Faculty Working Papers

PRELIMINARY EVIDENCE ON THE DESCRIPTIVE AND PREDICTIVE PROPERTIES OF GPPA EARNINGS DATA IN THE AIRLINES INDUSTRY

William A. Hopwood, Assistant Professor,
Department of Accountancy
William A. Hillison, Florida State University
Kenneth S. Lorek, Florida State University

#711

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 22, 1980

PRELIMINARY EVIDENCE ON THE DESCRIPTIVE AND PREDICTIVE PROPERTIES OF GPPA EARNINGS DATA IN THE AIRLINES INDUSTRY

William A. Hopwood, Assistant Professor,
Department of Accountancy
William A. Hillison, Florida State University
Kenneth S. Lorek, Florida State University

#711

Summary

This research provides some preliminary evidence on the predictive properties of general price level adjusted income. The results indicated that the price level adjustment process did not improve univariate predictability.

The accounting literature is replete with suggestions that the forecasting of future earnings is an important consideration in the investment choice process. The Security and Exchange Commission has even considered making the inclusion of supplementary earnings forecasts mandatory (Wall Street Journal, 1978). More recently, the FASB, in its Statement of Financial Accounting Concepts #2, has stressed the importance of forecasted values to users of accounting information: (1980, p. 24)

> Users can be expected to favor those sources of information and analytical methods that have the greatest predictive value in achieving their specific objectives.

In response to the importance of providing predictions of future earnings, extensive research efforts have been directed toward the specification of <u>both</u> the times series properties and predictive ability of historical cost (HC) earnings data.[1] Alternative parsimonious models have been suggested as descriptive of both <u>quarterly</u> and <u>annual</u> HC earnings numbers. But virtually no descriptive or predictive evidence on general purchasing power adjusted (GPPA) earnings data has been provided since FASB Statement #33 was promulgated. A primary reason for conducting such research concerns the possibility that changing price levels have induced noise in HC income series perhaps resulting in a detrimental effect on predictive ability. To the extent that noise reduction results in enhancement of predictive ability, users of accounting information might generate more accurate predictions of future earnings. These predictions could then be employed more efficiently in the investment decision making process.

Along these lines, Hendriksen (1970, p. 213) states:

> Historical dollar reporting serves a useful purpose in providing accountability for stewardship of cash funds, for cash flow analysis, and in tracing dollars through the business process. . . . Adjustments should be made for price-level changes in order to provide better information for measuring efficiency, to provide a basis for prediction of future income, and for managerial decision making." (emphasis ours)

FASB's Statement of Financial Accounting Concepts #2 (1980, p. 24) states:

> ". . . For Eg., the econometric models now used for economic forecasting are designed to use as data financial aggregates (among other things) as those aggregates are compiled at present. They might work less well if price level adjusted data were used. However, it might be possible to revise the model for use with that kind of data so that even better predictions could be made" (emphasis ours)

The objectives of this paper are to provide preliminary descriptive and predictive empirical evidence on the time series properties of GPPA data. The impact of general purchasing power adjustments on the income series is assessed via structural comparisons of GPPA and HC time series models. Additional information is obtained by comparing goodness of fit measures and monitoring the behavior of sample variances for both series. Comparisons of the predictive ability of GPPA and HC earnings data are also provided. Empirical evidence on this set of surrogate evaluators (descriptive and predictive ability) for HC and GPPA earnings data should provide valuable input to standard setting bodies and decision makers.

PREDICTIVE ABILITY CRITERION

The predictive ability criterion has long been championed as an operational test of the usefulness of alternative income measures. The 1966 ASOBAT Committee (1966, p. 24) discussed its importance as follows:

> The past earnings of the firm are considered to be the most important single item of information relevant to the prediction of future earnings.

Beaver, Kennelly and Voss (1968) suggested that the criterion has potential applicability in the assessment of the effects of alternative income measurements: (p. 685)

> . . . alternative accounting measurements are evaluated in terms of their ability to predict events of interest to decision-makers. The measure with the greatest predictive power with respect to a given event is considered to be the 'best' method for that purpose.

However, Revsine (1971) and Greenball (1971) have advanced convincing arguments concerning the artifactual nature of predicting future levels of income and the jointness inherent in evaluating the predictive ability of a set of data with a given prediction model. We agree with the spirit of these caveats which certainly preclude the rank ordering as to desirability of HC vis-a-vis GPPA via the predictive ability criterion. However, the FASB has made the importance of future earnings a primary consideration in their conceptual framework project. More specifically, the FASB has stated (1976, p. 55)

> Earnings from an enterprise for a period measured by accrual accounting are generally considered to be the most relevant indicator of relative success or failure of the earnings process of an enterprise in bringing in needed cash.

The reasoning behind this statement can be traced to four propositions discussed in the conceptual framework project:

(1) The primary interest of the investor is in a return on his investment in the form of cash flows (p.45).
(2) Earnings as measured by accrual accounting are generally thought to be the most relevant indicator of an enterprise's cash earning ability (p.45).
(3) Fundamental financial analysis focuses on the earning power of an enterprise in estimating the intrinsic value of the stock (p. 57).
(4) The most important single factor in determining a stock's value is now held to be the indicated average future earning power (p.57).

