Predictive Ability and Descriptive Validity of Earnings Forecasting Models
Author(s): Benoit Deschamps and Dileep R. Mehta
Source: *The Journal of Finance*, Vol. 35, No. 4 (Sep., 1980), pp. 933-949
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2327211

THE JOURNAL OF FINANCE • VOL. XXXV, NO. 4 • SEPTEMBER 1980

Predictive Ability and Descriptive Validity of Earnings Forecasting Models

BENOÎT DESCHAMPS and DILEEP R. MEHTA*

I. Introduction

VALIDATION OF SECURITY VALUATION models critically depends on obtaining reliable proxies for investors' expectations of future earnings. It is obvious that different values of these proxies in a given case are likely to lead to different, if not conflicting, results. In spite of significance of properly formulated proxies, they have not been paid requisite attention in the valuation studies. Nor, for that matter, have financial theories offered proper guidelines for selecting and processing relevant information for devising proxies for future earnings. This situation is not due to any oversight or neglect, but has arisen because of substantial problems involved, for instance, in identifying and measuring relevant factors for developing forecasts. Even when these forecasts are strictly based on earnings history, selection of an appropriate extrapolation method still remains a controversial and somewhat uncomfortable issue, as evidenced by inconsistent conclusions from studies conducted by Ball and Watts [3], Brooks and Buckmaster [6], and Foster [9]. The purpose of this study is to examine effectiveness of different extrapolation methods previously advanced in the literature.

In Section II, extrapolation models previously advanced in the literature are classified into four major hypotheses on earnings time-series. Section III deals with relevant data and specification of forecasting horizons for conducting empirical investigation. Section IV develops the criteria for the selection process. In Section V, empirical evidence on the performance of alternative processes is presented, and different evaluation measures are employed in order to establish whether a clearcut dominance of a specific process emerges. The final section concludes with major findings of this study.

II. Alternative Earnings Generating Process Hypotheses

One major approach traditionally utilized suggests that future earnings growth is the estimated slope of log-linear regression of the form of:

$$z_t = b_0 + b_1T + a_t \tag{1}$$

* Ecol des Hautes Etudes Commerciales, Montréal and Georgia State University, respectively. Research for this paper was supported by the Canada Council, the Ecole des Hautes Etudes Commerciales de Montréal, and the Georgia State University.

where: z_t is the logarithm of earnings per share;[1]
T is a time index;
a_t is a white noise random disturbance.

The constant growth model (1) is valid when earnings follow a steady and deterministic path over time. Lintner and Glauber [13] questioned this model and suggested that earnings changes are largely random. This hypothesis was more thoroughly investigated by Ball and Watts [3]. Based on results of several statistical tests, they concluded that the behavior of earnings over time is best described by a submartingale or drifting random walk process:

$$z_t = z_{t-1} + \theta_0 + a_t \tag{2}$$

where: θ_0 is a drift parameter.

The principal difference between the constant growth model (1) and the submartingale model (2) lies in the treatment of white noise random shocks a_t. In the constant growth model, the effect of a_t is strictly transitory; its impact is limited to contemporaneous earnings values. In the submartingale process, all future earnings values are affected by the random shocks a_t, which thus have a permanent effect; the ordered history of earnings is, hence, irrelevant for predictive purposes; and the best earnings forecast is current earnings adjusted for a drift.

An issue, first emphasized by Ball and Watts [3], is whether an earnings series can be viewed as a mixture of these two extreme models, whereby the random shocks a_t have a permanent as well as a transitory impact on future earnings. On annual earnings series, the exponentially weighted moving average (EWMA) model such as:

$$z_t - z_{t-1} = \theta_0 + (1 - \lambda)a_{t-1} + \mathrm{a}_t \tag{3}$$

has frequently been employed (without the drift θ_0). If the smoothing constant, λ, takes the value unity, equation (3) reduces to the submartingale model, while it describes the constant growth model when λ is zero [with θ_0 corresponding to b_1 in equation (1)].

A key problem of earlier studies was the length of earnings series. These series consisted of as few as 10 annual observations in some cases, and seldom more than 20 (in order to avoid likelihood of nonstationarity).[2] The resultant large sampling errors would likely favor acceptance of the null hypothesis. Of course, quarterly earnings data in place of the annual series can augment the number of observations. However, the pervasive seasonality of quarterly data, as noted by Lorek, McDonald and Patz [14], requires reformulation of the above three hypotheses without destroying their basic features. Thus, a seasonal adjustment for the constant growth model calls for a seasonal intercept $b_0(s)$ in equation (1),

[1] The model is based on the logarithm of $Z_t = ce^{gt}$, with $\ln(Z_t) = \ln(c) + gt$. A logarithmic transformation of raw earnings data has been utilized throughout this paper.

[2] Analysis of the annual earnings series for the period 1908–1974 led Watts and Leftwich [22] to conclude that misspecification due to non-stationarity of the annual series tilted the scale in favor of naive models such as random walk.

i.e.,

$$z_t = b_0(s) + b_1(T) + a_t \tag{4}$$

which requires specification of the quarterly intercepts represented by four dummy variables and a time index. When the assumptions of the ordinary least squares method are met, an estimate of the slope b_1 is an unbiased value of the deseasonalized quarterly growth rate in earnings per share.

