Improving the Use of XBRL Data – Through Imputation and Industry Dissemination

Amos Baranes,

Peres Academic Center

8 HaNeviim St., POB 328, Rehovot, Israel 76120

Rimona Palas

Accounting Department, School of Business, College of Law and Business

#24 Ben Gurion St., Ramat Gan, Israel 5110801

XBRL Financial Reporting Taxonomy is considered one of the most significant changes in the disclosure environments in U.S. capital markets. The ability to use the data is vital to its implementation. The main purpose of this study is not to show that XBRL provides more useful information (i.e. better prediction models) than COMPUSTAT, or to examine the validity of the financial analysis tools but only to examine whether XBRL data can be used for financial statement analysis.

The study analyzes companies' XBRL filings of quarterly data from 2011 to 2016, using a two-step Logit regression model. The model is then used to arrive at the probability of the directional movement of earnings between current quarter and subsequent quarter. The results classified the companies as ones that would realize an increase or a decrease in earnings.

Although the final model indicated an ability to predict subsequent earnings changes on average about 67% of the time (similar to previous studies), it based the models on about 23% of the entire sample examined, and could classify less than 10% of the entire sample.

The inability to model most of the companies is due to incompleteness of the data in XBRL filings. In order to overcome this problem Multivariate Imputation by Chained Equations (MICE) was implemented. This increased the number of observations in the final models by 150%. The models utilized 56% of the original companies (more than double) and classified 27% of the original companies (about triple), and still increased the accuracy of prediction to 68%.

Dissemination of the model, based on industry membership, did not improve the accuracy, however it utilized more of the original companies (59%) and classified a larger number of the companies (29%).

These results suggest that XBRL data with imputation and industry dissemination can be used to enhance its ability as a financial statement analysis tool.

*Keywords***:** Accounting information, Earnings prediction, Investment strategy, XBRL, Multivariate imputation, Industry Analysis

# INTRODUCTION

The currently mandated SEC XBRL filing data provides immediate availability and easy accessibility for both academics and practitioners. For this data to be useful, it should provide quality information for decisions, specifically, investment decisions. The objective of this study is to examine whether the XBRL database can be used with models, developed in previous studies(Ou & Penman, 1989), predicting the direction of movement of earnings. The study does not attempt to examine the validity of these models, only the ability to use the data in the analysis of financial statements based on these models, and how the data can be improved.

The quality of earnings information has been measured by its ability to predict future earnings based on past performance (Penman & Zhang, 2002) and while earnings announcements were found to provide only a modest amount of new information to the share market (Ball & Shivakumar, 2008), studies show that investors over rely on past earnings performance when predicting future earnings performance (Bloomfield, Libby, & Nelson, 2003).

XBRL (eXtensible Business Reporting Language) is a freely available and global standard for exchanging business information. XBRL allows the expression of [semantic meaning](https://en.wikipedia.org/wiki/Semantics#Computer_science) commonly required in [business reporting](https://en.wikipedia.org/wiki/Business_reporting). One use of XBRL is to define and exchange financial information, such as financial statements.

The SEC has created the XBRL U.S. 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 U.S. GAAP companies were required to file their financial reports using the XBRL reporting technology starting from June 15, 2011.

XBRL has several advantages over COMPUSTAT, which has been a popular source of financial information for both academics and practitioners. Among XBRL data advantages are the fact that it is freely available while COMPUSTAT is costly. XBRL filings also have a time advantage, it takes an average of 14 weekdays from the time a company files with the SEC for that data to appear in COMPUSTAT (D’Souza, Ramesh, & Shen, 2010), while XBRL data is published concurrently with the related PDF versions, and is immediately available. In addition, the reliability of COMPUSTAT has also been questioned, prior studies have shown that COMPUSTAT data may differ from the original corporate financial (Kinney & Swanson, 1993; Miguel, 1977; Tallapally, Luehlfing, & Motha, 2011) and data found in other accounting databases (Rosenberg & Houglet, 1974; Yang, Vasarhelyi, & Liu, 2003).

The quality of the newly mandated SEC XBRL data, used in presenting past earnings performance, is a key factor for the success of its use and implementation for both academics and practitioners. Quality of the data provided by XBRL filings has been measured in several ways, among them: the number of errors in the computation of the filings (Chychyla & Kogan, 2015; R. Debreceny, Farewell, Piechocki, Felden, & Gräning, 2010; Williams, 2015); in comparison with other sources of financial data (Boritz & No, 2013); and in assessing irregularities in accounting data (Henselmann, Ditter, & Scherr, 2015).

Another suggestion for quality measurement is to examine whether findings from prior research that relied on private vendor databases (such as COMPUSTAT), if replicated, will still hold using XBRL database (Vasarhelyi, Chan, & Krahel, 2012). The current study is an attempt to follow this suggestion, and examine the ability of earnings, presented in XBRL data, to indicate future earnings, and how the information may be enhanced.

The paper is organized as follows, the second section reviews academic literature examining research conducted on the validity of XBRL as a means for data, the limitations of XBRL data and the earnings prediction model using financial statement information. The third section outlines the method employed and the data used. Section four presents and discusses the results for the models developed to forecast future movements in earnings for the entire sample. The fifth section describes the methodology used for data imputation and presents the results. The next section presents and discusses the results for the models developed to forecast future movements in earnings utilizing the completed data. The seventh section attempts to examine the results based on industry disaggregation. The last section concludes the paper.

# ACADEMIC RESEARCH

In this section will be presented a review of relevant literature on three issues: the validity and quality of XBRL as a data source, the limitations of XBRL data and the earnings prediction model (Ou & Penman, 1989).

## 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 U.S. 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 (Henselmann et al., 2015).

There is no deviation from the expected digit distribution due to differences between varying database providers(Henselmann et al., 2015).

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 (Amrhein, Farewell, & Pinsker, 2009).

(Chychyla & Kogan, 2015)demonstrated, based on the filings of five thousand companies, how XBRL data can be utilized in an automated large-scale fashion to extract and process commonly used accounting numbers.

(Liu & O’Farrell, 2013) 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.

(Henselmann et al., 2015) 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.

Other studies found that XBRL is a useful tool not only for investors but for other financial decisions, such as loan decisions regarding loan size and interest rates (Kaya & Pronobis, 2016).

There have been a couple of attempts to use XBRL data in predicting changes in earnings.(Williams, 2015) investigated whether XBRL company filings, filed in the years 2007-2009 may be used in models attempting to predict future earnings. The study examined whether 70 accounting concepts, extracted from S&P 500 companies XBRL filings, provided adequate data needed to create earnings prediction models (Abarbanell & Bushee, 1998; Ou & Penman, 1989) and what modifications would make XBRL much more useful. The findings of the study were that XBRL filings, during the investigated period, could not be used to create earnings prediction models; however, adjusting the data, by populating any missing accounting concepts, did enable earnings prediction.

(Baranes & Palas, 2017), reexamined the usefulness of XBRL S&P 500 company filings in the prediction of future earnings, in the years 2011-2015. Their results show that these filings, without any modifications, were not only useful in predicting future earnings changes, based on the (Ou & Penman, 1989) model, but provided better predictions than previous models using COMPUSTAT data.

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.

## Limitations of XBRL data

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". However, early research has found inconsistencies, errors, or unnecessary extensions in the XBRL filings.

(Boritz & No, 2008) were among the first researchers to study this issue, they examined the XBRL data filed with the XBRL Voluntary Filing Program from its inception in 2007. Their results showed that while 89.4% of the filings passed their taxonomy validation tests, only 34.2% of the filings passed the instance document validation test, and none of the companies passed the Financial Reporting Instance Standards and Financial Reporting Taxonomies Architecture validation. The authors believed that a significant number of errors in the filings were due to the use of some form of extension taxonomy.

