Evaluation of Time-Series Models Using Predictive Values of Quarterly Earnings

*Lee Kyung-joo**

Contents

Ⅰ. Introduction

A large number of studies in the accounting literature have evaluated different time-series models of corporate earnings in an attempt to identify the model that best describes the earnings generating process. However, this attempt has not been very successful for the models of quarterly earnings because each of the Griffin [1977] and Watts [1975], Foster [1977], and Brown and Rozeff [1979] models has received some support in the previous studies. There are several different methods which are used in the literature to evaluate time-series models. The within sample methods, such as minimizing the Akaike Information Criterion (AIC), the Schwartz Bayesian Criterion (SBC) or the residual

* Department of Accounting, College of Economics & Commerce

variance, generally result in choosing the model which "best characterizes" the historical path of the data, where the different criteria use different definitions of "best characterizes". The out of sample methods, such as minimizing the Mean Squared Forecast Error (MSFE) or minimizing the out of sample AIC or SBC, generally result in choosing the model which "best forecasts". For example, the MSFE uses the square of the difference between the forecast and the actual value as a measure of how good a forecast is, while the AIC and SBC use the MSFE with different penalties imposed for overly elaborate models (or rewards for parsimonious models) as a measure of how good a forecast is.[1]

Clearly, however, these different evaluation criteria are not independent and should usually result in similar conclusions regarding the selection of a model. But because they each focus on different aspects of the models, they will not always result in identical conclusions. For example, an overly parameterized model would be rejected by the AIC and SBC but not by the MSFE criterion while overly parsimonious models tend to be rejected by AIC and MSFE but not by SBC. In a study by Dharan [1983], the within sample AIC criterion led to a selection of the Brown and Rozeff [1979] (100)×(011) model (the BR model) while the Foster [1977] (100)×(010) model (the F model) was chosen by the MSFE criterion.[2] Furthermore, mixed results have been reported in previous studies which employed the MSFE. For example, Collins and Hopwood [1980] and Bathke and Lorek [1984] have found an evidence supporting the BR model. In contrast, Benston and Watts [1978] and Lorek [1979] have provided results in favor of the F model and the Griffin [1979] and Watts [1975] (011)×(011) model (the GW model), respectively.

These mixed results suggest that sometimes one criterion is unable to distinguish between models. Therefore, we propose an alternative method of evaluating time-series models of quarterly earnings. Our method utilizes the predictive values of quarterly earn-

---

1) In addition to the methods described here, there is an additional method which evaluates the models in terms of the association between earnings forecast errors and unexpected changes in stock prices. Using this criterion, the model which generates the forecast errors most highly associated with stock price changes is considered to be the best model utilized by he market.

2) The three 'premier' models of quarterly earnings identified in the literature are described as a multiplication of regular and seasonal components, and can be parsimoniously designated as (pdq)×(PDQ) using the Box and Jenkins [1976] notation. The p, d, and q refer to the number of regular autoregressive, differencing and moving average terms, respectively. The P, D, and Q are the respective seasonal counterparts.

ings for improving the forcasts of annual earnings.[3] As demonstrated in the next section, particular time-series models of quarterly earnings imply particular theoretical predictive values, and we can use these theoretical predictive values to construct an alternative way of measuring how good a quarterly model is. In particular, a model for which there is little difference between the theoretical predictive values (TPV) and the observed or empirical predictive values (EPV) would be preferable to a model for which these values diverge substantially. The difference between these two values can be viewed as an out of sample measure of model misspecification, which makes this a reasonable criterion by which to evaluate a model. In a sense, then, the model which "best forecasts" according to this criterion is the one for which the observed reduction in forecast variance is closest to what it should be if the model were true.

Additional motivation for this study is provided by the lack of analytical work as to how the quarterly earnings reports lead to improvement in the forecasts of annual earnings. As noted by Ball and Foster (1982, p.123), most of previous studies have addressed this issue "almost exclusively at an empirical level with limited analysis of the conditions under which improvement in forecasting would (or would not) be expected". In this study, we demonstrate analytically that for a specific class of quarterly models, quarterly earnings reports always improve annual earnings forecasts and the degree of improvement (predictive value) is determined by the parameter value of the given quarterly earnings model.

Using a sample of 235 firms over 5 year period from 1980 to 1984, our empirical tests evaluate three time-series models of quarterly earnings (the BR, F and GW model). The result shows that our criterion selects the BR model over either the F model or the GW model. This finding is generally consistent with the results from the within sample methods (AIC and residual variance) and the MSFE criterion. However, our proposed method makes these rankings clearer.

The remainder of this paper is organized as follows. In the next section, we derive the theoretical relationship between time-series properties (parameter value) and the predictive values of quarterly earnings. Our research design including sample selection and model comparison procedure is described in Section 3. The results of the empirical analysis are

3) The term 'predictive value' is defined in this paper as the improvement in the forecasts of annual earnings from incorporating the first, second and/or third quarter's realized earnings over the forecasts made at the beginning of the year. Thus, the 'predictive value' and 'improvement' are used interchangeably.

presented in Section 4, and concluding remarks appear in the final section.

