The Superiority of Analyst Forecasts as Measures of Expectations: Evidence from Earnings
Author(s): Lawrence D. Brown and Michael S. Rozeff
Source: *The Journal of Finance*, Vol. 33, No. 1 (Mar., 1978), pp. 1-16
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2326346
Accessed: 17/03/2010 12:08

THE JOURNAL OF FINANCE · VOL. XXXIII, NO. 1 · MARCH 1978

The Journal of FINANCE

VOL. XXXIII MARCH 1978 No. 1

THE SUPERIORITY OF ANALYST FORECASTS AS MEASURES OF EXPECTATIONS: EVIDENCE FROM EARNINGS

LAWRENCE D. BROWN AND MICHAEL S. ROZEFF*

ACCURATE MEASUREMENT OF EARNINGS expectations is essential for studies of firm valuation, cost of capital and the relationship between unanticipated earnings and stock price changes. Under the rational expectations hypothesis [23], market earnings expectations should be measured by the best available earnings forecasts. Univariate time series forecasts are often used for this purpose ([1], [3], [4], [5], [12], [13], [14], [16], [18], [20]) instead of direct measures of earnings expectations such as security analysts' forecasts. Univariate time series forecasts neglect potentially useful information in other time series and therefore do not generally provide the most accurate possible forecasts [24]. Since security analysts process substantially more data than the time series of past earnings, their earnings forecasts *should* be superior to time series forecasts and provide better measures of market earnings expectations.

However, the mere existence of analysts as an employed factor in long run equilibrium means that analysts *must* make forecasts superior to those of time series models. To reach this conclusion, one need only assume that participants in the market for forecasts act in their own best interests and that both forecast producers and consumers demand forecasts solely on the basis of their predictive ability.[1] Since analysts' forecasts cost more than time series forecasts, the continued employment of analysts by profit-maximizing firms implies that analysts' forecasts must be superior to those of the lower cost factor, time series models.

Past comparisons of analysts' forecasts to sophisticated time series models conclude that analysts' forecasts are not more accurate than time series forecasts (Cragg and Malkiel (CM) [9]; Elton and Gruber (EG) [11]). This evidence plainly conflicts with basic economic theory. Hence, the predictive accuracy of analysts' forecasts is re-examined in this paper. In contrast with other studies, the results overwhelmingly favor the superiority of analysts over time series models.

Part I considers statistical tests and experimental design. Part II contains the empirical results. Summary and implications appear in Part III.

* College of Business Administration, The University of Iowa, Iowa City.

1. We assume that forecast purchasers do not derive nonmonetary benefits from forecasts.

I. Experimental Design

A. *Statistical Evaluation of Forecast Methods*

Without direct information on the costs of imperfect forecasts to forecast users, comparative forecast accuracy is usually evaluated by comparing the error distributions of different forecast methods statistically. However, statistical comparisons in past studies ([9], [11]) utilize test statistics improperly, particularly Theil's U [25] and Student's t. In this section, after discussing the defects of these statistics for evaluating two or more forecast methods, the alternative statistical methods used in this study are introduced.[2]

Theil's U-statistic (applied to earnings) is the square root of

$$U_{ij}^2 = \frac{\sum_{t=1}^{T} (\dot{P}_{ijt} - \dot{A}_{it})^2}{\sum_{t=1}^{T} \dot{A}_{it}^2},$$

where $\dot{A}_{it}$ = change in actual earnings per share of firm i from $t-1$ to t,
$\dot{P}_{ijt}$ = predicted change in earnings per share of firm i from $t-1$ to t by forecast method j, and
T = total number of time series observations.

For its computation, it requires *time series* data on a firm's earnings *changes*.[3] Given forecast method j and earnings time series data on firm i, Theil's U compares the forecast accuracy of method j to that of a naive, no change, earnings forecast model.[4,5] Since analysts' earnings forecasts are currently available only in short time series, use of Theil's U for comparative forecast evaluation necessarily relies on small samples.[6] Larger sample sizes are possible by testing forecast methods on a cross-section of firms. Finally, no procedure is available with tests of significance which uses Theil's U to compare two forecast methods when neither is a no-change method. Direct hypothesis tests are preferable to inferences drawn from ranking the U statistics of different forecast methods.

For hypothesis tests of two forecast methods, an appropriate design is a one-sample or matched pairs case with self-pairing by firm. The members of each pair

2. Past studies also contain experimental biases: CM compare analysts' five-year forecasts with realizations over three and four-year horizons; EG compare analysts' forecasts with the "best" of nine time series models selected from the same time period in which comparisons with analysts' forecasts are made. This procedure introduces *ex post* selection bias.

3. EG computed "Theil's U" using earnings *levels* rather than *changes*. This statistic has unknown sampling properties.

4. $\dot{P}_{ijt} = \dot{A}_{it}$ and $U_{ij} = 0$ if prediction is perfect in every period. If no change is predicted in each period (i.e., $\dot{P}_{ijt} = 0$), $U_{ij} = 1$; $0 < U_{ij} < 1$ if prediction is less than perfect but better than the no-change prediction and $U_{ij} > 1$ if forecast method j is less accurate than the no-change prediction.

5. CM used *cross-sectional* rather than temporal data. This "Theil's U" statistic has unknown sampling properties because each error is drawn from a different error distribution, one for each firm.

6. EG's sample size in computing Theil's U varied between two and six.

are the errors from the two methods; the matched pair is reduced to a single observation by taking the difference in the errors. The usual parametric test of the mean difference is the paired t-test [17]. An alternative non-parametric test of the median difference is the Wilcoxon Signed Ranks test [8].

