2015

The Prediction Of Future Earnings Using Financial Statement Information: Are Xbrl Company Filings Up To The Task?

Kelly Lawrence Williams
University of Mississippi

THE PREDICTION OF FUTURE EARNINGS USING FINANCIAL STATEMENT INFORMATION: ARE XBRL COMPANY FILINGS UP TO THE TASK?

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Patterson School of Accountancy
The University of Mississippi

Kelly L. Williams

August 2015

ABSTRACT

Financial statement data for large companies became available to the public in XBRL format starting in 2009 in the United States. Proponents of XBRL, along with the SEC, argue that XBRL filings offer several advantages over data provided by data aggregators, such as lower cost, faster availability, and broader coverage. The purpose of this study was to contribute to the common body of knowledge by investigating whether current XBRL company filings are useful in the prediction of future earnings and to attempt to interactively obtain the balances of 70 accounting concepts needed to create an earnings prediction model from a sample of XBRL filings. Current XBRL filings do not allow for interactive extraction of required accounting elements because too many accounting elements are missing from the XBRL filings. Accordingly, an additional data set was created by manually populating missing accounting concepts within the XBRL filings if sufficient component accounting concepts existed within the same XBRL filing (e.g., if current liabilities and long-term liabilities were tagged in the XBRL filing, total liabilities could be calculated). This process mimicked what could be performed by added functionality built directly into the XBRL taxonomy. This functionality would not create any excess time, effort, or cost for preparers or users. This fully populated XBRL data set allows the user to create earnings prediction models interactively, whereas the current XBRL data set does not. This indicates that current XBRL company filings are likely to be limited in their usefulness in other areas as well, while a more fully populated set of XBRL company filings that includes additional data has the potential to improve the usefulness of XBRL data.

DEDICATION

This dissertation is dedicated to my wonderful husband, Jimmy Williams, who has supported me throughout this entire process, been patient when I am sure that was not easy, and made me laugh when I got discouraged. Thank you, Jimmy.

LIST OF ABBREVIATIONS

AICPA	American Institute of Certified Public Accountants
CEASA	Center for Excellence in Accounting and Security Analysis
GAAP	Generally Accepted Accounting Principles
HTML	HyperText Markup Language
PDF	Portable Document Format
SEC	Securities and Exchange Commission
USFR-AR	US Financial Reporting - Accountants Report
XBRL	eXtensible Business Reporting Language

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Dr. Rick Elam. Dr. Elam is not only my dissertation chair but also my mentor. Since very early in the program, he has been amazingly generous with his time and resources and has taught me so many things about research, career choices, and my own development. These are things that will serve me well for the rest of my life. I would also like to thank Dr. Mitchell Wenger. Dr. Wenger has helped me so many times with research ideas and turning those ideas into a completed study. His creativity is inspiring. Drs. Elam and Wenger have spent countless hours teaching me about research and the publishing process. I am also grateful to Dr. Vicki Dickinson for motivating me with so many great research ideas. Dr. Dickinson played an instrumental role in shaping the topic studied in this dissertation. I would also like to thank Dr. John Bentley for his invaluable assistance with the statistical methods used in this dissertation as well as for being one of the nicest people anyone could ever work with. These four exceptional people make up my dissertation committee, and without each and every one of them, this dissertation would never have come to be.

I am so grateful to my fellow accountancy Ph.D. students (those who are no longer students and those who are). The comradery that exists among all of us will last a lifetime. Never have I had friends who would go to such great lengths to help in any way they could. I am especially grateful to Dr. Dereck Barr-Pulliam, my war buddy. Because of our pact and all of his help, support, and encouragement, I have successfully come through this process.

I would like to thank all the faculty members at the Patterson School of Accountancy. Many of them taught me, either in or outside the classroom, about research and areas of accounting I never even knew about. So many of them helped me through challenges or gave me advice.

I am especially grateful to my father, Dr. Howard Lawrence, who himself is a Ph.D. graduate of the Ole Miss accounting program. He encouraged me for years to pursue this journey and always believed in me. He has helped me in so many ways to achieve what he always knew I could do. I would like to thank my mother for her unwavering belief in me. She has been a huge source of support and has taken up slack in other areas in order to allow me to study.

I would definitely like to thank my husband, Jimmy Williams, and my son, Bryce Williams, for their part in this process. They have been amazing in their support, patience, and encouragement. I am indebted to them for their sacrifice.

Last, but certainly not least, I would like to thank God, whom I have called on countless times and who has always come through.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The Securities and Exchange Commission (SEC) requires all public companies to report financial statements using eXtensible Business Reporting Language (XBRL). According to the SEC, there are a number of reasons why XBRL company filing data (which it refers to as “interactive data”) is well suited for gathering large amounts of company financial data, making tasks that utilize company data easier and more efficient. Specifically, the SEC states:

> Interactive data can provide investors quicker access to the information they want in a form that’s easily used and can help companies prepare the information more quickly and more accurately.
>
> Using today’s disclosure documents, investors who seek specific information directly from the source must often manually search lengthy corporate annual reports or lengthy mutual fund documents. Even if these documents are online, they are often in a plain-text format with limited search capability. The need to search for and extract particular information in such documents can be time consuming.
>
> Interactive data allows investors and others to pinpoint facts and figures within today’s often lengthy disclosure documents. Using interactive data, an investor can immediately pull out specific information and compare it to information from other companies, performance in past years, and industry averages.... Meanwhile, for the financial professionals and financial publishers, analyzing companies could become cheaper and easier. Interactive data also may help filers improve their reporting processes. (SEC, 2010)

XBRL company filing data is a new source of company financial data and has many potential advantages for a broad constituency, including analysts, investors, researchers, data aggregators, and others. Most users are forced to rely on data aggregators in order to collect company financial data. A large amount of accounting research relies on the use of Compustat, a private vendor database. Compustat is used to gather company financial data and is published by Standard & Poor's. According to D'Souza, Ramesh, and Shen (2010), Compustat is "an important information intermediary that acts as a key supplier in the market for corporate accounting information" (p. 160). An advantage of XBRL is that access to XBRL US GAAP 10-K filings to the SEC (hereafter referred to as "XBRL company filings") is free, while subscriptions to proprietary databases, such as Compustat, are very expensive. Still another potential advantage of XBRL is that manual collection of large amounts of accounting data from paper, Portable Document Format (PDF), or even HyperText Markup Language (HTML) 10-Ks requires considerable time and effort. Ideally, a user could interactively extract just the needed data from an XBRL company filing to meet his or her specific needs. Vasarhelyi, Chan, and Krahel (2012) state that "the very purpose of XBRL [is the] automation of data parsing" (p. 157). Using accounting data from XBRL company filings also has a significant time advantage. XBRL 10-K and 10-Q filings are published concurrently with the related HTML filing versions. However, it takes an average of 14 weekdays from the time a company files with the SEC for that accounting data to appear in Compustat's Research Insight database (D'Souza et al., 2010). Finally, XBRL company filings include more companies than are represented in the Compustat database. Compustat and Thomson ONE Banker together only cover about 70% of the companies that file with the SEC. Many of the smaller companies' data are not included in the

Compustat database; therefore, XBRL company filing data cover a much broader group of companies (SEC, 2009).

A variety of things were investigated in this study, each with a specific goal in mind. The first focus of the study was to determine whether current XBRL company filings achieve the SEC's promise of interactive data access. Because achievement of the goals of the current XBRL company filings were not fully met in this study, a more complete set of XBRL company filings was created. This data set is referred to as "fully populated XBRL company filings," where missing accounting elements were manually calculated when there was sufficient additional information available in the XBRL company filings. The specific process for the creation of the fully populated XBRL company filing data set is described in more detail below. The purpose of this study was to investigate whether fully populated XBRL company filings would improve the usefulness of XBRL company filing data, i.e., if fully populated XBRL company filings could provide increased interactive data access compared to the current XBRL company filings. The investigation was conducted in order to determine if the usefulness of XBRL company filings could be improved with changes to the US GAAP Financial Reporting Taxonomy.[1] Functionality could then be built into the XBRL taxonomy that would automatically populate XBRL company filings without any additional effort on the preparer's or user's part.

In the last part of the study, fully populated XBRL company filing data were compared to Compustat data in regard to each data set's accuracy in predicting future earnings. Most researchers in the area of earnings prediction have used Compustat data. Like most proprietary data aggregators, Standard & Poor's employs standardization techniques for the company data

[1]A taxonomy is an "electronic dictionary of business reporting elements used to report business data" (XBRL US, 2008, p. 111). "The XBRL US GAAP [Financial Reporting] Taxonomy describes and classifies elements representing US GAAP reporting concepts" (XBRL US, 2008, p. 13).

that it collects, which can make the reported balances significantly different from company-reported data.

Vasarhelyi et al. (2012) made suggestions for new research opportunities as a result of the evolving XBRL technology. One such suggestion was:

> If replicated, will findings from prior research that relied on private vendor databases using pre-XBRL tagged filings still hold? Private databases often have proprietary aggregation and labeling methods, so results may change due to the use of more granular and/or differently standardized data. (p. 163)

Following the Vasarhelyi et al. (2012) suggestion, an attempt was made to create two earnings prediction models that originated in the Ou and Penman (1989a, 1989b), Lev and Thiagarajan (1993), and Abarbanell and Bushee (1997, 1998) studies. In this study, the models were modified slightly because some of the variables required data going back to 2007, and XBRL company filings are not available for the years prior to 2009. In the first step, an effort was made to use current XBRL company filings to interactively capture the necessary accounting concepts needed to create the earnings prediction models. This proved not to be possible using current XBRL filings because far too many accounting concepts were not tagged within those XBRL company filings. Next, an effort was made to use the fully populated XBRL company filings to interactively capture the necessary accounting concepts needed to create the earnings prediction models. Using this set of data made it possible to create the two earnings prediction models. As an interesting test of the fully populated XBRL company filing data, the two models created using fully populated XBRL company filing data were compared with the same two earnings prediction models created using Compustat data in order to compare the ability of the accounting data from the different sources to predict future earnings. The distinction between these differing types of company filing data is discussed in greater detail below.

1. **Current XBRL company filing data**: This data was interactively extracted from the XBRL 10-K company filings submitted to the SEC. “Interactively” means that only the required information was extracted from the XBRL company filings without an attempt to manually calculate any missing balances. For example, the balance reported in each XBRL company filing for total liabilities was extracted, if it existed, but no additional steps, such as manual calculations, were performed to attain this balance. Some of the reasons why an accounting element may not have been interactively extractable from an XBRL company filing are: the preparer erroneously did not tag[2] 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. According to SEC rules, XBRL company filings should mimic exactly the related paper/PDF/HTML filings. For instance, a company might choose to display its liabilities section on its paper/PDF/HTML 10-K as follows:

Current liabilities:	
Accounts payable	$ 100,000
Other current liabilities	25,000
Total current liabilities	125,000
Long-term debt	75,000

[2]A tag is “markup information that describes a unit of data in an instance document and encloses it in angle brackets (“<>” and “</>”). All facts in an instance document are enclosed by tags that identify the element of the fact” (XBRL US, 2008, p. 111). An example would be: (<Cash> </Cash>).

Note that in this example, the company did not explicitly include a subtotal for long-term liabilities, nor did it include an amount for total liabilities, which is purely a formatting preference on the company's part. Although no reader of a paper/PDF/HTML 10-K would interpret this as the company having no total liabilities, a computer without the proper software would not be able to make this determination. The problem is that the SEC requires that XBRL company filing preparers not tag any amount for total liabilities because the XBRL company filings are required to mimic the paper filings. Accordingly, if a user attempts to extract the balance for total liabilities from this company's XBRL filing, as the SEC claims users of XBRL company filings can do, there would be no balance returned. XBRL company filings, unlike paper/PDF/HTML statements, are designed for computers to read, not humans. A computer is not able to extract just the balance for total liabilities from this company's XBRL filing without sophisticated programming, although a computer running a well-written extraction program would be able to appropriately determine the existence and values of these accounts. This particular example is very simple, whereas the relationship hierarchy among the thousands of XBRL tags is very complicated, not to mention that the tags and relationships are modified on a continuous basis.

2. **Fully populated XBRL company filing data**: Because the current XBRL company filings did not allow for interactive extraction of required accounting concepts, fully populated XBRL company filing data were created in order to determine if the usefulness of XBRL company filings could be improved with changes to the US GAAP Financial Reporting Taxonomy. The process for creating these fully populated XBRL company filings mimics the functionality that could be built into the XBRL taxonomy, which

would automatically populate XBRL company filings without any additional effort on the preparer's or user's part. The fully populated XBRL company filings allow for the interactive extraction of required accounting elements because far more accounting concepts are available than there are in the current XBRL company filings.

Data for the fully populated XBRL company filing data were extracted from the XBRL 10-K company filings submitted to the SEC. However, unlike the current XBRL company filing data**,** if a required piece of data was not reported within that XBRL company filing, the balance was manually calculated, provided there was sufficient other information included in the XBRL company filing. This manual calculation process followed rigid rules designed to imitate a potential automated process. The resulting data set from filling in these additional tags is referred to herein as the "fully populated XBRL company filing data." Using the example above, the filing would be fully populated after filling in the total liabilities with a balance of $200,000. Note that the balances were not changed, only that missing tags were filled in if there was sufficient information tagged in the XBRL company filing to do so.

3. **Compustat data:** Compustat data was extracted from the Compustat Annual Fundamentals database.

After applying the above-described procedures, the ability of the fully populated XBRL company filing data and the Compustat data to predict future earnings changes were compared. Because a certain number of years of information are required for the models, only the companies required to begin filing with XBRL in 2009 (the first year of required filing in XBRL format with the SEC) were included in the sample. These companies have a market capitalization of more than $5 billion. Medium-sized and small companies did not begin filing in XBRL format until 2010 and 2011, respectively, and sufficient data is not yet available to examine these companies' reports.

The SEC describes XBRL data as allowing users to interactively extract just the information needed from within companies' lengthy financial statements. Based on this description, XBRL company filings have great potential for research and decision-making purposes when financial data is required. Because data for specific accounting elements are needed for earnings prediction models, XBRL company filings should be particularly well suited for this task. However, the earnings prediction models could not be created interactively with current XBRL company filings because too many required accounting elements were missing. Nevertheless, after creating a fully populated set of XBRL company filings, it appeared that the ability of XBRL company filings to be used to create earnings prediction models was enhanced. In addition, fully populated XBRL company filings predicted with a higher level of accuracy than Compustat for one of the earnings prediction models. There was no significant difference in the prediction accuracy between fully populated XBRL company filings and Compustat for the other future earnings prediction model.

The most significant finding in this study was that current XBRL company filings cannot be used to create earnings prediction models; however, fully populated XBRL company filings can. All XBRL company filings could be fully populated with functionality built directly into the XBRL taxonomy, and this would not create any excess cost for preparers or users or require any additional time or effort. Because current XBRL company filings cannot be used to create earnings prediction models but fully populated XBRL company filings can, there is a strong possibility that fully populated XBRL company filings would be more useful in other areas, such as bankruptcy prediction and stock price predictions.

II. BACKGROUND

XBRL

XBRL is a computer markup language designed to standardize business and financial reporting and aid in the preparation, analysis, and communication of business information. XBRL allows information to be exchanged between different computer systems, both internally and externally. XBRL does not require that an accountant know computer coding because the logic is built into the taxonomies and software programs; once configured, the software can translate the business information so that the data a user requires can be examined (Zarowin & Harding, 2000). This eliminates the need for rekeying data, which greatly reduces errors and labor hours. XBRL also increases accuracy because it can both verify and calculate.

XBRL is a vocabulary of eXtensible Markup Language (XML), which uses tags around pieces of information, giving them meaning (Debreceny & Farewell, 2010; Mahoney & White, 2007). The XBRL US GAAP Taxonomy Preparers Guide describes a tag as “markup information that describes a unit of data in an instance document and encloses it in angle brackets, i.e. (“<Cash>” and “</Cash>”). All facts in an instance document are enclosed by tags that identify the element of the fact”[3] (XBRL US, 2008, p. 111). These XBRL tags provide a standard format that can be used for the distribution of metadata associated with business reporting information (Piechocki, Felden, Gräning, & Debreceny, 2009). Hoffman, Kurt, and Koreto (1999) state that the term “markup” refers to the codes (tags) that identify pieces of information. XML is similar to HTML that websites use; however, HTML focuses on “how to display data,” while XML focuses on “what is the data.” HTML works because there is a standard, and that is why web browsers have the ability to display the content of millions of different sites around the world. The same ability can be applied to accounting data. With XML, any tag can be created to identify information. For instance,

<AccountsReceivable> 100 </ AccountsReceivable>

and

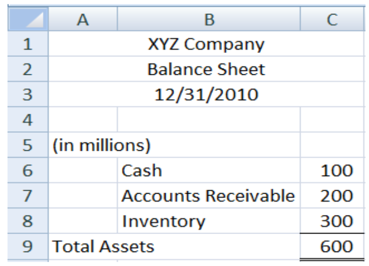
<AcctsRec> 100 </ AcctsRec>

are both valid.

XBRL has a defined set of tags for financial and business data so that every company theoretically reports the same piece of data with the same tag. This is important because computers are highly literal and do not interpret *AccountsReceivable* and *AcctsRec* as representing equivalent accounting concepts. Also, different computer applications do not effectively communicate well with each other. For instance, if the sample balance sheet of XYZ Company in Figure 1 were to be read by a human, it would probably be obvious that XYZ Company had $100,000,000 USD in cash at December 31, 2010. To a computer software application, this report may well have no meaning or an incorrect meaning. If another software

[3]An instance document is an “XML file that contains business reporting information and represents a collection of financial facts and report-specific information using tags from one or more XBRL taxonomies.” (XBRL US, 2008, p. 109).

application were to read the value of 100 in cell C6, it would not know what this value represented.



	A	B	C
1		XYZ Company	
2		Balance Sheet	
3		12/31/2010	
4			
5	(in millions)		
6		Cash	100
7		Accounts Receivable	200
8		Inventory	300
9	Total Assets		600

Figure 1. Sample balance sheet of XYZ Company.

XBRL instance documents can analyze and compare data much faster and more efficiently than could be done using HTML, PDF, or word processing documents (Srivastava & Liu, 2012). McNamar (2003) observed in 2003 that if the SEC had used XBRL for filings when Enron was committing fraud, Enron's financials would have been compared to industry standards and it would have been apparent how much they deviated from industry norms. Conceptually, XBRL can revolutionize the way companies disseminate and consume the voluminous amounts of data created in the financial world.

XBRL does have a defined set of tags; however, recall that the "X" in "XBRL" stands for "eXtensible." This is because an extension taxonomy can be produced, allowing for the creation of additional tags to be used in XBRL instance documents. Extensions in XBRL filings should only be used if the required tag does not exist in the XBRL taxonomy. The *definition* of a tag should be used to determine the "appropriate" tag, not the *label* of the tag. For example, if a company reports "Plant Assets, Net" on its balance sheet, the tag "PropertyPlantAndEquipmentNet" would be the appropriate tag from the 2013 US GAAP

Financial Reporting Taxonomy. Allowing companies to use extensions in their XBRL filings facilitates higher reporting flexibility but deteriorates comparability across companies.

A non-profit organization, XBRL International, manages the XBRL standard. XBRL International is made up of jurisdictions (e.g., XBRL Australia, XBRL Denmark, XBRL US, and XBRL Europe) and direct members (e.g., the American Institute of Certified Public Accountants, the Federal Deposit Insurance Corporation, and Fujitsu Ltd.) (XBRL International, 2013b). The Financial Accounting Standards Board and the International Accounting Standards Board are responsible for the development and maintenance of the US GAAP Financial Reporting Taxonomy and the International Financial Reporting Standards Taxonomy, respectively.

History of XBRL. Numerous authors have detailed the history of XBRL. According to Kernan (2009), XBRL was first envisioned by Charles Hoffman in 1998. Hoffman identified the problem of various computer applications' inability to effectively communicate with one another. Hoffman stated, "It was like having an e-mail system that could only create a message, not send or receive it. The financial world had become trapped in an electronic Tower of Babel, endlessly copying and pasting information from one system into another" (Kernan, 2009, p. 4). Hoffman believed that if there was some way to enter information into a computer only once and that information could be used throughout the business and never reentered, it would put an end to the inefficient way things had previously been done. The solution came to him in April 1998 while flipping through a book about XML. The book explained how XML was solving the problems of data sharing in other industries. Hoffman realized that the same type of markup language could be used to share financial statements without reentering or copying and pasting data multiple times (Kernan, 2009).

In July 1998, Hoffman informed the American Institute of Certified Public Accountants' (AICPA) High Tech Task Force about the potential of using XML in financial reporting, and based on this preliminary information, the High Tech Task Force proposed the creation of a prototype set of financial statements using XML. With the help of Mark Jewett (Erutech) and Jeffrey Ricker (XML Solutions), Hoffman was able to complete the prototype by December 1998 (XBRL International, 2011).

The XBRL International website states that perhaps the single biggest accomplishment in this early stage was the creation of a uniform language for financial statements. Eric Cohen was recruited to assist with this. Cohen agreed that there needed to be a common set of tags for financial statements, but he wanted to go beyond and create tags for all accounting information, such as payroll. (Cohen later developed XBRL Global Ledger.) In May 1999, a meeting was held to present eXtensible Financial Reporting Markup Language (XFRML), which was the original name for XBRL. It was Bob Elliot who announced that focusing on just financial reporting was not enough and that all business reporting should be covered (Kernan, 2009). In July 1999, the AICPA agreed to fund the XBRL effort. Twelve companies, including the (then) Big Five accounting firms, formed a steering committee to begin implementing the business plan and creating an XML financial reporting specification (XBRL International, 2011).