(R. Debreceny et al., 2010) found that one quarter of filers had computational errors, nearly half of the errors were due to errors in application of negative values (inappropriate treatment of credit\debit assumption), an additional quarter of the errors were where one or more members of calculation relationships were either missing or extraneous. (R. S. Debreceny et al., 2011) found that 40% of the extensions included in the XBRL filings were unnecessary, the appropriate tags were available in the US GAAP Financial reporting taxonomy.

(Boritz & No, 2013) compared financial information reported by three data aggregators (COMPUSTAT, Google Finance, and Yahoo Finance) with those reported by XBRL data. They found that more than 50% of the information provided by XBRL data were not available from the aggregators. Of the financial facts that were available from both sources they found that 4.8% of the data did not match when comparing from XBRL to the aggregators and 8% did not match when comparing from the aggregators to XBRL, 55.7% of these mismatches were materially different.

The values reported in COMPUSTAT were discovered to significantly differ from the values reported in XBRL SEC filings(Chychyla & Kogan, 2015). Although the study did not attempt to compare COMPUSTAT and XBRL SEC filings it does find that COMPUSTAT significantly alters numbers reported, specifically 17 (out of 30) variables reported by COMPUSTAT are different from values reported by XBRL SEC filings.

(Williams, 2015) observed in a study of S&P 500 companies that 23 of 70 accounting concepts examined had proportions of less than 0.50 complete data in a sample of 296 companys filings. 67% of the variables examined were found to be incalculable or to return erroneous results. The findings suggest that untouched XBRL filings cannot be used to create earnings prediction models. When using "fully populated" company filings, filings whose missing tags have been automatically populated based on component XBRL tags, the proportion of completed data improved from 0.50 to 0.80. While the untouched XBRL filings could not be used to create earnings prediction models fully populated filing could.

(Du, Vasarhelyi, & Zheng, 2011) find that the number of errors per filing is significantly decreasing as more quarters pass and when companies file more times. While this suggests that filers learn from their experience and therefore future filings will improve, a significant number of required accounting elements for financial statement analysis, is still expected to be missing from current XBRL filings.

## The Earnings Prediction Model

(Ou & Penman, 1989) is considered a foundation paper in accounting research literature (cited 124 times according to PROQUEST) The study was 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.

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 chi-squared 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 1973-1983. 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. 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 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.

There have been many replications of the study (Ou & Penman, 1989): over different time periods (Bird, Gerlach, & Hall, 2001; Holthausen & Larcker, 1992), in comparison to analysts' forecast (Stober, 1992), and across different countries (Bird et al., 2001; Setiono & Strong, 1998) with varying results. Results also varied when the models were disseminated to different industries (Alam & Brown, 2006; Jordan, Clark, & Donald, 2009).

The current study does not attempt to examine the validity of the earnings prediction model, but only to use it as a quality measure.

# DATA AND METHOD

## XBRL

XBRL uses meta information to describe data items and link them together through various relationships. In order for the data to be compared across companies the same taxonomy must be used by all filers. Therefore, the SEC has created the XBRL U.S. GAAP Financial Reporting Taxonomy. This taxonomy defines common rules on how to present standard accounting information in XBRL filings. For companies that wish to file information that is not standard (company specific filings) may do so through extensions. Extensions are an important part of XBRL filings that provide additional reporting flexibility, however research (R. S. Debreceny et al., 2011) found that 40 percent of all extensions were unnecessary because the corresponding elements exist in the U.S. GAAP Financial Reporting Taxonomy.

Using the NASDAQ company list (<http://www.nasdaq.com/screening/company-list.aspx>) all 6,726 tickers listed on all of the three major US stock exchanges (AMEX, NASDAQ, and NYSE) were found.

The quarterly financial data was obtained using XBRL Analyst (created by FinDynamics); an Excel plugin that allows users to access the company’s XBRL tagged data from its XBRL SEC filing via the XBRL US database. Using this software not only allows for easy access and analysis of the data but also for the calculation of any missing balances. For example, the balance reported in each XBRL filing for total liabilities is not available on the original XBRL filing but is extracted and calculated on the XBRL Analyst.

## Data

Of all 6,726 tickers, only 4,380 of the companies that were traded on Q1, 2016 filed with the SEC financial statements in XBRL format. The data is from quarterly filings from 1st quarter of 2011 to 3nd quarter of 2016 (23 quarters).

Of the 4,380 tickers listed on the different stock exchanges the following tickers were removed: 365 tickers for non-common stocks; 387 tickers for companies with IPO's between 2012 and 2017; and 25 tickers for companies with more than one ticker (the same CIK).

The final sample included 3,603 companies (53.6% of all tickers listed) that were publicly traded on Q1, 2016. These findings are compatible with previous studies where the final sample included 296 companies (59.2%) (Williams, 2015), 343 companies (68.6%) (Baranes & Palas, 2016) of the total population of S&P 500 companies. Table 1 lists descriptive data for these companies.

**Table 1 - Descriptive Data for the Study Sample**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | | N | | Frequency | Percent |
| Stock Exchange | AMEX | | 3,603 | | 12 | 0.33% |
| NASDAQ | 3,603 | | 1702 | | 47.24% |
| NYSE | | 3,603 | | 1889 | 52.43% |
| Size (Revenues) | < $10,000,000 | | 3,603 | | 612 | 16.99% |
| $10,000,000- $100,000,000 | | 3,603 | | 1062 | 29.48% |
| $100,000,000-$500,000,000 | | 3,603 | | 998 | 27.70% |
| $500,000,000-$1,000,000,000 | | 3,603 | | 358 | 9.94% |
| $1,000,000,000-$10,000,000,000 | | 3,603 | | 515 | 14.29% |
| $10,000,000,000-$100,000,000,000 | | 3,603 | | 57 | 1.58% |
| >$100,000,000,000 | | 3,603 | | 1 | 0.03% |
| Industry (SIC Code) | Agriculture, Forestry and Fishing (01-09) | | 3,603 | | 12 | 0.33% |
| Mining (10-14) | | 3,603 | | 181 | 5.02% |
| Construction (15-17) | | 3,603 | | 52 | 1.44% |
| Manufacturing (20-39) | | 3,603 | | 1329 | 36.89% |
| Transportation, Communications, Electric, Gas and Sanitary Services (40-49) | | 3,603 | | 310 | 8.60% |
| Wholesale Trade (50-51) | | 3,603 | | 104 | 2.89% |
| Retail Trade (52-59) | | 3,603 | | 204 | 5.66% |
| Real Estate (60-67) | | 3,603 | | 861 | 23.90% |
| Services (70-89) | | 3,603 | | 550 | 15.27% |
| Public Administration (91-99) | | 3,603 | | 0 | 0.00% |

In the attempt to duplicate the Ou and Penman (1989) study as closely as possible 68 variables were extracted from the XBRL filing data. It should be noted that some of the variables had to be calculated from the original filing, whereas some variables were already calculated as part of the XBRL Analyst tool. This database contained 79,191 records. To calculate growth variables and drifts, additional records were eliminated, which left 58 variables and 60,498 records.

Additional records were removed in three stages. In the first stage, every company that had more than 35% of the variables missing (20 variables) was removed, this stage removed 9.44% of the records and left 54,787 records. In the second elimination stage, every variable which had more than 15% missing data points was eliminated. This left 38 explanatory variables for the entire sample.