The parametric paired t-test is inappropriate for testing mean error differences of forecast methods applied to cross-section earnings data. If applied to error measures stated in level form (e.g., $|P_{ijt} - A_{it}|$, where P_{ijt} = firm i's forecasted earnings per share for period t by method j and A_{it} = firm i's actual earnings per share in period t), the test's assumption that paired differences are drawn from the same population is violated since each error difference depends upon each firm's earnings per share level. If applied to error measures stated in ratio form (e.g., $|P_{ijt} - A_{it}|/|A_{it}|$), the distributional assumptions of the paired t-test are also unlikely to be fulfilled since ratio measures applied to earnings per share data are dominated by outliers because actual earnings per share are often close to zero.[7]

Meaningful pairwise comparisons require test statistics which are insensitive to error definition and outliers. We adopt the Wilcoxon Signed Ranks test which meets these requirements and has power comparable to the parametric paired t-test [8, p. 213].

For tests of several forecast methods, the generalization of the paired t-test, two-way analysis of variance, is inapplicable.[8] The Friedman test [8], which is based on two-way analysis of variance by ranks and is independent of error definition, is used instead.

For an error measure, we choose relative error ignoring sign, $|P_{ijt} - A_{it}|/|A_{it}|$, a metric which is likely to be of interest to forecast purchasers.[9] In any event, the Wilcoxon test statistic is insensitive to error definition (see fn. 16).

B. *Forecast Horizon*

Because economic theory provides no guidance concerning the association of analyst superiority with a particular forecast horizon, several horizons should be investigated.[10] Our choice of horizons reflects the following considerations: (i) micro-level information obtained by analysts often concerns earnings of the following several quarters or fiscal year; (ii) current fiscal and monetary policies affect earnings of the subsequent one to five quarters; (iii) published forecasts are available mainly for short horizons. We thus investigate point estimates of quarterly earnings per share for forecast horizons of one to five quarters. We also examine annual earnings forecasts. The basic time series data are quarterly primary

7. EG's cross-section parametric t-test is inappropriate. Their use of an error measure stated in terms of levels squared (mean square error) appears to compound the inherent difficulty in applying the paired t-test to cross-section earnings data (see fn. 16).

8. Preliminary tests indicated serious violation of the homogeneity of variances and additivity assumptions, basically because of error outliers. Violation of the ANOVA assumptions also prevents application below of a factorial design with sample year and forecast horizon as factors, forecast method as treatment and firm as replication.

9. For a discussion of the deficiencies of using $|P_{ijt}|$ or $|P_{ijt} + A_{it}|/2$ in the denominator see [25].

10. The forecast horizons studied in the past have been five years (CM) and one year (EG).

earnings per share before extraordinary items, adjusted for stock splits, stock dividends and other capitalization changes for the years 1951–1975.

Ex ante conditional predictions of all forecast methods are determined as follows for a sample of 50 firms for each of the four years 1972–1975. Starting with third quarter 1971 earnings (III/1971), conditional earnings per share predictions for the ith firm by the jth method are obtained for the individual quarters of 1972. The forecasts of 1972 quarterly earnings, conditional on III/1971, are denoted $P_{ij}(\text{I}/1972 \mid \text{III}/1971)$, $P_{ij}(\text{II}/1972 \mid \text{III}/1971)$, $P_{ij}(\text{III}/1972 \mid \text{III}/1971)$ and $P_{ij}(\text{IV}/1972 \mid \text{III}/1971)$. Moving ahead one quarter, predictions are again obtained for each of the four quarters of 1972 made conditional upon IV/1971 earnings data. Again moving ahead one quarter, predictions are obtained for the last three quarters of 1972 conditional upon knowledge of I/1972 earnings, etc. Table 1 shows the set of 1972 predictions so obtained. With these conditional predictions, relative forecast errors ignoring sign are computed for each forecast method j over five distinct quarterly forecast horizons for use in the quarterly error comparisons. Annual earnings forecasts for 1972 are the sum of the forecasts $P_{ij}(\text{I}/1972 \mid \text{IV}/1971)$, $P_{ij}(\text{II}/1972 \mid \text{IV}/1971)$, $P_{ij}(\text{III}/1972 \mid \text{IV}/1971)$, and $P_{ij}(\text{IV}/1972 \mid \text{IV}/1971)$, that is, the one to four period ahead point forecasts made conditional upon knowledge of the prior year's fiscal earnings.[11] After obtaining analogous forecasts for the years 1973, 1974 and 1975, quarterly and annual comparisons are repeated for these years.

TABLE 1

SUMMARY OF PREDICTIONS BY FORECAST HORIZON FOR 1972[a,b]

1 Quarter Ahead	2 Quarters Ahead	3 Quarters Ahead	4 Quarters Ahead	5 Quarters Ahead[c]
$P_{ij}(\text{I}/1972 \mid \text{IV}/1971)$	$P_{ij}(\text{I}/1972 \mid \text{III}/1971)$			
$P_{ij}(\text{II}/1972 \mid \text{I}/1972)$	$P_{ij}(\text{II}/1972 \mid \text{IV}/1971)$	$P_{ij}(\text{II}/1972 \mid \text{III}/1971)$		
$P_{ij}(\text{III}/1972 \mid \text{II}/1972)$	$P_{ij}(\text{III}/1972 \mid \text{I}/1972)$	$P_{ij}(\text{III}/1972 \mid \text{IV}/1971)$	$P_{ij}(\text{III}/1972 \mid \text{III}/1971)$	
$P_{ij}(\text{IV}/1972 \mid \text{III}/1972)$	$P_{ij}(\text{IV}/1972 \mid \text{II}/1972)$	$P_{ij}(\text{IV}/1972 \mid \text{I}/1972)$	$P_{ij}(\text{IV}/1972 \mid \text{IV}/1971)$	$P_{ij}(\text{IV}/1972 \mid \text{III}/1971)$

[a] Predictions missing from the table (e.g., $P_{ij}(\text{I}/1972 \mid \text{II}/1971)$, $P_{ij}(\text{II}/1972 \mid \text{II}/1971)$) are absent because our source of analyst data does not contain these forecasts.