Ⅱ. Time-Series Properties and Predictive Values of Quarterly Earnings[4]

Consider five points in time during a year. The time 0 is the beginning of the fiscal year, while time-periods 1,2,3 and 4 correspond to the dates when actual earnings are available for the first, second, third and fourth quarter, respectively. Note that time 4 is the end of the fiscal period when annual earnings become known. At each point in time (0,1,2,3), the forecast of a firm's annual earnings (A) conditional on the available quarterly report (Q_τ) is defined as :

$$E(\tilde{A}|Q_\tau)=\sum_{t=1}^{4}E(\tilde{Q}_t|Q_\tau)$$

where E is an expectation operator, ~ denotes random variable, and $\tilde{Q}_t|Q_\tau$ is the earnings for quarter $t(=1,2,3,4)$ conditional on $\tau(=0,1,2,3)$ quarter's actual earnings. For example, $E(\tilde{Q}_3|Q_0)$, $E(\tilde{Q}_3|Q_1)$ and $E(\tilde{Q}_3|Q_2)$ are the forecast of the third quarter earnings conditional on zero, one and two quarterly reports, respectively. Note that for $t\leq\tau, E(\tilde{Q}_t|Q_\tau)=Q_t$ since actual quarterly earnings are known at these points in time.

To assess the degree of improvement in forecasting annual earnings from incorporating the quarterly reports, the variance of annual forecast error conditional on each quarterly earnings is compared with that generated before the release of the quarter's earnings.[5] Denoting V_τ as the variance of the forecast error for annual earnings conditional on τ quarter's earnings, we have :

$$V_\tau=\mathrm{var}[\tilde{A}-E(\tilde{A}|Q_\tau)]$$

4) The analysis is similar to that of Barnea, Dyckman and Magee [1972], but differs in the following way : First, they did not employ a time-series model governing the quarterly earnings, and second they examined the predictive value of the first quarter's earnings only.

5) The selection of forecast error variance to measure the improvement in forecast is based on (1) the popularity of its use in empirical literature as a forecast error metric, and (2) the assumption of quadratic loss function. Barnea, Dyckman and Magee [1972] also used the forecast error variance.

$$=\text{var}[\sum_{t=1}^{4}(\tilde{Q}_t-E(\tilde{Q}_t|Q_\tau))]$$

$$=\sum_{t=1}^{4}\text{var}[\tilde{Q}_t-E(\tilde{Q}_t|Q_\tau)]+\sum_{s}^{4}\sum_{\neq t}^{4}\text{cov}[(\tilde{Q}_s-E(\tilde{Q}_s|Q_\tau)),(\tilde{Q}_t-E(\tilde{Q}_t|Q_\tau))] \cdots\cdots (1)$$

Equation (1) shows that at the beginning of the year, the variance of annual earnings forecast error (V_0) consists of (ⅰ) error variances of one, two, three and four quarter ahead forecasts, and (ⅱ) their covariances. At the end of the first quarter, the release of actual earnings would lead to reduction in the error variance, i.e., more accurate forecast, for the forthcoming annual earnings. The improved forecast is due to (ⅰ) the substitution of the realized first quarterly earnings for their predicted values, and (ⅱ) the revision in the forecasts of the remaining three quarters.[6] In subsequent time periods, as additional quarterly earnings become available, the forecast accuracy would increase substantially compared with that at the beginning of the year.

To see how the time series behavior of quarterly earnings affects their predictive value for forecasting annual earnings, we begin by noting that any ARIMA process can be written as an infinite weighted sum of current and previous shocks a_j :

$$Q_{\tau+t}=a_{\tau+t}+\pi_1 a_{\tau+t-1}+\pi_2 a_{\tau+t-2}+\cdots\cdots$$

where a_τ is an independently and identically distributed 'white noise' process with mean zero and constant variance (σ^2), and π_j ($j=0,1,2,\cdots$) is the linear filter relating $Q_{\tau+t}$ to $a_{\tau+t}$. Note that if the process is stationary then the infinite series $\sum\pi_j$ will converge. Accordingly, a forecast $Q_\tau(t)$ of $Q_{\tau+t}$, which is said to be made at origin τ for lead time t, can be expressed as :

$$Q_\tau(t)=\pi_t a_t+\pi_{t+1}a_{\tau-1}+\cdots\cdots$$

The t-step ahead forecast error $e_\tau(t)$ is then defined as :

$$e_\tau(t)=Q_{\tau+t}-Q_\tau(t)$$
$$=\sum_{j=0}^{t-1}\pi_j a_{\tau+t-j}$$

where $\pi_0=1$. For the purpose of this study, $e_\tau(t)$ can be interpreted as the error of t-

6) Abdel-Khalik and Espejo [1978] provide an adaptive model which includes both "substitution effect" and "revision effect". See also Brown and Rozeff [1979] for the empirical evidence of separate contribution of each effect to the improvement in forecasting annual earnings.

step ahead quarterly earnings forecast conditional on τ quarter's realized earnings, i.e., $e_\tau(t) = [Q_t - E(\tilde{Q}_t | Q_\tau)]$ in terms of previous notation. Note that τ and t are bounded by [0,3] and [1,4], respectively and $t \le 4-\tau$.