Once these two stages were implemented a third stage, the removal of outliers (for both variables and stock returns) was implemented. Removal of outliers is important because it can drastically bias/change the fit estimates and predictions. To identify the outliers, Interquartile range (IQR) method (Barbato, Barini, Genta, & Levi, 2011) was used. Based on this method the data is arranged by value (from the lowest to the highest value) and is divided into four quartiles. The lowest quartile values (under 25%) is Q1 and the highest quartile (over 75%) is Q3, the interquartile range (IQR) is the range between Q1 and Q3, and therefore covers 50% of the data. The lower limit is computed as Q1 – 1.5xIQR, and the upper limit is Q3 + 1.5xIQR, any data value beyond these limits was recognized as an outlier and eliminated.

Once all discussed data was removed 36 variables (of the original 68 variables) remained (Appendix 1).

## Method

Similar to the Ou and Penman (1989) 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 (Appendix 1). The drift term was estimated as the mean earnings per share change over the four prior quarters to the estimated quarter (see Ou and Penman, 1989).

In the second step, a stepwise logistic regression model was 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. The AIC measure (Burnham & Anderson, 2004) 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 increased by adding another variable.

A different model was developed for each of the quarters for which a forecast was made, using quarterly data from the previous three years of observations – for example, the forecast period for Q3, 2015, is Q2, 2013 to Q2, 2015. This approach deviates from the method used by Ou and Penman (1989), who used the same model to arrive at a probability of the directional movement in EPS for all subsequent periods. The method adopted developed a different model for each of the periods the forecasts were made (Bird et al., 2001).

The logistic models, were then used to provide a forecast of the probability that the company's EPS for the next quarter will be 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 MODELS

In the first run all 36 variables were used (see appendix 1), a list of the variables found significant in each model is presented in Table 2.

**Table 2: Results of the Logistic Regressions for Predicting Q3 2015 through Q2 2016**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Q3/2015 | Q4/2015 | Q1/2016 | Q2/2016 |
| Δ Net Profit Margin | -0.5920 | -0.5030 | -0.5220 | -0.5232 |
| ROA | -37.2309 | -20.2390 | -20.3294 | -35.4367 |
| Δ Days sales to Accounts Recv. | 1.2420 | 1.4173 | 1.3564 | 1.2439 |
| Δ Quick Ratio | -0.5556 | -0.4951 | -0.6006 | -0.6457 |
| Operating Income to Total Assets | 14.1379 | 10.7761 | 10.4624 | 13.8259 |
| Δ Equity to Fixed Assets | -1.7640 | -1.5512 | -1.3140 | -1.7376 |
| Δ Capital Expenditures to Total Assets | 0.0719 |  |  | 0.0455 |
| Δ Total Revenue | -0.7836 | -0.7640 |  |  |
| Sales to Total Assets | -0.7746 |  |  | -0.2316 |
| Sales to Fixed Assets | 0.0687 |  |  | 0.0226 |
| Working Capital to Total Assets | 0.3499 |  |  | 0.2717 |
| Sales to Total Accounts Recv. |  | -0.0072 | -0.0125 |  |
| Δ Sales to Total Assets |  |  | -0.7362 | -0.8982 |
| Δ Working Capital |  |  | 0.2434 | 0.2114 |
| Δ Capital Expenditures to Total Assets | 0.1810 |  |  |  |
| Long Term Debt to Equity | -0.1107 |  |  |  |
| Equity to Fixed Assets | -0.0189 |  |  |  |
| Current Ratio | -0.0066 |  |  |  |
| ROCE |  | -7.1971 |  |  |
| Return on Operating Expenditures |  |  | -6.8534 |  |
| Δ Capital Expenditures to Total Assets |  |  |  | 0.0701 |

The number of variables found significant in the different models range from 9 to 15 (an average of 11.75) for each model, the total number of variables found significant for all models is 21. Ou and Penman (1989) found between 16-18 variables, and Bird et al. (2001) found 12 to 18 variables. Five of the variables (Δ Net Profit Margin, ROA, Δ Days sales to Accounts Recv, Δ Quick Ratio, Operating Income to Total Assets, and Δ Equity to Fixed Assets) were common for all the models, eight variables were common to two of the four models, and the other seven variables were specific to only one model.

Of the five prominent variables (variables which appear in all four models), only two (Δ Quick Ratio, and Operating Income to Total Assets) appear in the Ou and Penman (1989) model, and three (Δ Net Profit Margin, ROA, and Operating Income to Total Assets) appear in the Bird et al. (2001) models.

## The Model Forecasts

The accuracy of the forecasts is 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 3) ranges between 66% - 70%, with an average of 67.02%. These results are similar to the results presented by Ou and Penman (1989) which averaged 67% and those of Bird et al. (2001) which ranged between 60-67%.

**Table 3: Accuracy and Portfolio size**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Q3/2015** | **Q4/2015** | **Q1/2016** | **Q2/2016** |  | **Average** |
| Accuracy | 66.09% | 70.00% | 66.15% | 65.85% |  | 67.02% |
| Number of companies used in model | 836 | 826 | 853 | 816 |  | 832.75 |
| Portfolio Size | 329 | 336 | 374 | 354 |  | 348.25 |
| Percentage of Portfolio size | 39.35% | 40.68% | 43.85% | 43.38% |  | 41.81% |

However, it should be noted that a very small number of companies, 23.1% were utilized (an average of 833 companies) in determining the models, out of the entire sample (3,603 companies, see Table 1). Of the companies utilized, the models were only able to classify an average of 42%, that is less than 10% (348 companies) of the entire sample.

# DATA IMPUTATION

The main problem with the models presented is their inability to model many of the companies (only 23%) so that even if the models can classify approximately 42% of the companies utilized they create relatively small portfolios that include less than 10% of the entire sample (average of 348 out of 3,603).

One of the reasons that the models could not use more data is because the data was not available (Chychyla & Kogan, 2015; Williams, 2015). An accounting element may not be extractable from an XBRL company filing due to several reasons, among them: the preparer erroneously did not tag the accounting element, the preparer used the wrong tag for an accounting element, or the SEC’s protocol for the preparation of XBRL company filings set forth in the EDGAR Filer Manual did not permit or require a tag.

To overcome this problem, of complex incomplete data, multiple imputation is the best method to be employed (Rubin, 1996). There are several approaches for imputing multivariate data, Multivariate Imputation by Chained Equations (MICE) is considered to be a better alternative in cases where no suitable multivariate distribution can be found. MICE specifies the multivariate imputation model on a variable-by-variable basis by a set of conditional densities, one for each incomplete variable. Starting from an initial imputation, MICE draws imputations by iterating over the conditional densities.

For the purpose of this study the package of MICE in R was implemented, while the package provides five iterations for implementation, only the first one was used for the current analysis.

Table 4 presents changes from the original data (data) to that of the data with imputation (full data). The number of observations increased by about 10%, however this small change allowed for the most important change, and that is the number of companies that are were utilized by the models, which increased by an average of 144%. This means that more than twice as many companies may be examined by the models and used in the classification for prediction purposes.