[b] i and j refer to firm i and method j, respectively.

[c] Five quarter ahead are available for BJ and V only.

C. *Time Series Models and Analysts' Forecasts*

Within the class of univariate time series models, Box and Jenkins (BJ) [6] models are highly regarded for their ability to make the most efficient use of the time series data. The BJ modelling technique enables one to select the most appropriate time series model consistent with the process generating each firm's time series of quarterly earnings per share data. BJ models, by not making *a priori* assumptions about the processes generating the data, subsume autoregressive,

11. Beaver [1] concludes that a quarterly approach to predicting annual earnings is at least as good as an annual approach to predicting annual earnings. Also see [7], [19] and [22] for other aspects of the usefulness of quarterly earnings per share data.

moving average and mixed models as special cases.[12] Forecasts of individually fitted BJ models should, therefore, perform better than forecasts of a particular class of time series models applied to all firms' time series data. We adopt the BJ modelling technique in this paper. Two other time series models are also included, a "seasonal martingale" (denoted M) and a "seasonal submartingale" (S). These models have been used as standards of comparison in the earnings forecast literature and are available for forecast producers and users at minimal cost.

As a source of analysts' forecasts we choose the Value Line Investment Survey since it contains one to five quarter ahead earnings forecasts which can be accurately dated and measured. Value Line makes earnings forecasts for 1,600 firms in contrast with institutional research firms which provide fewer, more expensive forecasts. Our hypothesis test thus compares a relatively sophisticated time series model with an "average" source of analysts' forecasts.

BJ conditional forecasts are obtained by standard methods after identifying and estimating each firm's appropriate model [6].[13] Value Line's conditional forecasts are taken directly from individual issues of the Value Line Investment Survey. The Survey, published weekly, makes quarterly earnings predictions four times a year for each firm included.

To define conditional forecasts of the naive models for each firm i, let A_{it} denote the tth actual quarterly earnings per share for firm i, where $t=1,\dots,96$ (I/1951–IV/1974).

Seasonal submartingale (S) conditional one to four quarter ahead forecasts at time t are

one quarter ahead	$A_{it-3}+(A_{it}-A_{it-4})$
two quarters ahead	$A_{it-2}+(A_{it}-A_{it-4})$
three quarters ahead	$A_{it-1}+(A_{it}-A_{it-4})$
four quarters ahead	$A_{it}\quad+(A_{it}-A_{it-4})$.

Seasonal martingale (M) conditional one to four quarter ahead forecasts made in period t are A_{it-3}, A_{it-2}, A_{it-1}, and A_{it}. M's forecasts for a given quarter do not change as actual earnings per share data become available. S modifies M's forecasts with the change of the latest period's quarter over that of the previous year.

Actual quarterly earnings data are announced for most firms approximately five to six weeks into the subsequent quarter. Time series forecasts then become

12. The *ad hoc* time series models used in previous studies at a time when BJ techniques were unavailable are special cases of BJ models.

13. Recent research by Froeschle [15] and diagnostic tests of Dent and Swanson [10] were helpful in identifying the BJ models in addition to the standard diagnostic tests. As an aid to identifying the BJ models, most of which had multiplicative seasonal components, theoretical autocorrelation and partial autocorrelation functions for many quarterly multiplicative seasonal models were obtained. The coefficients of the BJ models, estimated with data through IV/1974, were not re-estimated with less data for earlier periods or more data for later periods. Foster [13] has shown that coefficient re-estimation of BJ quarterly earnings models is unnecessary due to its negligible effect on forecast errors. In any event, our procedure (no re-estimation) favors BJ in nearly all comparisons with Value Line.

possible and Value Line forecasts are published, on average, forty to fifty days later.[14]

The pattern of forecasts for all models is summarized in Table 1. Note that models M and S are not used to generate five quarter ahead forecasts.

II. Empirical Results

A. *Sample Selection*

Fifty firms were randomly selected from Moody's Handbook of Common Stocks. Each firm has complete quarterly earnings data available from 1951, is included in the Value Line Investment Survey since 1971 and has a December fiscal year. The resulting sample (Appendix A) is representative of the New York Stock Exchange firms included in Moody's and Value Line. Utilities were excluded due to insufficient quarterly earnings data. Sample sizes are reduced in those rare instances when the Value Line conditional forecasts are unavailable.

B. *Annual Comparisons*

The error distributions of relative annual forecast errors are shown in Table 2 for each of the years 1972–75 using the four forecast methods, seasonal martingale (M), seasonal submartingale (S), Box-Jenkins (BJ) and Value Line (V). Table 2 also contains Friedman test statistics (Chi-square with 3 degrees of freedom) and Wilcoxon test statistics (Student's t with $N-1$ degrees of freedom where N is sample size). The Friedman test statistic examines the null hypothesis that *all four* error distributions are identically distributed; the Wilcoxon statistic tests the null hypothesis that the median error difference of *two* methods being compared exceeds zero.