The benefits of XBRL became evident to companies and regulatory agencies, including the SEC, almost immediately, and less than three years after the AICPA funded the effort, the Australian Prudential Regulation Authority became the first regulator in the world to adopt and use XBRL. It used the language to facilitate data collection from 11,000 super funds, insurers, and banks required to report to it on a regular basis. The following month, Microsoft Corporation became the first technology company to report financials using XBRL (XBRL International, 2011).

Federal Deposit Insurance Corporation. The Federal Deposit Insurance Corporation (FDIC) of the United States began requiring XBRL in 2005 for its Call Reports.[4] This was the first large-scale implementation of XBRL in the U.S. and, at that time, the largest use of XBRL internationally (Federal Financial Institutions Examination Council, 2006). The FDIC's use of XBRL has proved to be extremely successful.

The FDIC worked closely with Call Report preparation software vendors even very early in the process of modernizing the Call Report collection process. As a result, the FDIC pre-approved software vendors that provided Call Report preparation software with the XBRL taxonomy built in. The software, not the preparer of the Call Report, tags the Call Report with the XBRL tags (Harris & Morsfield, 2012).

Immediate benefits were recognized once the FDIC started using XBRL—for example, cleaner data, more accurate data, faster data inflow, increased productivity, faster data access, and seamless throughput (Federal Financial Institutions Examination Council, 2006). Martin Gruenberg, FDIC vice chairman during the XBRL implementation (now FDIC chairman), stated:

> Key to these successes was the minimal disruption to banks. Bankers did not know they were using XBRL in the new system—it was transparent to them. This was due to our work with the software vendors that provided the bank filing software.

[4]According to the Federal Financial Institutions Examination Council, "the Call Report is a quarterly data series of a financial institution's condition and income that is used for multiple purposes, including assessing the financial health and risk profile of the institution" (p. 8).

> In short, XBRL has helped us achieve significant efficiencies and reduce operating costs. The standard has enabled us to improve the immediate quality of the data we receive. Our data quality standards are conveyed efficiently, requiring significantly less intervention from agency staff to reconcile and validate. The data are more timely and accurate, allowing us to make better-informed decisions every day.
>
> Interactive data and a common XBRL language have enabled us to dramatically improve the quality of communication between the regulatory agencies and reporting banks. Receiving data faster and more accurately strengthens our supervisory function and also improves the public transparency of the condition of the banking system. We've made an important investment in building this new system and it appears that the benefits have been well worth the cost. (Gruenberg, 2006)

Securities and Exchange Commission. Even before the development of XBRL, the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database tagged information; however, those tags did not provide enough information because they were not thoroughly detailed. For example, EDGAR documents might tag an entire table rather than the specific information within the table. But if the reader of the financials did not know what the table was describing, the year it was reporting, or the specific information it contained, it was not of significant use except for determining that the table does, in fact, exist (XBRL International, 2011).

The SEC mandated the use of XBRL for financial disclosures beginning in 2009 for companies with a market capitalization of more than $5 billion and beginning in 2011 for all public companies (Bartley, Chen, & Taylor, 2010a; Capozzoli & Farewell, 2010; Sledgianowski, Fonfeder, & Slavin, 2010). The rules apply to public companies and foreign private issuers that prepare their financial statements in accordance with U.S. generally accepted accounting principles (US GAAP) and foreign private issuers that prepare their financial statements using International Financial Reporting Standards as issued by the International Accounting Standards Board. Companies must provide their financial statements in XBRL not only to the commission but also on their corporate websites. If a company does not provide XBRL filings to the SEC or does not post the XBRL filings on its corporate website, the company will not be deemed to have sufficient current public information under Rule 144 of the Securities Act, which provides a resale exemption safe harbor from the registration requirements for the resale of restricted and control securities. In addition, the company will not be considered current with its Exchange Act reports and therefore will not be able to use the short Form S-3, F-3, or S-8 or to choose to provide information at a level prescribed by Form S-3 or F-3 on Form S-4 or F-4 (SEC, 2009). At this time, XBRL company filings submitted to the SEC do not have to be audited.

Unlike the preparers of Call Reports for the FDIC, the preparers of financial statements in XBRL that are to be submitted to the SEC must be proficient in the use and preparation of XBRL. There is no software that can create the XBRL company filings “behind the scenes.” XBRL US makes available a preparers guide (XBRL US GAAP Taxonomy Preparers Guide) that explains how the US GAAP Financial Reporting Taxonomy works and how to create an XBRL instance document using the US GAAP Financial Reporting Taxonomy. The US GAAP Financial Reporting Taxonomy currently contains approximately 17,000 tags. The SEC includes instructions and guidelines for preparers of XBRL company filings as part of the EDGAR Filer Manual. The EDGAR Filer Manual and the XBRL US GAAP Taxonomy Preparers Guide are very clear about only tagging the exact information that appears on the financial statements.

International. XBRL is being used around the world for a number of purposes, including financial reporting, banking, tax reporting, business registry, insurance, and governmental reporting. An example of international XBRL use is that of the National Bank of Belgium's Central Balance Sheet Office, which requires XBRL to be used in the reporting of annual accounts (XBRL International, 2013c).

A breakdown of international XBRL uses is shown in Figure 2.

Type of Implementation	Percentage of Total International Use of XBRL
Financial Regulator, Securities	47%
Financial Regulator, Banking	18%
Business Register	16%
(Standard) Business Reporting	9%
Tax Regulator	7%
Other	4%

Source: (XBRL International, 2013a)

Figure 2. International uses of XBRL.

Similar to the SEC, regulators of securities in other countries are using XBRL for reporting purposes. Examples are the Financial Services Agency of Japan, Bombay Stock Exchange, Tokyo Stock Exchange, Stock Exchange of Thailand, Spanish Securities Commission, Shenzhen Stock Exchange, Financial Supervisory Service of Korea, China Securities Regulatory Commission (for mutual funds), Israel Securities Authority, Korean Securities Dealers Automated Quotations, Shanghai Stock Exchange, and Oslo Stock Exchange. Even more countries are working toward requiring XBRL for financial reporting purposes (XBRL International, 2013c).

The Australian government utilizes XBRL for business-to-government reporting. It calls this initiative to simplify the reporting process Standard Business Reporting (SBR). The government agencies participating include the Australian Securities and Investments Commission, Australian Taxation Office, Australian Prudential Regulation Authority, all eight state and territory government revenue offices, and the Australian Bureau of Statistics (taxonomy only). Participants and new uses of SBR are constantly evolving with the Australian government. Businesses employ SBR-enabled software that interprets the financial data in their accounting systems, populates the required government forms, and sends the forms to the appropriate government agency (SEC, 2012). The Government of New Zealand and the Dutch Ministry of Finance are in the process of developing their own SBR using XBRL (XBRL International, 2013c).

The U.K.'s Her Majesty's Revenue and Customs department, the National Tax Agency of Japan, and the Federal Public Service Finance of Belgium all require corporate tax returns to be filed in XBRL (XBRL International, 2013c). Several more countries are in the process of mandating tax reporting using XBRL (XBRL International, 2012).

Banking is another area that has made use of XBRL internationally. Banks that have adopted XBRL for regulatory reporting include the Bank of Japan, Bank of Spain, Deposit Insurance Corporation of Ontario (for credit unions), Financial Supervisory Service of Korea, European Banking Authority, U.S. FDIC, and U.S. Federal Financial Institutions Examination Council.

The main sources of XBRL filings from all over the world are shown in Figure 3.

Source of XBRL Filings	**Percentage of Total International XBRL Filings**
United Kingdom Her Majesty's Revenue and Customs	27%
InfoCamere	20%
United Kingdom Companies House	18%
German Bundesbank	7%

National Bank of Belgium	7%
Belgium Ministry of Finance	6%
Australian Standard Business Reporting	2%
Danish Business Authority	1%
United States Securities and Exchange Commission	1%
Other	11%

Source: (XBRL International, 2013a)

Figure 3. International sources of XBRL filings.

There are a number of potential advantages to using data in XBRL format, and several areas have had success in utilizing it. Now that company-reported financial statement data filed with the SEC are available in XBRL format, a great deal of research is needed to investigate areas in which using this data would be more beneficial than what was previously used. The first of these areas was explored in the current study. Specifically, XBRL company filing data was used to predict future earnings and their prediction accuracy was compared to that of Compustat.

Future Earnings Prediction Using Financial Statement Information

An area that relies heavily on future earnings predictions using financial statement information is that of fundamental analysis. Fundamental analysis involves the use of information contained in financial statements to predict future earnings of a company and ultimately determine the company's intrinsic value. This intrinsic value can be compared to the market price, allowing mispriced securities to be identified. Because price will eventually gravitate toward value, identification of mispriced securities enables abnormal returns to be earned (Kothari, 2001; Richardson, Tuna, & Wysocki, 2010).

The research on fundamental analysis focuses on the ability of fundamentals (financial variables) to predict either future earnings or stock returns, but research on the prediction of future earnings has a number of important benefits. Richardson et al. (2010) point out that "given the multiple other users of general purpose financial reports (e.g., customers, suppliers, competitors, management, etc.) that make financial decisions, analysis to improve forecasting models of future earnings is invaluable" (p. 450). Abarbanell and Bushee (1997) state that "predicting accounting earnings, as opposed to explaining security returns, should be the central task of fundamental analysis" (p. 1). This belief is held by several researchers in the area of fundamental analysis. For instance, Graham, Dodd, and Cottle (1962) assert that the "most important single factor determining a stock's value is now held to be the indicated average future earning power, i.e., the estimated average earnings for a future span of years. Intrinsic value would then be found by first forecasting this earning power and then multiplying that prediction by an appropriate 'capitalization factor'" (p.28). Penman (1992) states that "the task of research is to discover what information projects future earnings and, from a financial statement analysis point of view, what information in the financial statements does this" (p.471).

Data Sources

Once proprietary databases of company financial data became available, they were often used as the data source for many empirical accounting studies because of the amount of time and effort saved over hand collection of company data. There are several proprietary databases containing company financial data. Access to some of these databases is costly; however, they allow for quick and easy access to large amounts of company information. Examples include Compustat and Value Line. There are also websites that provide company financial data at no charge, although using these financial websites is not a particularly fast method of gathering large amounts of company data. Examples of these financial websites include Google Finance, Yahoo! Finance, and MSN Money.

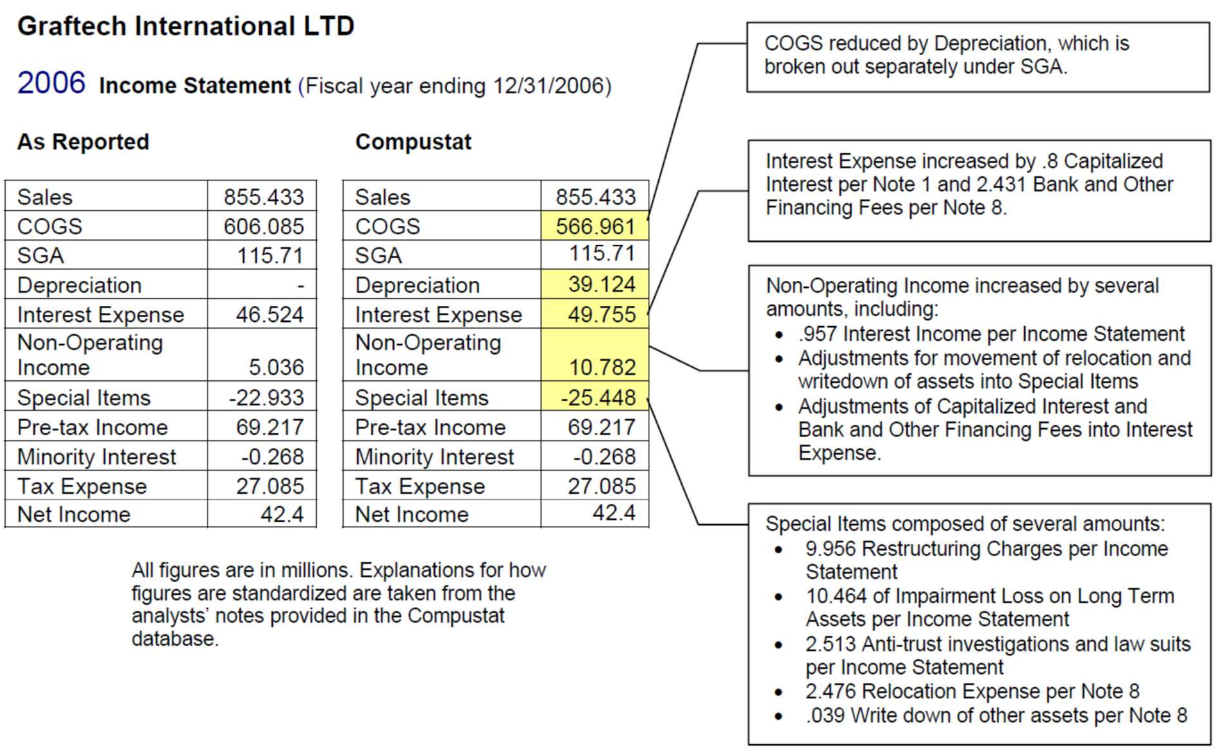
Although use of these proprietary databases saves time and effort over manual collection of company financial data, subscriptions to the databases are often very expensive. XBRL company filings are freely available, and because they are computer-readable, extraction of the data should save a great deal of time and effort over hand collection of data from company financial statements. Vasarhelyi et al. (2012) state that "the very purpose of XBRL [is the] automation of data parsing" (p. 157). XBRL company filings also have a significant time advantage. XBRL 10-K and 10-Q filings are published concurrently with the related PDF filing versions. However, it takes an average of 14 weekdays from the time a company files with the SEC for that accounting data to appear in Compustat's Research Insight database (D'Souza et al., 2010). XBRL company filings include more companies than most, if not all, data aggregators. Compustat and Thomson ONE Banker together only cover about 70% of the companies that file with the SEC. Many of the smaller companies' data are not included in the Compustat database; therefore, XBRL company filing data covers a much broader group of companies (SEC, 2009).

Compustat is widely used in accounting research today and specifically in earnings prediction research. Compustat is a database of financial, statistical, and market information on active and inactive global companies throughout the world. The service began in 1962, and since then, thousands of research studies have relied on the information obtained through Compustat. The database is published by Standard & Poor's and, like most proprietary data aggregators, it employs standardization techniques for the company data that it collects. The reason for this, according to Compustat, is that:

> Companies often present their financial results in a variety of formats, making it difficult to construct parallel company comparisons on an "apples-to-apples" basis. After collecting data from a diverse set of sources, we standardize it by financial statement and

by specific data item definition, preparing information that is comparable across companies, industries and time periods. (Capital IQ Compustat, 2013)

Some of these standardization techniques are illustrated in Figure 4, which uses GrafTech International LTD's 2006 income statement data as an example. Company-reported data (labeled "as-reported") is shown in the table to the left, while Compustat data is shown in the table to the right. Any amounts that differ between as-reported and Compustat are highlighted. Explanations for the differences in amounts are to the far right. All differences in this example are due to standardization techniques employed by Compustat.



As Reported	
Sales	855.433
COGS	606.085
SGA	115.71
Depreciation	-
Interest Expense	46.524
Non-Operating Income	5.036
Special Items	-22.933
Pre-tax Income	69.217
Minority Interest	-0.268
Tax Expense	27.085
Net Income	42.4

Compustat	
Sales	855.433
COGS	566.961
SGA	115.71
Depreciation	39.124
Interest Expense	49.755
Non-Operating Income	10.782
Special Items	-25.448
Pre-tax Income	69.217
Minority Interest	-0.268
Tax Expense	27.085
Net Income	42.4

Source: (Capital IQ Compustat)

Figure 4. Example of Compustat's standardization techniques.

Differences in data sources can have serious implications for research and decision-making. Kern and Morris (1994) warn that "[a]nalysts and researchers need to exercise great care when selecting databases and variables from those databases. These choices can affect the results of and the inferences drawn from empirical research in ways more than is anticipated by researchers" (p. 280). Differences between accounting data reported by a company and those reported by data aggregators such as Compustat may differ because of standardization techniques, as described above, because of erroneous data reported in the data aggregator database, or because data is missing in the data aggregator database.

During a speech about the potential uses of XBRL, former SEC Chairman Christopher Cox (2006) stated:

> Executives who have taken the time to double check the data that financial analysts following their companies are working with can sometimes get quite a shock. That's because some of them bear no resemblance to what the companies published. The truth is, too many CEOs have no idea what happens to their information after it leaves their control in the form of SEC-mandated financial statements. When they are asked, "Do you know where analysts get data on your companies to populate their valuation models?" they usually reply, "well, from our financial statements."
>
> BZZZZZ. Wrong answer. And then, their first reaction is surprise. That surprise turns to concern when they realize that the numbers the analysts are using in their valuation models can have an error rate of 28%, or higher still if the data in question comes from the footnotes.

Currently, data aggregators are required by anyone needing large amounts of company accounting data because manual collection of accounting data from paper or PDF financial statements is very labor-intensive and costly. XBRL is intended to be the tool that allows users to extract accounting data from company-reported financial statements without this manual collection. Understanding the implications of using any particular data source is extremely important.

III. RELATED RESEARCH

XBRL: Related Research

Deficiencies in XBRL for financial reporting. Much of the research on XBRL for financial reporting since the time XBRL began being used for filings to the SEC has focused on the quality of the XBRL filings. A common theme of these studies is that the quality has been low for XBRL filings to the SEC, and much of the blame for this has been put on the XBRL company filing preparers.

***Data quality due to preparers*.** Numerous researchers have stated that even though the SEC does not require an audit of the XBRL filings, it is obvious that these documents must be of high quality and free of errors (Fang, 2011; Srivastava & Kogan, 2010; Garbellotto, 2009; Alles & Gray, 2012). The AICPA's Statement of Position 09-1 states:

> In order for XBRL to be a useful tool for investors and other users of business information, the data contained in XBRL files must be accurate and reliable. Preparers of XBRL-tagged data may be issuers or non-issuers and responsible for providing accurate information in their XBRL files on which investors and other users of business information may rely. (AICPA, 2009, Item 4, p.2)

An XBRL US consortium survey discovered that of the estimated 500 companies from the first reporting phase,[5] only 340 had correctly converted their financial statements to

[5]The SEC ruled in January 2009 that all public companies must report financial statements in XBRL by June 2011. Companies were phased in over the periods 2009 through 2011.

XBRL filings. This indicates that companies were having difficulty correctly implementing their initial XBRL filing (Sledgianowski et al. 2010).

Bartley, Chen, and Taylor (2010a) examined early XBRL filings and found that the XBRL 2006 and 2008 voluntary 10-K report submissions of almost all the companies examined contained errors that would not be acceptable under the SEC's rules for mandatory submissions. Bartley et al. (2010a) point out that these errors occurred in all the various steps of the preparation, including mapping, extension, tagging, creating, and validating. In a subsequent study, Bartley, Chen, and Taylor (2011) found a dramatic decrease in the number of errors in XBRL instance documents in 2008 compared with 2006 XBRL instance documents. The authors attributed this largely to preparer experience.

Boritz and No (2008) gathered information about the quality of XBRL filings in the SEC's XBRL Voluntary Filing Program (VFP). The stated purpose of the paper was to determine whether there was a need for some type of quality assurance for XBRL-tagged information furnished to a regulatory authority and whether the XBRL-Related Documents furnished under the VFP conformed to the suggested XBRL taxonomies, specifications, and requirements of the VFP. Boritz and No (2008) performed validation tests on 304 XBRL filings furnished by 74 participants in the VFP from its inception in 2005 to December 31, 2007. While the researchers said that 272 of the 304 filings passed their taxonomy validation tests, only 104 of the 304 filings (34.2%) passed the instance document validation test without any errors or exceptions being noted. The authors observed that none of the companies passed Financial Reporting Instance Standards and Financial Reporting Taxonomies Architecture validation tests. They also found that all 304 of the companies examined used some form of extension taxonomies in order to customize labels, presentation, match-up subtotals, etc. Boritz and No (2008) believed this to be a source of a significant number of errors in the filings.

In a subsequent study, Boritz and No (2009) conducted a mock assurance engagement on United Technologies Corporation's (UTC) XBRL-Related Documents to address the implicitly posed research question: What assurance issues can arise in the conversion of paper paradigm financial statements to XBRL-Related Documents? After the mock assurance procedures on UTC's 10-Q XBRL-Related Documents, the authors had high assurance that the 10-Q XBRL-Related Documents were a complete and accurate reflection of UTC's 10-Q. A conclusion of fairness of the presentation in accordance with GAAP of the XBRL-Related Documents was not possible because of the lack of assurance standards or guidelines for making such an assessment for various sections. They also made the following observations:

- Fifty-four percent of the instance document was based on UTC's taxonomy extensions. Twelve redundant elements (i.e., elements that existed in the approved standard taxonomy and therefore did not need to be extended) were found in the UTC taxonomy.
- The contexts used by UTC had unorganized naming, numbering, and locations that increased the difficulty of analysis and could cause problems in subsequent periods to determine inter-period consistency.
- Titles and subtitles were not presented in a consistent manner. Labels were also missing or not the same in the label linkbase as in the taxonomy.
- Some of the subtotals which appeared in the 10-Q were omitted in the instance document.

- A disaggregation of textual narratives was found in both the notes and the Management Discussion and Analysis.
- UTC used taxonomy extensions for the “Report of Independent Registered Public Accounting Firm” instead of using the approved US Financial Reporting-Accountants Report (USFR-AR) taxonomy. It would have been more appropriate to use the standard taxonomy.

