MANAGEMENT SCIENCE
Vol. 12, No. 9, May, 1966
Printed in U.S.A.

THE EFFECTS OF PARAMETER MISSPECIFICATION AND NON-STATIONARITY ON THE APPLICABILITY OF ADAPTIVE FORECASTS*†

JOHN BOSSONS

Carnegie Institute of Technology

A "blowup factor" is defined for the measurement of the effect on forecast error variance of two types of misspecification which may be implicit in the choice of a particular adaptive forecasting scheme: (1) misspecification of the number of non-zero parameters of the stationary linear stochastic process generating the observed time series, and (2) misspecification arising from postulating stationarity when in fact the generating process is non-stationary in mean.

1. Introduction

Adaptive forecasts or exponentially-weighted moving averages have been proposed as normative predictors in numerous potential forecasting and control applications by Box and Jenkins [3], Brown [4] [5], Cox [7], Holt [12], Magee [15], and Winters [24] [25]. As seemingly "rational" predictions which are both simple to compute and similar to popular psychological reinforcement learning theory, their appeal is such that they have also seen widespread use in positive models of the way in which anticipations are formed. (See references in [2].)

The estimation of parameters of adaptive forecasting rules has been investigated by Box and Jenkins [3], Theil and Wage [22], Nerlove and Wage [17], and Leenders [14]. The optimality (i.e., minimum-variance unbiassedness) of the resultant adaptive forecasts has been demonstrated for specific stochastic processes by Box and Jenkins [3], Muth [16] and Nerlove and Wage [17].

The optimality of any forecasting rule can be easily demonstrated if it can be shown (1) that the rule corresponds to a linear transformation of the stochastic process assumed to generate the time series to which the rule is to be applied, and (2) that efficient estimates can be derived for the parameters of the stochastic model (and hence of the rule for "adapting" the forecasts). As Whittle [23, p. 97] has remarked, adaptive forecasting rules are nothing more than transformations of models of linear stationary stochastic processes. Since efficient estimates of parameters can be derived for any linear stationary stochastic process (Cf. Wold [26], Whittle [23], and Phillips [19]), adaptive forecasts are clearly optimal if the corresponding linear stationary stochastic process is in fact the process generating the time series being forecast. The extent to which adaptive forecasts are minimum-variance predictors is a question which can

* Received March 1965.

† A preliminary version of this paper was presented at the San Francisco TIMS meeting in February 1965. I am indebted to David Grether and to Marc Nerlove for their comments and suggestions. I am also indebted to two referees for useful suggestions. This paper has been written with assistance provided by National Science Foundation Grant GS-597.

thus be reduced to the simpler, or at least better-clarified, issue of the extent to which postulating the corresponding linear stochastic model introduces specification bias. Any attempt to test the relative accuracy of different adaptive forecasting mechanisms is thus implicitly if not explicitly an attempt to examine which of the models underlying each predictive mechanism is in fact the more accurate specification of the stochastic process generating the time series being predicted. That such attempts are typically framed only implicitly in terms of measuring model specification error is generally to the disadvantage of such studies.

In this paper, specification error is defined as the additional variance of forecast errors introduced by misspecification of the generating process. Specifically, a measure of specification error will be defined as

$$f = \text{var}\,[x_t - x_t^{**}]/\text{var}\,[x_t - x_t^{*}] - 1$$

where $\{x_t\}$ is the series being forecasted, $\{x_t^{**}\}$ is the series of forecasts generated by the predictive mechanism whose associated specification error is being measured, and $\{x_t^{*}\}$ is a series of optimal (i.e., minimum asymptotic variance) predictors of $\{x_t\}$. The lower bound on f is zero and is clearly obtained by a predictive scheme for which $x_t^{**} = x_t^{*}$ plus terms of order no greater than $N^{-1/2}$, where N denotes the number of realizations in a particular sample. There is no upper bound to f. The larger is f, the greater the extent to which the forecasts $\{x_t^{**}\}$ generated by a particular predictive mechanism is affected by misspecification. This measure may be used for two purposes: (1) to measure the effect of a known misspecification introduced by using a simple predictive scheme in an application where "reality" is more complex, and (2) to measure the extent to which the forecasts generated by a particular predictive scheme is in fact sensitive to various types of misspecification to which the forecasts may be subject. In this paper, two types of specification error will be analyzed: (1) that introduced by postulating a relatively simple linear predictive scheme where in fact the predicted series is generated as a more complex transformation of a stationary stochastic process, and (2) that introduced by postulating such underlying stationarity when in fact the underlying process is non-stationary in mean.