These factors suggest the importance of assessing the predictive ability of GPPA versus HC data. A-priori reasoning invoked by ASOBAT, FASB pronouncements and standards, and SEC releases has not provided convincing evidence regarding the potential utility of generating GPPA data. The empirical evidence provided in this study is not sufficient to prove or disprove the a-priori reasoning inherent in the above works, but it will provide preliminary empirical evidence on the descriptive and predictive properties of GPPA accounting measurements.

DATA

Data requirements were substantial due to the GPPA adjustment model and the Box-Jenkins methodology both of which were employed and have stringent data needs. In effect, these two requirements represent both cross-sectional as well as time series constraints. Detailed quarterly financial information was necessary for each sample firm to enable the generation of reasonable GPPA estimations. The time series requirement suggests that a sufficiently long data base of quarterly earnings numbers must be generated for each firm to apply the prediction model.

These requirements were met by the airline industry. Firms in this industry are required to report detailed quarterly financial statement information to the Civil Aeronautics Board (CAB). The information is consolidated and reported by the CAB in the Air Carrier Financial Statistics. Our sample is comprised of 24 firms in the airline industry on which sufficient cross-sectional and time series data were available. We began the analysis with 30 firms but 6 firms were eliminated due to data unavailability. Sixty-seven quarters of data were collected beginning with the first quarter 1962 and terminating with the third quarter, 1978. Although our sampling criteria and data requirements precluded a random sample from being drawn, we feel that a significant number of firms in the airlines industry was represented. A list of sample firms is provided in Appendix A.

GPPA MODEL CONSIDERATIONS

The Davidson and Weil Model (1975) was used to estimate GPPA quarterly "operating profit or (loss)" with some minor modifications. First, the Consumer Price Index for all Urban Consumers (CPI) was employed rather than the GNP Implicit Price Deflator Index. FASB has recently specified the use of the CPI due to its ready availability and the fact that it may be a better indicator of the effects of inflation for financial statement users.[2] Second, since the airline industry is predominantly service oriented, problems of inventory adjustment (i.e., FIFO vs. LIFO) were effectively avoided.

Third, the CAB Air Carrier Financial Statistics did not report the actual depreciation method employed by reporting firms. However, we were able to obtain the depreciation methods used by more than half the sample firms from alternative sources. In all cases the straight line method was used for external reporting purposes. We therefore assumed that the remaining sample firms also used the straight-line method. This assumption appears reasonable and was also employed by Parker (1977) in his GPPA study using COMPUSTAT data.

Fourth, unlike previous research which analyzed annual GPPA data, this study concentrates on quarterly data. By employing quarterly data, we have implemented certain refinements in the GPPA estimation process. Basically, all GPPA estimation models must invoke a proportionality or averaging assumption regarding the occurrence of revenue and expense items throughout the period. For example, it would be assumed that a firm which reported $10 million in annual sales would have generated those sales evenly throughout the year. Actual quarterly sales data allow more detailed specification of the seasonality patterns inherent in the data (i.e., perhaps actual sales were $1 million, $2 million, $4 million and $3 million per quarter). The extensive literature on the time series properties of interim accounting data is supportive of such seasonality patterns (see Foster (1977)). It is suggested that GPPA adjustments on quarterly data might benefit from these patterns.

GPPA MODEL VALIDATION

We estimated GPPA earnings because detailed financial data are not presently available to permit calculation of "actual" GPPA earnings on an extensive time series basis. Thus the relative predictive ability of GPPA earnings vis-a-vis HC earnings is an interesting research question only if the results of the estimation procedure are a reasonable surrogate for actual GPPA earnings. Although model validation for typical GPPA estimation procedures is incomplete, evidence indicates consistently that the estimation procedures provide satisfactory surrogates for actual GPPA numbers.

Davidson and Weil (1975), Basu and Hanna (1975), and Ketz (1978) provide evidence concerning the accuracy of models similar to the one used in this study. Basu and Hanna's evidence is the most comprehensive for earnings data providing 53 total years' observations over 23 firms. They report actual vs. estimated median difference at .1% with a 95% confidence interval -4.4% to +3.2%. Although these results must be considered relative to the assumptions made and biases induced the estimation procedure appears reasonable.

Ketz compared several similar estimation models, Petersen (1971), Davidson and Weil (1975) and Parker (1977), with actually calculated airline data. The actually calculated GPPA data were generated by McKenzie (1970) from Civil Aeronautic Board data. Only balance sheets were prepared, however, so the validation evidence is not as comprehensive as with Basu and Hanna. It was concluded by Ketz that

> All three of the algorithms were found to be good estimations of the general price level balance sheets . . . and any of them would be a valid tool to use in general price level studies. (p. 959)

Ketz's findings are particularly germane to the present study given our concentration on airline industry data.

JUSTIFICATION OF BOX-JENKINS

Since any predictive ability test jointly examines the data and the prediction model, we have selected a prediction model which has evidenced relatively high predictive power using quarterly earnings data. Foster (1977) and Lorek (1979), among others, have demonstrated the superiority of Box-Jenkins time series analysis in both one-step ahead and multi-step ahead predictions of quarterly earnings vis-a-vis sets of relatively simplistic predictors in the Green & Segall (1967) tradition. The impressive track record of these quarterly time series models partially obviates the impact of the jointness caveat in any predictive test.