Debreceny, Farewell, Piechocki, Felden, and Gräning (2010) examined the quality of the XBRL filing data in the EDGAR database repository and found that one-quarter of the initial 400 XBRL filings had errors. The authors reported that the primary cause of the errors was inappropriate treatment in the instance documents of underlying debit/credit assumptions in the taxonomy. Debreceny, Farewell, Piechocki, Felden, Gräning, and d’Eri (2011b) found that 40% of the extensions included in the XBRL filings from the first reporting phase were unnecessary (i.e., the appropriate tags were in the US GAAP Financial Reporting Taxonomy, and an extension tag did not need to be created).

Janvrin and No (2012) conducted a field study to examine the process of XBRL implementation. They identified four issues that still hinder XBRL implementation: lack of management support and involvement, conflicts about implementation approach, lack of organizational readiness and expertise, and lack of internal controls over the tagging process.

Du, Vasarhelyi, and Zheng (2013) found a significant decrease in the number of errors in XBRL company filings from June 2009 to December 2010. This finding supports the notion that preparers were learning from their experience in preparing XBRL filings with the SEC and therefore the subsequent filings were of higher quality.

Data quality due to regulation. The Center for Excellence in Accounting and Security Analysis (CEASA) released a white paper in January 2013 entitled *An Evaluation of the Current State and Future of XBRL and Interactive Data for Investors and Analysts* (Harris & Morsfield, 2012). The paper describes the project undertaken by the organization, which included interviews with analysts, investors, regulators, preparers, XBRL developers, data aggregators, and XBRL filing and consumption tool vendors. CEASA also surveyed and had discussions with investors and analysts. The authors of the study stated, "The survey and interview questions, and our conclusions, were organized around the original vision for interactive data—i.e., that data in this format would provide incrementally more relevant, timely, and reliable information to more end users, who could then manipulate and organize the data according to their own purposes at a lower cost" (p. i).

Aggregated financial statements were once necessary because of limited data processing capabilities. With much more sophisticated data processing capabilities, this limitation of data is no longer necessary. Harris and Morsfield (2012) found that analysts and investors wanted more disaggregated information so that they could manipulate the data in a way to perform specific analyses, stating that the "ability to query high levels of detail when desired is the power of interactive data" (p. 7). Currently, this objective cannot be achieved because the focus of building an XBRL filing is to mimic a portion of a regulatory filing (i.e., XBRL tagging follows traditional paper filings too closely). Much of the information users require does not appear on the face of the financial statements or in an SEC filing. This narrow focus when creating XBRL filings does not provide adequate incremental value over the HTML filings that companies were already providing. The XBRL users in the CEASA study believed that the usability of the financial statement data was compromised due to the requirement that the XBRL financial statement data be presented just as it is in the paper financial statements (Harris & Morsfield, 2012).

The authors of the CEASA study found that several investors and analysts felt that XBRL data has the potential to replace manual collection of the data they need. However, those investors and analysts who tried to use XBRL data for a large number of company furnishings were very dissatisfied with the usefulness and usability of the XBRL data (Harris & Morsfield, 2012). The authors observed an "expectations gap," where the gap between preparers' expectations of how much the XBRL data would be used and how much it is actually used was significant. Similarly, the gap between investors' expectations about the usability of XBRL data and its actual usability was also significant (Morsfield, 2012). Users also expressed a desire for XBRL consumption and analysis tools that do not require programming or query language knowledge before they would be willing to integrate XBRL into their workflow (Harris & Morsfield, 2012).

Harris and Morsfield (2012) described the following three detrimental decisions as the reasons why, in their opinion, XBRL for financial reporting has not worked well thus far:

a. **The decision to frame the regulation so that it appeared to many filers and to the XBRL development community that filers had to create an XBRL-tagged reproduction of the paper or HTML presentations of their filings:**

   - We believe this presentation-centric step hindered or diverted what should have been an important evolution from a paper presentation-centric view of financial reporting information to a far more transparent and effective data-centric one.
   - Valuable resources were spent on learning the details of a technology rather than on its use for enhanced financial reporting processes, leading to better

analysis and decision-making, both within the filing firms and for end users of their data.

b. The design of XBRL filing and consumption technology such that it requires extensive and detailed technical knowledge to input or to extract data:

- We are not technologists and we believe that one should not have to become a technologist to the level required by XBRL in order to either input or extract financial data.
- We believe this contributes to data and tagging errors by filers, as well as to lack of interest on the part of investors and analysts to date.

c. The reticence (or inability) of regulators and filers to ensure that the interactive filings data are accurate and correctly-tagged from day one of their release to the public and forward (or, to communicate to the market for this information that they were not insisting on this and why):

- We believe this is a key reason that the data are not being used as much as their potential would suggest.
- The regulation, as written, provided numerous loopholes that permitted filers to submit filings with low-quality data and tagging
    - Limited liability for filers
    - No external auditor requirement.
- Filers were unwilling or unable to ensure the quality of their data
    - Interactive filings data did not match the related EDGAR filing data
    - Incorrect tags were utilized
    - Unnecessary and extensive custom tags were used (p. 37)

Debreceny, d'Eri, Felden, Farewell, and Piechocki (2011a) assessed the existence of 38 accounting concepts in the US GAAP Financial Reporting Taxonomy required for calculating 63 different financial ratios. They found that not all accounting concepts had a direct match but that some of the concepts could be calculated by combining more than one XBRL tag. They also captured various levels of alternative matches. These were "close" but not exact matches and had the potential to lower the information quality of the ratios when used. Debreceny et al. (2011a) then assessed the existence of the same 38 accounting concepts in the 2010 XBRL filings for direct matches, combinations of XBRL tags, and alternative matches. They found that some of the accounting concepts were rarely available, while others were available in most filings. Finally, Debreceny et al. (2011a) assessed the existence of accounting concepts in the 2010 XBRL filings for calculating 19 of the 63 financial ratios using a "tight" requirement and a "loose" requirement. They explained the differences between these two requirements using the debt/equity ratio as an example. The "tight" requirement for the calculation of debt required both *LongTermDebtNoncurrent* and *LongTermDebtCurrent* to appear in the instance document (the authors determined that adding these two accounting concepts together was the most direct way to arrive at total debt). The "loose" requirement allowed either of these elements to appear in the instance document. The percentage of ratios that could be calculated under the "tight" and "loose" requirements ranged from 0% to 98% and 45% to 99%, respectively.

Debreceny et al. (2011a) stated that one of the research questions in the study was: "[W]here automatic creation of ratios from the filings is not feasible, what are the causes? Is this a function of the way corporations report in the financial statements that are the foundation for the XBRL filings or are the roots in the details of the XBRL implementation?" (p. 2). However, the authors did not answer this question in the latest version of their working paper.

Prior to the SEC's mandate that XBRL be used, R. D. Plumlee and Plumlee (2008) wrote a commentary, before XBRL was required by the SEC, to provide direction regarding assurance issues in relation to XBRL filings. The authors felt confident that once filings in XBRL were mandatory, users would demand assurance on the tagging process. They stated that while XBRL was only voluntary, the focus was only to "agree" the XBRL filing to the related official filing, and they referred to this as a paper-oriented reporting paradigm. The authors went on to state:

> However, once filing in XBRL becomes required, the power of XBRL to allow individual financial datum to be extracted from the SEC's financial database will be realized. This "data-centric" idea is a crucial extension of the traditional reporting paradigm that will alter the way financial and nonfinancial data can be used. (p. 353)

This insightful observation of what the authors thought the future of XBRL company filings would be was, and still is, what users are waiting for. Unfortunately, the SEC and other XBRL regulatory bodies have not yet made a shift from this paper-oriented reporting paradigm or paper presentation–centric view of financial reporting information to a data-centric view of financial reporting information. Only then will the true power of XBRL be realized in the XBRL company filings.

Future Earnings Prediction Using Financial Statement Information: Related Research

Several studies have used financial ratios to predict earnings. For example, Ou and Penman (1989a, 1989b) developed a summary measure that estimates the probability of a future earnings change. This summary measure was calculated based on financial ratios combining large amounts of accounting data. Sixty-eight financial ratios were initially chosen to potentially include in the summary measure, based on a survey of financial accounting and financial analysis texts as to which items might predict future earnings. These 68 ratios were pared down to a parsimonious set of ratios that were the most effective at predicting future earnings change. The logic behind Ou and Penman's model begins with a simple valuation model:

$$V = E(d) / r,$$

where V is a stock's value, E (d) is expected future dividends, and r is the rate at which future dividends are discounted. Financial statement items that are correlated with future dividends (the numerator in the above equation) were identified. This step was justified based on the fact that, intuitively, future dividends are paid out of future earnings, and Ball and Brown (1968) showed that higher earnings imply higher firm values. Therefore, future earnings are value-relevant. Given this fact, Ou and Penman identified accounting data that predicts earnings increases and earnings decreases one year in advance. Logistic regression was used to estimate the summary measure ($\widehat{Pr}$) based on selected accounting data. The $\widehat{Pr}$ value predicted earnings changes over 60% of the time and with more accuracy than time-series models.

The original purpose of Ou and Penman's study (1989b) was to predict stock returns. In a later paper, Penman (1992) expressed some regret that the study extended this far because he felt doing so distracted from the financial statement analysis.

Ou (1990) found that non-earnings accounting data contains information about future earnings that current and prior earnings do not. Ou (1990) described the relation between non-earnings accounting data (defined as all accounting data except earnings) and stock returns as a two-link process, depicted in Figure 5. A predictive information link exists between non-earnings accounting data and future earnings changes, while a valuation link exists between predicted future earnings changes and stock returns.

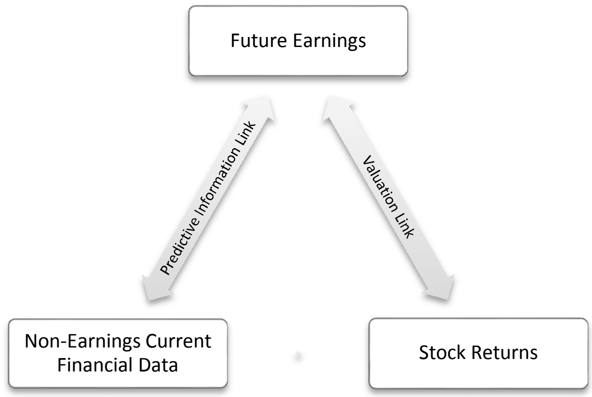


Figure 5. Non-earnings accounting data and stock returns as a two-link process.

Lev and Thiagarajan (1993) used accounting information to determine excess returns, employing a "guided search procedure" to identify those ratios and other fundamental signals that are used by investors to determine the quality and growth of earnings. Ou and Penman (1989b, 1989a) used a "statistical search" for the appropriate ratios, employing sources such as *The Wall Street Journal*, *Barron's*, Value Line publications, newsletters of major securities firms, and professional commentaries on corporate financial reporting and analysis. The 12 ratios and other fundamental signals they found to be used by investors to predict future earnings were related to inventory; accounts receivable; capital expenditures; research and development; gross margin; selling and administrative; provision for doubtful receivables; effective tax rate; order backlog; labor force; last-in, first-out earnings; and audit qualification. Lev and Thiagarajan found that most of the ratios and other fundamental signals predicted contemporaneous abnormal returns beyond that of current earnings and that this incremental explanatory power increased when the ratios and other fundamental signals were conditioned on macroeconomic variables. The macroeconomic variables used were inflation, gross national product growth, and business inventories.

Unlike Lev and Thiagarajan (1993), who studied the link between fundamentals and contemporaneous abnormal returns, Abarbanell and Bushee (1997) first studied the link between fundamentals and future earnings changes. They used the ratios and other fundamental signals that were used in Lev and Thiagarajan's (1993) full sample[6] to determine if there was economic justification for analysts and investors to rely on those signals. They also determined how efficiently analysts used the fundamental signals in their forecasts. Abarbanell and Bushee did find an association between some, but not all, of the fundamental signals and future earnings changes. They also found that analysts did not fully use the information in the fundamental signals when making earnings forecasts. In a subsequent study, Abarbanell and Bushee (1998) found an association between these fundamentals and future abnormal returns.

Differences Among Data Sources: Related Research

Prior research on data sources used in accounting and finance has found significant variances among data sources—some that have caused results of empirical studies to be vastly different, depending on the choice of the data source. Differences between accounting data reported by a company and those reported by data aggregators such as Compustat may occur because of standardization techniques, because of erroneous data reported in the data aggregator database, or because data is missing in the data aggregator database. The choice of data source can have a serious impact on research results and decision-making.

Zimmerman (1983) and Porcano (1986) did studies examining the relationship between firm size and effective tax rates and found conflicting results. One difference in their studies was

[6]The full sample included nine of the original 12 fundamental signals in order to substantially increase the sample size and make the sample more representative. The three fundamental signals that were eliminated were related to research and development; provision for doubtful receivables; and order backlog.

that Zimmerman had used Compustat to gather his data, while Porcano had used Value Line. Kern and Morris (1994) replicated Porcano's (1986) study using both Compustat and Value Line data in order to investigate if the two different data sources could have been a factor in the conflicting results. They included only the firms that were common to both databases and found that the average annual effective tax rate measure was significantly different when calculated using Compustat data versus Value Line data for 12 of the 14 years in the study. Kern and Morris (1994) warned that "[a]nalysts and researchers need to exercise great care when selecting databases and variables from those databases. These choices can affect the results of and the inferences drawn from empirical research in ways more than is anticipated by researchers" (p. 280).

Philbrick and Ricks (1991) examined earnings per share as it was reported in Value Line, I/B/E/S, and Compustat from 1984 to 1986 and found differences across the databases. They stated that the source of actual EPS data, but not analysts' forecast data, had a significant impact on empirical associations between forecast errors and stock returns.

Yang, Vasarhelyi, and Liu (2003) evaluated data reported in Compustat and Value Line. First, they gathered seven variables, found to be most often used in empirical research, from each of the two data sources for 1,479 companies. They found that 12.4% of the comparisons had discrepancies larger than 1% and that 5.02% had missing fields. Because the discrepancies were much higher than expected, further analysis was done. Two hundred companies were selected, and the same seven variables from the first analysis were obtained from Compustat, Value Line, and the original financial statements. The authors categorized discrepancies as explainable (foreign currency differences, industry factors, and definitional factors) or unexplainable (non-disclosed coding rule differences and coding errors). One hundred and thirty five unexplainable discrepancies were found; 99 of those discrepancies were found in Compustat, and 36 were found in Value Line. A final analysis was done to examine the impact of using Compustat and Value Line on the cross-sectional distributional properties of ratios. Ten of the 11 ratios examined by Deakin (1976) were calculated by Yang et al. with data from Compustat and Value Line. Yang et al. found that two of the 10 ratios produced significantly different mean, variance, skewness, and kurtosis measures depending on which database was used.

Tallapally, Luehlfing, and Motha (2011) found differences in the amounts for cost of goods sold reported in Compustat and EDGAR Online for every company except one in their sample of 26 manufacturing companies. In a subsequent study, Tallapally et al. (2012) compared the sales/revenue amounts reported by both Compustat and XBRL for 27 companies for the years ending after June 30, 2009 but before July 1, 2010 and found that differences did exist. They stated that reconciliations of the differences were not obvious.

Boritz and No (2013) analyzed and compared financial facts reported by three data aggregators—Compustat, Yahoo! Finance, and Google Finance—with those reported in the associated interactive data renderings.[7] They found that more than half of the financial facts reported in the interactive data renderings were not available from the data aggregators. In addition, of the financial facts that were available from both interactive data renderings and data aggregators, 4.8% of the facts did not match when comparing from the rendering to the aggregators and 8% did not match when comparing from the aggregators to the rendering. Of these mismatches, 55.7% were materially different.

---

[7]Examining the interactive data, or XBRL, renderings are (or should be) equivalent to examining the PDF version of the 10-K and not equivalent to examining XBRL instance documents. Renderings only show XBRL labels and formatting, which are required to mirror the PDF 10-K filings. All the interactive data renderings examined by Boritz and No (2013) were, in fact, identical to the related PDF 10-K filing.

Chychyla and Kogan (2013) examined 30 accounting items as reported in Compustat and plain-text 10-K filings. This was done by starting with XBRL 10-K filings and then removing differences between XBRL 10-K filings and plain-text 10-K filings. Five thousand companies were examined from the period of October 1, 2011 to September 30, 2012. The authors found that 22 of the 30 accounting items were significantly different between the two data sources.

IV. RESEARCH QUESTIONS AND HYPOTHESES DEVELOPMENT

Based on prior research and preliminary investigations, a significant number of required accounting elements were expected to be missing from the current XBRL company filings. Therefore, the expectation was that interactively calculating the variables required to create earnings prediction models would not be possible. Harris and Morsfield (2012) comment on this issue:

> ... for financial reports, the SEC had mandated filings through the EDGAR system and went through the multi-year pain of getting issuers onto a web-based filing platform (in HTML format). Yet, even with this there was potentially a need for structuring the data in a way that it could be easily used interactively. This was the presumed motivation of the SEC mandating that filings were also done in XBRL. But, as articulated above, this is one of the problems. The SEC's XBRL mandate had a presentation (filing)-centric focus rather than a data-centric focus. That is, the focus became that of formatting data to accommodate a specific filing or presentation, rather than on making individual data points available for the end user to utilize or present as they required. (p. 41)

This leads to the first research question:

> **Research question 1 (RQ_1)**: What proportion of the accounting elements needed to create the earnings prediction models is tagged in current XBRL company filings?

An important step in analyzing the ability of XBRL company filings to provide interactive use of company financial statement data is to compare the XBRL tags used in XBRL company filings to the explicitly listed accounting concepts in companies' audited 10-Ks. This makes it possible to make some determinations about what types of issues or obstacles exist in making XBRL company filings interactive. Specifically, it is important to determine whether accounting concepts tagged by companies in their XBRL filings were explicitly listed in their audited 10-Ks—and in turn, whether accounting concepts explicitly listed by companies in their audited 10-Ks were appropriately tagged in their XBRL filings. This is the procedure that the SEC expects XBRL company filing preparers to follow. Therefore, the first hypothesis (in null form) is:

> **Hypothesis 1 (H_1)**: The number of accounting concepts appropriately tagged in XBRL company filings will not be significantly different from the number of accounting concepts explicitly reported in companies' audited 10-Ks.

The deficiencies in the current XBRL company filings are presumed by many to be due to the SEC's protocol, which requires that current XBRL company filings mimic the related paper filings. Much of the information required by users of company filing data is not explicitly tagged in XBRL company filings. Many of these missing tags could be automatically populated based on component XBRL tags by a function incorporated directly into the XBRL taxonomy. This more complete set of XBRL filings is referred to as "fully populated XBRL company filings." Fully populated XBRL company filings are XBRL filings whose missing tags have been automatically populated based on component XBRL tags, potentially making XBRL financial statement data more useful by providing a more complete collection of accounting concepts that could be extracted without manual calculation. For example, if a company tagged an amount for current liabilities and noncurrent liabilities—but not for total liabilities—in its XBRL filing, the amount for total liabilities could be calculated by functionality built into the XBRL taxonomy that sums the amounts for current liabilities and noncurrent liabilities. The reason this can be done is because the hierarchy of the tags and how they relate to one another (parent elements, child elements, sibling elements, etc.) is already built into the XBRL taxonomy. Figure 6 displays the relationship of these three tags as shown in the XBRL taxonomy.

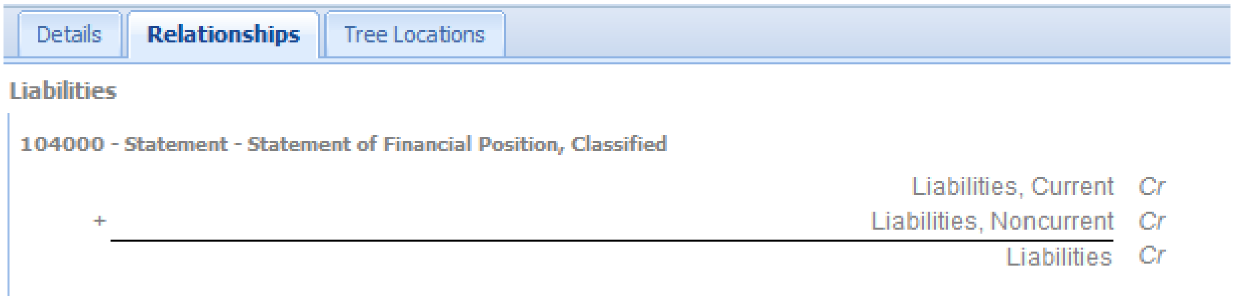


Figure 6. Relationships view for liabilities within the XBRL taxonomy**.**

Although not all missing tags can be automatically populated, fully populated XBRL filings are expected to have improved usability compared to the current XBRL filings. Fully populating XBRL company filings was achieved by first identifying the required XBRL elements needed from the XBRL company filings in order to create the two earnings prediction models used in this study. Any missing XBRL elements were investigated to see if they could be manually calculated solely based on the existence of sufficient component elements. This was done to mimic the process that functionality built directly into the XBRL taxonomy could perform. If any XBRL element required human subjectivity to calculate, it was not manually populated, and none of the balances tagged by a preparer were altered in any way.

To illustrate, consider the previously cited example of a company choosing to display its liabilities section on its balance sheet of the paper/PDF 10-K as follows:

Current liabilities:	
Accounts payable	$ 100,000
Other current liabilities	25,000
Total current liabilities	125,000
Long-term debt	75,000

According to the SEC's protocol for the preparation of XBRL company filings, the preparer of this company's XBRL 10-K filing should not tag an amount for total liabilities because it does not explicitly appear on the face of the paper/PDF 10-K, which is purely a formatting issue. In order to make the XBRL company filing fully populated, the XBRL tag for total liabilities would be filled in with the missing $200,000. If the company had erroneously tagged long-term debt with anything other than a positive 75,000, that tag would not be filled in for the analysis because this would be an error created by the preparer, not by the SEC's protocol for the preparation of XBRL company filings.