The effect of specification error can be easily estimated if it can be assumed that misspecification is limited to incorrect specification of the number of nonzero parameters in the linear stationary process generating the series.

This latter form of misspecification is of course highly specific. It is easy to estimate the effects of misspecification of the number of parameters in a linear stationary model only because it is assumed that the stochastic process represented by the model is in fact linear and stationary. An example of such parameter misspecification and of the potential usefulness of estimates of the effect of the resultant specification bias is analyzed in the first two sections of this paper. In Section 2, the model analyzed in [22], [17], and [14] is extended to include the seasonality disregarded in the Theil-Nerlove-Wage analysis. In Section 3, this model is then used to derive estimates of the amount of forecast error introduced by disregarding the seasonality of a time series.

Any estimate of specification bias effects is of course contingent upon some further specification which is *not* tested. The estimates presented in Section 3 are based on postulating that the extended model containing a linear model being tested is in fact a valid representation of the time series being forecast. This postulate is of course contingent on the specification that the underlying stochastic process is in fact both linear and stationary. Some effects of non-stationarity are examined in the final section of this paper.

2. The Extended Theil-Wage Model

The stochastic model formulated by Theil and Wage in equations (5.1) through (5.4) of their paper [22, p. 201] is consistent with the stochastic specification implicitly underlying their predictive equation (2.7) only if the seasonal component of that equation is disregarded. In this section their model is extended to incorporate this disregarded component.

It is assumed by Theil and Wage that a reasonable model of the stochastic process generating a variable such as corporate sales or new orders can be postulated by decomposing that variable into multiplicative seasonal, trend, and irregular factors and specifying that the seasonal and trend factors are each generated as a multiplicative random walk. The specific decomposition is as follows: Let the series being predicted be X_t, and let Z_t, Ξ_t, H_t, and W_t denote respectively a multiplicative seasonal component, the expected value of the non-seasonal component of X_t, the ratio of successive values of this (non-seasonal) expected value, and a residual stochastic variable. Then $X_t/Z_t = \Xi_t W_t$ and $\Xi_t/\Xi_{t-1} = H_t$, with W_t, Z_t/Z_{t-L}, and H_t/H_{t-1} being realizations of stationary stochastic processes, and with L denoting the number of periods within a year. The specification that Z_t and H_t are random walks is consistent with the notion that changes in both the seasonal effect and the underlying trend are typically unpredictable; their multiplicative compounding is consistent with empirical evidence (Cf. e.g., [21]) supporting the hypothesis that rates of change in corporate sales are symmetrically distributed rather than the absolute changes themselves.

As Alexander [1] has shown, the assumption of symmetry in the distribution of first differences in the logarithm of a stochastic variable whose rate of change is in fact symmetrically distributed introduces only a very minor specification bias. Consequently, expressing the above decomposition in logarithms of the original variables not only yields a linear model but also yields conveniently-distributed stochastic variables. Using lower case symbols to denote natural logarithms of the original variables, we thus have

$$x_t = \xi_t + \zeta_t + w_t \tag{1}$$

$$\xi_t = \xi_{t-1} + \eta_t \tag{2}$$

$$\eta_t = \eta_{t-1} + v_t \tag{3}$$

$$\zeta_t = \zeta_{t-L} + z_t \tag{4}$$

where L denotes the length of the seasonal cycle. The residuals are all postulated to be realizations of stationary stochastic processes with zero mean and zero covariances. Thus $\mathcal{E}w_t = \mathcal{E}v_t = \mathcal{E}z_t = 0$, $\mathcal{E}w_tv_j = \mathcal{E}w_tz_j = \mathcal{E}v_tz_j = 0$ for all t and j, and $\mathcal{E}w_tw_j = \mathcal{E}v_tv_j = \mathcal{E}z_tz_j = 0$ for all $j \neq t$. Note that as a result of this specification x_t is implicitly assumed to have no deterministic component.