We have specifically selected quarterly (rather than annual) GPPA and HC earnings data for several reasons: 1) By looking at shorter data bases (1960's-1970's) the potential for structural change is reduced, 2) Seasonality patterns in quarterly data may lead to refinements in the GPPA transformations and 3) Predictive ability tests using Box-Jenkins analysis on quarterly data bases have proved more encouraging than applications employing annual data (See Albrecht, Lookabill and McKeown (1977) and Watts and Leftwich (1977)).

DESCRIPTIVE EVIDENCE

The time series properties of GPPA quarterly earnings data are virtually unknown. However, time series properties of HC quarterly earnings have been studied extensively. Several findings are summarized from this rapidly expanding literature: 1) HC quarterly earnings exhibit quarter to quarter and quarter by quarter relationships, 2) Seasonality is manifested either by seasonal differencing of the data or combinations of seasonal autoregressive and moving average parameters, and 3) Parsimonious models have generally outperformed firm-specific models in predictive testing on holdout samples. In this paper, we report several different kinds of descriptive evidence on GPPA and HC quarterly earnings for our sample firms. These include information on model structure, goodness of fit criteria and variance information.

These sources provide collectively some preliminary information of the impact of GPPA adjustments on the time series properties of quarterly earnings in the airline industry. Further research will determine the generalizability of these results across other firms, industries, and time periods.

Box-Jenkins time series models were identified for each of the 24 sample firms using the algorithm reported by Hopwood (1980)[3]. Firm specific models were identified for both GPPA and HC data since a suggested parsimonious structure for GPPA quarterly earnings data is presently unavailable. Due to the updating evidence provided by McKeown and Lorek (1978), and certain constraints induced by the predictive ability testing reported in a later section of this paper, 15 different models were identified for each stream (HC and GPPA) with number of observations (NOB) ranging from 52-66. In Tables 1 and 2 we summarize the underlying structures of the firm-specific autoregressive-integrated-moving average (ARIMA) models for HC and GPPA respectively.[4]

Table 1 reports that the most commonly identified structure for HC Quarterly Earnings Data is one which includes regular parameters <u>and</u> seasonal parameters.[5] 260 of 360 models are in this combination mode: a) AR(1) and seasonal AR(1) with 177 models, b) AR(1) and seasonal MA(1) with 54 models, c) MA(1) and seasonal AR(1) with 10 models and, finally, sundry multiplicative regular and seasonal forms with 19 models. The

- -

Insert Tables 1 and 2

- -

most frequently identified non-multiplicative model: a simple AR(1) accounts for 76 of the remaining 100 models. Recent empirical evidence by Griffin (1977), Foster (1977), Lorek (1979), and Collins and Hopwood (1980) supports the combinational structure for HC quarterly earnings. The results for the airline industry HC quarterly earnings are generally consistent with this characterization in that quarter-to-quarter movements (regular parameters) and quarter-by-quarter movements (seasonal parameters) were most frequently identified.

An examination of the P,Q columns of Table 1 indicates that there were only 79 of the 360 identified models that did <u>not</u> have a seasonality parameter, whether it was autoregressive or moving average. However, 18 of these 79 models required seasonal differencing (D) of the data, so that seasonality, in the form of explicit parameters (P,Q) and/or seasonal differencing (D) is present in 299 out of 360 models (83%). These findings are consistent with the pervasive seasonality evidence provided by Coates (1972), Foster (1977) and Lorek (1979).

As Table 1 indicates, the firm-specific models for HC quarterly earnings were relatively parsimonious in nature. Due to the recent proliferation of alternative structures for parsimonious models for quarterly earnings, we did not expect the Foster (100)X(010), Griffin (011)X(011) or the Brown Rozeff (100)X(011) models to dominate the firm-specific models. However, alternative parsimonious characterizations did. The two most likely candidates for generally representative parsimonious models for the airline industry were (100)X(110) with 84 cases and (100)X(100) with 70.[6]

Table 2 presents descriptive evidence on the GPPA models. Some interesting patterns are apparent when comparing the GPPA models with the HC models. First, 263 or 360 models included regular and seasonal parameters: a) AR(1) and seasonal AR(1) with 216 models; b) AR(1) and

Table 1

HC Time Series Models

I Parameter Combinations:

p	q	P	Q	Frequency
1	1	1	1	3
1	1	0	1	1
1	1	1	0	1
1	1	0	0	2
0	1	1	1	0
0	1	0	1	7
0	1	1	0	10
0	1	0	0	0
1	0	1	1	7
1	0	0	1	54
1	0	1	0	177
1	0	0	0	76
0	0	1	1	6
0	0	0	1	6
0	0	1	0	9
0	0	0	0	1
				360

