NORTHERN ARIZONA UNIVERSITY
College of Business Administration

The Contextual Nature of the Predictive Power of Statistically-Based Quarterly Earnings Models

Working Paper Series—05-18 | December 2005

Kenneth S. Lorek*
G. Lee Willinger**

Current version as of November 15, 2005

* *Corresponding author:*
Kenneth S. Lorek
Northern Arizona University
College of Business Administration
P.O. Box 15066
Flagstaff, AZ 86011-5066
U.S.A.
(928) 523-7406
Ken.Lorek@nau.edu
fax : (928) 523-7331

** Price College of Business
University of Oklahoma
Norman, OK 73019
(405) 325-5780
lwilling@ou.edu

The authors are indebted to the participants of the brown bag series at Northern Arizona University and the accounting research colloquium at The Florida State University for comments on an earlier draft. We are especially indebted to comments received from Bruce Billings and Rick Morton.

The Contextual Nature of the Predictive Power of Statistically-Based Quarterly Earnings Models

Kenneth S. Lorek and G. Lee Willinger

I. Introduction

The time-series properties and predictive ability of quarterly earnings numbers have evoked considerable research efforts in accounting since the late 1970s [See Foster (1977), Brown and Rozeff (1979), Lorek (1979), and Brown (1993), among others]. Salient empirical findings from the time-series literature suggest that the time-series properties of quarterly earnings may be well described by both quarter-to-quarter (adjacent) and quarter-by-quarter (seasonal) effects. "Premier" ARIMA processes such as those originally popularized by Foster (1977) [(100) X (010) with drift]; Griffin (1977) and Watts (1975) [(011) X (011)] and Brown and Rozeff (1979) [(100) X (011)] were identified to model simultaneously the well documented, adjacent and seasonal dependencies in quarterly earnings data.[1] These so-called "premier" ARIMA models have consistently outperformed simplistic random walk and seasonal random walk models in predictive-ability tests on holdout samples in the aforementioned works. In fact, this dominance of the ARIMA models versus simplistic statistically-based alternatives has been a common thread across three decades of time-series research in accounting.

Recent evidence provided by Baginski et al. (2003) that quarterly earnings persistence has declined through time in addition to the increased incidence of loss quarters experienced by our sample firms raises doubts about the current ability of ARIMA models to track autocorrelation such that predictive ability is enhanced. Whether these ARIMA models have maintained superiority over naïve alternatives in more recent time periods is an important issue to those seeking efficient quarterly earnings expectation models. We provide current empirical evidence on this issue.

We explicitly invoke the predictive ability criterion originally popularized by Beaver, Kennelly, and Voss (1968) in accounting. While alternative evaluative criteria may certainly be employed to assess the usefulness of accounting data (i.e., capital market association testing and/or behavioral experimentation, among others), Bernard (1995) has argued forcefully for continued use of the predictive ability criterion in empirical financial-based research. Our current empirical work is consistent with the spirit of Bernard's recommendations.

There are several reasons to examine the efficacy of these "premier" ARIMA models using current data bases. First, the decade of the 1990s experienced earnings growth that was pervasive across most industries and sectors of the U. S. economy. Baldwin (2002) suggests that "….in the wild bull market of the 1990s, they were reacting to the uptick in earnings growth" (p. 114) (where "they" refers to "investors"). This explosive growth in earnings may have altered the time-series properties of quarterly earnings and/or the efficacy of the premier ARIMA models that were identified nearly three decades ago. At a minimum, the predictive power of the respective ARIMA models may be differentially affected.

Second, close scrutiny of the sampling filters used in extant time-series works reveals that certain firms were systematically precluded from entering test samples, namely high-technology and regulated firms. This may simply have been a function of when the original time-series studies in accounting were performed since many high-technology firms (e.g., Dell, Cisco Systems, among others) were nonexistent when the premier ARIMA models for quarterly earnings were originally identified in the mid-1970s. In fact, the relatively long quarterly earnings data bases necessary for time-series analysis have implicitly biased samples toward including larger, manufacturing-oriented, New York Stock Exchange firms (i.e., so-called "old economy firms"). On the other hand, regulated firms (i.e., banks, insurance companies and utilities) were typically omitted due to concerns pertaining to the potential effect that the rate regulation process might have on the income generating process. For example, complex reimbursement formulas mandated by regulatory agencies impact both the sales and earnings streams of utility companies. Researchers did not want these potentially idiosyncratic quarterly earnings numbers to affect the cross-sectional sample autocorrelation functions (SACFs) that they used to identify the structure of the "premier" ARIMA models. High-technology and regulated firms are of particular interest to us in the current study since their quarterly earnings time-series properties are not well understood.[2]

Third, insights regarding the time-series properties of quarterly earnings have been attained incrementally in a piecemeal fashion given researchers' inability to identify an optimal ARIMA structure that describes the quarterly earnings series for all firms across time. Examples of this approach include Lorek and Bathke (1984) and Brown and Han (2000) who successfully identified or employed ARIMA structures for nonseasonal firms as well as Lorek, Stone and Willinger (1999) who assessed the differential predictive ability of opaque and transparent firms' quarterly earnings series. We continue in this tradition of partitioning sample firms into high-technology, regulated and default samples and assess predictive performance separately for each sample of firms.

Our primary objective is to examine the time-series properties and predictive ability of the quarterly earnings numbers of the aforementioned sample of firms using a database that begins with the first quarter of 1990 and ends with the fourth quarter of 2002. In this manner, the inter-temporal stability of the "premier" ARIMA model structures may be assessed relative to certain naïve benchmark models. Consistent with extant time-series work, we find that the premier ARIMA models are well specified descriptively for all three samples of firms. However, in marked contrast to extant work, high-technology firms exhibit predictive ability results consistent with a simplistic random walk with drift expectation model. These findings are at variance from those that we report for regulated firms where we provide predictive ability results that demonstrate the superiority of the Griffin-Watts ARIMA model.

We also provide supplementary evidence that the predictive performance of ARIMA models is sensitive to partitioning the regulated firm sample further into two distinct subsamples of firms: 1) utilities, and 2) banks, savings and loans, and insurance companies (i.e., financial institutions).

We present empirical results pertaining to three additional variables: analyst coverage, the number of loss quarters experienced by our sample firms, and earnings persistence. The effect of analyst coverage on predictive ability is salient given the need for capital market researchers to employ quarterly earnings expectations for firms that are not covered by analysts. Additionally, we assess whether the statistical extrapolations of quarterly earnings are affected by the extent to which a firm is covered by analysts (i.e., no coverage, moderate coverage or extensive coverage). We also provide empirical data documenting the frequency with which our sample firms report loss quarters, a proxy for the volatility of the income generating process. Our findings reveal a dramatic increase in the number of loss quarters experienced by our sample firms across the holdout period that we employ (i.e., 2000 – 2002), especially for high-technology firms. Finally, consistent with Baginski et al. (2003), we provide evidence of reduced levels of earnings persistence for all of our sample firms relative to the levels of earnings persistence reported for firms using data from previous decades.

We begin our paper with a background section which discusses related time-series work followed by discussion of our sampling procedures and a descriptive analysis of our sample firms, modeling considerations, predictive results, and supplementary analyses. We end by providing a discussion on the study's limitations, some concluding remarks, and suggestions for future research.