The time series x_t generated by this model is non-stationary. However, since the generating process is stationary, the series can be transformed into a stationary series for which standard results apply. (Cf. [23], section 8.5.) Specifically, it follows from (1) through (4) that the first differences of first differences linked over L periods of x_t are in fact stationary.[1] Denoting these L-period differences of first differences of x_t by $Y_t = \Delta_L\Delta x_t$, we have

$$Y_t = \sum_{j=0}^{L-1} v_{t-j} + \Delta z_t + \Delta_L\Delta w_t \tag{5}$$

with covariance function

$$\phi_{YY}(j) = \begin{cases} [LG^2 + 2K^2 + 4]\sigma_w^2 & \text{for } j = 0 \\ [(L-1)G^2 - K^2 - 2]\sigma_w^2 & \text{for } |j| = 1 \\ (L-j)G^2\sigma_w^2 & \text{for } 1 < |j| < L-1 \\ (G^2+1)\sigma_w^2 & \text{for } |j| = L-1 \\ -2\sigma_w^2 & \text{for } |j| = L \\ \sigma_w^2 & \text{for } |j| = L+1 \\ 0 & \text{for } |j| > L+1 \end{cases} \tag{6}$$

where $G = \sigma_v / \sigma_w$ and $K = \sigma_z/\sigma_w$. Since Y_t is stationary, it may be defined either in a moving average representation or as an autoregressive process. Choosing the latter because of its advantages for estimation, we have

$$Y_t + \sum_{k=1}^{\infty} d_kY_{t-k} = \epsilon_t \tag{7}$$

where $\{\epsilon_t\}$ is a purely random process. From Durbin [8] (Cf. also Whittle [23], section 3.4), efficient estimates of the autoregressive coefficients d_k are given by the least-squares estimates $\hat{d}_k$ obtained from

$$\sum_{k=0}^{P} \hat{d}_k\phi_{YY}(j-k) = \delta_j\hat{\sigma}_\epsilon^2, \qquad j = 0, \cdots, P \tag{8}$$

where $\delta_j = 1$ if $j = 0$ and $\delta_j = 0$ otherwise, and where P is sufficiently large so that using higher values of P does not significantly reduce $\hat{\sigma}_\epsilon^2$, the estimate of the variance of the underlying purely random process.[2]

As in [17], adaptive expectations consist of the following rules for adjusting estimates of the current expected values of the components ξ_t, η_t, and ζ_t.

[1] In the original version of this paper the transformed process was written as $\Delta^2\Delta_Lx_t$ rather than $\Delta_L\Delta x_t$. I am indebted to David Grether for pointing out that $\Delta_L\Delta x_t$ is in fact stationary so that $\Delta^2\Delta_Lx_t$ is overdetermined.

[2] A sequential algorithm due to Durbin [9] for solving (8) by successively incrementing P provides an efficient means of testing the rate of convergence of estimates of σ_ϵ as P is increased.

Denoting these estimates respectively by $\xi_t^{\sim}$, $\eta_t^{\sim}$, and $\zeta_t^{\sim}$, we have[8]

$$x_{t+1}^* = \xi_t^{\sim} + \eta_t^{\sim} + \zeta_{t-L+1}^{\sim} \tag{9}$$

where

$$\xi_t^{\sim} = \xi_{t-1}^{\sim} + \eta_{t-1}^{\sim} + \alpha[x_t - x_t^*] \tag{10}$$

$$\begin{aligned} \eta_t^{\sim} &= \eta_{t-1}^{\sim} + \beta[\xi_t^{\sim} - \xi_{t-1}^{\sim} - \eta_{t-1}^{\sim}] \\ &= \eta_{t-1}^{\sim} + \alpha\beta[x_t - x_t^*] \end{aligned} \tag{11}$$

$$\begin{aligned} \zeta_t^{\sim} &= \zeta_{t-L}^{\sim} + \gamma[x_t - \xi_t^{\sim} - \zeta_{t-L}^{\sim}] \\ &= \zeta_{t-L}^{\sim} + (1-\alpha)\gamma[x_t - x_t^*] \end{aligned} \tag{12}$$

Substituting (10)-(12) in (9), we have

$$x_{t+1}^* = (x_t^* - \zeta_{t-L}^{\sim}) + \eta_{t-1}^{\sim} + \alpha(1+\beta)[x_t - x_t^*] + \zeta_{t-L+1}^{\sim} \tag{13}$$