A similar adjustment in the submartingale model (2) incorporates seasonality in the quarterly drift factor δ_i (the average change from the ith quarter to the next one), whereby:

$$z_{t-i} - z_{t-i-1} = \delta_i + a_{t-i} \quad \text{for} \quad i = 1, 2, 3, 4 \tag{5a}$$

and

$$\delta_1 + \delta_2 + \delta_3 + \delta_4 = \theta_0 \tag{5b}$$

An alternative formulation is obtained by differencing the quarterly series over an interval of one year rather than one quarter (as in (5)):

$$z_{t-1} - z_{t-i-4} = \theta_0 + e_{t-i} \tag{6}$$

where

$$e_{t-i} \equiv a_{t-i} + a_{t-i-1} + a_{t-i-2} + a_{t-i-3}$$

Even when a series is a true submartingale, the alternative formulations (5) and (6) would not lead to identical conclusions. To see this, let equations (5a) and (6) be adjusted one quarter forward:

$$z_{t-i+1} - z_{t-i} = \delta_{i+1} + a_{t-i+1} \tag{5a$'$}$$

and

$$z_{t-i+1} - z_{t-i-3} = \theta_0 + e_{t-i+1} \tag{6$'$}$$

Given the submartingale character of the series, the random shocks a_{t-i} and a_{t-i+1} in (5a) and (5a′) will not be autocorrelated. In contrast, e_{t-i} and e_{t-i+1} in (6) and (6′) will have an overlap of three quarterly random shocks, a_{t-i}, a_{t-i-1}, and a_{t-i-2}. Hence, the disturbances e_t cannot be assumed serially independent.[3] In fact, the

[3] A model similar to (6) has been widely used in the literature; see, for instance, Foster [9], Joy et al [12], and Brown and Rozeff [7]. Because these studies have apparently not recognized the spurious autocorrelation created by (6) and reflected in (8), their results cannot be taken at face value to dismiss the hypothesis that the true generating process is well approximated by the submartingale model. Another model sharing the similar characteristics of serially dependent disturbances even when a true process is submartingale is obtained by differencing (6), whereby:

$$(z_{t-\iota} - z_{t-\iota-4}) - (z_{t-\iota-1} - z_{t-\iota-5}) = e'_{t-i} \tag{7}$$

where

$$e'_{t-\iota} = a_{t-\iota} - a_{t-\iota-4}$$

model (6) will exhibit an autocorrelation structure:

$$\rho k = \begin{cases} (0.75)/k & \text{for} \quad k \leq 3 \\ 0 & \text{for} \quad k \leq 3 \end{cases} \tag{8}$$

where ρk is the autocorrelation between $(z_{t-i} - z_{t-i-4})$ and $(z_{t-i-k} - z_{t-i-k-4})$.

Of course, once the spurious autocorrelation is properly recognized, models (5) and (6) should lead to identical inferences. Indeed, the deseasonalized drift parameter θ_0 in (6) allows us to use augmented observations of quarterly data more effectively than is possible in (5).[4] The same argument also applies to the constant growth model (4). Expressing (4) in terms of annual differences yields:

$$z_t - z_{t-4} = \sum_{s=1}^{4} b_0(s) + a_t - a_{t-4} \tag{9}$$

or, when B is defined as a lag operator such that $B^k z_t = z_{t-k}$,

$$(1 - B^4)\, z_t = \theta_0 + (1 - \theta_4 B^4) a_t \tag{10}$$

where

$$\theta_0 = \sum_{s=1}^{4} b_0(s) \quad \text{and} \quad \theta_4 = 1.0$$

Now the deseasonalized submartingale (6) can be well approximated by:

$$(1 - \phi B)(1 - B^4) z_t = \theta_0 + a_t \tag{11}$$

because, for $\phi \simeq 0.65$, it has an autocorrelation structure

$$\rho_k = \phi^k$$

that is strikingly similar to the one generated by (6) for $k \leq 3$.

Inspection of (10) and (11) shows that both these models can be easily accommodated in:

$$(1 - \phi B)(1 - B^4) z_t = \theta_0 + (1 - \theta_4 B^4) a_t \tag{12}$$

This model is a deseasonalized counterpart of EWMA for quarterly series, or a Mixture of Constant Growth and Submartingale (MCGS), where the autoregressive parameter ϕ registers the permanent impact of random shocks a_t, and the moving average parameter θ_4 reflects their transitory effect. In fact, the MCGS is much more flexible than EWMA, since it accommodates a series other than a weighted average of the constant growth and the submartingale models. Thus, four situations are pertinent in the present instance:[5]

[4] Alternatively, there will be a larger number of variables to estimate in (5), which would lead to a larger loss of degrees of freedom and thereby reduce the reliability of statistical results.