Using the above definition of forecast error and rewriting equation (1), we get:

$$V_\tau = var[\sum_{t=1}^{4-\tau} e_\tau(t)]$$
$$= \sum_{t=1}^{4-\tau} var[e_\tau(t)] + 2\sum_{s=1}^{3-\tau}\sum_{t>s}^{4-\tau} cov[e_\tau(s), e_\tau(t)] \cdots\cdots (2)$$

The variance of and covariance between forecast errors of same origin but different lead times are shown to be as follows (Box and Jenkins [1976], p.160):

$$var[e_\tau(t)] = \sigma^2 \sum_{h=1}^{t} \pi_{h-1}^2$$

and

$$cov[e_\tau(s), e_\tau(t)] = \sigma^2 \sum_{i=0}^{s-1} \pi_i \pi_{i+t-s}$$

where $t>s$ and $\pi_0 = 1$. By substituting these into (2), the forecast error variance of annual earnings conditional on τ quarter's earnings can be written as:

$$V_\tau = [\sum_{t=1}^{4-\tau}\sum_{h=1}^{t} \pi_{h-1}^2 + 2\sum_{s=1}^{3-\tau}\sum_{t>s}^{4-\tau}\sum_{i=0}^{s-1} \pi_i \pi_{i+t-s}]\sigma^2 \quad , \tau = 0,1,2,3 \cdots\cdots (3)$$

The following two points should be noted for equation (3). First, only $\pi_j (j=0,\cdots,3)$ are relevant because the maximum forecast horizon is 4 quarters within a year. Second, error variance of the forecast made after the third quarter's earnings (V_3) is constant (σ^2) and hence not affected by σ_j for any j.

An assumption is made regarding the time-series model. We assume that the time-series process of quarterly earnings is described by single parameter in its regular and seasonal component with maximum of first-order differencing in regular series.[7] Under this

7) In a box and Jenkins [1976] notation of (pdq)×(PDQ), this assumption requires that quarterly earnings processes belong to the class of the models with (1) $p+q \le 1$, (2) $P+Q \le 1$, and (3) $d \le 1$. Note that no restriction is imposed on the seasonal differencing (D). All the models suggested by Foster [1977] (100)×(010), Griffin [1977] and Watts [1975] (011)×(011), and Brown and Rozeff [1977] (100)×(011) belong to this class. Thus, the assumption does not appear to be an unrealistic one.

assumption, all π_j $(1 \le j \le 3)$ can be expressed as a function of the parameter Π (either AR or MA) of the regular component. To see this, first note that the class of seasonal time-series models assumed may be written in the following general form :

$$\phi(B)\phi_s(B^4)(1-B)^d(1-B^4)^D Q_t = \theta_0 + \theta(B)\theta_s(B^4)a_t,$$

where $\phi(B)=1-\phi B$, $\phi_s(B^4)=1-\phi_s B^4$, $\theta(B)=1-\theta B$, and $\theta_s(B^4)=1-\theta_s B^4$; ϕ and ϕ_s are regular and seasonal AR parameter, respectively; θ and θ_s are regular and seasonal MA parameter, respectively; θ_0 is a constant term; and B is a backward shift operator such that $B^k Q_t = Q_{t-k}$. By ignoring the constant term, this form can be rewritten as :

$$\begin{aligned}(1-B)^d(1-B^4)^D Q_t &= \phi(B)^{-1}\phi_s(B^4)^{-1}\theta(B)\theta_s(B^4)a_t \\ &= (1+\phi B+\phi^2B^2+\cdots)(1+\phi_s B^4+\phi_s^2 B^8+\cdots)(1-\theta B)(1-\theta_s B^4)a_t\end{aligned}$$

Since the π_j's for this study are relevant only three periods back (a_{t-3}), we have the following form for the non-differenced series by ignoring the terms associated with π_j for $j \ge 4$:

$$(1-B)^d(1-B^4)^D Q_t = a_t + (\phi-\theta)a_{t-1} + \phi(\phi-\theta)a_{t-2} + \phi^2(\phi-\theta)a_{t-3} + \cdots$$

Depending on (i) the order of regular defferencing (d) and (ii) whether AR or MA process is considered, we have the following four types of models :[8]

Model 1 : when $d=0$ and $\theta=0$,

$Q_t = a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 a_{t-3}$, and $\pi_j = \phi^j$ for $j=1,2,3$.

Model 2 : when $d=0$ and $\phi=0$,

$Q_t = a_t - \theta a_{t-1}$, and $\pi = -\theta$ for $j=1$: $\pi_j = 0$ for $j=2,3$.

Model 3 : when $d=1$ and $\theta=0$,

$Q_t = a_t + (1+\phi)a_{t-1} + (1+\phi+\phi^2)a_{t-2} + (1+\phi^2+\phi^3)a_{t-3}$, and $\pi_j = 1+\sum_{s=1}^{j}\phi^s$ for $j=1,2,3$.

Model 4 : when $d=1$ and $\phi=0$,

$Q_t = a_t + (1-\theta)a_{t-1} + (1-\theta)a_{t-2} + (1-\theta)a_{t-3}$, and $\pi_j = 1-\theta$ for $j=1,2,3$.

It should be emphasized that for all the above models, π_j are expressed as a function of the regular parameter Π (either ϕ or θ) and lag j,i.e., $\pi_j = f(\Pi, j)$. Furthermore, note

8) The orders of seasonal differencing (D) and seasonal component (ϕ_s and θ_s) are irrelevant because they habe no impact on a_{t-j} and thus on π_j for $j \epsilon [0,3]$.

that Brown and Rozeff [1979] and Foster [1977] models are special cases of Model 1, while Griffin [1977] and Watts [1975] models belong to Model 4.