Using the Friedman test, the null hypothesis is rejected at the 1% level in 1972, 1973 and 1975. In the 12 pairwise hypothesis tests of V's errors against those of M, S, and BJ, the sign of the Wilcoxon test statistic favors Value Line in every instance. Statistical significance occurs 8 times; 6 times at the 1% level and twice at the 5% level. Thus, V generally produces smaller annual errors than the three time series models suggesting that Value Line annual earnings forecasts are superior to those of time series models.

As argued earlier, BJ forecasts should be superior to forecasts of *ad hoc* time series models. The annual comparisons show that the BJ models generally yield smaller forecast errors than the other time series models studied. In 8 comparisons with M and S, the Wilcoxon test favors BJ 7 times with statistical significance 3 times. These findings suggest that BJ's forecasts are superior to those of *ad hoc* naive time series models.

While the annual results provide strong support for the hypothesis of analyst superiority, they use only a fraction of the data. More powerful tests are achieved using the larger sample sizes of the quarterly data and many more comparative tests can be performed with these data. We turn next to quarterly comparisons.

14. The time interval from announcement to forecast varies from approximately 7 to 70 days for our sample firms. The fact that the Investment Survey, published in 13 installments, makes forecasts for different firms each week accounts for the variation.

TABLE 2

WILCOXON AND FRIEDMAN TEST STATISTICS AND ERROR DISTRIBUTIONS, ANNUAL COMPARISONS OF VALUE LINE AND TIME SERIES MODEL PREDICTION ERRORS, 1972–1975[c]

1972

Error Distribution[d]	<.05	.05–.10	.10–.25	.25–.50	.50–.75	.75–1.00	>1.00
M	3	7	14	17	4	3	2
S	11	6	12	10	3	1	7
BJ	10	6	12	12	4	1	5
V	13	7	17	12	0	0	1

SAMPLE SIZE = 50
Friedman Statistic = 27.10[a]
Wilcoxon Statistics[e]

	S	BJ	*V*
M	−.55	.24	4.46[a]
S		.46	3.50[a]
BJ			3.45[a]

1973

Error Distribution[d]	<.05	.05–.10	.10–.25	.25–.50	.50–.75	.75–1.00	>1.00
M	2	6	16	18	6	0	2
S	11	8	14	9	4	1	3
BJ	8	6	15	16	3	0	2
V	10	9	13	16	0	0	2

SAMPLE SIZE = 50
Friedman Statistic = 33.19[a]
Wilcoxon Statistics[e]

	S	BJ	*V*
M	3.15[a]	2.51[a]	4.61[a]
S		−1.89[b]	0.34
BJ			2.17[b]

1974

Error Distribution[d]	<.05	.05–.10	.10–.25	.25–.50	.50–.75	.75–1.00	>1.00
M	8	6	12	15	4	1	4
S	12	3	11	12	6	2	4
BJ	5	8	16	13	4	0	4
V	6	7	15	13	5	0	4

SAMPLE SIZE = 50
Friedman Statistic = 4.68
Wilcoxon Statistics[e]

	S	BJ	*V*
M	−.21	2.37[a]	2.23[b]
S		1.24	1.44
BJ			0.61

TABLE 2 (*continued*)

	1975 Error Distribution[d]						
	<.05	.05–.10	.10–.25	.25–.50	.50–.75	.75–1.00	>1.00
M	4	7	13	10	2	3	11
S	3	5	12	7	9	4	10
BJ	7	3	13	12	2	3	10
V	7	5	18	5	3	3	9

SAMPLE SIZE = 50
Friedman Statistics = 12.84[a]
Wilcoxon Statistics[e]

	S	BJ	*V*
M	−1.77[b]	0.86	3.29[a]
S		2.99[a]	3.11[a]
BJ			1.28

[a] Significant at the 1% level, one-tailed test.

[b] Significant at the 5% level, one-tailed test.

[c] *V* = Value Line, *M* = Seasonal Martingale, *S* = Seasonal Submartingale, BJ = Box-Jenkins.

[d] Each entry below designates the number of observations for a given model whose relative error ignoring sign is within the stated fractiles.

[e] Each Wilcoxon test statistic below results from comparing the method at the top with the method on the side. Thus, positive Wilcoxon statistics indicate superiority of model on top.

C. *Quarterly Comparisons*

In each year, 1972 to 1975, quarterly forecasts are obtained for the forecast methods in the manner shown in Table 1. Relative forecast errors of all four methods are compared over 1–4 quarter forecast horizons; BJ and *V* are also compared over 5 quarter horizons. In each of the four years, sample sizes are approximately 200 for the 1 and 2 quarter ahead comparisons, 150 for the 3 quarter ahead comparisons, and 100 for the 4 quarter ahead comparisons. Test results over all horizons appear in Table 3 and are summarized in Table 4.

With minor exceptions (3 and 4 quarter horizons in 1974), the Friedman statistics are highly significant when the four methods are tested as a group; the null hypothesis of identically distributed distributions is rejected in 14 of the 16 Friedman tests. Using Wilcoxon test statistics, *V*'s errors are tested pairwise against *M*'s and *S*'s errors 16 times each and against BJ's errors 20 times. The resulting 52 hypothesis tests of *V* against *M*, *S* and BJ are summarized in Table 4A. In the 34 instances of significant Wilcoxon test statistics, *V* is statistically superior 33 times. In the remaining 18 tests, the sign of the *t*-statistic favors *V* 12 times. In total, *V* is favored 45 times out of 52, revealing an overwhelming dominance of *V* over the time series models.