[5] Griffin [11] had some success with the model $(1 - B)(1 - B^4)z_t = (1 - \theta_1 B)(1 - \theta_4 B^4)a_t$. If expressed into another fashion, this model could approximate the MCGS:

$$(1 + \theta_1 B + \theta_1^2 B^2 + \theta_1^3 B^3 + \cdots)(1 - B)(1 - B^4) z_t = (1 - \theta_4 B^4) a_t, \quad \text{for} \quad \theta_1 B < 1$$

or:

$$[1 - (1 - \theta_1)B - \theta_1(1 - \theta_1^2)B^2 - \theta_1(1 - \theta_1^3)B^3 - \cdots](1 - B^4) z_t = (1 - \theta_4 B^4) a_t$$

Defining $\phi = (1 - \theta_1)$, if the terms in B with higher power than 1 in $[\cdot]$ can be ignored, the model reduces to the MCGS (see also footnote 3).

1) When $\phi = 0.0$ and $\theta_4 = 1.0$, the MCGS represents the constant growth model (10).
ii) When $\theta_4 = 0.0$, the MCGS reduces to the autoregressive model (11) which, in turn, approximates the submartingale model (6) for $\phi = 0.65$.
iii) When the estimated values of ϕ and θ_4 fulfill the relationship

$$\phi = 0.65x \tag{13a}$$

$$\theta_4 = 1.00(1 - x) \tag{13b}$$

for $0 < x < 1$, the MCGS represents a weighted average of the constant growth (10) and submartingale (6) models.
iv) When both parameters are nonzero, $\phi \neq 0.65$, and $\theta_4 \neq 1$, the MCGS is inconsistent with all of the three hypotheses previously advanced in the literature.

In brief, the application of the MCGS model will permit a cross check on validity of the two extremes or their mixture.

Along with the three hypotheses formalized in equations (10), (11), and (12), a fourth method, developed by Box and Jenkins [4] has been recently studied on earnings time-series (for instance, by Foster [9], Griffin [11], Albrecht et al. [1], Brown and Rozeff [7], and Watts and Leftwich [21]). The class of models Box and Jenkins (BJ) proposed are called integrated autoregressive moving averages (ARIMA):

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d z^t = \theta_0 + (1 - \theta_1 B - \cdots - \theta_q B^q)a_t \tag{14a}$$

$$\theta_p(B)(1 - B)^d z^t = \theta_0 + \theta_q(B)a_t \tag{14b}$$

where $\phi_p(B)$ is an autoregressive polynomial of order p in the lag operator B, $(1 - B)^d$ represents successive differences of order d, and $\theta_q(B)$ is a moving average polynomial of order q in the lag operator B. To simplify the notation, only differences on adjacent observations are specified in equation (14), although this is often not the case. In fact, the ARIMA models encompass as special cases the models (10), (11), and (12) except that the methodology requires selecting a particular model on the basis of sample data, and this process is cumbersome and expensive for ARIMA.[6] Indeed, recent evidence concerning earnings forecast suggests that simpler models may be more effective than the ARIMA.[7] Nevertheless, this evidence is ambiguous on at least three counts:

a) Only short forecasting horizons were examined. As Elton and Gruber noted [8], the performance of naively specified extrapolations deteriorates with the length of the lead time.
b) When two or more base periods were employed (which was seldom done), parameter estimates were not revised to reflect potential for gradual changes

[6] Whether costs of data gathering and processing outweigh the benefit of using an appropriate specification that would not be reasonably approximated by simpler (and cheaper) processes cannot be established a priori.

[7] However, Granger and Newbold [10, section 5–6] show that overall the ARIMA has proven extremely successful in predicting a wide range of economic time series.

in the generating process (see, for instance, Foster's reporting [9, p. 13] of poor performance of ARIMA models that were estimated only once).

c) Statistical tests were inadequate. Sometimes the models were assessed solely on the basis of their "fit" rather than their predictive ability on a holdout sample.

III. Data Selection and Model Estimation

A. *Sources of Data and Selection of Time-Horizons*

The data consist of quarterly earnings per share for periods ranging from 11 to 16 years (1961–1976) for 91 electric utility firms.[8] The major source of data is the Value Line Data Bases tapes (see Arnold Bernhard & Co. [2]), which has been supplemented with various issues of the Value Line Investment Service.

For the Box-Jenkins modeling methodology, a rule of thumb requires at least 50 observations.[9] However, since it is desirable to assess the accuracy of forecasts by considering more than one forecasting origins, and since data availability was limited, the forecasting models were estimated with 40 quarterly observations. This procedure, in turn, implies that the earliest time origin from which the forecasts can be made is the end of 1970, leaving a maximum of 6 years over which the accuracy of the earnings forecasts can be ascertained.

Another problem lies in the choice of a sample period during which the parameters of a forecasting model may vary over time to reflect changes in the underlying earnings generating process. When this results in only a marginal modification of the parameters of a statistical model constructed to approximate the true process, the problem can be solved by re-estimating the model at regular time-intervals. In other situations, however, it may be necessary to re-specify the model on the basis of new information. Hence, it was decided to re-specify and re-estimate, at the firm level, the ARIMA and other three models for each forecasting time origin.