**Table 4 – Changes in Data due to Imputation**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Q3/ 2015 | | | Q4/ 2015 | | | Q1/ 2016 | | | Q2/ 2016 | | |  | | Average |
|  | Data | | Full Data | Data | | Full Data | Data | | Full Data | Data | | Full Data |  | |  | |
|  |  | |  |  | |  |  | |  |  | |  |  | |  | |
| total observations | 23,403 | | 25,895 | 23,631 | | 26,160 | 23,760 | | 26,329 | 23,917 | | 26,537 |  | |  | |
| **change** |  | | **10.65%** |  | | **10.70%** |  | | **10.81%** |  | | **10.96%** |  | | **10.78%** | |
|  |  | |  |  | |  |  | |  |  | |  |  | |  | |
| observations in final model | 9,011 | | 25,895 | 12,611 | | 26,160 | 11,300 | | 26,329 | 9,779 | | 26,537 |  | |  | |
| **change** |  | | **187.37%** |  | | **107.44%** |  | | **133.00%** |  | | **171.37%** |  | | **149.79%** | |
|  |  | |  |  | |  |  | |  |  | |  |  | |  | |
| # variables in model | 15 | | 13 | 9 | | 14 | 10 | | 18 | 13 | | 18 |  | |  | |
| **change** |  | | **-13.33%** |  | | **55.56%** |  | | **80.00%** |  | | **38.46%** |  | | **40.17%** | |
|  |  | |  |  | |  |  | |  |  | |  |  | |  | |
| # Companies in model | 836 | | 2,214 | 826 | | 1,972 | 853 | | 2,016 | 816 | | 1,920 |  | |  | |
| **change** |  | | **164.83%** |  | | **138.74%** |  | | **136.34%** |  | | **135.29%** |  | | **143.80%** | |

# THE MODELS BASED ON FULL DATA

In the first run all 36 variables were used (see appendix 2), a list of the variables found significant in each model is presented in Table 5.

**Table 5: Results of Logistic Regressions for Predicting Q3 2015 through Q2 2016 Full Data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables | Q3/2015 | Q4/2015 | Q1/2016 | Q2/2016 |
| Δ Net Profit Margin | -0.4525 | -0.4453 | -0.4402 | -0.4519 |
| ROA | -15.5055 | -16.8892 | -18.5010 | -19.5109 |
| Δ Equity to Fixed Assets | -2.0725 | -1.7947 | -1.7543 | -2.4537 |
| Δ Days sales to Accounts Recv. | 0.8071 | 0.8706 | 0.8763 | 0.9137 |
| Δ Quick Ratio | -0.5683 | -0.5909 | -0.5485 | -0.5687 |
| Operating Income to Total Assets | 7.6220 | 6.4203 | 8.5289 | 10.1506 |
| Net Profit Margin | -1.6839 | -1.5142 | -1.4182 | -0.8134 |
| Return on Operating Expenditures | -4.2742 | -4.2676 | -3.8261 |  |
| Sales to Total Assets | -0.3197 |  | -0.3104 | -0.3690 |
| Pretax Income to Sales | 0.5457 | 0.7506 | 0.6266 |  |
| Quick Ratio | -0.0191 | -0.0236 | -0.0278 |  |
| Δ Capital Expenditures to Total Assets |  | -0.0509 | -0.0444 | -0.0521 |
| Δ Sales to Total Assets | -0.7386 | -0.6431 |  |  |
| Sales to Fixed Assets |  | -0.0178 | -0.0153 |  |
| Days Sales Accounts Recv. |  | 0.0002 | 0.0002 |  |
| EBITDA to Sales |  |  | -0.3173 | -0.3096 |
| Δ Total Assets |  |  | 1.5068 | 1.9125 |
| Δ Total Revenue |  |  | -0.8867 | -0.9298 |
| Δ Capital Expenditures to Total Assets | 0.0765 |  |  |  |
| Δ Production |  |  | 0.1760 |  |
| ROCE |  |  |  | -3.2999 |
| Δ Total Depreciation |  |  |  | 1.8295 |
| Working Capital to Total Assets |  |  |  | -0.1158 |
| Δ Operating Income to Total Assets |  |  |  | -0.0950 |
| Δ Pretax Income to Sales |  |  |  | 0.0966 |
| Δ Production |  |  |  | 0.1667 |

The number of variables found significant in the different models range from 13 to 18 (an average of 16) for each model, the total number of variables found significant for all models is 26, with the original data only 21 variables were found significant for all models. Seven of the variables (Δ Net Profit Margin, ROA, Δ Equity to Fixed Assets, Δ Days sales to Accounts Recv, Δ Quick Ratio, Operating Income to Total Assets, and Net Profit Margin) were common for all the models, five variables were common to three models, six variables were common to two of the four models, and the other eight variables were specific to only one model.

Of the seven variables, common to all four models, five were also common to the previous models (before imputation). The stability of the model can be measured by the fact that of the 26 variables in the model twelve variables (46%) were common to 3-4 of the models, compared to five out of 21 (24%) variables from the previous models (before imputation).

## The Model Forecasts

The accuracy of the models (presented in Table 6) ranges between 66% - 73%, with an average of 68.15% compared to the models based on the original data which averaged 67.02%.

**Table 6 – Accuracy and Portfolio Size– Full Data**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Q3/2015** | **Q4/2015** | **Q1/2016** | **Q2/2016** |  | **Average** |
| Accuracy | 66.32% | 72.96% | 66.73% | 66.57% |  | 68.15% |
| Number of companies used in model | 2,214 | 1,972 | 2,016 | 1,920 |  | 2,030.5 |
| Portfolio Size | 966 | 957 | 952 | 957 |  | 958 |
| Percentage of Portfolio size | 43.63% | 48.53% | 47.22% | 49.84% |  | 47.31% |

However, the principal issue is the significant change is in the number of companies utilized. The number of companies utilized in the models increased on average by 144% (see table 4) and now the model utilized about 56% (an average of 2,030 companies) of the entire sample (3,603 companies, see Table 1), where with the previous models only 23.1% was used (an average of 833 companies) in determining the models. The models with the full data were also able to classify an average of 47% of the companies utilized, compared to the previous models which classified only 42%. Of the entire sample of 3,603 companies, more than 26% were classified by the models with the full data, as opposed to less than 10% classified by the previous models.

The implication of these results is that not only were the models with full data able to classify more companies, they were able to this without losing the ability to accurately classify the companies as increasing or decreasing in earnings.

# DISSMEINATION INTO INDUSTRIES

Disseminating the data into industry, and only then applying the imputation method described, did not increase the accuracy of the predictive ability of the model, however, it did increase the number of companies used in the portfolio and in the final models.

The average accuracy of the eight models, based on the eight industries, was only 67.32%, which, although similar to previous models (Bird et al., 2001; Ou & Penman, 1989) is lower than the accuracy of the entire full sample which was 68.15% (presented in table 6).

However, here as well, the principal issue is the change in the number of companies utilized. The number of companies utilized in all the industry models increased from 2,030 companies to 2,144 companies, 60% of the entire sample (3,603 companies, see Table 1). The industry models with the full data were also able to classify an average of 49.1% of the companies utilized, compared to the previous models which classified only 43.7%. Of the entire sample of 3,603 companies, more than 29% were classified by the models with the full data, as opposed to 26% classified by the previous models.

**Table 7 – Accuracy and Portfolio Size– Full Data – Industry Membership**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total Sample | Mining (10-14) | Construction (15-17) | Manufacturing (20-39) | Transportation, Communications, Electric, Gas and Sanitary Services (40-49) | Wholesale Trade  (50-51) | Retail Trade (52-59) | Real Estate (60-67) | Services (70-89) |
| Accuracy | 67.32% | 69.90% | 71.21% | 68.69% | 67.59% | 61.35% | 67.10% | 65.72% | 67.00% |
| Number of companies used in model | 2,144 | 72 | 26 | 931 | 208 | 78 | 164 | 327 | 340 |
| Portfolio Size | 1,053 | 41 | 12 | 463 | 109 | 48 | 102 | 137 | 142 |
| Percentage of Portfolio size | 49.10% | 56.45% | 47.57% | 49.74% | 52.53% | 60.90% | 62.54% | 41.81% | 41.66% |

# CONCLUSIONS

The focus of this study has been to examine the use of the newly mandated accounting data format of XBRL on a large scale in previously researched earning prediction models (Bird et al., 2001; Ou & Penman, 1989). The use of XBRL allows not only easier and less costly 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, and especially small investors.