The data are also summarized in Table 4 by the mean Wilcoxon *t*-value ($\bar{t}$), the estimated standard deviation of the mean *t*-value ($s(\bar{t})$) and the ratio $\bar{t}/s(\bar{t})$. The latter ratio is itself a *t*-statistic only if each *t*-value being averaged is drawn from the same distribution. Since the distribution of *t*-values is likely to depend upon the horizon, model and/or year that the experiment is conducted, we refrain from

TABLE 3

WILCOXON AND FRIEDMAN TEST STATISTICS, QUARTERLY COMPARISONS OF VALUE LINE AND TIME SERIES MODEL PREDICTION ERRORS, 1972–1975[c,d]

		Forecast Horizon												
		One Quarter			Two Quarter			Three Quarter			Four Quarter			Five Quarter
		S	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*V*
	M	2.14[b]	6.87[a]	8.15[a]	0.79	5.41[a]	6.87[a]	−1.09	2.50[a]	5.77[a]	−3.09[a]	1.41	5.22[a]	—
1972	*S*	—	4.62[a]	5.25[a]	—	4.62[a]	5.57[a]	—	3.03[a]	5.42[a]	—	3.38[a]	5.30[a]	—
	BJ	—	—	1.75[b]	—	—	2.51[a]	—	—	4.09[a]	—	—	3.93[a]	3.11[a]
		Sample Size = 200			Sample Size = 200			Sample Size = 150			Sample Size = 100			Sample Size = 50
		Friedman Stat. = 73.45[a]			Friedman Stat. = 60.54[a]			Friedman Stat. = 41.14[a]			Friedman Stat. = 43.43[a]			
		S	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*V*
	M	8.02[a]	8.98[a]	10.66[a]	5.81[a]	6.41[a]	8.70[a]	4.81[a]	3.52[a]	6.31[a]	2.55[a]	1.69[b]	4.63[a]	—
1973	*S*	—	−0.60	1.62	—	−1.83[b]	1.04	—	−3.57[a]	−0.02	—	−1.59	1.04	—
	BJ	—	—	2.48[a]	—	—	3.47[a]	—	—	3.34[a]	—	—	2.79[a]	1.66
		Sample Size = 199			Sample Size = 200			Sample Size = 150			Sample Size = 100			Sample Size = 50
		Friedman Stat. = 173.51[a]			Friedman Stat. = 119.91[a]			Friedman Stat. = 75.22[a]			Friedman Stat. = 29.12[a]			
		S	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*V*
	M	3.35[a]	6.29[a]	6.19[a]	0.84	4.88[a]	3.78[a]	−0.25	2.59[a]	1.29	−2.69[a]	1.41	0.29	—
1974	*S*	—	2.34[a]	2.95[a]	—	2.31[b]	1.50	—	1.53	0.97	—	2.67[a]	2.80[a]	—
	BJ	—	—	1.16	—	—	−1.45	—	—	−1.04	—	—	−0.92	−2.20[b]
		Sample Size = 199			Sample Size = 199			Sample Size = 149			Sample Size = 100			Sample Size = 50
		Friedman Stat. = 47.57[a]			Friedman Stat. = 22.63[a]			Friedman Stat. = 5.40			Friedman Stat. = 2.92			
		S	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*S*	BJ	*V*	*V*
	M	2.07[b]	5.76[a]	8.22[a]	−2.64[a]	3.63[a]	5.29[a]	−4.49[a]	2.93[a]	2.95[a]	4.89[a]	−0.78	−0.05	—
1975	*S*	—	4.70[a]	6.36[a]	—	6.02[a]	6.14[a]	—	6.13[a]	5.14[a]	—	3.62[a]	3.28[a]	—
	BJ	—	—	3.51[a]	—	—	1.62	—	—	−0.22	—	—	0.08	0.45
		Sample Size = 199			Sample Size = 199			Sample Size = 149			Sample Size = 100			Sample Size = 50
		Friedman Stat. = 80.32[a]			Friedman Stat. = 44.49[a]			Friedman Stat. = 33.25[a]			Friedman Stat. = 15.66[b]			

[a] Significant at the 1% level, one-tailed test.

[b] Significant at the 5% level, one-tailed test.

[c] V = Value Line, M = Seasonal Martingale, S = Seasonal Submartingale, BJ = Box-Jenkins.

[d] Each Wilcoxon test statistic entered in the table results from comparing method at the top with method on the side. Thus, positive Wilcoxon statistics indicate superiority of model on top.