This process was not subjective on the part of the researcher, as missing amounts were filled in only if:

1. The necessary component information was available.
    - For example, if current liabilities and long-term liabilities were given, then total liabilities could be determined by adding them together.
2. The preparer did not fill in a tag for that balance—and should not have, according to the SEC.
    - This would be the case if the balance was not explicitly listed on the audited 10-K.

This leads to the next research question:

Research question 2 (RQ_2): What proportion of the accounting elements needed to create the earnings prediction models is tagged in fully populated XBRL company filings?

The evaluation of the advantages of the fully populated XBRL company filings are important because, as previously stated, the SEC and other proponents of XBRL argue that XBRL company filing data offers a number of advantages over the data provided by data aggregators, such as lower cost, quicker availability, and broader coverage of companies. Therefore, examining the abilities of XBRL company filing data compared to that of data provided by data aggregators is worthwhile. There are a number of possible areas in research and practice that could be explored. One important area is earnings prediction. Understanding how the different data sources affect the prediction accuracy of an earnings prediction model is important in order to improve upon later earnings forecasting models. It may also illustrate how disparities in research results exist depending on the data source used. This leads to the final hypothesis (in null form):

Hypothesis 2 (H_2): Earnings prediction models created using fully populated XBRL company filing data will not predict earnings with a different accuracy as earnings prediction models created using Compustat data.

V. RESEARCH DESIGN

The sample for this study was drawn from the companies that made up the Standard & Poor's 500 Index (S&P 500) as of December 31, 2012. These large companies were chosen as the sample because XBRL company filing data became available in 2009 only for companies with a market capitalization of more than $5 billion. Data from 2009 was required for this study. Seventy of these 500 companies were financial institutions, and because their disclosure and presentations standards differ from other types of companies, they were eliminated from the sample. Of the remaining 430 companies, 134 were eliminated because they were not first-year filers and therefore insufficient information was available as of the time of the study (a first-year filer is a company required to begin filing its 10-K using XBRL on or after September 15, 2009). The final main sample included 296 companies that were part of the S&P 500 on December 31, 2012, and these companies were used to evaluate RQ_1. Table 1 lists descriptive statistics for the 296 companies in the sample.

Table 1

Descriptive Statistics for Companies Included in Sample

		N	Frequency	Percent
Stock Exchange	New York Stock Exchange	296	236	79.73
	NASDAQ-NMS Stock Market	296	60	20.27
Size (Revenues)	< $1,000,000,000	296	1	0.34
	$1,000,000,000 - $5,000,000,000	296	63	21.28
	$5,000,000,000 - $10,000,000,000	296	71	23.99
	$10,000,000,000 - $50,000,000,000	296	123	41.55
	$50,000,000,000 - $100,000,000,000	296	24	8.11
	< $100,000,000,000	296	14	4.73
Industry (SIC Code)	Agriculture, Forestry & Fishing (01-09)	296	1	0.34
	Mining (10-14)	296	28	9.46
	Construction (15-17)	296	3	1.01
	Manufacturing (20-39)	296	137	46.28
	Transportation, Communications, Electric, Gas & Sanitary Services (40-49)	296	52	17.57
	Wholesale Trade (50-51)	296	6	2.03
	Retail Trade (52-59)	296	26	8.78
	Finance, Insurance & Real Estate (60-67)	296	5	1.69
	Services (70-89)	296	37	12.50
	Public Administration (91-99)	296	1	0.34

Descriptive statistics for the 296 companies in the sample used to evaluate RQ_1.

A sub-sample (50 companies) of the 296 companies described above was also taken for additional analyses. A sub-sample of 50 was used for the additional analyses due to the fact that this process involved a great deal of manual data collection and calculations. This sub-sample was chosen at random by assigning each of the 296 companies a random number, after which the list was sorted in ascending order by the random number. The first 50 companies on the list were included in the sub-sample. This sub-sample was used to address H_1, RQ_2, and H_2.

XBRL financial statement data was obtained from the 296 companies' XBRL filings using XBRLAnalyst. XBRLAnalyst is an Excel add-in created by FinDynamics that allows the import of specific XBRL data elements into Excel spreadsheets using download features and

Excel function calls. Audited financial statement data was also obtained from the HTML 10-K reports in the EDGAR database repository, and the Compustat data was extracted from the Compustat Annual Fundamentals database for the 50-company sub-sample.

RQ_1 addresses the availability of accounting elements that can be interactively obtained with current XBRL company filings. Specifically, RQ_1 asks the following:

> What proportion of the accounting elements needed to create the earnings prediction models is tagged in current XBRL company filings?

The variables needed for the earnings prediction models are listed in Table 2. The exact calculations for each of the variables, including the Compustat data items and XBRL tags that were used in this study, are listed in Appendix A.

Table 2

Variables Required to Create the Earnings Prediction Models

1	% Δ in current ratio
2	% Δ in inventory turnover
3	Inventory / total assets
4	% Δ in (inventory / total assets)
5	% Δ in inventory
6	% Δ in sales
7	Δ in dividend per share
8	% Δ in (capital expenditures / total assets)
9	% Δ in debt-equity ratio
10	Return on closing equity
11	Gross margin ratio
12	Sales to total cash
13	% Δ in total assets
14	Working capital / total assets
15	Operating income / total assets
16	Cash dividend as % of cash flows
17	Δ in earnings per share - drift
18	% Δ in inventory - % Δ in sales
19	% Δ in industry capital expenditures - % Δ in firm capital expenditures
20	% Δ in sales - % Δ in gross margin
21	% Δ in selling and administrative expenses - % Δ in sales
22	0 for LIFO, 1 for FIFO or other
23	Δ in earnings per share

The variables needed to create the O&P Earnings Prediction Model and the A&B Earnings Prediction Model.

To evaluate RQ_1, the proportion of accounting concepts exactly matching the 70 identified concepts that were found in the 296 XBRL company filings were calculated.

H_1 states that the number of accounting concepts tagged appropriately in XBRL filings is not significantly different from the number of explicitly reported accounting concepts in companies' audited 10-K reports for the 70 accounting concepts required to calculate the variables needed for the earnings prediction models.

To test H_1, a search was performed in each company's audited 10-K report to determine which of the 70 accounting concepts required to compute the 46 variables used in the analysis were explicitly listed. Each company's XBRL filing was searched to identify which of the 70 XBRL elements were present. A paired-samples t-test was conducted to analyze differences between the number of explicit accounting concepts found in the audited 10-K reports and the number found in the XBRL filings. An alpha level of 0.05 was used for this test. The 50 company sub-sample was used for this analysis.

RQ_2 addresses the availability of accounting elements that can be interactively obtained with fully populated XBRL company filings. Specifically, RQ_2 asks the following:

> What proportion of the accounting elements needed to create the earnings prediction models is tagged in fully populated XBRL company filings?

Fully populated XBRL company filings are XBRL filings whose missing tags have been automatically populated based on component XBRL tags. This is a function that could be incorporated into the XBRL taxonomy, potentially making XBRL financial statement data more useful by providing a more complete collection of accounting concepts that could be extracted without manual calculation. For example, if a company tagged an amount for current liabilities and noncurrent liabilities—but not for total liabilities—in its XBRL filing, the amount for total liabilities could be calculated by functionality built into the XBRL taxonomy that sums the amounts for current liabilities and noncurrent liabilities. The reason this can be done is because the hierarchy of the tags and how they relate to one another (parent elements, child elements, sibling elements, etc.) is already built into the XBRL taxonomy. Figure 7 displays the relationship of these three tags as shown in the XBRL taxonomy.

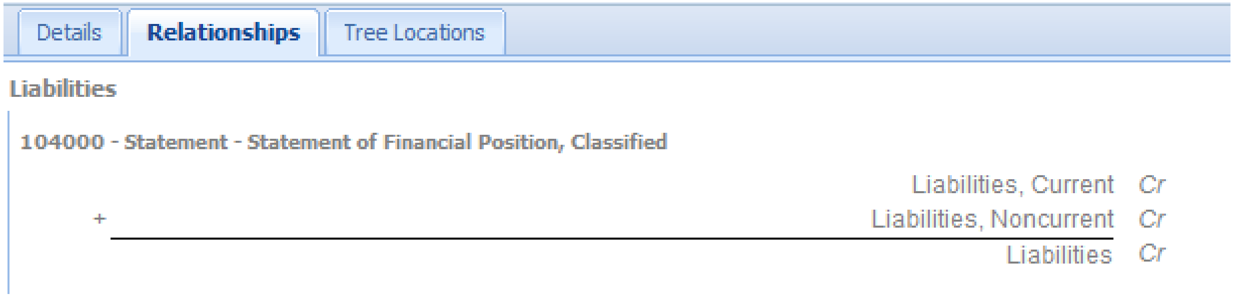


Figure 7. Relationships view for liabilities within the XBRL taxonomy.

Although not all missing tags can be automatically populated, fully populated XBRL filings are expected to have improved usability compared to the current XBRL filings.

To evaluate RQ_2, the required XBRL elements needed from the XBRL company filings to create the two earnings prediction models were identified. Any missing XBRL elements were investigated to see if they could be manually calculated solely based on the existence of sufficient component elements. This was done to mimic the process that functionality built directly into the XBRL taxonomy could perform. If any XBRL element required human subjectivity to calculate, it was not manually populated, and none of the balances tagged by a preparer were altered in any way. As illustrated previously, consider the example of a company choosing to display its liabilities section on its balance sheet of the paper/PDF 10-K as follows:

Current liabilities:	
Accounts payable	$ 100,000
Other current liabilities	25,000
Total current liabilities	125,000
Long-term debt	75,000

According the SEC's protocol for preparation of XBRL company filings, the preparer of this company's XBRL 10-K filing should not tag an amount for total liabilities because it does not explicitly appear on the face of the paper/PDF 10-K, which is purely a formatting issue. In order to make the XBRL company filing fully populated, the XBRL tag for total liabilities would be filled in with the missing $200,000. If the company had erroneously tagged long-term debt with anything other than a positive 75,000, that tag would not be filled in for the analysis because this would be an error created by the preparer, not by the SEC's protocol for the preparation of XBRL company filings.

This process was not subjective on the part of the researcher, as missing amounts were filled in only if:

1. The necessary component information was available.
    - For example, if current liabilities and long-term liabilities were given, then total liabilities could be determined by adding them together.
2. The preparer did not fill in a tag for that balance—and should not have, according to the SEC.
    - This would be the case if the balance was not explicitly listed on the paper 10-K.

The data set created from this process is referred to as "fully populated XBRL filings."
The proportion of accounting concepts exactly matching the 70 identified concepts that were found in the fully populated XBRL company filings were calculated.

H_2 states that earnings prediction models created using fully populated XBRL company filing data will not predict earnings with a different accuracy as earnings prediction models created using Compustat data. The SEC requires public companies to file financial statements in XBRL format so that users can access company data easily. Compustat has commonly been used as a data source for users to access company data. Therefore, it is worthwhile to explore differences in these data sources—in particular, the ability of XBRL company data and Compustat data to predict future earnings.

Two earnings prediction models were used in the current study. The variables included in the first earnings prediction model (hereafter, the "O&P Earnings Prediction Model") were those determined by Ou and Penman (1989a, 1989b) to be optimal in predicting future earnings, with some removed due to a lack of observations and high correlations. The second earnings prediction model (hereafter, the "A&B Earnings Prediction Model") included the variables used by Abarbanell and Bushee (1997, 1998) in their earnings prediction model, with some variables excluded due to a lack of observations.

Ou and Penman's (1989a, 1989b) final earnings prediction model included 26 variables (25 independent variables and one dependent variable). The way in which they determined these 26 variables was through a statistical approach. Ou and Penman calculated a summary measure that predicted the direction of future earnings for each company in their sample. This summary measure, denoted $\widehat{Pr}$, was calculated based on 68 ratios. Ou and Penman then calculated univariate logistic regression estimations on each of the 68 ratios from the estimation period to determine which variables predicted future earnings changes. Those variables that exhibited predictive ability of one-year ahead earnings changes in the univariate logistic regression were then included in a multivariate model. Any variables not significant in the multivariate model were removed. Finally, each remaining variable was estimated step-wise, retaining only the variables still significant. These variables and their associated weights calculated during the estimation period were used to calculate $\widehat{Pr}$ for every company in their sample during the prediction period as follows:

$$\widehat{Pr}_{it} = \left[1 + \exp(-\hat{\theta}' \mathrm{X}_{it})\right]^{-1}$$

where X_{it} is the set of ratios for firm i in fiscal year t and $\hat{\theta}$ is the set of estimated coefficient weights.

The O&P Earnings Prediction Model created in this study includes 12 of Ou and Penman's (1989a, 1989b) variables (11 independent variables and one dependent variable). Two of the variables included in Ou and Penman's (1989a, 1989b) earnings prediction model were excluded because there were no observations in the sample, and six additional variables were excluded due to very few observations in the sample. Six more variables were then removed because of high correlations with other variables in the model. The final 12 variables used in the O&P Earnings Prediction Model are listed in Table 3. The exact calculations for each of the variables, including the Compustat data items and XBRL tags that were used in this study, are listed in Appendix A.

Table 3

Variables Required to Create the O&P Earnings Prediction Model

1	% Δ in current ratio
2	Inventory / total assets
3	% Δ in inventory
4	% Δ in sales
5	Δ in dividend per share
6	% Δ in (capital expenditures / total assets)
7	% Δ in debt-equity ratio
8	Return on closing equity
9	% Δ in total assets
10	Working capital / total assets
11	Cash dividend as % of cash flows
12	Δ in earnings per share - drift

The final 12 variables used in the O&P Earnings Prediction Model.

Abarbanell and Bushee's (1997, 1998) final earnings prediction model included 10 variables (nine independent variables and one dependent variable). These variables were originally identified by Lev and Thiagarajan (1993) using a guided search. Lev and Thiagarajan

(1993) identified the ratios used most by investors to assess the quality and growth of earnings by searching *The Wall Street Journal*, *Barron's*, Value Line publications, newsletters of major securities firms, and professional commentaries on corporate financial reporting and analysis. They determined that these variables were related to inventory; accounts receivable; capital expenditures; research and development; gross margin; selling and administrative; provision for doubtful receivables; effective tax rate; order backlog; labor force; last-in, first-out earnings; and audit qualification. However, Lev and Thiagarajan (1993) eliminated the ratios related to order backlog, provision for doubtful receivables, and research and development in their final sample in order to substantially increase the sample size and make the sample more representative. Abarbanell and Bushee (1997, 1998) also used just the final nine ratios in their studies. Each of these ratios was "specifically motivated by arguments for why these signals would be expected, *a priori*, to be related to future earnings changes" (Abarbanell and Bushee, 1998, p. 22).

The current study used six of these 10 variables (five independent variables and one dependent variable), eliminating the audit qualification variable because it was not available in XBRL company filings, the effective tax rate variable because there were no observations in the sample, and the accounts receivable and labor force variables because there were very few observations in the sample. The variables used in the A&B Earnings Prediction Model are listed in Table 4. The exact calculations for each of the variables, including the Compustat data items and XBRL tags that were used, are listed in Appendix A.

Table 4

Variables Required to Create the A&B Earnings Prediction Model

1	% Δ in inventory - % Δ in sales
2	% Δ in industry capital expenditures - % Δ in firm capital expenditures
3	% Δ in sales - % Δ in gross margin
4	% Δ in selling and administrative expenses - % Δ in sales
5	0 for LIFO, 1 for FIFO or other
6	Δ in earnings per share

The final six variables used in the A&B Earnings Prediction Model.

To test H_2, the fully populated XBRL filing data set created for RQ_2 was used to create the two earnings prediction models. These same models were then created using data from the Compustat database. The models were created by conducting a multiple regression to estimate annual earnings based on Compustat data and then on fully populated XBRL company data for the year 2009. This was done using the variables for the O&P Earnings Prediction Model as well as the A&B Earnings Prediction Model. Four models in all were created (O&P Earnings Prediction Model using fully populated XBRL company data, O&P Earnings Prediction Model using Compustat data, A&B Earnings Prediction Model using fully populated XBRL company data, and A&B Earnings Prediction Model using Compustat data). After the models were created, the coefficients calculated during the estimation period (2009) were used to predict future earnings during the prediction period (2011) for each company in the sample. For each observation, the squared residual was computed by squaring the difference between the actual earnings value and the regression-based prediction. A comparison of the mean of the two sets of squared residuals for each model was conducted using a paired t-test. A level of significance of 0.05 was used in the paired t-test. There would be a significant difference in the squared residuals if the *p* value of the paired t-test result was less than or equal to the level of significance value of 0.05. If they were significantly different, then this result would suggest that one of the sets of data (XBRL or Compustat) had a lower mean squared error and thus a higher accuracy than the other set of data. On the other hand, a non-significant difference would indicate that neither set of data was more accurate at predicting future earnings than the other set of data.

VI. RESULTS

An attempt was made to collect XBRL financial statement data for the 70 accounting concepts that are required to calculate the 46 variables needed for the two earnings prediction models utilized in this study in order to determine if a sufficient number of required accounting concepts were available from current XBRL company filings. In addition, the comparability of information in the XBRL filings and the audited 10-K reports was explored. This was accomplished by comparing explicitly listed accounting concepts required in this study in the companies' audited 10-K reports and the tagged information in their XBRL filings. There was also an attempt made to "fully populate" the current XBRL filings. In other words, for any accounting concepts required for this study that were not tagged in the XBRL filings, a structured attempt based on a specific set of rules was made to calculate and populate those accounting concepts based on other accounting concepts tagged within the same XBRL filings. This was done in order to mimic a process that the XBRL taxonomy itself could perform and to evaluate the increased usability of an XBRL filing that contains a more complete set of data. Finally, the two earnings prediction models were created using both Compustat data and fully populated XBRL company filing data. A comparison was made as to which data set could predict future earnings changes more accurately.

RQ_1 poses the question: "What proportion of the accounting elements needed to create the earnings prediction models is tagged in current XBRL company filings?" Seventy XBRL accounting concepts were required to calculate the 46 variables needed for the two earnings prediction models utilized in this study.

To evaluate RQ_1, the number of explicitly tagged XBRL accounting concepts required to compute the 46 variables found in the current XBRL company filings were counted and the proportion of accounting concepts exactly matching the 70 identified concepts was calculated. Table 5 contains the results of RQ_1.

Table 5

Results of RQ₁: Proportion of Data Complete in Current XBRL Company Filings

	2012			**2011**			**2010**		
		95% Confidence Interval			**95% Confidence Interval**			**95% Confidence Interval**	
XBRL Element Names	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**
Assets				0.993	0.976	0.999	0.993	0.976	0.999
AssetsCurrent				0.983	0.961	0.994	0.983	0.961	0.994
CashAndCashEquivalentsAtCarryingValue				0.986	0.966	0.996			
CommonStockDividendsPerShareDeclared				0.581	0.523	0.638	0.578	0.519	0.635
CostOfRevenue				0.182	0.140	0.231	0.193	0.149	0.242
GrossProfit				0.412	0.356	0.471	0.416	0.359	0.474
IncomeLossFromContinuingOperationsPerBasicShare	0.476	0.418	0.535	0.497	0.438	0.555	0.493	0.435	0.552
InventoryNet				0.774	0.722	0.820	0.774	0.722	0.820
InventoryPolicyTextBlock				0.764	0.711	0.811			
Liabilities				0.561	0.502	0.618	0.574	0.516	0.631
LiabilitiesCurrent				0.983	0.961	0.994	0.980	0.956	0.993
NetCashProvidedByUsedInOperatingActivities				0.882	0.839	0.916			
OperatingIncomeLoss				0.858	0.813	0.896			
PaymentsOfDividends				0.334	0.281	0.391			
PaymentsToAcquirePropertyPlantAndEquipment				0.743	0.690	0.792	0.740	0.686	0.789
ProfitLoss				0.672	0.616	0.725			
SalesRevenueNet				0.486	0.428	0.545	0.490	0.432	0.548
SellingGeneralAndAdministrativeExpense				0.639	0.581	0.693	0.645	0.588	0.700
StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest				0.716	0.661	0.767	0.747	0.693	0.795

Proportions and 95% confidence intervals of accounting elements needed for the O&P and A&B earnings prediction models found in current XBRL company filings of 296 companies.
Note. Some cells are blank because those variables were not needed for certain years to create the earnings prediction models.

Table 5 continued

Results of RQ_1: Proportion of Data Complete in Current XBRL Company Filings

	2009			2008			2007		
		95% Confidence Interval			95% Confidence Interval			95% Confidence Interval	
XBRL Element Names	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**
Assets	0.993	0.976	0.999	0.983	0.961	0.994			
AssetsCurrent	0.983	0.961	0.994	0.983	0.961	0.994			
CashAndCashEquivalentsAtCarryingValue	0.990	0.971	0.998						
CommonStockDividendsPerShareDeclared	0.497	0.438	0.555	0.476	0.418	0.535			
CostOfRevenue	0.189	0.146	0.239	0.193	0.149	0.242			
GrossProfit	0.426	0.369	0.484	0.365	0.310	0.423	0.341	0.287	0.398
IncomeLossFromContinuingOperationsPerBasicShare	0.507	0.448	0.565	0.470	0.412	0.528	0.419	0.362	0.477
InventoryNet	0.770	0.718	0.817	0.743	0.690	0.792			
InventoryPolicyTextBlock	0.044	0.024	0.074						
Liabilities	0.561	0.502	0.618	0.534	0.475	0.592			
LiabilitiesCurrent	0.983	0.961	0.994	0.980	0.956	0.993			
NetCashProvidedByUsedInOperatingActivities	0.929	0.894	0.956						
OperatingIncomeLoss	0.855	0.809	0.893						
PaymentsOfDividends	0.361	0.307	0.419						
PaymentsToAcquirePropertyPlantAndEquipment	0.750	0.697	0.798	0.743	0.690	0.792	0.716	0.661	0.767
ProfitLoss	0.689	0.633	0.741						
SalesRevenueNet	0.470	0.412	0.528	0.446	0.388	0.505	0.412	0.356	0.471
SellingGeneralAndAdministrativeExpense	0.642	0.584	0.697	0.628	0.571	0.684	0.611	0.553	0.667
StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest	0.757	0.704	0.805	0.764	0.711	0.811			

Proportions and 95% confidence intervals of accounting elements needed for the O&P and A&B earnings prediction models found in current XBRL company filings of 296 companies.
Note. Some cells are blank because those variables were not needed for certain years to create the earnings prediction models.

