OPTIMISING PEER SELECTION FOR RELATIVE VALUATION

A COMPARATIVE STUDY OF INDUSTRY CLASSIFIERS AND AN IDIOSYNCRATIC PEER GROUP SIZES APPROACH

LUCAS NILSSON

MARCUS SALOMON-SÖRENSEN

Master Thesis
Stockholm School of Economics
2024

Optimising Systematic Peer Selection for Relative Valuation: A Comparative Study of Industry Classifiers and an Idiosyncratic Peer Group Size Approach

Abstract:

This study enhances the widely used relative valuation technique by conducting a comparative analysis of conventional and product-description-based classifiers and introducing a novel idiosyncratic peer group size approach. Focusing on U.S. companies from 2010 to 2021, we applied two non-linear peer ranking methods – one univariate and one multivariate – to identify peers used to estimate the forward P/E for the firm in focus. The findings revealed that the alternative product-description-based classifier TNIC achieved the highest valuation accuracy compared to the conventional classifiers SIC and GICS, as well as the alternative product-description-based D2V classifier. We attribute TNIC's superior performance to its dynamic structure, which aligns with a more flexible and adaptive approach to grouping companies, moving away from the rigid, predefined structures of conventional classifiers. Additionally, we introduced a new approach to optimise peer group sizes, by accounting for differences in industry classifier characteristics and industry size. We conclude that this idiosyncratic approach produced 20-30% higher valuation accuracy than the fixed peer group size approach, widely used in prior studies. This approach is entirely novel, and while we urge future research to test its robustness, our findings highlight the importance of accounting for industry size for systematic peer selection.

Keywords:
Relative valuation, systemic peer selection, industry classification, peer group size

Authors:
Lucas Nilsson (50743)
Marcus Salomon-Sörensen (42521)

Tutor:
Stina Skogsvik, Assistant Professor, Department of Accounting

Master Thesis
Master Program in Accounting, Valuation and Financial Management
Stockholm School of Economics

Table of contents

1. Introduction

> *"Finding similar and comparable firms is often a challenge and we have to often accept firms that are different from the firm being valued on one dimension or the other."* (Damodaran, 2002, p.11)

In private markets, where financial information and forecasts are often limited, reducing the applicability of intrinsic value models, relative valuation stands out as a widely used technique (Asquith et al., 2005; Berk & DeMarzo, 2017; Pinto, et al., 2018; Forte et al., 2020). Relative valuation relies on the use of multiples, which represent the ratio between the estimated value – such as equity or enterprise value (EV) – and a relevant value driver, like revenue or earnings. In practice, when conducting firm valuations for M&A transactions or IPOs you typically rely on two sources: comparable public companies and valuations from similar transactions, (Berk & DeMarzo, 2017). These sources allow the identification of valuation benchmarks, enabling practitioners (professionals actively engaged in valuations) to address the challenge of valuing private companies in the absence of readily available market data.

The core principle of relative valuation is that comparable assets should have similar valuation (Fama & French, 1992), making the selection of peers with similar characteristics a crucial step. However, finding "perfect" peers is difficult, as no two companies are identical (Skogsvik & Skogsvik, 2009). Additionally, banks may have incentives to choose peers that favourably influence the valuation to secure deals (Eaton et al., 2021), meaning valuation accuracy may not always be the primary objective. In response to the limited practical focus on valuation accuracy, academic research has explored more systematic approaches to peer selection. While much of the research on peer selection has concentrated on improving valuation accuracy (Knudsen et al., 2017), peer selection has also been applied to areas such as sales growth forecasting (Theising et al., 2023), credit risk assessment (Pandey & Bandhu, 2022), and organisational benchmarking (Kennedy et al., 2021).

Previous research on systematic peer selection models for relative valuation identifies peers using two key dimensions. (i) Financial profile, capturing elements such as profitability, growth, and risk (Bhojraj & Lee, 2002) through metrics like return on equity (ROE) and operating margin. (ii) Business profile, typically is based on industry classifications as firms within the same industry are assumed to have similar risk profiles and earnings growth dynamics (Alford, 1992). Commonly used industry classification systems, such as SIC and GICS, rely on static definitions that are rarely updated, with firm's categorisation remaining largely stable over time. But as Lee et al. (2015, p. 411) note, industry classifications are inherently imprecise: "*With no conceptual guidance on what constitutes an 'industry', the choice of industry benchmarks ultimately relies on subjective judgement*."

