The Prediction of Earnings Movements Using Accounting Data: Using XBRL

Amos Baranes[1]* and Rimona Palas[2]

[1]Peres Academic Center, Rehovot, Israel
[2]Accounting Department, School of Business, College of Law and Business, Ramat Gan, Israel

Abstract

The usefulness of accounting information as a basis for a profitable investment strategy is an important issue.

The objective of this study is to repeat the original Ou et al. study using the XBRL database, standardized financial reporting system required by the SEC.

The study analyzes XBRL quarterly data, from the first quarter of 2011 to the fourth quarter of 2015, using a two-step Logit regression model to determine the variables to be included in the prediction model. The prediction model was then used to arrive at the probability of the directional movement of earnings between the current quarter and the subsequent quarter.

The results of the final models' indicated a significant ability to predict subsequent earnings changes. The predictions appear to be correct on average about 72.4% of the time. However, these forecasts were not able to provide a basis for a profitable investment strategy.

Keywords: Accounting information; Earnings prediction; Investment strategy; XBRL

Introduction

The ability of earnings to indicate future earnings has been recognized as a measure of earnings quality Penman et al. and while Shivakumar et al. conclude that earnings announcements provide only a modest amount of new information to the share market, Bloomfield et al. show those investors over relies on old earnings performance when predicting future earnings performance [1-3].

These studies highlight the necessity to develop a tool to better predict future earnings and help develop various investment strategies.

Many research papers have concentrated on the importance of earnings announcements and forecasts in the determination of investment decisions. While earlier research has only been able to show relatively low informativeness of earnings later studies were able to show the incremental information content of specific components of the financial statements [4-7]. For example, Finger et al. show that earnings provide information for future earnings and cash flows, and predict sign changes in the earnings per share using forecasting models developed from various income statement and balance sheet components [8-10,14]. Shroff et al. assess the predictive ability of a "composite" model, which forecasts as a function of current earnings and current security prices, against three univariate benchmarks: a random walk model, a random walk with drift model, and a first order autoregressive/moving average (ARIMA) model [11]. The findings indicate that the composite model obtains significantly lower forecast errors relative to the benchmark models. Sletten et al. find that earnings informativeness is higher in bad-news periods than in good-new periods [12]. Alam et al. were able to show that disaggregated earnings data were better able to predict next period's earnings in the banking industry [13].

Ou et al. were the first researchers to focus on the usefulness of accounting information to predict the direction of movement of earnings relative to trend adjusted current earnings [10,14]. The study is important because it evaluates whether accounting information can consequently be used as the basis for profitable investment strategy. Given investors' reliance on earnings this could be a valuable tool for a profitable investment strategy. The authors found that financial statement analysis can provide a measure that is an indicator of future earnings which in turn is used as a successful investment strategy. However, the evidence from subsequent studies has been mixed [15-19].

One objective of this study is to repeat the original Ou et al. study over a more recent time period and provide a viable tool for investment decisions. However, the main objective is to examine the methodology using, not the original COMPUSAT database, but the XBRL database. XBRL (extensible Business Reporting Language) is a freely available and global standard for exchanging business information. XBRL allows the expression of semantic meaning commonly required in business reporting. One use of XBRL is to define and exchange financial information, such as a financial statement [10,14].

The SEC has created the XBRL US GAAP Financial Reporting Taxonomy. This taxonomy is a collection of accounting data concepts and rules that enables companies to present their financial reports electronically. The SEC's deployment was launched in 2008 in phases, and all public US GAAP companies were required to file their financial reports using the XBRL reporting technology starting from June 15, 2011.

While COMPUSTAT is a popular source of financial information for both academics and practitioners, it has been questioned how

Citation: Baranes A, Palas R (2016) The Prediction of Earnings Movements Using Accounting Data: Using XBRL. Int J Account Res 5: 143. doi:10.4172/2472-114X.1000143

reliable the data are. Prior studies have shown that COMPUSTAT data may differ from the original corporate financial data and data found in other accounting databases [20-24].

On the other hand, while there is still not enough research regarding the reliability of XBRL data, studies up to date seems positive: Boritz et al. find that when examining the quality of interactive data XBRL tagged information it is the most complete and most accurate source of company data compared with COMPUSTAT, Yahoo Finance and Google Finance; Chychyla, although did not attempt to compare COMPUSTAT and XBRL 10-K reports, found that COPUSTAT significantly alters numbers reported on the 10-K filings; and Heselmann et al. suggest that XBRL analysis is a useful tool in assessing irregularities in accounting data. The important advantages of the XBRL data, is that it allows easy and quick access, and provides up to date information to users [25-27].