II Differencing Combinations

d	D	Frequency
0	0	199
1	0	8
0	1	143
1	1	10
		360

0 for p, q, P, Q implies no identified parameter

1 for p, q, P, Q implies at least one identified parameter

p = autoregressive parameter

q = moving average parameter

P = seasonal autoregressive parameter

Q = seasonal moving average parameter

d = regular differencing

D = seasonal differencing

Table 2

GPPA Time Series Models

I Parameter Combinations

p	q	P	Q	Frequency
1	1	1	1	0
1	1	0	1	4
1	1	1	0	0
1	1	0	0	7
0	1	1	1	6
0	1	0	1	2
0	1	1	0	1
0	1	0	0	1
1	0	1	1	12
1	0	0	1	22
1	0	1	0	216
1	0	0	0	55
0	0	1	1	9
0	0	0	1	14
0	0	1	0	7
0	0	0	0	4
				360

II Differencing Combinations

d	D	Frequency
0	0	166
1	0	37
0	1	118
1	1	39
		360

All parameters defined in Table 1.

seasonal MA(1) with 22 models; and c) AR(1) and both seasonal AR(1) and MA(1) with 12 and sundry multiplicative models with 13 models. Overall, the GPPA transformations had a markedly small impact upon these general combinational forms. We do note, however, that there was a greater concentration (216 vs. 177) in the AR(1) and seasonal AR(1) mode.

Second, we also observe that the most frequently encountered non-multiplicative model was still the simple AR(1) process with a frequency of 55. Third, non-seasonal models accounted for 67 occurrences. Of these, seven used seasonal differencing which resulted in 300 out of 360 models (83%) exhibiting some form of seasonality. These findings are virtually identical to the HC results. Finally, the same candidates for possible airline parsimonious models were identified: (100)X(110) with 82 cases and (100)X(100) with 80.

To summarize this section, the overall tenor of the results indicates that the structure of the identified models did not change substantially when moving from HC to GPPA data. In fact, we note that the (pdq)X(PDQ) model structures were identical for 137 of the 360 total models identified (38%). On the other hand, we recognize that parameter values may still vary significantly even if the structures were comparable. Finally, we note that all Box-Pierce Q statistics for HC and GPPA indicated no significant residual autocorrelation (α .05). Comparisons of Box-Pierce Q Statistics further revealed similar goodness of fit information for HC and GPPA. For 172 (47.8%) models the HC Q statistics were smaller than the corresponding GPPA Q statistics for the same firm and NOB. For 188 models (52.2%) the reverse was true.

VARIANCE INFORMATION

An interesting empirical question concerns whether GPPA transformations serve as a variance reduction mechanism vis-a-vis HC numbers.[7] This is an important consideration in time series research because the Box-Jenkins method assumes that the variance of the input data is constant over time [Box and Jenkins, 1973, p. 26]. To the extent that temporal variance explosion occurs, there can be a loss in predictive power of the BJ statistical model. Evidence is presented in Tables 3 and 4 that the constant variance assumption is violated for both HC <u>and</u> GPPA quarterly earnings data although the GPPA transformations do serve as a <u>relative</u> variance reduction mechanism.

Tables 3 and 4 present information on the sample variances for the raw data (d=0, D=0) and the seasonally differenced data (d=0, D=1) for both GPPA and HC quarterly earnings for NOB = 52, 56, 60 and 64. For each of these data base lengths, the total number of observations were divided into two equal parts. Sample variances were generated for both subperiods and the ratio:

$$\left(\frac{S_2^{\ 2}}{S_1^{\ 2}}\right)$$

was derived. S Ratios greater than one indicate increasing temporal variance since the variance of the more recent data is in the numerator while the variance of the older data is in the denominator. Overall, the results indicate that the GPPA transformation process decreases the degree to which the constant variance assumption is violated; however, even the GPPA data indicate an excessive number of firms which exhibited increased variances. Note that for the raw form HC data in Table 3, 92 of 96 data bases yielded increasing (>1) S ratios. Raw form undifferenced

Table 3

HC and GPPA S Ratios - Raw Data

Firm #	52 H/C	52 GPPA	56 H/C	56 GPPA	60 H/C	60 GPPA	64 H/C	64 GPPA
1	2.11	2.10	1.39	1.34	1.03	.83	1.29	.75
2	7.50	4.01	6.66	3.10	8.52	3.67	8.27	3.47
3	25.61	15.21	30.74	12.95	19.38	6.57	31.23	9.52
4	4.54	2.63	3.68	2.02	2.43	1.22	2.67	1.24
5	3.23	2.66	3.49	2.44	3.40	1.73	3.55	1.76
6	2.69	1.78	2.45	1.46	3.10	1.40	3.39	1.42
7	2.78	1.01	3.22	2.02	2.98	1.92	3.90	2.06
8	1.59	1.37	1.41	1.12	1.72	1.19	1.91	1.20
9	2.65	1.93	3.00	2.23	3.11	2.15	4.19	2.49
10	2.56	1.65	2.18	1.30	2.12	1.06	2.08	1.14
11	2.90	1.71	3.02	1.81	3.58	1.96	3.49	1.75
12	1.44	1.11	1.07	.80	.89	.52	.86	.41
13	25.92	11.06	20.29	8.51	18.09	5.85	9.49	3.25
14	1.41	1.15	1.30	.95	1.73	.89	1.89	.71
15	20.64	12.22	14.51	7.15	9.95	3.14	12.58	2.82
16	2.36	1.80	2.58	1.64	2.79	1.55	2.80	1.40
17	17.92	6.11	22.35	6.59	18.64	4.83	12.16	2.80
18	.55	.34	.74	.42	1.16	.54	1.63	.65
19	7.39	4.73	9.52	4.84	10.07	4.53	6.92	2.61
20	19.19	6.22	26.04	7.70	27.52	7.71	31.67	7.36
21	30.76	9.12	13.78	4.90	15.76	4.15	8.76	2.12
22	4.36	2.69	4.64	2.62	5.14	2.50	4.98	2.29
23	4.52	2.53	4.87	2.57	5.72	3.21	5.44	2.97
24	5.74	3.42	4.23	1.99	3.33	1.18	2.37	.74