Lagging (13) one period, collecting terms, lagging the result L periods, and again collecting terms, we have

$$\begin{aligned} x_{t+1}^* = {} & [2 - \alpha(1+\beta)]x_t^* - (1-\alpha)x_{t-1}^* + \alpha(1+\beta)x_t - \alpha x_{t-1} \\ & + [1 - (1-\alpha)\gamma]x_{t-L+1}^* - [2(1 - [1-\alpha]\gamma) \\ & - \alpha(1+\beta)]x_{t-L}^* + (1-\alpha)(1-\gamma)x_{t-L-1}^* \\ & + (1-\alpha)\gamma x_{t-L+1} - [2(1-\alpha)\gamma + \alpha(1+\beta)]x_{t-L} \\ & + [\alpha + (1-\alpha)\gamma]x_{t-L-1} \end{aligned} \tag{14}$$

To simplify this equation, we use the lag operator $zX_t = X_{t-1}$ and set

$$\begin{aligned} A &= \alpha(1+\beta) - 2 \\ B &= 1 - \alpha \\ C &= (1-\alpha)\gamma - 1 \\ M(z) &= C - [2(1+C) + A]z + (C - B + 1)z^2 \\ H(z) &= A + Bz + z^{L-1}M(z) \end{aligned}$$

so that (14) becomes

$$x_{t+1}^* = [H(z) + 2 - z + z^{L-1}(1-z)^2]x_t - zH(z)x_{t+1}^* . \tag{15}$$

Since $Y_{t+1}^* = x_{t+1}^* - x_t - x_{t-L+1} + x_{t-L}$, equation (15) thus implies

$$\begin{aligned} Y_{t+1}^* &= [1 - zH(z)]^{-1}[H(z) + (1-z)(1-z^L)]x_t \\ &= J(z)Y_t \end{aligned} \tag{16}$$

[8] Note that $\zeta_t^{\sim}$ as defined here and in [22] is not a full-information estimate. It would be appropriate in equation (11) to adjust $\zeta_t^{\sim}$ to take into account the constraint that

$$\textstyle\sum_{j=0}^{L-1} \exp\,(\zeta_{t-j}^{\sim}) = 1.$$

This adjustment can be made by simply multiplying $\zeta_{t-j}^{\sim}$ by $[\sum_0^{L-1} \exp\,(\zeta_{t-j}^{\sim})]^{-1}$ for $0 \leq j \leq 11$. Note that the superscript tilde used here is equivalent to the overbar notation in [17].

where

$$J(z) = [1 - zH(z)]^{-1}[1 + (1 - z)^{-1}(1 - z^L)^{-1}H(z)].$$

The adaptive forecasting scheme of [14] thus reduces to an autoregressive function defined on $\{Y_t\}$. From the definition of Y_t made in (7), it is evident (Cf. [23]) that the Y^*_{t+1} is a minimum variance predictor of Y_{t+1} if and only if

$$J(z) = \hat{d}(z)[\hat{d}(z)^{-1}/z]_+ \tag{17}$$

where the operator $[\ ]_+$ denotes the exclusion of negative powers of z. Equation (17) consequently provides a basis for defining the optimal values of the smoothing parameters α, β, and γ by expanding $J(z)$ in partial fractions. Since $\hat{d}(z)$ can be estimated either from subjective *a priori* estimates of G and K or from observed estimates $\hat{\phi}_{YY}(j)$ of the autocovariance function for Y_t, equation (17) provides a general form for the determination of optimal smoothing parameters.

3. Evaluation of the Effect of Parameter Misspecification

Given linearity in any generating model or forecasting mechanism, we can compute coefficients of the extrapolative function $\mu(z)$ which generates the predicted forecast $x^*_{t+k} = \mu(z)x_t$, provided that the x_t are realizations of a stationary stochastic process.[4] The extrapolative function $\mu(z)$ can be calculated either from estimates of the autoregressive coefficients $\alpha(z)$ calculated from the covariance function $\phi_{xx}(z)$ (which in turn can either be estimated from data or defined from known or predicted characteristics of the model) or can be derived by transforming the forecasting mechanism into an extrapolative function. In the former case we have

$$\mu(z) = \alpha(z)[\alpha(z)^{-1}/z^k]_+ \tag{18}$$

as in equation (17) above. Given $\mu(z)$ and given that the underlying model is in fact a correct specification of the stochastic process generating x_t, the error variance associated with the model or predictive mechanism is

$$\text{var}\,(x_{t+k} - x^*_{t+k}) = \sigma_e^2 \textstyle\sum_{j=0}^{k-1} \beta_j^2 \tag{19}$$