The findings of the study suggest that XBRL data can be used in a large scale financial statement analysis, for both research and investment, as viable data source. While the models developed with the original data provided a similar accuracy rate to that of previous studies, they were only able to classify a relatively small portion of the companies (less than 10%).

The reason the models could only utilize a small portion of the data, and therefore classify an even smaller portion, is that many observations were incomplete or unavailable. To overcome this problem a Multivariate Imputation by Chained Equations (MICE) was employed. The method was able not only to provide more robust models which were able to classify a much larger number of companies (more than 26% of the original companies), but to do so at a higher accuracy rate than previously.

The ability of the models to utilize and classify companies into investment portfolios increased to 29% (of the original sample) when the sample was disseminated into industries, based on SIC codes.

This study contributes to previous research by expanding the scope of XBRL filings data used to all company filings and to by enhancing the original data by multivariate imputation. The attempt of the study is not to examine the validity of the prediction models presented, but to see if XBRL data filings may be used in this type of financial statement analysis.

The main limitation of this study 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 (R. S. Debreceny et al., 2011; Du, Vasarhelyi, & Zheng, 2013). However, given that there are indications that XBRL quality increases over time (Du et al., 2013), the methodology may be tested again in the future.

There are several possible extensions of this study among them developing other methods of populating missing components and implementing more advanced methodologies for the ratio analysis. The passage of time, which will allow higher quality filings, will also enhance the use of the XBRL data.

1. BIBLIOGRAPHY

Abarbanell, J., & Bushee, B. J. (1998). Abnormal Returns to a Fundamental Analysis Jeffery S . Abarbanell University of North Carolina at Chapel Hill. *The Accounting Review*, *73*(1), 19–45.

Alam, P., & Brown, C. a. (2006). Disaggregated earnings and the prediction of ROE and stock prices: a case of the banking industry. *Review of Accounting and Finance*, *5*(4), 443–463. https://doi.org/10.1108/14757700610712480

Amrhein, D. G., Farewell, S., & Pinsker, R. (2009). REA and XBRL GL : Synergies for the 21 st Century Business Reporting System. *The International Journal of Digital Accounting Research*, *9*, 127–152. https://doi.org/10.4192/1577-8517-v9

Ball, R., & Shivakumar, L. (2008). How much new information is there in earnings? *Journal of Accounting Research*, *46*(5), 975–1016. https://doi.org/10.1111/j.1475-679X.2008.00299.x

Baranes, A., & Palas, R. (2016). The Prediction of Earnings Movements Using Accounting Data: Using XBRL. *International Journal of Accounting Research*, *4*(2), 1–7. https://doi.org/10.4172/2472-114X.1000143

Baranes, A., & Palas, R. (2017). The Prediction of Earnings Movements Using Accounting Data: Using XBRL. *International Journal of Accounting Research*. https://doi.org/10.4172/2472-114X.1000143

Barbato, G., Barini, E. M., Genta, G., & Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics*, *38*(10), 2133–2149. https://doi.org/10.1080/02664763.2010.545119

Bird, R., Gerlach, R., & Hall, A. D. (2001). “The prediction of earnings movements using accounting data: An update and extension of Ou and Penman” -- a response. *Journal of Asset Management*, *2*(2), 180–195. https://doi.org/10.1057/palgrave.jam.2240044

Bloomfield, R. J., Libby, R., & Nelson, M. W. (2003). Do Investors Overrely on Old Elements of the Earnings Time Series? *Contemporary Accounting Research*, *20*(1), 1–31. https://doi.org/10.1506/N8T8-9QR7-YUCX-91X2

Boritz, J. E., & No, W. G. (2008). SEC’s XBRL voluntary program on EDGAR: A case for quality assurance. *Current*, *2*(2), A36–A50.

Boritz, J. E., & No, W. G. (2013). The Quality of Interactive Data: XBRL versus Compustat, Yahoo Finance, and Google Finance, 0–51. https://doi.org/10.1007/s10551-015-2769-z.For

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference\rUnderstanding AIC and BIC in Model Selection. *SOCIOLOGICAL METHODS & RESEARCH*, *33*(2), 261–304.

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*, *29*(1), 37–72. https://doi.org/10.2308/isys-50922

D’Souza, J. M., Ramesh, K., & Shen, M. (2010). The interdependence between institutional ownership and information dissemination by data aggregators. *Accounting Review*. https://doi.org/10.2308/accr.2010.85.1.159

Debreceny, R., Farewell, S., Piechocki, M., Felden, C., & Gräning, A. (2010). Does it add up? Early evidence on the data quality of XBRL filings to the SEC. *Journal of Accounting and Public Policy*, *29*(3), 296–306. https://doi.org/10.1016/j.jaccpubpol.2010.04.001

Debreceny, R. S., Farewell, S. M., Piechocki, M., Felden, C., Gräning, A., & D’Eri, A. (2011). Flex or break? Extensions in XBRL disclosures to the SEC. *Accounting Horizons*, *25*(4), 631–657. https://doi.org/10.2308/acch-50068

Du, H., Vasarhelyi, M. a., & Zheng, X. (2013). XBRL Mandate: Thousands of Filing Errors and So What? *Journal of Information Systems*, *27*(1), 61–78. https://doi.org/10.2308/isys-50399

Du, H., Vasarhelyi, M. A., & Zheng, X. (2011). XBRL Mandate: Thousands of Filing Errors and So What? Retrieved from http://eycarat.faculty.ku.edu//myssi/\_pdf/3-Du et al. XBRL Mandate.pdf

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*, *12*(1), 117–151. https://doi.org/10.2308/jeta-51247

Holthausen, R. W., & Larcker, D. F. (1992). The prediction of stock returns using financial statement information. *Journal of Accounting and Economics*, *15*(2–3), 373–411. https://doi.org/10.1016/0165-4101(92)90025-W

Jordan, C. E., Clark, S. J., & Donald, M. (2009). Using financial statement analysis to explain the variation in firms’ earnings-price ratios. *Academy of Accounting and Financial Studies Journal*, *13*(1), 91–102.

Kaya, D., & Pronobis, P. (2016). The benefits of structured data across the information supply chain: Initial evidence on XBRL adoption and loan contracting of private firms. *Journal of Accounting and Public Policy*, *35*(4). https://doi.org/10.1016/j.jaccpubpol.2016.04.003

Kinney, M. R., & Swanson, E. P. (1993). The Accuracy and Adequacy of Tax Data in COMPUSTAT. *Journal of the American Taxation Association*, *15*(1), 121. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=6147470&site=ehost-live

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*, *21*(4), 297–313. https://doi.org/10.1108/LODJ-03-2012-0043

Miguel, J. G. S. (1977). The Reliability of R&D Data in COMPUSTAT and 10-K Reports. *The Accounting Review*, *52*(3), 638–641. Retrieved from http://links.jstor.org/sici?sici=0001-4826(197707)52:3%3C638:TRORDI%3E2.0.CO;2-#

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, *11*(4), 295–329. https://doi.org/10.1016/0165-4101(89)90017-7

Penman, S. H., & Zhang, X. J. (2002). Accounting conservatism, the quality of earnings, and stock returns. *Accounting Review*, *77*(2), 237–264. https://doi.org/10.2308/accr.2002.77.2.237

Rosenberg, B., & Houglet, M. (1974). ERROR RATES IN CRSP AND COMPUSTAT DATA BASES AND THEIR IMPLICATIONS. *The Journal of Finance*, *29*(4), 1303–1310. https://doi.org/10.1111/j.1540-6261.1974.tb03107.x

Rubin, D. B. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1996.10476908

SEC. (2009). *Interactive Data to Improve Financial Reporting. Release Nos. 33-9002, 34-59324; 39-2461; IC-28609.* Washington D.C.