Two important results arise from equation (3). First, ***V_τ is a decreasing function of τ.*** That is, $V_0>V_1>V_2>V_3$ or $dV_\tau/d\tau<0$, which indicates that quarterly reports always improve the forecast of annual earnings if quarterly earnings are generated by the class of models assumed. Alternatively stated, quarterly reports have predictive value for forecasting annual earnings. This result is intuitively apparent because the forecasts become more accurate as the forecast horizon $(4-\tau)$ decreases. This is also consistent with empirical evidence that the accuracy of annual earnings forecast increases as the end of the year approaches (e.g., Lorek [1979] and Collins and Hopwood [1980]).

Proof: If $\pi_j\geq0$, it is obvious from equation (3) that V_τ is always decreasing in τ. Note that $\pi_j=1+\sum_{s=1}^{j}\phi^s>0$ (Model 3) and $\pi_j=1-\theta>0$ (Model 4) because $|\phi|<1$ and $|\theta|<1$ from the stationarity and invertability condition, respectively. Hence, $dV_\tau/d\tau<0$ for the Model 3 and Model 4. Next, consider Model 1 when the AR parameter ϕ is negative so that $\pi_j<0$ for some j. Using the relation $\pi_j=\phi^j$, we have : $V_0-V_1=(1+\phi)^2(1+\phi^2)^2>0$; $V_1-V_2=\phi^2+(1+\phi)^2(1+\phi^2))>0$; and $V_2-V_3=(1+\phi)^2>0$. Hence, $V_0>V_1>V_2>V_3$. Finally, the comparisons among V_τ's for the Model 2 give $V_0-V_1=V_1-V_2=V_2-V_3=(1-\theta)^2>0$, showing $dV_\tau/d\tau<0$.

Second, ***the predictive values of quarterly earnings for forecasting annual earnings are a function of the parameter Π of a given model.*** To see this, note that the contribution of the first quarter's report to the increased forecast accuracy can be measured by the difference between the forecast error variances conditional on zero and one quarter's earnings, V_0-V_1. Likewise, the contribution of the second (third) quarter over the first (second) quarter is measured by V_1-V_2 (V_2-V_3). Using equation (3), the ***total improvement***(TI), relative to the forecast error at the beginning of the year (V_0), during a year in the forecast of annual earnings from incorporating realized quarterly earnings can be defined as :

$$TI=(V_0-V_3)/V_0$$
$$=[\sum_{t=2}^{4}\sum_{h=1}^{t}\pi_{h-1}^2+2\sum_{s=1}^{3}\sum_{t>s}^{4}\sum_{i=0}^{s-1}\pi_i\pi_{i+t-s}]\ /V_0 \quad\cdots\cdots\cdots (4)$$

Since the ***relative improvement*** (RI) contributed by each quarter τ (=1,2,3) to the total improvement can be defined as $RI(Q_\tau)=(V_{\tau-1}-V_\tau)/(V_0-V_3)$, we have from equation (3) :

$$RI(Q_1)=[\sum_{h=1}^{4}\pi_{h-1}^2+2\sum_{s=1}^{3}\sum_{i=0}^{s-1}\pi_i\pi_{i+4-s}]/(V_0-V_3)\quad\cdots\cdots\cdots (5a)$$

and $$RI(Q_2) = [\sum_{h=1}^{3}\pi_{h-1}^2 + 2\sum_{s=1}^{2}\sum_{i=0}^{s-1}\pi_i\pi_{i+3-s}]/(V_0-V_3) \quad (5b)$$

$$RI(Q_3) = [\sum_{h=1}^{2}\pi_{h-1}^2 + 2\pi_1]/(V_0-V_3) \quad (5c)$$

The predictive values (both TI and RI) of quarterly earnings as defined above clearly show that they are a function of π_i's. Since π_i's are uniquely determined by the parameter value (θ or ϕ), the predictive values are a function of the parameter value of a given model. Therefore, for a model and its parameter value, the theoretical predictive values can be obtained from equations (4), (5a), (5b) and (5c). For example, the theoretical TI for the BR and F models is $(3+6\phi+7\phi^2+6\phi^3+4\phi^4+2\phi^5+\phi^6)/(4+6\phi+7\phi^2+6\phi^3+4\phi^4+2\phi^5+\phi^6)$, while TI for the GW model is given by $(3+12\pi+14\pi^2)/(4+12\pi+14\pi^2)$ where $\pi=(1-\theta)$. Also, $RI(Q_1)$ for the BR and F models is $(1+2\phi+3\phi^2+4\phi^3+3\phi^4+2\phi^5+\phi^6)/(3+6\phi+7\phi^2+6\phi^3+4\phi^4+2\phi^5+\phi^6)$ and $RI(Q_1)$ for the GW model is $(1+6\pi+9\pi^2)/(3+12\pi+14\pi^2)$.

Ⅲ. Research Design

1. Sample Selection

Our sample consists of 235 COMPUSTAT-CRSP firms which satisfy the following selection criteria : (1) quarterly earnings per share (EPS) data is available in the *Value Line Investment Survey* over the period 1967~1984; (2) each firm has a fiscal year ending on December throughout the period 1967~1984; and (3) each firm must be in the manufacturing industry with two-digit SIC code between 10 and 39.

The first criterion is used to have enough EPS data for estimating the time-series models by the BJ methodology. The second and third criteria are imposed to ensure the comparability of earnings series across firms. The firms in the regulated industries such as Banking, Utilities and Transportation are excluded because they may have earnings processes quite different from the manufacturing firms. As is typical with time-series research in accounting, the familiar 'survivorship bias' applies to the sample because it includes only those firms that have existed for at least 18 years.