2. Background

We specifically examine three distinct categories of firms: high-technology, regulated, and default sample firms. Francis and Schipper (1999) provide a convincing rationale for being interested in examining the earnings expectation models of high-technology firms by stating that: "….to the extent that earnings predictions become more sophisticated and *accurate (*emphasis added) over time, the news content of earnings announcements per se will be reduced" (p. 326). Since high-technology firms were not widely represented in the samples originally examined by time-series researchers, the accuracy of statistically-based quarterly earnings expectation models for these firms has not been thoroughly assessed. Moreover, high-technology firms typically operate in an environment of rapid change in their factor input and product output markets, rely heavily upon technological innovations, and invest heavily in research and development making the predictive ability of quarterly earnings series increasingly more challenging for such firms.[3]

To our knowledge, we are the first researchers to assess the predictive power of ARIMA-based models using an identification database of quarterly earnings that begins with the first quarter of 1990. This allows us to include proportionately more newly-formed firms with lesser concentration on the "old economy" firms that have dominated extant time-series work. By partitioning sample firms in the manner that we suggest, we also avoid the quixotic search for a singular optimal ARIMA model that has superior predictive power across all firms and time. Partitioning also allows us to assess whether seasonal autocorrelation (forecast error metrics) of the quarterly earnings series of regulated firms is greater (lower) than that of high-technology firms due to the impact of the rate regulation process described above and the seasonal demands for the output of such firms.

We were also influenced by Brown (1999) who believes that earnings surprise research can conduct more powerful tests by not assuming that all firms quarterly earnings series follow the same time-series model. In fact, Bathke, Lorek and Willinger (2004) refine inferences regarding the earnings-return relationship by controlling for the descriptive goodness-of-fit of the seasonal random walk with drift model (i.e., firms were partitioned into good-fit and bad-fit subsets based on the value of the Box-Ljung Q-statistic computed on the residual series of each sample firm conditioned upon the aforementioned quarterly earnings expectation model).[4]

Our results suggest that the performance of alternative statistically-based quarterly earnings expectation models is highly contextual based on the type of firm under examination. This is a sobering finding when viewed from the perspective that over one-third of our sample firms are not covered by security analysts, which we document in the supplementary analysis section of the paper. For these firms, researchers must use a statistically-based quarterly earnings expectation model. The selection of the expectation model for these firms takes on added importance given Walther's (1997) finding that investors in lightly followed firms overrely on the predictions of earnings generated from statistically-based models, relative to earnings predictions of financial analysts.. Our findings suggest that the choice of the best statistically-based model is not an obvious one.

3 . *Method*

3.1 Sampling Procedures

We obtained a subpopulation of 1,216 calendar year-end firms that had complete quarterly earnings data available for the 1990 to 2002 period on the Quarterly Compustat File.[5] We used undeflated quarterly net income numbers before extraordinary items similar to the series analyzed by Foster (1977), Griffin (1977), Lorek (1979), and Bathke, Lorek and Willinger (1989), among others. From this subpopulation we derived three subsets of firms: high-technology firms (n=202), regulated firms (n=218), and default firms (n=796), which serve as our primary test samples. The default sample simply includes all firms with complete time-series data other than the high-technology and regulated firms in the first two subsets.

We employed the partitioning scheme originally popularized by Francis and Schipper (1999) to identify high-technology firms. Basically, the high technology sample (see table 1) includes firms in the computer, electronics, pharmaceuticals and telecommunications industries. The five industries with the greatest concentration of high-technology firms were: SIC #283 drugs (n=54); SIC #737 computer programming, software and data processing (n=34); SIC #366 communications equipment (n=26); SIC #367 electronic components & semiconductors (n=24); and SIC #481 telephone communications (n=21). These five industries comprised approximately 79% of the high-technology sample listed in table 1.[6]

Table 1
Industries Included in High-Technology Sample
(n=202)

SIC CODE	NUMBER OF SAMPLE FIRMS	INDUSTRY
283	54	Drugs
357	19	Computer and Office Equipment
362	8	Electrical Industrial Apparatus
363	4	Household Appliances
364	7	Electrical Lighting and Wiring Equipment
365	2	Household Audio, Video Equipment, Audio Receiving
366	26	Communications Equipment
367	24	Electronic Components, Semiconductors
481	21	Telephone Communications
737	34	Computer Programming, Software, Data Processing
873	3	Research, Development, Testing Services
	202	

Table 2 lists the sample firms in the regulated sample (n=218). We included firms in the banking, electrical services, gas production, and insurance industries in this sample. The five industries with the greatest concentration of regulated firms were: SIC #602 commercial banks (n=55); SIC #491 electric services (n=32); SIC #493 combinational utility services (n=24); SIC #633 fire, marine and casualty insurance (n=24); and SIC #492 gas production and distribution (n=15). These five industries comprised approximately 69% of the regulated firm sample listed in Table 2.

Table 2
Industries Included in Regulated Sample
(n=218)

SIC CODE	NUMBER OF SAMPLE FIRMS	INDUSTRY
491	32	Electric Services
492	15	Gas Production and Distribution
493	24	Combinational Utility Services
494	10	Water Supply
602	55	Commercial banks
603	7	Savings Institutions
631	10	Life Insurance
632	12	Medical Service & Health Insurance
633	24	Fire, Marine & Casualty Insurance
	29	Miscellaneous Industries with 4 or less firms
	218	

Firms with SIC codes not reflected in Tables 1 and 2 were assigned to a default sample. As expected, the default sample evidenced considerable variety in industry representation with 46 different industries with 5 or more sample firms and 153 different industries with 4 or less sample firms using three-digit SIC codes. In general, the default sample firms provide a comparative benchmark against which the time-series properties and predictive ability of the quarterly earnings numbers of high-technology and regulated sample firms might be assessed.

3.2 Profile Analysis

Table 3 presents sample profile information on firm size, profitability, and the number of loss quarters experienced by firms in the test samples: high-technology (n=202), regulated (n=218), and default (n=796). We are interested in firm size since Bathke et al. (1989) provide evidence that predictive ability of statistically-based ARIMA models is positively related to firm size (i.e., larger firms have significantly smaller earnings forecast errors than smaller firms). Inspection of table 3 reveals sizable differences across test samples in the median market value of common stock equity computed on December 31, 1999, the end of the model identification period. Regulated firms are, on average, the largest with a median market value of common stock equity of $1,548.8 million, followed by high-technology firms with $429.6 million, and the default sample with $299.1 million.

We also report the median and range of annual income numbers computed across the entire identification period for firms in our test samples. Regulated firms reported the largest median annual income of $71.6 million with default firms reporting $8.8 million and high-technology firms generating only $3.6 million. Due to the size-effect evidence reported in Bathke et al. (1989), it was expected that regulated firms would have the largest median annual income since these firms were, on average, approximately five times larger than the default firms and 3.6 times larger than the high-technology firms. On the other hand, the high-technology firms exhibited lower median annual income than the default firms despite exhibiting larger, median firm size ($429.6 million versus $299.1 million).