TABLE 4

SUMMARY OF WILCOXON TEST COMPARISONS

A: Value Line vs. Time Series Models[a]

	Total	Forecast Horizon 1Q	2Q	3Q	4Q	5Q	Forecast Model *M*	*S*	BJ	Year 1972	1973	1974	1975
Number of Comparisons	52	12	12	12	12	4	16	16	20	13	13	13	13
Comparisons Favorable to V[b]	45	12	11	9	10	3	15	15	15	13	12	9	11
Comparisons Statistically Favorable to V[c]	33	10	8	7	7	1	13	10	10	13	8	4	8
Comparisons Statistically Unfavorable to V	1	0	0	0	0	1	0	0	1	0	0	1	0
Mean Wilcoxon Test Statistic ($\bar{t}$)	3.25	4.86	3.75	2.83	2.37	.76	5.27	3.40	1.51	4.84	3.67	1.18	3.29
$\bar{t}/s(\bar{t})$[d]	8.27	5.45	4.51	3.81	3.72	.67	5.65	6.24	3.48	9.98	4.18	1.81	4.24

B: BJ vs. Naive Time Series Models

	Total	Forecast Horizon 1Q	2Q	3Q	4Q	Forecast Model *M*	*S*	Year 1972	1973	1974	1975
Number of Comparisons	32	8	8	8	8	16	16	8	8	8	8
Comparisons Favorable to BJ[b]	27	7	7	7	6	15	12	8	4	8	7
Comparisons Statistically Favorable to BJ[c]	24	7	7	6	4	13	11	7	4	6	7
Comparisons Statistically Unfavorable to BJ	2	0	1	1	0	0	2	0	2	0	0
Mean Wilcoxon Test Statistic ($\bar{t}$)	3.15	4.87	3.93	2.33	1.48	3.97	2.34	3.98	1.63	3.00	4.00
$\bar{t}/s(\bar{t})$[d]	6.37	4.70	4.16	2.41	2.25	6.23	3.25	6.46	1.05	4.99	4.96

[a] V = Value Line, M = Seasonal Martingale, S = Seasonal Submartingale, BJ = Box-Jenkins.

[b] Comparisons are favorable if Wilcoxon statistic in Table 3 is positive.

[c] Comparisons are statistically favorable if Wilcoxon statistic in Table 3 is positive and significant at the 5% level or better.

[d] Both $\bar{t}$ and $s(\bar{t})$ are computed using the number of comparisons in each column of the Table.

hypothesis tests on $\bar{t}$ and present $\bar{t}$ and $\bar{t}/s(\bar{t})$ without formal tests of significance. For the 52 comparisons involving V, the mean Wilcoxon test statistic is 3.25 and $\bar{t}/s(\bar{t})$ is 8.27.

Table 4A also decomposes the 52 comparisons of V with the time series models by forecast horizon, model and year.[15] The data show that Value Line's forecast superiority holds over all horizons studied with a tendency for its superiority to decline as horizon lengthens. V's predominance model-by-model is, as hypothesized, quite evident with somewhat less superiority over BJ than over M and S. Turning our attention to the 20 comparisons between V and BJ, V is superior in 10 of 11 cases in which the test statistic is significant. In 5 of the remaining 9 comparisons, the sign of the Wilcoxon test statistic favors V. For completeness, Table 4A summarizes Wilcoxon tests by year. Again we expect V to be superior, on average, but have no hypothesis concerning particular years. Comparisons unfavorable to V tend to be confined to 1974, but even in this year, 4 of the 5 statistically significant comparisons favor Value Line.

In summary, the evidence strongly supports the hypothesis that Value Line consistently makes significantly better predictions than time series models. The statistically significant experiments overwhelmingly favor Value Line. In the remaining experiments the majority of the Wilcoxon tests also favor Value Line, providing additional support for the hypothesis of analyst superiority.

Table 4B summarizes the 32 comparisons of BJ with the naive time series models. The mean Wilcoxon test statistic is 3.15 and $\bar{t}/s(\bar{t})$ equals 6.37. In 26 cases, there are significant differences with BJ statistically superior 24 times. BJ is superior to M and S in 3 of the remaining 6 comparisons. Hence, BJ is favored in 27 of 32 comparisons, providing strong support for the hypothesis that BJ predicts earnings better than *ad hoc* time series models.

Table 4B also summarizes comparisons involving BJ by horizon, model and year. BJ's superiority over the naive models is clearly evident over each forecast horizon with a tendency for its superiority to decline as horizon lengthens. In comparison to individual models, BJ outperforms both M and S with somewhat less dominance over S. Turning to comparisons by year, the superiority of BJ is consistent over time, with most of the comparisons unfavorable to BJ occurring in 1973. Even in this year, the mean Wilcoxon test statistic is 1.63 and 4 of the 6 significant comparisons favor BJ.[16]

In conclusion, the quarterly and the annual comparisons provide convincing evidence both of Value Line's superiority over each of the three time series models and BJ's superiority over the naive models. The quarterly results also show that V's superiority over the time series models and BJ's superiority over the naive models

15. The decomposition is an alternative to analysis of variance which is inapplicable to the error distribution (see fn. 8).

16. As noted earlier, the Wilcoxon tests should be insensitive to error definition. Wilcoxon test statistics were recomputed on annual and selected quarterly comparisons using three additional error measures, mean square error, root mean square error and relative error squared. The small changes in the test statistics left the results virtually unchanged. Parametric t-tests were also applied to the four error measures. Both the sign and magnitude of these test statistics were highly sensitive to error definition. The hypothesis tests using the parametric t-test most often gave results in disagreement with the Wilcoxon test when mean square error was chosen as the error definition. This may account for EG's results differing from ours.

are not confined to particular models, horizons, or years. The very general character of Value Line's superiority in predicting earnings, evidenced over all models, horizons, and years in 64 separate hypothesis tests involving sample sizes averaging 125, lends extraordinary support to the hypothesis of analyst superiority.