Vasarhelyi et al. made suggestions for new research opportunities as a result of the evolving XBRL technology. Their suggestion was to examine whether findings from prior research that relied on private vendor databases (such as COMPUSTAT), if replicated, will still hold using XBRL database. This paper is an attempt to follow their suggestion [28].

The paper is organized as follows, Section II reviews academic literature evaluating Ou et al. and subsequent studies and examining research conducted on the validity of XBRL as a means for data. Section III outlines the method employed and the data used. Section IV presents and discusses the results for the model developed to forecast future movements in earnings, in terms of accuracy and as the basis for profitable investment strategy. The last section concludes the paper [10,14].

Academic Research

In this section will be presented a review of relevant literature on three issues evaluation of the Ou et al. [10,14] study, an evaluation of subsequent studies, and an examination of the validity of XBRL as a means for data comparison. The three issues will be examined separately.

Evaluation of the Ou et al. and Consequent Studies [10,14].

Ou & Penman (1989): Ou is considered a foundation paper in accounting research literature (cited 124 times according to PROQUEST) because they were the first to focus on the usefulness of accounting information to predict the direction of the movement of earnings relative to trend adjusted current earnings [10,14].

Using an extensive financial statement analysis (68 accounting variables) the study modeled the direction of movements (increase/decrease) in earnings per share (EPS) one year out. The sample was obtained from the 1984 COMPUSTAT annual report files and the study was conducted in several stages. In the first stage a χ^2 test was applied to a univariate LOGIT estimation and conducted for 68 accounting variables using annual report data over the period 1965-1972 and then again over the period 1973-1977. In both periods 34 (50%) of the coefficients estimated had p-values less than 0.10. In the second stage a multivariate model was used, on the variables found in the first stage, using a step-wise procedure, deleting descriptors not significant at the 0.10 level with all other descriptors included. In this stage, stage two, additional descriptors were dropped resulting in a model with 16 explanatory variables (for the 1965-1972 period) and 18 variables (for the 1973-1977 period). The results of both time periods were then used

to forecast the probability of a company's EPS lying above its trend-adjusted EPS in each of the years from 1973 to 1983. The companies were classified with a probability above 0.5 (the test was then repeated with p>0.6) as one that would realize an increase in EPS or a company with a probability below 0.5 (the test was then repeated with p<0.4) as one that would realize a decrease in EPS.

Although the two models only had 6 descriptors which appear in both time periods, many of the descriptors captured similar operating characteristics. For example, inventories, sales and deflated earnings appear in more than one descriptor. An estimation of the correlation of the prediction ability for both time periods, provided a mean for the 11 years of 0.62, the two models classified the firms consistently 78.7% of the time (for a classification of above or below 0.5).

The results of the final models' indicated a significant ability of the descriptors to jointly describe subsequent earnings changes. The χ_i^2 values from the 2X2 contingency table are highly significant and the predictions appear to be correct about 60% of the time for a probability cutoff of (0.5, 0.5) and 66% of the time for a (0.6, 0.4) cutoff.

Ou et al. continued to develop a trading strategy based on these predictions. Stocks were assigned long and short investment positions based on their probability. They purchased an equally weighted portfolio of all stocks whose estimated probability was in access of 0.6 (long position), and sold an equally weighted portfolio of all stocks whose probability was below 0.4 (short position). This strategy realized a return of 8.3% over a one year holding period, an incremental 5.7% in the second year, and 5.5% in the third year [10,14].

Replication of Ou and Penman (1989): There have been many replications of the Ou et al. study over different time periods, different countries, different industries, in comparison with analysts' predictions, and with additional methodologies, with mixed results [10,14].