Table 3
Profile of Sample Firms

	Market Value Of Common Stock Equity* (in millions)		Annual Income for 1990 – 1999 period (in millions)		Number of Loss quarters	
	Median	Range	Median	Range	Mean	Range
High Technology (n=202)	$429.6	$.084 to $274,596	$3.6	$-7,987 to $7,712	11.83	0 to 40
Regulated (n=218)	$1,548.8	$.044 to $207,809	$71.6	$ -662.4 to $11,370	2.83	0 to 25
Default (n=796)	$299.1	$.111 to $508,044	$ 8.8	$-2,566.8 to $22,071	8.27	0 to 39

*computed as of December 31, 1999

Table 3 also presents the mean and range of the average number of loss quarters experienced by our sample firms across the forty quarterly income numbers in the identification period of 1990 – 1999. We observe that the regulated (high-technology) firms have the smallest (largest) number of mean loss quarters with 2.83 (11.83). The default firms had 8.27 mean loss quarters during this time period. In the supplementary analysis section, we provide additional empirical evidence suggesting that increases in the number of loss quarters during the holdout period may help explain the decline in predictive power of ARIMA models, especially for high-technology firms.

Table 4
Cross-sectional Sample Autocorrelations: 1990-1999
(Means and Standard Deviations)

Panel A: Levels of quarterly earnings (d=0, D=0)

	Lags										
1	2	3	4	5	6	7	8	9	10	11	12
High-technology firms (n=202)											
.382	.296	.239	.242	.156	.108	.087	.094	.039	.008	-.010	-.007
(.279)	(.278)	(.258)	(.248)	(.217)	(.215)	(.189)	(.182)	(.154)	(.142)	(.123)	(.128)
Regulated firms (n=218)											
318	.249	.237	.390	.159	.105	.100	.245	.036	-.004	-.008	.124
(.390)	(.395)	(.320)	(.272)	(.280)	(.294)	(.227)	(.228)	(.190)	(.209)	(.145)	(.214)
Default firms (n=796)											
.322	.223	.186	.248	.105	.059	.051	.109	.011	-.021	-.023	.026
(.288)	(.293)	(.249)	(.266)	(.215)	(.213)	(.179)	(.193)	(.146)	(.148)	(.121)	(.147)

Panel B: Seasonal differences of quarterly earnings (d=0, D=1)

	Lags										
1	2	3	4	5	6	7	8	9	10	11	12
High-technology firms (n=202)											
.231	.128	.037	-.290	-.034	-.034	-.020	-.022	-.012	-.003	-.015	-.023
(.245)	(.159)	(.161)	(.211)	(.177)	(.169)	(.151)	(.146)	(.127)	(.122)	(.120)	(.144)
Regulated firms (n=218)											
.151	.076	.030	-.315	-.008	.010	.008	.001	-.006	-.029	-.031	-.011
(.248)	(.165)	(.185)	(.206)	(.166)	(.154)	(.145)	(.133)	(.131)	(.116)	(.114)	(.115)
Default firms (n=796)											
229	.116	.026	-.295	-.052	-.040	-.035	-.036	-.019	-.018	-.019	-.017
(.236)	(.167)	(.164)	(.215)	(.157)	(.143)	(.141)	(.149)	(.125)	(.121)	(.126)	(.133)

Table 4 (Continued)

Panel C: Consecutive and seasonal differences of quarterly earnings (d=1, D=1)

	Lags											
	1	2	3	4	5	6	7	8	9	10	11	12
High-technology firms (n=202)												
	-.350	-.004	.148	-.364	.124	-.007	.007	-.009	.002	.013	.004	-.004
	(.207)	(.122)	(.173)	(.172)	(.178)	(.169)	(.149)	(.155)	(.145)	(.134)	(.127)	(.134)
Regulated firms (n=218)												
	-.387	-.022	.165	-.378	.141	.016	.010	-.002	.008	-.010	-.016	.008
	(.167)	(.122)	(.181)	(.167)	(.165)	(.156)	(.145)	(.143)	(.134)	(.117)	(.129)	(.131)
Default firms (n=796)												
	-.346	-.008	.142	-.353	.110	.003	.004	-.014	.010	.003	-.001	.006
	(.202)	(.124)	(.175)	(.192)	(.171)	(.146)	(.142)	(.154)	(.136)	(.127)	(.125)	(.130)

3.3 Descriptive Analysis of Sample Firms

Table 4 presents the average cross-sectional SACFs across lags 1 through 12 for the three samples: high-technology (n=202), regulated (n=218) and default firms (n=796) computed for the quarterly earnings data over the 1990-1999 identification period.[7] Panel A presents these values for the raw data or levels of quarterly net income (d=0, D=0). We observe that all three samples exhibit relatively slow decay in the SACF values across lags consistent with potential nonstationarity. Consistent with our expectations, we also note that the regulated firms exhibit considerably greater values at the seasonal lags of 4, 8, and 12 (.390, .245, .124) vis-à-vis those exhibited by the high-technology (.242, .094, -.007) and default firms (.248, .109, .026), respectively.

The SACF values of the consecutively differenced data (d=1, D=0) did not exhibit any systematic differences across samples and are, therefore, not reported. Panel B contains the SACF values of the seasonally-differenced series (d=0, D=1). Inspection of this panel reveals that all three samples exhibit autoregressive decay across the first three lags of their respective SACFs: high-technology (.231, .128, .037), regulated (.151, .076, .030), and default firms (.229, .116, .026) coupled with negative spikes at the seasonal lag of four. This pattern is consistent with the identification of the Brown-Rozeff, premier ARIMA model (100) X (011).[8] Finally, panel C contains the SACFs of the consecutively and seasonally-differenced series (d=1, D=1) which exhibit spikes at lags one and four across subsamples consistent with the Griffin-Watts, premier ARIMA model (011) X (011). We discuss these ARIMA model structures more thoroughly in the next section.

3.4 Prediction Models

We assessed the forecast accuracy of three parsimonious ARIMA time-series models advocated in the accounting literature [See Brown (1993), among others] and two naïve benchmark models. The structural forms of these models, as well as our rationale for selecting them, are provided below.

Model 1): Random Walk with drift: (RWD)

$$E(Q_t) = Q_{t-1} + \delta \qquad (1)$$

where:

Q_t = quarterly earnings at time t

Q_{t-1} = quarterly earnings at time t - 1

δ = deterministic trend constant

The RWD model suppresses seasonality, does not require firm-specific parameter estimation aside from the deterministic trend constant, and serves as a control against potential structural changes in the holdout period. Although the RWD model has been systematically outperformed in extant time-series work on the predictive ability of quarterly earnings, the volatile period over which we assess predictive performance (i.e., 2000-2002), the decline in quarterly earnings persistence reported by Baginski et al. (2003), and the RWD's parsimonious nature make it an interesting candidate model to examine.

Model 2): Seasonal Random Walk with drift: (SRWD)

$$E(Q_t) = Q_{t-4} + \delta \quad (2)$$

where:

Q_{t-4} = quarterly earnings at time t - 4

We include the SRWD model for several reasons. First, it resulted in superior market association metrics (i.e., cumulative average residuals) in Foster (1977) despite the fact that it was not the best in terms of predictive ability. Second, the SRWD model has been employed extensively in the earnings-return literature by Bernard and Thomas (1990) and Ball and Bartov (1996), among others, as the primary proxy for the security market's expectation of quarterly earnings. Third, the SRWD model does not require firm-specific parameter estimation aside from the deterministic trend constant and is relatively parsimonious in nature. Finally, rather than suppressing seasonal effects like the RWD model, its expectations are based entirely upon seasonal patterns in the data. In this sense, it is the opposite of the RWD model which bases its expectations exclusively on adjacent (i.e., quarter-to-quarter) effects.