The above selection criteria yielded a sample of 235 firms. Table 1 shows the breakdown of the sample firms by industry (two-digit SIC code). Twenty three industries are represented in the sample. There is clustering in particular industries, notably Chemicals (SIC=28) and Electric Machinery (SIC=36), which account for 15.7% and 13.6% respectively, of the sample firms.

2. Measuring Predictive Values of Quarterly Earnings

For a given time-series model (the BR, F, or GW model), the estimated parameter values will determine the *theoretical* predictive values (TPV). TPVs were obtained by substituting the estimated parameter values for the true ones (ϕ or θ) in the functional forms specified in the Section 2. This procedure was repeated for each model, each firm, and each year over the period from 1980 to 1984. Each time-series model was estimated initially using 52 quarters' EPS data (1967~1979) in order to obtain the parameter value estimates for the year 1980. The use of 52 observations is based on the suggestion by Box

Table 1. Industry Classifications of Sample Firms

Two-Digit SIC Code	Industry Description	Number of Firms
10	Metal Mining	9
12	Coal Mining	3
13	Oil and Gas Extraction	5
14	Nonmetal Mineral	1
16	Heavy Construction	2
20	Food and Kindred	10
21	Tobacco	3
22	Textile Mill	3
24	Lumber and Wood	2
25	Furniture and Fixtures	2
26	Paper	11
27	Printing and Publishing	7
28	Chemicals	37
29	Petroleum Refining	18
30	Rubber	7
32	Stone, Clay and Glass	11
33	Primary Metal	15
34	Fabricated Metal	9
35	Industrial Machinery	21
36	Electric Machinery	32
37	Transportation Equipment	19
38	Instruments	7
39	Miscellaneous Goods	1
	Total	235

and Jenkins [1976, p.18] that at least 50 observations should be used to estimate a preliminary model.[9] The reestimation procedure was then used by adding additional four quarters' data to the data base to estimate parameter values for the year 1981 through 1984.

Empirical (observed) predictive values (EPV) were measured using annual earnings forecast errors conditional on the available quarterly reports. As a forecast error metric, we used squared forecast error (SFE) which is specified as :[10]

$$SFE(Q_\tau)_{iy} = (A_{iy} - E(A|Q_\tau)_{iy})^2$$

where

A_{iy} = actual annual earnings for firm i and year y, and

$E(A|Q_\tau)_{iy}$ = forecasted annual earnings conditional on τ quarter's earnings for firm i and year y, $\tau=0,1,2,3$.

Closely following the definitions in Section 2, the total improvement (TI) during a year relative to the beginning of the year in the accuracy of annual earnings forecasts from incorporating realized quarterly earnings is measured by:

$$TI_{iy} = [SFE(Q_0)_{iy} - SFE(Q_3)_{iy}]/SFE(Q_0)_{iy}$$

Similarly, the relative improvement (RI) in the forecast accuracy contributed by each quarter is measured by:

$$RI(Q_\tau)_{iy} = \frac{SFE(Q_{\tau-1})_{iy} - SFE(Q_\tau)_{iy}}{SFE(Q_0)_{iy} - SFE(Q_3)_{iy}}, \qquad \tau=1,2,3$$

The forecasts of annual earnings at the end of each quarter $E(A|Q_\tau)$ are obtained by summing the remaining quarterly forecasts of the year and the actual earnings of current and previous quarters. The quarterly earnings forecasts (one through up to four-quarter-

9) See Lorek and McKeown [1978] for a preliminary evidence regarding the effect of number of observations on the predictive ability. In general, they found a significant increase in predictive ability as the number of observations increases.

10) Although the use of SFE as a forecast error metric is consistent with the analytical results in Section 2, absolute forecast error (AFE) metric was also used (1) to examine the sensitivity of the results to different measures of forecast error, and (2) to be comparable with previous studies which employed this measure. The results using AFE were not materially different from those reported.

ahead forecasts) for the year 1980 are generated by estimating each time-series model using 52 quarters' EPS data (1967~1979). Each model is reestimated by adding the first quarter's actual earnings to the data base in order to generate one through three-quarter-ahead forecasts for 1980. This reestimation procedure continues by sequentially including the most recent actual earnings data until the one-quarter-ahead forecast for the fourth quarter of 1984 is generated.[11]

The theoretical (TPV) and empirical (EPV) predictive values were then averaged over 5 year period to get corresponding mean values for each firm and each model :

$$MTPV_{im} = 1/5 \sum_{y=80}^{84} TPV_{imy}$$

and

$$MEPV_{im} = 1/5 \sum_{y=80}^{84} EPV_{imy}$$

where

$MTPV_{im}$ = mean theoretical predictive values,

$MEPV_{im}$ = mean empirical predictive values,

i = firm index, i = 1, ···, 235,

m = time-series model index, m = BR, F, or GW, and

y = year index, y = 1980, ···, 1984.

3. Model Comparison Procedure

The primary objective of this study is to evaluate three time-series models of quarterly earnings (the BR, F, and GW models) by comparing across the models how the empirical predictive values are close to the theoretical predictive values. The absolute deviation (DPV_{im}) of the theoretical predictive values from corresponding empirical values was used to measure how close the empirical predictive values are to the theoretical values :

$$DPV_{im} = |MTPV_{im} - MEPV_{im}|$$

The magnitude of DPV_{im} represents the extent of misspecification of model m for firm i;

11) The rational for adopting the reestimation procedure over the simpler adaptive forecasting technique is due to the empirical evidence suggesting the superiority of reestimation over adaptive forecasting in predictive ability (McKeown and Lorek [1978]).

the model is more misspecified, the larger is the magnitude of DPV[illegible]. Therefore, this measure can be used to evaluate different time-series models using the following procedure. First, for a given firm, three time-series models were ranked using the magnitude of DPV : a rank of one was assigned to the model yielding the smallest DPV, while a rank of three was assigned to the model yielding the largest DPV.