Holthausen et al. reexamined Ou using a different time period (1978-1988), including Over-the-Counter firms, and using only 60 of the original 68 ratios. The study estimated four different logit models (two exchanges: NYSE/AMEX and OTC, and two time periods: 1973-1977 and 1978-1982) which retained 15 ratios (the original Ou study had 18 ratios). The correlation in the probability scores between 1973 and 1977 model, and 1978 and 1982 model for NYSE/AMEX (OTC) firms was 0.70 (0.58). The predictive ability of their models were qualitatively similar, using a cut-off of 0.5 the overall accuracy is 60.1% (compared to 60%) and using cut-offs of 0.4 and 0.6 had an overall predictive accuracy of 65.0% (compared to 67%). However, the profitability of the trading strategy realized little value added over the period of their study; that is the Ou et al. strategy worked well in 1978-1982 period (a common period for both studies) regardless of exchange with an excess return varying from 6.9% to 10.3% (8.0% to 11.4% on OTC firms). However, the strategy performed poorly in 1983-1988 period, where returns were negative (ranging from -4% to -5%) regardless of the exchange [10,14].

Bernard et al. replicated the Ou study using the same logit model to make predictions for the same years (1973-1977 and 1978-1983) and re-estimate logit model (using their approach over a previous estimation period) to produce probabilities for earnings increase for the 1984-1988 and 1989-1992 periods. The mean profitability of their investment strategies produced excess return of 4.74% in the first year and 1.24% in the second year [10-16].

Stober et al. compared the Ou et al. model prediction ability to that of analysts' forecasts of earnings. Using the same time period as Ou et

al. they found that the model accurately predicts the signs of one-year-ahead EPS 46% of the time, analysts' forecasts are correct about 54% of the time but a combined model correctly predicted the sign 78% of the time [10,14,17].

Setiono et al. examined the Ou et al. model using a UK sample over a period from 1980 to 1988 and found that a portfolio based on the forecasted probabilities realized abnormal returns [10,14,18].

Bird et al. extended the Ou et al. model by covering a later time period (the years 1983-1997) and by encompassing the UK and Australian markets in addition to the US market. Their results found 12 variables (compared to Ou et al. and using a cut-off of 0.5 showed an accuracy of 57.5%-62% (compared to 60%) and using cut-offs of 0.4 and 0.6 had an average predictive accuracy of 60.5%-66.5% (compared to 67%) depending on the country examined. Their investment strategy, based on the Ou et al. model yielded negative returns [10,14,19].

In examining specific industries Jordan et.al applied simple regression analysis to each of 25 of the variables used by Ou et al. in order to explain variations in the E/P ratios of publicly traded oil and gas firms during the years 2005-2006. Their results showed that three independent variables were significant in relation to the E/P ratio when examined individually and remain statistically significant when combined in a multiple regression model. The model was able to explain almost 62% of the variation in firms' E/P ratios [29,10,14].

Alam et al. examined the ability of disaggregated earnings to predict ROE in the banking industry. The results show that the mean adjusted R-square significantly increased from 0.576 to 0.623 with the progressive disaggregation of earnings during the years 1979-1996. The results also demonstrate that disaggregated components are better able to predict next period earnings than aggregated earnings [13].

All of these studies suggest that while there might be validity to using financial information to predicting earnings a more finely tuned and timely tool is necessary.

Validity of XBRL

Extensible Business Reporting Language (XBRL) is a business and financial reporting technology that is being implemented to enhance internal and external reporting, electronic filing, and sharing of information.

Beginning in 2009 the SEC requires that all publicly traded companies must submit financial reports in a standardized structure using XBRL to the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system under a three-year phase-in schedule. In the first phase, as of June 15, 2009 large accelerated filers that have a worldwide public common equity float above $5 billion as of the end of the second fiscal quarter of the most recently completed fiscal year, and who prepare their financial statements according to U.S. GAAP (Generally Accepted Accounting Principles), are subject to XBRL quarterly filings. In the second phase, as of June 15, 2010 all other large accelerated filers are required to comply. In the last phase, which started on June 15, 2011, all remaining filers, including smaller reporting companies, are required to file XBRL quarterly reports as an exhibit to the traditional filings (SEC 2009).

The novelty of the XBRL structured financial reports is that the reporting content is marked up with standardized elements (XBRL tags) from a publicized list of pre-defined items (XBRL taxonomy). For example, the 2013 US GAAP taxonomy contain approximately 19,000 XBRL tags that allow the user to easily extract the desired information for analysis purposes.

Literature suggests that there are several advantages of using SEC XBRL filings both for the adopting companies as well as the capital markets and research:

The XBRL structure enables unique identification and reliable extraction of accounting numbers from the financial reports – additional information comes tagged and there are no distortions due to the use of different display formats [27].