GPPA data yield 83 of 96 increasing S ratios.[8] However, on a relative basis, the GPPA transformations served to reduce the S ratio from its HC level on every occasion.[9] Table 4 reveals S ratios greater than one for 89 of 96 seasonally differenced HC series and 84 of 96 GPPA series. GPPA and HC relative S ratio comparisons revealed that the GPPA transformations reduced the S-ratio again on every occasion. Thus, the empirical evidence supports the relative variance reduction properties of the GPPA transformations. In the next section the predictive performance of these models is assessed.

- -

Insert Tables 3 and 4

- -

PREDICTIVE RESULTS

A primary motivation for this study was to assess the relative predictive ability of HC and GPPA data. To accomplish this objective, we employed the time series models for HC and GPPA discussed previously in a predictive context. The specific predictive hypothesis in null format follows:

> Ho: There is no difference in mean absolute percentage error between one to eight step-ahead forecasts of HC quarterly earnings generated from HC time series models and forecasts of GPPA quarterly earnings generated from GPPA time series models.

We employed the MAPE and MSE metrics to assess the accuracy of the predictions where:

$$MAPE = \left| \frac{(P-A)}{A} \right|$$

$$MSE = \left| \frac{(P-A)}{A} \right|^2$$

Table 4

HC and GPPA S Ratios - Seasonal Differences

	52		56		60		64	
Firm #	H/C	GPPA	H/C	GPPA	H/C	GPPA	H/C	GPPA
1	2.02	1.59	1.59	1.21	1.45	1.03	1.29	.77
2	4.36	2.60	2.58	1.64	1.78	.98	1.08	.58
3	13.01	7.81	16.02	5.44	26.45	10.25	23.19	9.50
4	4.26	2.94	5.75	3.56	3.66	2.23	3.97	2.15
5	17.13	9.59	13.03	7.04	15.75	6.04	12.33	5.57
6	2.77	1.56	3.26	1.53	3.81	1.66	3.96	1.71
7	1.97	.99	5.83	2.44	7.04	3.01	7.65	3.18
8	2.56	1.52	3.36	1.59	3.29	1.61	3.15	1.52
9	2.21	1.50	2.05	1.54	2.03	1.56	2.53	1.61
10	6.46	5.91	7.89	4.36	3.30	1.44	3.79	1.59
11	6.17	3.60	6.79	3.56	6.64	3.21	3.62	1.79
12	.96	.66	.85	.58	.94	.60	1.36	.69
13	16.79	8.30	24.81	10.22	16.10	7.56	15.52	7.09
14	4.87	2.32	4.19	1.87	4.90	2.14	4.73	2.05
15	12.77	9.61	7.40	4.35	9.97	5.23	9.24	4.71
16	2.74	2.57	2.89	2.60	2.14	1.87	2.23	1.18
17	5.42	4.64	16.61	6.29	9.20	3.80	8.64	3.78
18	.20	.12	.23	.13	.43	.19	.74	.28
19	5.93	3.97	6.61	3.82	7.37	3.60	7.28	3.48
20	9.95	5.28	19.92	8.20	24.60	9.48	17.95	7.06
21	12.59	7.93	19.41	9.20	6.69	3.17	6.20	2.68
22	2.43	1.66	2.61	1.65	3.71	2.13	2.75	1.69
23	2.52	1.38	4.07	1.68	5.02	1.91	3.16	1.59
24	7.67	5.37	4.42	2.22	2.40	1.07	2.53	1.10

P = predicted quarterly earnings

A = actual quarterly earnings

MAPE gives equal weight to all forecast errors and is consistent with a linear loss function. MSE provides proportionately greater weight to large forecast errors and is consistent with a quadratic loss function.[10] Since the results were not affected by choice of error metric, MAPE is used for exposition purposes.

Overall data base length for both the HC and GPPA series was 67 quarters. It began with first quarter 1962 and ended with third quarter, 1978.[11] Firm-specific time series models were identified for 15 different data base lengths for each firm with the NOB ranging from 52-66. Modeling was performed by an improved version of the algorithm discussed by Hopwood (1980). We selected 52 as the minimum NOB for model identification to minimize the impact of small sample bias in parameter estimation.[12]

For each of the first eight models identified per firm (NOB 52-59), we generated one to eight step-ahead forecasts inclusively. Since total data base length was limited to 67 observations, we were unable to generate all 8 step-ahead forecasts for the remaining seven models per firm (NOB = 60-66). One to 7 step-ahead forecasts were obtained for NOB = 60, one to 6 for NOB = 61 and analogously a single one step-ahead forecast was obtained for NOB = 66. In summary, beyond NOB = 59, the forecast horizon decreased as the number of observations increased on a one for one basis.

