CLUSTER ANALYSIS OF SHARE PRICE: HOW FIRM CHARACTERISTICS RELATE TO ACCOUNTING METRICS

Mfon Akpan [*], Guneet Dhillon [**], Kim Trottier [***]

* *Corresponding author,* Department of Accounting, Methodist University, Fayetteville, the USA
Contact details: Department of Accounting, Methodist University, 5400 Ramsey St., Fayetteville, NC 28311, the USA
** Faculty of Computer Science Alumni, Dalhousie University, Halifax, Canada
*** Department of Accounting, HEC Montréal, Québec, Canada

OPEN ACCESS

How to cite this paper: Akpan, M., Dhillon, G., & Trottier, K. (2021). Cluster analysis of share price: How firm characteristics relate to accounting metrics. *Risk Governance and Control: Financial Markets & Institutions, 11*(4), 8–25. https://doi.org/10.22495/rgcv11i4p1

ISSN Online: 2077-4303
ISSN Print: 2077-429X

Received: 06.09.2021
Accepted: 26.11.2021

JEL Classification: C8, G120, M41
DOI: 10.22495/rgcv11i4p1

Abstract

The purpose of this paper is to improve our understanding of the relationship between share price and accounting information. Much of the literature utilizes the earnings number to reflect firm value. However, the revenue number seems more relevant for high-tech firms (Xu, Cai, & Leung, 2007), and cash flow figures are more informative for internet companies (Romanova, Helms, & Takeda, 2012). We build on this notion that share price may map out to different accounting numbers for different firms. We collect 629 accounting metrics for 3,365 firms in the U.S. and estimate their correlation with the firms' share price. We analyze these correlations and find that many firms exhibit a low correlation between share price and earnings. Other accounting numbers are important for these firms, including book value of net assets, retained earnings, stock options, gain or loss items, special or non-recurring items, and dividend rates. We are curious to learn what causes firms to anchor onto different metrics, therefore perform a cluster analysis to group similar firms together along three key accounting metrics. We examine the composition of each cluster and find that capital structure, dividend patterns, the persistence of operations, age, and industry can influence which accounting number is correlated with firm value. We encourage other researchers to continue this exploration as there are many interesting questions to answer.

Keywords: Share Price, Earnings, Financial Accounting, EPS, Cluster Analysis, Accounting Numbers, Industry, Text Mining

Authors' individual contribution: Conceptualization — M.A. and K.T.; Methodology — G.D.; Software — G.D.; Validation — G.D.; Formal Analysis — K.T.; Investigation — G.D.; Data Curation — K.T.; Writing — Original Draft — M.A. and K.T.; Writing — Review & Editing — K.T.

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

1. INTRODUCTION

The investment community is always keen to assess firm value, and the accounting profession is often interested in how their numbers map out to this value. The most common proxy for firm value is share price, and the most common accounting metric related to it is the earnings number. Some research suggests the earnings-to-price relationship does not hold for all firms. Xu, Cai, and Leung (2007) show that revenues-to-price is a better metric for high-tech firms. Romanova, Helms, and Takeda (2012) show that cash flows may be a stronger indicator of value for internet companies. More recently, Barth, Li, and McClure (2021) explore how share price is associated with various accounting amounts such as net income, book value of equity, intangible assets, growth opportunities, revenue, operating cash flows, special items, other comprehensive income, dividends, capital expenditures, cost of goods sold, total assets, and selling, general and administrative expenses. However, there has been little exploration beyond these metrics.

This paper uses an exploratory approach with a large dataset to identify other accounting numbers that are strongly associated with share price, and then explores characteristics that differentiate firms based on which of their accounting metrics are connected to firm value. We calculate the correlation between the share price and 629 accounting metrics over 40 accounting periods for the 3,365 U.S. firms in our sample. Our data spans from 2009–2018 to encompass a time period when detailed accounting numbers were publicly available. We then identify the most important accounting metrics and use cluster analysis to group the firms according to the strength of their correlation with these metrics. Using various tools, we unpack some of the differentiating characteristics that lead firms to cluster together.

Our analysis of the correlations suggests that earnings, revenues, and cash flows are not among the top correlations for most firms. Other measures such as book value of equity, volatility, and unusual items often outperform the traditional metrics. For many firms, the highest share-price correlation comes from nonrecurring or special items, or items representing gains and losses. We perform a deeper dive into the companies where gains and losses are prevalent and discover that the share price of these firms is strongly connected to accounting numbers that report on volatility. We explore the performance of accounting metrics across the four financial reporting quarters and find that earnings and dividends correlate with share price in quarters 1, 2, and 3 whereas non-recurring items are reflected in firm value in the fourth quarter. Sales revenue seems important in quarters 1 and 4 but less so in the middle of the fiscal year.

We perform a cluster analysis to group similar firms together. After assessing several configurations, we end up with a model of 3 accounting metrics and 2 clusters. Two of the accounting metrics are consistent with the extant literature (revenue and earnings) while the third metric is a relatively new consideration (retained earnings). Our overall findings suggest that capital structure, dividend patterns, persistence of operations, age, and industry have bearing on whether firms are best represented by their earnings or their revenues.

Our research adds to the literature on accounting and firm value by exploring a very large set of new metrics and recognizing that a multitude of firm characteristics might determine what metric is best. While we concur with prior literature that earnings and revenues perform well globally, we identify many characteristics that cause one metric to outperform the other. Most of these characteristics go beyond the industry effects that have been documented in the literature. We recommend further research to explore these characteristics in more depth, to mine the data for more patterns, and to use case analysis to understand why these patterns exist.

The remainder of this paper consists of a literature review in Section 2, followed by a description of our research methodology in Section 3. Section 4 presents and discusses the results from the correlation analysis (Subsection 4.1), cluster formation (Subsection 4.2), cluster analysis (Subsection 4.3), and analysis of differentiating characteristics between clusters (Subsection 4.4). Section 5 concludes the paper.

2. LITERATURE REVIEW

Accounting research in capital markets has generally been underpinned by the association between earnings and share price (earnings-valuation hereafter). However, research has progressed towards investigating how other financial statement information relates to firm value (Jegadeesh & Livnat, 2006). Barth et al. (2021) examine the association between equity price and accounting amounts that include earnings and book value of equity along with 14 other amounts from financial statements and 10 industry indicators. They conclude that the relationship between accounting and share price has evolved and become more nuanced in our current economy. Moving away from a one-size-fits-all valuation approach requires considering variations in firm characteristics and accounting metrics.

2.1. Accounting metrics

Valuation models are based on predicting future cash flows. Out of the many numbers provided in financial statements, earnings are most often perceived as being associated with value. However, analysis has shown that the correlation of earnings to a firm's value has diminished over time (Chandra & Ro, 2008). For example, companies such as Amazon.com, Inc. (Amazon) experienced high stock valuations during periods when they had low or no earnings.

The earnings number is equal to revenues minus expenses plus or minus gains, losses, and special items; therefore some potentially useful information may be lost by looking at earnings instead of its components. Alternatively, some components may add noise in the valuation process (such as special items) therefore analysts often remove them from earnings when estimating equity value (Pope & Wang, 2005). The components of earnings have the potential to contain business performance information that may be diluted or lost in the earnings calculation (Chandra & Ro, 2008), particularly when they follow different processes due to their persistence, or susceptibility to manipulation and variations in accounting practices.

Revenue is generally more persistent than earnings (Jegadeesh & Livnat, 2006) and more consistent (Chandra & Ro, 2008). Analysts consider revenue to be more predictable and controllable by the company (Barker & Imam, 2008). Consequently, investors respond to information in revenues. When revenue reports beat forecasted expectations, stock returns tend to be higher than when the earnings beat forecasts (Jegadeesh & Livnat, 2006).

Revenue can contain information on cash flows that can be lost in the earnings calculation (Chandra & Ro, 2008), but some revenues are more informative about future cash flows than others. Barker and Imam (2008) report that most analysts consider earnings to have a higher quality when derived from the primary operating activities of a business versus non-operating activities, such as comprehensive income. Analysts have pointed out that the natural growth of revenue from operating activities is the predominant source of quality earnings (Barker & Imam, 2008). The natural growth of revenue from operations is more sustainable as time moves forward (Barker & Imam, 2008).

However, revenues can be complex in certain industries. E-businesses have disrupted fundamental financial accounting revenue recognition rules, which dictate that sales revenue can only be reported when products or services are delivered and accepted by customers at an agreed-upon price and collection of payment is reasonably certain (Raisinghani, Shoemaker, & Schkade, 2004). Some e-business practices potentially break these rules. E-businesses that provide subscription-based services have been found to report the subscription fees before the services are rendered, therefore accelerating sales revenue. This acceleration of revenue violates the accounting principle of delivery of service to constitute a sale. The provision of future services may also involve costs that are not matched to the reported accelerated sales revenue (Raisinghani et al., 2004).

Jegadeesh and Livnat (2006) propose that expenses are easier to manipulate than revenues. There is evidence that companies are less likely to practice earnings management when revenue and earnings trend in the same direction (Jegadeesh & Livnat, 2006). According to Raisinghani et al. (2004) revenue becomes a particularly important financial performance measure when a business has an upward trend in earnings while revenue is constant or flat. The earnings growth may not be based on improved sales, since the revenue is flat, and this may be an indicator of accounting earnings management.

The recognition of expenses may also be inconsistent. Accounting convention requires refunds to be netted from the sales revenue figures before being reported as net sales on the financial accounting income statement (Raisinghani et al., 2004). E-businesses have been found to report refunds on the income statement as expenses, therefore overstating both expenses and revenues. Sales discounts should also be netted from the sales price such as revenue of $80 in the case of a $100 sale with a $20 discount. Yet some e-businesses report the transaction as sales revenue of $100 with $20 of marketing expenses, which would overstate sales revenues and overstate marketing expenses (Raisinghani et al., 2004).

According to Raisinghani et al. (2004), the online retailer Amazon utilized earnings management techniques to increase its gross profit by understating the costs of goods sold. Amazon reported fulfillment costs for the maintenance and preparation of merchandise for delivery to customers as operating expenses. This earnings management practice presented overstated operating expenses. Amazon was able to present a higher gross profit than its competitors at that time (Raisinghani et al., 2004).

Given the potential issues with earnings and its components, one might consider the use of other accounting metrics. Financial statements contain other information not related to revenues and expenses, which could be critical when evaluating earnings to measure past performance and to predict future earnings persistence and earnings quality (Barker & Imam, 2008). Since valuation is based on expected future cash flows, some measures of the firm's current cash flows may be informative. The adjustments of accruals contained in earnings, such as allowance for doubtful accounts, warranties, refunds, and inventory can reduce the quality of cash flow from operations found on cash flow statements (Saito, 2012).