There is no deviation from the expected digit distribution due to differences between varying database providers [27].

XBRL has the potential to streamline internal accounting practices leading to cost savings and improved efficiency and effectiveness in the accounting and finance function as well as enhanced internal control leading to cost savings and improved efficiency [30].

The aim of the SEC XBRL mandate is to decrease information asymmetry by improving the information processing capability of regulatory filings (SEC 2009). XBRL-structured SEC filings are expected to improve data gathering and analyses by reducing manual data entries, and bringing all filings to a "common ground". Although early research has found inconsistencies, errors, or unnecessary extensions in the XBRL filings most recent studies found XBRL data to be not only with less errors than other forms of data, but to also provide higher quality information [31,28].

Boritz et al. compared XBRL data filed with the SEC with the data provided by three data aggregators: COMPUSTAT, Google Finance, and Yahoo Finance. Their results find a significant rate of omission of more than 50% in the financial statement items provided by the aggregators compared with the interactive SEC XBRL data. For items that are not omitted they find between 5-8% mismatches, with approximately 56% differences being greater than conventional materiality. The implications of their study is that XBRL information is the most complete and most accurate source of company data [25].

Chychyla et al. found that the values reported in COMPUSTAT significantly differ from the values reported in XBRL SEC filings. Although they do not attempt to compare COMPUSTAT and XBRL SEC filings they find that COMPUSTAT significantly alters numbers reported, specifically 17 (out of 30) variables reported by COMPUSTAT are differ from values reported by XBRL SEC filings. They are able to demonstrate how XBRL data can be utilized in an automated large-scale fashion to extract and process commonly used accounting numbers [26].

O'Farrel et al. examine the ability of XBRL data in terms of improving transparency and quality of financial accounting information as proxied by forecast accuracy. Their results found a significant improvement in analyst forecast accuracy since XBRL mandates [32,33].

Henselmann et al. state that the XBRL data may provide the SEC and investors a simple measure to flag financial reports carrying higher probability of human interaction. Their study, which was based on XBRL 10-K filings submitted to the SEC between July 2009 and March 2013, measured a firm-year-level of abnormal digit frequency and explored its association with earnings quality. Their findings are consistent with the underlying assumption that higher manipulation of earnings is reflected in higher irregularities in the frequency of digits in accounting numbers reported in the financial reports, which may indicate lower earnings quality [27].

Although XBRL data and its study is still at the early stage these studies suggest that XBRL data is a useful and accurate tool for financial statement analysis and may be used to predict the direction of future movement in earnings.

Data and Method

Data

Quarterly financial data were obtained using XBRL Analyst; an Excel plugin that allows users to access the company's XBRL tagged data from its XBRL SEC filing via the XBRL US database. The sample is of US companies included in the S&P 500 on March 31, 2016 who filed with the SEC financial statements in XBRL format. These large firms were all part of the phase 1 adoption (see validity of XBRL), which ensured that the longest time frame could be used for the analysis. The quarterly data used is from 1st quarter of 2011 to 1st quarter of 2016 (21 quarters).

Of the list of 500 companies only 400 were on the S&P 500 list for the whole time period. Of those 400 companies, 55 were financial institutions, because their disclosure and presentations standards differ from other types of companies, they were eliminated from the sample. Two additional companies had two types of stock on the list and therefore one of the shares was eliminated.

The final sample included 343 companies that were part of the S&P 500 on March 31, 2016. Table 1 lists descriptive data for these companies.

In the attempt to duplicate the Ou et al. study as closely as possible 60 variables Appendix 1 were used from the original 68 variables. The only variables not included in the study were those who were not available for a large number of companies [10,14].

Method

Similar to the Ou et al. method, a two-step approach was used to develop the model. In the first step a logistic regression univariate model was used to evaluate the significance of each explanatory variable. Only variables which were found to be associated significantly (at a 10% level) with the direction of earnings per share, above the drift, were maintained. The drift term was estimated as the mean earnings per share change over the four prior quarters to the estimated quarter [10,14].