Six forecasting time origins were selected, allowing evaluation of the accuracy of the forecasting models on the basis of six yearly forecasts, the first at the end of 1970 covering the four quarters of 1971, the second at the end of 1971, and so on, until the sixth at the end of 1975 for the 1976 earnings. At each base period, only the most recent 40 quarterly earnings were utilized to estimate the models. The number of firms with an available 10-year record of quarterly earnings varied for each forecasting base period:

1970:	79 firms	1973:	88 firms
1971:	82 firms	1974:	91 firms
1972:	87 firms	1975:	91 firms

[8] The list of firms included in the sample is available upon request from the authors.

[9] Box and Jenkins [4, p. 33] suggested this rule on the basis that the largest lag k at which an estimate of an autocorrelation function is computed should not exceed the number of observations N divided by 4, i.e., $k < N/4$. Preliminary investigation revealed that autocorrelation after lag 8 seldom exceeded 0.15 for the quarterly earnings series, which is not significant in light of the fact that their standard error is approximated by $(1/N^{1/2})$.

With these numbers of firms for each forecasting origin, it is possible to evaluate the accuracy of 518 one-year ahead earnings forecasts. When longer lead times are considered, the number of point estimates is reduced. For instance, when the lead time is two years, the last available forecasting origin is the year-end of 1974, and hence, the number of two-year lead time forecasts that can be investigated is the number of one-year forecasts, less 91. Thus the number of matched observations that can serve to investigate the predictive ability of the models is as follows:

One-year lead time:	518 matched observations
Two-year lead time:	427 matched observations
Three-year lead time:	336 matched observations
Four-year lead time:	248 matched observations
Five-year lead time:	161 matched observations

B. Estimation of the Models

In estimating the models, when a logarithmic transformation of earnings was not possible due to negative earnings per share for a given period, it was arbitrarily decided to add the amount of the maximum loss encountered, plus 10¢, to the specific earnings series. This procedure resulted in a linear shift of the earnings series toward positive values.

The process of identifying and estimating the ARIMA models was not straightforward, because outliers were found especially during the aftermath of the credit market contraction in 1969 as well as during the energy crisis of 1974. Because outliers could not be eliminated in time-series modeling, identification and estimation of a parsimonious model became very difficult.[10]

Several general characteristics of the 518 ARIMA specifications may be noted:

a) In nearly all instances, differencing was required to achieve stationarity. As only one level of differencing was generally necessary, the dominant nonstationary factor appears to be a linear trend.
b) In all but one cases, the differencing required was at the seasonal level. Because autocorrelations of nonseasonal differences usually exhibited nonsystematic patterns or positive and negative estimates that did not dampen quickly, they were considered either nonstationary or stationary but unsusceptible to being modeled parsimoniously.
c) In almost every instance, the order of the parameters (power of the lag operator B) is not higher than 6, implying that recent values of the series are much more important in determining future earnings levels than old ones.
d) The models have a parsimonious parametrization, requiring (apart from one level of differencing) one trend parameter and never more than three autoregressive and/or moving average parameters.

[10] The iterative procedure was undertaken at least five times for each of the 518 firm-time specifications.

e) There is no indication that a single specification is an appropriate representation for all series. A comparison across firms did not reveal any common denominator except seasonal differencing and presence of a trend. More often than not, however, the structure of the identified model resembles the MCGS process.
f) Comparison of the selected models for the same firm with different forecasting origins, reveals that non-stationarity is a plausible phenomenon. In spite of changes in the models over time, their theoretical autocorrelations still shared a common pattern, a feature which was reflected in the powers of the lag operator B specified in various models for the same firm.[11]

With respect to the MCGS model (12), examination of estimated parameters may provide tentative evidence on the validity of the three competing hypotheses on the underlying earnings generating process. Table I shows the deciles distribution of each of the parameter estimates for the MCGS model.[12]

Estimates reported in Table I are inconsistent with the constant growth as well as the submartingale models, since the estimated values of ϕ and θ_4 are substantially different from their theoretical values under these two processes, as indicated in Section II. In contrast, because the mean estimates are $\bar{\phi} = 0.32$ and $\bar{\theta}_4 = 0.54$, or about half of the hypothesized values for the constant growth and the submartingale, the parameter estimates are consistent with the viewpoint that a weighted average of these two processes is a valid representation (on the aggregate) of the earnings generating process.[13]

IV. Performance Criteria

The rational expectations theory developed by Muth [17, 18], Nerlove [19], and Mincer [15] states that expectations are rational if they fully incorporate all of the scarce information available at the time of the forecast. Although different processes may be justified on the basis of the rational expectations theory—and we believe that the Box-Jenkins methodology comes closest[14]—, a direct implementation of this theory is not feasible to date. Nevertheless, the desirable properties of mathematical surrogates for rational expectations may be identified as follows.[15] First, in generating unbiased one-period ahead forecasts, information contained in the realizations of the series must be efficiently utilized; and second, the same information must be consistently applied in making multi-span forecasts.

[11] This is not surprising given the data overlap at different sample periods.

[12] A firm need not be represented in the same decile column for each of its three parameters.

[13] Because the mean estimated value of θ_4 is above unity in the highest decile, estimated MCGS models are sometimes non-invertible, i.e., they do not have a convergent autoregressive representation. Problems, mainly for estimation, posed by non-invertibility are discussed by Box and Jenkins [4, section 6.4] and Granger and Newbold [10, section 4.9].

