ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
Bachelor Thesis Economics & Business
Specialization: Financial Economics

Enhancing Stock Valuation Accuracy: A Five-Factor Regression Model for Selecting Comparable Companies

Author: Mushegh Harutyunyan
Student number: 617777
Thesis supervisor: Erik Fernau
Second reader: Fabrizio Core
Finish date: 24/07/2024

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second reader, Erasmus School of Economics or Erasmus University Rotterdam.

ABSTRACT

In this thesis, I study the effects of a newly developed five-factor regression model for selecting peers on the accuracy of stock valuation estimates. To study that, panel data on firms listed on the New York Stock Exchange (NYSE) from 2014 to 2019 is collected. The paper finds that this novel model for selecting peers offers improvements in the accuracy of valuation estimates when this approach is used within industries. Consequently, when the selected fundamental value drivers used in the regression model as explanatory variables are incorporated with the industry classification (ICOMP) the most accurate valuation estimates are obtained. Nevertheless, the warranted, or predicted multiple (WPE/WPB) alone does not yield very accurate valuation estimates showing similar results to the industry classification approach. Therefore, while both the WPE/WPB and the ICOMP are part of our model, only the latter consistently yields accurate valuation estimates and is recommended to be applied in practice.

TABLE OF CONTENTS

CHAPTER 1 Introduction

Bob Dylan's "Do not Think Twice, It is All Right" has endured as a brilliant song and a terrible principle for valuation—especially when it comes to multiples (Rehm & Karnani, 2023). Managers and finance practitioners should *always* think twice about multiples. The proper application of multiples and the correct selection of peer companies, can serve as a quick estimation tool and provide a comparison to validate the accuracy and reliability of the discounted cash flow model (Rehm & Karnani, 2023). Furthermore, research indicates that the multiples approach is the most commonly utilized technique for evaluating the value of stocks (Demirakos, Strong, & Walker, 2004). However, choosing the correct peer companies turns out to be very challenging. One famous method for choosing peer companies is the industry classification approach where the industry average for a particular multiple is calculated and compared to the multiple of the target company. However, the use of the industry average overlooks the fact that companies, even in the same industry, can have drastically different expected growth rates, returns on invested capital, and capital structures (Koller et al., 2005). As a result, often the multiples are misunderstood and misapplied by analysts (Koller et al., 2005). Because of the drawbacks of the industry classification approach, researchers developed new, alternative methods for selecting peer companies. One of these approaches, namely the regression approach by Bhojraj & Lee (2002), is the inspirational article of this paper and a model similar to the one developed by Bhojraj & Lee (2002) will be developed in this paper.

Researchers, particularly in the field of Finance have tested the existing methods and developed alternative or combined methods for selecting comparable companies. The goal was to examine the performance of existing methods in estimating the multiple of the target firm and to find methods or combinations of methods for selecting more comparable companies. The approach that yields the lowest absolute difference between the observed price and the predicted price of a target firm is considered to be the method that selects the closest, or most comparable companies. As noted by Lee et al. (2015) without any conceptual guidance on what constitutes an "industry", the industry classification approach relies on subjective judgment, hence, benchmarking through this approach can potentially be an inaccurate approximation. Similarly, Knudsen et al. (2017) argued its unreliability in selecting peers. Furthermore, Knudsen et al. (2017) find that peers selected using the SARD approach both within and across industries yield more accurate valuation estimates compared to the industry classification approach. Additionally, they found that the SARD within industry yields more accurate estimates compared to SARD across industries which suggests that the novel SARD method performs best when incorporated with the industry classification approach. These findings were robust across time, company size, and number of peers. Finally, Bhojraj & Lee (2002) ran a series of annual cross-sectional regressions of either EV/Sales or P/B ratio on eight explanatory variables using data from 1982 to 1998. To predict the company's multiple, the estimated coefficients from last year's regressions are used, in conjunction with each company's current year information, (Bhojraj & Lee, 2002). Peer firms are identified as those having the closest predicted, or so-called “warranted” multiple. The results show that comparable companies selected with this approach provide significant improvements over the peers selected based on other methods, such as the industry classification approach and size matches (Bhojraj & Lee, 2002).

While most studies focused on finding new methods for identifying comparable companies, no study modified the regression approach. The regression approach suggested by Bhojraj and Lee (2002), may be considered cumbersome to use by practitioners (Knudsen et al., 2017). This paper aims to alter the regression approach to make it easily applicable while producing highly accurate valuation estimates. The five fundamental value drivers suggested by Knudsen et al. (2017) are used as explanatory variables in the regression model instead of the eight used by Bhojraj and Lee (2002). This will ease the data collection procedure and regression analysis. The methodology for identifying peers is replicated from the paper by Bhojraj and Lee (2002). After identifying the peers and obtaining the results for the accuracy of valuation estimates for the target firm based on the identified peers, the robustness of the results across time is tested. Furthermore, the accuracy of valuation estimates obtained by the regression method is compared to the accuracy of valuation estimates obtained by the industry classification approach. Therefore, this thesis aims to study: *“How does the simplified regression approach developed in this paper for identifying peers influence the accuracy of stock valuation estimates of companies listed in NYSE in absolute terms and relative to other methods?”*

In the sample, I use companies that are traded on the New York Stock Exchange (NYSE) which mainly consists of large-cap companies but also includes mid-cap companies, and small-cap companies. NYSE mainly consists of United States stocks. The entire dataset is collected for each firm as of the 31st of December using the last twelve months (LTM) timeframe. Shocks to an industry or the broader market make the use of multiples and, generally, relative valuation fully inapt (Rehm & Karnani, 2023). Therefore, the data from 2020 onwards are not included in the sample because the COVID-19 pandemic and the following lockdown are one type of shock (Rehm & Karnani, 2023). The data for fundamental value drivers is obtained from Compustat and CRSP, and only the data for analysts’ forecasts of twelve months forward earnings per share is collected from I/B/E/S. For the regression approach, the dependent variable is the predicted, or warranted multiple, and the explanatory variables are the five variables used by Knudsen et al. (2017) which are empirical proxies for fundamental value drivers. A simple classic linear regression model (CLRM) is applied for analysis using a t-test to test the significance of the estimated coefficients.

I hypothesize that applying the simplified regression approach with five explanatory variables to identify peers will result in highly accurate valuation estimates. Furthermore, I hypothesize that the regression approach yields significant improvements in the precision of valuation estimates when compared to the industry classification approach. While industry classification assumes that companies operating within the same industry share fundamental value drivers, the regression approach regresses the multiples on these fundamentals and chooses peers based directly on these fundamentals. Therefore, I believe the regression approach implements a more detailed analysis which is why I anticipate that the regression approach may provide significant improvements in the precision of valuation estimates when compared to the industry classification approach.