In the second step, a stepwise logistic regression model was then used to determine the variables to be included in the final model. A two-ways (backward and forward) process of adding and removing variables to minimize the Akaike Information Criterion (AIC) measure of goodness of fit was used and implemented with the R software version 3.2.2. As discussed in Anderson et al. the AIC measure has several advantages over the Bayesian Information Criterion (BIC). The first part of the process (backwards) involved a cycle of including all the remaining variables in a single regression, and then progressively removing those that did not prove significant based on the AIC measure of goodness. The same process was repeated (forward) by starting with one variable, measuring the AIC and then adding another variable. A variable was considered insignificant if the total AIC score of the model decreased by adding another variable [34].

A different model was developed for each of the quarters for which a forecast was made, using quarterly data from all previous four years of observations – the forecast period being quarters 2, 2015 through 1, 2016. This approach deviated from the method used by Ou et al., who also developed a model but used the same model to arrive at a probability of the directional movement in EPS for all subsequent periods. The method adopted the method used by Bird et al. who developed a different model for each of the periods the forecasts were made [10,14,19].

Models

A list of the variables found significant in each model is presented in Table 2. The number of variables found significant in the different models range from 3 to 9 for each model; an average of 6 variables per model. The total number of variables found significant for all models is 12, Ou et al. found between 16-18 variables and Bird et al. found 12 to 18 variables. Three of the variables were common for all the models (Pretax Income/Sales, Gross Profit Margin and % Change in Total Revenues) and four variables were specific to only one model (% Change in Total Debt to Equity, % Change in Operating income to Total Assets, % Change in Total Assets and % Change in Sales/total Assets) [10,14,19].

Of the three variables which appear on all models only % Change in EBITDA/Sales appears in 17 of 22 models of Bird et al. [19].

All of the model's nine variables (or similar ones) were found to be

		N	Frequency	Percent
Size (Revenues)	< $1,000,000,000	343	70	20.4
	$1,000,000,000-$5,000,000,000	343	182	53.1
	$5,000,000,000-$10,000,000,000	343	42	12.2
	$10,000,000,000-$50,000,000,000	343	46	13.4
	$50,000,000,000-$100,000,000,000	343	2	0.6
	>$100,000,000,000	343	1	0.3
Industry (SIC Code)	Agriculture, Forestry & Fishing (01-09)	343	0	0
	Mining (10-14)	343	21	6.1
	Construction (15-17)	343	5	1.5
	Manufacturing (20-39)	343	129	37.6
	Transportation, Communications, Electric, Gas & Sanitary Services (40-49)	343	52	15.2
	Wholesale Trade (50-51)	343	5	1.5
	Retail Trade (52-59)	343	30	8.7
	Finance, Insurance & Real Estate (60-67)	343	64	18.7
	Services (70-89)	343	37	10.8
	Public Administration (91-99)	343	0	0

Table 1: Descriptive data for the study sample.

significant in the Ou et al. [10,14] model. In the Bird et al. models most of the model variables were found to be significant depending mainly on the country (models were created for the US, UK and Australia) [19].

The accuracy of the forecasts are judged on the basis of the percentage of companies classified as 'long' that actually experienced an increase in EPS and those classified as 'short' that actually experience a decrease in EPS. The accuracy of the models (presented in Table 2) ranges between 66% and 77%, with an average of 72.4%. These results are better than those presented by Ou et al. which averaged 67% and those of Bird et al. which ranged between 60% and 67% [10,14,19].

The Model Forecasts

The logistic models, describe in the previous section were then used to provide a forecast of the probability for each company of it EPS for the next quarter being above its current EPS. Based on these probabilities the stock can be classified. A company stock is assigned to a 'long' position (EPS are expected to increase) if the probability is greater than 0.6, and to a 'short' position (EPS are expected to decrease) if the probability is less 0.4.

The accuracy of the forecasts are judged on the basis of the percentage of companies classified as 'long' that actually experienced an increase in EPS and those classified as 'short' that actually experience a decrease in EPS. The accuracy of the models (presented in Table 1) ranges between 66% and 77%, with an average of 72.4%. These results are better than those presented by Ou et al. which averaged 67% and those of Bird et al. which ranged between 60-67%. That could be partially explained by the fact that XBRL data is more reliable then COMPUSTAT data [10,14,19].