Setiono, B., & Strong, N. (1998). Predicting stock returns using financial statement information. *Journal of Business Finance and Accounting*, *25*(5–6), 631–657. https://doi.org/10.1111/1468-5957.t01-1-00205

Stober, T. L. (1992). Summary financial statement measures and analysts’ forecasts of earnings. *Journal of Accounting and Economics*, *15*(2–3), 347–372. https://doi.org/10.1016/0165-4101(92)90024-V

Tallapally, P., Luehlfing, M. S., & Motha, M. (2011). The Partnership Of EDGAR Online And XBRL - Should Compustat Care? *The Review of Business Information Systems*, *15*, 39–46. Retrieved from http://search.proquest.com/docview/900720360?accountid=11262

Vasarhelyi, M. a., Chan, D. Y., & Krahel, J. P. (2012). Consequences of XBRL Standardization on Financial Statement Data. *Journal of Information Systems*, *26*(1), 155–167. https://doi.org/10.2308/isys-10258

Williams, K. L. (2015). *The Prediction of Future Earnings Using Financial Statement Information: Are XBRL Company Filings up to the Task?*

Yang, D. C., Vasarhelyi, M. a., & Liu, C. (2003). A note on the using of accounting databases. *Industrial Management & Data Systems*, *103*(3), 204–210. https://doi.org/10.1108/02635570310465689

1. Appendix 1 - Univariate LOGIT estimation results for all accounting descriptors selected.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | | Accounting Descriptor | Q3/ 2015 | | | | | | Q4/ 2015 | | | | | Q1/ 2016 | | | | Q2/ 2016 | | | |
|  |  | | | Coeff | Chi | p-val | Nobs | Coeff | | Chi | p-val | Nobs | Coeff | | Chi | p-val | Nobs | Coeff | Chi | p-val | Nobs |
| 1 | Current Ratio | | | -0.019 | 3.79 | 0.052 | 23,209 | -0.016 | | 2.68 | 0.102 | 23,441 | -0.017 | | 2.84 | 0.092 | 23,570 | -0.016 | 2.7 | 0.1 | 23,756 |
| 2 | Quick Ratio | | | -0.053 | 15.92 | 0 | 22,710 | -0.053 | | 16.25 | 0 | 22,886 | -0.053 | | 16.1 | 0 | 22,933 | -0.054 | 16.95 | 0 | 23,050 |
| 3 | Days Sales Accounts Recv. | | | 0 | 8.99 | 0.003 | 22,666 | 0 | | 12.79 | 0 | 22,857 | 0 | | 12.35 | 0 | 22,943 | 0 | 11.15 | 0.001 | 23,075 |
| 4 | Depreciation to PP&E | | | -0.015 | 0.75 | 0.387 | 23,680 | -0.008 | | 0.24 | 0.623 | 23,947 | -0.008 | | 0.24 | 0.625 | 24,100 | -0.007 | 0.19 | 0.664 | 24,336 |
| 5 | Return on Operating Expenditures | | | -12.671 | 814.19 | 0 | 23,254 | -12.61 | | 813.91 | 0 | 23,405 | -12.33 | | 788.99 | 0 | 23,508 | -12.212 | 783.33 | 0 | 23,629 |
| 6 | Long Term Debt to Equity | | | 0.044 | 3.01 | 0.083 | 22,230 | 0.015 | | 0.35 | 0.555 | 22,459 | 0.024 | | 0.9 | 0.344 | 22,563 | 0.034 | 1.92 | 0.166 | 22,753 |
| 7 | Equity to Fixed Assets | | | -0.004 | 2.96 | 0.085 | 24,298 | -0.004 | | 1.95 | 0.162 | 24,558 | -0.004 | | 2.8 | 0.094 | 24,689 | -0.003 | 1.24 | 0.266 | 24,900 |
| 8 | Sales to Total Assets | | | -0.457 | 28.14 | 0 | 24,670 | -0.39 | | 20.77 | 0 | 24,952 | -0.373 | | 19.2 | 0 | 25,114 | -0.393 | 21.59 | 0 | 25,369 |
| 9 | ROA | | | -25.9 | 796.03 | 0 | 23,692 | -26.045 | | 811.49 | 0 | 23,921 | -26.032 | | 819.37 | 0 | 24,083 | -26.008 | 823.88 | 0 | 24,280 |
| 10 | ROCE | | | -12.337 | 786.57 | 0 | 23,135 | -12.312 | | 791.55 | 0 | 23,278 | -12.056 | | 767.99 | 0 | 23,374 | -11.935 | 763.33 | 0 | 23,487 |
| 11 | Gross Profit Margin | | | -0.138 | 8.58 | 0.003 | 24,283 | -0.174 | | 13.85 | 0 | 24,550 | -0.198 | | 18.21 | 0 | 24,708 | -0.197 | 18.42 | 0 | 24,894 |
| 12 | EBITDA to Sales | | | -0.871 | 146.32 | 0 | 24,521 | -0.925 | | 169.7 | 0 | 24,782 | -0.985 | | 195.2 | 0 | 24,918 | -0.962 | 190.19 | 0 | 25,131 |
| 13 | Pretax Income to Sales | | | -2.29 | 517.09 | 0 | 24,797 | -2.291 | | 530.36 | 0 | 25,012 | -2.336 | | 559.49 | 0 | 25,129 | -2.32 | 563.29 | 0 | 25,284 |