[14] At the firm level, specification of ARIMA constitutes an optimal extraction of information contained in the historical series because residuals of the estimated model must be free of any systematic patterns. In contrast, the other three models cannot provide proper surrogates of rational earnings expectations unless they are a very close replication of the known stochastic structure of the underlying earnings generating process.

[15] This classification is suggested by Pesando [20].

Table I

Summary of the Distribution of the Parameter Estimates for the MCGS Model

$$(1 - \emptyset B)(1 - B^4)z_t = \theta_0 + (1 - \theta_4 B^4)a_t$$

Parameter	Sample Mean	Mean Estimate for the Decile 1	2	3	4	5	6	7	8	9	10
$\emptyset$	.32	−.09	.07	.15	.22	.28	.34	.41	.48	.57	.74
θ_0	.027	−.011	.005	.011	.017	.021	.027	.035	.046	.055	.070
θ_4	.54	−.08	.23	.35	.44	.52	.59	.66	.75	.87	1.07

Forecasts are efficient when one-period ahead predictions share a common stochastic pattern with the realizations of the series; and multi-period forecasts are consistent when they are obtained recursively. Thus, when forecasts are efficient and consistent, they have validity in the context of the rational expectations theory.

In this light, it is clear that the forecasting methodology proposed by Box and Jenkins represents a structured search for an efficient forecasting model, and consistency is then achieved by utilizing this model recursively.

V. Comparison of the Models

A. *Predictive Ability*

In order to avoid discrepancies due to the fact that the earnings per share for firms differ in magnitude, proportional forecasting errors were considered:

$$_iu_T(l) = \frac{_iz_{t+l} - {_iF_T(l)}}{_iz_{t+l}} \tag{15}$$

where $_iu_T(l)$ is the lead-time-l forecasting error made in predicting the realized (untransformed) annual earnings $_iz_{t+l}$ of firm i from the time origin t, and $_iF_T(l)$ is the extrapolation provided by a specific model.

Table II contains results on the ex-post accuracy of the four alternative models in predicting future earnings for the five lead-time periods. Several observations may be made on Table II:

a) Performance of the submartingale model is inferior to that of the others. Most often, for the submartingale model, the proportions of forecasts falling within the tolerance limits are lower than those for the other models.
b) There appears to be no clear dominance among the constant growth, the ARIMA and the MCGS predictors.
c) The accuracy of the forecasting methods deteriorates as the lead time increases. This is not surprising since theoretically the variance of the forecasting errors is positively correlated with the length of the forecasting horizon (Box and Jenkins, [4]).
d) The proportions of the forecasts that fall within the 10% tolerance limits tend to stablize beginning with the 3-year lead time.

In general, the performance of the models is not very impressive, especially for

Table II

Sample Proportions of the Predictions with Forecasting Errors within 5%, 10% and 15% Tolerance Limits

Lead Times	Tolerance Limits	Model 1 Constant Growth	Model 2 Submartingale	Model 3 ARIMA	Model 4 MCGS
1-Year	$-5\% < u < 5\%$	.322	.214	.295	.293
($N = 518$)	$-10\% < u < 10\%$	.529	.392	.515	.521
	$-15\% < u < 15\%$	.666	.519	.666	.689
2-Year	$-5\% < u < 5\%$	.251	.189	.237	.255
($N = 417$)	$-10\% < u < 10\%$	.445	.311	.426	.433
	$-15\% < u < 15\%$	.602	.424	.583	.571
3-Year	$-5\% < u < 5\%$	.167	.182	.170	.185
($N = 336$)	$-10\% < u < 10\%$	.345	.271	.324	.339
	$-15\% < u < 15\%$	.503	.387	.488	.491
4-Year	$-5\% < u < 5\%$	.157	.145	.153	.165
($N = 248$)	$-10\% < u < 10\%$	.306	.234	.323	.315
	$-15\% < u < 15\%$	.456	.359	.444	.472
5-Year	$-5\% < u < 5\%$	.155	.124	.124	.186
($N = 161$)	$-10\% < u < 10\%$	.286	.255	.261	.286
	$-15\% < u < 15\%$	.379	.348	.360	.367

lead time of five years. It is not possible, however, to determine the extent of deterioration attributable to unpredictable cross-sectional factors such as the energy crisis.

A rigorous assessment of the predictive ability of the different models requires an explicit consideration of the loss an investor is expected to incur for using a wrong forecasting method. It is impossible to specify accurately such a "cost function," but a sound procedure should be relatively insensitive to small changes in its specification of the cost function.[16] A commonly employed cost function is a quadratic transformation of forecasting errors u:

$$L(u) = cu^2 \tag{16}$$

where L (*) is the cost function, and c, a positive constant. The quadratic cost criterion (16) leads to selection of a forecasting model having the minimum mean squared forecasting errors. Potential biases are avoided by computing the mean in relative terms. Thus the performance measure

$$E(u_i^2 - u_j^2)$$

reflects the expected difference in the cost of errors related to the use of two forecasting techniques i and j across the sample firms. This procedure has an additional advantage of eliminating from the analysis the effect of economy-wide random shocks that are unpredictable, but which may have a contemporaneous impact on the firm's earnings.