[4] As Whittle [23, section 8.5] has shown, stationarity of the x_t is a sufficient but not necessary condition for the existence of a forecast $x_{t+k} = \mu(z)x_t$ which is best linear unbiased. The necessary condition is that the model generating the x_t be stationary. Consequently equation (18) applies to the Theil-Wage model with $\beta(z) = \alpha(z)^{-1}$ defined as

$$\beta_j = \textstyle\sum_{k=0}^{j} (k + 1)b_{j-k} + \sum_{k=L}^{j} (k + 1)b_{j-k}\,,$$

where $b(z) = d(z)^{-1}$. (Cf. [23], p. 95.) It should be noted that in some applications the empirical fit of the Theil-Wage model can be improved by specifying that

$$Y_t = (1 - z^L)(1 - \rho z)X_t$$

with Y_t defined as before in (5) but with $\rho < 1$ instead of $\rho = 1$ as in the Theil-Wage model. This modified model has been used by Nerlove and Grether in a forthcoming analysis of unemployment data.

where $\beta(z) = \alpha(z)^{-1}$ and where σ_ϵ^2 is the variance of $\{\epsilon_t\}$, the purely random process of which $\{x_t\}$ is a transformation. Given a second linear model with an associated extrapolative function $\mu^*(z)$ and autoregressive representation $\alpha^*(z)x_t = \epsilon_t$, the forecast errors associated with the forecasts $x_{t+k}^{*\prime}$ generated by this second model have a variance

$$\text{(20)} \qquad \operatorname{var}(x_{t+k} - x_{t+k}^{*\prime}) = \sigma_\epsilon^2 \left[\textstyle\sum_{j=0}^{k-1} \beta_j^2 + \sum_{j=0}^{\infty} (\beta_j^* - \beta_{j+k})^2\right]$$

where $\beta^*(z) = \alpha^*(z)^{-1}\mu^*(z)$. The second summation on the right-hand side of (20), in practice summed over the number of terms actually computed for $\mu(z)$ and $\mu^*(z)$, represents the effect on the forecast error variance of the parameter misspecification which would result from using the second model or predictive scheme if in fact the first model is the correct specification of the generating mechanism.

For the Theil-Wage model, equation (20) can be used to examine the sensitivity of forecast errors to the disregarding of the seasonal component for different values of t and k. For the original Theil-Nerlove model (Cf. [17]), we have

$$\text{(21)} \qquad y_t = \Delta^2 x_t$$

with covariance function

$$\text{(22)} \qquad \begin{aligned} \phi_{yy}(j) &= (G^2 + 6)\sigma_w^2 && \text{for } j = 0 \\ &\quad -4\sigma_w^2 && \text{for } |j| = 1 \\ &\quad \sigma_w^2 && \text{for } |j| = 2 \\ &\quad 0 && \text{for } |j| > 2 \end{aligned}$$

and with an estimated autoregression function $\hat{A}(z)$ given by

$$\text{(23)} \qquad \textstyle\sum_{k=0}^{P} \hat{A}_k \phi_{yy}(j - k) = \delta_j \hat{\sigma}_{\epsilon'}^2, \qquad j = 0, \cdots, P,$$

with $\delta_j = 1$ if $j = 0$ and $\delta_j = 0$ otherwise and with P the same number as in (8). Consequently, substituting in (20), it is evident that the factor by which the error variance of the forecasts is increased by disregarding the seasonal component is

$$\text{(24)} \qquad f = \textstyle\sum_{j=0}^{M} (\hat{B}_{j+1} - \hat{C}_j)^2$$

where $\hat{C}(z) = \hat{a}(z)^{-1}J(z)$ and $\hat{B}(z) = \hat{A}(z)^{-1}$ and where M is the number of lags for which the inverse polynomials are computed. Programs can be written to provide estimates of f for different values of G and K.

This same analysis can easily be extended to other applications. For the stock prices forecasted by multiple smoothing in Brown and Meyer [5, pp. 681–682], for example, it is clear that substantial error variance is introduced by their triple-smoothing extrapolative mechanism if in fact stock prices are a random walk as is conventionally assumed. (Cf. [6].) The specific amount by which the error variance is increased is

$$\text{(25)} \qquad f = \textstyle\sum_1^{\infty} N_j^2$$

where $N(z)$ is the moving average specification for first differences of prices which is implicit in their triple-smoothing adaptive mechanism.