Table 5 illustrates that 23 of the 70 accounting concepts had proportions of less than 0.50 complete data in the current XBRL company filings for this sample. This suggests that current XBRL company filings cannot be used to interactively capture the accounting elements necessary to calculate the ratios required to create earnings prediction models in this study.

H_1 states that the number of accounting concepts tagged appropriately in XBRL filings will not be significantly different from the number of explicitly reported accounting concepts in companies' audited 10-K reports for the 70 accounting concepts required to calculate the variables needed for the earnings prediction models. The results of the paired-samples t-test of H_1 are listed in Table 6.

Table 6

Results of H_1: Comparison of Explicitly Reported Accounting Concepts in the Audited 10-K Reports to Explicitly Tagged Accounting Concepts in the Company XBRL Filings

	2012			2011			2010		
XBRL Element Names	***t***	***df***	***p* (two-tailed)**	***t***	***df***	***p* (two-tailed)**	***t***	***df***	***p* (two-tailed)**
Assets				1.000	49	0.322	NA	NA	NA
AssetsCurrent				1.000	49	0.322	NA	NA	NA
CashAndCashEquivalentsAtCarryingValue				-1.000	49	0.322			
CommonStockDividendsPerShareDeclared				1.288	49	0.204	-3.280	49	0.002**
CostOfRevenue				-5.755	49	< 0.001**	-4.950	49	< 0.001**
GrossProfit				1.769	49	0.083	1.769	49	0.083
IncomeLossFromContinuingOperationsPerBasicShare	-5.715	49	< 0.001**	-5.715	49	< 0.001**	-5.715	49	< 0.001**
InventoryNet				-1.429	49	0.159	-1.000	49	0.322
InventoryPolicyTextBlock				-0.629	49	0.533			
Liabilities				1.769	49	0.083	2.064	49	0.044*
LiabilitiesCurrent				NA	NA	NA	NA	NA	NA
NetCashProvidedByUsedInOperatingActivities				-2.585	49	0.013*			
OperatingIncomeLoss				1.769	49	0.083			
PaymentsOfDividends				-5.715	49	< 0.001**			
PaymentsToAcquirePropertyPlantAndEquipment				-2.447	49	0.018*	-1.950	49	0.057
ProfitLoss				-5.480	49	< 0.001**			
SalesRevenueNet				-3.562	49	0.001**	-2.682	49	0.010**
SellingGeneralAndAdministrativeExpense				NA	NA	NA	NA	NA	NA
StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest				-4.149	49	<0.001**	-4.149	49	< 0.001**

Note. A significant result would indicate that there is a significant difference in the number of explicitly reported accounting concepts in the audited 10-K reports compared to the number of explicitly tagged accounting concepts in the current XBRL company filings. Some cells are blank because those variables were not needed for certain years to create the earnings prediction models.
N/A - The *t* could not be computed because the standard error of the difference is zero.
* *p* value is significant at the 0.05 level (two-tailed).
** *p* value is significant at the 0.01 level (two-tailed).

Table 6 continued

Results of H_1: Comparison of Explicitly Reported Accounting Concepts in the Audited 10-K Reports to Explicitly Tagged Accounting Concepts in the Current XBRL Company Filings

	2009			2008			2007		
XBRL Element Names	***t***	***df***	***p* (two-tailed)**	***t***	***df***	***p* (two-tailed)**	***t***	***df***	***p* (two-tailed)**
Assets	NA	NA	NA	NA	NA	NA			
AssetsCurrent	NA	NA	NA	NA	NA	NA			
CashAndCashEquivalentsAtCarryingValue	-1.000	49	0.322						
CommonStockDividendsPerShareDeclared	-3.280	49	0.002**	-2.189	49	0.033*			
CostOfRevenue	-6.461	49	< 0.001**	-6.205	49	< 0.001**			
GrossProfit	2.064	49	0.044*	1.429	49	0.159	NA	NA	NA
IncomeLossFromContinuingOperationsPerBasicShare	-5.715	49	< 0.001**	-5.480	49	< 0.001**	-5.715	49	< 0.001**
InventoryNet	-1.000	49	0.322	-0.573	49	0.569			
InventoryPolicyTextBlock	-14.941	49	< 0.001**						
Liabilities	2.064	49	0.044*	1.429	49	0.159			
LiabilitiesCurrent	NA	NA	NA	NA	NA	NA			
NetCashProvidedByUsedInOperatingActivities	-2.333	49	0.024*						
OperatingIncomeLoss	NA	NA	NA						
PaymentsOfDividends	-5.480	49	< 0.001**						
PaymentsToAcquirePropertyPlantAndEquipment	-2.824	49	0.007**	-2.447	49	0.018*	-3.500	49	0.001**
ProfitLoss	-5.024	49	< 0.001**						
SalesRevenueNet	-3.562	49	0.001**	-3.988	49	< 0.001**	-4.365	49	< 0.001**
SellingGeneralAndAdministrativeExpense	NA	NA	NA	-1.000	49	0.322	NA	NA	NA
StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest	-4.149	49	< 0.001**	-4.149	49	< 0.001**			

Note. A significant result would indicate that there is a significant difference in the number of explicitly reported accounting concepts in the audited 10-K reports compared to the number of explicitly tagged accounting concepts in the current XBRL company filings. Some cells are blank because those variables were not needed for certain years to create the earnings prediction models.

N/A - The *t* could not be computed because the standard error of the difference is zero

* *p* value is significant at the 0.05 level (2-tailed)

** *p* value is significant at the 0.01 level (2-tailed)

These results indicate a significant difference in the number of concepts explicitly reported in XBRL filings compared to the number of concepts explicitly reported in audited 10-K reports for 36 of the 70 accounting concepts. These 36 accounting concepts are components of 31 of the 46 variables, which means that 67% of the 46 variables would be incalculable or would return erroneous results.

An example of this difference is that a company may have total revenues listed in its audited 10-K report but have revenues tagged in its XBRL filing. The definition for the tag SalesRevenueNet, the accounting concept required for the earnings prediction models, in the XBRL taxonomy is: “Total revenue from sale of goods and services rendered during the reporting period, in the normal course of business, reduced by sales returns and allowances, and sales discounts” (Financial Accounting Standards Board, 2012). The definition for the tag Revenues in the XBRL taxonomy is: “Aggregate revenue recognized during the period (derived from goods sold, services rendered, insurance premiums, or other activities that constitute an entity’s earning process). For financial services companies, also includes investment and interest income, and sales and trading gains.” (Financial Accounting Standards Board, 2012) Therefore, the tag Revenues could be used for a nonfinancial company and/or a financial company, whereas the tag SalesRevenueNet should only be used for a nonfinancial company. Although using the tag Revenues is a perfectly acceptable way for companies to prepare their XBRL filings, this example illustrates the hurdles present for investors who wish to automate their analysis activities using XBRL filings.

RQ_2 poses the question: “What proportion of the accounting elements needed to create the earnings prediction models is tagged in fully populated XBRL company filings?” Fully populated XBRL company filings are XBRL filings whose missing tags have been automatically populated based on component XBRL tags. This is a function that could be incorporated into the XBRL taxonomy, potentially making XBRL financial statement data more useful by providing a more complete collection of accounting concepts that could be extracted without manual calculation. Although not all missing tags could be automatically populated, fully populated XBRL filings were expected to have improved usability compared to the current XBRL filings.

To evaluate RQ_2, the number of XBRL accounting concepts required to compute the 46 variables found in the fully populated XBRL company filings were counted and the proportion of accounting concepts exactly matching the 70 identified concepts were calculated. Table 7 contains the results of RQ_2.

Table 7

Results of RQ2: Proportion of Data Complete in Fully Populated XBRL Company Filings

	2012			2011			2010		
		95% Confidence Interval			95% Confidence Interval			95% Confidence Interval	
XBRL Element Names	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**
Assets				1.000	0.929	1.000	1.000	0.929	1.000
AssetsCurrent				1.000	0.929	1.000	1.000	0.929	1.000
CashAndCashEquivalentsAtCarryingValue				0.980	0.894	0.999			
CommonStockDividendsPerShareDeclared				1.000	0.929	1.000	1.000	0.929	1.000
CostOfRevenue				0.880	0.757	0.955	0.880	0.757	0.955
GrossProfit				0.860	0.733	0.942	0.860	0.733	0.942
IncomeLossFromContinuingOperationsPerBasicShare	0.940	0.835	0.987	0.940	0.835	0.987	0.940	0.835	0.987
InventoryNet				1.000	0.929	1.000	1.000	0.929	1.000
InventoryPolicyTextBlock				1.000	0.929	1.000			
Liabilities				1.000	0.929	1.000	1.000	0.929	1.000
LiabilitiesCurrent				1.000	0.929	1.000	1.000	0.929	1.000
NetCashProvidedByUsedInOperatingActivities				1.000	0.929	1.000			
OperatingIncomeLoss				0.900	0.782	0.967			
PaymentsOfDividends				1.000	0.929	1.000			
PaymentsToAcquirePropertyPlantAndEquipment				0.940	0.835	0.987	0.940	0.835	0.987
ProfitLoss				1.000	0.929	1.000			
SalesRevenueNet				1.000	0.929	1.000	1.000	0.929	1.000
SellingGeneralAndAdministrativeExpense				0.800	0.663	0.900	0.800	0.663	0.900
StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest				1.000	0.929	1.000	1.000	0.929	1.000

Proportions and 95% confidence intervals of accounting elements needed for the O&P and A&B earnings prediction models found in fully populated XBRL company filings of 50 companies.
Note. Some cells are blank because those variables were not needed for certain years to create the earnings prediction models.

Table 7 continued

Results of RQ_2: Proportion of Data Complete in Fully Populated XBRL Company Filings

	2009			2008			2007		
		95% Confidence Interval			95% Confidence Interval			95% Confidence Interval	
XBRL Element Names	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**	**Proportion**	**Lower**	**Upper**
Assets	1.000	0.929	1.000	1.000	0.929	1.000			
AssetsCurrent	1.000	0.929	1.000	1.000	0.929	1.000			
CashAndCashEquivalentsAtCarryingValue	0.980	0.894	0.999						
CommonStockDividendsPerShareDeclared	1.000	0.929	1.000	0.920	0.808	0.978			
CostOfRevenue	0.880	0.757	0.955	0.880	0.757	0.955			
GrossProfit	0.860	0.733	0.942	0.860	0.733	0.942	0.860	0.733	0.942
IncomeLossFromContinuingOperationsPerBasicShare	0.940	0.835	0.987	0.940	0.835	0.987	0.940	0.835	0.987
InventoryNet	1.000	0.929	1.000	1.000	0.929	1.000			
InventoryPolicyTextBlock	1.000	0.929	1.000						
Liabilities	1.000	0.929	1.000	1.000	0.929	1.000			
LiabilitiesCurrent	1.000	0.929	1.000	1.000	0.929	1.000			
NetCashProvidedByUsedInOperatingActivities	1.000	0.929	1.000						
OperatingIncomeLoss	0.900	0.782	0.967						
PaymentsOfDividends	1.000	0.929	1.000						
PaymentsToAcquirePropertyPlantAndEquipment	0.960	0.863	0.995	0.960	0.863	0.995	0.940	0.835	0.987
ProfitLoss	1.000	0.929	1.000						
SalesRevenueNet	1.000	0.929	1.000	1.000	0.929	1.000	1.000	0.929	1.000
SellingGeneralAndAdministrativeExpense	0.800	0.663	0.900	0.800	0.663	0.900	0.800	0.663	0.900
StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest	1.000	0.929	1.000	1.000	0.929	1.000			

Proportions and 95% confidence intervals of accounting elements needed for the O&P and A&B earnings prediction models found in fully populated XBRL company filings of 50 companies.
Note. Some cells are blank because those variables were not needed for certain years to create the earnings prediction models.

Table 7 illustrates that none of the 70 accounting concepts had proportions of less than 0.50 complete data in the fully populated XBRL company filings for this sample. The lowest proportion of data was 0.80, which was for selling, general, and administrative expense. Although not all of the accounting concepts required for the two earnings prediction models could have been automatically calculated by software, the fully populated XBRL filings provides more information to users than the current XBRL filings.

H_2 states that earnings prediction models created using fully populated XBRL company filing data will not predict earnings with a different accuracy as earnings prediction models created using Compustat data. Although fully populating the XBRL company filings eliminated many of the deficiencies inherent in current XBRL company filings, it is believed that the values reported in Compustat have been changed (standardized) enough to be significantly different from the values reported in the XBRL company filings. It is unclear, based on the previous literature, if this standardization of financial statement information improves or deteriorates the data's usefulness. This portion of the study investigates the accuracy to which XBRL company filing data can predict future earnings compared to that of Compustat data.

To test H_2, the fully populated XBRL filing data set created for RQ_2 was used to create the two earnings prediction models. These same models were then created using data from the Compustat database. The models were created by conducting a multiple regression to estimate annual earnings based on Compustat data and then based on fully populated XBRL company data for the year 2009. This was done using the variables for the O&P Earnings Prediction Model as well as the A&B Earnings Prediction Model. Four models in all were created (O&P Earnings Prediction Model using fully populated XBRL company data, O&P Earnings Prediction Model using Compustat data, A&B Earnings Prediction Model using fully populated XBRL company data, and A&B Earnings Prediction Model using Compustat data). After the models were created, the coefficients calculated during the estimation period (2009) were used to predict future earnings during the prediction period (2011) for each company in the sample. For each observation, the squared residual was computed by squaring the difference between the actual earnings value and the regression-based prediction. A comparison of the mean of the two sets of squared residuals for each model was conducted using a paired t-test. A level of significance of 0.05 was used in the paired t-test. There would be a significant difference in the squared residuals if the *p* value of the paired t-test result was less than or equal to the level of significance value of 0.05. If they were significantly different, then this result would suggest that one of the sets of data (XBRL or Compustat) had a lower mean squared error and thus a higher accuracy than the other set of data. On the other hand, a non-significant difference would indicate that neither set of data was more accurate at predicting future earnings than the other set of data. This process, along with the associated results, are described in more detail below. Most important to note, however, is that it was possible to create these earnings prediction models with fully populated XBRL company filing data, while it was not possible to do so using current XBRL company filing data.

O&P Earnings Prediction Model

A multiple regression was conducted to estimate annual earnings using the fully populated XBRL filing data based on the year 2009 (the estimation period) for the O&P Earnings Prediction Model. The original earnings prediction model created by Ou and Penman (1989a, 1989b) included 25 predictor variables. Two of these variables were excluded because of no observations in the XBRL company filings and because of no observations for one of the variables and only five observations for the other variable in the Compustat database for the sample in the study. An additional variable was excluded because it required 2007 equity, which was not available in XBRL company filings. However, when running the multiple regression for these 22 variables with the fully populated XBRL company filing data, SPSS only included 16 out of the 22 predictors in the regression model. The SPSS statistical software was unable to run the regression model if all of the 22 predictors were included in the regression model. Six predictors were removed because of too much missing data. They included:

- % Δ in (capital expenditures / total assets), one-year lag
- % Δ in sales / total assets
- Return on total assets
- % Δ in (pretax income / sales)
- Cash flow to total debt
- Repayment of LT debt as % of total LT debt

An additional five variables were removed due to high correlations between independent variables found in either the fully populated XBRL company filing data set or the Compustat data set. The six variables removed were:

- % Δ in inventory turnover
- % Δ in (inventory / total assets)
- Gross margin ratio
- Operating income / total assets
- Sales to total cash

The remaining 11 independent variables included:

- % Δ in current ratio

- Inventory / total assets
- % Δ in inventory
- % Δ in sales
- Δ in dividend per share
- % Δ in debt-equity ratio
- % Δ in (capital expenditures / total assets)
- Return on closing equity
- % Δ in total assets
- Working capital / total assets
- Cash dividend as % of cash flows

Table 8 reports summary statistics for the variables used in the O&P Earnings Prediction Model using fully populated XBRL company filing data. Table 9 presents the Pearson/Spearman correlation coefficients. The correlation coefficients measure the strength and direction of the relationship between variables. Table 10 summarizes the results of the regression.

Table 8

Summary Statistics for the Variables Used in the O&P Earnings Prediction Model Based on the Fully Populated XBRL Company Filing Data for the Year 2009

Variable	**N**	**Minimum**	**25%**	**Median**	**Mean**	**75%**	**Maximum**	**Standard Deviation**
% Δ in current ratio	49	-0.118	0.026	0.137	0.186	0.270	0.803	0.221
Inventory / total assets	46	0.003	0.048	0.085	0.097	0.140	0.235	0.063
% Δ in inventory	46	-0.274	-0.151	-0.087	-0.039	0.062	0.391	0.169
% Δ in sales	50	-0.364	-0.174	-0.054	-0.070	0.024	0.364	0.144
Δ in dividend per share	46	-0.762	0.000	0.000	-0.030	0.060	0.169	0.209
% Δ in debt-equity ratio	50	-0.704	-0.297	-0.138	-0.143	-0.003	0.330	0.225
% Δ in (capital expenditures / total assets)	47	-0.658	-0.448	-0.317	-0.280	-0.174	0.350	0.247
Return on closing equity	50	-0.013	0.087	0.129	0.172	0.218	0.656	0.153
% Δ in total assets	50	-0.069	0.011	0.056	0.092	0.140	0.638	0.148
Working capital / total assets	50	-0.009	0.044	0.134	0.165	0.259	0.483	0.134
Cash dividend as % of cash flows	41	0.024	0.126	0.218	0.235	0.322	0.563	0.142
EPSt$_{+1}$ - EPSt - drift$_{t+1}$	47	-2.060	-0.040	0.580	1.003	2.045	4.275	1.675

Summary statistics for the 2009 variables used in the O&P Earnings Prediction Model using fully populated XBRL company filing data.

72

Table 9

Pearson/Spearman Correlation Coefficients for the O&P Earnings Prediction Model Based on Fully Populated XBRL Company Filing Data for the Year 2009

	% Δ in current ratio	Inventory/ total assets	% Δ in inventory	% Δ in sales	Δ in dividends per share	% Δ in debt equity ratio	% Δ in capital expenditures/ total assets	Return on closing equity	% Δ in total assets	Working capital / total assets	Cash dividend as % of cash flows
% Δ in current ratio	1.000 ---	-0.137	-0.051	0.002	-0.112	-0.208	-0.137	0.116	-0.129	-0.211	0.320
Inventory / total assets	-0.028	1.000 ---	0.029	0.129	0.028	-0.193	0.026	0.182	0.001	0.353	-0.297
% Δ in inventory	0.014	0.137	1.000 ---	0.410	-0.027	-0.066	0.200	0.075	0.462	-0.172	0.325
% Δ in sales	0.055	0.156	0.435	1.000 ---	0.158	-0.104	0.129	0.240	0.316	0.114	0.114
Δ in dividends per share	-0.163	0.080	0.030	-0.005	1.000 ---	0.346	0.040	0.168	-0.111	0.090	-0.034
% Δ in debt equity ratio	-0.321	-0.350	-0.147	-0.185	0.162	1.000 ---	0.070	0.137	-0.043	0.173	-0.125
% Δ in capital expenditures / total assets	-0.119	-0.068	0.198	0.079	0.278	0.101	1.000 ---	0.186	-0.340	0.073	0.132
Return on closing equity	-0.059	0.332	0.095	0.269	0.292	0.011	0.250	1.000 ---	0.043	0.024	0.143
% Δ in total assets	-0.195	0.054	0.404	0.410	-0.136	0.234	-0.174	0.219	1.000 ---	0.064	0.097
Working capital / total assets	-0.100	0.489	-0.143	0.115	0.002	0.086	0.078	0.342	0.169	1.000 ---	-0.179
Cash dividend as % of cash flows	0.284	-0.244	0.138	0.085	0.059	-0.054	0.191	0.050	-0.157	-0.174	1.000 ---

Pearson (Spearman) correlation coefficients are presented above (below) the diagonal.

73

Table 10

Results of the Multiple Regression for the O&P Earnings Prediction Model Based on the Fully Populated XBRL Company Filing Data for the Year 2009

	B	***t***	***p***
Constant	1.398	1.384	0.182
% Δ in current ratio	1.859	0.733	0.472
Inventory / total assets	10.923	1.877	0.076
% Δ in inventory	-0.359	-0.141	0.889
% Δ in sales	-3.201	-1.281	0.215
Δ in dividend per share	-1.083	-0.827	0.418
% Δ in debt-equity ratio	4.827	2.890	0.009
% Δ in (capital expenditures / total assets)	0.125	0.087	0.931
Return on closing equity	-4.566	-1.450	0.164
% Δ in total assets	0.364	0.111	0.913
Working capital / total assets	-5.302	-1.703	0.105
Cash dividend as % of cash flows	0.839	0.327	0.747

Regression results estimating annual earnings for the O&P Earnings Prediction Model using fully populated XBRL company filing data for the year 2009.
Note. N = 31; Adjusted R^2 = 0.229.

A multiple regression was conducted to estimate annual earnings using the Compustat data based on the year 2009 (the estimation period) for the O&P Earnings Prediction Model. The same 11 independent variables and one dependent variable were included in this regression as in the one described above using fully populated XBRL company filing data. Table 11 reports summary statistics for the variables used in the O&P Earnings Prediction Model using Compustat data. Table 12 presents the Pearson/Spearman correlation coefficients, and Table 13 summarizes the results of the regression.