However, even the cash flow process varies by industry. For example, the liquidity requirements for e-businesses are not the same as their brick-and-mortar competitors (Raisinghani et al., 2004). There have been e-businesses that have produced negative earnings and, at the same time, have generated abundant cash flow from minimal capital investments. Raisinghani et al. (2004) explain that the computer technology company Dell, Inc. reported $20 million of sales per day while at the same time having poor short-term liquidity due to negative working capital.

2.2. Accounting rules

Accounting rules cause further complexities when using accounting metrics as a measure of firm value. Accounting conservatism has an impact on several measures used by analysts. Lacina (1998) explains that the principle of conservatism academically lends itself to the undervaluing of revenue-generating assets on the balance sheet. Pope and Wang (2005) explain that valuation on the book value and earnings capitalization factor increase with accounting conservatism. The impact of conservatism can be reflected in the stock price (Lacina, 1998), causing the book value to be less than market value as in Pope and Wang's (2005) model. According to Lacina (1998), companies with high research and development (R&D) expenses are prime examples of how accounting conservatism can impact the stock market results.

Some items, such as internally generated intangible assets, are difficult to measure and verify therefore are excluded from the financial statements. The modern accounting system is not adept at capturing the value of intangible assets, which reduces its ability to determine a firm value (Darrough & Ye, 2007). The funds spent on these assets are expensed, which diminishes earnings for firms investing in this type of intangibles (Saito, 2012).

Darrough and Ye (2007) explore the knowledge-based economy and conclude that firms with sustained losses tend to be those that invest heavily in R&D. Conversely, firms that spend heavily on R&D tend to have high financial earnings losses. There has been an increase in smaller firms that are investing heavily in R&D, and accounting rules related to R&D activity-related expenses have been detrimental to them (Darrough & Ye, 2007).

Unless these investments are uniform over the years, expensing them will also result in earnings that are more variable than would be obtained from capitalizing and systematically amortizing intangible assets. This is the case for e-businesses and other technology companies (Saito, 2012), which weakens their earnings-valuation relationship. However, as the business environment has moved from a traditional model to a more e-business model, the valuation of intangible assets has become increasingly important for many companies (Darrough & Ye, 2007).

All accounting metrics contain some subjectivity. The market trust in accurate reporting is such that accounting changes initiated by companies are viewed negatively by analysts due to the impact they may have on earnings quality (Barker & Imam, 2008)

Barker (2004) states that earnings do not have a practical definition, and the reporting of depreciation, interest income, and interest expense according to Generally Accepted Accounting Principles (GAAP) cannot be easily predicted. Bao and Bao (2004) state that management is responsible for financial reporting and at the core of financial reporting are accounting estimates known as accruals. Although the decision regarding the size and scope of accounting estimates should not be looked upon negatively if the purpose is to produce an accurate financial report, management's choice of accounting estimates can be viewed negatively if it is found to deviate from the goal of accurate financial reporting (Bao & Bao, 2004).

It may be difficult to ascertain whether accruals improved the financial results. As stated in Barker (2004), it is not possible to quantify income and expenses that are under the control of management versus those that come from external areas. External impacts are measured by management estimates, which have a strong effect on financial performance measures and earnings (Barker, 2004). However, in their sample of over 12,000 firm-years, Bao and Bao (2004) did not find earnings management to be problematic for financial reporting and stock valuation.

Barker and Imam (2008) assert that accounting changes initiated by companies are viewed negatively by analysts due to the impact they may have on earnings quality.

2.3. Economics

In addition to variations in accounting estimates and practices, firms experience economic factors that can distort earnings-valuation. The correlation between earnings and share price is problematic when firms incur a loss. In their exploration of e-businesses, Xu and Cai (2009) find that share price is more highly correlated with revenues for businesses with negative earnings. Yet losses are not a rare occurrence. The number of companies reporting negative earnings increased to over 40% in 2000 from approximately 3% in the 1960s (Darrough & Ye, 2007). Many of these companies did not present with a risk of distress or going concern. Many businesses sustain losses for many years as they invest heavily in assets that give future — and not immediate — benefits (Darrough & Ye, 2007).

Both elements of the earnings-valuation relationship can evolve through the business life cycle. Saito (2012) states that financial accounting earnings are not a strong indicator of future performance for new and yet-to-be-established companies. Darrough and Ye (2007) give the example of large pharmaceutical companies that focus heavily on the research and development of new drugs. The pharmaceutical companies use the capital raised through equity to finance the research and keep the businesses afloat during periods of low sales generation. According to Darrough and Ye (2007), these companies focus on revenue generation by putting their capital in areas that could potentially reap future returns. Revenue-focused businesses are more likely to generate positive future earnings than companies that are facing financial and operational losses without the same revenue-driven strategies (Darrough & Ye, 2007).

Firms in a growth stage should be valued differently. Firms with large investments in R&D and advertising have a stronger possibility of generating high positive earnings growth in the future therefore these aspects should be included in valuation (Tokic, 2002). Financial analysts sometimes perform stock valuation and recommendations based on the price-to-earnings-growth (PEG) ratio. Bradshaw (2000) notes that individual financial analysts tend to issue greater positive stock recommendations when their earnings forecasts are above the consensus earnings forecast, and concludes that financial analysts utilize PEG as a tool to calculate their earnings forecasts and formulate stock recommendations. PEG is not a formal stock valuation theory but is often utilized for the evaluation of firms that are part of the high-growth sector of businesses (Bradshaw, 2000).

Accounting metrics may vary by industry. According to Markman (2017), the usefulness of a traditional earnings-valuation model comes into question when analyzing internet businesses. Saito (2012) explains that immediate negative earnings are not a predictor of future earnings in a knowledge-based economy (Darrough & Ye, 2007).

Some accounting practices are industry-specific. For example, Dutta, Caplan, and Marcinko (2014) describe a common business practice for e-businesses in which products are sold to consumers before being bought from the manufacturer or vendor; which is known as drop shipping. A consumer places an order with the e-business, which then places an order to its supplier or manufacturer, and the manufacturer or vendor sends the product directly to the consumer. The e-business does not take possession of the product, which is different from traditional business retailers (Dutta et al., 2014). The way an e-business records sales transactions could greatly impact its financial statements (Dutta et al., 2014). E-businesses can record sales transactions on a gross or net basis. The gross method presents the entire amount received from the consumer as sales revenue. The net method presents sales revenue as the difference between the money received from the consumer and the money paid to the supplier for the cost of the product. The difference between the two is that under the gross method a higher revenue and cost of sales are presented on the financial statements (Dutta et al., 2014).

In response to Priceline.com, which used the gross method to overstate sales revenue, the U.S. Securities and Exchange Commission issued Staff Accounting Bulletin No. 101 — Revenue Recognition in Financial Statements[1] (Dutta et al., 2014). The bulletin describes how firms are required to report revenue on a net basis when the firm acts as a broker or agent without assuming ownership of the products or risk for payment default (Dutta et al., 2014).

[1] https://www.sec.gov/interps/account/sab101.htm

2.4. High-tech industry

Several studies examine the earnings-valuation relationship in the high-tech industry. Businesses that provide digital products and services have unique financial characteristics that differentiate them (Raisinghani et al., 2004). The digital products and services are indestructible, easily modified, and easily replicated; their indestructible nature refers to the depreciation of quality and the method of merchandising; and modification refers to product development, customization, and differentiation (Raisinghani et al., 2004).

King (2000) mentions some downsides to the valuation of startup e-businesses, namely the high startup costs that revolve around technology and marketing. Marketing has the greatest drain on capital, as electronic businesses seek to gain market share and compete against rival companies. Technology and market costs contribute to low revenue and negative profits. For more established firms, King (2000) explains that e-businesses have inherently low operating expenses and a high potential for profitability due to the global scope and constant availability of the internet to connect customers to businesses.

Xu and Cai (2009) analyze firms in e-businesses and find they have stock values that are more highly correlated with their revenues than with their earnings, with an effect that is more pronounced for businesses with negative earnings. More generally, Chandra and Ro (2008) find that e-businesses do not fit the traditional earnings-based valuation models, such as discounted cash flows or the dividend discount model. After investigating the pertinence of earnings, book value, sales, and R&D expenses for high-tech companies during 1990–1999 and 2000–2021, the data revealed a relevant link between sales revenues and R&D expenses in the valuation of high-tech firms and high-tech loss firms, in particular. Saito (2012) states that current accounting guidelines possibly decrease the earnings quality and informational value of financial accounting information for high-tech companies that rely on intangible assets for operations.

Chandra and Ro (2008) mention three reasons for the technology sector to be greatly impacted by accounting-based valuation:

- The technology industry was in a state of constant development and quick evolution, which the existing financial reporting system is not well equipped to reflect.
- Technology companies typically have large initial and ongoing R&D expenditures that cannot be capitalized and depreciated or amortized, impacting earnings.
- Technology companies have greater growth rates than other businesses due to their rapidly evolving nature.

Revenue recognition can present particular challenges in technology-based companies as illustrated by the following Chester Games Corporation (CGC) example, although reporting precision and comparability has improved with the adoption of a new accounting standard in 2018 on revenue from contracts with customers (Ferreira, 2020).

CGC was an online gaming business founded in 2011 (Conrod & Cumby, 2016). The company had between eight and 12 games available at any one time via the CGC gaming smartphone application. The users played the games for free. The players were able to enhance their game-playing experience by purchasing virtual currency in order to purchase virtual goods within the game. The majority of CGC's revenue was generated from the purchase of virtual currency and then the subsequent sale of virtual game items to customers that play the games hosted on CGC's servers (Conrod & Cumby, 2016).

Conrod and Cumby (2016) explain that consumable goods, such as energy or life, are consumed while playing the game. The consumable goods are short-lived, must be used immediately, and have no residual value. The durable goods have residual value for the player and can be used for an extended period. A tank is an example of a durable good, which will stay with the character for the life of the game unless it is lost, wrecked, or abandoned. Revenue recognition for consumable and durable goods is recorded separately for each game (Conrod & Cumby, 2016). The revenue from a consumable is recognized immediately. The revenue from durable goods is recognized over the average life of the player, as was standard industry policy in 2013 (Conrod & Cumby, 2016). However, the average life of the player was not easy to ascertain. There were players that played the game only once and others that played the game for years. There were various player upgrades and modifications that could impact player life. It took several years for CGC to collect data for these estimations as well as be able to differentiate between revenue from consumables or durable goods (Conrod & Cumby, 2016).