4. Effects of Non-Stationarity

The analytic procedures presented in Sections 2 and 3 depend heavily on the assumption that the underlying process $\{\epsilon_t\}$ is in fact stationary. If it is not stationary, then the estimation theorems relied upon in deriving extrapolative predictors are not applicable. Furthermore, if underlying stochastic variables are non-stationary, it is by no means clear that the type of parameter misspecification discussed in Section 3 for the Theil-Wage model will necessarily increase the error variance of forecasts, for it is possible that the greater simplicity of the model misspecified as to parameters may make it less sensitive to the non-stationarity.

To introduce non-stationarity into the Theil-Wage model, we may either posit periodic shifts in the means of the stochastic variables w_t, v_t, and z_t or occasional perturbations in the variances of these three variables. The former will result in non-stationarity in the mean of $Y_t = \Delta_L \Delta x_t$; the latter will result in non-stationarity of the covariance function $\phi_{YY}(j)$. It is likely that mean non-stationarity is more important in most applications to sales forecasting problems than covariance non-stationarity; in any case attention will here be restricted to cases of mean non-stationarity with covariance function both stationary and known.

For the Theil-Wage model, it will be assumed that the most likely form of mean non-stationarity is obtained by postulating occasional shifts in the expected value of w_t so that

$$w_t = \lambda_t + u_t, \tag{26}$$

with $\sigma_v/\sigma_u = G$, $\sigma_z/\sigma_u = K$, and $\mathcal{E}u_t = \mathcal{E}u_t u_{t-j} = 0$ for all t and $j \neq t$, where

$$\lambda_t = \lambda_i \tag{27}$$

for $t_i \leqq t < t_{i+1}$. It will be assumed that the t_i (the periods in which λ_i changes) are generated by a Poisson process with a rate parameter θ. Substituting in (1), it is evident that the effect on the Theil-Wage model of postulating (26) is to introduce intervals during which a company's sales tend to be systematically higher or lower than otherwise. (It may be possible to identify the sources of changes in λ_i with particular events such as the obtaining of an important contract, the launching of a new product, or a competitor's action. If so, such identification may yield the basis for useful procedures by which to incorporate a decision-maker's *a priori* predictions about such events into mechanical adaptive forecasting rules.)

Substituting (26) in (1), it is evident that u_t (the mean-adjusted value of w_t) is equivalent to w_t as defined in (1) so that (7) becomes

$$d(z)[Y_t - \Delta_L \Delta \lambda_t] = \epsilon_t \tag{28}$$

Thus, even though the process $\{Y_t\}$ is non-stationary as a result of the non-stationarity of $\mathcal{E}w_t$, it can be transformed into the stationary process

$$\{Y_t - \Delta_L \Delta\lambda_i\}$$

which has the same autoregressive representation as (5). This suggests that the Theil-Wage adaptive forecasting scheme presented above in (9)–(12) be changed by substituting

$$\xi_t^{\sim} = \xi_{t-1}^{\sim} + \eta_{t-1}^{\sim} + \alpha[x_t - \Delta\hat{\lambda}_t - x_t^*] \tag{29}$$

$$\eta_t^{\sim} = \eta_{t-1}^{\sim} + \alpha\beta[x_t - \Delta\hat{\lambda}_t - x_t^*] \tag{30}$$

$$\zeta_t^{\sim} = \zeta_{t-L}^{\sim} + (1 - \alpha)\gamma[x_t - \Delta\hat{\lambda}_t - x_t^*] \tag{31}$$

$$x_{t+1}^* = \xi_t^{\sim} + \hat{\lambda}_t + \eta_t^{\sim} + \zeta_{t-L+1}^{\sim} \tag{32}$$

for these equations, where $\hat{\lambda}_t$ is the best available estimate of λ_t. Substituting (29), (30), and (31) in (32) and collecting terms yields the extrapolative relationship

$$Y_{t+1}^* = J(z)[Y_t - \Delta_L \Delta\hat{\lambda}_t] + \hat{\lambda}_t(1 + [(1 - z^2)(1 - z^L)/z]_+) \tag{33}$$

which provides an asymptotically best linear predictor of Y_{t+1} if (17) is satisfied, and if $\hat{\lambda}_t$ is an efficient estimate of λ_t (and hence of λ_{t+1}).