Model 3): Foster's ARIMA model (100) X (010) with drift

$$E(Q_t) = Q_{t-4} + \varphi_1(Q_{t-1} - Q_{t-5}) + \delta \quad (3)$$

where:

φ_1 = autoregressive parameter

Q_{t-5} = quarterly earnings at time t - 5

Foster's premier ARIMA model is a simple autoregressive process with a drift term superimposed upon the seasonally-differenced series. Due to the ease with which its parameters may be estimated, it has become the most popular of the premier ARIMA models.[9] It serves as a bridge between the simpler naïve benchmark models discussed above and the more complex ARIMA models that we discuss next.

Model 4): Brown and Rozeff's ARIMA model (100) X (011)

$$E(Q_t) = Q_{t-4} + \varphi_1(Q_{t-1} - Q_{t-5}) - \acute{\O}_1 a_{t-4} \quad (4)$$

where:

$\acute{\O}_1$ = seasonal moving average parameter

a_{t-4} = disturbance term at time t - 4

Brown and Rozeff (1979) provide evidence that Foster's premier ARIMA model does not fully account for residual autocorrelation at lag 4. Therefore, they added a seasonal moving-average parameter ($\acute{Ø}_1$) to remedy this potential deficiency. Inspection of the autocorrelation patterns of the quarterly earnings series of our three samples of firms reveals behavior consistent with this ARIMA process [see Panel B) of Table 4].

Model 5): Griffin-Watts ARIMA model (011) X (011)

$$E(Q_t) = Q_{t-4} + (Q_{t-1} - Q_{t-5}) - \acute{ø}_1 a_{t-1} - \acute{Ø}_1 a_{t-4} - \acute{ø}_1 \acute{Ø}_1 a_{t-5} \quad (5)$$

where:

$\acute{ø}_1$ = regular moving-average parameter

a_{t-1} = disturbance term at time t

a_{t-5} = disturbance term at time t - 5

Watts (1975) and Griffin (1977) identified this relatively complex ARIMA process that contains both regular and seasonal differences as well as regular ($\acute{ø}_1$) and seasonal ($\acute{Ø}_1$) moving-average parameters combined with a series of distributed-lag, disturbance terms (i.e., a_{t-1}, a_{t-4}, a_{t-5}). Inspection of the autocorrelation patterns of the quarterly earnings series of our three samples of firms also reveals behavior consistent with this ARIMA process [see panel c) of Table 4].

The holdout period for predictive testing consisted of the three-year period: 2000-2002. The parameters of the aforementioned expectation models were estimated using a quarterly earnings data base beginning with the first quarter, 1990 and ending with the fourth quarter, 1999 in order to generate the first quarter, 2000 predictions. The parameters were then re-estimated prior to making each additional one-step-ahead prediction over the 2000-2002 period. Therefore, the forecast profile contains twelve one-step-ahead quarterly earnings forecasts for each of the five expectation models. That is, the high-technology sample of 202 firms yields 2,424 origin-date forecasts, the regulated sample of 218 firms yields 2,616 origin-date forecasts, and the default sample of 796 firms yields 9,552 origin-date forecasts.

The accuracy of the forecasts was assessed using the mean absolute percentage error (MAPE):[10]

$$MAPE = 1/n \ \Sigma \ | (A - F) / A | \quad (6)$$

where:

n = number of sample firms

A = actual quarterly earnings

F = forecasted quarterly earnings

This error metric has been used extensively in extant time-series work [See Foster (1977) and Bathke and Lorek (1984), among others] facilitating inter-temporal comparisons of the predictive findings.[11]

4. *Predictive Results*

Panel A of Table 5 contains the MAPE metrics for the high-technology sample for the one-step-ahead quarterly earnings predictions across the five expectation models (RWD, SRWD, F, BR, GW) for each individual quarter (1st, 2nd, 3rd, 4th), as well as on a pooled basis across all quarters and years. The accuracy of the predictions was assessed by using the Friedman ANOVA ranks test (Hollander and Wolfe 1973). For each firm, the prediction model yielding the smallest absolute percentage error was given a rank of one, the next smallest error was given a rank of two and so on until the model yielding the largest error was given a rank of five. Panel A also provides the average rank of each prediction model and Friedman's S-statistic and its associated level of significance for each individual quarter and on a pooled bases across quarters and years.

Table 5

Mean Absolute Percentage Errors of One-Step-Ahead Quarterly Earnings Predictions (2000-2002) - High-technology firms (n=202)

Panel A: Overall Results

Model	1st Qtr Avg Rank	1st Qtr MAPE	2nd Qtr Avg Rank	2nd Qtr MAPE	3rd Qtr Avg Rank	3rd Qtr MAPE	4th Qtr Avg Rank	4th Qtr MAPE	Pooled Avg Rank	Pooled MAPE
RWD	3.04	.594	2.70	.500	2.69	.510	2.80	.547	2.81	.538
SRWD	3.12	.592	3.26	.607	3.29	.618	3.20	.621	3.22	.610
F	2.96	.591	3.08	.591	3.07	.591	3.03	.604	3.03	. 594
BR	2.99	.580	3.11	.576	3.07	.586	3.10	.605	3.07	.587
GW	2.89	.578	2.85	.549	2.89	.556	2.87	.591	2.88	.568
Friedman ANOVA S-statistic	6.92		47.95		49.39		26.66		102.72	
Significance Level	.14		.0001		.0001		.0001		.0001	

where:

RWD = random walk with drift
SRWD = seasonal random walk with drift
F = Foster's (100) X (010) with drift ARIMA model
BR = Brown and Rozeff's (100) X (011) ARIMA model
GW = Griffin-Watts' (011) X (011) ARIMA model

Table 5 (Continued)

Panel B: Paired Comparisons Based on Ranks of Prediction Models on a Pooled Basis – High-Technology Firms

Model (Average rank)	SRWD (3.22)	F (3.03)	BR (3.07)	GW (2.88)
RWD (2.81)	RWD***	RWD***	RWD***	RWD**
SRWD (3.22)		F***	BR***	GW***
F (3.03)			—	GW***
BR (3.07)				GW***

where:

***	= significant at .001
**	= significant at .01
*	= significant at .05
--	= non-significant

Inspection of the results presented in Panel A of Table 5 reveals the following: First, the RWD model provides the lowest pooled MAPE (.538) as well as the lowest MAPE for quarters two (.500), three (.510), and four (.547).[12] The RWD model, however, provides the highest MAPE for quarter one (.594).[13] Second, the RWD model provides the lowest average ranks among the five prediction models for the same periods where it exhibits the lowest MAPEs. Third, the SRWD model provides the highest pooled MAPE (.610). Fourth, the levels of MAPEs reported in Table 5 are markedly higher across models than those reported by Foster (1977) and Bathke and Lorek (1984), among others. In fact, Foster's best-performing model was the F ARIMA model with MAPEs across quarters ranging from .218 to .287 for the 1962-1974 holdout period. Bathke and Lorek's best-performing model, on the other hand, was the BR ARIMA model which had a pooled MAPE of .369 across the 1975-1977 holdout period. It is evident that there appears to be a substantial deterioration in predictive power of statistically-based quarterly earnings expectation models, especially for the high-technology sample firms. We provide some reasons why this might be happening in the supplementary analysis section.