- -

Insert Table 5

- -

Table 5

Forecast Error Matrix For Each Sample Firm

NOB \ Forecast Horizon	Step-Ahead Predictions							
	1	2	3	4	5	6	7	8
52	53	54	55	56	57	58	59	60
53	54	55	56	57	58	59	60	61
54	55	56	57	58	59	60	61	62
55	56	57	58	59	60	61	62	63
56	57	58	59	60	61	62	63	64
57	58	59	60	61	62	63	64	65
58	59	60	61	62	63	64	65	66
59	60	61	62	63	64	65	66	67
60	61	62	63	64	65	66	67	
61	62	63	64	65	66	67		
62	63	64	65	66	67			
63	64	65	66	67				
64	65	66	67					
65	66	67						
66	67							
	—	—	—	—	—	—	—	—

Table 5 presents the structure for the matrix of forecast errors for each sample firm. Cell values in this matrix represent the particular predicted values for each model. This matrix may be partitioned into an 8x8 matrix for NOB 50-59 and a triangular matrix for NOB 60-66. Casting the forecast errors in this manner results in a comparison of 92 forecasts (HC vs. GPPA) for each sample firm and 2,208 forecasts in total (24x92).

The null hypothesis was tested by using the non-parametric Wilcoxon-Signed Ranks test discussed by Siegel (1956). We were particularly interested in a comparison of the differences in the MAPE for HC and GPPA. Since the statistical test was performed on MAPE differences rather than levels, the impact of cross-sectional dependencies in the models across NOB's and the forecast horizon was mitigated. We also note that the MAPE error metric was truncated at 100% for those explosive errors generated either by the HC or GPPA models.[13]

Table 6 provides the results of testing the predictive hypothesis. It assesses predictive ability across the entire 8-step ahead forecast horizon for all 24 sample firms. Negative cell values indicate superior predictive performance ($MAPE_{GPPA} > MAPE_{HC}$) for the HC models while positive cell values indicate superior predictive performance ($MAPE_{HC} > MAPE_{GPPA}$) for the GPPA models.

- -

Insert Table 6

- -

Table 6

MAPE Error Differences - HC vs. GPPA

NOB \ Forecast Horizon	Step-Ahead Prediction Errors 1	2	3	4	5	6	7	8
52	.074	-.089	-.205	-.014	-.011	-.158	-.162	-.022
53	-.042	-.044	.032	.031	-.095	-.082	-.095	.037
54	-.175	-.062	.008	-.095	-.105	-.022	.014	-.197
55	-.037	-.025	-.071	-.173	-.117	-.105	-.136	-.154
56	.023	-.039	-.183	-.035	-.093	-.050	-.166	-.071
57	-.135	-.205	.091	-.016	-.095	-.140	-.154	-.040
58	-.164	.077	.027	-.118	-.110	-.061	-.136	-.251
59	.033	-.053	-.184	-.064	.004	-.077	-.253	-.041
60	-.015	-.123	-.040	.037	.005	-.234	.002	
61	-.108	.018	.028	.057	-.188	-.016		
62	-.044	-.049	-.082	-.206	-.014			
63	-.086	-.135	-.264	-.006				
64	-.203	-.261	-.088					
65	-.355	-.093						
66	-.055							
$\bar{\bar{X}}$	-.086	-.077	-.071	-.050	-.075	-.095	-.121	-.092
Wilcoxon Test Statistic	-4.32[a]	-3.52[a]	-3.47[a]	-2.59[b]	-3.03[b]	-4.82[a]	-3.94[a]	-3.51[a]

a = significant at α = .001, two tailed test.

b = significant at α = .01, two tailed test.

Note that the HC predictions were significantly more accurate than the GPPA predictions for all 8 step-ahead forecasts. The significance levels were α = .001 for 1-3, 6-8 and α = .01 for 4 and 5 steps ahead.[14] As the grand mean row suggests, one-step ahead HC predictions were .086 percent more accurate than the GPPA predictions. The smallest difference was .050 for 4 step-ahead predictions while the greatest difference was .121 for 7-step-ahead predictions.

Although the negative signs of the grand mean cell values and the reported α levels both underscore the pervasive pattern of HC predictive dominance, we do note particular (NOB/Horizon) combinations in the body of the matrix with positive cell values. For example, when NOB=52 one-step-ahead GPPA predictions were .074 more accurate than HC predictions. In summary, only 18 of 92 cell values indicate such positive values.

Supplemental tests were conducted to assess the sensitivity of the results to the particular truncation value (100%) invoked. Table 7 provides some descriptive results on the magnitude of the explosive error problem.[15]

Table 7

Truncation %	H/C # Explosive MAPE / Total MAPE	GPPA # Explosive MAPE / Total MAPE
100%	683/2208 = 30.9%	998/2208 = 45.1%
200%	335/2208 = 15.2%	480/2208 = 21.7%
300%	202/2208 = 9.1%	347/2208 = 15.7%
400%	160/2208 = 7.2%	257/2208 = 11.6%
500%	135/2208 = 6.1%	209/2208 = 9.5%
1000%	86/2208 = 3.9%	103/2208 = 4.7%

Table 7 highlights the relative impact of these explosive MAPE's on HC and GPPA predictions. It shows the greater percentage of explosive errors for GPPA relative to HC. We have constructed MAPE error matrices analogous to Table 6 for the alternative truncation percentages, and the reported results are insensitive to these alternative truncation percentages.