We then utilized these ranks to test whether there is a significant difference in the DPVs across three time-series models.[12] The Friedman test was used because (1) the test involves comparison of DPVs from different time-series model for the same firms (related sample) and (2) the distributional properties of the variable DPV are unknown. The Friedman test is a nonparametric version of ANOVA (Conover [1980]) and has been utilized in the time-series literature (e.g., Bathke and Lorek [1984] and Brown, Hagerman, Griffin and Zmijewski [1987]). Finally, if the Friedman test rejects the null hypothesis of no difference in DPVs across the three models, the matched-paired t-test based on the ranks was conducted to test the significance of difference in each model pair and to determine which model is the "best" one in terms of its specification.

IV. Empirical Results

Table 2 presents the results of evaluating three quarterly earnings models based on their predictive values, using squared percentage forecast errors. Panel A of Table 2 summarizes the average rank and mean DPV for each model and for both total (TI) and relative (RI) predictive values. The F-statistics from the Friedman test reveal that the null hypothesis can be rejected in favor of the alternative hypothesis at the α level of 0.001 when TI was used. Except for RI(Q_1), the null hypothesis was also rejected for RI(Q_2) and RI(Q_3) with α level of 0.10 and 0.001, respectively. This result suggests that there is a significant difference across three premier models.

Panel B of Table 2 provides the t-statistics and the associated levels of significance for the multiple comparisons of three time-series models. For TI, all the pairwise comparisons were statistically significant. The significance levels of two-tail tests for (BR, F), (BR,

12) Formally, the null and alternative hypotheses can be stated as :

$H_0 : E(DPV_{BR}) = E(DPV_F) = E(DPV_{GW})$

$H_a : E(DPV_j) \neq E(DPV_k)$ for any $j \neq k \varepsilon BR, F,$ or GW

where E denotes an expectation operator.

GW) and (GW, F) pairs were 0.001, 0.002 and 0.023, respectively. This result indicates that the BR model has the smallest specification error, while the F model is the most misspecified one. Although the significance levels are not high, the BR model is still better than the F or GW models when the comparisons are based on RI. Overall, these results indicate that the BR model is the best one (i.e., the least misspecified model) under our evaluation criterion. This result is also consistent with the evidence in previous studies based on the within sample method (Brown and Rozeff [1979a] and Dharan [1983]) and the out of sample method (Collins and Hopwood [1980] and Bathke and Lorek [1984]). To investigate whether this consistency holds for the sample used in this study, both methods were applied to our sample.

Table 2. Comparisons of Time-Series Models Based on Predictive Values[a]

Panel A. Summary Statistics and Overall Comparisons[b]

	TI		$RI(Q_1)$		$RI(Q_2)$		$RI(Q_3)$	
Model	Mean Rank	Mean DPV	Mean Rank	Mean DPV	Mean Rank	Mean DPV	Mean Rank	Mean DPV
BR	1.785	0.093	1.928	0.259	1.902	0.223	1.719	0.164
F	2.330	0.112	2.013	0.273	2.026	0.240	2.255	0.211
GW	2.021	0.111	2.060	0.267	2.072	0.243	2.026	0.190
Friedman F	21.67		1.58		2.74		27.38	
p-value	0.001		0.207		0.066		0.001	

Panel B. Pairwise Comparisons[c]

Model Pair	TI	$RI(Q_1)$	$RI(Q_2)$	$RI(Q_3)$
BR-F	-5.21	-0.90	-1.31	-4.05
	(0.001)	(0.370)	(0.192)	(0.001)
BR-GW	-3.19	-1.42	-1.85	-1.24
	(0.002)	(0.158)	(0.066)	(0.215)
GW-F	-2.28	0.53	0.52	-0.16
	(0.023)	(0.598)	(0.603)	(0.871)

a : The forecast error metric used is squared forecast error.
The statistics are based on 235 sample firms using average predictive values over 5 years.

b : For each firm, a rank of one (three) is assigned to the model yielding the smallest (largest) value of absolute deviation between average theoretical predictive value and average empirical predictive value (DPV).

c : The matched-pair t-tests based on ranks are used. Associated p-values are in parentheses.

Table 3 reports the results of model comparison based on the within sample method. Three measures of the goodness-of-fit were used : (1) the Ljung-Box x^2 statistic; (2) the Akaike Information Criterion (AIC); and (3) the residual variance. The Friedman F-statistics (Panel A) indicate that there is a significance difference ($\alpha=0.001$) across models in their ability to fit the data regardless of which goodness-of-fit measure was used. The pairwise comparison results (Panel B) suggest that both the BR and GW models are superior to the F model, while the BR-GW pair comparisons provide mixed results which depend on the choice of the goodness-of-fit measures. These results are consistent with those in Brown and Rozeff [1979a] and Dharan [1983].