| 14 | Net Profit Margin | | | -3.395 | 647.38 | 0 | 24,574 | -3.369 | | 652.88 | 0 | 24,771 | -3.396 | | 678.22 | 0 | 24,906 | -3.356 | 675.77 | 0 | 25,046 |
| 15 | Sales to Total Accounts Recv. | | | -0.024 | 3.48 | 0.062 | 22,121 | -0.027 | | 4.44 | 0.035 | 22,323 | -0.026 | | 4.18 | 0.041 | 22,443 | -0.031 | 5.74 | 0.017 | 22,624 |
| 16 | Sales to Fixed Assets | | | -0.026 | 9.02 | 0.003 | 23,475 | -0.025 | | 8.19 | 0.004 | 23,733 | -0.021 | | 5.87 | 0.015 | 23,877 | -0.017 | 3.74 | 0.053 | 24,102 |
| 17 | Working Capital to Total Assets | | | -0.107 | 3.18 | 0.074 | 25,650 | -0.082 | | 1.86 | 0.172 | 25,921 | -0.082 | | 1.89 | 0.169 | 26,028 | -0.107 | 3.25 | 0.071 | 26,241 |
| 18 | Operating Income to Total Assets | | | -14.497 | 497.08 | 0 | 24,017 | -14.53 | | 504.44 | 0 | 24,260 | -14.709 | | 516.11 | 0 | 24,388 | -14.941 | 534.33 | 0 | 24,579 |
| 19 | Δ Working Capital | | | -0.742 | 102.07 | 0 | 22,487 | -0.76 | | 108.74 | 0 | 22,743 | -0.666 | | 85.14 | 0 | 22,853 | -0.656 | 83.79 | 0 | 22,982 |
| 20 | Δ Current Ratio | | | -1.115 | 111.32 | 0 | 22,824 | -1.132 | | 115.9 | 0 | 23,048 | -1.112 | | 112.76 | 0 | 23,164 | -1.077 | 107.2 | 0 | 23,327 |
| 21 | Δ Quick Ratio | | | -1.098 | 202.35 | 0 | 23,370 | -1.147 | | 221.51 | 0 | 23,546 | -1.107 | | 209.65 | 0 | 23,603 | -1.102 | 209.33 | 0 | 23,690 |
| 22 | Δ Days sales to Accounts Recv. | | | 1.238 | 156.65 | 0 | 23,270 | 1.299 | | 172.35 | 0 | 23,512 | 1.323 | | 178.34 | 0 | 23,674 | 1.384 | 194.88 | 0 | 23,834 |
| 23 | Δ Total Revenue | | | -2.099 | 273.25 | 0 | 23,959 | -1.98 | | 246.1 | 0 | 24,202 | -2.016 | | 258.97 | 0 | 24,403 | -2.148 | 295.13 | 0 | 24,562 |
| 24 | Δ Total Depreciation | | | 0.086 | 0.07 | 0.787 | 22,840 | 0.208 | | 0.43 | 0.512 | 23,105 | 0.189 | | 0.36 | 0.551 | 23,214 | 0.211 | 0.45 | 0.503 | 23,375 |
| 25 | Δ ROCE | | | -0.587 | 763.33 | 0 | 22,736 | -0.577 | | 745.16 | 0 | 22,927 | -0.58 | | 767.52 | 0 | 23,090 | -0.57 | 745.19 | 0 | 23,215 |
| 26 | Δ Capital Expenditures to Total Assets | | | 0.051 | 4.33 | 0.037 | 23,603 | 0.034 | | 1.96 | 0.162 | 23,830 | 0.031 | | 1.71 | 0.191 | 24,000 | 0.04 | 2.72 | 0.099 | 24,096 |
| 27 | Δ Equity to Fixed Assets | | | -4.275 | 382.54 | 0 | 23,235 | -4.136 | | 363.67 | 0 | 23,446 | -4.012 | | 344.86 | 0 | 23,596 | -3.974 | 340.62 | 0 | 23,720 |
| 28 | Δ Sales to Total Assets | | | -1.982 | 252.55 | 0 | 24,034 | -1.879 | | 229.25 | 0 | 24,267 | -1.928 | | 243.48 | 0 | 24,459 | -2.064 | 280.02 | 0 | 24,611 |
| 29 | Δ Pretax Income to Sales | | | -0.58 | 543.87 | 0 | 22,568 | -0.566 | | 522.56 | 0 | 22,764 | -0.568 | | 532.45 | 0 | 22,900 | -0.556 | 513.81 | 0 | 22,983 |
| 30 | Δ Net Profit Margin | | | -0.654 | 799.14 | 0 | 22,614 | -0.629 | | 748.69 | 0 | 22,805 | -0.626 | | 754.52 | 0 | 22,958 | -0.62 | 737.85 | 0 | 23,052 |
| 31 | Δ Sales to Working Capital | | | 0.073 | 2.09 | 0.149 | 23,259 | 0.088 | | 3.03 | 0.082 | 23,527 | 0.034 | | 0.47 | 0.492 | 23,604 | 0.014 | 0.08 | 0.775 | 23,673 |
| 32 | Δ Production | | | -0.77 | 58.26 | 0 | 22,582 | -0.717 | | 51.02 | 0 | 22,785 | -0.701 | | 49.12 | 0 | 22,860 | -0.805 | 66.3 | 0 | 22,990 |
| 33 | Δ Total Assets | | | -1.951 | 34.08 | 0 | 24,069 | -1.897 | | 32.6 | 0 | 24,305 | -1.581 | | 22.92 | 0 | 24,488 | -1.791 | 29.52 | 0 | 24,656 |
| 34 | Δ Working Capital to Total Assets | | | -0.684 | 71.47 | 0 | 22,270 | -0.73 | | 83.03 | 0 | 22,528 | -0.646 | | 66.09 | 0 | 22,608 | -0.625 | 62.46 | 0 | 22,704 |
| 35 | Δ Operating Income to Total Assets | | | -0.603 | 449.85 | 0 | 22,391 | -0.596 | | 443.6 | 0 | 22,607 | -0.594 | | 446.97 | 0 | 22,753 | -0.594 | 451.36 | 0 | 22,892 |
| 36 | Δ Capital Expenditures to Total Assets | | | -0.058 | 5.86 | 0.015 | 23,428 | -0.061 | | 6.38 | 0.012 | 23,716 | -0.055 | | 5.23 | 0.022 | 23,873 | -0.057 | 5.77 | 0.016 | 24,096 |