Investment Strategy

Table 2 demonstrates that financial statement analysis may be used to predict the sign of future earnings change. However, we would like to find out if this information be used as an investment strategy which will provide better returns than a price based strategy. The investment strategy was implemented as follows:

i. For each of the four models Q2 2015-Q1 2016, stocks are assigned to investment positions 45 days after the end of the quarter for which the accounting ratios were reported Table 2. It is assumed that quarterly report information from XBRL is available at this time.

ii. Stocks are purchased (long position) if the probability is greater

than 0.6 and sold (short position) if the probability is less than or equal to 0.4. Strategy based on values of probability above and below 0.5 were also examined however not yield bettered results, similar to previous research (Ou et al. [1]; Baird et al. 2001.

iii. Stocks are held for a period of 1 quarter and mean return differences to the long and short positions are observed at the end of the period.

Two sets of investment strategies are examined. The first is based on the Ball and Brown (1968) strategy, the return for each firm is defined as the firm's observed return for the quarter.

The second investment strategy, which will be used as a benchmark, reflects the result of an investment strategy that could have been executed at the time, Perfect Foresight strategy (Ou and Penmna, 1989).

Firms are separated into long positions and short positions based on actual change in EPS in the next quarter. Long positions are taken on stocks whose actual EPS for the next quarter are above trend and short positions in all stocks whose actual EPS are below trend. Positions are taken at the same time as those for the probability model and on the same firms used by each model. This strategy attempts to examine whether earning predictions are relevant for determining firms' values and therefore may be used to determine a profitable investment strategy.

For each model the same amount of money is invested in the long and short positions for zero net investment, ignoring transaction costs. The results of the different investment strategies are presented in Table 3.

The first question is whether predictive power in forecasting the movement in a company's earnings for the next period would be sufficient to identify mispriced stock. Using the Perfect Foresight strategy provides the answer to this question. Over the four quarters investment period the short strategy yielded an average monthly return of 3.6%, however the long strategy yielded a negative return of 3.6%. This indicates very little value for the information about the directional movement of a company's earnings for the next quarter. This performance is very different than the average monthly return of 1.83% (annual return of 22%) realized by the same portfolio in the Ou et al. study, and the average monthly return of 1.18% (annual return of 14.2%) realized by the Bird et al. study [10,14,19].

The investment portfolios based on the earnings prediction models yielded a negative return for all periods, with the short strategy providing less of a loss than the long strategy. This is in line with the Holthausen et al. study that found the strategy performed poorly in the 1983-1988 period, where returns were negative (ranging from -4% to -5%) regardless of the exchange [15].

In conclusion, earnings prediction does not seem sufficient to form the basis for a profitable investment strategy for one quarter ahead. However, this does not seem to be related to the forecasting abilities of the model, even with perfect foresight investment strategy the ability to create investment value is questionable.

Conclusion

The focus of this study has been on developing models to forecast the direction of movement in EPS, using the newly mandated accounting data format of XBRL. The use of XBRL allows not only easier access to the data but also the ability to adjust the models almost immediately as current information is posted, thus providing a much more relevant tool for investors [10,14,19].

Variables	Q2/2015	Q3/2015	Q4/2015	Q1/2016
Pretax Income/Sales	-2.628	-2.828	-2.565	-2.131
Gross Profit Margin	0.271	0.286	0.158	0.173
% Change in Total Revenues	-0.770	-0.787	-0.969	-0.962
Equity/Fixed Assets			0.011	
Sales/Fixed Assets			-0.028	
% Change in Current Ratio	-0.019	-0.018		
% Change in EBITDA/Sales	-0.054		-0.028	
% Change in Equity/Fixed Assets		-0.059	-0.065	
% Change in Total Debt to Equity		-0.005		
% Change in Operating Income to Total Assets		-0.004		
% Change in Total Assets			0.025	
% Change in Sales/Total Assets			-0.014	
Accuracy %	0.773	0.656	0.734	0.732

Table 2: Results of the logistic regressions for predicting Q2 2015 through Q1 2016.