If technique i dominates technique j for every firm, all differences $(u_i^2 - u_j^2)$ will be negative. Such a clear-cut situation is unlikely to occur; in general, some

[16] Granger and Newbold [10, pp. 279–280] discuss cost of error functions and their specification.

of the differences would be negative, others positive, and the resultant sample mean $E(u_i^2 - u_j^2)$ different from zero. Whether this mean is nonzero can be statistically assessed in the following fashion. Because of the Central Limit Theorem (Mood and Graybill [16]), the sample mean is normally distributed, no matter what is the original distribution, with a mean equal to the mean difference, and a variance equal to the variance of the differences divided by the number of observations. Therefore, a statistical test:

$$T = \frac{E(u_i^2 - u_j^2) - 0.0}{\sigma(u_i^2 - u_j^2)/N^{1/2}} \tag{17}$$

can be performed in order to determine if one of two techniques i and j dominates the other. The results for the test statistics (17) are presented in Table III, for each of the 5 different lead times.

Two observations are pertinent on the results reported in Table III:

a) The submartingale is systematically dominated by all of the other models.
b) In spite of dominance over the submartingale, no clearcut dominance emerges among the remaining three approaches for all lead times.

An alternative cost function based on absolute forecasting error was employed to test the sensitivity of results reported in Table III. The essentially similar results obtained with the alternative cost function suggested that additional specifications are not likely to be meaningful.

B. *Descriptive Validity*

In the past, performance of forecasting models has been evaluated solely in terms of their predictive ability, i.e., through a comparison with the realized values of the variable in question.[17] In the present instance, the predictive ability of a model has an economic meaning in the context of the rational expectations theory whereby predictors are distributed identically with the values of the economic variable being predicted. But the rational expectations theory also requires that forecasts incorporate all of the relevant information contained in historically observed values. Hence, superior predictive ability is a necessary but not sufficient criterion for dominance.

The criterion of incorporating all relevant information signifies in the present context that time-series residuals of the model have a zero mean and are serially uncorrelated. The statistical models utilized in this study, in all four instances, explicitly require that the mean residual is (close to, if not exactly) zero. Only the ARIMA model systematically considers the autocorrelation requirement. Table IV summarizes the distribution of the serial correlation of the residuals from the four models. Examination of the reported mean estimates reveals that the residuals of the constant growth model are substantially and positively autocorrelated at lag one. Similarly, the residuals of the submartingale model exhibit an

[17] A serious shortcoming of some of the previous studies (e.g., Ball and Watts [3, p. 675]) is that they did not have a validation sample, and attempted to establish a model's ability to predict values which were already utilized in estimating its parameters.

Table III

Dominance among the Models for Predictions at Various Lead Times: T-Statistics on Mean Differences

	Lead Time = 1 Year $N = 518$			Lead Time = 2 Years $N = 427$			Lead Time = 3 Years $N = 336$			Lead Time = 4 Years $N = 248$			Lead Time = 5 Years $N = 161$			
	Model			Model			Model			Model			Model			
	2	3	4	2	3	4	2	3	4	2	3	4	2	3	4	
Model 1	−1.37	0.83	0.53	−5.25	−1.53	−2.35	−2.05	1.02	−0.14	−0.95	1.72	0.82	−1.53	1.78	0.64	1
Model 2		1.66	3.05		5.72	4.30		3.49	2.67		2.43	1.48		2.67	2.08	2
Model 3			0.25			−2.01			−1.86			−1.21			−0.96	3

Explanatory notes:

a) the models are:
Model 1: Constant Growth
Model 2: Submartingale
Model 3: ARIMA
Model 4: MCGS

b) The test statistics are normal variates. The 1%, 5% and 10% critical values (two-tail) are 2.58, 1.96 and 1.64.

c) The tests are always performed between the lowest coordinate and the highest coordinate. For instance, for the 2-year lead time, the normal statistic T measuring the relative accuracy of model 2 (submartingale) against model 3 (ARIMA),

$$T = \frac{E(u_2^2 - u_3^2)}{\sigma(u_2^2 - u_3^2)/N^{1/2}} = 5.72$$

indicates that model 2 forecasting errors are larger than model 3's, and hence that model 2 is dominated by model 3.

Table IV

Residual Autocorrelations

Model	Decile	Distribution of Autocorrelations at Lag											
		1	2	3	4	5	6	7	8	9	10	11	12
1. Constant Growth	Lowest	−.11	−.38	−.30	−.23	−.32	−.32	−.32	−.31	−.28	−.26	−.31	−.35
	Highest	.63	.45	.36	.44	.28	.21	.18	.16	.14	.16	.17	.11
	Mean	.25	.04	.04	.13	−.01	−.06	−.06	−.05	−.07	−.07	−.07	−.11
2. Submartingale	Lowest	−.56	−.51	−.36	−.26	−.33	−.30	−.27	−.25	−.24	−.24	−.24	−.27
	Highest	.02	.24	.27	.47	.26	.23	.28	.25	.25	.22	.29	.11
	Mean	−.29	−.15	−.06	.14	−.04	−.04	−.01	.01	.01	−.01	.03	−.06
3. MCGS	Lowest	−.15	−.32	−.30	−.14	−.29	−.25	−.25	−.22	−.22	−.26	−.23	−.22
	Highest	.16	.24	.29	.17	.27	.23	.21	.19	.17	.18	.19	.19
	Mean	.00	−.04	.00	.01	.00	−.02	−.03	−.02	−.02	−.04	−.02	−.01
4. ARIMA	Lowest	−.12	−.16	−.18	−.18	−.18	−.19	−.21	−.24	−.20	−.24	−.22	−.21
	Highest	.13	.18	.17	.11	.16	.17	.20	.17	.18	.15	.17	.18
	Mean	.01	.01	.00	−.03	−.01	−.01	−.02	−.03	−.03	−.03	−.02	−.01