Table 3. Comparison of Goodness-of-Fit Statistics Across Time-Series Models[a]

Panel A. Summary Statistics and Overall Comparisons[b]

	Ljung-Box x^2		AIC		Res Variance	
Model	Mean Rank	Mean Value	Mean Rank	Mean Value	Mean Rank	Mean Value
BR	1.785	11.609	1.545	-56.489	1.677	0.177
F	2.657	17.699	2.536	-48.654	2.617	0.299
GW	1.957	10.166	1.919	-55.319	1.706	0.193
Friedman F	179.09		117.42		219.07	
p-value	0.001		0.001		0.001	

Panel B. Pairwise Comparisons[c]

Model Pair	Ljung-Box x^2	AIC	Res Variance
BR-F	-12.08(0.001)	-12.39(0.001)	-15.19(0.001)
BR-GW	2.82(0.005)	-4.92(0.001)	-0.48(0.634)
GW-F	-15.23(0.005)	-7.37(0.001)	-14.38(0.001)

a : The three goodness-of-fit statistics are obtained by estimating three time-series models for each of 235 sample firms using 52 quarters' EPS data (1967~1979).

b : For each firm, a rank of one (three) is assigned to the model yielding the smallest (largest) goodness-of-fit statistic.

c : The matched-pair t-tests based on the ranks are used. Associated p-values are in parentheses.

Table 4 provides the results of model evaluation based on the relative accuracy of forecasting annual earnings, using squared percentage errors (SPE).[13] Panel A of Table

13) The forecast error metric, squared forecast error (SFE), was deflated by actual earnings in

4 summarizes the average ranks and forecast errors for each model, the Friedman F-statistics, and the associated levels of significance. Two results are worth noting. First, as shown theoretically in Section 2, quarterly earnings have predictive value for improving the forecasts of annual earnings. This can be seen from the monotonic decrease in mean values of APE as additional quarterly earnings become available. To examine whether the improvements in forecasts are statistically significant, we tested the null hypothesis of no difference in APEs across quarters in which annual forecasts are generated. The F-statistics were 41.16, 38.97 and 53.03 for the BR, F and GW model respectively, resulting

Table 4. Comparisons of Time-Series Models Based on Annual Earnings Forecast Errors[a]

Panel A. Summary Statistics and Overall Comparisons[b]								
	Q_0		Q_1		Q_2		Q_3	
Model	Mean Rank	Mean SPE	Mean Rank	Mean SPE	Mean Rank	Mean SPE	Mean Rank	Mean SPE
BR	1.950	0.546	1.970	0.406	1.920	0.294	1.975	0.198
F	1.986	0.607	2.035	0.467	2.084	0.368	2.040	0.229
GW	2.064	0.619	1.995	0.438	1.996	0.323	1.985	0.199
Friedman F	6.65		2.02		12.46		2.22	
p-value	0.001		0.133		0.001		0.109	

Panel B. Pairwise Comparisons[c]				
Model Pair	Q_0	Q_1	Q_2	Q_3
BR-F	-0.90	-1.59	-4.00	-1.88
	(0.368)	(0.112)	(0.001)	(0.060)
BR-GW	-2.81	-0.62	-1.96	-0.43
	(0.005)	(0.537)	(0.050)	(0.668)
GW-F	2.13	-1.03	-2.15	-1.51
	(0.034)	(0.302)	(0.032)	(0.131)

a : The statistics are based on pooling data across 235 sample firms and over 5 years (1,175 observations). The forecast error metric used is squared percentage error.
b : For each firm-year, a rank of one (three) is assigned to the model yielding the smallest (largest) SPE.
c : The matched-pair t-tests based on ranks are used.
Associated p-values are in parentheses.

---

order to ensure relative comparability of forecast errors among firms because earnings numbers in absolute scale are different across firms. Also, all forecast errors greater than 300 percent were truncated to 300 percent to avoid the problem of outliers.

in the rejection of the null hypothesis at $\alpha < 0.001$ regardless of which model is used for forecasting.[14] This result is consistent with the empirical findings in Lorek [1979], Brown and Rozeff [1979] and Collins and Hopwood [1980].

Second, there is a significant difference in the accuracy of annual earnings forecasts across three time-series models at the α level of 0.001 for the forecasts made at Q_0 and Q_2 while the difference is insignificant when the conditioning quarters are Q_1 and Q_3. Pairwise comparison results (Panel B) indicate that the BR model has the smallest forecast errors. Except for the forecasts at Q_3, however, the superiority of the BR model over the F and GW models is statistically insignificant. Furthermore, the results of the GW-F pair comparison depend on forecast horizon; the GW model is superior to the F model except for the forecast made at the beginning of the year. Overall, these results indicate that the out of sample method provides no clear-cut rankings of the three models for our sample. In contrast, our proposed method gives clearer rankings; the dominance of the BR model.

V. Conclusion

In this study, we attempt to provide additional evidence regarding the relative superiority among three quarterly univariate time-series models. Different from prior studies, our approach utilizes the theoretical relationship between predictive values and time-series properties (parameter values) of quarterly earnings to examine this issue by comparing across models the magnitude of differences between theoretical and empirical predictive values.

The results suggest that the BR model dominates the GW or F model. Furthermore, this finding is robust with respect to different forecast error metrics (APE or SPE) and predictive value measures (TI or RIs). Although this result is generally consistent with that based on the within sample method (AIC and residual variance), the ranking of the models is clearer under our method than the out of the sample method (MSFE) which has been extensively used in previous studies to evaluate the time-series models of quarterly earnings.

14) A nonparametric test, the Kruskal-Wallis test (Connover [1980, pp.229~237]), was also employed to test the null hypothesis. The x^2 statistics of 391.33(BR), 300.06(F) and 363.31 (GW) provide the same inference.