Time Period	N	Portfolio size	Return on Portfolio	N Long strategy	Return on Long Strategy	N Short strategy	Return on Short Strategy
0.4<Probability<0.6							
Q2 2015	294	75	-0.0378	47	0.0000	28	0.0013
Q3 2015	273	61	-0.3081	32	-0.0112	29	0.0002
Q4 2015	297	64	-0.1029	35	-0.0048	29	-0.0007
Q1 2016	316	82	-0.2163	58	-0.0106	24	-0.0043
Average					-0.0066		-0.0009
0.5 < Probability < 0.5							
Q2 2015	294	294	0.0082	176	0.0007	116	0.0007
Q3 2015	273	273	-0.0424	152	-0.0019	120	-0.0009
Q4 2015	297	297	-0.0962	171	-0.0039	125	-0.0014
Q1 2016	316	316	-0.0869	196	-0.0049	119	-0.0037
Average					-0.0025		-0.0013
Perfect Foresight Strategy							
Q2 2015					0.0118		-0.0173
Q3 2015					-0.0148		0.0415
Q4 2015					-0.0514		0.0515
Q1 2016					-0.0883		0.0694
Average					-0.0357		0.0363

Table 3: Returns from investment in stocks on the basis of estimated probability of an earnings change.

The findings of the study suggest that XBRL data can be used in financial statement analysis and in research as viable data source. The models developed provided a higher accuracy rate than that of previous studies.

However, when attempting to create profitable investment portfolios, based on these earnings predictions, the study was not able to replicate the Ou et al. study. Even the perfect foresight strategy was not able to produce a clearly positive return. These findings suggest, that either there is no correlation between the ability to predict future earnings and the ability to identify mispriced stock, or there is a different issue. Since previous studies were able to identify such a correlation, other explanations, of the limitations of this study, might be relevant [10,14,19].

The first limitation of the study is size and relative uniformity of the data, S&P 500 companies. Another limitation is the relatively short time period data (from 2011) of the SEC XBRL mandate. The short time period not only limits the amount of data available but may also cause other problems such as inconsistencies, errors, or unnecessary extensions in the XBRL filings [31,28]. However, given that there are indications that XBRL quality increases over time the methodology may be tested again in the future.

A significant limitation of this study is the inherent deficiencies in the current XBRL filings, where much of the data is not explicitly tagged. However, Williams et al. found that by populating missing components better prediction models can be created. Fully populating the data, with functionality built directly into the XBRL taxonomy, would not create any excess time, effort, or cost for preparers or users [35-40].

There are several possible extensions of this study among them increasing the data size, developing methods of populating missing components and implementing more advance methodologies for the ratio analysis.

References

1. Penman SH, Zhang X (2002) Accounting conservatism, the quality of earnings, and stock returns. Account Rev 77: 237-264.

2. Ball R, Shivakumar L (2008) How Much New Information Is There in Earnings? J Accounting Res 46: 975-1016.

3. Bloomfield RJ, Libby R, Nelson MW (2003) Do investors overrely on old elements of the earnings time series. Contemp Accounting Res 20: 1-31.

4. Ball R, Brown P (1968) An Empirical Evaluation of Accounting Income Numbers. J Accounting Res 159-178.

5. Beaver WH (1968) The Information Content of Annual Earnings Announcements. J Accounting Res pp: 67-92.

6. Foster G, Olsen C, Shevlin, T (1984) Earnings Releases Anomalies and the Behavior of Security Returns. Account Rev 59: 574.

7. Bernard VL, Thomas JK (1990) Evidence That Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings. J Account Econ 13: 305.

8. Finger CA (1994) The ability of earnings to predict future earnings and cash flow.,J Accounting Res 32: 210.

9. Ou JA (1990) The Information Content of Non earnings Accounting Numbers as Earnings Predictors. J Accounting Res 28: 144.

10. Ou JA, Penman SH (1989) Financial Statement Analysis and the Prediction of Stock Returns. J Account Econs 11: 295-329.

11. Shroff PK (1999) The variability of earnings and non-earnings information and earnings prediction. J Bus Financ & Account 26: 863-882.

12. Roychowdhury S, Sletten E (2012) Voluntary Disclosure Incentives and Earnings Informativeness. Account Rev 87: 1679-1708.

13. Alam P, Brown CA (2006) Disaggregated earnings and the prediction of ROE and stock prices: a case of the banking industry. Rev Account Financ 5: 443-463.

14. Ou JA, Penman SH, Larcker DF (1989) Accounting Measurement, Price-Earnings Ratio and the Information Content of Security Prices; Discussion. J Accounting Res 27: 111.

15. Holthausen RW, Larcker DF (1992) The Prediction of Stock Returns Using Financial Statement Information. J Account Econ 15: 373.