Raisinghani et al. (2004) sum up the main earnings-valuation issues in the high-tech industry. They assert that traditional earnings-based valuation methods — such as economic value added (EVA); earnings before interest, taxes, depreciation, and amortization (EBITDA); and price-to-earnings (PE) ratios — are not effective valuation tools for many e-businesses. The majority of e-businesses do not have positive earnings or maintain low earnings (Raisinghani et al., 2004). The e-business sector is prototypical of intellectual property and proprietary technology. Due to the lack of comparability of intellectual property value, for the most part, it cannot be listed on an e-business's balance sheet (Raisinghani et al., 2004).

These concerns are quickly spreading to other industries as well due to innovation. The first-to-market (FTM) businesses have traditionally commanded higher valuations. The first business to enter a market will receive most of the revenue stream and investment support in its particular area of expertise. Disruptive businesses utilize technology that drastically impacts an existing industry or creates a completely new industry (Raisinghani et al., 2004). E-businesses develop proprietary technology and intellectual property to increase efficiency and develop new markets. The advent of e-business has created FTM disruptive businesses. As Raisinghani et al. (2004) state, the first business to introduce disruptive technology will capture market share, sales revenues, and investment capital.

2.5. Analysts

Financial analysts are among the primary users of financial accounting information (Barker & Imam, 2008), and many have adapted their use of earnings-valuation. Demirakos, Strong, and Walker (2004) explain that conventional accounting models do a better job at valuing more traditional forms of business. For example, earnings forecasting is predominantly used by financial analysts in more stable business sectors, such as the beverage sector (Demirakos et al., 2004). Analysts prefer to use single-period valuation models for traditional businesses, whereas multi-period valuation models are used for more complex businesses (Demirakos et al., 2004).

Accounting information is heavily based on past estimates, making accurate current valuations challenging. Investors tend to focus on potential future performance and not past performance. As a result, financial analysts have increased their usage of current or market data to supplement financial accounting information (Raisinghani et al., 2004).

Ratios such as the PE ratio and the return on investment (ROI) are widely used for stock valuation (Ho, Liao, & Kim, 2011). However, earnings-based ratios are not practical for businesses that may have negative earnings for several years before earning a profit (Ho et al., 2011). Bailey, Brown, Potter, and Wells (2008) explain that valuations based on free cash flow (FCF) are problematic since FCF is derived from earnings. For firms with losses or the potential for high growth, their revenue is more relevant than cash flow for valuation. In these cases, the price-to-sales (PS) ratio often replaces the PE ratio for business valuation. The PS ratio can also be more meaningful for businesses that operate with few tangible assets and utilize capital to purchase intangible assets and advertising (Ho et al., 2011).

Research has shown that analysts tend to choose varying valuation methods based on the company's business sector (Demirakos et al., 2004). For example, traditional approaches to valuation are inappropriate for the valuation of e-businesses (Ho et al., 2011). According to Raisinghani et al. (2004), the valuation of e-businesses faces challenges in the areas of financial measurement tools, subjectivity, and strategy development. Investor confidence was diminished with the expansion of the technology sector and the emergence of e-business (Raisinghani et al., 2004). Chandra and Ro (2008) show that industry analysts predominantly use revenue and revenue ratios to value technology companies. King (2000) states that financial analysts evaluate existing e-businesses based on revenue trends, negative earnings trends, and the position of the company in the marketplace relative to its closest competitors. The FCF of e-businesses may be negative or unpredictable (Bailey et al., 2008). FCF volatility stems from large capital expenditures that may be found on the income statement, which is typical of e-businesses (Bailey et al., 2008).

Tokic (2002) concludes that the valuation of e-business is based on expected earnings and a suitable risk-adjusted discount rate. E-businesses have high growth and can invest capital at a higher-than-average rate of return. E-businesses invest capital in intangible assets, such as R&D and advertising which can strongly indicate future cash flows and increases in market value. The valuation of an e-business is a function of both earnings and growth. Growth strategies neglect earnings (Tokic, 2002).

Bradshaw (2004) demonstrates that analysts tend not to follow earnings-based models, such as the PEG model, and may recommend growth stocks without examining how much growth is already priced into the stock. Financial analysts use more than the present values of financial models to make buy or sell stock recommendations; they also incorporate personal opinions and qualitative research (Bradshaw, 2004).

The literature suggests the relationship between earnings and share price (value) may be affected by accounting practices, economics, and company or industry characteristics. This warrants an exploration of whether other accounting metrics relate better to firm value for specific firms.

3. RESEARCH METHODOLOGY

Our research approach is to collect a large set of accounting metrics for a variety of companies, and then use cluster analysis to group companies based on how their share price correlates with these metrics. We then analyze common factors within each group. We collect all 680 accounting variables from Compustat Quarterly during 2009–2018 for U.S. listed firms with 10 years of data, where the 10-year time series is to calculate correlations. We then collect share price and returns information from the Center for Research in Security Prices (CRSP) for these firms. Since the data is quarterly, this generates 134,600 observations on 3,365 firms. Our choice of time period is grounded on the theory of market efficiency, which states that share price reflects all publicly available information about the firm. We focus on the years 2009–2018 since detailed accounting information was digitally accessible to all market participants in this period due to the use of XBRL (Cong, Du, & Vasarhelyi, 2018).

When calculating correlations, we remove the 65 variables that are textual, such as address, state, and phone number, leaving us with 615 numeric variables. We incorporate an additional 14 constructs connected to prior literature (Table 1) for a total of 629 variables.

Table 1. Additional firm-specific variables constructed and explored

Variable	***Description***
ADA: Average discretionary accruals	The residuals from estimating the cross-sectional Jones model by industry (2-digit SIC when there were more than 8 observations, and 1-digit SIC for the rest), averaged over the 10-year period.
pREV: Persistence of revenues	The coefficient obtained in firm-specific regressions of revenues against prior quarter revenues, estimated with no intercept.
pEPS: Persistence of Earnings per Share	The coefficient obtained in firm-specific regressions of earnings per share (EPS) against prior quarter EPS, estimated with no intercept.
pDIV: Persistence of dividends	The coefficient obtained in firm-specific regressions of common dividends declared against prior quarter common dividends declared, estimated with no intercept.
pOPCFLOW: Persistence of operating cash flows	The coefficient obtained in firm-specific regressions of operating cash flows against prior quarter operating cash flows, estimated with no intercept.
pFINCFLOW: Persistence of financing cash flows	The coefficient obtained in firm-specific regressions of financing cash flows against prior quarter financing cash flows, estimated with no intercept.
pINVCFLOW: Persistence of investing cash flows	The coefficient obtained in firm-specific regressions of investing cash flows against prior quarter investing cash flows, estimated with no intercept.
growREV: Growth based on revenues	An indicator variable equal to one if *pREV* > 1 and zero otherwise.
growFLOW: Growth based on positive operating cash flows	An indicator variable equal to one if *pOPCFLOW* > 1 and zero otherwise.
growFIN: Growth based on negative financing cash flows	An indicator variable equal to one if *pFINCFLOW* < 0 and zero otherwise.
growINVEST: Growth based on negative investing cash flows	An indicator variable equal to one if *pINVCFLOW* < 0 and zero otherwise.
Dilution effect	Calculated as [common shares used to calculate basic EPS/common shares used to calculate diluted EPS] - 1).
Age	The number of years since the firm went public (had their Initial Public Offering) according to IPODATE, but noting the data is missing in much of the dataset (missing for 1,514/3,365 firms = 45% of observations).
Industry	Industry groupings based on SIC code ranges.

The 14 constructs include discretionary accruals from the earnings management literature (Acar & Coskun, 2020) and persistence variables pertaining to residual income valuation models (Fullana, González, & Toscano, 2021). The *Growth* and *Industry* variables relate to Barth et al. (2021), to which we add *Dilution* (McEnroe & Mindak, 2020) and *Age* (Hauser & Thornton, 2017). We scale the numeric variables that are not reported on a per-share basis by the number of shares outstanding at quarter-end so that both sides of the equation (i.e., share price and the accounting metrics) are on a per-share basis. We use the number of shares outstanding for the purpose of calculating basic EPS, but we also explore the share quantity from the diluted EPS number.

For sensitivity analysis, we perform our tests using share price on four different dates. Our analysis is performed separately using share price on the earnings announcement date or using the average share price over the three days following the earnings announcement date. We also link our results to prior literature with an assumption that filings are released 60 days following year-end (Impink, Lubberink, van Praag, & Veenman, 2012) or 3 months (90 days) after year-end (Xu et al., 2007). We report the results from using the average share price over the three days following the earnings announcement date, but verify they are not sensitive to our choice of share price metric.

Each firm in our sample is assumed to be best represented by an accounting metric, but we initially make no assumptions about which metric is best. We use simple correlation to identify which of the 629 numerical accounting variables are most highly associated with firm value through share price. Each firm has 40 observations (10 years × 4 quarters per year), which provides sufficient power for calculating correlations between their accounting metrics and firm value. We identify the most prevalent correlations, then group our sample of firms along these measures using cluster analysis.

Clustering algorithms have a wide variety of applications in accounting and auditing. Researchers have used clustering as an anomaly detection algorithm to recognize fraudulent practices (Issa & Vasarhelyi, 2011; Thiprungsri & Vasarhelyi, 2011). Gupta and Huefner (1972) employed clustering with financial ratios to identify similar industry characteristics. The recent studies of Hoberg and Phillips (2010, 2016) identify firm competitors by using text-based cluster analysis.

Frades and Matthiesen (2010) categorize clustering algorithms into 4 broad categories based on how these algorithms define clusters:

- Hierarchical clustering algorithms produce tree-shaped clusters called dendrograms. They can be either made by combining smaller clusters or splitting up bigger clusters, for example, clustering using representatives (CURE) (Guha, Rastogi, & Shim, 1998) and balanced iterative reducing and clustering using hierarchies (BIRCH) (Zhang, Ramakrishnan, & Livny, 1996).
- Partitional clustering algorithms aim to divide the given data into integral numbers of clusters. It does so in an iterative manner where the cluster sets change in every iteration as the algorithm approaches near local or global minima. *K*-means (Hartigan & Wong, 1979) is the most common example for unsupervised clustering.
- Density-based clustering leverages the density of data and patterns of neighboring data to segregate the data set. These algorithms are immune to outliers as they can easily detect clusters of arbitrary shapes. Density-based spatial clustering of applications with noise (DBSCAN) (Ester, Kriegel, Sander, & Xu, 1996) is a very popular example of density-based clustering algorithms.
- In grid-based clustering, the total space is segregated into a specific number of rectangular regions called cells. These algorithms are more concerned with the space between the data points. They are also computationally efficient especially when dealing with large multi-dimensional data (Frades & Matthiesen, 2010). An example of this is a statistical information grid approach to spatial data mining (STING) (Wang, Yang, & Muntz, 1997).