Table 11

Summary Statistics for the Variables Used in the O&P Earnings Prediction Model Based on the Compustat Data for the Year 2009

Variable	**N**	**Minimum**	**25%**	**Median**	**Mean**	**75%**	**Maximum**	**Standard Deviation**
% Δ in current ratio	50	-0.114	0.026	0.136	0.183	0.283	0.773	0.217
Inventory / total assets	45	0.005	0.050	0.085	0.100	0.140	0.236	0.064
% Δ in inventory	45	-0.264	-0.132	-0.090	-0.037	0.065	0.313	0.153
% Δ in sales	50	-0.319	-0.202	-0.057	-0.081	0.023	0.134	0.132
Δ in dividend per share	50	-0.592	0.000	0.028	0.028	0.086	0.393	0.177
% Δ in debt-equity ratio	33	-0.489	-0.327	-0.136	-0.123	0.006	0.729	0.251
% Δ in (capital expenditures / total assets)	50	-0.680	-0.441	-0.310	-0.286	-0.176	0.338	0.231
Return on closing equity	50	0.010	0.088	0.126	0.181	0.238	0.641	0.159
% Δ in total assets	50	-0.069	0.011	0.057	0.092	0.140	0.636	0.148
Working capital / total assets	50	-0.011	0.044	0.134	0.165	0.259	0.483	0.134
Cash dividend as % of cash flows	42	0.024	0.128	0.217	0.232	0.318	0.559	0.139
EPS_{t+1} - EPS_t - $drift_{t+1}$	50	-1.225	-0.047	0.578	1.019	1.881	4.249	1.494

Summary statistics for the 2009 variables used in the O&P Earnings Prediction Model using Compustat data.

Table 12

Pearson/Spearman Correlation Coefficients for the O&P Earnings Prediction Model Based on the Compustat Data for the Year 2009

	% Δ in Current ratio	Inventory / total assets	% Δ in inventory	% Δ in sales	Δ in dividends per share	% Δ in debt equity ratio	% Δ in capital expenditures / total assets	Return on closing equity	% Δ in total assets	Working capital / total assets	Cash dividend as % of cash flows
% Δ in current ratio	1.000 ---	-0.120	-0.101	0.072	-0.102	-0.100	-0.171	0.197	-0.142	-0.226	0.306
Inventory / Total assets	-0.018	1.000 ---	0.053	0.243	0.099	0.095	0.120	0.221	-0.029	0.337	-0.275
% Δ in inventory	-0.070	0.135	1.000 ---	0.221	0.018	-0.025	0.215	0.032	0.447	-0.198	0.242
% Δ in sales	0.075	0.247	0.295	1.000 ---	0.120	-0.053	0.136	0.374	0.342	0.200	0.218
Δ in dividends per share	0.032	0.077	0.101	0.052	1.000 ---	0.031	0.223	0.153	-0.168	0.113	0.004
% Δ in debt equity ratio	-0.224	-0.213	-0.118	-0.216	-0.141	1.000 ---	0.098	0.083	0.048	0.175	-0.100
% Δ in capital expenditures / total assets	-0.162	-0.057	0.195	0.092	0.231	0.158	1.000 ---	0.152	-0.257	0.106	0.096
Return on closing equity	0.048	0.386	0.081	0.434	0.274	-0.090	0.217	1.000 ---	0.004	0.015	0.166
% Δ in total assets	-0.220	0.018	0.384	0.398	-0.078	0.391	-0.027	0.196	1.000 ---	0.063	0.116
Working capital / total assets	-0.111	0.467	-0.199	0.195	0.043	0.173	0.077	0.360	0.159	1.000 ---	-0.156
Cash dividend as % of cash flows	0.279	-0.247	0.071	0.163	0.260	-0.050	0.149	0.118	-0.123	-0.155	1.000 ---

Pearson (Spearman) correlation coefficients are presented above (below) the diagonal.

Table 13

Results of the Multiple Regression for the O&P Earnings Prediction Model Based on the Compustat Data for the Year 2009

	B	***t***	***p***
Constant	0.890	0.873	0.396
% Δ in current ratio	0.094	0.079	0.938
Inventory / total assets	9.072	1.827	0.088
% Δ in inventory	-0.437	-0.146	0.886
% Δ in sales	-4.363	-1.878	0.080
Δ in dividend per share	-1.435	-1.227	0.239
% Δ in debt-equity ratio	3.520	3.940	0.001
% Δ in (capital expenditures / total assets)	-0.693	-0.555	0.587
Return on closing equity	-2.849	-1.958	0.069
% Δ in total assets	-4.477	-1.222	0.241
Working capital / total assets	-4.766	-1.800	0.092
Cash dividend as % of cash flows	4.546	1.649	0.120

Regression results estimating annual earnings for the O&P Earnings Prediction Model using Compustat data for the year 2009.
Note. $N = 27$; Adjusted $R^2 = 0.641$.

The coefficients calculated for each of the 11 variables and the one constant (using fully populated XBRL data and Compustat data) were used to create the O&P earnings prediction models.

O&P Earnings Prediction Model Using Fully Populated XBRL and Compustat Data

Δ in earnings per share = α

+ β_1* % Δ in current ratio

+ β_2 * Inventory / total assets

+ β_3 * % Δ in inventory

+ β_4 * % Δ in sales

+ β_5 * Δ in dividend per share

+ β_6 * % Δ in debt-equity ratio

+ β_7 * % Δ in (capital expenditures/total assets)

+ β_8 * Return on closing equity

+ β_9 * % Δ in total assets

+ β_{10} * Working capital / total assets

+ β_{11} * Cash dividend as % of cash flows

Year 2011 (the prediction period) data was used to test the prediction accuracy of each of the O&P earnings prediction models. The mean squared error was calculated for each observation in the sample for each model based on the year 2011 data. A paired sample t-test was conducted to compare the two sets of squared residuals—one based on the model using fully populated XBRL data and the other based on Compustat data. A level of significance of 0.05 was used in the paired sample t-test.

The result of the paired sample t-test was $p = 0.310$. This indicated that there was a not a significant difference in the prediction accuracy of the O&P Earnings Prediction Model using fully populated XBRL data (M = 7.8972) and the model using Compustat data (M = 5.9513). As stated above, six additional variables had to be excluded from the O&P earnings prediction models because of too much missing data in the fully populated XBRL company filing data set. The inclusion of these six variables would not allow SPSS to run the regression using the fully populated XBRL company filing data set, while SPSS could run the regression using the Compustat data set with these six variables included. Table 14 summarizes the percent of data complete on each of the 25 independent variables and the one dependent variable included in the original Ou and Penman (1989a, 1989b) earnings prediction model in the 2009 fully populated XBRL company filing data set and the 2009 Compustat data set.

Table 14

Percent of Data Complete for the Original Variables in the Ou and Penman Earnings Prediction Model in the 2009 Fully Populated XBRL Company Filing Data Set and the 2009 Compustat Data Set

	Fully Populated XBRL Company Filing Data Set	**Compustat Data Set**
% Δ in current ratio	98%	100%
% Δ in inventory turnover	100%	100%
Inventory / total assets	100%	96%
% Δ in (inventory / total assets)	90%	90%
% Δ in inventory	90%	90%
% Δ in sales	100%	100%
Δ in dividend per share	80%	100%
Δ in return on opening equity	0%	66%
% Δ in (capital expenditures / total assets)	94%	100%
% Δ in (capital expenditures / total assets), one-year lag**	68%	100%
% Δ in debt-equity ratio	96%	66%
% Δ in sales / total assets**	100%	100%
Return on total assets**	76%	98%
Return on closing equity	100%	100%
Gross margin ratio	86%	100%
% Δ in (pretax income / sales)**	66%	100%
Sales to total cash	100%	100%
% Δ in total assets	100%	100%
Cash flow to total debt**	46%	78%
Working capital / total assets	100%	100%
Operating income / total assets	84%	100%
Repayment of LT debt as % of total LT debt**	32%	88%
Cash dividend as % of cash flows	98%	100%
% Δ in depreciation***	0%	10%
% Δ in (depreciation / plant assets)***	0%	0%
Δ in earnings per share	94%	100%

The percent of data complete on each of the 25 independent variables and the one dependent variable included in the original Ou and Penman (1989a, 1989b) earnings prediction model in the 2009 fully populated XBRL company filing data set and the 2009 Compustat data set.

*Excluded from the O&P earnings prediction models because 2007 equity is not available in XBRL company filings

**Excluded from the O&P earnings prediction models because of too much missing data in the fully populated XBRL company filing data set, but not in the Compustat data set.

***Excluded from the O&P earnings prediction models because of too much missing data in the fully populated XBRL company filing data set and in the Compustat data set.

A&B Earnings Prediction Model

A multiple regression was conducted to estimate annual earnings using the fully populated XBRL filing data based on the year 2009 (the estimation period) for the A&B Earnings Prediction Model. The original earnings prediction model created by Abarbanell and Bushee (1997, 1998) included nine predictor variables. The variable related to audit qualification was eliminated in the current study because it was not available in XBRL company filings, the variable related to effective tax rate was eliminated because there were no observations in the sample, and the variables related to accounts receivable and labor force were eliminated because there were very few observations in the sample.

The remaining five independent variables used in this regression included:

- % Δ in inventory - % Δ in sales
- % Δ in industry capital expenditures - % Δ in firm capital expenditures
- % Δ in sales - % Δ in gross margin
- % Δ in selling and administrative expenses - % Δ in sales
- 0 for LIFO, 1 for FIFO or other

Table 15 reports summary statistics for the variables used in the A&B Earnings Prediction Model using fully populated XBRL company filing data. Table 16 presents the correlation coefficients, and Table 17 summarizes the results of the regression.

Table 15

Summary Statistics for the Variables Used in the A&B Earnings Prediction Model Based on the Fully Populated XBRL Company Filing Data for the Year 2009

Variable	**N**	**Minimum**	**25%**	**Median**	**Mean**	**75%**	**Maximum**	**Standard Deviation**
% Δ in inventory - % Δ in sales	46	-0.217	-0.053	0.003	0.079	0.130	0.853	0.245
% Δ in industry capital expenditures - % Δ in firm capital expenditures	47	-0.495	-0.105	0.104	0.051	0.231	0.377	0.233
% Δ in sales - % Δ in gross margin	43	-0.930	-0.099	-0.026	-0.084	0.013	0.125	0.220
% Δ in selling and administrative expenses - % Δ in sales	40	-0.053	-0.024	0.082	0.106	0.187	0.486	0.147
0 for LIFO, 1 for FIFO or other	50	0	1.00	1.00	0.78	1.00	1	0.418
EPS_{t+1} - EPS_t	47	-1.494	0.090	0.450	0.716	1.350	3.450	1.172

Summary statistics for the 2009 variables used in the A&B Earnings Prediction Model using fully populated XBRL company filing data.

Table 16

Pearson/Spearman Correlation Coefficients for the A&B Earnings Prediction Model Based on Fully Populated XBRL Company Filing Data for the Year 2009

	% Δ in inventory - % Δ in sales	% Δ in industry capital expenditures - % Δ in firm capital expenditures	% Δ in sales - % Δ in gross margin	% Δ in selling and administrative expenses - % Δ in sales	0 for LIFO, 1 for FIFO or other
% Δ in inventory - % Δ in sales	1.000 ---	-0.171	-0.017	0.280	-0.451
% Δ in industry capital expenditures - % Δ in firm capital expenditures	-0.106	1.000 ---	-0.132	0.214	-0.221
% Δ in sales - % Δ in gross margin	0.036	-0.299	1.000 ---	-0.310	0.232
% Δ in selling and administrative expenses - % Δ in sales	0.263	0.279	-0.242	1.000 ---	-0.365
0 for LIFO, 1 for FIFO or other	-0.378	-0.248	0.200	-0.413	1.000 ---

Pearson (Spearman) correlation coefficients are presented above (below) the diagonal.

Table 17

Results of the Multiple Regression for the A&B Earnings Prediction Model Based on the Fully Populated XBRL Company Filing Data for the Year 2009

	B	***t***	***p***
Constant	0.434	0.656	0.518
% Δ in inventory - % Δ in sales	-1.008	-0.702	0.489
% Δ in industry capital expenditures - % Δ in firm capital expenditures	0.117	0.112	0.912
% Δ in sales - % Δ in gross margin	-0.002	-0.002	0.999
% Δ in selling and administrative expenses - % Δ in sales	2.198	1.330	0.196
0 for LIFO, 1 for FIFO or other	-0.049	-0.071	0.944

Regression results estimating annual earnings for the A&B Earnings Prediction Model using fully populated XBRL company filing data for the year 2009.
Note. $N = 31$; Adjusted $R^2 = -0.059$. (the negative sign is not an error)

Another multiple regression was conducted to estimate annual earnings using the Compustat data based on the year 2009 (the estimation period) for the A&B Earnings Prediction Model. The same five independent variables and one dependent variable were included in this regression as in the one described above using fully populated XBRL company filing data. Table 18 reports summary statistics for the variables used in the O&P Earnings Prediction Model using Compustat data. Table 19 presents the Pearson/Spearman correlation coefficients, and Table 20 summarizes the results of the regression.

Table 18

Summary Statistics for the Variables Used in the A&B Earnings Prediction Model Based on the Compustat Data for the Year 2009

Variable	**N**	**Minimum**	**25%**	**Median**	**Mean**	**75%**	**Maximum**	**Standard Deviation**
% Δ in inventory - % Δ in sales	45	-0.221	-0.060	-0.002	0.058	0.127	0.483	0.189
% Δ in industry capital expenditures - % Δ in firm capital expenditures	50	-0.455	-0.123	0.097	0.042	0.229	0.456	0.252
% Δ in sales - % Δ in gross margin	50	-0.267	-0.074	-0.029	-0.039	0.018	0.075	0.081
% Δ in selling and administrative expenses - % Δ in sales	44	-0.079	-0.016	0.068	0.082	0.168	0.311	0.109
0 for LIFO, 1 for FIFO or other	50	0	1.00	1.00	0.78	1.00	1	0.419
EPS_{t+1} - EPS_t	50	-0.729	0.145	0.450	0.774	1.290	3.368	1.002

Summary statistics for the 2009 variables used in the A&B Earnings Prediction Model using Compustat data.

Table 19

Pearson/Spearman Correlation Coefficients for the A&B Earnings Prediction Model Based on Compustat Data for the Year 2009

	% Δ in inventory - % Δ in sales	% Δ in industry capital expenditures - % Δ in firm capital expenditures	% Δ in sales - % Δ in gross margin	% Δ in selling and administrative expenses - % Δ in sales	0 for LIFO, 1 for FIFO or other
% Δ in inventory - % Δ in sales	1.000 ---	-0.041	-0.040	0.344	-0.388
% Δ in industry capital expenditures - % Δ in firm capital expenditures	0.020	1.000 ---	-0.057	0.269	-0.088
% Δ in sales - % Δ in gross margin	0.058	-0.035	1.000 ---	-0.289	0.084
% Δ in selling and administrative expenses - % Δ in sales	0.272	0.316	-0.220	1.000 ---	-0.362
0 for LIFO, 1 for FIFO or other	-0.352	-0.122	0.089	-0.364	1.000 ---

Pearson (Spearman) correlation coefficients are presented above (below) the diagonal.

Table 20

Results of the Multiple Regression for the A&B Earnings Prediction Model Based on the Compustat Data for the Year 2009

	B	***t***	***p***
Constant	0.376	0.816	0.420
% Δ in inventory - % Δ in sales	-0.355	-0.338	0.737
% Δ in industry capital expenditures - % Δ in firm capital expenditures	-0.676	-0.956	0.346
% Δ in sales - % Δ in gross margin	2.607	0.878	0.386
% Δ in selling and administrative expenses - % Δ in sales	4.782	2.627	0.013
0 for LIFO, 1 for FIFO or other	0.244	0.534	0.597

Regression results estimating annual earnings for the O&P Earnings Prediction Model using Compustat data for the year 2009.
Note. $N = 41$;Adjusted $R^2 = 0.063$.

The coefficients calculated for each of the five variables and the one constant (using fully populated XBRL data and Compustat data) were used to create the A&B earnings prediction models.

A&B Earnings Prediction Model Using Fully Populated XBRL Data and Compustat Data

Δ in earnings per share = α

+ β_1 * % Δ in inventory - % Δ in sales

+ β_2 * % Δ in industry capital expenditures - % Δ in firm capital expenditures

+ β_3 * % Δ in sales - % Δ in gross margin

+ β_4 * % Δ in selling and administrative expenses - % Δ in sales

+ β_5 * 0 for LIFO, 1 for FIFO or other

Year 2011 (the prediction period) data was used to test the prediction accuracy of each model. The mean squared error was calculated for each observation in the sample for each model based on the year 2011 data. A paired sample t-test was conducted to compare the two sets of squared residuals—one based on the model using fully populated XBRL data and the other based on Compustat data. A level of significance of 0.05 was used in the paired sample t-test.

The result of the paired sample t-test was $p = 0.039$. This indicated that there was a significant difference in the prediction accuracy of the model using fully populated XBRL data and the model using Compustat data. Upon comparing the mean squared residuals between the two sources of data, it was observed that the squared residuals of the fully populated XBRL data (M = 0.1725) were lower than those of the Compustat data (M = 2.8795). However, because of the low R_2 of the regression using fully populated XBRL company filings data and Compustat data, it cannot be assumed that the fully populated XBRL data had higher prediction accuracy than the Compustat data for the A&B Earnings Prediction Model.

The most significant finding in this study was that current XBRL company filings cannot be used to create earnings prediction models; however, fully populated XBRL company filings can. All XBRL company filings could be fully populated with functionality built directly into the XBRL taxonomy, and this would not create any excess time, effort, or cost for preparers or users. Because current XBRL company filings could not be used to create earnings prediction models but fully populated XBRL company filings could, there is a strong possibility that fully populated XBRL company filings would be more useful in other areas as well. It must be noted that this study did not determine that fully populated XBRL company filing data predicts at a higher level than Compustat data. Nonetheless, the inherent timing and cost advantages of XBRL data collection potentially makes fully populated XBRL company filing data a useful data source.

VII. CONCLUSIONS

The SEC requires all public companies to report financial statements using XBRL. The availability of company-reported financial statement data in a computer-readable format offers a number of potential uses, and a great deal of research is needed to explore these opportunities. At a more basic level, however, deficiencies that might cause the current XBRL company data to be inadequate in its use must first be investigated. There has been a great deal of research highlighting the limited usefulness of current XBRL company filings and the rarity of their use, yet very few studies have attempted to delve deeper into the source of the deficiencies.

This study contributes to the common body of knowledge in accounting by investigating (1) if current XBRL company filings provide adequate interactive data access and (2) how modest changes to the functionality in the XBRL taxonomy could make XBRL much more useful. This was accomplished by first attempting to interactively obtain the balances of 70 accounting concepts from a sample of current XBRL company filings. What is meant by "interactively" is that only the required information was extracted from the XBRL company filings without any attempt to manually calculate any missing balances. The SEC states that XBRL company filings allow for interactive use of the accounting data and, in fact, refers to XBRL company filing data as interactive data (SEC, 2009). It was found that current XBRL company filings do not allow for interactive extraction of required accounting elements because too many accounting elements are not tagged in the XBRL company filings.

In order to demonstrate the potentially improved usefulness of XBRL company filings, a fully populated set of XBRL company filings was created. This was accomplished by manually populating any missing accounting concepts in the XBRL company filings if there were sufficient component accounting concepts tagged with the XBRL company filings. This was done to mimic a process that could be accomplished by functionality built directly into the XBRL taxonomy and possibly a few changes to the rules for XBRL filing preparation. It was found that many more accounting concepts could be interactively captured with the fully populated XBRL company filings.

The SEC and other proponents of XBRL argue that XBRL company filing data offers a number of advantages over the data provided by data aggregators, such as lower cost, quicker availability, and broader coverage of companies. Compustat is a leading provider to the market for accounting information and has been used in a considerable number of research studies. Prior research has shown that significant differences exist among the data reported by companies and the data reported in Compustat, largely because standardization techniques have been applied to the data in the Compustat database. Because XBRL company filing data may be different from the data reported by Compustat, research should be done to identify areas where research results differ when using standardized data rather than company-reported data and also where research and practice could be improved by using XBRL company filing data. As an interesting test of the fully populated XBRL company filing data, two earnings prediction models were created using fully populated XBRL company filing data and then the same two earnings prediction models were created using Compustat data. The predictive ability of each data set was compared in regard to the prediction of future earnings. The results indicated that, for one of the models, fully populated XBRL company filings predicted future earnings with a higher level of accuracy than did Compustat. There was no significant difference in the prediction accuracy between fully populated XBRL company filings and Compustat for the other future earnings prediction model.

The most important result in this study was that current XBRL company filings cannot be used to create earnings prediction models, but fully populated XBRL company filings can. XBRL company filings could be transformed into a fully populated XBRL company filing with functionality built directly into the XBRL taxonomy. This functionality would not create any excess cost for preparers or users or require any additional time or effort. The fact that current XBRL company filings could not be used to create earnings prediction models but fully populated XBRL company filings could indicates that current XBRL company filings are likely to be limited in their usefulness in other areas as well, while fully populated XBRL company filings would greatly improve their usefulness. The findings of this study are of interest to a broad constituency, including regulators such as the SEC and the Financial Accounting Standards Board, data aggregators, analysts, investors, researchers, XBRL US, XBRL International, and others.

LIST OF REFERENCES

Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research*, *35*(1), 1–24.

Abarbanell, J. S., & Bushee, B. J. (1998). Abnormal returns to a fundamental analysis strategy. *Accounting Review*, 19–45.