The paper finds that the five-factor regression model for selecting comparable companies is significantly more accurate when applied within industries compared to applying it across industries, and compared to the industry classification approach. However, the second technique used in the paper, which involves applying the previous year's coefficients along with the current year's information to predict the multiple for the target firm (WPE/WPB), and the industry classification approach on average produce similar accuracy of valuation estimates. Consequently, the findings suggest that the five-factor regression model is the most efficient when comparable companies are selected from the industry where the target firm operates. All in all, while the five-factor regression model provides valuable insights into identifying truly comparable companies for the target firm, it is more effective as a complementary model, yielding highly accurate stock valuation estimates when used in combination with industry classification.

The remainder of the paper is structured as follows. Chapter 2 discusses relevant prior literature on relative valuation. Chapter 3 in detail describes the data collected for the analysis. Chapter 4 describes the methodology used in the analysis. Chapter 5 introduces the results of the study, tests the proposed hypothesis, and compares the results to the findings of prior literature. Chapter 6 summarizes the research and discusses the limitations of the paper. Finally, Appendix A provides additional supportive materials in the form of tables.

CHAPTER 2 Theoretical Framework

Prior studies by Demirakos, Strong, and Walker (2004), as well as Asquith Mikhail, and Au (2005), have pinpointed that in practice relative valuation techniques, particularly the multiples approach for valuing stocks are widely applied. Multiples-based valuation serves not only as an additional valuation method but also as a feasible substitute for more complicated valuation methods. The multiples approach assumes that the basic economic concept that perfect substitutes should sell for the same price holds (Knudsen et al., 2017). Therefore, papers in the field of multiples-based valuation have developed various models for identifying the most comparable companies, or perfect substitutes. To test whether the chosen firms are really comparable, the majority of papers have compared the price obtained by the use of the identified peers to the observed price for the target company. The model that yields the lowest error, or absolute difference between the predicted and observed price scaled by the observed price is considered to be the model that identifies the closest substitutes for the target company.

Alford (1992) examines the P/E multiple for the years 1978, 1982, and 1986 for companies that are listed on the NYSE, ASE, and NASDAQ. Alford (1992) chooses comparable firms on the basis of industry, defined by the first three digits of SIC, risk, proxied by total assets, and earnings growth both in pairs and individually. The paper finds that selecting peers using industry classification and the combination of risk and earnings growth are the two methods that yield the most accurate valuation estimates. The results suggest that much of the cross-sectional variation in P/E that is explained by risk and earnings growth is also explained by industry (Alford, 1992). Therefore, the assumption that companies operating in the same industry are likely to share fundamental value drivers holds in this particular paper.

Baker and Ruback (1999) study S&P500 firms in 1995. Unlike the majority of papers, Baker and Ruback (1999) do not develop a novel approach for identifying peers. Instead, they use industry classification and test which of the four approaches for estimating the industry multiple performs the best. The four approaches are the arithmetic mean, the harmonic mean, the value-weighted mean, and the median. Secondly, the authors test which multiples perform best for 22 industries. The study finds that the harmonic mean should be used for estimating the industry multiple. Hence, a key insight from this paper is the adoption of the harmonic mean for determining the average of the multiples. Furthermore, Baker and Ruback (1999) find that EBITDA is a better basis for substitutability compared to EBIT and revenue. In other words, when deriving the value from multiples EBITDA multiples should be used for 10 of the 22 industries. For 9 of 22 industries, EBIT is the measure that minimizes the spread across multiples within a specific industry. Revenue is the worst-performing measure.

Cheng and McNamara (2000) study the performance of P/E and P/B multiples where comparable companies are identified based on industry classification, size, profitability, proxied by return on equity, and a combination of these factors. The paper includes all firm observations available on the Industrial Compustat database from 1973 to 1992. Similar to Alford (1992), Cheng and McNamara (2000) find that industry classification alone yields the second most accurate valuation estimates. Nevertheless, in contrast to Alford (1992), they find that industry classification in combination with return on equity performs significantly better than the industry classification approach alone for P/E and P/B multiples. Furthermore, they find that valuation accuracy increases with the number of firms in the target firm's industry and with firm size.

Eberhart (2001) studies how the amount of information available for the comparable companies of the target firm affects the accuracy of the target firm's stock value estimates. The sample consists of the firms in the intersection of Compustat files and the Center for Research in Security Prices (CRSP) files from 1986 to 1995. The final sample includes 2,890 observations. The paper defines comparable companies as those in the same four-digit SIC code as the target firm. The number of comparable firms is used as a simple proxy for the information availability assuming that the higher the number of comparable companies the more information is available. Eberhart (2001) finds that when more comparable firms are identified, the precision of the valuation estimates for the target firm increases. This finding is similar to the finding of Cheng and McNamara (2000) who show that the valuation accuracy increases with the size of the target firm's industry.

Liu, Nissim, and Thomas (2002) study a list of value drivers and conclude which of the studied value drivers yields the most accurate valuation estimates. The paper uses 11% to 18% of companies listed on the NYSE, AMAX, and NASDAQ including 26,613 company-year observations from 1982 to 1999. Lie, Nissim, and Thomas (2002) find that multiples derived from forward, or forecasted earnings per share (EPS) perform the best in explaining stock prices. Furthermore, regarding relative performance, historical earnings are the second best-performing value driver, cash flow and book value measures rank third, and sales perform the worst. This ranking appears robust across different statistical methods used, sample years, and industries. The finding that forecasted EPS outperforms historical, or reported EPS is in line with the findings of Kim and Ritter (1999), Lie and Lie (2002), and Schreiner and Spremann (2007). Therefore, a key insight from this paper is the adoption of the forward earnings per share instead of historical earnings as one of the fundamental value drivers in my five-factor regression model.

Lie and Lie (2002) study various multiples used for valuing a company. The sample includes 8,621 active companies available in the Compustat database. The financial data is collected from the fiscal year 1998 and the earnings forecasts are for the fiscal year 1999. Firstly, the findings suggest that the asset multiple, which is calculated by dividing the market value by the book value of assets, typically yields more accurate valuation estimates relative to the sales and earnings multiples. Secondly, the incorporation of forecasted earnings instead of historical earnings does improve the estimates of the P/E multiple. Moreover, in line with Baker and Ruback (1999), Lie and Lie (2002) find that the EBITDA multiple overall produces more precise valuation estimates than the EBIT multiple. Finally, they show that the precision of the valuation estimates differs across company size, profitability, and the extent of intangible value showing that the valuation estimates are more precise for large firms. The finding that larger companies have more accurate valuation estimates can be explained by the idea that large firms tend to have a higher P/E meaning that the error calculated by taking the absolute value of the predicted price minus the actual price scaled by the actual price is smaller for large firms compared to small firms (Cheng and McNamara, 2000).

