}{f}'(\ell) = [u_1^{-\ell},\ldots,u_n^{-\ell}] \quad \text{with } |u_i| \geq 1;\ u_i \neq u_j \text{ for } i \neq j;$$
$$\text{and } |u_i u_j| < 1/\beta \text{ for } 1 \leq i,\ j \leq n.$$

The coefficients of the forecast function

$$\hat{p}(t,\ell) = b_1(t)u_1^{-\ell} + b_2(t)u_2^{-\ell} + \ldots + b_n(t)u_n^{-\ell} \qquad (A.1)$$

are estimated by discounted least squares with smoothing constant β; $0 < \beta < 1$.

Model B: ARIMA model

$$\prod_{i=1}^{n} (1 - \frac{1}{u_i}B)z_t = \prod_{i=1}^{n} (1 - u_i\beta B)a_t \ . \qquad (A.2)$$

Then, Brown's forecasting procedure using the forecast function (A.1) in model A will provide optimal forecasts in terms of minimizing the mean square forecast error if and only if the underlying stochastic process follows the ARIMA model in (A.2).

Proof. In the proof of theorem 1 we use the following result about the inverse of a Vandermonde matrix:

$$A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ a_1 & a_2 & \dots & a_n \\ \vdots & \vdots & & \\ a_1^{n-1} & a_2^{n-1} & \dots & a_n^{n-1} \end{bmatrix} \quad \text{where } a_i \neq a_j \text{ (for } i \neq j).$$

It is easily shown that the inverse of A is given by

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

where a_{ij} are the coefficients in the expansion

$$P_i(x) = \prod_{\substack{k=1 \\ k\neq i}}^{n} \frac{x-a_k}{a_i-a_k} = a_{i1}x^0 + a_{i2}x^1 + \dots + a_{in}x^{n-1} \quad (1 \leq i \leq n) \tag{A.3}$$

Theorem 1 is proved by showing that model A and model B have

i) the same <u>form</u> of the forecast function
ii) the same <u>updating formula</u> for the coefficients of the forecast function.

ad (i). The eventual forecast function for model B is the solution of

$$\prod_{i=1}^{n} (1 - \frac{1}{u_i}B)\, \hat{z}_t(\ell) = 0 \qquad \text{for} \quad \ell > n$$

and is given by

$$\hat{z}_t(\ell) = b_1^*(t)u_1^{-\ell} + b_2^*(t)u_2^{-\ell} + \dots + b_n^*(t)u_n^{-\ell} \quad \text{for } \ell > 0 \tag{A.4}$$

which coincides with the forecast function of model A given in (A.1).

ad (ii). The updating algorithm for the coefficients of the forecast function $\underset{\sim}{b}'(t)\underset{\sim}{f}(\ell)$ for model A is given by

$$\underset{\sim}{b}'(t) = L'\underset{\sim}{b}(t-1) + \underset{\sim}{h}[z_t - \hat{p}(t-1,1)] \quad . \tag{A.5}$$

Dobbie [5] showed that for the case of exponential fitting functions $\underset{\sim}{h} = (h_1,h_2,\ldots,h_n)'$ is given by

$$h_i = (1 - \beta u_i^2) \prod_{\substack{k=1 \\ k\neq i}}^{n} \frac{1-\beta u_i u_k}{1 - \frac{u_i}{u_k}} \qquad 1 \leq i \leq n \tag{A.6}$$

The updating algorithm for the coefficients of the eventual forecast function of the ARIMA model in (A.2) is given by:

$$\underset{\sim}{b}^*(t) = L'\underset{\sim}{b}^*(t-1) + \underset{\sim}{g}[z_t - \hat{z}_{t-1}(1)] \quad . \tag{A.7}$$