The goal of any clustering algorithm is to classify *N*-dimensional points into *K* clusters so that intra-cluster distance among the points is minimized. *K*-means clustering is one of the simplest and most efficient clustering algorithms (Hartigan & Wong, 1979). The motivation for using *K*-means for this research is the ease of implementation, guarantee of convergence, and the fact that the algorithm generalizes to different shapes of clusters. The working of this algorithm is as follows. Let (x_t) be the observation of a single firm in a multi-dimensional space of accounting metrics. The algorithm initializes K cluster centers $\{C_1, C_2, \dots, C_k\}$ for K clusters. The next part is calculating the Euclidean distance of each data point from each cluster center. The general formula for computing the Euclidean distance is,

$$E(p,q) = \sqrt{\sum_{i=1}^{m} (p_i, - q_i)^{2}} \quad (1)$$

where, p & q are Euclidean vectors of data points and cluster centers, and m is the number of dimensions of vector space.

Once data points are clustered in the first iteration, it is followed by calculation of new means or cluster centers $\{C_1, C_2, \dots, C_k\}$ from the respective data points. If new means or cluster centers are identical to the ones in the previous iteration, the algorithm stops. Otherwise, all data points are reassigned to new clusters using Euclidean distance, and the process repeats.

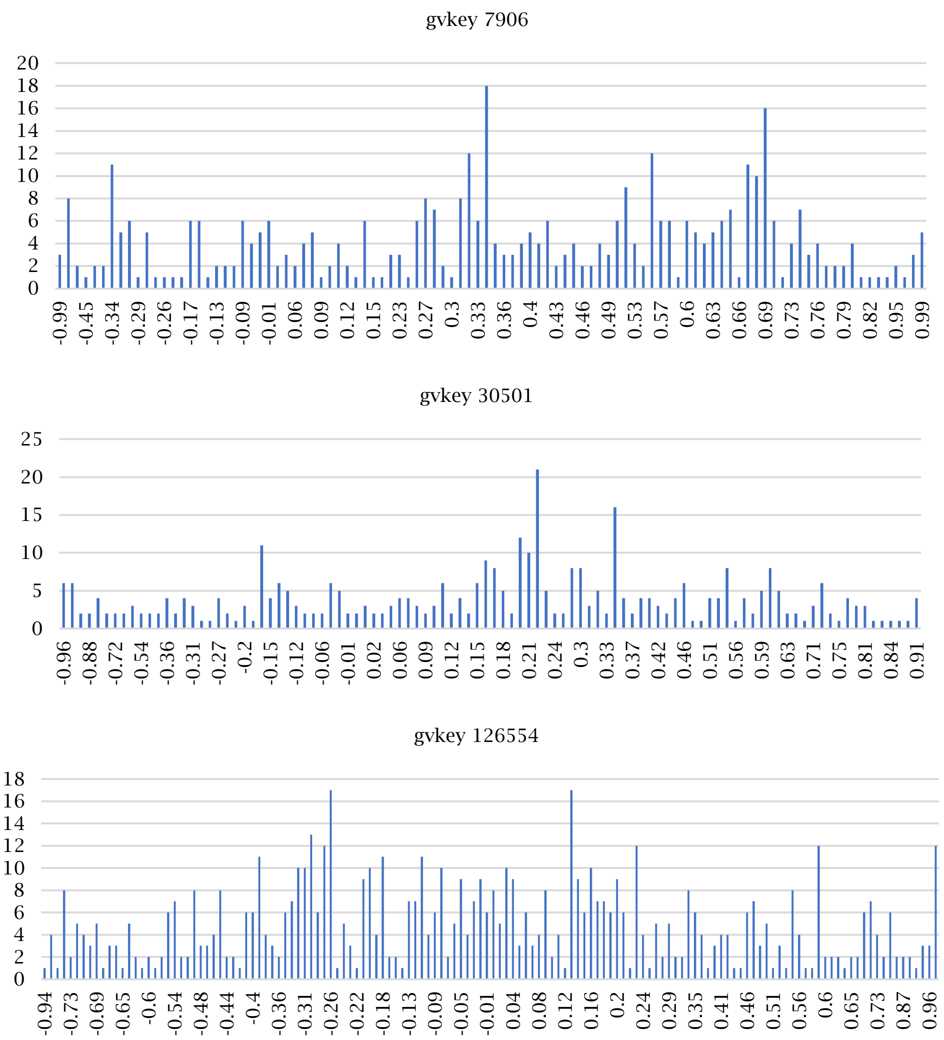
The technique divides the entire dataset into K clusters, but the choice of K can affect the efficiency of the algorithm. We apply the silhouette coefficient (SC) technique to design our best cluster analysis. This coefficient measures how close each point in one cluster is to points in the neighboring clusters. A value of 1 indicates the data can be categorized perfectly among the clusters, whereas a value of 0 tells us the clusters are meaningless (Aranganayagi & Thangavel, 2007). We calculate the silhouette coefficient's value for clusters $K = 2, 3, 4, \dots 10$. Once the firms are assigned to clusters, we explore firm characteristics that connect the firms within clusters.

4. RESULTS AND DISCUSSION

4.1. Correlations

The distribution of correlations ranges from large negative values to large positive values, as expected. We show the distribution of correlations for three randomly selected firms (with *gvkey* = 7906, 30501, and 126554) in Figure 1.

Figure 1. Examples of firms' distribution of correlation of share price with the 629 accounting metrics



Here we can empirically observe idiosyncrasies among the three firms. The firm with *gvkey* 126554 has 12 metrics that are highly correlated with its share price, and many negative correlations. The firm with *gvkey* 7906 has many metrics that exhibit a moderately high correlation with share price compared to the other two firms.

We explore how various correlations feature among the firms. For this analysis only, we remove annual data and "near-duplications" of variables for a cleaner assessment. A "near-duplication" occurs when two or more variables are in the list but they are essentially adjusted versions of one another, where the adjustment might be basic versus diluted, pre-tax versus after-tax, preliminary versus final, etc. For example, the following three variants might concurrently appear in the top 10 list because they are closely related to each other: *Gain/Loss basic EPS effect*, *Gain/Loss diluted EPS effect*, *Gain/Loss pre-tax* (scaled), *Gain/Loss preliminary*. This reduces our analysis in this section to 151 variables. For each firm, we use their 10 years of quarterly data to calculate share price correlation with each of these 151 accounting variables. We then rank these variables for each firm and select their top 10 values to get a sense of how consistent the correlations might be across firms.

We explore how frequently the traditional metrics (e.g., earnings, revenue, etc.) exhibit a high correlation with share price by examining how often they appear in the highest 10 correlations for a firm. Table 2 tabulates the 20 variables that appear most frequently in a firm's top 10 list.

Table 2. Variables most frequently appearing in a firm's top 10 list of correlations

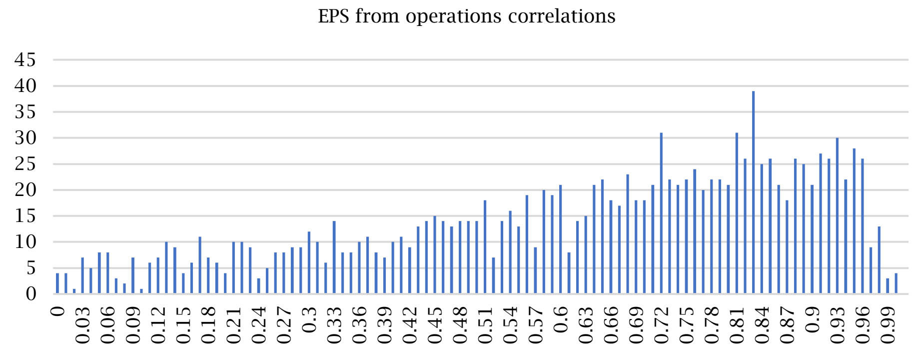
No.	*Variable*	*No. of firms*	*Mean absolute correlation*
1	Common/Ordinary equity -Total	772	0.83
2	Options - Fair value of options granted	758	0.86
3	EPS from operations	745	0.81
4	Unadjusted retained earnings	703	0.85
5	Retained earnings	666	0.84
6	Stockholders equity - Total	658	0.83
7	Stockholders equity > Parent > Index fundamental > Quarterly	615	0.83
8	Depreciation, depletion and amortization (accumulated)	604	0.84
9	Nonrecurring income taxes - After-tax	575	0.93
10	Life of options - Assumption (# yrs)	535	0.88
11	Property, plant, and equipment - Total (gross) - Quarterly	525	0.84
12	Common shares used to calculate EPS	506	0.83
13	Common shares used to calculate EPS - Fully diluted	500	0.82
14	Common shares issued	496	0.83
15	Volatility - Assumption (%)	494	0.86
16	Assets - Total	466	0.85
17	EPS - Excluding extraordinary items	461	0.78
18	Common shares for diluted EPS	446	0.82
19	S&P core earnings	435	0.82
20	Other special items after-tax	431	0.93

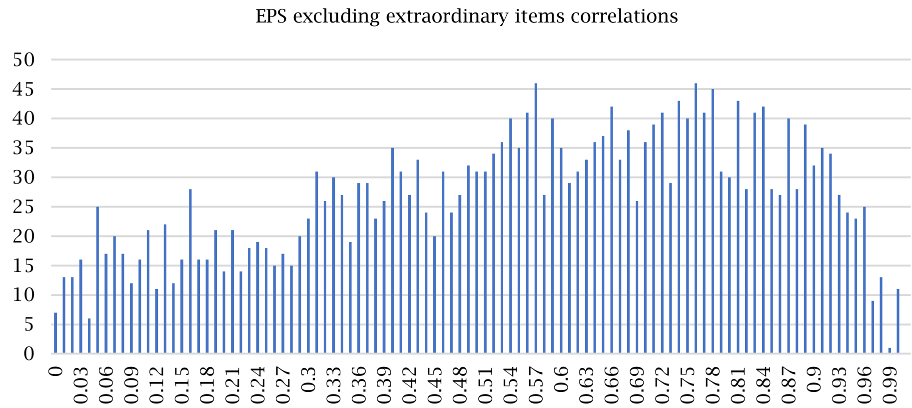
Table 2 suggests there are many important accounting metrics associated with share price. Although three variants of earnings often appear as a top 10 variable (in positions 3, 17, and 19), they appear at most in the top 10 lists of 745 out of 3,365 firms (22% of firms). Other contenders are variants of the book value of equity or net assets (in positions 1, 4, 5, 6, and 7), volatility (in positions 2, 10, and 15), and unusual items (in positions 9 and 20). Moreover, out of the 151 variables analyzed, a total of 144 (95%) appear at least once in a firm's top 10 list (not tabulated), which further suggests a wide variation in how firms connect with their share price.