Table V

Q-Statistics for Residuals Autocorrelations

	Models			
Decile	Constant Growth	Submartingale	MCGS[c]	ARIMA
1	5.19	6.95	2.37	1.92
2	7.91	9.80	3.64	2.81
3	9.74	11.73	4.48	3.35
4	11.43	13.68	5.15	3.79
5	13.37[a]	15.18[b]	5.92	4.24
6	15.55[a]	17.13[b]	6.70	4.78
7	18.22[a]	19.34[b]	7.67	5.27
8	22.39[a]	22.03[b]	9.02	5.80
9	29.44[a]	25.75[b]	10.67	6.54
10	46.12[a]	36.55[b]	15.23	7.92
Mean	17.91	17.80	7.08	4.64

[a] Significant at 10%; critical value of 13.36.
[b] Significant at 10%; critical value of 14.68.
[c] Significant at 10% if the statistic exceeds the critical value of 15.99.

important negative serial correlation at lag one, but the mean autocorrelations for the MCGS model are about zero. This evidence further substantiates the indications given in Section III that the true underlying earnings generating process is a mixture of the constant growth and submartingale models. But an individual lag correlation is subject to considerable sampling errors, especially when small sample sizes are utilized. A more efficient approach, suggested by Box and Pierce [5], is to perform a statistical test of the hypothesis that an entire sequence of lag correlations be estimated on purely random data (white noise):

$$Q = n \sum_{k=1}^{m} r_k^2 \tag{21}$$

where n is the number of residuals, r_k is the estimated autocorrelation at lag k, and m is large enough to assure that the residuals are truly uncorrelated at higher lags. In the present instance, m may be reasonably assumed to be 12.

The Q-statistic follows approximately a chi-square distribution with $(m - p)$ degrees of freedom, where p is the number of estimated parameters excluding a constant such as θ_0 in (10).[18] Using 12 autocorrelations of the residuals, a Q-statistic was computed for each of the 518 firm-base periods from each of the four models.[19] The decile distribution of these Q-statistics is presented in Table V. Data in Table V indicate that in more than 60% of the cases, one cannot accept the hypothesis concerning validity of either the constant growth or the submar-

[18] This is because the Q-statistics which are derived from autocorrelations are unaffected by the level of a series.

[19] Two degrees of freedom are lost when the MCGS is employed. For the constant growth and the submartingale models, the number of degrees of freedom lost is the number of parameters estimated (including seasonal drifts or intercepts) minus one since the seasonal factors can be expressed as deviations from an average level or drift.

tingale models. On the other hand, the hypothesis that the MCGS model is consistent with the underlying earnings generating process cannot be rejected.

VI. Conclusion

In this paper, two conventional extrapolation methods, the constant growth and the submartingale, were examined. The third model investigated was specified at the firm level in accordance with the time series modeling methodology developed by Box and Jenkins. In order to enhance statistical reliability of the results and, at the same time, to avoid nonstationarity biases of long annual series, quarterly earnings series were employed. Potential presence of seasonality in quarterly series, however, required reformulation of the first two models, and, more important, rendered the exponentially weighted moving average (EWMA) model ineffectual. Consequently, a mixture of constant growth and submartingale (MCGS) process was devised in the place of EWMA. The MCGS is similar to the EWMA in that it allows an a priori verification scheme of validating the two extrapolation methods. The MCGS is, however, more robust than the EWMA in that, apart from allowing the use of seasonal data, a) it permits the drift parameter, and b) it does not constrain parameters to be strictly weighted. On the basis of sample results, this study shows that the underlying earnings generating process is on the average consistent with a weighted average of the submartingale and constant growth processes.

In order to avoid potential ambiguities and sample-specific results, forecasting horizon was stretched to as much as 5 years. Further, more than one base periods were utilized, and all models were re-estimated for these base periods. Finally, a comprehensive battery of tests based on not only statistical (descriptive validity) but also economic (predictive ability) considerations—rather than any ad hoc basis—were employed.[20]

The submartingale model was initially suggested for the description of earnings series on purely empirical grounds (Lintner and Glauber [13]) as an alternative to the constant growth model. Results of this study show that the submartingale is decidedly inferior to the alternative models.[21] It is worth noting that this lack of empirical support for the submartingale has no implication on the efficiency of the capital market, although knowledge of the valid process is required to test whether stock prices reflect the information contained in realized earnings. Indeed, the submartingale model has no theoretical justification for earnings that it has for economic sequences such as interest rates (see Pesando [21]) that are determined by the forces of supply and demand in the competitive capital markets.