Bhojraj and Lee (2002), which is my inspirational article, present an approach for selecting comparable companies. The authors develop a multivariable regression model which they argue are the determinants of the multiples and explain the variation in the multiples. Using the regression model the multiple of each company is predicted which they call “warranted multiples.” The companies with the closest predicted, or warranted multiples to the target firm’s warranted multiple are chosen as peers for the target firm. The approach of choosing the four companies with the closest warranted multiples to the target firm’s warranted multiple is implemented both across and within industries. The performance of this approach is tested by how well the selected peers predict current-to-three-year-ahead multiples. The study incorporates companies that are present in both the Compustat industrial and research files as well as the I/B/E/S database of earnings forecasts. The analysis is conducted as of June 30th from 1982 to 1998. The final analysis shows that peers chosen based on warranted multiples provide more accurate predictions of valuation estimates compared to other methods for selecting peers such as industry and size matches (Bhojraj and Lee, 2002). More importantly, the paper demonstrates that warranted multiple approach produces significantly better results when applied within industries compared to when it is applied across different industries. The authors show that the warranted multiples within industries which they call “ICOMP” is the best-performing method for choosing peers. The method that yields the second most accurate valuation estimates is simply using the warranted multiple of the target firm. The warranted multiple is obtained by taking the previous year’s estimated coefficients and using them in conjunction with the current year's information for the target firm. They call this method WPE and WEVS for price over earnings per share and enterprise value divided by sales multiples respectively. In conclusion, since the WPE/WEVS and ICOMP are the two methods that yield the most accurate results, they are used for my novel five-factor regression model.

Dittmann and Weiner (2005) conduct a study to determine which method for selecting peers yields the most accurate results by using enterprise value to EBIT multiple. The paper uses European and United States companies. The data is collected from Worldscope and Datastream resulting in 67,433 firm-year observations between 1993 and 2002. Similar to Alford (1992) companies at the intersection of the 14% most similar firms in terms of return on assets (ROA), and at the intersection of the 14% most similar firms in terms of total assets (TA) are selected as peers. The authors find that peers selected based only on ROA, and ROA and TA are the two methods that yield the most accurate valuation estimates. In addition, selecting comparable companies from the same country yields the most accurate results for the United States, United Kingdom, Greece, and Denmark. For the remaining European countries, valuation estimates are more accurate when peers are selected from the 15 European Union member states.

Nel et al. (2014) examine which valuation fundamentals or combination of those fundamentals for peer group selection provides the highest valuation accuracy. The study uses around 71% of the South

African companies listed on the Johannesburg Stock Exchange (JSE) from 2001 to 2010. Nel et al. (2014) use the intersection of three sets of companies selecting peers as the companies whose fundamentals deviate within +/- 30% from the target company's fundamentals. The paper finds that the peers selected based on a combination of profitability, proxied by return on equity, and revenue growth yield the most accurate valuation estimates. Nel et al. (2014) notice that one limitation of the results in the paper is that it focuses solely on fundamentals as a selection method and overlooks industry analysis. However, the incorporation of industry analysis or a third fundamental variable is not possible due to the limited depth of the South African market.

Lee, Ma, and Wang (2015) develop a unique and novel school of thought suggesting that companies searched consecutively, or co-searched by the same user exhibit fundamental similarities across various dimensions. They call the method for selecting peers "Search-Based Peers" (SBPs). The study limits searches to firms listed on S&P1500 from January 2008 to December 2011. Lee, Ma, and Wang (2015) find that companies that are often co-searched by multiple users are economically similar. The paper states that search-based peers explain significantly more of the cross-sectional variations in target firms' valuation multiples, growth rates, R&D expenditures, leverage, and profitability compared to the industry peers identified by GICS6. Even though the method of peer identification used in this paper yields promising results, there can potentially be limited data availability in practice (Lee, Ma, Wang, 2015).

Young and Zeng (2015) use warranted multiples suggested by Bhojraj and Lee (2002) for identifying peer companies. Moreover, the paper not only studies economic comparability for identifying peers but also financial reporting decisions, or accounting methods used by companies. The study consists of 21,205 European firm-year observations retrieved from Extel and Reuters Knowledge Sample from the 1st of January 1997 to the 31st of December 2011. Young and Zeng (2015) find that pricing accuracy increases when the four companies with the closest warranted multiple to the target firm's warranted multiple are selected as peers. Additionally, the paper shows that similarities in accounting practices among cross-border European firms increase valuation accuracy by improving the comparability of value determinants.

Knudsen et al. (2017) develop a novel approach for identifying peers based on fundamental value drivers. The authors use five variables as proxies for fundamental value drivers and develop the sum of absolute rank differences (SARD). The companies are ranked by those five variables and the firms that have the lowest SARD relative to the target firm are selected as peers. The idea of the paper is that firms with the closest fundamental value drivers to the target firms should have a similar value in the market. The sample consists of the S&P Composite 1500 analysis conducted as of the 31st of March from 1995 to 2014 resulting in 12,350 firm-year observations. Knudsen et al. (2017) find that the SARD approach outperforms the industry classification approach. Furthermore, they find that the SARD approach when peers are chosen from the same industry in which the target company operates, yields the most accurate valuation estimates, outperforming SARD across industries, the industry classification approach, and the regression approach suggested by Bhojraj and Lee (2002). The latter finding is robust across the majority of industries, time, number of peers, and company size.

Overall the first school of thought insists on the efficiency and simplicity of the industry classification approach for choosing peers (Alford, 1992). Another school of thought studies the fundamental value drivers and some of the papers show that the selection of peers based on fundamental value drivers alone leads to highly accurate equity valuation estimates (Dittmann and Weiner, 2005; Nel et al., 2014; Knudsen et al. (2017). Finally, a novel and very unique approach for identifying comparable companies was suggested by Lee, Ma, and Wang (2015) However, some of the papers also use the industry classification approach as a complementary model and incorporate it with the fundamental value drivers. For instance, Bhojraj and Lee (2002) run a series of regressions using value drivers as explanatory variables and the various multiples as their dependent variables. One technique they use to select peers is constraining the selection of peers to the four firms with the closest warranted multiples to the predicted multiple of the target firm and that operate within the same industry as the target firm. They find that this technique along with the technique of choosing peers merely based on the warranted multiples are the two best performing techniques.