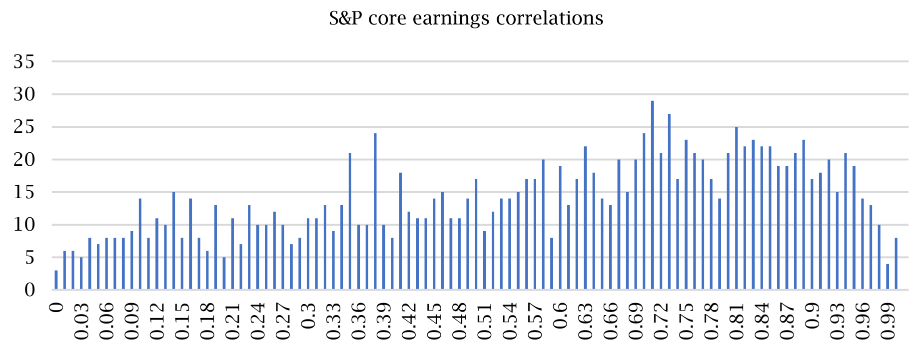
Figure 2 shows the distribution of correlations for the three variants of earnings identified in Table 2. Although the absolute correlations are skewed towards high values, a large proportion of firms have share prices that are not strongly connected to earnings. This supports our research objective of exploring how other accounting metrics are associated with share price.

Whereas Table 2 explores how often metrics are in the top 10 correlations (capturing both the dimensions of magnitude and frequency) we note that, once we remove the condition that it be a measure in a firm's "top 10", the earnings constructs become less important. We examine this in Table 3, which tabulates the variables with the highest mean absolute value of correlations (Panel A). This is followed by the highest positive (Panel B) and negative (Panel C) mean correlations, along with the number of firms included in the calculation.

Figure 2. Distribution of 3,365 firm correlations of the earnings metrics (from Table 2) with share price*







*Notes: * The total sample is not always 3,365 since some firms are missing certain Compustat variants of earnings.*

Table 3. Highest values of correlations (corr.) (Part 1)

No.	***Variable***	***Mean abs. (corr.)***	***Mean (corr.)***
Panel A: Absolute value of correlations			
1	Nonrecurring income taxes - After-tax	0.93	-0.05
2	Gain/Loss on sale (Core earnings adjusted) after-tax	0.93	-0.04
3	Other special items after-tax	0.93	-0.06
4	Gain/Loss after-tax	0.92	-0.04
5	Cost & earnings in excess of billings	0.92	0.06
6	Billings in excess of cost & earnings	0.91	0.00
7	Unbilled receivables - Quarterly	0.91	0.02
8	Extinguishment of debt after-tax	0.91	0.01
9	Acquisition/merger after-tax	0.88	-0.11
10	Restructuring cost after-tax	0.88	0.02
11	Life of options - Assumption (# yrs)	0.88	0.02
12	Dividend rate - Assumption (%)	0.88	-0.18
13	Risk free rate - Assumption (%)	0.87	0.06
14	Options - Fair value of options granted	0.86	0.60
15	Volatility - Assumption (%)	0.86	-0.27
16	Liabilities and stockholders equity - Total	0.86	0.79
17	Dividends per share - Pay date - Quarter	0.85	0.81
18	Capital surplus/Share premium reserve	0.85	0.48
19	Assets - Total	0.85	0.78
20	Unadjusted retained earnings	0.85	0.71

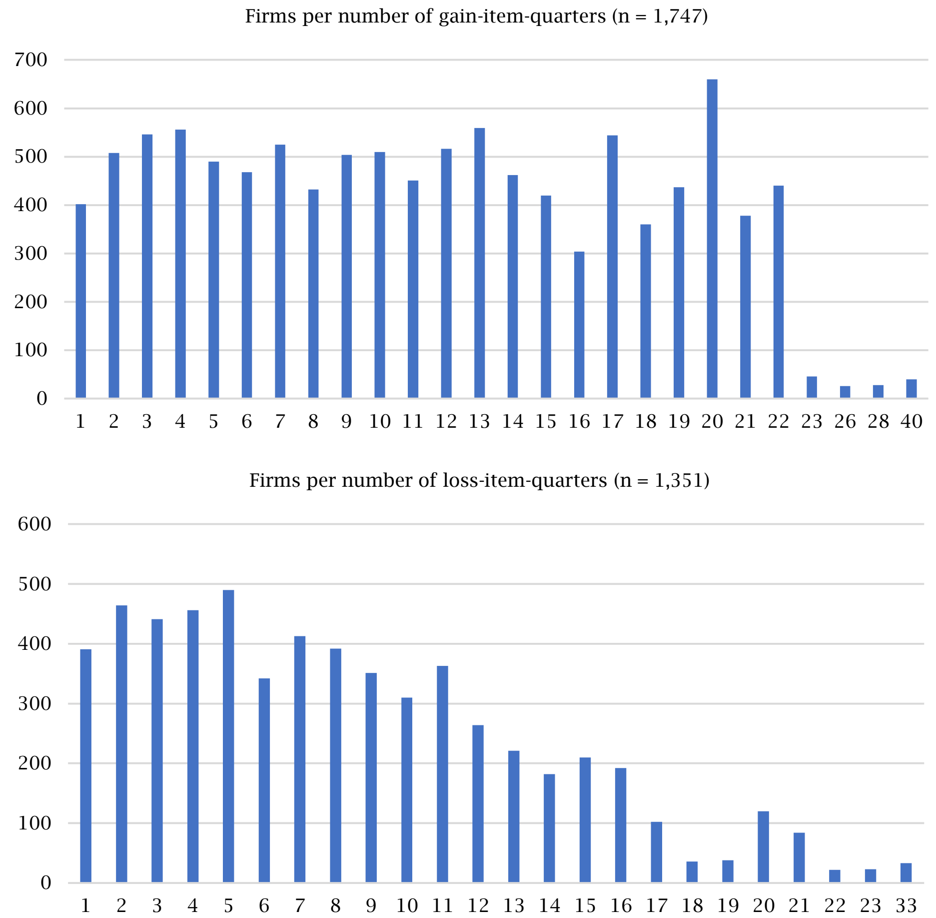
Table 3. Highest values of correlations (corr.) (Part 2)

No.	Variable	Mean (corr.)	No. of firms
Panel B: Highest positive correlations			
1	Cost & earnings in excess of billings	0.92	105
2	Gain/Loss after-tax	0.92	185
3	Nonrecurring income taxes - After-tax	0.90	296
4	Billings in excess of cost & earnings	0.90	135
5	Extinguishment of debt after-tax	0.89	169
6	Other special items after-tax	0.89	224
7	Gain/Loss on sale (core earnings adjusted) after-tax	0.89	168
8	Unbilled receivables - Quarterly	0.88	95
9	Acquisition/merger after-tax	0.87	126
10	Dividend rate - Assumption (%)	0.86	100
11	Life of options - Assumption (# yrs)	0.85	317
12	Property, plant and equipment - Total (gross) - Quarterly	0.85	459
13	Liabilities and stockholders equity - Total	0.84	499
14	Depreciation, depletion and amortization (accumulated)	0.83	494
15	Dividends per share - Pay date - Quarter	0.83	359
16	Options - Fair value of options granted	0.83	752
17	Assets - Total	0.83	571
18	Risk free rate - Assumption (%)	0.82	236
19	Liabilities - Total and noncontrolling interest	0.82	296
20	Dividends per share - Ex-date - Quarter	0.82	429
Panel C: Highest negative correlations			
1	Nonrecurring income taxes - After-tax	-0.82	466
2	Cost & earnings in excess of billings	-0.82	137
3	Billings in excess of cost & earnings	-0.81	192
4	Gain/Loss after-tax	-0.80	306
5	Unbilled receivables - Quarterly	-0.80	137
6	Other special items after-tax	-0.79	361
7	Options - Fair value of options granted	-0.77	188
8	Order backlog	-0.77	37
9	Dividend rate - Assumption (%)	-0.76	310
10	Gain/Loss on sale (core earnings adjusted) after-tax	-0.75	299
11	Extinguishment of debt after-tax	-0.75	262
12	Restructuring cost after-tax	-0.75	213
13	Volatility - Assumption (%)	-0.74	871
14	Liabilities and stockholders equity - Total	-0.72	49
15	Life of options - Assumption (# yrs)	-0.72	648
16	Assets - Total	-0.71	56
17	Common shares used to calculate EPS - 12 months moving (MM)	-0.69	1010
18	Common shares outstanding	-0.69	922
19	Com shares for diluted EPS	-0.69	852
20	Common shares used to calculate EPS - Fully diluted - 12 MM	-0.68	924

We note from Table 3 a greater emphasis on unusual and nonrecurring items, such as *Nonrecurring income taxes* and *Gain/Loss* numbers. *Other special items* also figure more prominently, appearing highly positive for 224 firms and highly negative for 361 firms. We perform sensitivity analysis on the results from Tables 2 and 3 (not tabulated). When considering all earnings metrics concurrently, and only examining firms with non-missing values of this construct (n = 3,311 firms), the earnings metrics jump to position 2, with 776 firms including it in their top 10. However, they still fail to make any of the lists in Table 3. Although this result is surprising, it confirms that a one-size-fits-all approach to valuation does not work, and idiosyncrasies should be considered.

We dive deeper into one of the correlations from Table 3 to explore the emphasis on the *Gain/Loss* items. For each firm in our sample, we count the number of quarters where they included a gain or loss as a line item in their statement of income. Where gains are concerned, we find 1,747 of the 3,365 firms report a gain item in at least one of their 40 quarters. For these firms, the mean (median) number of quarters where a gain is reported is 6.1 (4), and the range is 1–40 quarters. Where losses are concerned, we find 1,351 firms report a loss item in at least one quarter, with the mean (median) number of quarters being 4.3 (3) and a range of 1–33 quarters. These firms accumulate gains (losses) per share amounts with a mean of 0.932 (-0.237), a median of 0.126 (-0.034), and ranges from nearly zero to 102.68 (-22.74). Therefore, losses seem to be relatively less prevalent and at lower amounts. The distribution of firms reporting gains/loss items is shown in Figure 3.