Alles, M. G., & Gray, G. L. (2012). A relative cost framework of demand for external assurance of XBRL filings. *American Accounting Association*.

American Institute of Certified Public Accountants (AICPA). (2009). Statement of position 09-1: Performing agreed-upon procedures engagements that address the completeness, accuracy, or consistency of XBRL-tagged data. New York.

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 159–178.

Bartley, J., Chen, A., & Taylor, E. (2010a). Avoiding common errors of XBRL implementation. *Journal of Accountancy*, *209*(2), 46–51.

Bartley, J., Chen, A. Y. S., & Taylor, E. Z. (2011). A comparison of XBRL filings to corporate 10-Ks—Evidence from the voluntary filing program. *Accounting Horizons*, *25*(2), 227–245.

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

Boritz, J. E., & No., W. G. (2009). Assurance on XBRL-related documents: The case of United Technologies Corporation. *Journal of Information Systems*, *23*, 49.

Boritz, J. E., & No, W. G. (2013). The Quality of Interactive Data: XBRL Versus Compustat, Yahoo Finance, and Google Finance. (Working Paper).

Capital IQ Compustat. (2013). The Impact of Disparate Data Standardization on Company Analysis. [cited May 2013]

Capital IQ Compustat. (2013). *Compustat standardization & data quality*. [cited May 2013] Retrieved from http://www.compustat.com/Compustat_Standardization/.

Capozzoli, E., & Farewell, S. (2010). SEC XBRL filing requirements: An instructional case on tagging financial statement disclosures. *Issues in Accounting Education, 25*(3), 489–511.

Chychyla, R., & Kogan, A. (2013). *Using XBRL to conduct a large-scale study of discrepancies between the accounting numbers in Compustat and SEC 10-K filings*. Available at SSRN: http://ssrn.com/abstract=2304473 or http://dx.doi.org/10.2139/ssrn.2304473.

Cox, C. (2006). Opening remarks to the Practising Law Institute's SEC speaks series, a speech delivered at SEC Speaks, Washington, D.C., March 3. Retrieved from http://www.sec.gov/news/speech/spch030306cc.htm.

D'Souza, J. M., Ramesh, K., & Shen, M. (2010). The interdependence between institutional ownership and information dissemination by data aggregators. *The Accounting Review, 85*(1), 159–193.

Deakin, E. B. (1976). Distributions of financial accounting ratios: Some empirical evidence. *The Accounting Review, 51*(1), 90–96.

Debreceny, R., d'Eri, A., Felden, C., Farewell, S., & Piechocki, M. (2011a). Feeding the information value chain: Deriving analytical ratios from XBRL filings to the SEC. Paper read at American Accounting Association Annual Meeting, Aug 6-10, at Denver, Colorado, USA.

Debreceny, R., & Farewell, S. (2010). Adios! Airways: An assignment on mapping financial statements to the US GAAP XBRL taxonomy. *Issues in Accounting Education*, *25*, 465.

Debreceny, R., Farewell, S., Piechocki, M., Felden, C., & Graning, 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.

Debreceny, R. S., Farewell, S. M., Piechocki, M., Felden, C., Gräning, A., & d'Eri, A. (2011b). Flex or break? Extensions in XBRL disclosures to the SEC. *Accounting Horizons, 25*(4), 631–657.

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.

Fang, J. (2011). Why is the U.S. XBRL conversion process so slow? *The CPA Journal*, *81*(1), 6, 8–10.

Federal Financial Institutions Examination Council. (2006). Improved business process through XBRL: A use case for business reporting. [cited May 2013] Retrieved from http://xbrl.us/wp-content/uploads/2007/12/20060202FFIECWhitePaper.pdf

Financial Accounting Standards Board. (2012). *US GAAP Financial Reporting Taxonomy*. Retrieved from http://xbrlview.fasb.org/yeti/resources/yeti-gwt/Yeti.jsp#tax~%28id~52*v~235%29!net~%28a~1058*l~251%29!lang~%28code~en-us%29!rg~%28rg~32*p~12%29.

Garbellotto, G. (2009). XBRL implementation strategies: The deeply embedded approach. *Strategic Finance, 91*(5), 56–57, 61.

Graham, B., Dodd, D. L., & Cottle, S. (1962). *Security Analysis, Principles and Technique*. (4th ed.) New York: McGraw-Hill.

Gruenberg, M. J. (2006). XBRL at the FDIC, a speech delivered at the 14th Annual XBRL International Conference, Philadelphia, Pennsylvania, December 4. Retreived from http://www.fdic.gov/news/news/speeches/archives/2006/chairman/spdec0406.html.

Harris, T. S., & Morsfield, S. G. (2012). An evaluation of the current state and future of XBRL and interactive data for investors and analysts.

Hoffman, C., Kurt, C., & Koreto, R. (1999). The XML files. *Journal of Accountancy*, *187*(5), 71–77.

Janvrin, D. J., & No, W. G. (2012). XBRL implementation: A field investigation to identify research opportunities. *Journal of Information Systems*, *26*(1), 169–197.

Kern, B. B., & Morris, M. H. (1994). Differences in the compustat and expanded value line databases and the potential impact on empirical research. *Accounting Review*, 274–284.

Kernan, K. (2009). *The Story of Our New Language*. American Institute of Certified Public Accountants, Inc. Retrieved from http://www.aicpa.org/interestareas/frc/accountingfinancialreporting/xbrl/downloadabledocuments/xbrl_09_web_final.pdf.

Kothari, S. (2001). Capital markets research in accounting. *Journal of Accounting and Economics, 31*(1), 105–231.

Lev, B., & Thiagarajan, S. R. (1993). Fundamental information analysis. *Journal of Accounting Research, 31*(2), 190–215.

Mahoney, L., & White, C. (2007). Creating XBRL instance documents in Excel. *CPA Journal*, *77*(7), 66.

McNamar, R. (2003). New technology can help avoid a second Enron. *Regulation*, *26*, 62.

Morsfield, S. (2012). Q&A with an expert: How investors and analysts are using XBRL-tagged data (part 1). *Dimensions*. Retrieved from http://merrilldisclosuresolutions.org/2012/07/16/qa-with-an-expert-how-investors-and-analysts-are-using-xbrl-tagged-data-part-1/

Ou, J. A. (1990). The information content of nonearnings accounting numbers as earnings predictors. *Journal of Accounting Research, 28*(1), 144–163.

Ou, J. A., & Penman, S. H. (1989a). Accounting measurement, price-earnings ratio, and the information content of security prices. *Journal of Accounting Research, 27*, 111–144.

Ou, J. A., & Penman, S. H. (1989b). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics, 11*(4), 295–329.

Penman, S. H. (1992). Return to fundamentals. *Journal of Accounting, Auditing & Finance, 7*(4), 465–483.

Philbrick, D. R., & Ricks, W. E. (1991). Using value line and IBES analyst forecasts in accounting research. *Journal of Accounting Research, 29*(2), 397–417.

Piechocki, M., Felden, C., Gräning, A., & Debreceny, R. (2009). Design and standardisation of XBRL solutions for governance and transparency. *International Journal of Disclosure and Governance, 6*(3), 224–240.

Plumlee, R. D., & Plumlee, M. A. (2008). Assurance on XBRL for financial reporting. *Accounting Horizons, 22*(3), 353–368.

Porcano, T. (1986). Corporate tax rates: Progressive, proportional, or regressive. *Journal of the American Taxation Association, 8*(1), 17–31.

Richardson, S., Tuna, İ., & Wysocki, P. (2010). Accounting anomalies and fundamental analysis: A review of recent research advances. *Journal of Accounting and Economics, 50*(2), 410–454.

Securities and Exchange Commission (SEC). (2013). *Interactive Data to Improve Financial Reporting* 2009 [cited June 2013]. Retrieved from http://www.sec.gov/rules/final/2009/33-9002.pdf.

Securities and Exchange Commission (SEC). (2013). *What Is Interactive Data and Who's Using It?* 2010 [cited May 2013]. Available from http://www.sec.gov/spotlight/xbrl/what-is-idata.shtml.

Sledgianowski, D., Fonfeder, R., & Slavin, N. S. (2010). Implementing XBRL reporting. *The CPA Journal, 80*(8), 68–72.

Srivastava, R. P., & Kogan, A. (2010). Assurance on XBRL instance document: A conceptual framework of assertions. *International Journal of Accounting Information Systems, 11*(3), 261–273.

Srivastava, R. P., & Liu, Q. (2012). Special issue of JIS on XBRL. *Journal of Information Systems, 26*(1), 97–101.

Standard Business Reporting (SBR). (2013). *About SBR* 2012 [cited June 2013]. Retrieved from http://www.sbr.gov.au/about-sbr.

Tallapally, P., Luehlfing, M. S., & Motha, M. (2011). The partnership of EDGAR Online and XBRL—Should Compustat care? *Review of Business Information Systems, 15*(4), 39–46.

Tallapally, P., Luehlfing, M. S., & Motha, M. (2012). Data differences—XBRL versus Compustat. *Journal of Technology Research, 3*.

Vasarhelyi, M. A., Chan, D. Y., & Krahel, J. (2012). Consequences of XBRL standardization on financial statement data. *Journal of Information Systems, 26*(1), 155–167.

XBRL International. (2011). *XBRL's History* 2011 [cited March 2011]. Available from http://www.xbrl.org/history.aspx.

XBRL International. (2012). XBRL International Annual Report 2012, 32.

XBRL International. (2013a). Tagging Is Sharing: Enriched Data for Use and Analysis. In *XBRL 26 International Conference*. Dublin, Ireland.

XBRL International. (2013). *XBRL International* 2013b [cited June 2013]. Retrieved from http://www.xbrl.org/xbrl-international-1.

XBRL International. (2013). *XBRL Projects* 2013c [cited June 2013]. Retrieved from http://www.xbrl.org/knowledge_centre/projects/list.

XBRL US, Inc. (2012). *XBRL US GAAP Taxonomy Preparers Guide* 2008 [cited June 25 2012]. Retrieved from http://xbrl.us/documents/preparersguide.pdf.

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.

Zarowin, S., & Harding, W. E. (2000). Finally, business talks the same language. *Journal of Accountancy,190*(2), 24–31.

Zimmerman, J. L. (1983). Taxes and firm size. *Journal of Accounting and Economics, 5*, 119–149.

APPENDIX

Appendix A – Calculations for the Variables Included in the O&P and A&B Earnings Prediction Models

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
Cash Dividend as Percent of Cash Flows	Cash Dividends $\text{Paid}_{\mathbf{t}}$ ÷ Cash Provided by $\text{Operations}_{\mathbf{t}}$	Cash $\text{Dividends}_{\mathbf{t}}$ *divided by* Operating Activities – Net Cash $\text{Flow}_{\mathbf{t}}$	$DV_{\mathbf{t}}$ ÷ $OANCF_{\mathbf{t}}$	$\text{PaymentsOfDividends}_{\mathbf{t}}$ ÷ $\text{NetCashProvidedByUsedInOperatingActivities}_{\mathbf{t}}$
Change in Dividends per Share	(Dividends $\text{Declared}_{\mathbf{t}}$ ÷ Common Shares $\text{Outstanding}_{\mathbf{t}}$) - (Dividends $\text{Declared}_{\mathbf{t-1}}$ ÷ Common Shares $\text{Outstanding}_{\mathbf{t-1}}$)	(Dividends $\text{Common/Ordinary}_{\mathbf{t}}$ *divided by* (Common Shares Used to Calculate Earnings Per Share $\text{Basic}_{\mathbf{t}}$ *multiplied by* Adjustment Factor (Company) - Cumulative by $\text{Ex-Date}_{\mathbf{t}}$)) *minus* (Dividends $\text{Common/Ordinary}_{\mathbf{t-1}}$ *divided by* (Common Shares Used to Calculate Earnings Per Share $\text{Basic}_{\mathbf{t-1}}$ *multiplied by* Adjustment Factor (Company) - Cumulative by $\text{Ex-Date}_{\mathbf{t-1}}$))	($DVC_{\mathbf{t}}$ ÷ ($CSHPRI_{\mathbf{t}}$ X $AJEX_{\mathbf{t}}$)) - ($DVC_{\mathbf{t-1}}$ ÷ ($CSHPRI_{\mathbf{t-1}}$ X $AJEX_{\mathbf{t-1}}$))	$\text{CommonStockDividendsPerShareDeclared}_{\mathbf{t}}$ - $\text{CommonStockDividendsPerShareDeclared}_{\mathbf{t-1}}$
Change in Earnings Per Share	Adjusted $EPS_{\mathbf{t+1}}$ - Adjusted $EPS_{\mathbf{t}}$	(Earnings Per Share (Basic) – Excluding Extraordinary $\text{Items}_{\mathbf{t+1}}$ *divided by* Adjustment Factor (Company) - Cumulative by $\text{Ex-Date}_{\mathbf{t+1}}$) *minus* (Earnings Per Share (Basic) – Excluding Extraordinary $\text{Items}_{\mathbf{t}}$ *divided by* Adjustment Factor (Company) - Cumulative by $\text{Ex-Date}_{\mathbf{t}}$)	($EPSPX_{\mathbf{t+1}}$ ÷ $AJEX_{\mathbf{t+1}}$) - ($EPSPX_{\mathbf{t}}$ ÷ $AJEX_{\mathbf{t}}$)	IncomeLossFromContinuingOperationsPerBasi $\text{cShare}_{\mathbf{t+1}}$ - IncomeLossFromContinuingOperationsPerBasi $\text{cShare}_{\mathbf{t}}$
Change in Earnings per Share Minus Drift[i]	$EPS_{\mathbf{t+1}}$ - $EPS_{\mathbf{t}}$ - $\text{drift}_{\mathbf{t+1}}$ (drift_{t} is estimated as the mean earnings-per-share change over the four	(Earnings Per Share (Basic) – Excluding Extraordinary $\text{Items}_{\mathbf{t+1}}$ ÷ Adjustment Factor (Company) - Cumulative by Ex-$\text{Date}_{\mathbf{t+1}}$) *minus* Earnings Per Share (Basic) – Excluding Extraordinary $\text{Items}_{\mathbf{t}}$ ÷ Adjustment Factor (Company) - Cumulative by $\text{Ex-Date}_{\mathbf{t+1}}$) *minus*	($EPSPX_{\mathbf{t+1}}$ ÷ $AJEX_{\mathbf{t+1}}$) - ($EPSPX_{\mathbf{t}}$ ÷ $AJEX_{\mathbf{t}}$) - (((($EPSPX_{\mathbf{t}}$ ÷ $AJEX_{\mathbf{t}}$) - ($EPSPX_{\mathbf{t-1}}$ ÷ $AJEX_{\mathbf{t-1}}$)) +	IncomeLossFromContinuingOperationsPerBasi $\text{cShare}_{\mathbf{t+1}}$ - IncomeLossFromContinuingOperationsPerBasi $\text{cShare}_{\mathbf{t}}$ - (((IncomeLossFromContinuingOperationsPerBa $\text{sicShare}_{\mathbf{t}}$ - IncomeLossFromContinuingOperationsPerBasi $\text{cShare}_{\mathbf{t-1}}$) +

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
	years prior to year t+1) ***Except in 2009:*** EPS_{t+1} - EPS_{t} - $drift_{t+1}$ ($drift_{t}$ is estimated as the mean earnings-per-share change over the three years prior to year t+1)	((((Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t}$) *minus* (Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t-1}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t-1}$)) *plus* ((Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t-1}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t-1}$) *minus* (Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t-2}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t-2}$)) *plus* ((Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t-2}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t-2}$) *minus* Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t-3}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t-3}$))) *divided by* 3) ***Except in 2009:*** (Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t+1}$ ÷ Adjustment Factor (Company) - Cumulative by $Ex\text{-}Date_{t+1}$) *minus* Earnings Per Share (Basic) – Excluding Extraordinary $Items_{t}$	(($EPSPX_{t-1}$ ÷ $AJEX_{t-1}$) - ($EPSPX_{t-2}$ ÷ $AJEX_{t-2}$)) + (($EPSPX_{t-2}$ ÷ $AJEX_{t-2}$) - ($EPSPX_{t-3}$ ÷ $AJEX_{t-3}$))) ÷ 3) ***Except in 2009:*** ($EPSPX_{t+1}$ ÷ $AJEX_{t+1}$) - ($EPSPX_{t}$ ÷ $AJEX_{t}$) - (((($EPSPX_{t}$ ÷ $AJEX_{t}$) - ($EPSPX_{t-1}$ ÷ $AJEX_{t-1}$)) + (($EPSPX_{t-1}$ ÷ $AJEX_{t-1}$) - ($EPSPX_{t-2}$ ÷ $AJEX_{t-2}$))) ÷ 2)	($IncomeLossFromContinuingOperationsPerBasicShare_{t-1}$ - $IncomeLossFromContinuingOperationsPerBasicShare_{t-2}$) + ($IncomeLossFromContinuingOperationsPerBasicShare_{t-2}$ - $IncomeLossFromContinuingOperationsPerBasicShare_{t-3}$)) ÷ 3) ***Except in 2009:*** $IncomeLossFromContinuingOperationsPerBasicShare_{t+1}$ - $IncomeLossFromContinuingOperationsPerBasicShare_{t}$ - ((($IncomeLossFromContinuingOperationsPerBasicShare_{t}$ - $IncomeLossFromContinuingOperationsPerBasicShare_{t-1}$) + ($IncomeLossFromContinuingOperationsPerBasicShare_{t-1}$ - $IncomeLossFromContinuingOperationsPerBasicShare_{t-2}$)) ÷ 2)

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
		÷ Adjustment Factor (Company) - Cumulative by Ex-Date$_{t+1}$) *minus* ((((Earnings Per Share (Basic) – Excluding Extraordinary Items$_{t}$ ÷ Adjustment Factor (Company) - Cumulative by Ex-Date$_{t}$) *minus* (Earnings Per Share (Basic) – Excluding Extraordinary Items$_{t-1}$ ÷ Adjustment Factor (Company) - Cumulative by Ex-Date$_{t-1}$)) *plus* ((Earnings Per Share (Basic) – Excluding Extraordinary Items$_{t-1}$ ÷ Adjustment Factor (Company) - Cumulative by Ex-Date$_{t-1}$) *minus* (Earnings Per Share (Basic) – Excluding Extraordinary Items$_{t-2}$ ÷ Adjustment Factor (Company) - Cumulative by Ex-Date$_{t-2}$))) *divided by* 2)		
Gross Margin Ratio	Gross Profit$_{t}$ ÷ Net Sales$_{t}$	Gross Profit (Loss)$_{t}$ *divided by* Sales/Turnover (Net)$_{t}$	GP$_{t}$ ÷ SALE$_{t}$	GrossProfit$_{t}$ ÷ SalesRevenueNet$_{t}$
Inventory	Inventory Valuation Method$_{t}$	Inventory Valuation Method$_{t}$	INVVAL$_{t}$	InventoryPolicyTextBlock$_{t}$
Inventory / Total Assets	Inventory$_{t}$ ÷ Total Assets$_{t}$	Inventories – Total$_{t}$ *divided by* Assets – Total$_{t}$	INVT$_{t}$ ÷ AT$_{t}$	InventoryNet$_{t}$ ÷ Assets$_{t}$
Operating Income / Total Assets	Operating Income$_{t}$ ÷ Total Assets$_{t}$	Operating Income After Depreciation$_{t}$ *divided by* Assets – Total$_{t}$	OIADP$_{t}$ ÷ AT$_{t}$	OperatingIncomeLoss$_{t}$ ÷ Assets$_{t}$
Percent Change in (capital expenditures / total assets)	((Capital Expenditures$_{t}$ ÷ Total Assets$_{t}$) - (Capital	((Capital Expenditures$_{t}$ *divided by* Assets – Total$_{t}$) *minus* (Capital Expenditures$_{t-1}$ *divided by* Assets – Total$_{t-1}$)) *divided by*	((CAPX$_{t}$ ÷ AT$_{t}$) - (CAPX$_{t-1}$ ÷ AT$_{t-1}$)) ÷	((PaymentsToAcquirePropertyPlantAndEquipment$_{t}$ ÷ Assets$_{t}$) - (PaymentsToAcquirePropertyPlantAndEquipment$_{t-1}$ ÷ Assets$_{t-1}$)) ÷

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
	Expenditures$_{t-1}$ ÷ Total Assets$_{t-1}$)) ÷ (Capital Expenditures$_{t-1}$ ÷ Total Assets$_{t-1}$)	(Capital Expenditures$_{t-1}$ *divided by* Assets – Total$_{t-1}$)	(CAPX$_{t-1}$ ÷ AT$_{t-1}$)	(PaymentsToAcquirePropertyPlantAndEquipment$_{t-1}$ ÷ Assets$_{t-1}$)
Percent Change in Current Ratio	((Current Assets$_{t}$ ÷ Current Liabilities$_{t}$) - (Current Assets$_{t-1}$ ÷ Current Liabilities$_{t-1}$)) ÷ (Current Assets$_{t-1}$ ÷ Current Liabilities$_{t-1}$)	((Current Assets – Total$_{t}$ *divided by* Current Liabilities – Total$_{t}$) *minus* (Current Assets – Total$_{t-1}$ *divided by* Current Liabilities – Total$_{t-1}$)) *divided by* (Current Assets – Total$_{t-1}$ *divided by* Current Liabilities – Total$_{t-1}$)	((ACT$_{t}$ ÷ LCT$_{t}$) - (ACT$_{t-1}$ ÷ LCT$_{t-1}$)) ÷ (ACT$_{t-1}$ ÷ LCT$_{t-1}$)	((AssetsCurrent$_{t}$ ÷ LiabilitiesCurrent$_{t}$) - (AssetsCurrent$_{t}$ ÷ LiabilitiesCurrent$_{t}$)) ÷ (AssetsCurrent$_{t-1}$ ÷ LiabilitiesCurrent$_{t-1}$)
Percent Change in Debt - Equity Ratio	((Total Liabilities$_{t}$ ÷ Total Stockholder's Equity$_{t}$) - (Total Liabilities$_{t-1}$ ÷ Total Stockholder's Equity$_{t-1}$)) ÷ (Total Liabilities$_{t-1}$ ÷ Total	((Liabilities – Total$_{t}$ *divided by* Stockholders' Equity – Total$_{t}$) *minus* (Liabilities – Total$_{t-1}$ *divided by* Stockholders' Equity – Total$_{t-1}$)) *divided by* (Liabilities – Total$_{t-1}$ *divided by* Stockholders' Equity – Total$_{t-1}$)	((LT$_{t}$ ÷ TEQ$_{t}$) - (LT$_{t-1}$ ÷ TEQ$_{t-1}$)) ÷ (LT$_{t-1}$ ÷ TEQ$_{t-1}$)	((Liabilities$_{t}$ ÷ StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest$_{t}$) - (Liabilities$_{t-1}$ ÷ StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest$_{t-1}$)) ÷ (Liabilities$_{t-1}$ ÷ $_{t-1}$)