Advancements in statistical tools, such as AI-driven language analysis, have opened new possibilities for refining how companies can be grouped together. Eaton et al. (2021) demonstrated the effectiveness of a dynamic, product-description-based industry classifier developed by Hoberg & Phillips (2016) – known as the Text-Based Network Industry Classifier (TNIC). They show that TNIC more accurately reflected peer groups used for relative valuation in M&A transactions compared to traditional classifiers like SIC. Similarly, Adebäck & Haqués (2022), in a master's thesis at the Stockholm School of Economics, found that TNIC outperformed conventional classifiers in systematic peer selection. Building on this progress, Hoberg & Phillips (2023) introduced the D2V-TNIC-3 (D2V) industry classifier, representing a further refinement in product-description-based classification. While TNIC identifies peers directly based on the similarity of product descriptions, D2V creates 300 fixed industry definitions from product descriptions and assigns companies to these predefined industries. Although the study did not primarily focus on profitability accuracy (a critical driver for valuations) it found that D2V enhance profitability prediction accuracy by 26% compared to TNIC, which itself was 46% more accurate than the conventional Standard Industry Classifier (SIC).

Moreover, using data from actual M&A transactions, Eaton et al. (2021) observed that peer groups typically range from eight to ten firms in practice. However, the peer group size has received little attention in previous research which is acknowledged by Dittmann & Weiner (2005, p.5), stating, "*The choice of the number five [peers] is arbitrary, and we are not aware of any study that investigates which minimal number of firms is optimal*". Knudsen et al. (2017) examined peer group sizes and found that the most accurate results were achieved with 6–16 peers, depending on the multiple used. However, their analysis assumed a fixed peer group size for all 12,350 firms valued over their study period, which seems overly simplistic. Moreover, Cheng & McNamara (2000) found that larger industries tend to generate higher valuation accuracy, yet no prior research has aptly accounted for the idiosyncratic aspect of industry size in peer group construction methodologies.

By using data from U.S. companies spanning from 2010 to 2021, we investigate peer group accuracy for forward price-to-earnings (P/E) multiples using the non-linear peer ranking methods developed by Cheng & McNamara (2000) and Knudsen et al. (2017). Firstly, our analysis compares the performance of several industry classifiers: Market[1], GICS, SIC, D2V, and TNIC. Consistent with previous research, we found that all classifiers outperformed the Market, with TNIC generally achieving the highest accuracy and robustness over the entire study period. The performance of the other classifiers varied, with relative performance rankings shifting depending on the test. Second, we introduce a novel idiosyncratic peer group size approach that uses heuristic measures, i.e. rules of thumb, to estimate the optimal size for each peer group. Applying the optimal ratio – one of the heuristic measures which accounts for differences in classifier characteristics and

[1] From this point forward, "Market," spelled with a capital M, will refer to the classifier that treats the entire market as a single industry, i.e. no industry delimitation applied in the peer ranking.

industry size – valuation accuracy improved significantly across all classifiers, achieving a 20–30% increase compared to the traditional fixed peer group size approach commonly used in previous studies.

The remaining text is structured as follows: Section 2. reviews the relevant previous literature; Section 3. outlines the methods employed in our study; Section 4. presents the results and includes a discussion of the findings; and finally, Section 5. provides conclusion and practical implications.

2. Prior research and theory

Conducting a relative valuation using multiples and peer selection requires addressing several important considerations. To approach this from an academic perspective, we systematically examine the steps outlined in prior research, which often differ from the more flexible methods used in practice. This structured analysis provides a basis for understanding relative valuation and supports the methodology detailed in Section 3.

The following structure is presented in this section, starting in Section 2.1. with an introduction to valuation and how it is done in practice. Followed by a presentation of systemic peer selection in 2.2 and peer ranking methods in 2.3. Furthermore, in 2.4 we elaborate on different industry classifiers and different ways of selecting the number of peers in 2.5. Lastly, in 2.6 we further elaborate on the thesis contribution and present a summary of the most central previous literature.

2.1. Introduction to valuation

Valuation is a central process in finance for determining the value of, for example a firm, an asset or an investment. It provides a foundation for decision-making in areas such as M&A, IPOs, and financial reporting. The two main approaches for firm valuation are (i) intrinsic valuation, and (ii) relative valuation. Intrinsic valuation is a method of determining the value of a firm based on its fundamental financial characteristics (Berk & DeMarzo, 2017). Koller et al. (2020) elaborate that intrinsic valuation, for example the discounted cash flow model (DCF), relies on three principal components: the estimation of future cash flows, the determination of an appropriate discount rate, and the assumption of a terminal value that captures cash flows beyond the explicit forecast horizon. A second intrinsic valuation model is the Dividend Discount Model (DDM). Berk & DeMarzo (2017) describe the DDM Model as a valuation approach that determines a company's value by calculating the present value of its expected future dividend payments. It operates on the principle that an asset's value is derived from the cash flows it generates for its owners.