For the case where θ is known and small (in the sense that θ^{-1}, the mean interval between shifts in λ_i, is larger than the number of realizations of Y_t required to estimate λ_i), efficient estimates of λ_i and λ_{i-1} have been derived by Hinich and Farley [11] using a modified form of a test statistic developed by Page [18]. The Hinich-Farley estimators are

$$\begin{aligned} \hat{\lambda}_i &= \hat{\lambda}_{i-1} + (6\theta^{-1}/N(N-1))[(2/(N+1))\textstyle\sum_{j=1}^N jx_j - \sum_{j=1}^N x_j] \\ \hat{\lambda}_{i-1} &= (2/N(N-1))[(2N+1)\textstyle\sum_{j=1}^N x_j - 3\sum_{j=1}^N jx_j] \end{aligned} \tag{34}$$

where N is the sample size. A test for defining the dating of t_i may be performed by computing sequential values of $\hat{\lambda}_i(t)$ and $\hat{\lambda}_{i-1}(t)$ for fixed N. Alternatively, expressions derived by Rice [20] and Hinich [10] for the asymptotic mean and variance of the number of zero-axis crossings of a stationary Gaussian process may be used as a detection device to test whether the number of zero-axis crossings has dropped sufficiently significantly to warrant the conclusion that the mean of the process has shifted.

Given estimates of θ, $J(z)$, and the variances of λ_i and of $\Delta\lambda_i$, we can define estimates of the effect on forecast error variance of not taking non-stationarity into account in specifying a forecasting rule. For the forecasts defined in (33), we have

$$\operatorname{var}(Y_{t+1}^* - Y_{t+1}) = C_0^2\sigma_\epsilon^2 + \theta \operatorname{var}(\lambda_i - \lambda_{i-1}) \tag{35}$$

where, as before, $C(z) = d(z)^{-1}J(z)$. The extent by which this error variance is increased by specifying the forecasting scheme discussed in Section 2 (i.e. by setting $Y_{t+1}^* = J(z)Y_t$) is

$$f = \operatorname{var}\lambda[16 + \textstyle\sum_{j=0}^\infty M_j^2]/(C_0^2\sigma_\epsilon^2 + \theta\operatorname{var}(\lambda_i - \lambda_{i-1})) \tag{36}$$

where $M(z) = d(z)^{-1}J(z)(1-z)^2(1-z^L)$. Furthermore, it is necessary to emphasize that this estimate of f is based on known values of $J(z)$. If, as in general, it is necessary to estimate $J(z)$, the estimation bias introduced by misspecification will result in an even greater error variance for the forecasts.

5. Conclusion

Much discussion of adaptive forecasting schemes has suffered from the fact that such schemes have too often not been analyzed in terms of the model implicitly assumed to generate the series being forecast. Though it is clear that the optimality of any linear predictive scheme such as an adaptive procedure is conditional upon the accuracy of the underlying specification, it is rare that the amount of forecast error likely to be introduced by misspecification of different types is explicitly analyzed.

In this paper two types of potential misspecifications have been discussed. For misspecification limited to incorrect specification of the number of non-zero parameters in the linear process generating the series, the error introduced by misspecification can be easily computed given estimates of the coefficients of the moving average representation of each model. These estimates can in turn be obtained by procedures which can in part be dated back to Wold's 1938 contribution [26] and which are summarized in Section 2.

For misspecification due to incorrect postulating of stationarity in the generating process, only partial results are presented here. For the case of mean nonstationarity with the rate of mean shifts known, some results derived by Hinich [11] have been used to supply estimates of parameters from which measures of the effect on forecast error of specifying complete wide-sense stationarity are derived.

The measurement of misspecification error can be important to a forecaster for several reasons. It enables a forecaster to evaluate the extent to which the cost savings introduced by use of a simple forecasting scheme is offset by greater inaccuracy of the resultant forecasts in cases where it is known that the simpler scheme is a misspecification of the underlying stochastic process. It similarly enables the analyst to pinpoint those types of likely misspecification to which the accuracy of the forecasts is likely to be most sensitive, and so to indicate the areas in which improved specification may be necessary in order to reduce forecast error to a given tolerated level. Beyond this, estimates of the effect of likely forms of misspecification on forecast accuracy may be used to increase estimates of forecast error derived from standard estimating procedures. Such "blowup factors" can be useful in obtaining improved estimates of error variance for use in subsequent analysis of managerial decisions. Though adaptive forecasting and detection mechanisms have widespread potential applicability—cf. for instance, the striking empirical success of the Jakowatz filter [13]—their usefulness can be enhanced by analysis of the nature of the associated forecast errors.