Panel A of Table 5 also reveals a statistically significant difference in the average ranks of the prediction models (p=.0001) for the second, third, and fourth quarters in addition to the pooled results across quarters and years. Panel B provides all pairwise comparisons between each of the prediction models using the pooled predictions.[14] Recall that lower rankings imply a lower MAPE and greater predictive ability. Each cell value in Panel B contains the superior prediction model for the row-column comparison as well as its corresponding significance level. Specifically, the RWD model outperforms the SRWD, F and BR ARIMA models at p=.001 and the GW ARIMA model at p=.01. The statistical dominance of the RWD model over the ARIMA-based models in the pooled predictions is a unique finding in the time-series literature in accounting, standing in marked contrast to extant time-series work.[15] The GW ARIMA model outperforms the SRWD, F, and BR ARIMA models at p=.001. Finally, the SRWD model is outperformed by the rest of the models at p=.001.

Table 6

Mean Absolute Percentage Errors of One-Step-Ahead Quarterly Earnings Predictions (2000-2002) - Regulated firms
(n=218)

Panel A: Overall Results

Model	1st Qtr Avg Rank	1st Qtr MAPE	2nd Qtr Avg Rank	2nd Qtr MAPE	3rd Qtr Avg Rank	3rd Qtr MAPE	4th Qtr Avg Rank	4th Qtr MAPE	Pooled Avg Rank	Pooled MAPE
RWD	3.11	.433	3.06	.437	2.92	.410	2.85	.485	2.98	.441
SRWD	2.99	.375	3.09	.418	3.15	.418	3.63	.426	3.09	.426
F	2.90	.387	2.95	.414	2.92	.414	2.99	.425	2.94	.425
BR	3.08	.379	3.06	.419	3.05	.409	3.17	.490	3.09	.424
GW	2.92	.393	2.84	.412	2.96	.402	2.86	.471	2.89	.419
Friedman ANOVA S-statistic	9.36		11.32		10.41		23.50		33.39	
Significance Level	.053		.023		.033		.0001		.0001	

where:

RWD = random walk with drift
SRWD = seasonal random walk with drift
F = Foster's (100) X (010) with drift ARIMA model
BR = Brown and Rozeff's (100) X (011) ARIMA model
GW = Griffin-Watts' (011) X (011) ARIMA model

Table 6 (Continued)

Panel B: Paired Comparisons Based on Ranks of Prediction Models on a Pooled Basis – Regulated Firms

Model (Average rank)	SRWD (3.09)	F (2.94)	BR (3.09)	GW (2.89)
RWD (2.98)	RWD***	--	RWD***	GW***
SRWD (3.09)		F***	--	GW***
F (2.94)			F***	--
BR (3.09)				GW***

where:
*** = significant at .001
** = significant at .01
* = significant at .05
-- = non-significant

Table 6 reports similar predictive data for the regulated firm sample. Unlike the predictive performance evidenced by the high-technology sample, we find, on average, a much smaller range in pooled MAPEs with the GW ARIMA model exhibiting the lowest pooled MAPE (.419) and the RWD model exhibiting the highest pooled MAPE (.441).[16] There is also a 22% reduction in pooled MAPEs when comparing the best prediction models for the high-technology and regulated samples of firms (i.e., .538 to .419).

Panel A of Table 6 reveals a statistically significant difference in the average ranks of the prediction models (p-values ranging from .033 to .0001) for the second, third, and fourth quarters as well as on a pooled basis across quarters and years. Panel B provides the pooled pairwise comparisons which indicate that the GW ARIMA model outperforms the RWD, SRWD, and BR ARIMA models at p=.001, but is insignificantly different from the F ARIMA model. The dominance of the GW ARIMA model over the three other expectation models is consistent with the more stable earnings environment in which these regulated firms operate relative to say, high-technology firms. Evidently, the GW ARIMA model is able to track autocorrelation in quarterly earnings across the estimation period more efficiently for regulated firms vis-à-vis high-technology firms. The F ARIMA model also outperforms the SRWD model and the BR ARIMA model at p=.001, but is insignificantly different from the RWD model (as well as the GW ARIMA model). Finally, the RWD model outperforms the SRWD and BR ARIMA models at p=.001, is insignificantly different from the F ARIMA model and is outperformed by the GW ARIMA model.

Table 7

Mean Absolute Percentage Errors of One-Step-Ahead Quarterly Earnings Predictions (2000-2002) Default firms (n=796)

Panel A: Overall Results

Model	1st Qtr Avg Rank	1st Qtr MAPE	2nd Qtr Avg Rank	2nd Qtr MAPE	3rd Qtr Avg Rank	3rd Qtr MAPE	4th Qtr Avg Rank	4th Qtr MAPE	Pooled Avg Rank	Pooled MAPE
RWD	3.17	.610	2.92	.518	2.82	.509	2.82	.583	2.93	.555
SRWD	3.01	.575	3.22	.570	3.25	.583	3.26	.646	3.18	.594
F	2.95	.574	2.93	.538	3.03	.563	3.02	.629	2.98	.576
BR	2.94	.560	3.03	.533	2.98	.538	3.06	.621	3.00	.563
GW	2.94	.576	2.90	.529	2.91	.543	2.84	.600	2.90	.562
Friedman ANOVA S-statistic	35.76		68.48		98.15		122.24		187.25	
Significance Level	.0001		.0001		.0001		.0001		.0001	

where:

RWD = random walk with drift
SRWD = seasonal random walk with drift
F = Foster's (100) X (010) with drift ARIMA model
BR = Brown and Rozeff's (100) X (011) ARIMA model
GW = Griffin-Watts' (011) X (011) ARIMA model

Table 7 (Continued)

Panel B: Paired Comparisons Based on Ranks of Prediction Models on a Pooled Basis – Default Firms

Model (Average rank)	SRWD (3.18)	F (2.98)	BR (3.00)	GW (2.90)
RWD (2.93)	RWD***	RWD***	RWD***	--
SRWD (3.18)		F***	BR***	GW***
F (2.98)			--	GW***
BR (3.00)				GW***

where:

***	= significant at .001
**	= significant at .01
*	= significant at .05
--	= non-significant

Table 7 reports predictive information pertaining to the default sample. Similar to the high-technology sample, we observe that the RWD model exhibits the lowest pooled MAPE (.555) while the SRWD model has the highest pooled MAPE (.594).[17] Additionally, the pooled MAPEs are relatively higher than those exhibited by the regulated firms, very similar to those reported earlier for the high-technology firms. Panel A of Table 7 reports significant Friedman ANOVA statistics (p=.0001) for each quarter and across quarters and years on a pooled basis. Panel B provides the pairwise comparisons on the ranks of the pooled MAPEs which indicate that the RWD and GW ARIMA models outperform the remaining models (p=.001) and are insignificantly different from each other. The SRWD model is outperformed by all models (p=.001). In the next section we provide information on: 1) the extent of analyst following, 2) the number of loss quarters reported by our sample firms during the holdout period, 3) the levels of quarterly earnings persistence computed for our sample firms, and 4) the effect on predictive ability of a finer partitioning of regulated firms into utilities and financial institutions to provide a contextual setting against which our predictive findings might be better assessed.