In spite of the dominance by the other three models over the submartingale process individually, none of them consistently exhibited superior predictive

[20] Results of additional tests, such as analysis of ranking of the models on the basis of the magnitude of forecasting errors and Wilcoxon Matched-Pairs Signed Rank Test were consistent with results reported in this paper, and have not been reported due to space limitations.

[21] Although the scope of our investigation was confined to the electrical utility industry, it does constitute a major part of actively traded firms on the U.S. stock exchanges.

performance. Further, the predictive ability of all models was rather mediocre and it deteriorated (albeit at a decreasing rate) over longer horizons. Ambiguity concerning ranking of the three models other than the submartingale is resolved to a major extent when the criterion of efficient processing of historical information is used: it revealed that the constant growth (as well as the submartingale) model did not incorporate systematic patterns of historical earnings.

As for the remaining two models, test results were more favorable (but only marginally) to the ARIMA than the MCGS. Conceptually, the ARIMA embodies a superior methodology. Operationally, the MCGS has an edge over the ARIMA due to simplicity in its application. Selection of one of these two models thus involves a trade-off between the cost of implementation (especially when forecasts are needed for a large number of firms) and the risk that essentially a mechanical process may not correspond to the underlying earnings generating process.

REFERENCES

1. W. Steve Albrecht, Larry Lookabill and James C. McKeown. "The Time-Series Properties of Annual Earnings." *Journal of Accounting Research,* (Autumn 1977).
2. Arnold Bernhard & Co., Inc.: *The Value Line Data Bases* (*A User's Manual*), New York (1976).
3. Ray Ball and Ross Watts. "Some Time-Series Properties of Accounting Income." *Journal of Finance,* (June 1972).
4. George E. P. Box and Gwilyn M. Jenkins. *Time-Series Analysis*: *Forecasting and Control,* Rev. Ed. (San Francisco: Holden-Day 1976).
5. ——— and David A. Pierce. "Distribution of Residual Autocorrelation in Autoregressive-Integrated Moving Average Time-Series Models." *Journal of the American Statistical Association,* (December 1970).
6. Leroy D. Brooks and Dale A. Buckmaster. "Further Evidence of the Time-Series Properties of Accounting Income." *Journal of Finance,* (December 1976).
7. Lawrence D. Brown and Michael S. Rozeff. "The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings." *Journal of Finance,* (March 1978).
8. Edwin J. Elton and Martin J. Gruber. "Earnings Estimates and the Accuracy of Expectational Data." *Management Science,* (April 1972).
9. George Foster. "Quarterly Accounting Data: Time-Series Properties and Predictive-Ability Results." *The Accounting Review,* (January 1970).
10. Clive W. J. Granger and Paul Newbold. *Forecasting Economic Time-Series,* (New York; Academic Press 1977).
11. Paul A. Griffin. "The Time-Series Behavior of Quarterly Earnings. Preliminary Evidence." *Journal of Accounting Research,* (Spring 1977).
12. O. Maurice Joy, Robert H. Litzenberger and Richard W. McEnally. "The Adjustment of Stock Prices to Announcements of Unanticipated Changes in Quarterly Earnings." *Journal of Accounting Research,* (Autumn 1977).
13. John Lintner and Robert Glauber. "Higgledy Piggledy Growth in America," Paper presented to the Seminar on the Analysis of Security Prices, 1967, University of Chicago, Reprinted in *Modern Developments in Investment Management,* edited by James Lorie and Richard Brealey, 1972 (New York: Praeger Publishers; 1972).
14. Kenneth S. Lorek, Charles L. McDonald and Dennis H. Patz. "A Comparative Examination of Management Forecasts and Box-Jenkins Forecasts of Earnings." *The Accounting Review,* (April 1976).
15. Jacob Mincer. "Models of Adaptive Forecasting." *In Economic Forecasts and Expectations—Analysis of Forecasting Behavior and Performance,* J. Mincer Ed., (New York NBER and Columbia University Press, 1969).
16. Alexander M. Mood and Franklin Graybill. *Introduction to the Theory of Statistics,* 2nd edition (New York McGraw-Hill Book Co., 1963).

17. John F. Muth. "Optimal Properties of Exponentially Weighted Forecasts." *Journal of the American Statistical Association*, (June 1960).
18. ———. "Rational Expectations and the Theory of Price Movements." *Econometrica*, (July 1961).
19. Marc Nerlove. "Distributed Lags and Unobserved Components in Economic Time-Series." In *Ten Economic Studies in the Tradition of Irwin Fisher*, W. Fellner et al. Eds (New York: John Wiley and Sons, 1967).
20. James E. Pesando, "Rational Expectations and Distributed Lag Expectations Proxies." *Journal of the American Statistical Association*, (March 1976).
21. ———. "On the Efficiency of the Bond Market: Some Canadian Evidence." *Journal of Political Economy*, (December 1978).
22. Ross L. Watts and Richard W. Leftwich. "The Time-Series of Annual Accounting Earnings." *Journal of Accounting Research*, (Autumn 1977).