REFERENCES

1. Alexander, Sidney S., "Price Movements in Speculative Markets: Trends or Random Walks," *Industrial Management Review*, 2 (May 1961), pp. 7–26.

2. Bossons, John, "Some Evidence on the Validity of the Adaptive Expectations Hypotheses," *Capital Markets Workshop Working Paper 13*, Carnegie Institute of Technology, February 1966.
3. Box, G. E. P. and Jenkins, G. M., "Some Statistical Aspects of Adaptive Optimization and Control," *Journal of the Royal Statistical Society* (*Series B*), 24 (1962), pp. 297–331.
4. Brown, Robert G. *Statistical Forecasting for Inventory Control*, New York: McGraw-Hill, 1959.
5. —— and Meyer, R. F., "The Fundamental Theorem of Exponential Smoothing," *Operations Research*, 9 (September–October, 1961), pp. 673–687.
6. Cootner, Paul (ed.), *The Random Character of Stock Market Prices*, Cambridge, Mass.: M.I.T. Press, 1964.
7. Cox, D. R., "Prediction by Exponentially-Weighted Moving Averages and Related Methods," *Journal of the Royal Statistical Society* (*Series B*), 23 (1961), pp. 414–422.
8. Durbin, J., "Efficient Estimation of Parameters in Moving-Average Models," *Biometrika*, 46 (December 1959), pp. 306–316.
9. ——, "The Fitting of Time Series Models," *Revue de l'Institut International de Statistique*, 28 (December 1960), pp. 233–243.
10. Hinich, Melvin J., "Estimation of Spectral Characteristics for a Class of Non-Stationary Time Series," *Management Sciences Report No. 34*, Carnegie Institute of Technology, February 1965.
11. ——, and Farley; J. U., "Theory and Application of an Estimation Model for Time Series with Nonstationary Means," *Management Science*, forthcoming.
12. Holt, Charles C., "Forecasting Trends and Seasonals by Exponentially Weighted Moving Averages," *O.N.R. Memorandum No. 52*, Carnegie Institute of Technology, April 1957, dittoed.
13. Jakowatz, C., Shuey, R. and White, J., "Adaptive Waveform Recognition," *Fourth International Symposium on Information Theory*, London, 1960.
14. Leenders, C. T., "Adaptive Forecasting under Simple Probabilistic Assumptions," Mimeo, January 1964.
15. Magee, John F., *Production Planning and Inventory Control*, New York: McGraw-Hill, 1958.
16. Muth, John F., "Optimal Properties of Exponentially Weighted Forecasts of Time Series with Permanent and Transitory Components," *Journal of the American Statistical Association*, 55 (June 1960), pp. 299–306.
17. Nerlove, Marc and Wage, S., "On the Optimality of Adaptive Forecasting," *Management Science*, 10 (January 1964), pp. 207–224.
18. Page, E., "A Test for a Change in a Parameter Occurring at an Unknown Point," *Biometrika*, 42 (December 1955), pp. 523–527.
19. Phillips, A. W., "Analysis, Estimation, and Control of Linear Systems," dittoed, December 1962.
20. Rice, S. O., "Mathematical Analysis of Random Noise," in N. Wax (ed.), *Selected Papers on Noise and Stochastic Processes*, New York: Dover, 1947, pp. 189–213.
21. Simon, Herbert A. and Bonini, Charles P., "The Size Distribution of Business Firms," *American Economic Review*, 48 (September 1958), pp. 607–617.
22. Theil, Henri and Wage, S., "Some Observations on Adaptive Forecasting," *Management Science*, 10 (January 1964), pp. 198–206.
23. Whittle, Peter, *Prediction and Regulation*, Princeton: Van Nostrand, 1964.
24. Winters, Peter R., "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 6 (April 1960), pp. 324–342.
25. ——. Chapter 14 of Charles C. Holt, Franco Modigliani, John F. Muth, and Herbert A. Simon, *Planning Production, Inventories, and Work Force* (Englewood Cliffs, N. J.: Prentice-Hall, 1960).
26. Wold, Herman, *A Study in the Analysis of Stationary Time Series*, Stockholm: Almquist and Wiksell, 1938.