All in all, the incorporation of the two methods, namely the industry classification and value drivers, on average leads to more accurate valuation predictions which is intuitively logical because this approach makes the model more complete and takes more factors into consideration. At the same time, it makes the model cumbersome and not practical but multiples-based valuation is mainly used because of not only its efficiency in valuation estimates or as a useful comparison to the DCF model but also for its simplicity (Keun Yoo, 2006). That is why the approach suggested by Bhojraj and Lee (2002) is not widely used by practitioners even though it yields highly accurate valuation estimates (Knudsen et al., 2017). My paper uses the same approach but develops a simplified model that eases the data retrievability and data analysis process and at the same time may yield reasonably precise equity valuation estimates. This research idea is potentially interesting because the paper uses simple fundamental value drivers that were shown by Knudsen et al. (2017) to perform very well in evaluating the stock's value. Moreover, Knudsen et al. (2017) assign equal weights to each of the five value drivers meaning that each variable is equally important for every multiple, but my study with the regression approach gives each value driver a different weight for various multiples which is much more realistic. Therefore, I formulate the following hypothesis:

H1: *The five-factor regression method developed in this paper for selecting comparable companies positively affects the accuracy of stock valuation estimates both in absolute terms and relative to other methods.*

CHAPTER 3 Data

3.1 Sample description

I collected panel data on firms listed on the New York Stock Exchange (NYSE) from 2014 to 2019. The reason the data starts from 2014 is that the stock market bounced back from the 2008 global financial crisis at the beginning of 2013. For instance, by March 2013 the S&P500 achieved its first new all-time high since 2007 (Duggan, 2023). Therefore, to make sure that the stock market was fully recovered I collect data starting from year 2014. The data is collected as of the 31st of December 2014 to the 31st of December 2019 using the last twelve months (LTM) timeframe. Similarly, the data after 2019 is not used because during shocks to an industry or the broader market, the application of multiples is inapt and COVID-19 can be considered a type of shock (Rehm & Karnani, 2023). The accounting data is collected from Comustat, market data from CRSP, and analyst earnings forecasts from I/B/E/S. In line with Bhojraj & Lee (2002), companies with a stock price lower than $3 per share and sales below $100 million are excluded from the sample. While the paper acknowledges that NAICS and SIC are the most widely used industry codes in the United States, there are many firms in the paper with missing SIC codes (Columbia College, 2024). Therefore, an alternative industry classification method is used. Firms are categorized into their respective industries based on the first two digits of a four-digit 'Industry Group' code. This variable is obtained from Compustat and the first two digits of it indicate the major industry group in which the company operates. Industries with less than five firms are dropped from the sample. The finalized sample consists of 6,656 company-year observations and 23 industry groups.

3.2 Variables

My regression model has two dependent variables that are equity value multiples namely, *price divided by earnings per share (P/E)* and *price divided by book value of equity per share (P/B).* These multiples have advantages and disadvantages. The advantage is that equity value multiples allow for a direct valuation of equity. One of the disadvantages is that the comparison between firms is made difficult because of leverage. In other words, differences in the use of debt financing, or leverage complicate the direct comparison of financial ratios between firms (Bregonje, 2024, slide 33). Another disadvantage is that P/E and P/B multiples are sensitive to outliers due to one-off income or costs. Finally, there can be potential discrepancies and inconsistencies in the accounting methods and practices used by different companies to record their assets and liabilities leading to incomparability of book values of equity between firms (Bregonje, 2024, slide 33). The latter is one of the reasons for the poor performance of the P/B ratio in relative valuation (Liu, Nissim, and Thomas, 1999). On the other hand, the P/E ratio is a relatively good performing multiple and is the most commonly used multiple (CFI Team, 2023). Therefore, both multiples are used in the analysis to see how my novel approach for choosing peers performs for these two multiples that are distinct with regard to their accuracy in relative valuation.

The regression model includes five explanatory, or independent variables that prior literature suggests are good proxies for future growth, risk, and profitability. The explanatory variables are the following:

Market capitalization is defined as the current price multiplied by the total number of shares outstanding. In my regression model, it is a continuous variable measured in billion dollars. The data for market capitalization is collected from Compustat and it is taken as a proxy for size (Alford, 1992). The size of companies is very different in the sample, with market capitalization ranging from $0.101 billion to $569 billion. For market capitalization and all other variables, there are 6,656 observations. Size to some extent also serves as a proxy for risk because smaller firms are overall less liquid compared to larger firms and are therefore traded at a lower multiple (Knudsen et al., 2017).

Return on equity (roe) is calculated by dividing the net income by shareholders' equity. The data for return on equity is obtained from Compustat. Alford (1992) and Nel et al. (2014) show that return on equity is a well-performing variable for choosing comparable companies. Furthermore, return on equity is a good proxy for profitability and an important determinant of P/E and P/B ratios as shown by Knudsen et al. (2017). The variable in the regression model is a ratio expressed in proportion rather than a percentage and is not subject to any adjustments. Finally, firms with negative returns on equity are not excluded from the sample, meaning that loss firms are also studied.

Net debt/EBIT is simply defined as net debt divided by Earnings Before Interest and Taxes and similar to return on equity, is a ratio expressed as a proportion. Data for both net debt and EBIT are retrieved from Compustat. There are multiple definitions for net debt, but the paper uses the most commonly used definition which is subtracting cash & cash equivalents from total debt (CFI Team, 2024). This variable is a crucial element in credit analysis (Palepu et al., 2013). In the model, net debt/EBIT is used as a proxy for risk due to its effectiveness in evaluating a company's ability to repay debt.

EBIT margin is determined by dividing EBIT by sales, or revenue. Data for sales is obtained from Compustat. EBIT margin is proxied by Even though Knudsen et al. (2017) show that EBIT margin is particularly useful and is a major determinant of enterprise value multiples, it is still used for equity valuation multiples in this paper. Moreover, the results show that there is a significant correlation between the equity valuation multiples and the EBIT margin for some years enhancing its usefulness by providing insights into the cross-sectional variation in P/E and P/B ratios.

Future growth is defined as the growth rate in earnings per share. However, prior literature such as Lie, Nissim, and Thomas (2002) show that multiples derived from forward, or forecasted earnings per share (EPS) perform better relative to historical earnings per share in predicting stock prices. Therefore, in the paper analysts' 12-month forward forecasts of earnings per share serve as a proxy for a company's future growth rate which is obtained from I/B/E/S. Similar to Knudsen et al. (2017), the growth rate is calculated with the following formula:

$$Growth\ rate_{i,t} = \frac{Forecasted\ EPS_{i,t+1} - Forecasted\ EPS_{i,t}}{Forecasted\ EPS_{i,t}}$$

where '$Forecasted\ EPS_{i,t+1}$' is the analysts' forecasts of company i's earnings per share for period t+2, because the prediction is 12-month forward and '$\text{Forecasted EPS}_{i,t}$' is the analysts' forecasts of company i's earnings per share for period 't+1'. Hence, for instance, future, or implied growth in 2015 is the forecasted earnings per share in 2017 minus forecasted earnings per share in 2016 divided by the forecasted earnings per share in 2016.

3.3 Descriptive statistics

By definition, a firm is considered a large-cap firm if its market capitalization exceeds $10 billion (Ross, 2024). The average market capitalization is $18.3 billion meaning that on average the paper studies and focuses on large-cap companies (see Table 1). Furthermore, it is important to notice that the minimum values of multiples are negative, large numbers meaning that loss firms are not excluded from the sample.