5. *Supplementary Analysis*

5.1 Analyst Following

We provide additional evidence on the extent to which analysts followed our sample firms. Analyst coverage is salient for our study for three reasons. First, recent evidence by Elgers, Lo, and Pfeiffer (2001) suggests that analyst following is positively associated with the efficiency with which investors process firm-specific information. Second, Walther (1997) reports that investors in lightly followed firms tend to overrely on time-series predictions of earnings, relative to earnings predictions of analysts. Third, to the extent that over one-third of our sample firms have no analyst coverage, empirical evidence on the MAPEs of statistically-based earnings forecast models is especially germane. Table 8 gives information on analyst coverage for firms in our three samples. We distinguish between firms with no analyst coverage at the end of the identification period, (i.e., December 31, 1999) moderate coverage (1-5 analysts), and extensive coverage (greater than 5 analysts) as reported by the *Institutional Brokers' Estimate System* (I/B/E/S).

Table 8
Analyst Coverage Information – I/B/E/S

Panel A: Frequency Distribution of Analyst Coverage

Sample	No-Coverage*	Moderate Coverage**	Extensive Coverage***	Total
High-technology	75 (37.1%)	60 (29.7%)	67 (32.2%)	202
Regulated	39 (17.9%)	63 (28.9%)	116 (53.2%)	218
Default	318 (39.9%)	217 (27.3%)	261 (32.8%)	796
	432 (35.5%)	340 (28%)	444 (36.5%)	1,216

Panel B: MAPEs by Size of Firm and Extent of Analyst Coverage

Sample		No-Coverage*		Moderate Coverage**		Extensive Coverage***		Total
		Small	Large	Small	Large	Small	Large	
High-technology	N	64	11	35	25	2	65	202
	MAPE	.609	.649	.580	.590	.767	.528	
		(p= .166)		(p= .704)		(p= .001)		
Regulated	N	31	8	56	7	22	94	218
	MAPE	.535	.462	.456	.539	.373	.376	
		(p=. 046)		(p= .064)		(p= .711)		
Default	N	255	63	131	86	12	249	796
	MAPE	.644	.547	.611	.484	.630	.504	1,216
		(p= .001)		(p= .001)		(p= .001)		

where:

*	= no analyst following of sample firm at end of identification period
**	= between 1-5 analysts following of sample firm at end of identification period
***	= greater than 5 analysts following of sample firm at end of identification period

Panel A of Table 8 reveals that across samples, 432 firms (35.5%) had no analyst coverage. This finding underscores the importance of identifying efficient statistically-based, quarterly earnings expectation models given the lack of analyst coverage for these firms. Moderate coverage (i.e., 1-5 analysts) was afforded to 340 firms (28%) while extensive coverage (i.e., greater than 5 analysts) was provided on 444 firms (36.5%). Recall that regulated firms (see table 3) were, on average, larger than high-technology or default sample firms, so it was not surprising that a greater percentage of regulated firms (53.2%) had extensive analyst coverage versus high-technology (32.2%) and default sample firms (32.8%).

While we expected that size and analyst coverage would be related as evidenced by the increased analyst coverage of the larger regulated firms, we were also interested in the extent to which the MAPEs of statistically-based, quarterly earnings expectation models were affected by levels of analyst coverage, especially given the recent evidence that investors in lightly followed firms overrely on statistically-based earnings forecasts.[18] Panel B presents pooled MAPEs computed across the five prediction models and the twelve quarters in the holdout period for each sample (high-technology, regulated, and default). MAPEs were computed across the varying levels of analyst coverage (none, moderate, and extensive) on each of the samples partitioned by size. That is, we divided each sample of firms into an equal number of small and large firm subsets based on the fair market value of common stock equity at the end of the model identification period. We then assigned sample firms into small and large categories nested under the extent of analyst coverage. Consistent with our expectations, Panel B reveals proportionately more large firms with extensive coverage as well as proportionately more small firms with no coverage across the three samples.

We used the Mann-Whitney U-test to determine which pairwise comparisons of MAPEs in Panel B were significantly different. We report significantly lower MAPEs for large firms in five of the pairwise comparisons: extensive coverage firms for the high-technology sample (p=.001), no-coverage firms for the regulated sample (p=.046) and no-coverage, moderate and extensive coverage firms for the default sample (p=.001). However, we also observe insignificant MAPE differences for the no-coverage (p=.166) and moderate coverage (p=.704) firms for the high-technology sample and the extensive coverage firms for the regulated sample (p=.711). There was also one instance, the comparison of moderate-coverage firms in the regulated sample, where the MAPEs of small firms were actually significantly smaller (p=.064). These results are consistent with analysts deciding, in certain instances, to cover firms that exhibit time-series properties more conducive to statistical extrapolation, independently of firm size. Perhaps time-series forecasts may be combined more readily with firm, industry, and macroeconomic information by analysts to generate more accurate, multivariate earnings forecasts for certain firms. Additional research is necessary to fully examine this issue.

5.2 Number of Loss Quarters

Table 3 presented profile information on the average number of loss quarters experienced by each sample of firms during the identification period (1990-1999). High-technology (default, regulated) sample firms experienced an average number of loss quarters of 11.83 (8.27, 2.83) during the 40 quarter identification period. Alternatively, high-technology (default, regulated) sample firms experienced quarterly net losses 29.6% (20.7%, 7.1%) of the time during the identification period. We computed similar percentages for the holdout period (2000-2002) that were 41.4% (24.9%, 11%) for the three samples, respectively. These increased incidences of loss quarters in the holdout period vis-à-vis the identification period, especially for the high-technology sample, may have contributed to the inflated MAPEs that we report. Moreover, the relatively strong performance of the RWD model, an extremely parsimonious model that does not require firm-specific parameter estimation aside from the deterministic trend constant, may partially explain why ARIMA-based models did not fare as well on the high-technology firms. Short-memory models, like the RWD model, appear to perform better in this setting.

5.3 Persistence

Baginski et al. (2003) document a significant decline in quarterly earnings persistence computed over a 35-year period for a sample of 172 New York Stock Exchange (NYSE) firms which were specifically selected by them because they were representative of “old economy”, capital-intensive firms. They derived an ARIMA-based proxy for quarterly earnings persistence based upon the BR ARIMA model structure and valuation theory.[19] Specifically, their persistence proxy declined significantly from a median of 19.49 in the 1967-1978 subperiod to 8.12 in the 1991-2001 subperiod for their “old economy” firms.. We employed the same persistence proxy used by Baginski et al. and computed it across our model identification period (i.e., 1990-1999) for the three separate samples of firms that we examine. The persistence proxy had median values of 8.17 for the high-technology firm sample, 8.07 for the default firm sample, and 9.54 for the regulated firm sample. These results serve to reinforce the findings of Baginski et al. that quarterly earnings persistence has declined through time and we provide new evidence that this phenomenon is not confined to the “old economy” firms that they examined. The relatively higher persistence proxy obtained for the regulated firm sample (9.54) is consistent with its enhanced level of predictive ability vis-à-vis the high-technology (8.17) and default samples (8.07). These reduced persistence levels, versus those obtained by Baginski et al. computed during earlier time periods, help explain why the RWD prediction model outperforms more sophisticated statistically-based, ARIMA expectation models for high-technology firms.[20]

5.4 Regulated Firms

We disaggregated the regulated firm sample further into two categories: utilities (n=87) comprised primarily of electric services, gas production, and water supply companies and financial institutions (n=131) comprised primarily of banks, savings and loans, and insurance companies. While both types of firms are indeed subject to federal and/or state regulations pertaining to rates and/or expense reimbursements, their income generating processes (and SIC codes) are dissimilar. Further disaggregation of the regulated sample is also consistent with the partitioning philosophy that we employed successfully in the earlier sections of the paper.