References

Abdel-Khalik, R. and J.Espejo, "Expectations Data and the Predictive Value of Interim Reporting," *Journal of Accounting Research* (Spring 1978), pp.1~13.

Ball, R. and G.Foster, "Corporate Financial Reporting : A Methodological Review of Financial Research," *Supplement to Journal of Accounting Research* (1982), pp.161~234.

Barnea, A., T.Dyckman and R.Magee, "Discussion of the Predictive Content of Interim Reports-A Time-series Analysis," *Empirical Research in Accounting : Selected Studies* (1972), pp.145~155.

Bathke, A. and K.Lorek, "The Relationship Between Time-Series Models and the Security Market's Expectation of Quarterly Earnings," *The Accounting Review* (April 1984), pp.49~68.

Box, G. and G.Jenkins, *Time-Series Analysis : Forecasting and Control* Revised ed. Holden-Day, 1976.

Brown, L. and M.Rozeff, "Univariate Time-Series Models of Quarterly Accounting Earnings per Share : A Proposed Model," *Journal of Accounting Research* (Spring 1979a), pp.179~189.

Brown, L. and M.Rozeff, "The Predictive Value of Interim Reports for Improving Forecasts of Future Quarterly Earnings," *The Accounting Review* (July 1979b), pp.585~591.

Brown, L., R.Hagerman, P.Griffin and M.Zmijewski, "Security Analyst Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings," *Journal of Accounting and Economics* (April 1987), pp.61~87.

Collins, W. and W. Hopwood, "A Multivariate Analysis of Annual Earnings Forecasts Generated from Quarterly Forecasts of Financial Analysts and Univariate Time-Series Models," *Journal of Accounting Research* (Autumn 1980), pp.390~406.

Collins, W., W. Hopwood and J. McKeown, "The Predictablity of Interim Earnings over Alternative Quarters," *Journal of Accounting Research* (Autumn 1984), pp.467~479.

Dharan, B., "Empirical Identification Procedures for Earnings Models," *Journal of Accounting Research* (Spring 1983), pp.256~270.

Foster, G., "Quarterly Earnings Data : Time-Series Properties and Prdictive-Ability Results," *The Accounting Review* (January 1977), pp.1~21.

Griffin, P., "The Time-Series Behavior of Quarterly Earnings : Preliminary Evidence," *Journal of Accounting Research* (Spring 1977), pp.71~83.

Lorek, K., "Predicting Annual Net Earnings with Quarterly Earnings Time-Series Models," *Journal of Accounting Research* (Spring 1979), pp.190~204.

Lorek, K. and J.McKeown, "The Effect on Predictive Ability of Reducing the Number of Observations on a Time-Series Analysis of Quarterly Earnings Data," *Journal of Accounting Research* (Spring 1978), pp.204~214.

McKeown, J. and K.Lorek, "A Comparative Analysis of the Predictive Ability of Adaptive Forecasting, Re-estimation, and Re-identification using Box-Jenkins Time-Series Analysis on Quarterly Earnings Data," *Decision Sciences* (October 1978), pp.658~672.

Watts, R., "The Time-Series Behavior of Quarterly Earnings," working paper, University of Newcastle, April 1975.

〈국문초록〉

예측가치에 의한 분기별회계이익 시계열모형의 평가

이 경 주

본 연구의 목적은 분기별회계이익(quarterly earnings)의 시계열특성을 설명하는 대표적 모형으로 제시되고 있는 Foster의 (100×(010)모형, Griffin과 Watts의 (011)×(011)모형, Brown과 Rozeff의 (100)×(011)모형을 상호 비교 · 평가하는데 있다. 예측오차 또는 시장기대치(market expectations)로서의 이익예측능력 등을 평가기준으로 사용한 과거의 연구결과는 상반된 경우가 많아 분기별회계이익의 최적 시계열모형에 관한 명백한 규명이 아직까지도 이루어지고 있지 않다. 과거의 연구와는 달리 본 연구에서는 연간회계이익(annual earnings)을 예측함에 있어서 분기별회계이익정보가 공헌하는 정도를 나타내는 「예측가치(predictive value)」를 모형간의 비교 · 평가기준으로 사용하였다. 즉, 특정 모형에 대해 이론적(theoretical) 예측가치를 도출하고 이에 대응하는 실제 또는 경험적(empirical) 예측가치를 계산한 후 이 두 수치의 차이를 이용하여 각 모형들을 상호 비교하고 가장 작은 차이를 보이는 모형을 선택하는 것이다.

본 연구의 결과는 다음과 같이 요약된다. 첫째, 이론적 분석을 통하여 분기별회계이익정보는 항상 연간회계이익에 대한 예측의 정확성을 높여준다는 점에서 유용하며 유용성의 정도 즉, 예측가치는 특정 분기별회계이익모형의 모수값(parameter value)에 의해 결정된다는 것이 증명되었다. 둘째, 235개의 미국 상장기업을 표본으로 하고 18년(1967~1984) 동안의 주당순이익자료를 이용한 실증분석에 의하면 Brown과 Rozeff모형이 분기별회계이익의 시계열특성을 가장 잘 설명하는 것으로 나타났다. 과거 연구에서 사용된 기준을 적용한 경우에도 이와 유사한 결과를 얻었으나, 본 연구에서 사용된 예측가치기준을 적용하였을 때에 보다 명확한 결론을 얻을 수 있었다.