Table 1. *Descriptive statistics for dependent and explanatory variables*

Variable	Obs.	Mean	Std. dev.	Min	Max
PE_close	6,656	16.056	146.734	-5899.700	2033.730
PB_close	6,656	3.844	25.308	-956.580	830.500
marketcap_	6,656	18.300	40.600	0.101	569.000
roe_	6,656	0.202	4.064	-70.678	315.600
netdebt_EBIT	6,656	4.595	50.957	-1577.667	2226.291
EBIT_margin	6,656	0.156	0.272	-5.523	8.045
Future_growth_rate	6,656	0.110	2.471	-55.333	107.000

Note: Market capitalization is given in billion dollars and all the other variables are given in proportions rather than percentages

CHAPTER 4 Methodology

Even though panel data is collected, the data is treated in a cross-sectional manner. To analyze the collected data classic linear regression model (CLRM) is applied for each year separately from 2014 to 2017. In the first part of the analysis, the CLRM assumptions that are in practice possible to test are tested. To test the second assumption of CLRM, which is that the data is homoscedastic, a white test is used. The white test is implemented twice once with P/E being the dependent variable and once when the P/B is the dependent variable. The white test for P/E yields a P-value of 0.9999. Therefore, the null hypothesis of homoscedasticity cannot be rejected and it is assumed that the second assumption of CLRM holds. However, for P/B the white test yields a P-value approximating zero. Therefore for all significance levels, the null hypothesis is rejected and the data is heteroskedastic. Furthermore, the third assumption of CLRM imposes that the error terms are not correlated with each other. Nevertheless, after analyzing the data the results suggest that the error terms are likely to be correlated. Because for P/B the second assumption is violated and the third assumption is violated for both multiples, structure needs to be imposed. Hence, clustered standard errors are used which is a common solution for the violation of the second and third assumptions.

In the later stage of the analysis, CLRM is run for every year separately from 2014 to 2017 and the obtained coefficients are used with the current year information to generate a predicted multiple for the current year for company i. The significance of coefficients is tested by a simple two-tailed t-test. The following regression model is used:

$$P/E_{i,t} = \beta_{0,t-1} + \beta_{1,t-1}\, market\ capitalization_{i,t} + \beta_{2,t-1}\, return\ on\ equity_{i,t} + \beta_{3,t-1}\, netdebt\_EBIT_{i,t} + \beta_{4,t-1}\, EBIT\ margin_{i,t} + \beta_{5,t-1}\, future\ growth\ rate_{i,t} + \epsilon_{i,t}$$

where market capitalization is given in dollars and all the other variables are expressed as ratios. All the coefficients are obtained from the previous year and used in conjunction with the current year's explanatory variables. Four different techniques are used to obtain a predicted multiple for the target firm and those techniques are the following:

COMP takes the four closest firms with their predicted, or warranted multiples to the target firm as comparable firms. Next, the harmonic mean of the actual multiples for these four firms is calculated.

ICOMP is the same as COMP, but with the added requirement that the firms must belong to the same industry.

WPE/WPB is the warranted multiple for the target firm. It is calculated by simply taking the previous year’s estimated coefficients and using them in conjunction with the current year's information for the target firm to generate the multiple.

Industry Classification calculates the harmonic mean and of the P/E or P/B ratios of all firms operating within the same industry group as the target firm. How the industry groups are defined in this paper is explained and discussed in the data section.

ICOMP and WPE/WPB are the two best-performing techniques in Bhojraj and Lee (2002). Therefore, these two methods are used for my novel approach and COMP and industry classification are mainly used for comparison to ICMOP and WPE/WPB and for gaining insights into the differences between the results obtained by the four different methods. Baker and Ruback (1999) show that the harmonic mean is the best approach for estimating the multiple and that is the reason the harmonic mean is used rather than any other type of a mean for estimating the multiple. Nevertheless, the sample includes firms with negative P/E or P/B ratios, making the use of harmonic and geometric means not always possible. Therefore, if one or more of the identified peers have negative P/E or P/B, the arithmetic mean is used as an alternative. The obtained multiple is then multiplied by the target company's earnings per share for the P/E multiple and by the book value of equity for the P/B ratio. A price for the target firm is obtained which we call the predicted price. Finally, in line with Knudsen et al. (2017) the error term is calculated by taking the absolute difference between the predicted price and the actual price, scaled by the actual price.

CHAPTER 5 Results & Discussion

As already mentioned the classic linear regression model (CLRM) is used for analysis. Return on equity, net debt divided by earnings before interest and taxes (EBIT), EBIT margin, and future growth rate are all in ratios therefore, the interpretation is the following: when one of these variables increases by 1 unit, the P/E or P/B ratio changes by the coefficient of the corresponding explanatory variable. The interpretation is slightly different for market capitalization. When market capitalization increases by $1, the multiple changes by the coefficient for market capitalization (see Table 2).

Table 2. Regression results for the years from 2014 to 2018, P/E multiple being the dependent variable

	P/E				
	2014	2015	2016	2017	2018
Market capitalization ($)	1.77×10^{-11} (0.000)	3.36×10^{-11} (0.000)	1.94×10^{-11} (0.000)	2.82×10^{-11} (0.000)	-1.35×10^{-11} (0.000)
Return on equity	0.668*** (0.248)	3.393*** (1.272)	3.702 (2.761)	1.608* (0.623)	2.280** (1.121)
Net debt/EBIT	-0.121 (0.094)	-0.048 (0.055)	-0.033 (0.040)	-0.117 (0.174)	-0.039 (0.031)
EBIT margin	5.152 (7.008)	13.344*** (4.236)	36.416*** (9.940)	16.622** (8.548)	44.652* (23.249)
Future growth rate	1.824 (3.830)	-0.406 (0.681)	1.943 (2.661)	-0.166 (0.638)	-0.842 (0.760)
Constant	18.624 (4.628)	12.411 (3.986)	12.628 (4.807)	18.154 (4.705)	-1.115 (8.660)
Observations	1,038	1,038	1,109	1,154	1,155
R^2	0.017	0.036	0.057	0.028	0.025
Adjusted R^2	0.031	0.031	0.053	0.024	0.021

N*ote:* Standard errors are given in parentheses. *p < 0.1; **p < 0.05 ***p < 0.01

I only discuss the year 2015 from Table 2 because the results are similar across all regression years. Furthermore, the table with regression results for the P/B ratio is in Appendix A. The R-square for P/E multiple in 2015 is 0.036, or 3.6% which is quite low (see Table 2). It means that only 3.6% of the variation in the P/E ratio is explained by the five variables included in the regression model. However, for P/B ratio the R-squares are higher relative to the R-squares observed for P/E with the highest value of 23.6% for 2017 (see Appendix A). Therefore, the five-factor model explains the variation in P/B ratio more, hence, is implied to be a more suitable model for P/B multiple.