We discovered that the best-performing (worst-performing) expectation model for the utilities was the BR ARIMA model (RWD model) with a pooled MAPE across quarters and years in the holdout period of .459 (.578). The superior performance of an ARIMA-based model (i.e., BR) over the RWD model was expected given the highly seasonal nature of the utility firms' income generating process.[21] A short-memory model like the RWD is evidently unable to track these seasonal patterns effectively across time.

The differential predictive performance of the financial-institution firms also lends credence to further partitioning of the regulated firm sample. The best-performing (worst-performing) expectation model for the financial-institution firms was the RWD model (SRWD model) with a pooled MAPE of .350 (.402). The superiority of the RWD model is suggestive of a potentially volatile income generating process for financial-institution firms during the holdout period. This notion is underscored by the superiority of a short-memory expectation model like the RWD versus the long-memory ARIMA models. We also observe that the pooled MAPEs of the best expectation model are substantially reduced for the financial-institution firms (.350) versus the utilities (.459). These findings are supportive of the need to finer partition the regulated firm sample to further enhance the predictive power of quarterly earnings expectation models.

6. *Limitations, Conclusions and Suggestions for Future Research*

Prior to discussing our conclusions, we detail the limitations associated with the research design that we employ. First, our sampling procedures operationalized across the 1990-2002 time period are subject to a survivorship bias endemic to all time-series work. Second, while we have examined a very wide set of ARIMA-based (i.e., BR, F, GW) and simplistic (i.e., RWD, SRWD) benchmark models, other expectations models could yield different results. Third, the examination of conditioning variables (i.e., number of loss quarters, analyst following, and earnings persistence), while providing an interesting background against which the predictive findings might be assessed, is not completely exhaustive.

We provide new empirical evidence on the statistical dominance of the RWD model over ARIMA-based models in the pooled one-step-ahead quarterly earnings predictions for high-technology firms. This finding stands in marked contrast to the extant time-series literature where ARIMA-based models have systematically dominated simplistic, statistically-based, benchmark models like the RWD model. We also find that the predictive power of our tested models is highly contextual given the dominance of the GW ARIMA model over the RWD model for regulated firms as well as the joint dominance of the RWD model and the GW ARIMA model over the three remaining expectation models for the default sample of firms. Partitioning sample firms into high-technology, regulated and default samples has allowed this contextual predictive performance to surface. In fact, further partitioning of the regulated sample into utilities and financial institutions provides further predictive insights and underscores the contextual nature of predictive ability.

We observe a considerable deterioration in the predictive power of all statistically-based quarterly earnings expectation models that we tested across the 2000-2002 holdout period relative to earlier time-series work, especially for the high-technology and default sample firms. We speculate that the parsimonious nature of the RWD model (i.e., it does not require firm-specific autoregressive or moving-average parameter estimation) coupled with the volatile earnings environment experienced in the holdout period enable the RWD model to outperform more complex, ARIMA-based models for the high-technology firms.

We document a dramatic increase in the number of loss quarters experienced by sample firms across the holdout period (2000-2002) relative to the identification period (1990-1999). Specifically, high-technology (default, regulated) sample firms experienced quarterly net losses 29.6% (20.7%, 7.1%) of the time during the model identification period. The frequency of loss quarters shifted upwards to 41.4% (24.9%, 11%) for the three samples, respectively, during the holdout period. The dramatic increase in the frequency of loss quarters for the high-technology firms in particular (i.e., 40%) provides an intuitive explanation for the superior performance of the RWD model on such firms. Rather than extrapolating predictions based on autocorrelation patterns observed over the identification period like the ARIMA-based models, the RWD model invokes a short-memory assumption and bases its predictions exclusively on the most recent quarterly earnings figure.

We provide evidence that 35.5% of all sample firms that we examine had no analyst coverage reported by *I/B/E/S*. Recent passage of the Sarbanes-Oxley Act in 2002, which prohibits firm-specific management disclosures of expectational data to sell-side analysts, unless such disclosures are made known to the general public, may further reduce the number of firms covered by analysts. The Act serves to refocus analysts' efforts to emphasizing fundamental financial analysis in forming quarterly earnings expectations rather than being guided exclusively by management tips. Since fundamental financial analysis is considerably more time consuming than relying exclusively on management tips and guidance, analysts may elect to follow fewer firms. In any case, the development of accurate, statistically-based quarterly earnings expectation models is crucial for such firms.

We extend Baginski et al.'s finding of a significant decline in levels of earnings persistence for a sample of 172 NYSE firms to the high-technology, regulated, and default sample firms that we examine. Using an ARIMA-based persistence proxy, Baginski et al. report median persistence levels of 19.49 in the 1967-1978 period and 8.12 in the 1991-2001 period. We report similarly reduced persistence levels (i.e., median values of 8.17 for high-technology firms, 8.07 for default firms and 9.54 for regulated firms) for our sample firms. The reduced levels of earnings persistence reported for our sample firms provide an intuitive explanation for the increased levels of MAPEs that we report relative to previous time-series work, especially for the high-technology and default sample firms. It appears that current quarterly earnings numbers are less persistent than they were in previous decades.

Finally, our results are also salient for capital market researchers assessing the earnings-return relationship. First, statistically-based quarterly earnings expectation models must be used for more than one-third of the firms that we examine since these firms are not covered by analsyts. Second, we provide new empirical evidence consistent with the notion that researchers may conduct more powerful tests by *not* assuming that all firms have quarterly earnings series that follow the *same* time-series model. Moreover, the practice of using the SRW model as the singular proxy for the market's expectation of quarterly earnings [See Bernard and Thomas (1990) and Ball and Bartov (1996), among others] may need to be reexamined. Our results suggest that different expectation models provide superior predictive ability for different sets of firms.

REFERENCES

Albrecht, W. S., L. L. Lookabill, and J. C. McKeown. "The time-series properties of annual Earnings." *Journal of Accounting Research* 15 (1977): 226-244.

Baginski, S. P., B. Branson, K. Lorek and G. Lee Willinger. "A time-series approach to measuring the decline in quarterly earnings persistence." *Advances in Accounting* 20 (2003): 23-42.

Baldwin, W. "Coping with a crazy market." *Forbes* (Sept. 2, 2002): 114.

Ball, R. and E. Bartov. "How naive is the stock market's use of earnings information?" *Journal of Accounting and Economics* 21 (1996): 319-337.

Bathke, A. W. and K. S. Lorek. "The relationship between time-series models and the security market's expectation of quarterly earnings." *The Accounting Review* 59 (1984): 163-176.

Bathke, A. W., K. S. Lorek, and G. L. Willinger. "Differential earnings behavior and the security market assessment of variation in seasonal earnings patterns." Forthcoming, *Journal of Accounting, Auditing, and Finance* (Fall, 2004): 463-482.

Bathke, A. W., K. S. Lorek, and G. L. Willinger. "Firm-size and the predictive ability of quarterly earnings data." *The Accounting Review* 64 (1989): 49-68.