D. *Further Analysis*

The superiority of Value Line over time series models follows from the rational behavior of forecast producers and consumers and should be generalizable to other sources of analyst forecasts and other time periods. As a preliminary test of the sensitivity of our results to choice of analyst, we obtained predictions of 1975 annual earnings per share made by the Standard and Poor's Earnings Forecaster (SP) for each firm included in the 1975 annual earnings sample.[17] Wilcoxon tests of SP against M, S, and BJ favored SP, yielding t-statistics of 3.18, 2.85 and 1.45 respectively. These results are remarkably similar to those using Value Line.[18] This evidence suggests that Value Line's forecast superiority over time series models is not unique.

To ascertain whether the sample period posed unusual difficulties for time series earnings forecasting, a BJ model was fitted to the Quarterly Earnings Index of the Dow Jones Industrial Average over the 1951–1975 time period.[19] Average quarterly percentage errors ignoring sign produced by the BJ model for 1972–1975 were 7.31%, 6.61%, 9.99%, and 15.47% respectively. Since the mean and standard deviation of average percentage forecast errors over the 1951–1975 period were 10.14% and 4.38%, it appears that the 1972–1975 period was not a particularly difficult one in which to predict earnings. Indeed, from this standpoint, the 1972–1975 period is comparable to the "stable" years of the sixties, 1962–1967, studied by CM and EG.[20]

These results indicate that if appropriate hypothesis tests are applied to other analysts and time periods, the results are likely to parallel those using Value Line and the 1972–1975 time period.

E. *A Brief Investigation of Value Line Superiority*

To produce forecasts superior to time series models, Value Line must utilize information not contained in the time series of quarterly earnings. During the period between the most recent quarterly earnings announcement and the subsequent Value Line prediction, Value Line acquires incremental information which, if an important part of its total information set, may explain Value Line's

17. SP, published weekly, contains annual predictions made by Standard and Poor's and other investment firms. The SP prediction for each firm is that made by Standard and Poor's on the date closest to the Value Line prediction date.

18. V's t-statistics versus M, S, and BJ were 3.29, 3.11, and 1.28 respectively (See Table 2). A direct Wilcoxon test between V and SP favored $V(t=.77)$.

19. The sample period, 1972–1975, may appear "unusual" since it includes peacetime wage and price controls, high inflation and inventory profits, large changes in employment and new accounting requirements. If events arising during the sample period caused the earnings generating process to change, the forecast ability of the BJ modelling technique may be hampered, unintentionally favoring the analyst.

20. The average percentage errors were 12.67%, 10.71%, 7.03%, 4.93%, 6.08% and 5.26%, respectively for 1962–1967.

superiority. Information arising during this interval is likely to be most important for predicting next quarter's earnings. Assuming that the generation of this incremental information is positively related to the passage of time, earnings should be relatively easier to predict the further Value Line's prediction date is from the most recent earnings announcement date, and one quarter horizon forecast errors should be negatively related to the corresponding intervals.

To test this hypothesis, we obtained for the firms in the 1975 one quarter horizon sample their Value Line errors and the time intervals (7–70 days) since their most recent earnings announcements. A rank correlation was applied to these variables. The insignificantly negative Spearman rho which was obtained suggests that information obtained by Value Line during this interval has a negligible effect on its ability to predict next quarter's earnings.[21] This evidence is consistent with the hypothesis that Value Line's superiority can be attributed to its use of the information set available to it on the quarterly earnings announcement date, and not to the acquisition of information arising after the quarterly earnings announcement date.

III. Summary and Implications

Basic economic theory and the equilibrium employment of analysts, a higher cost factor than time series models, imply that analysts must produce better forecasts than time series models. Past studies ([9], [11]) of comparative earnings forecast accuracy have concluded otherwise but use inappropriate parametric tests and contain experimental biases. Using nonparametric statistics which provide proper yet powerful tests, we find that (1) BJ models consistently produce significantly better earnings forecasts than martingale and submartingale models; (2) Value Line Investment Survey consistently makes significantly better earnings forecasts than the BJ and naive time series models. The findings are in accord with rationality in the market for forecasts and the long-run equilibrium employment of analysts.

If market earnings expectations are rational [23], it follows that the best available earnings forecasts should be used to measure market earnings expectations. Given rational market expectations, our evidence of analyst superiority over time series models means that analysts' forecasts should be used in studies of firm valuation, cost of capital and the relationship between unanticipated earnings and stock price changes until forecasts superior to those of analysts are found.[22] Past findings ([2], [21]) that share price levels are significantly better explained by analysts' earnings

21. The lack of a significant negative correlation between prediction error and time since last announcement date may occur if the interval is intentionally lengthened by Value Line in order to acquire more information about the firms whose earnings are more difficult to predict. To test this possibility, we measured each firm's prediction "difficulty" by its average one quarter horizon percentage error ignoring sign yielded by its BJ model. No significant correlation was found between this variable and the time interval between the most recent quarterly earnings announcement and the Value Line prediction date.

22. In examining the relationship between unanticipated earnings and stock price changes, for example, the sign of the forecast error from a time series is often used ([7], [12], [13]) as a device for classifying unanticipated earnings into "favorable" or "unfavorable" categories. With this methodology, BJ and *V* classify earnings differently 213 times out of the 797 one quarter ahead forecasts in our sample.

forecasts than by those of time series models are consistent with our evidence and with market rationality.

The hypothesis of analyst superiority versus univariate time series models is derived from basic economic theory and is not limited to the case of earnings. It is therefore applicable to all types of forecasts subject to the market test. There is no presumption that other, non-market forecasts such as those made by corporate executives or government agencies should be better (or worse) than those generated by univariate time series models.