In the regression model, the general-to-specific method is used. In other words, first, the five-factor regression model is run and then one variable at a time is removed until all variables are significant at a 5% significance level. The same process is implemented for all years. The finalized model for 2015 only includes two explanatory variables and those are the return on equity and EBIT margin. The other three are removed because they are insignificant at a 5% significance level. Furthermore, the finalized regression model for 2015 with the two abovementioned variables has an R-squared of 0.03 or 3%. This means that the explanatory power of market capitalization, Net debt/EBIT, and future growth rate for the variation in P/E ratio is very small. This holds for all the other years as well. Furthermore, similar to 2015 for all the other years either return on equity or EBIT margin, or both of them have a significant correlation with P/E ratio. The return on equity coefficient of 3.393 for 2015 means that a 1 unit increase in return on equity on average increases the P/E ratio by 3.393 units. The coefficient of EBIT margin for 2015 is 13.344. One unit increase in EBIT margin on average leads to an increase of 13.344 units in P/E ratio (see Table 2). It is really hard to assess whether the magnitude by which return on equity and EBIT margin correlate with the P/E ratio is the expected magnitude, but these variables have the expected sign. The fact that a positive correlation between the multiple and return on equity is observed is in line with Bhojraj and Lee (2002). Furthermore, the EBIT margin shows how efficient a firm is in converting its revenues into operating profit. Hence, I expected a positive correlation between the EBIT margin and the P/E multiple. Therefore, even though it is really hard to assess whether the magnitude by which return on equity and EBIT margin correlate with the P/E ratio is the expected magnitude, these variables have the expected sign.

Table 3 shows the distribution of absolute valuation errors obtained by four different techniques that are explained in the methodology section. While the regression is run using the complete data including 23 industry groups, the absolute errors are obtained for two major industry groups namely, Aerospace and Healthcare. Furthermore, the actual price is used as a benchmark for the true intrinsic value as the errors are calculated by taking the absolute difference between the predicted price per share and the actual price per share, scaled by the actual price per share. Similar to Bhojraj and Lee (2002) while the paper recognizes that the market is not necessarily efficient, and may not reflect all the publicly available information, the paper still believes that the market price contains information useful for valuation purposes.

It is important to note that P/E ratios systematically yield significantly lower average and median absolute errors compared to P/B ratios supporting the idea that the P/B ratio is a poor-performing multiple in relative valuation analysis (Liu, Nissim, and Thomas, 2002). Furthermore, the three-year ahead predictions are calculated only for the years 2015 and 2016. The reason for that is that if a three-year-ahead absolute error is calculated for 2017, 2018, and 2019, prices from 2020 or later should be used. However, it was already discussed that the paper does not use any information from 2020 onwards because during that period a shock in the form of a COVID-19 pandemic followed by a lockdown took place. Additionally, all the estimated coefficients for the P/B multiple for 2016 are insignificant (see Appendix A). Therefore, the warranted multiple cannot be obtained and ICOMP, COMP, and WPB cannot be applied. For this reason, for the three-year-ahead prediction of the P/B ratio only 2015 could have been used in the data. Nevertheless, I decided not to use only one year for the prediction of a three-year-ahead ratio as it may yield biased results and not be representative of the average error that would have been obtained if more years were used.

To understand what is defined to be a low absolute error, the results of other papers are studied. For instance, Knudsen et al. (2017) report a mean absolute error of 0.306, or 30.6%, for the P/E ratio, which they consider to be low, indicating that their model performs well. Moreover, Bhojraj and Lee (2002) obtain errors for the P/B ratio using ICOMP that fluctuates around 0.55 or 55% and they argue that ICOMP is a well-performing method. Table 3 shows that the absolute error for the P/E ratio is 0.291, or 29.1% and for the P/B ratio, an average of 53.5% and 62.8% errors are obtained using ICOMP. Taking into consideration what is considered to be a low absolute error by the prior literature, I state that the five-factor regression method developed in this paper for selecting peers yields accurate valuation estimates in absolute terms.

To get insights into how the model performs relative to other approaches Table 3 is closely analyzed. Table 3 shows that for both multiples the ICOMP always yields a smaller mean absolute error compared to COMP. The only difference between ICOMP and COMP is that in ICOMP comparable firms are selected from the same industry and for COMP there is no such restriction. This finding suggests that there are similarities between firms operating within the same industry that are not captured in my five-factor regression model. Furthermore, on average ICOMP outperforms industry classification. Nevertheless, the WPE/WPB and industry classification on average yield similar absolute errors. Therefore, since ICOMP and WPE/WPB are both part of the model, there is only partial support for the hypothesis that the five-factor regression method developed in this paper for relative valuation yields more accurate valuation estimates relative to other approaches, namely the most popular industry classification approach. Among these four approaches, ICOMP is the only one that incorporates industry classification with the fundamental value drivers and it produces the most accurate valuation estimates. This finding suggests that industry classification and the methods that only analyze fundamental value drivers, such as the WPE/WPB are both complementary approaches rather than substitutes for each other. Therefore, the most accurate valuation estimates are obtained when the two are incorporated into a model.

In conclusion, based on the results, there is only partial support for the hypothesis which states that the five-factor regression method developed in this paper for obtaining a multiple for the target firm positively affects the accuracy of valuation estimates both in absolute terms and relative to other methods. While the model performs well in absolute terms and the ICOMP yields more accurate valuation estimates relative to industry classification, the WPE/WPB does not outperform the industry classification approach.

Table 3. Distribution of absolute valuation errors for various measurement techniques, the absolute error being the absolute difference between the predicted price and the actual price, scaled by the actual price

		Absolute Error		
		Mean	Median	Interquartile Range
Panel A: Price-to-earnings-per-share				
PE0	*ICOMP*	0.295	0.238	0.233
	WPE	0.336	0.382	0.268
	COMP	0.334	0.354	0.184
	Ind. Class.	0.350	0.356	0.231
PE1	*ICOMP*	0.262	0.318	0.334
	WPE	0.364	0.463	0.356
	COMP	0.464	0.491	0.222
	Ind. Class.	0.265	0.279	0.091
PE3	*ICOMP*	0.317	0.289	0.410
	WPE	0.363	0.335	0.530
	COMP	0.579	0.553	0.272
	Ind. Class.	0.326	0.356	0.156
Panel B: Price-to-book-value-per-share				
PB0	*ICOMP*	0.503	0.609	0.377
	WPB	0.536	0.598	0.396
	COMP	0.665	0.59	0.322
	Ind. Class.	0.521	0.505	0.310
PB1	*ICOMP*	0.609	0.713	0.246
	WPB	0.602	0.705	0.258
	COMP	0.644	0.635	0.283
	Ind. Class.	0.678	0.682	0.204

Note: All values are given as a proportion of the actual price per share rather than a percentage. The PE0, PE1, and PE3 are for current, one-year-ahead and three-year-ahead predictions respectively. The absolute errors are based on two industry groups, namely Aerospace and Healthcare

Finally, the robustness of the results across time is tested (see Appendix A). The same model is estimated from 2015 to 2018 separately for each year. The results from Tables 5 and 6 in Appendix A on average align well with those reported in Table 3, suggesting that my findings are robust across time.