DISCUSSION OF RESULTS AND CONCLUSIONS

Descriptive findings suggest that GPPA transformations serve as a relative variance reduction mechanism and do not radically alter the quarterly earnings time series properties. Since firm-specific BJ models were identified, clustering of (pdq)X(PDQ) models across firms/horizons around the (100)X(110) and (100)X(100) structures suggests that these represent possible parsimonious models for the airline industry. Note that these models were identified frequently for both the HC and GPPA data.

A surprising result centered on the statistical dominance of the HC predictions over the GPPA predictions. Our priors suggested the opposite and the descriptive results reported above appeared to confirm this - i.e., GPPA transformations reduced the variance in the quarterly earnings series vis-a-vis the HC series. An ex-post justification for our predictive findings rests on the particular time series pattern of variance reduction. Since the BJ prediction model simply extrapolates past behavior into the future, haphazard variance reduction in the GPPA data base could have swamped the GPPA forecasts with an error component if the actual variance reduction pattern changed during the holdout period. In fact, the erratic pattern of inflation across the holdout data base periods is consistent with this interpretation.

The descriptive-predictive paradox reported herein may not be as surprising as we suggest given the lack of empirical evidence on GPPA data. In fact, a prominent Big-8 CPA firm has even suggested that surprising results are to be expected in this area. Peat, Marwick, Mitchell & Co. states: (1980, p. 1)

> The story is now beginning to unfold, and indeed the numbers are rough, controversial and perhaps confusing.

Although our results are data, sample and industry specific as well as being subject to the propriety of the GPPA adjustment model employed, we feel that the results are of importance to standard setting bodies like FASB and SEC as well as users in general. If the recent call for supplementary GPPA data by FASB in Statement #33, is based explicitly or implicitly on supposedly improved predictive ability, our results suggest the opposite. Future research on "actual" GPPA data will address the generalizability of the reported findings.

APPENDIX A

Sample Firms

1. Airlift International
2. Alaska Airlines
3. Allegheney Airlines
4. Aloha Airlines
5. American Airlines
6. Braniff Airways
7. Continental Airlines
8. Delta Airlines
9. Eastern Airlines
10. Tiger International Airlines
11. Hawaiian Airlines
12. National Airlines
13. North Central Airlines
14. North West Airlines
15. Ozark Airlines
16. Pan American Airways
17. Piedmont Airlines
18. Reeve Airlines
19. Seaboard World Airlines
20. Southern Airways
21. Texas International Airlines
22. Trans World Airlines
23. UAL (United Airlines)
24. Western Airlines

Footnotes

1. See Foster (1977), Brown & Rozeff (1979) and Lorek (1979), for examples of recent work in the time series area.

2. See FASB Statement #33.

3. Several enhancements were made to the Hopwood algorithm. These include:
   1) APTRNID and PPTRNID variables were **not set to one** due to the seasonality of the data (criteria 3, p. 293, Hopwood, 1980).
   2) Lower order models were examined prior to consideration of higher order models.
   3) Parameters for lags other than one or two (or seasonal multiples thereof) were not considered due to perceived sampling variation.
   4) Invertability and stationarity requirements were invoked.

4. The specific parameter values and the related ACF and PACF information are available from the authors upon request.

5. Since each sample firm had 15 models identified (NOB = 52-66), there are 360 models (24 sample firms X 15 models each) summarized in Tables 1 and 2.

6. We leave for future research the determination of industry specific parsimonious time series models. Abdulkhader, Icerman and Lorek (1980) have examined several industries and provide preliminary results on industry specific time series models for earnings and sales data.

7. See Hopwood (1979) for a discussion of this issue.

8. The null hypothesis of equal variances between the two subperiods (S_1 and S_2) can be rejected for both the GPPA and HC data using the sign test at $\alpha = .01$. This applies to the raw data in Table 3 and the seasonally differenced data in Table 4.

9. We only report on the raw and seasonally differenced data because of the high concentration of (pdq)X(PDQ) structures which these transformations summarized i.e., 95% of the HC models and 79% of the GPPA models were represented by these series.

10. See Demski and Feltham (1972) for a discussion of surrogate criteria used in the evaluation of alternative forecast models.

11. Data availability dictated the particular starting and ending points in the data base.

12. See Lorek and McKeown (1978) for a discussion of the tradeoff between structural change and small sample bias in this context.

13. Foster (1977) and Brown and Rozeff (1979) employed similar truncation schemes in predictive tests employing the MAPE metric.

14. Note that the data reported in Table 6 are potentially dependent across adjacent time periods. Since we conducted eight pairwise comparisons, a more conservative test would be to require the level of significance to be 1 percent per test to avoid drawing invalid inferences. This makes the probability of accepting the hypothesis of significant difference when none exists less than 8 percent. The null hypothesis is similarly rejected using this criterion. Brown and Rozeff (1979) used a similar analysis. See Miller (1966) for a discussion of the Bonferroni inequality in this context.