APPENDIX A

Sample Firms

Abbott Laboratories
Allegheny Ludlum Industries, Inc.
American Airlines, Inc.
Anaconda Company
Boeing Company
Borg-Warner Corporation
Braniff International Corporation
Caterpillar Tractor Company
Champion International Corporation
Chrysler Corporation
Clark Equipment Company
Colgate-Palmolive Company
Continental Can Company, Inc.
Curtiss-Wright Corporation
Cutler-Hammer, Inc.
Eastern Airlines, Incorporated
Eastman Kodak Company
Flintkote Company
Freeport Minerals Company
Fruehauf Corporation
GATX Corporation
General Electric Company
Goodrich (B. F.) Company
Gulf Oil Corporation
Homestake Mining Company
International Business Machines Corporation
International Paper Co.
Kennecott Copper Corporation
Leheigh Portland Cement Co.
Ligget Group Inc.
Lowenstein (M.) & Sons, Inc.
Nabisco, Inc.
National Distillers & Chemical Corporation
National Steel Corporation

Pan American World Airways, Inc.
Pepsico, Inc.
Phelps Dodge Corporation
Phillips Petroleum Co.
Pullman, Incorporated
Raybestos-Manhattan, Inc.
Republic Steel Corporation
Standard Brands, Inc.
Standard Oil Company of Indiana
Sterling Drug, Incorporated
St. Regis Paper Company
Timken Company
United States Gypsum Company
United States Steel Corporation
United Technologies Corp.
Wrigley (W. M.) Jr. Company

REFERENCES

1. W. H. Beaver. "The Information Content of the Magnitude of Unexpected Earnings," Unpublished Manuscript, 1974.
2. F. W. Bell. "The Relation of the Structure of Common Stock Prices to Historical, Expectational and Industrial Variables," *Journal of Finance* (March 1974), pp. 187–197.
3. H. Benishay. "Variability in Earnings-Price Ratios of Corporate Equities," *American Economic Review* (March 1961), pp. 81–94.
4. G. Benston. "Published Corporate Accounting Data and Stock Prices," *Empirical Research in Accounting: Selected Studies, 1967*. Supplement to Vol. 5, *Journal of Accounting Research*, pp. 1–54.
5. R. S. Bower and D. H. Bower. "Risk and the Valuation of Common Stock," *Journal of Political Economy* (May/June 1969), pp. 349–362.
6. G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day, 1970).
7. P. Brown and J. Kennelly. "The Informational Content of Quarterly Earnings: An Extension and Some Further Evidence," *Journal of Business* (July 1972), pp. 403–415.
8. W. J. Conover. *Practical Nonparametric Statistics* (New York: John Wiley & Sons, Inc., 1971).
9. J. G. Cragg and B. G. Malkiel. "The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings," *Journal of Finance* (March 1968), pp. 67–84.
10. W. T. Dent and J. A. Swanson. "On Forecast Error in Autoregressive Integrated Moving Average Models," Working Paper No. 75-3, College of Business Administration, The University of Iowa, February 1975.
11. E. J. Elton and M. J. Gruber. "Earnings Estimates and the Accuracy of Expectational Data," *Management Science* (April 1972), pp. 409–424.
12. G. Foster. "Stock Market Reaction to Estimates of Earnings per Share by Company Officials," *Journal of Accounting Research* (Spring 1973), pp. 25–37.
13. ———. "Quarterly Accounting Data: Time Series Properties and Predictive-Ability Results," *Accounting Review*, (January 1977), pp. 1–21.
14. I. Friend and M. Puckett. "Dividends and Stock Prices," *American Economic Review* (September 1964), pp. 656–682.
15. J. R. Froeschle. "The Analysis of Seasonal Time Series," Doctoral Dissertation, The University of Iowa, 1975.
16. N. Gonedes. "Properties of Accounting Numbers: Models and Tests," *Journal of Accounting Research* (Autumn 1973), pp. 212–237.

17. W. L. Hays and R. L. Winkler. *Statistics: Probability, Inference and Decision* (New York: Holt, Rinehart and Winston, Inc., 1971).
18. R. C. Higgins. "Growth, Dividend Policy and Capital Costs in the Electric Utility Industry," *Journal of Finance* (September 1974), pp. 1189–1201.
19. J. E. Kiger. "An Empirical Investigation of NYSE Volume and Price Reactions to the Announcement of Quarterly Earnings," *Journal of Accounting Research* (Spring 1972), pp. 113–128.
20. R. H. Litzenberger and C. U. Rao. "Estimates of the Marginal Time Preference and Average Risk Aversion of Inventors in Electric Utility Shares 1960–66," *The Bell Journal of Economics and Management Science* (Spring 1971), pp. 265–277.
21. B. G. Malkiel and J. G. Cragg. "Expectations and the Structure of Share Prices," *American Economic Review* (September 1970), pp. 601–617.
22. R. G. May, "The Influence of Quarterly Earnings Announcements in Investor Decisions as Reflected in Common Stock Price Changes," *Empirical Research in Accounting: Selected Studies, 1971*. Supplement to Vol. 9, *Journal of Accounting Research*, pp. 119–163.
23. J. F. Muth. "Rational Expectations and the Theory of Price Movements," *Econometrica* (July 1961), pp. 315–335.
24. C. R. Nelson. "Rational Expectations and the Predictive Efficiency of Economic Models," *The Journal of Business* (July 1975), pp. 331–343.
25. H. Theil, *Applied Economic Forecasting* (Amsterdam: North-Holland Publishing Company, 1966).