The obtained results are really similar to the results obtained by Knudsen et al. (2017), where they also find that their sum of absolute rank differences (SARD) approach yields the lowest absolute errors outperforming the industry classification and SARD approach across industries when the fundamental value drivers are analyzed within industries.

Furthermore, Bhojraj and Lee (2002) find that ICOMP and WPE/WPB are the two best-performing techniques for predicting the value of the target company yielding the lowest absolute errors. While my model finds support for the superior performance of ICOMP, WPE/WPB does not perform very well and also does not outperform the industry classification approach. Therefore, my findings are only partially similar to the findings of Bhojraj and Lee (2002). Nevertheless, the explanation for the divergence in WPE/WPB between this paper and the one by Bhojraj and Lee (2002) can be explained by the difference in the regression models. Bhojraj and Lee (2002) developed a regression model with eight explanatory variables with an average R-square of approximately 51% which is significantly high compared to the R-square of my regression model. WPE/WPB is determined by only using the regression model. Therefore, it is potentially the case that because more of the variation in the ratios is explained by their model, their model yields more precise valuation estimates when applying WPE/WPB than my model does. From a practical point of view, the ICOMP used based on the five-factor regression model is more easily applicable than the ICOMP based on the eight-factor model suggested by Bhojraj and Lee (2002) because in my model simpler variables are used easing the data retrievability and data analysis process. More importantly, the simplification of the model did not decrease the accuracy of valuation estimates and the model still produces reasonably low absolute errors (see Table 3). In conclusion, not only does the model aim to achieve efficiency in valuation estimates but also it aims for simplicity because one of the reasons for the wide application of multiples-based valuation is its simplicity (Keun Yoo, 2006).

The findings are also in line with Cheng and McNamara (2000) who suggest that the industry classification approach in combination with return on equity (which is one of the fundamental value drivers in my regression model) yields the most accurate valuation estimates supporting the concept that industry classification incorporated with fundamental value drivers is the best-performing approach in relative valuation. On the other hand, my findings contradict the findings of Liu, Nissim, and Thomas (2002) and Lie and Lie (2002) who show that the use of forward, or forecasted earnings per share (EPS) improves the valuation estimates. Nevertheless, in my regression model for both P/E and P/B multiples, the future growth rate, calculated based on future EPS, is found to be insignificant across all years. Therefore, future earnings per share do not improve the accuracy of my valuation estimates as suggested by the two abovementioned papers. In conclusion, despite the fact that the main findings are supported by prior literature, there are components in my study that contradict prior literature as discussed above.

CHAPTER 6 Conclusion

In this thesis, I look at how a newly developed five-factor regression model affects the accuracy of stock valuation estimates. Bhojraj and Lee (2002) had the same purpose for their study, but their approach was perceived as cumbersome and not applicable in practice (Knudsen et al., 2017). Overall, prior literature suggests inconclusive, or mixed results. Some papers argue that the industry classification alone is a simple and efficient approach for valuing a stock, while others insist on the use of fundamental value drivers or a combination of those drivers with industry classification for finding the most comparable firms and therefore obtaining more accurate valuation estimates. Therefore, the purpose of the study is to develop a systematic, consistent, and at the same time simple model for identifying the most comparable firms for the target company and increasing the accuracy of the target firm's stock valuation estimates. Thus, the paper aims to answer the following research question: *"How does the simplified regression approach developed in this paper for identifying peers influence the accuracy of stock valuation estimates of companies listed in NYSE in absolute terms and relative to other methods?"*

To answer this research question, data for my five fundamental value drivers and P/E and P/B multiples is collected for stocks that are listed on the New York Stock Exchange (NYSE). The data is collected from 2014 to 2019. The analysis of this data shows that there is only partial support for the hypothesis which states that the five-factor regression method developed in this paper positively affects the accuracy of valuation estimates both in absolute terms and relative to other methods.

This study concludes that although fundamental value drivers are an important factor for identifying comparable companies, industry classification still provides additional accuracy in valuation estimates when incorporated with the value drivers which is evidenced by the significant difference between ICOMP and COMP. This finding can be explained by the idea that firms operating within the same industry share similarities that are not captured in my five-factor regression model consisting of fundamental value drivers. Furthermore, the industry classification alone has a mediocre performance and its performance on average is similar to the one of the warranted, or predicted multiple (WPE/WPB). Therefore, using industry classification or WPE/WPB as a single factor for identifying peers will not yield as accurate valuation estimates as the model that selects comparable firms based on the combination of industry classification and fundamental value drivers. Therefore, the paper concludes that while the five-factor regression model offers some insights into identifying truly comparable companies for the target firm, it functions more effectively as a complementary model and should be used in combination with industry classification to yield highly accurate stock valuation estimates. Additionally, the finding that using the five-factor regression model to generate the predicted multiples and then identifying peers with the obtained predicted multiples within an industry is the model that yields the most accurate stock valuation estimates, can serve as a recommendation for investors to use this approach when valuing a stock using relative valuation analysis.

In the paper, the most common approach for calculating the absolute error is applied. It is defined as the absolute difference in predicted price per share and actual price per share, scaled by the actual price per share. However, when the absolute error is defined this way, large-cap companies, or companies with a high share price tend to have downward biased absolute errors (Cheng and McNamara, 2000). Furthermore, an average firm on my data is a large-cap firm. Therefore, it may be the case that the absolute errors obtained in the paper are downward biased. Future research is encouraged to replicate my model and try to develop a more objective definition of absolute error which may potentially lead to less biased absolute errors. Another suggestion for future research may be to use a sample that mainly consists of middle and small-cap firms which can help avoid the limitation of potentially biased errors driven by large-cap firms.