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
	Stockholder's Equity$_{t-1}$)			
Percent Change in Industry Capital Expenditures **-** Percent Change in Firm Capital Expenditures[ii]	((Industry Capital Expenditures$_t$ - Industry Capital Expenditures$_{t-1}$) ÷ Industry Capital Expenditures$_{t-1}$) - ((Firm Capital Expenditures$_t$ - Firm Capital Expenditures$_{t-1}$) ÷ Firm Capital Expenditures$_{t-1}$)	(**Industry Average** Capital Expenditures$_t$ *minus* ((**Industry Average** Capital Expenditures$_{t-1}$ *plus* **Industry Average** Capital Expenditures$_{t-2}$) *divided by* 2)) *divided by* ((**Industry Average** Capital Expenditures$_{t-1}$ *plus* **Industry Average** Capital Expenditures$_{t-2}$) *divided by* 2)) *minus* ((Capital Expenditures$_t$ *divided by* ((Capital Expenditures$_{t-1}$ *plus* Capital Expenditures$_{t-2}$) *divided by* 2)) *divided by* ((Capital Expenditures$_{t-1}$ *plus* Capital Expenditures$_{t-2}$) *divided by* 2)))	(**Industry Average**CAPX$_t$ - ((**Industry Average**CAPX$_{t-1}$ + **Industry Average**CAPX$_{t-2}$) ÷ 2)) ÷ ((**Industry Average**CAPX$_{t-1}$ + **Industry Average**CAPX$_{t-2}$) ÷ 2)) - ((CAPX$_t$ - ((CAPX$_{t-1}$ + CAPX$_{t-2}$) ÷ 2)) ÷ ((CAPX$_{t-1}$ + CAPX$_{t-2}$) ÷ 2)))	(**Industry Average**PaymentsToAcquirePropertyPlantAndEquipment$_t$ **-** ((**Industry Average**PaymentsToAcquirePropertyPlantAndEquipment$_{t-1}$ + **Industry Average**PaymentsToAcquirePropertyPlantAndEquipment$_{t-2}$) ÷ 2)) ÷ ((**Industry Average**PaymentsToAcquirePropertyPlantAndEquipment$_{t-1}$ + **Industry Average**PaymentsToAcquirePropertyPlantAndEquipment$_{t-2}$) ÷2)) **-** ((PaymentsToAcquirePropertyPlantAndEquipment$_t$ **-** ((PaymentsToAcquirePropertyPlantAndEquipment$_{t-1}$ + PaymentsToAcquirePropertyPlantAndEquipment$_{t-2}$) ÷ 2)) ÷ ((PaymentsToAcquirePropertyPlantAndEquipment$_{t-1}$) + PaymentsToAcquirePropertyPlantAndEquipment$_{t-2}$) ÷ 2)))
Percent Change in Inventory	(Inventory$_t$ - Inventory$_{t-1}$) ÷ Inventory$_{t-1}$	(Inventories – Total$_t$ *minus* Inventories – Total$_{t-1}$) *divided by* Inventories – Total$_{t-1}$	(INVT$_t$ - INVT$_{t-1}$) ÷ INVT$_{t-1}$	(InventoryNet$_t$ - InventoryNet$_{t-1}$) ÷ InventoryNet$_{t-1}$
Percent Change in Inventory -	((Inventory$_t$ - ((Inventory$_{t-1}$ +	((Inventory$_t$ *minus* ((Inventory$_{t-1}$ *plus* Inventory$_{t-2}$) *divided by* 2)) *divided by* ((Inventory$_{t-1}$ *plus* Inventory$_{t-2}$) *divided*	((INVT$_t$ - ((INVT$_{t-1}$ + INVT$_{t-2}$) ÷	((InventoryNet$_t$ - ((InventoryNet$_{t-1}$ + InventoryNet$_{t-2}$) ÷ 2)) ÷ ((InventoryNet$_{t-1}$ + InventoryNet$_{t-2}$) ÷ 2)) **-**

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
Percent Change in Sales	Inventory_{t-2}) ÷ 2)) ÷ ((Inventory_{t-1} + Inventory_{t-2}) ÷ 2)) - ((Net Sales_{t} - ((Net Sales_{t-1} + Net Sales_{t-2}) ÷ 2)) ÷ ((Net Sales_{t-1} + Net Sales_{t-2}) ÷ 2)) ****Except in 2009:*** ((Inventory_{t} - Inventory_{t-1}) ÷ Inventory_{t-1}) - ((Net Sales_{t} - Net Sales_{t-1} ÷ Net Sales_{t-1})	*by* 2)) *minus* ((Sales/Turnover $(\text{Net})_{t}$ *minus* ((Sales/Turnover $(\text{Net})_{t-1}$ *plus* Sales/Turnover $(\text{Net})_{t-2}$) *divided by* 2)) *divided by* ((Sales/Turnover $(\text{Net})_{t-1}$ *plus* Sales/Turnover $(\text{Net})_{t-2}$) *divided by* 2)) ****Except in 2009:*** ((Inventory_{t} *minus* Inventory_{t-1}) *divided by* Inventory_{t-1}) *minus* ((Sales/Turnover $(\text{Net})_{t}$ *minus* Sales/Turnover $(\text{Net})_{t-1}$) *divided by* Sales/Turnover $(\text{Net})_{t-1}$)	2)) ÷ ((INVT_{t-1} + INVT_{t-2}) ÷ 2)) - ((SALE_{t} - ((SALE_{t-1} + SALE_{t-2}) ÷ 2)) ÷ ((SALE_{t-1} + SALE_{t-2}) ÷ 2)) ****Except in 2009:*** ((INVT_{t} - INVT_{t-1}) ÷ INVT_{t-1}) - ((SALE_{t} - SALE_{t-1}) ÷ SALE_{t-1})	(($\text{SalesRevenueNet}_{t}$ - (($\text{SalesRevenueNet}_{t-1}$ + $\text{SalesRevenueNet}_{t-2}$) ÷ 2)) ÷ (($\text{SalesRevenueNet}_{t-1}$) + $\text{SalesRevenueNet}_{t-2}$) ÷ 2)) ****Except in 2009:*** ((InventoryNet_{t} - $\text{InventoryNet}_{t-1}$) ÷ $\text{InventoryNet}_{t-1}$) - (($\text{SalesRevenueNet}_{t}$ - $\text{SalesRevenueNet}_{t-1}$) ÷ $\text{SalesRevenueNet}_{t-1}$)
Percent Change in Inventory / Total Assets	((Inventory_{t} ÷ Total Assets_{t}) - (Inventory_{t-1} ÷ Total Assets_{t-1})) ÷	((Inventories – Total_{t} *divided by* Assets – Total_{t}) *minus* (Inventories – Total_{t-1} *divided by* Assets – Total_{t-1})) *divided by* (Inventories – Total_{t-1} *divided by* Assets – Total_{t-1})	((INVT_{t} ÷ AT_{t}) - (INVT_{t-1} ÷ AT_{t-1})) ÷ (INVT_{t-1} ÷ AT_{t-1})	((InventoryNet_{t} ÷ Assets_{t}) - ($\text{InventoryNet}_{t-1}$ ÷ Assets_{t-1})) ÷ ($\text{InventoryNet}_{t-1}$ ÷ Assets_{t-1})

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
	(Inventory$_{t-1}$ ÷ Total Assets$_{t-1}$)			
Percent Change in Inventory Turnover Ratio	((Cost of Goods_Sold$_{t}$ ÷ Average Inventory$_{t}$) - (Cost of Goods Sold$_{t-1}$ ÷ Average Inventory$_{t-1}$)) ÷ (Cost of Goods Sold$_{t-1}$ ÷ Average Inventory$_{t-1}$) ***Except in 2009:*** ((Cost of Goods_Sold$_{t}$ ÷ Inventory$_{t}$) - (Cost of Goods Sold$_{t-1}$ ÷ Inventory$_{t-1}$)) ÷ (Cost of Goods Sold$_{t-1}$ ÷ Inventory$_{t-1}$)	((Cost of Goods Sold$_{t}$ *divided by* ((Inventories – Total$_{t}$ *plus* Inventories – Total$_{t-1}$) *divided by* 2)) *minus* (Cost of Goods Sold$_{t-1}$ *divided by* ((Inventories – Total$_{t-1}$ *plus* Inventories – Total$_{t-2}$) *divided by* 2))) *divided by* (Cost of Goods Sold$_{t-1}$ *divided by* ((Inventories – Total$_{t-1}$ *plus* Inventories – Total$_{t-2}$) *divided by* 2)) ***Except in 2009:*** ((Cost of Goods Sold$_{t}$ *divided by* Inventories – Total$_{t}$) *minus* (Cost of Goods Sold$_{t-1}$ *divided by* Inventories – Total$_{t-1}$)) *divided by* (Cost of Goods Sold$_{t-1}$ *divided by* Inventories – Total$_{t-1}$)	((COGS$_{t}$ ÷ ((INVT$_{t}$ + INVT$_{t-1}$) ÷ 2)) - (COGS$_{t-1}$ ÷ ((INVT$_{t-1}$ + INVT$_{t-2}$) ÷ 2))) ÷ (COGS$_{t-1}$ ÷ ((INVT$_{t-1}$ + INVT$_{t-2}$) ÷ 2)) ***Except in 2009:*** ((COGS$_{t}$ ÷ INVT$_{t}$) - (COGS$_{t-1}$ ÷ INVT$_{t-1}$)) ÷ (COGS$_{t-1}$ ÷ INVT$_{t-1}$)	((CostOfRevenue$_{t}$ ÷ ((InventoryNet$_{t}$ + InventoryNet$_{t-1}$) ÷ 2)) - (CostOfRevenue$_{t-1}$ ÷ ((InventoryNet$_{t-1}$ + InventoryNet$_{t-2}$) ÷ 2))) ÷ (CostOfRevenue$_{t-1}$ ÷ ((InventoryNet$_{t-1}$ + InventoryNet$_{t-2}$) ÷ 2)) ***Except in 2009:*** ((CostOfRevenue$_{t}$ ÷ InventoryNet$_{t}$) - (CostOfRevenue$_{t-1}$ ÷ InventoryNet$_{t-1}$)) ÷ (CostOfRevenue$_{t-1}$ ÷ InventoryNet$_{t-1}$)
Percent Change in Sales	(Net Sales$_{t}$ - Net Sales$_{t-1}$) ÷ Net Sales$_{t-1}$	(Sales/Turnover (Net)$_{t}$ *minus* Sales/Turnover (Net)$_{t-1}$) *divided by* Sales/Turnover (Net)$_{t-1}$	(SALE$_{t}$ - SALE$_{t-1}$) ÷ SALE$_{t-1}$	(SalesRevenueNet$_{t}$ - SalesRevenueNet$_{t-1}$) ÷ SalesRevenueNet$_{t-1}$
Percent Change in Sales - Percent Change	((Net Sales$_{t}$ - ((Net Sales$_{t-1}$ +	((Sales/Turnover (Net)$_{t}$ *minus* ((Sales/Turnover (Net)$_{t-1}$ *plus* Sales/Turnover (Net)$_{t-2}$) *divided by* 2))	((SALE$_{t}$ - ((SALE$_{t-1}$ + SALE$_{t-2}$) ÷	((SalesRevenueNet$_{t}$ - ((SalesRevenueNet$_{t-1}$ + SalesRevenueNet$_{t-2}$) ÷ 2)) ÷ ((SalesRevenueNet$_{t-1}$ + SalesRevenueNet$_{t-2}$) ÷

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
in Gross Margin	Net Sales$_{t-2}$) ÷ 2)) ÷ ((Net Sales$_{t-1}$ + Net Sales$_{t-2}$) ÷ 2)) - ((Gross Profit$_{t}$ - ((Gross Profit$_{t-1}$ + Gross Profit$_{t-2}$) ÷ 2)) ÷ ((Gross Profit$_{t-1}$ + Gross Profit$_{t-2}$) ÷ 2))	*divided by* ((Sales/Turnover (Net)$_{t-1}$ *plus* Sales/Turnover (Net)$_{t-2}$) *divided by* 2)) *minus* ((Gross Profit$_{t}$ *minus* ((Gross Profit$_{t-1}$ *plus* Gross Profit$_{t-2}$) *divided by* 2)) *divided by* ((Gross Profit$_{t-1}$ *plus* Gross Profit$_{t-2}$) *divided by* 2))	2)) ÷ ((SALE$_{t-1}$ + SALE$_{t-2}$) ÷ 2)) - ((GP$_{t}$ - ((GP$_{t-1}$ + GP$_{t-2}$) ÷ 2)) ÷ ((GP$_{t-1}$ + GP$_{t-2}$) ÷ 2))	2)) - ((GrossProfit$_{t}$ – ((GrossProfit$_{t-1}$ + GrossProfit$_{t-2}$) ÷ 2)) ÷ ((GrossProfit$_{t-1}$ + GrossProfit$_{t-2}$) ÷ 2))
Percent Change in Selling and Administrative Expenses - Percent Change in Sales	((Selling, General and Administrative Expenses$_{t}$ - ((Selling, General and Administrative Expenses$_{t-1}$ + Selling, General and Administrative Expenses$_{t-2}$) ÷ 2)) ÷ ((Selling, General and Administrative Expenses$_{t-1}$ +	((Selling, General and Administrative Expense$_{t}$ *minus* ((Selling, General and Administrative Expense$_{t-1}$ *plus* Selling, General and Administrative Expense$_{t-2}$) *divided by* 2)) *divided by* ((Selling, General and Administrative Expense$_{t-1}$ *plus* Selling, General and Administrative Expense$_{t-2}$) *divided by* 2)) *minus* ((Sales/Turnover (Net)$_{t}$ *minus* ((Sales/Turnover (Net)$_{t-1}$ *plus* Sales/Turnover (Net)$_{t-2}$) *divided by* 2)) *divided by* ((Sales/Turnover (Net)$_{t-1}$ *plus* Sales/Turnover (Net)$_{t-2}$) *divided by* 2))	((XSGA$_{t}$ - ((XSGA$_{t-1}$ + XSGA$_{t-2}$) ÷ 2)) ÷ ((XSGA$_{t-1}$ + XSGA$_{t-2}$) ÷ 2)) - ((SALE$_{t}$ - ((SALE$_{t-1}$ + SALE$_{t-2}$) ÷ 2)) ÷ ((SALE$_{t-1}$ + SALE$_{t-2}$) ÷ 2))	((SellingGeneralAndAdministrativeExpense$_{t}$ - ((SellingGeneralAndAdministrativeExpense$_{t-1}$ + SellingGeneralAndAdministrativeExpense$_{t-2}$) ÷ 2)) ÷ ((SellingGeneralAndAdministrativeExpense$_{t-1}$ + SellingGeneralAndAdministrativeExpense$_{t-2}$) ÷ 2)) - ((SalesRevenueNet$_{t}$ – ((SalesRevenueNet$_{t-1}$ + SalesRevenueNet$_{t-2}$) ÷ 2)) ÷ ((SalesRevenueNet$_{t-1}$ + SalesRevenueNet$_{t-2}$) ÷ 2))

Variable	Variable Calculation	Compustat Concept	Compustat Concept Calculation	US GAAP XBRL Taxonomy Element Calculation
	Selling, General and Administrative Expenses$_{t-2}$) ÷ 2)) - ((Net Sales$_t$ - ((Net Sales$_{t-1}$ + Net Sales$_{t-2}$) ÷ 2)) ÷ ((Net Sales$_{t-1}$ + Net Sales$_{t-2}$) ÷ 2))			
Percent Change in Total Assets	(Total Assets$_t$ - Total Assets$_{t-1}$) ÷ Total Assets$_{t-1}$	(Assets – Total$_t$ *minus* Assets – Total$_{t-1}$) *divided by* Assets – Total$_{t-1}$	(AT$_t$ - AT$_{t-1}$) ÷ AT$_{t-1}$	(Assets$_t$ - Assets$_{t-1}$) ÷ Assets$_{t-1}$
Return on Closing Equity	Net Income$_t$ ÷ Ending Stockholders' Equity$_t$	Net Income$_t$ *divided by* StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest$_t$	NI$_t$ ÷ TEQ$_t$	ProfitLoss$_t$ ÷ StockholdersEquityIncludingPortionAttributableToNoncontrollingInterest$_t$
Sales to Total Cash	Net Sales$_t$ ÷ Total Cash$_t$	Sales/Turnover (Net)$_t$ *divided by* Cash$_t$	SALE$_t$ ÷ CH$_t$	SalesRevenueNet$_t$ ÷ CashAndCashEquivalentsAtCarryingValue$_t$
Working Capital / Total Assets	Working Capital$_t$ ÷ Total Assets$_t$	(Current Assets – Total$_t$ *minus* Current Liabilities – Total$_t$) *divided by* Assets – Total$_t$	(ACT$_t$ - LCT$_t$) / AT$_t$	(AssetsCurrent$_t$ - LiabilitiesCurrent$_t$) ÷ Assets$_t$

[i]Some of the 2009 calculations had to be slightly altered because the XBRL company filing data only goes back to 2008 for balance sheet items listed on the 2009 financial statements. 2009 was the first year that large companies (companies with a market capitalization of more than $5 billion) had to file financial statements in XBRL format.

[ii]**Industry Average**PaymentsToAcquirePropertyPlantAndEquipment was gathered using Compustat because gathering enough data in XBRL to get an industry average was not feasible.

VITA

KELLY LAWRENCE WILLIAMS

EDUCATION

Christian Brothers University, Master of Business Administration, December 2005

Christian Brothers University, Bachelor of Science in Accountancy (Psychology minor), August 2001

ACADEMIC AND PROFESSIONAL EXPERIENCE

Assistant Professor of Accounting, **Middle Tennessee State University** *August 2014 – Present*

Research Assistant/Instructor, **University of Mississippi** *August 2009 – May 2014*

Senior Accountant, **Marsh USA Inc.** *April 2006 – August 2009*

Staff Accountant, **Marsh USA Inc.** *August 2003 – April 2006*

Staff Accountant, **Dixon Odom, PLLC (now Dixon Hughes Goodman)** *September 2001 – August 2003*

Lead Instructor, **Becker Professional Education** *May 2006 – October 2006*

Instructor, **Becker Professional Education** *May 2005 – May 2006*

Classroom Facilitator, **Becker Professional Education** *1994 – 2001*

Accounting Intern, **Trammell Crow** *July 1998 – March 1999*

RESEARCH PAPERS

Tour of Five XBRL Tools: Products That Help Make Tagged Data Work for You and Your Clients (With Mitchell R. Wenger and Rick Elam; *Journal of Accountancy*; April 2013)

XBRL Tagging of Financial Statement Data Using XMLSpy: The Small Company Case (With Rick Elam and Mitchell R. Wenger; *Issues in Accounting Education*; August 2012)

Multilingual Meetings and the Time Value of Accurate Translations (With Milam Aiken and Will Pepper; 2012 *Business Research Yearbook*)
Global Ledger: The Next Step for XBRL (With Rick Elam and Mitchell R. Wenger; *The CPA Journal*; September 2011)

WORKING PAPERS

Analysis of Actual Company Filings Using XBRL: An Empirical Investigation into Interactive Use of XBRL Financial Statement Data (With Rick Elam and Mitchell R. Wenger)

SCHOLARLY PRESENTATIONS

PhD Rookie Recruiting & Research Camp – Miami, FL, 2013

University of Mississippi – University, MS, 2012

AAA Information Systems Section Midyear Meeting – Atlanta, GA, 2011

OTHER CONFERENCE ACTIVITIES

AAA Annual Meeting – San Francisco, CA, 2010

AAA Southeast Regional Meeting – Mobile, AL, 2010

AAA Information Systems Section Midyear Meeting – Atlanta, GA, 2011

AAA Annual Meeting – Anaheim, CA, 2013

TEACHING EXPERIENCE

Middle Tennessee State University *Fall 2014 – Present*

- Principals of Accounting I
  3 hours
- Principals of Accounting II
  21 hours

The University of Mississippi *Fall 2010 – Spring 2014*

- Principals of Accounting I
  21 hours
- Principals of Accounting II
  9 hours
- Cost Accounting
  6 hours
- Accounting Information Systems
  3 hours

Becker Professional Education

Lead Instructor *May 2006 – October 2006*

Instructor *May 2005 – May 2006*

FELLOWSHIPS, HONORS, & AWARDS

AAA/Deloitte Foundation/J. Michael Cook Doctoral Consortium Fellow, *2012*

University of Mississippi Graduate Achievement Award for Ph.D. Degree in Accountancy, *2011*

SERVICE

Student Success Committee *Fall 2014 – Present*

Awards & Scholarship Committee *Fall 2014 – Present*