Figure 3. Histogram of reported gain items or loss items in 1 to 40 quarters



To better understand the impact of *Gains/Losses* on share price, we select only the firms with at least eight quarters of *Gains/Losses*, reducing our sample from 1,747 to 495 for firms with gain items and from 1,351 to 240 for firms with loss items. We re-run the most frequent correlations for these firms, expecting *Gains/Losses* to prevail. As shown in Table 4, for the subset of firms that frequently report gain/loss items, their share price is associated with measures of volatility such as *Options*, *Volatility Assumptions*, and the *Risk free rate*.

Table 4. Variables most frequently appearing in a firm's top 10 list

No.	***Firms with 8+ quarters of gains (n = 495)***	***Firms with 8+ quarters of losses (n = 240)***
1	Income taxes - Deferred	Income taxes - Deferred
2	Depreciation, depletion and amortization (accumulated)	Options - Fair value of options granted
3	Life of options - Assumption (# yrs)	Life of options - Assumption (# yrs)
4	Property, plant and equipment - Total (gross) - Quarterly	Depreciation, depletion and amortization (accumulated)
5	Options - Fair value of options granted	Property, plant and equipment - Total (gross)
6	Unadjusted retained earnings	Risk free rate - Assumption (%)
7	Volatility - Assumption (%)	Volatility - Assumption (%)
8	Common/Ordinary equity - Total	Deferred taxes - Balance sheet
9	Risk free rate - Assumption (%)	Unadjusted retained earnings
10	Retained earnings	Common shares used to calculate EPS - 12 MM
11	Deferred taxes - Balance sheet	Accrued expenses
12	Stockholders equity - Total	Common shares issued
13	Accrued expenses	Retained earnings
14	Stockholders equity > Parent > Index fundamental > Quarterly	Common shares used to calculate EPS - Fully diluted - 12 MM
15	Acquisition/Merger after-tax	Common/Ordinary equity - Total
16	Capital surplus/Share premium reserve	Stockholders equity - Total
17	Income before extra items - Adjusted for common stock equivalents - 12MM	Other special items after-tax
18	Current deferred tax asset	Common shares outstanding
19	Dividend rate - Assumption (%)	Acquisition/Merger after-tax
20	Common shares issued	Total fair value assets

We re-run the analysis from Table 2 with different time periods (not tabulated). We look for differences when performing the analysis on firms in the early years (from 2009 to 2013) versus later years (2014 to 2019). The results are consistent with Table 2 except that *Total assets* and *S&P core earnings* are dropped, while *Accrued expenses* is picked up in early years, and both subsets pick up slightly different variants of *Taxes* and *Options*.

Finally, when each of the four quarters is analyzed separately (not tabulated), we find *Earnings* metrics rank higher in quarters 1, 2, and 3, while *Nonrecurring items* have a stronger association with share price in quarter 4. Each quarter picks up some items not featured in the full dataset in Table 2. *Sales* and *Revenue* are important in quarters 1 and 4, while *Mergers/Acquisitions* and *Operating income* metrics are highly correlated with share price in quarters 2 and 3. *Dividends per share* feature in the first three quarters but is no longer important in quarter 4.

4.2. Vector selection and cluster formation

Our preliminary analysis of correlations suggests share price may be connected to different accounting representations across firms and over time. However, the extant literature suggests earnings (and sometimes revenues) are sufficient. In this section, we use silhouette coefficient (SC) to guide us to the optimal trade-off between having many idiosyncratic measures or just one or two general ones.

We begin by considering the 10 accounting metrics listed in Table 5, which represent most of the accounting dimensions in Tables 2 and 3.

Table 5. Accounting metrics (per share)*

No.	*Accounting metric*	*Variable Compustat*
1	Dividends	*DVPSXQ*
2	Earnings	*EPSPIQ*
3	Revenue	*REVTY*
4	Earnings from Operations	*OPEPSQ*
5	Net Operating Cash Flow	*OANCFY*
6	Property Plant and Equipment	*PPENTQ*
7	Intangible Assets	*INTANQ*
8	Retained Earnings	*REQ*
9	Common Equity	*CEQQ*
10	Net Assets = Total Assets – Total Liabilities	*ATQ — LTQ*

*Notes: * Variables in levels have been scaled by the number of shares the firm used to calculate its basic EPS.*

These metrics capture various ways of representing firm value, as follows: 1 *Dividends* reflect cash flowing immediately to shareholders; 2 *Earnings* and 3 *Revenue* represent future cash flows that should eventually be accessible to shareholders; 4 *Earnings from operations* and 5 *Net Operating cash flow* provide a going concern perspective, as profits are re-invested in assets; and the last five constructs explore how share price connects to *Asset* value (6 and 7) and *Net Asset* value (8, 9, 10).

Our dataset is now 3,365 firm-specific vectors of values for the 10 metrics listed in Table 5. We calculate the SC obtained from allocating firms to clusters, iteratively, with the cluster number ranging from 2 to 10. According to Table 6, the highest SC value is 0.228, which is quite low. The goal is an SC as close to 1 as possible. This suggests that using all 10 accounting metrics prevents us from creating sufficiently distinct clusters for analysis.

Table 6. Silhouette coefficients for clusters based on the 10 accounting metrics in Table 5

No. of clusters	*SC*
2 clusters	0.228
3 clusters	0.173
4 clusters	0.171
5 clusters	0.160
6 clusters	0.157
7 clusters	0.165
8 clusters	0.145
9 clusters	0.142
10 clusters	0.146

We, therefore, approach the cluster formation from a different angle by examining two metrics at a time. The SC is estimated for each accounting construct listed in Table 5, in a pairwise fashion. Table 7 reports pairings that yield SC values equal to or greater than 0.5. The most differentiating constructs appear to be 8 (*Retained earnings*) and 3 (*Revenue*), followed by 2 (*Earnings*) and 10 (*Net assets*). We examine how vectors with these four elements would perform for clusters $K = 1$ to 10. Since retained earnings and net assets are somewhat similar, we also examine the clustering performance without Net Assets, hence with three elements.

Table 8 tabulates the SC from clustering our sample along vectors with either three or four elements. The highest value of 0.441 is achieved with two clusters of three elements; therefore we choose to implement this structure. We note that the SC is slightly lower than some of the values reported in Table 7, but although the clustering will be slightly less distinctive, it allows for insight along three interesting dimensions of a firm's reported numbers: *Revenue*, *Earnings*, and *Retained earnings*.

Table 7. Pairings of accounting constructs that resulted in values of SC of at least 0.50

No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1								0.50		
2				0.50				0.50		
3						0.50		0.54	0.52	0.51
4										
5								0.50		
6										
7										
8									0.52	0.51
9										0.63
10										

Notes: see Table 5 for the accounting metric associated with each number.

Table 8. SC for clusters based on vectors of 3 accounting elements (*Retained earnings*, *Revenue*, *Earnings*) or 4 elements (also including *Net assets*)

No. of clusters	*SC 3 elements*	*SC 4 elements*
2 clusters	0.441	0.391
3 clusters	0.295	0.274
4 clusters	0.283	0.293
5 clusters	0.290	0.250
6 clusters	0.279	0.258
7 clusters	0.275	0.253
8 clusters	0.272	0.254
9 clusters	0.274	0.236
10 clusters	0.249	0.241

4.3. Cluster analysis

Our cluster analysis groups n = 2,343 firms in Cluster 1 and n = 1,022 firms in Cluster 2. We examine the distribution of our accounting metrics in Figures 4a, 4b, and 4c.

Figure 4a. Boxplots of values of correlations of share price to revenue per share, by cluster

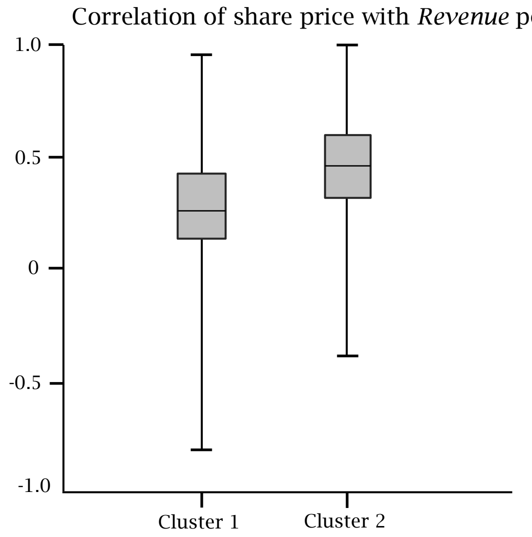


Figure 4b. Boxplots of values of correlations of earnings per share, by cluster

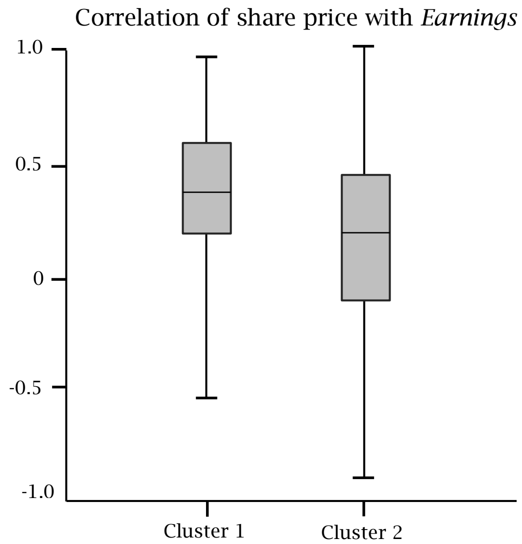
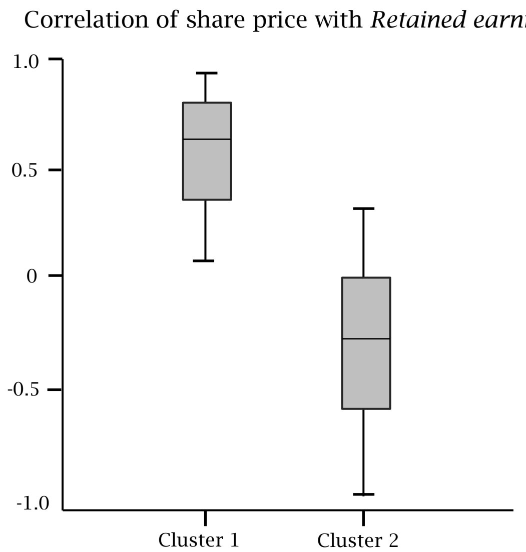


Figure 4c. Boxplots of values of correlations of retained earnings per share, by cluster



We can infer a few differences in the cluster groupings from the boxplots in Figures 4a, 4b, and 4c (where the differences in means are significant at the 1% level). Figure 4a on *Revenue* shows that Cluster 1 exhibits a larger range of correlation values than Cluster 2, with a slightly lower overall mean and median. The reverse pattern is observed in Figure 4b on *Earnings*. However, the difference between the clusters is most pronounced when examining *Retained earnings* in Figure 4c. Here we see that share prices of firms in Cluster 1 have a strong and positive connection to *Retained earnings*, while the relationship is weaker and negative for firms in Cluster 2.