REFERENCES

Alford, A. W. (1992). The Effect of the Set of Comparable Firms on the Accuracy of the Price-Earnings Valuation Method. *Journal of Accounting Research*, *30*(1), 94. https://doi.org/10.2307/2491093

Asquith, P., Mikhail, M. B., & Au, A. S. (2005). Information content of equity analyst reports. *Journal of Financial Economics*, *75*(2), 245-282. https://doi.org/10.1016/j.jfineco.2004.01.002

Bhojraj, S., & Lee, C. M. C. (2002). Who Is My Peer? A Valuation-Based Approach to the Selection of Comparable Firms. *Journal of Accounting Research*, *40*(2), 407–439. https://doi.org/10.1111/1475-679X.00054

Bregonje, A. (2024). *Introduction to Valuations* [PowerPoint slides]. Retrieved from https://canvas.eur.nl/courses/44066/pages/session-a1-valuation-assignment-materials philips?module_item_id=1138938

Cheng, C. S. A., & McNamara, R. (2000). The Valuation Accuracy of the Price-Earnings and Price-Book Benchmark Valuation Methods. *Review of Quantitative Finance and Accounting*, *15*(4), 349–370. https://doi.org/10.1023/A:1012050524545

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

Dittmann, I., & Weiner, C. (2005). Selecting Comparables for the Valuation of European Firms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.644101

Duggan, W. (2023, June 21). *A short history of the Great Recession*. Forbes Advisor. Retrieved from https://www.forbes.com/advisor/investing/great-recession/

Eberhart, A. C. (2001). Comparable firms and the precision of equity valuations. *Journal of Banking & Finance*, *25*(7), 1367–1400. https://doi.org/10.1016/S0378-4266(01)00171-6

Karnani, P., & Rehm, W. (2023, May 4). *The times for multiples: Five situations when multiples need more than a second look*. McKinsey & Company. Retrieved from https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/the-times-for-multiples-five-situations-when-multiples-need-more-than-a-second-look#/

Keun Yoo, Y. (2006). The valuation accuracy of equity valuation using a combination of multiples. *Review of Accounting and Finance*, *5*(2), 108–123. https://doi.org/10.1108/14757700610668958

Koller, T. et al, (2017). The right role for multiples in valuation. *McKinsey on Finance*, (15), 7-11.

Kim, M., & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics*, *53*(3), 409–437. https://doi.org/10.1016/S0304-405X(99)00027-6

Knudsen, J. O., Kold, S., & Plenborg, T. (2017). Stick to the Fundamentals and Discover Your Peers. *Financial Analysts Journal*, *73*(3), 85–105. https://doi.org/10.2469/faj.v73.n3.5

Lee, C. M. C., Ma, P., & Wang, C. C. Y. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, *116*(2), 410–431. https://doi.org/10.1016/j.jfineco.2015.02.003

Lie, E., & Lie, H. J. (2002). Multiples Used to Estimate Corporate Value. *Financial Analysts Journal*, *58*(2), 44–54. https://doi.org/10.2469/faj.v58.n2.2522

Liu, J., Nissim, D., & Thomas, J. (2002). Equity Valuation Using Multiples. *Journal of Accounting Research*, *40*(1), 135–172. https://doi.org/10.1111/1475-679X.00042

Nel, S., Bruwer, W., & Le Roux, N. (2014). An emerging market perspective on peer group selection based on valuation fundamentals. *Applied Financial Economics*, *24*(9), 621–637. https://doi.org/10.1080/09603107.2014.894629

Palepu, K. G., Healy, P. M., & Peek, E. (2013). *Business analysis and valuation: Text & cases* (3. ed., IFRS ed). Cengage Learning.

Schreiner, A., & Spremann, K. (2007). Multiples and Their Valuation Accuracy in European Equity Markets. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.957352

Stafford Library (2024). *Company and Industry Information Resource Guide: NAICS & SIC Codes.* Retrieved from https://library.ccis.edu/company-industry-info/naics-sic

Team, C. (2023, December 11). *Types of valuation multiples*. Corporate Finance Institute. Retrieved from https://corporatefinanceinstitute.com/resources/valuation/types-of-valuation-multiples/

Team, C. (2024, February 20). *Net Debt.* Corporate Finance Institute. Retrieved from https://corporatefinanceinstitute.com/resources/valuation/net-debt/

Young, S., & Zeng, Y. (2015). Accounting Comparability and the Accuracy of Peer-Based Valuation Models. *The Accounting Review*, *90*(6), 2571–2601. https://doi.org/10.2308/accr-51053

APPENDIX A

Table 4. Regression results for the years from 2014 to 2018, P/B multiple being the dependent variable

	P/E				
	2014	2015	2016	2017	2018
Market capitalization	-3.45×10^{-13} (0.000)	3.99×10^{-12} (0.000)	1.60×10^{-11} (0.000)	1.07×10^{-11} (0.000)	5.87×10^{-11} (0.000)
Return on equity	1.747*** (0.453)	4.926 (6.392)	12.875*** (4.364)	7.774* (4.257)	3.967 (4.016)
Net debt/EBIT	-0.002 (0.005)	0.000 (0.002)	-0.001 (0.001)	-0.010 (0.007)	-0.001 (0.001)
EBIT margin	-4.364* (2.401)	-1.409 (3.198)	-6.078*** (2.383)	-3.237 (3.072)	-4.849 (3.792)
Future growth rate	0.190 (0.195)	0.010 (0.262)	0.116 (0.950)	0.031 (0.059)	0.008 (0.021)
Constant	5.247 (1.415)	2.017 (0.526)	2.744 (0.696)	2.462 (0.748)	2.583 0.552
Observations	997	1,038	1,109	1,154	1,155
R^2	0.017	0.031	0.236	0.064	0.076
Adjusted R^2	0.012	0.026	0.233	0.060	0.072

Note: Standard errors are given in parentheses. *p < 0.1; **p < 0.05 ***p < 0.01

Table 5. Robustness of the results across time for the P/E ratio

		Mean Absolute Error			
		2015	2016	2017	2018
PE0	*ICOMP*	0.108	0.384	0.370	0.318
	WPE	0.082	0.448	0.454	0.360
	COMP	0.311	0.287	0.371	0.408
	Ind. Classification	0.279	0.410	0.427	0.285
PE1	*ICOMP*	0.043	0.318	0.418	0.270
	WPE	0.229	0.363	0.415	0.449
	COMP	0.306	0.457	0.509	0.584
	Ind. Classification	0.109	0.280	0.406	0.266
PE3	*ICOMP*	0.289	0.345		
	WPE	0.335	0.392		
	COMP	0.553	0.605		
	Ind. Classification	0.285	0.366		

Note: All values are given as a proportion of the actual price per share rather than a percentage. The PE0, PE1, and PE3 are for current, one-year-ahead and three-year-ahead predictions respectively

Table 6. Robustness of the results across time for the P/B ratio

		Mean Absolute Error		
		2015	2017	2018
PB0	*ICOMP*	0.415	0.341	0.753
	WPB	0.358	0.514	0.735
	COMP	0.409	0.590	0.996
	Ind. Class.	0.623	0.505	0.435
PB1	*ICOMP*	0.494	0.510	0.822
	WPB	0.502	0.487	0.817
	COMP	0.487	0.780	0.666
	Ind. Class.	0.673	0.632	0.729

Note: All values are given as a proportion of the actual price per share rather than a percentage. The PB0, PB1 are for current and one-year-ahead predictions respectively