Choosing $\ell = 1$ in (1.9), it is seen that

$$\underset{\sim}{g} = F_1^{*-1}\underset{\sim}{\psi}_1 = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ a_1^2 & a_2^2 & \ldots & a_n^2 \\ \vdots & & & \\ a_1^n & a_2^n & \ldots & a_n^n \end{bmatrix}^{-1} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{bmatrix} \tag{A.8}$$

where $a_i = \frac{1}{u_i} \qquad 1 \leq i \leq n$

and $\psi_k (1 \leq k \leq n)$ are the ψ-weights in

$$\psi(B) = \frac{\prod_{k=1}^{n}(1-\beta u_k B)}{\prod_{k=1}^{n}(1-\frac{1}{u_k}B)} = \frac{\prod_{k=1}^{n}(1-\frac{\beta}{a_k}B)}{\prod_{k=1}^{n}(1-a_k B)} \qquad (A.9)$$

In order to prove theorem 1 we have to show that

$$\underset{\sim}{g} = \underset{\sim}{h} \quad \text{or equivalently that}$$

$$\underset{\sim}{d} \doteqdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ \vdots & & & \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \end{bmatrix}^{-1} \underset{\sim}{\psi}_1 = \begin{bmatrix} a_1 h_1 \\ a_2 h_2 \\ \vdots \\ a_n h_n \end{bmatrix}$$

where $\underset{\sim}{d} = [d_1,\ldots,d_n]'$; $d_i = a_{i1}\psi_1 + a_{i2}\psi_2 + a_{in}\psi_n$ and a_{ij} are the elements of the inverse of matrix A. d_i is the coefficient of x^0 in $\frac{1}{x}P_i(\frac{1}{x})\psi(x)$, where $P_i(x)$ is given in (A.3).

Using (A.3) and (A.9)

$$\frac{1}{x}P_i(\frac{1}{x})\psi(x) = \frac{1}{x^n(1-a_i x)} \frac{\prod_{k=1}^{n}(1-\frac{\beta}{a_k}x)}{\prod_{k\neq i}(a_i - a_k)} \qquad (A.10)$$

It therefore remains to show that the coefficient of x^0 in

$$V_i(x) \doteqdot \frac{\prod_{k=1}^{n}(1-\frac{\beta}{a_k}x)}{x^n(1-a_i x)} \quad \text{equals} \quad a_i h_i \prod_{k\neq i}(a_i - a_k)$$

$$= a_i^n(1-\frac{\beta}{a_i^2}) \prod_{k\neq i}(1-\frac{\beta}{a_i a_k})$$

(A.11)

However,

$$V_i(x) = \frac{\prod_{k=1}^{n} \left(\frac{1}{x} - \frac{\beta}{a_k}\right)}{(1 - a_i x)} = [1+a_i x+a_i^2 x^2+\ldots] \times \left[\left(\frac{1}{x}\right)^n + c_1 \left(\frac{1}{x}\right)^{n-1} + \ldots + c_{n-1}\left(\frac{1}{x}\right) + c_n\right] \tag{A.12}$$

where the c_j's are the coefficients in the expansion of $\prod_{k=1}^{n} \left(\frac{1}{x} - \frac{\beta}{a_k}\right)$ given by

$$\begin{cases} c_1 = -\beta \sum_{k=1}^{n} \frac{1}{a_k} \\ c_2 = \beta^2 \sum_{k<\ell} \frac{1}{a_k a_\ell} \\ c_3 = -\beta^3 \sum_{k<\ell<m} \frac{1}{a_k a_\ell a_m} \\ \vdots \\ c_n = (-1)^n \frac{\beta^n}{a_1 a_2 \ldots a_n} \end{cases} . \tag{A.13}$$

Furthermore

$$\prod_{k\neq i} \left(1 - \frac{\beta}{a_i a_k}\right) = 1 + e_1^{(i)} + \ldots + e_{n-1}^{(i)} \tag{A.14}$$

where the coefficients $e_j^{(i)}$ $(1 \leq j \leq n-1)$ are given by

$$
\begin{cases}
e_1^{(i)} = -\dfrac{1}{a_i}\displaystyle\sum_{k\neq i}\dfrac{\beta}{a_k} \\
e_2^{(i)} = \dfrac{1}{a_i^2}\displaystyle\sum_{\substack{k<\ell \\ k\neq i \quad \ell\neq i}}\dfrac{\beta^2}{a_k a_\ell} \\
e_3^{(i)} = \dfrac{1}{a_i^3}\displaystyle\sum_{\substack{k<\ell<m \\ k\neq i,\ell\neq i,m\neq i}}\dfrac{\beta^3}{a_k a_\ell a_m} \\
\vdots \\
e_{n-1}^{(i)} = (-1)^{n-1}\dfrac{1}{a_i^{n-1}}\beta^{n-1}\displaystyle\prod_{k\neq i}\dfrac{1}{a_k} \quad .
\end{cases}
\tag{A.15}
$$