Beaver, W. H., J. W. Kennelly, and W. M. Voss. "Predictive ability as a criterion for the evaluation of accounting data." *The Accounting Review 43* (1968) 675-683.

Bernard V. L. and J. K. Thomas. "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings." *Journal of Accounting and Economics* 13 (1990): 305-340.

Bernard, V. L. "The Feltham-Ohlson framework: Implications for empiricists." *Contemporary Accounting Research* Spring 11 (1995): 733-747.

Brown, L. and M. Rozeff. "Univariate time-series models of quarterly earnings per share: a proposed model." *Journal of Accounting Research* 17 (1979) 179-189.

Brown, L. "Earnings forecasting research: Its implications for capital markets research." *International Journal of Forecasting* 9 (1993): 295-320.

Brown,L. "Discussion of post-earnings announcement drift and the dissemination of predictive information." *Contemporary Accounting Research* 16 (1999): 341-345.

Brown, L. and J. Han. "Do stock prices fully reflect the implications of current earnings for future earnings for AR1 firms?" *Journal of Accounting Research* 38 (2000): 149-164.

Das, S. C. Levine, and K. Sivaramakrishnan. "Earnings predictability and bias in analysts' earnings forecasts." *The Accounting Review* 73 (1998): 277-294.

Elgers, P. T., M. H. Lo and R. J. Pfeiffer, Jr. "Delayed security price adjustments to financial analysts' forecasts of annual earnings" *The Accounting Review* 76 (2001): 616-632.

Foster, G. "Quarterly accounting data: time-series properties and predictive-ability results." *The Accounting Review* 52 (1977): 1-21.

Francis, J. and K. Schipper. "Have financial statements lost their relevance?" *Journal of Accounting Research* 37 (1999): 319-352.

Griffin, P. A. "The time-series behavior of quarterly earnings: preliminary evidence." *Journal of Accounting Research* 15 (1977): 71-83.

Hollander. M., and D. Wolfe. *Nonparametric Statistical Methods* New York, NY: John Wiley And Sons, Inc. 1973.

Lorek, K. S. "Predicting annual net earnings with quarterly earnings time-series models." *Journal of Accounting Research* 17 (1979): 190-204.

Lorek, K. S. and A. W. Bathke. "A time-series analysis of nonseasonal quarterly earnings data." *Journal of Accounting Research* 22 (1984): 369-379.

Lorek, K. S., M. Stone, and G. L. Willinger. "The differential predictive ability of opaque and transparent firms." *Quarterly Journal of Business and Economics* 38 (1999): 3-20.

Soffer, L. C. and T. Lys. "Post-earnings announcement drift and the dissemination of Predictable information." *Contemporary Accounting Research* 16 (1999): 305-331.

Walther, B. "Investor sophistication and market earnings expectations." *Journal of Accounting Research* 35 (1997): 157-179.

Watts, R. "The time-series behavior of quarterly earnings." Manuscript, (1975) University of New Castle, NSW.

Watts, R. and R. W. Leftwich. "The time series of annual accounting earnings." *Journal of Accounting Research 15 (*1977): 253-271.

ENDNOTES

[1] We employ the customary (pdq) X (PDQ) notation where (p,P) variables represent the number of autoregressive and seasonal autoregressive parameters, (d,D) represent the levels of consecutive and seasonal differencing, and (q,Q) represent the number of moving-average or seasonal moving-average parameters.

[2] We present empirical evidence that sizable numbers of these firms are also not covered by security analysts.

[3] The transition from specialty item to commodity status has resulted in the rapid decline in the price of desktop computers and is consistent with the notion of rapid change among high-technology firms.

[4] Additionally, Bathke et al. (2004) computed cumulative average residuals measured across short, intermediate, and long windows in a manner consistent with Soffer and Lys (1999).

[5] All sample firms were also required to have the fair market value of common stock equity available at the end of the identification period, December 31, 1999.

[6] It could be argued that telecommunication firms should be classified as regulated firms rather than high-technology firms. For comparability purposes, we choose to follow the classification scheme of Francis and Schipper (1999) which assigns such firms to the high-technology sample. To the extent that such firms are misclassified, the power to differentiate between high-technology and regulated firms is reduced.

[7] The last 12 quarters of data, 2000-2002, were withheld for predictive testing.

[8] Values of the partial autocorrelation function cut off or spike at the first lag consistent with the behavior of an autoregressive process. We suppress the presentation of the partial autocorrelation functions to streamline the exposition.

[9] See Brown (1993) for an expanded discussion of the ARIMA models.

[10] All forecast errors greater than 100% were truncated to 100% to reduce the effects of explosive outliers.

[11] We also assessed the accuracy of the forecasts via the mean squared error metric with overall results that were qualitatively similar to those that we report in the paper.

[12] Across predictions in Table 5 (high-technology firms) the RWD model's predictions were truncated least frequently (30.5%) while the SRWD model's predictions were truncated most frequently (38.4%).

[13] Since the conditioning quarter for the RWD model is the previous quarter, the first quarter expectations are conditioned
entirely upon the fourth quarter results of the previous year. We observe that fourth quarter earnings are heavily influenced by the so-called "settling-up" effect where annual accruals are brought into correspondence with less rigorously derived quarterly estimates. This source of measurement error may serve to reduce the efficacy of the first quarter extrapolations of the RWD model. See Bathke and Lorek (1984), among others, for a discussion of the "settling-up" effect.

[14] Analysis of the paired comparisons for the remaining quarters does not offer any incremental insights beyond those reported for the pooled paired comparisons.

[15] It is true that both Albrecht, Lookabill and McKeown (1977) and Watts and Leftwich (1977) find a lack of superiority of ARIMA-based models versus naïve benchmark models like the RWD. However, these studies differ from the current effort in two important aspects. First, they both employed annual earnings data unlike the quarterly earnings data that we employ. Second, they did not test the premier quarterly earnings ARIMA models.

[16] Across predictions in Table 6 (regulated firms) the BR ARIMA model's predictions were truncated least frequently (19.8%) while the RWD model's predictions were truncated most frequently (24.2%).

[17] Across predictions in Table 7 (default firms) the BR ARIMA model's predictions were truncated least frequently (31.6%) while the SRWD model's predictions were truncated most frequently (36.9%).

[18] Das, Levine and Sivasamakrishnan (1998) find that analyst following and size are positively correlated (ρ = .70).

[19] See Baginski et al. (2003: pp. 40-42) for the mathematical derivation of the ARIMA-based proxy for quarterly earnings persistence.

[20] Another possible explanation for our predictive findings pertains to the presence of nonseasonal firms in the samples that we examine. For example, the high-technology sample is comprised of 53% nonseasonal firms using the identification procedures popularized by Lorek and Bathke (1984) and Brown and Han (2000) to detect such firms. These works provide evidence that seasonal ARIMA models do not predict quarterly earnings very accurately for nonseasonal firms. This may help explain why a nonseasonal ARIMA model like the RWD model outperformed the premier, seasonal ARIMA models for the high-technology firms.

[21] In fact, the cross-sectional SACF of the raw data (i.e., d=D=0) reveals substantially higher levels of autocorrelation at the seasonal lags (i.e., n=4, 8, and 12) for the utilities versus the financial institution firms.