This could be interpreted as the firms in Cluster 1 (Cluster 2) have share prices that are strongly and consistently connected to their *Earnings* (*Revenues*) and exhibit (do not exhibit) the expected positive correlation with *Retained earnings*. For ease of explication, we call firms in Cluster 1 "conventional" and firms in Cluster 2 "unusual."

4.4. Differentiating between clusters

Our next step is to explore characteristics that may have led firms to get grouped into their respective clusters. We perform an analysis based on 4 statistics: means, medians, ranges, and kurtosis. For each statistic, the 10 variables with the highest difference (between the two clusters) are shown in Table 9.

Table 9. Differentiating variables between clusters

Variable	*Cluster 1*	*Cluster 2*
Panel A: Means		
Preferred/preference stock - Nonredeemable	5.972	0.056
Fair value of options granted	0.790	59.192
Maintenance expense	0.105	0.003
Extraordinary items and discontinued operations (*CFlow*)	0.002	0.046
Other stockholders - Equity adjustments	0.033	0.945
Income before extraordinary items	2.136	-0.087
Income taxes - Deferred	0.010	-0.251
Quarterly EPS	0.012	0.271
Long-term debt	0.067	0.003
Dividends - Preferred/Preference	0.115	0.006
Panel B: Medians		
Treasury stock	0.044	0.001
Maintenance expense	0.076	0.003
Current deferred tax asset	0.031	0.002
Discretionary accruals	-0.112	-0.010
Other long-term liabilities	0.009	0.076
Income taxes	0.094	0.012
Inventory - Decrease (increase)	-0.008	0.000
Long-term debt	0.013	0.002
Inventory - Finished goods	0.097	0.012
Dividends - Common	0.070	0.009
Panel C: Ranges		
Implied option EPS diluted	2.682	0.000
Preferred/Preference stock - Nonredeemable	19.791	2910.842
Long-term debt	0.658	0.005
Options - Fair value of options granted	455.389	46243.786
Dividends - Preferred/Preference	1.198	110.579
Excise taxes	367.230	5.770
Life of options - Assumption (# yrs)	550.606	8.997
Long-term debt - Current	0.527	0.009
Receivables - Trade	751.406	38294.488
Interest paid - Net	26.852	1189.076
Panel D: Kurtosis		
Life of options - Assumption (# yrs)	1662.298	1.392
Other assets and liabilities - Net change (*CFlow*)	45.069	0.215
Utility plant - Gross additions (*CFlow*)	19.868	-0.130
Gross income (Income before interest charges)	4.748	-0.084
Goodwill (net)	22.407	800.135
Provision for loan/Asset losses	23.502	0.705
Cumulative adjustment factor by ex-date	31.330	1004.621
Interest expense - Total	18.180	0.702
Common stock - Carrying value	800.674	31.312
Equity in earnings (income statement) - Unconsolidated subsidiaries	11.513	-0.513

The means analysis reported in Panel A shows that firms in Cluster 1 have more *Nonredeemable stock*, *Debt*, and *Deferred taxes*, as well as higher *Income* (before extraordinary items) and *Maintenance expense*. Firms in Cluster 2 have higher valued stock *options* and *Equity adjustments* and record more *Extraordinary items*. Surprisingly, the quarterly EPS are higher for firms in Cluster 2 than Cluster 1. The medians in Panel B show firms in Cluster 1 additionally have more *Treasury stock* and *Finished goods inventory*, pay higher *Dividends* on the common stock, and have more negative *Discretionary accruals*. Taken together, this suggests the "conventional" firms tend to be manufacturing or merchandising, with regular income, traditional debt, and equity capital and managed earnings (through accruals). The "unusual" firms experience more extraordinary items and provide compensation through stock options.

Panels C and D offer insight into the spread of the variables within clusters. The ranges of values suggest firms in Cluster 2 exhibit more variation in the accounting metrics listed, except for the *Life of options* and the level of *Excise taxes*. The deviations in kurtosis suggest certain variables have distributions with much thicker tails than a normal distribution, such as *Life of options* and the *Carrying value of common stock* (for Cluster 1) and *Goodwill* and the *Adjustment factor* (for Cluster 2) as a result of changes due to new stock issuance or stock splits.

Taken together, this suggests wider variability in the "unusual" firm fundamentals and different distributions with respect to equity variables and goodwill.

None of the firm-specific variables that we added from Table 1 feature as having extremely disparate means in Panel A of Table 9, although "conventional" firms record relatively more negative discretionary accruals than "unusual" firms according to the medians in Panel B. Keep in mind that Table 9 tabulates only the most extreme differences between the two clusters. There are likely many others among the 600+ variables that are significantly different.

With this in mind, we revisit the variables in Table 1 which were based on prior literature. Although none of these variables appeared in our analysis thus far, we explore them in Table 10 to see whether they differ by cluster.

Table 10. Test of significance in the difference in key constructs between clusters

*Construct**	*Cluster 1*	*Cluster 2*	*p-value*
ADA	-0.426	-1.766	0.384
pREV	0.96	0.90	***0.001
pEPS	0.57	0.48	***0.001
pDIV	0.51	0.38	***0.001
pOPCFLOW	0.67	0.67	0.924
pFINCFLOW	0.65	0.59	0.125
pINVCFLOW	0.60	0.59	0.896
growREV	0.47	0.36	***0.001
growFLOW	0.00	0.01	0.331
growFIN	0.01	0.02	0.235
Dilution	-0.02	-0.01	**0.032
Age	20	19	**0.016

*Notes: * See construct definition in Table 1.*

The p-values in Table 10 suggest the mean persistence of revenue (*pREV*), earnings per share (*pEPS*), and dividends (*pDIV*) is significantly higher for firms in Cluster 1. Firms in Cluster 2 are slightly younger (*Age*) and exhibit less growth in revenue (*growREV*) as well as a higher dilution effect from their financial instruments (*Dilution*). Taken together, this suggests different capital structures between the two groups and that the "conventional" firms are more stable and possibly more established than their counterparts in Cluster 2, which may have led to a stronger relationship between their earnings and share price.

The clusters exhibit a few significant differences in industry composition. We allocate firms to industry groups as per Table 11a. Table 11b reports p-values from a Chi-square test of differences (from expected counts) in each industry. The expected counts are obtained by prorating the total from each cluster sample ($n = 2,331$ for Cluster 1 and $n = 1,021$ for Cluster 2) according to the overall likelihood of a firm being in that particular industry (% of total). Finance firms appear more often than expected in Cluster 1, while retail firms are slightly more likely to be in Cluster 2.

Table 11a. Industry groupings

Range of SIC Codes	*Industry*
0100–0999	Agriculture, forestry and fishing
1000–1499	Mining
1500–1799	Construction
1800–1999	Not used
2000–3999	Manufacturing
4000–4999	Transportation, communications, electric, gas and sanitary service
5000–5199	Whole trade
5200–5999	Retail trade
6000–6799	Finance, insurance and real estate
7000–8999	Services
9100–9729	Public administration
9900–9999	Nonclassifiable

Table 11b. Industry test of significant differences

Industry	*Number of observations*	*Cluster 1*	*Cluster 2*	*Total*	*% of total*	*Chi-square p-value*
Construction	Actual	31	10	41	1%	0.398
	Expected	29	12			
Finance	Actual	609	218	827	25%	*** 0.010
	Expected	575	252			
Manufacturing	Actual	853	410	1263	38%	0.122
	Expected	878	385			
Mining	Actual	96	46	142	4%	0.616
	Expected	99	43			
Retail	Actual	111	64	175	5%	* 0.079
	Expected	122	53			
Service	Actual	325	155	480	14%	0.383
	Expected	334	146			
Wholesale	Actual	306	118	424	13%	0.239
	Expected	295	129			
Cluster sum		2,331	1,021			

Our dataset has several text variables that are not included in the analysis. Our last investigation explores clustering differences among some of these, namely city, state, balance sheet presentation, and accounting standards. The firms in our sample are headquartered in 1,136 cities. Most of these cities only contain only a few firms, so we also categorize the observation as "few in city" if it is located in a city with fewer than 9 firms and "many in city" if at least 10 other firms are headquartered in their city. We look for differences in distribution by cluster (not tabulated) and find 1) no difference in the portion of firms in "few in city" vs. "many in city", 2) significantly more Cluster 1 firms in Houston and Minneapolis, and 3) significantly more Cluster 2 firms in Denver and San Diego. Otherwise, the clusters are fairly represented among cities. We do the same for states, using a cut-off value of 100 to differentiate between "few in state" and "many in state". We find 1) a disproportionate amount of Cluster 2 firms residing in states with relatively few firms, 2) Cluster 2 firms overrepresented in California and New Jersey, and 3) Cluster 1 firms overrepresented in Texas.

The firms in our sample are incorporated or legally registered in 49 different countries, with the majority (83%) being from the United States. The next five most common countries of origin are Canada (4%), China (2%), Israel (2%), Bermuda (1%), and the British Virgin Islands (1%). We find a slightly higher-than-expected number of Cluster 2 firms are from Canada but no differences otherwise.

Where accounting standards followed is concerned, the majority of firms (89%) follow their "domestic standards", while fewer followed "domestic standards generally in accordance with US GAAP" (7%) or "domestic standards in accordance with Inter-Agency Standing Committee (IASC)" (4%), and the distribution was similar for both clusters.

5. CONCLUSION

In this paper, we revisit the notion that accounting earnings are always representative of firm value. We explore 628 other accounting metrics and demonstrate that firms experience a wide variety of relationships between their accounting metrics and share prices. This supports the notion that earnings are not always the best indicator of value, and we need to understand when to consider an alternative accounting measure. For each of the 3,365 firms in our sample, we calculate correlations between share price and their accounting metrics. Using silhouette coefficient (SC) analysis, we examine how to best structure a cluster analysis that groups our sample of firms in a way that captures their different share-price relationships. Three metrics are identified through this process: revenue, earnings, and retained earnings, where the first two are consistent with prior literature but retained earnings is a novel addition.