Therefore, the right hand side of (A.11) is given by

$$
a_i^n\left(1-\frac{\beta}{a_i^2}\right)\prod_{k\neq i}\left(1-\frac{\beta}{a_i a_k}\right) = a_i^n\left(1-\frac{\beta}{a_i^2}\right)\left(1+e_1^{(i)}+\ldots+e_{n-1}^{(i)}\right) \quad .
\tag{A.16}
$$

Since the relation between the coefficients in (A.13) and (A.15) is

$$
\begin{cases}
c_1 = a_i e_1^{(i)} - \dfrac{\beta}{a_i} \\
c_2 = a_i^2 e_2^{(i)} - \beta e_1^{(i)} \\
c_3 = a_i^3 e_3^{(i)} - \beta a_i e_2^{(i)} \\
\vdots \\
c_{n-1} = a_i^{n-1} e_{n-1}^{(i)} - \beta a_i^{n-3} e_{n-2}^{(i)} \\
c_n = -\beta a_i^{n-2} e_{n-1}^{(i)}
\end{cases}
\tag{A.17}
$$

the coefficient of x^0 in $V_i(x)$ is given by

$$a_i^n + a_i^{n-1}c_1 + \ldots + a_i c_{n-1} + c_n = a_i^n\left(1 - \frac{\beta}{a_i^2}\right)\left[1 + e_1^{(i)} + e_2^{(i)} + \ldots + e_{n-1}^{(i)}\right]. \quad \text{(A.18)}$$

This equals (A.16), thus proving the claim that $\underset{\sim}{g} = \underset{\sim}{h}$.

q.e.d.

Furthermore it is easily seen that Brown's forecast procedure with forecast function $\underset{\sim}{b}(t)'\underset{\sim}{d}(\ell)$, where $\underset{\sim}{d}(\ell)$ is any non singular linear combination of the exponential fitting functions $\underset{\sim}{f}(\ell)$ of theorem 1,

$$\underset{\sim}{f}(\ell) = R\underset{\sim}{d}(\ell)$$

will provide minimum mean square error forecasts if and only if the underlying process follows the ARIMA model B.

Sinusoidal fitting functions which are frequently considered by Brown can be written as linear combinations of such exponential functions.

For example the fitting functions $d(\ell) = \begin{Bmatrix} 1 \\ \sin\omega\ell \\ \cos\omega\ell \end{Bmatrix}$ in $\hat{p}(t,\ell) = b_1(t) + b_2(t)\sin\omega\ell + b_3(t)\cos\omega\ell$ can be written as

$$\underset{\sim}{f}(\ell) = R\underset{\sim}{d}(\ell)$$

where

$$\begin{bmatrix} 1 \\ e^{-i\omega\ell} \\ e^{i\omega\ell} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -i & 1 \\ 0 & i & 1 \end{bmatrix}\begin{bmatrix} 1 \\ \sin\omega\ell \\ \cos\omega\ell \end{bmatrix} \quad \text{(A.19)}$$

Equation (A.19) shows that the roots of the characteristic equation for the ARIMA model which is implied by the sinusoidal

fitting functions of Brown's forecasting scheme are lying on the unit circle.

In the following corollary we use the fact that if the characteristic equation of the ARIMA model has a complex root, the conjugate complex will be a solution, too.

Corollary 1:

Model B: We consider the ARIMA model

$$\varphi(B)z_t = \varphi(\beta B)a_t$$

where the coefficients of $\varphi(B)$ are real.