# Appendix 2 - Univariate LOGIT estimation results for all accounting descriptors with Full Data

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | | Accounting Descriptor | Q3/ 2015 | | | | | | Q4/ 2015 | | | | | Q1/ 2016 | | | | | Q2/ 2016 | | | |
|  |  | | | Coeff | Chi | p-val | Nobs | Coeff | | Chi | p-val | Nobs | Coeff | | Chi | p-val | Nobs | Coeff | | Chi | p-val | Nobs |
| 1 | Current Ratio | | | -0.011 | 1.52 | 0.217 | 25,895 | -0.012 | | 1.95 | 0.162 | 26,160 | -0.013 | | 2.37 | 0.124 | 26,329 | -0.014 | | 2.56 | 0.11 | 26,537 |
| 2 | Quick Ratio | | | -0.038 | 12.14 | 0 | 25,895 | -0.043 | | 15.46 | 0 | 26,160 | -0.044 | | 16.61 | 0 | 26,329 | -0.04 | | 14.12 | 0 | 26,537 |
| 3 | Days Sales Accounts Recv. | | | 0 | 7.37 | 0.007 | 25,895 | 0 | | 10.92 | 0.001 | 26,160 | 0 | | 12.86 | 0 | 26,329 | 0 | | 10.37 | 0.001 | 26,537 |
| 4 | Depreciation to PP&E | | | -0.021 | 1.97 | 0.161 | 25,895 | -0.015 | | 1.09 | 0.297 | 26,160 | -0.013 | | 0.8 | 0.37 | 26,329 | -0.007 | | 0.24 | 0.624 | 26,537 |
| 5 | Return on Operating Expenditures | | | -11.806 | 948.95 | 0 | 25,895 | -11.73 | | 952.48 | 0 | 26,160 | -11.554 | | 936.42 | 0 | 26,329 | -11.433 | | 931.55 | 0 | 26,537 |
| 6 | Long Term Debt to Equity | | | 0.047 | 4.35 | 0.037 | 25,895 | 0.027 | | 1.5 | 0.221 | 26,160 | 0.037 | | 2.93 | 0.087 | 26,329 | 0.047 | | 4.62 | 0.032 | 26,537 |
| 7 | Equity to Fixed Assets | | | -0.003 | 1.6 | 0.206 | 25,895 | -0.003 | | 1.35 | 0.245 | 26,160 | -0.003 | | 2.11 | 0.146 | 26,329 | -0.003 | | 1.25 | 0.263 | 26,537 |
| 8 | Sales to Total Assets | | | -0.448 | 32.54 | 0 | 25,895 | -0.397 | | 25.81 | 0 | 26,160 | -0.384 | | 24.27 | 0 | 26,329 | -0.397 | | 26.1 | 0 | 26,537 |
| 9 | ROA | | | -24.311 | 964.04 | 0 | 25,895 | -24.498 | | 987.58 | 0 | 26,160 | -24.478 | | 991.73 | 0 | 26,329 | -24.655 | | 1010.28 | 0 | 26,537 |
| 10 | ROCE | | | -11.503 | 917.44 | 0 | 25,895 | -11.416 | | 919.21 | 0 | 26,160 | -11.319 | | 914.83 | 0 | 26,329 | -11.256 | | 919.39 | 0 | 26,537 |
| 11 | Gross Profit Margin | | | -0.151 | 11.09 | 0.001 | 25,895 | -0.165 | | 13.55 | 0 | 26,160 | -0.189 | | 17.98 | 0 | 26,329 | -0.186 | | 17.6 | 0 | 26,537 |
| 12 | EBITDA to Sales | | | -0.827 | 152.75 | 0 | 25,895 | -0.886 | | 179.25 | 0 | 26,160 | -0.957 | | 210.88 | 0 | 26,329 | -0.947 | | 209.59 | 0 | 26,537 |
| 13 | Pretax Income to Sales | | | -2.117 | 520.88 | 0 | 25,895 | -2.09 | | 522.2 | 0 | 26,160 | -2.144 | | 558.16 | 0 | 26,329 | -2.133 | | 565.28 | 0 | 26,537 |
| 14 | Net Profit Margin | | | -3.185 | 696.77 | 0 | 25,895 | -3.143 | | 699.83 | 0 | 26,160 | -3.196 | | 737.44 | 0 | 26,329 | -3.135 | | 729.03 | 0 | 26,537 |
| 15 | Sales to Total Accounts Recv. | | | -0.018 | 2.82 | 0.093 | 25,895 | -0.019 | | 3 | 0.084 | 26,160 | -0.018 | | 2.83 | 0.092 | 26,329 | -0.02 | | 3.56 | 0.059 | 26,537 |
| 16 | Sales to Fixed Assets | | | -0.026 | 12.88 | 0 | 25,895 | -0.025 | | 11.32 | 0.001 | 26,160 | -0.023 | | 10.06 | 0.002 | 26,329 | -0.02 | | 7.51 | 0.006 | 26,537 |
| 17 | Working Capital to Total Assets | | | -0.105 | 3.11 | 0.078 | 25,895 | -0.082 | | 1.92 | 0.166 | 26,160 | -0.091 | | 2.36 | 0.124 | 26,329 | -0.113 | | 3.69 | 0.055 | 26,537 |
| 18 | Operating Income to Total Assets | | | -14.095 | 596.21 | 0 | 25,895 | -14.255 | | 615.33 | 0 | 26,160 | -14.305 | | 618.45 | 0 | 26,329 | -14.463 | | 634.83 | 0 | 26,537 |
| 19 | Δ Working Capital | | | -0.658 | 116.05 | 0 | 25,895 | -0.672 | | 122.62 | 0 | 26,160 | -0.593 | | 97.11 | 0 | 26,329 | -0.582 | | 94.41 | 0 | 26,537 |
| 20 | Δ Current Ratio | | | -0.958 | 106.62 | 0 | 25,895 | -0.95 | | 105.97 | 0 | 26,160 | -0.911 | | 98.45 | 0 | 26,329 | -0.901 | | 97.62 | 0 | 26,537 |
| 21 | Δ Quick Ratio | | | -1.058 | 220.13 | 0 | 25,895 | -1.077 | | 229.67 | 0 | 26,160 | -1.038 | | 216.83 | 0 | 26,329 | -1.047 | | 223.4 | 0 | 26,537 |
| 22 | Δ Days sales to Accounts Recv. | | | 1.221 | 176.65 | 0 | 25,895 | 1.258 | | 187.44 | 0 | 26,160 | 1.293 | | 197.5 | 0 | 26,329 | 1.34 | | 212.53 | 0 | 26,537 |
| 23 | Δ Total Revenue | | | -2.045 | 327.12 | 0 | 25,895 | -1.952 | | 301.44 | 0 | 26,160 | -1.966 | | 308.12 | 0 | 26,329 | -2.109 | | 356.68 | 0 | 26,537 |
| 24 | Δ Total Depreciation | | | 0.314 | 1.19 | 0.275 | 25,895 | 0.407 | | 2.02 | 0.156 | 26,160 | 0.439 | | 2.36 | 0.125 | 26,329 | 0.598 | | 4.41 | 0.036 | 26,537 |
| 25 | Δ ROCE | | | -0.556 | 864.55 | 0 | 25,895 | -0.55 | | 857.73 | 0 | 26,160 | -0.552 | | 872.6 | 0 | 26,329 | -0.546 | | 863.01 | 0 | 26,537 |
| 26 | Δ Capital Expenditures to Total Assets | | | 0.044 | 3.76 | 0.053 | 25,895 | 0.035 | | 2.36 | 0.125 | 26,160 | 0.027 | | 1.46 | 0.226 | 26,329 | 0.035 | | 2.5 | 0.114 | 26,537 |
| 27 | Δ Equity to Fixed Assets | | | -4.076 | 408.66 | 0 | 25,895 | -3.919 | | 384.77 | 0 | 26,160 | -3.861 | | 375.54 | 0 | 26,329 | -3.814 | | 370.45 | 0 | 26,537 |
| 28 | Δ Sales to Total Assets | | | -1.883 | 276.69 | 0 | 25,895 | -1.791 | | 253.26 | 0 | 26,160 | -1.832 | | 266.86 | 0 | 26,329 | -1.967 | | 310.03 | 0 | 26,537 |
| 29 | Δ Pretax Income to Sales | | | -0.553 | 666.1 | 0 | 25,895 | -0.542 | | 647.31 | 0 | 26,160 | -0.543 | | 655.38 | 0 | 26,329 | -0.535 | | 641.91 | 0 | 26,537 |
| 30 | Δ Net Profit Margin | | | -0.641 | 1009.73 | 0 | 25,895 | -0.629 | | 986.69 | 0 | 26,160 | -0.626 | | 988.35 | 0 | 26,329 | -0.616 | | 962.86 | 0 | 26,537 |
| 31 | Δ Sales to Working Capital | | | 0.101 | 5.26 | 0.022 | 25,895 | 0.118 | | 7.29 | 0.007 | 26,160 | 0.074 | | 2.94 | 0.086 | 26,329 | 0.055 | | 1.61 | 0.205 | 26,537 |
| 32 | Δ Production | | | -0.755 | 69.91 | 0 | 25,895 | -0.684 | | 58.23 | 0 | 26,160 | -0.666 | | 55.86 | 0 | 26,329 | -0.734 | | 69.33 | 0 | 26,537 |
| 33 | Δ Total Assets | | | -1.914 | 40.08 | 0 | 25,895 | -1.876 | | 39.13 | 0 | 26,160 | -1.589 | | 28.37 | 0 | 26,329 | -1.74 | | 34.16 | 0 | 26,537 |
| 34 | Δ Working Capital to Total Assets | | | -0.633 | 96.02 | 0 | 25,895 | -0.652 | | 103.59 | 0 | 26,160 | -0.583 | | 84.12 | 0 | 26,329 | -0.569 | | 81.08 | 0 | 26,537 |
| 35 | Δ Operating Income to Total Assets | | | -0.555 | 523.83 | 0 | 25,895 | -0.544 | | 508.34 | 0 | 26,160 | -0.544 | | 516.52 | 0 | 26,329 | -0.55 | | 532.37 | 0 | 26,537 |
| 36 | Δ Capital Expenditures to Total Assets | | | -0.059 | 6.75 | 0.009 | 25,895 | -0.058 | | 6.75 | 0.009 | 26,160 | -0.052 | | 5.35 | 0.021 | 26,329 | -0.054 | | 5.92 | 0.015 | 26,537 |