16. Bernard V, Jacob T, Wahlen J (1997) Accounting-based stock price anomalies: Separating market inefficiencies from risk. Contemp Accounting Res 14: 89-136.

17. Stober TL (1992) Summary Financial Statement Measures and Analysts' Forecasts of Earnings. J Account Econ 15: 347.

18. Setiono B, Strong N 1998 Predicting stock returns using financial statement information. Journal of Business Finance & Accounting 25: 631-657.

19. Bird R, Gerlach R, Hall AD (2001) The prediction of earnings movements using accounting data: An update and extension of ou and penman. Journal of Asset Management 2: 180-195.

20. San Miguel JG (1977) The Reliability of R&D Data in Compustat and 10-K Reports. Account Rev 52: 638.

21. Kinney MR, Swanson EP (1993) The accuracy and adequacy of tax data in COMPUSTAT. The Journal of the American Taxation Association 15: 121.

22. Tallapally P, Luehlfing MS, Motha M (2011) The partnership of EDGAR online and XBRL - should compustat care?. The Review of Business Information Systems 15 39-46.

23. Rosenberg B, Houglet M (1974) Error Rates in CRSP and Compustat Data Bases and their Implications. The Journal of Finance 29 1303.

24. Yang DC, Vasarhelyi MA, Liu C (2003) A note on the using of accounting databases. Industrial Management and Data Systems 103 204-210.

25. Boritz J E, No WG (2013) The Quality of Interactive Data: XBRL Versus COMPUSTAT, Yahoo Finance, and Google Finance.

26. Chychyla R, Kogan A (2015) Using XBRL to conduct a large-scale study of discrepancies between the accounting numbers in compustat and SEC 10-K filings. Journal of Information Systems pp:29-37.

27. Henselmann K., Ditter D, Scherr E (2015) Irregularities in Accounting Numbers and Earnings Management - A Novel Approach Based on SEC XBRL Filings. Journal of Emerging Technologies in Accounting vol 12 no 1 117.

28. Du H, Vasarhelyi MA, Zheng X. 2013 XBRL Mandate: Thousands of Filing Errors and So What?. Journal of Information Systems vol 27 no 1 61.

29. Jordan CE, Clark SJ, Donald M, 2009 USING FINANCIAL STATEMENT ANALYSIS TO EXPLAIN THE VARIATION IN FIRMS' EARNINGS-PRICE RATIOS. Academy of Accounting and Financial Studies Journal vol 13 no. 1 91-101.

30. Amrhein DG, Farewell, S, Pinsker R 2009 REA and XBRL GL: Synergies for the 21st Century Business Reporting System. International Journal of Digital Accounting Research vol 9 127-152.http://www.uhu.es/ijdar/10.4192/1577-8517-v9_5.pdf

31. Debreceny RS, Farewell SM, Piechocki M, Felden C, Gräning A, d'Eri A 2011 Flex or Break? Extensions in XBRL Disclosures to the SEC. Accounting Horizons vol 25 no 4 631-657.

32. Boritz JE, No WG, 2008 The SEC's XBRL Voluntary Filing Program on EDGAR: A Case for Quality Assurance. Current Issues in Auditing, vol 2 no. 2 A36-A50.

33. Liu C, O'Farrell G 2013 The role of accounting values in the relation between XBRL and forecast accuracy. International Journal of Accounting and Information Management vol 21 no. 4 297-313.

34. Burnham KP, Anderson DR 2004 Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research vol 33 no 2 261-304.

35. Williams KL (2015) The prediction of future earnings using financial statement information: Are XBRL company filings up to the task? PHD dissertation, Patterson School of Accountancy, The University of Mississippi, USA.

36. Penman SH 1992 Return to Fundamentals. Journal of Accounting, Auditing & Finance vol 7 no. 4, 465.

37. Securities and Exchange Commission 2009 Interactive Data to Improve Financial Reporting. Washington DC.

38. Securities and Exchange Commission 2011 Response of the Division of Corporation Finance; Response of the Office of the Chief Accountant. Division of Corporation Finance No-Action Letter.

39. Vasarhelyi MA, Chan DY, Krahel JP (2012) Consequences of XBRL standardization on financial statement data. Journal of Information Systems 26 155-167.

40. Wahlen JM, Wieland MM 2011 Can financial statement analysis beat consensus analysts' recommendations?. Review of Accounting Studies vol 16 no. 1 89-115.