Furthermore it is assumed that the roots of $\varphi(B) = 0$ are distinct and lie on the unit circle, and that the eventual forecast function is given by

$$\hat{z}_t(\ell) = b_1^*(t)f_1(\ell)+\ldots+b_n^*(t)f_n(\ell) \quad .$$

Model A: Consider the Brown forecasting procedure with fitting functions

$$\underset{\sim}{f}'(\ell) = [f_1(\ell),\ldots,f_n(\ell)] \quad .$$

The coefficients of the forecast function

$$\hat{z}_t(\ell) = \underset{\sim}{b}'(t)\underset{\sim}{f}(\ell)$$

are fitted by discounted least squares with smoothing coefficient β, $0 < \beta < 1$. Then the Brown forecasting procedure with fitting functions as specified in model A will provide minimum mean square error forecasts if and only if the underlying stochastic process is given by the ARIMA model B.

In the following theorem we relax the assumption of distinct roots of $\varphi(B) = 0$.

Theorem 2:

Model B: $\varphi(B)z_t = \varphi(\beta B)a_t$ as specified in corollary 1, however we allow the possibility of multiple roots of $\varphi(B) = 0$ on the unit circle.

Model A: Brown model as specified in corollary 1.

Then model A and model B are equivalent in terms of having

(i) the same form of forecast function
(ii) the same recursive updating formula for the coefficients of the forecast function $\underset{\sim}{b}'(t)$.

Proof: The only part which remains to show is (ii).

We can write

$$\varphi(B) = \prod_{i=1}^{s} \varphi_i(B) \quad \text{where} \quad \varphi_i(B) \quad 1 \leq i \leq s$$

are real valued polynomials in B with distinct roots on the unit circle;

Model B can be written as

$$\varphi_1(B)\,\varphi_2(B)\ldots\varphi_s(B)z_t = \varphi_1(\beta B)\,\varphi_2(\beta B)\ldots\varphi_s(\beta B)a_t$$

or

$$a_t = \frac{\varphi_1(B)}{\varphi_1(\beta B)}\,\frac{\varphi_2(B)}{\varphi_2(\beta B)}\cdots\frac{\varphi_s(B)}{\varphi_s(\beta B)}\,z_t$$

We define $\alpha_t^{(s)} = \dfrac{\varphi_s(B)}{\varphi_s(\beta B)}\,z_t$ and through continued application of corollary 1 the claim is proved.

q.e.d.

References

[1] Bachelier, L. (1900), "Theory of Speculation," reprinted in Cootner, P.H. (ed.), The Random Character of Stock Market Prices," MIT Press, Massachusetts, (1964).

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

[3] Brown, R.G., Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice-Hall, New Jersey, (1962).

[4] Brown, R.G. and Meyer, R.F., "The Fundamental Theorem of Exponential Smoothing," Operations Research, 9, 673-685, (1961).

[5] Dobbie, J.M., "Forecasting Periodic Trends by Exponential Smoothing," Operations Research, 11, 908-918, (1963).

[6] Holt, C.C., "Forecasting Trends and Seasonals by Exponentially Weighted Moving Averages," O.N.R. Memorandum, Carnegie Institute of Technology, No. 52, (1957).

[7] Kolmogorov, A., "Interpolation und Extrapolation von stationären zufälligen Folgen", Bull. Acad. Sci. (Nauk) U.S.S.R., Ser. Math., 5, 3-14, (1941).

[8] Kolmogorov, A., "Stationary Sequences in Hilbert Space," Bull. Math., University of Moscow, 2, No. 6, (1941).

[9] Muth, G.F., "Optimal Properties of Exponentially Weighted Forecasts of Time Series with Permanent and Transitory Components," JASA, 55, 299-306, (1960).

[10] Pandit, S.M., "Data Dependent Systems: Modelling, Analysis and Optimal Control via Time Series," Ph.D. thesis, University of Wisconsin, Madison, (1973).

[11] Winters, P.R., "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, 6, 324-342, (1960).

[12] Wold, H.O., A Study in the Analysis of Stationary Time Series, Almquist and Wicksell; Uppsala (2nd ed. 1954), (1938).

[13] Yaglom, A.M., "The Correlation Theory of Processes whose n'th Difference Constitutes a Stationary Process," Matem. Sb. 37, 141, (1955).

[14] Yule, G.U., "On a method of Investigating Periodicities in Disturbed Series, with Special Reference to Wölfer's Sunspot Numbers," Phil. Trans., 226, 267, (1927).