Our analysis suggests firms fall nicely into two clusters based on these three metrics. Cluster 1, which we label "conventional" firms exhibits a strong association between share price and both earnings and retained earnings. Cluster 2, which we call "unusual" firms has share price more closely aligned with revenues and negatively related to retained earnings. Using several techniques to delve into how the two clusters are different, we observe that "conventional" firms tend to be more mature and established, have more persistent operations, and have a capital structure that is less reliant on stock options. They have higher net income, debt, dividends, and inventories, and more negative discretionary accruals. They are more likely to be in the finance industry. The "unusual" firms are more likely to be in retail, have more stock options and dilutive financial instruments, report more extraordinary items, and operate in states with fewer corporate headquarters but are overrepresented in California and New Jersey.

The main contribution of our research is an extension of previous work examining how accounting metrics map out to firm value. Our findings show that the relationship is not as simple as it has been proposed in the literature and that different variables can play a role for different firms. The implication is that a one-size-fits-all approach is unlikely to be effective in valuation. Based on our analysis, a better understanding of the firm characteristics should enable better matching of their accounting numbers to firm value. Our work is exploratory and provides a foundation for data-driven research on the complexities of valuation. Our analysis followed a few of many possible paths, therefore we have not completed an exhaustive assessment of the relationships that exist. Future research could include deeper dives into some of the correlations or further assessments of differentiating firm characteristics.

REFERENCES

1. Acar, G., & Coskun, A. (2020). A comparison of models for predicting discretionary accruals: A cross-country analysis. *The Journal of Asian Finance, Economics, and Business, 7*(9), 315–328. https://doi.org/10.13106/jafeb.2020.vol7.no9.315
2. Aranganayagi, S., & Thangavel, K. (2007). Clustering categorical data using silhouette coefficient as a relocating measure. In *Proceedings of the International conference on computational intelligence and multimedia applications (ICCIMA 2007)* (Vol. 2, pp. 13–17). https://doi.org/10.1109/ICCIMA.2007.328
3. Bailey, P., Brown, P. J., Potter, M., & Wells, P. A. (2008). A practical comparison of firm valuation models: Cash flow, dividend and income. *JASSA: The Finsia Journal of Applied Finance, 2*, 22–28. Retrieved from https://opus.lib.uts.edu.au/bitstream/10453/10199/1/2007003268OK.pdf
4. Bao, B.-H., & Bao, D.-H. (2004). Income smoothing, earnings quality and firm valuation. *Journal of Business Finance & Accounting, 31*(9–10), 1525–1557. https://doi.org/10.1111/j.0306-686X.2004.00583.x
5. Barker, R. (2004). Reporting financial performance. *Accounting Horizons, 18*(2), 157–172. https://doi.org/10.2308/acch.2004.18.2.157
6. Barker, R., & Imam, S. (2008). Analysts' perceptions of 'earnings quality'. *Accounting and Business Research, 38*(4), 313–329. https://doi.org/10.1080/00014788.2008.9663344
7. Barth, M. E., Li, K., & McClure, C. (2021). *Evolution in value relevance of accounting information* (Stanford University Graduate School of Business Research Paper No. 17-24). https://doi.org/10.2139/ssrn.2933197
8. Bradshaw, M. T. (2000). *The articulation of sell-side analysts' earnings forecasts, common stock valuations and investment recommendations* (PhD thesis, University of Michigan). Retrieved from https://abidahasker.xyz/download/Z3kfAQAAMAAJ-the-articulation-of-sell-side-analysts-earnings-forecasts-common-stock-valuations-and-investment-recommendations
9. Bradshaw, M. T. (2004). How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review, 79*(1), 25–50. https://doi.org/10.2308/accr.2004.79.1.25

10. Chandra, U., & Ro, B. T. (2008). The role of revenue in firm valuation. *Accounting Horizons, 22*(2), 199–222. https://doi.org/10.2308/acch.2008.22.2.199
11. Cong, Y., Du, H., & Vasarhelyi, M. A. (2018). Are XBRL files being accessed? Evidence from the SEC EDGAR log file dataset. *Journal of Information Systems, 32*(3), 23–29. https://doi.org/10.2308/isys-51885
12. Conrod, J. D., & Cumby, J. (2016). On-line gaming, financial reporting, and audit: Chester Games Corp. *Issues in Accounting Education, 31*(4), 431–437. https://doi.org/10.2308/iace-51221
13. Darrough, M., & Ye, J. (2007). Valuation of loss firms in a knowledge-based economy. *Review of Accounting Studies, 12*(1), 61–93. https://doi.org/10.1007/s11142-006-9022-z
14. Demirakos, E. G., Strong, N. C., & Walker, M. (2004). What valuation models do analysts use? *Accounting Horizons, 18*(4), 221–240. https://doi.org/10.2308/acch.2004.18.4.221
15. Dutta, S. K., Caplan, D. H., & Marcinko, D. J. (2014). Growing pains at Groupon. *Issues in Accounting Education, 29*(1), 229–245. https://doi.org/10.2308/iace-50595
16. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD, 96*(34), 226–231. Retrieved from http://www.cs.ecu.edu/~dingq /CSCI6905/readings/DBSCAN.pdf
17. Ferreira, P. (2020). *The liquidity, precision, and comparability effects of ASC 606: Revenue from contracts with customers*. https://doi.org/10.2139/ssrn.3522765
18. Frades, I., & Matthiesen, R. (2010). Overview on techniques in cluster analysis. In R. Matthiesen (Ed.), *Bioinformatics methods in clinical research*. (Vol. 593, pp. 81–107). https://doi.org/10.1007/978-1-60327-194-3_5
19. Fullana, O., González, M., & Toscano, D. (2021). The role of assumptions in Ohlson model performance: Lessons for improving equity-value modeling. *Mathematics, 9*(5), 513. https://doi.org/10.3390/math9050513
20. Guha, S., Rastogi, R., & Shim, K. (1998). CURE: An efficient clustering algorithm for large databases. *ACM SIGMOD Record, 27*(2), 73–84. https://doi.org/10.1145/276305.276312
21. Gupta, M. C., & Huefner, R. J. (1972). A cluster analysis study of financial ratios and industry characteristics. *Journal of Accounting Research, 10*(1), 77–95. https://doi.org/10.2307/2490219
22. Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society, 28*(1), 100–108. https://doi.org/10.2307/2346830
23. Hauser, R., & Thornton, J. H., Jr. (2017). Dividend policy and corporate valuation. *Managerial Finance, 43*(6), 663–678. https://doi.org/10.1108/MF-05-2015-0157
24. Ho, C.-T. B., Liao, C.-K., & Kim, H.-T. (2011). Valuing internet companies: A DEA-based multiple valuation approach. *Journal of the Operational Research Society, 62*(12), 2097–2106. https://doi.org/10.1057/jors.2010.191
25. Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies, 23*(10), 3773–3811. https://doi.org/10.1093/rfs/hhq053
26. Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy, 124*(5), 1423–1465. https://doi.org/10.1086/688176
27. Impink, J., Lubberink, M., van Praag, B., & Veenman, D. (2012). Did accelerated filing requirements and SOX Section 404 affect the timeliness of 10-K filings? *Review of Accounting Studies, 17*(2), 227–253. https://doi.org/10.1007/s11142-011-9172-5
28. Issa, H., & Vasarhelyi, M. A. (2011). *Application of anomaly detection techniques to identify fraudulent refunds*. https://doi.org/10.2139/ssrn.1910468
29. Jegadeesh, N., & Livnat, J. (2006). Revenue surprises and stock returns. *Journal of Accounting & Economics, 41*(1–2), 147–171. https://doi.org/10.1016/j.jacceco.2005.10.003
30. King, A. M. (2000). Valuing red-hot internet stocks. *Strategic Finance, 81*(10), 28–34.
31. Lacina, M. J. (1998). *Accounting conservatism, firm growth, earnings persistence, and earnings-based valuation* (Ph.D. dissertation, Purdue University). Retrieved from https://docs.lib.purdue.edu/dissertations/AAI9953737/
32. Markman, J. (2017, May 23). The Amazon era: No profits, no problem. *Forbes*. Retrieved from https://www.forbes.com/sites/jonmarkman/2017/05/23/the-amazon-era-no-profits-no-problem
33. McEnroe, J. E., & Mindak, M. (2020). An empirical analysis of an application of an alternative measurement model on international accounting standard 33, earnings per share. *Accounting Research Journal, 33*(2), 381–390. https://doi.org/10.1108/ARJ-01-2019-0002
34. NASDAQ. (n.d.). *Stock quotes, stock charts, and stock research*. Retrieved from https://www.nasdaq.com/market-activity/stocks
35. Pope, P. F., & Wang, P. (2005). Earnings components, accounting bias and equity valuation. *Review of Accounting Studies, 10*(4), 387–407. https://doi.org/10.1007/s11142-005-4207-4
36. Raisinghani, M., Shoemaker, B., & Schkade, L. (2004). Accounting issues in electronic commerce: An USA perspective regarding valuations and implications for corporate governance. *Journal of Applied Management Accounting Research, 2*(2), 53–68. Retrieved from https://www.proquest.com/docview/200861052
37. Romanova, N., Helms, M. M., & Takeda, M. (2012). Valuing internet companies: A more accurate, comprehensive financial model. *Academy of Entrepreneurship Journal, 18*(1), 35–53. Retrieved from https://cutt.ly/OTYbGu5
38. Saito, Y. (2012). The demand for accounting information: Young NASDAQ listings versus S&P 500 NYSE listings. *Review of Quantitative Finance and Accounting, 38*(2), 149–175. https://doi.org/10.1007/s11156-010-0223-y
39. Thiprungsri, S., & Vasarhelyi, M. A. (2011). Cluster analysis for anomaly detection in accounting data: An audit approach. *The International Journal of Digital Accounting Research, 11*, 69–84. https://doi.org/10.4192/1577-8517-v11_4
40. Tokic, D. (2002). *R&D, advertising and the market value of internet firms* (Ph.D. dissertation, University of Texas-Pan American). Retrieved from https://cutt.ly/mTTnjjH
41. Wang, W., Yang, J., & Muntz, R. (1997). STING: A statistical information grid approach to spatial data mining. *VLDB, 97*, 186–195. Retrieved from https://www.vldb.org/conf/1997/P186.PDF
42. Xu, L., & Cai, F. (2009). Before and after 2000: Revenue and high tech valuation. *Competitiveness Review, 19*(1), 26–35. https://doi.org/10.1108/10595420910929040
43. Xu, L., Cai, F., & Leung, C. K. (2007). The role of book value in high-tech valuation. *Advances in Competitiveness Research, 15*(1–2), 71–81.
44. Xu, L., Cai, F., & Paterson, W. (2016). Value relevance of earnings, book value, revenue, and R&D. *Business Review, 24*(1), 91–97. Retrieved from https://cutt.ly/dTTnWrc
45. Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. *ACM SIGMOD Record, 25*(2), 103–114. https://doi.org/10.1145/235968.233324