15. We examined the potential causes for the explosiveness of the MAPE metric: 1) denominator values approaching zero and 2) explosive forecasts in the numerator. The primary reason for these explosive errors is due to the former cause.

References

Abdulkhader, A., J. D. Iceman and K. S. Lorek, "Time Series Properties of Quarterly Sales and Earnings Data: The Industry Effect." Unpublished Paper, Florida State University, 1980.

Albrecht, W. S., L. L. Lookabill and J. C. McKeown, "The Time Series Properties of Annual Earnings: An Analysis of Individual Firms," Journal of Accounting Research (Autumn 1977), 226-44.

American Accounting Association, A Statement of Basic Accounting Theory (Evanston, IL: AAA), 1966.

Basu, S. and J. H. Hanna, Inflation Accounting: Alternatives, Implementation Issues and Some Empirical Evidence. Toronto, Society of Industrial Accountants of Canada, 1975.

Beaver, W. H., J. W. Kennelly, and W. M. Voss, "Predictive Ability as a Criterion for the Evaluation of Accounting Data," The Accounting Review (October, 1968), 675-83.

Box, G. E. P. and G. M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, 1970.

Brown, L. D. and M. S. Rozeff, "Univariate Time Series Models of Quarterly Earnings Per Share: A Proposed Premier Model," Journal of Accounting Research (Spring, 1979), 179-189.

Coates, R., "The Predictive Content of Interim Reports - A Time Series Analysis," Empirical Research in Accounting: Selected Studies, 1972. Supplement to Journal of Accounting Research, 132-44.

Collins, William A. and William S. Hopwood, "A Multivariate Analysis of Annual Earnings Forecasts Generated From Quarterly Forecasts of Financial Analysts and Univariate Time Series Models," Forthcoming Journal of Accounting Research, (Autumn, 1980).

Davidson, S. and R. Weil, "Inflation Accounting," Financial Analysts Journal, (January/February, 1975), 27-31, 70-84.

Demski, J. S., and G. A. Feltham, "Forecast Evaluation," The Accounting Review, (July, 1972), pp. 533-48.

Financial Accounting Standards Board, An Analysis of Issues Related to Conceptual Framework for Financial Accounting and Reporting. Elements of Financial Statements and Their Measurements. December, 1966.

__________, Tentative Conclusions on Objectives of Financial Statements of Business Enterprises, 1976.

__________, Statement of Financial Accounting Standards No. 1: Objectives of Financial Reporting and Elements of Financial Statements of Business Enterprises. 1977.

________. Statement of Financial Accounting Standards No. 33, Financial Reporting and Changing Prices. September, 1979.

________. Statement of Financial Accounting Concepts No. 2: Quantitative Characteristics of Accounting Information. May, 1980.

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

Gonedes, Nicholas J. and Nicholas Dopuch, "Economic Analyses and Accounting Techniques: Perspectives and Proposals" Journal of Accounting Research (Autumn, 1979), 384-450.

Green, D., Jr. and J. Segall, "The Predictive Power of First Quarter Earnings Reports," Journal of Business (January, 1967), 44-55.

Greenball, M. N., "The Predictive-Ability Criterion: Its Relevance in Evaluating Accounting Data," Abacus (June, 1971), 1-7.

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

Hendriksen, Eldon S., Accounting Theory, Homewood, IL: Richard D. Irwin, Inc., 1970.

Hillison, W. A., "Empirical Investigation of General Purchasing Power Adjustments on Earnings Per Share and the Movement of Security Prices," Journal of Accounting Research (Spring, 1979), 60-72.

Hopwood, William S., "An Empirical Investigation into the Effect of Changes in the General Price Level on the Time Series Properties of Quarterly Earnings Per Share," Working Paper #581, University of Illinois (1979).

Hopwood, W. S., "On the Automation of the Box-Jenkins Modeling Procedures: An Algorithm With An Empirical Test," Journal of Accounting Research (Spring, 1980), 278-288.

Ketz, J. E., "The Validation of Some General Price Level Estimating Models," The Accounting Review (October, 1978), 952-959.

Lorek, K. S. and J. C. 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), 204-214.

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

McKenzie, P. B., "The Relative Usefulness to Investors of Price-Level Adjusted Financial Statements: An Empirical Study," (Ph.D. Dissertation, Michigan State University, 1970).

McKeown, J. C. and K. S. 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), 658-672.

Miller, Rupert G., Simultaneous Statistical Inference (McGraw-Hill, 1966).

Parker, J. R., "Impact of Price Level Accounting," The Accounting Review (January, 1977), 69-96.

Peat, Marwick, Mitchell & Co., The New Inflation Data, A Survey of Annual Reports, (June, 1980).

Peterson, R., "Interindustry Estimation of General Price Level Impact on Financial Information," The Accounting Review (January, 1973), 34-43.

Revsine, L., "Predictive Ability, Market Prices, and Operating Flows," The Accounting Review (July, 1971), 480-89.

"SEC Again to Urge Profit Projections, Other Operating Forecasts by Companies," The Wall Street Journal (February 16, 1978), p. 6.

Watts, R. L. and R. W. Leftwich, "The Time Series of Annual Accounting Earnings," Journal of Accounting Research (Autumn, 1977), 253-271.