The alternative method, relative valuation, is widely used in finance due to its simplicity and reliance on observable market data. It involves estimating a firm's value by comparing it to similar companies, based on the main assumption that comparable companies should be valued similar (Fama & French, 1992). It involves using market-based multiples; for example, taking a focus firm's equity or enterprise value to an income, balance sheet, or cash flow item by using multiples such as the P/E, EV/EBITDA, and P/B. This method is particularly prevalent in M&A, IPOs, and equity research, where quick and practical valuation insights are required (Berk & DeMarzo, 2017). A key distinction in relative valuation lies between the two primary data sources: (i) publicly listed companies and (ii) previous similar transactions, both of which serve to assess a focus firm's value by evaluating similarities to potential peers based on growth, profitability, and risk. The following formula shows the relationship between the multiple, firm value and value driver:

$$Valuation\ multiple = \frac{\text{Firm value}}{\text{Value driver}}$$

2.1.1. Linkage and contrast between intrinsic and relative valuation

Intrinsic and relative valuation methods are interconnected to an extent as the DDM model can be used to drive equity-based multiples for relative valuation. According to Knudsen et al. (2017), by assuming a constant dividend growth rate in perpetuity (g_D), the DDM model can be used as a simple expression to drive equity-based multiples:

$$P = \frac{D}{r_e - g_D}$$

Here (P) is equivalent to the value of equity according to the market, (D) stands for dividend and lastly cost of equity (r_e). Further, by taking net earnings (E) times the Payout ratio, the following equation is presented:

$$P = \frac{E \times \text{Payout ratio}}{r_e - g_D}$$

If we instead of E take ROE times the book value of equity (B), the following equation is presented:

$$P = \frac{\text{ROE} \times B \times \text{Payout ratio}}{r_e - g_D}$$

The payout ratio can further be described by the retention rate, i.e. the part of net earnings reinvested in the business. If we then divide it with B, we have the price-to-book multiple (P/B):

$$\frac{P}{B} = \frac{\text{ROE} \times (1 - \text{Retention rate})}{r_e - g_D} = \frac{\text{ROE} - g_D}{r_e - g_D}$$

Lastly, by taking the denominator times ROE, it equals the P/E multiple:

$$\frac{P}{E} = \frac{\text{ROE} - g_D}{r_e - g_D} \times \frac{1}{\text{ROE}}$$

While intrinsic and relative valuation methods share some conceptual foundations, they differ fundamentally in their approaches and underlying principles. Intrinsic valuation focuses on the firm itself, relying on assumptions about factors such as interest rates and growth projections to estimate values. In contrast, relative valuation benchmarks a group of comparable firms, drawing on their observed market values rather than relying on model-generated valuations (Berk & DeMarzo, 2017; Koller et al., 2020).

One key critique against intrinsic valuation models, as mentioned by Miller & Modigliani (1961), is that they are sensitive to underlying assumptions, such as growth rates and discount rates, which may introduce bias or imprecision. Furthermore, Koller et al. (2020) emphasise the difficulty of accurately forecasting future cash flows, particularly for firms in volatile industries, which can lead to substantial variability in valuations. Additionally, some scholars argue that intrinsic valuation models often fail to capture the dynamic interplay of market sentiment and macroeconomic factors influencing asset prices in the short term (Zhang, 2024).

Relative valuation is a favoured approach for IPO-valuation as it has a high effectiveness in capturing prevailing market dynamics (Roosenboom, 2007). This is especially relevant in public markets, where sentiment-driven fluctuations can significantly impact investor perceptions. Its ease of implementation combined with its ability to benchmark a focus company against comparable peers using observable market multiples are other advantages of relative valuation (Forte et al., 2020). However, it rests on the assumption that the law of one price holds thus equity markets must be efficient (Fama, 1991). According to Maleki (2003) there are times when the stock market is less efficient in utilising information. For example, in 1999 when the stock market was in a "bubble" in certain sectors, intrinsic models could be more favourable due to their lower reliance on market dynamics. Furthermore, Skogsvik & Skogsvik (2009) emphasise the difficulty in identifying a firm that is a "perfect twin", as firms do not tend to be identical on all dimensions.

2.1.2. Valuation in practice

Practitioners often favour relative valuation for its simplicity and ease of use in benchmarking. As shown by Bradshaw (2001), out of 103 sell-side analyst reports from U.S. investment banks, 87% use earnings multiples, primarily P/E. This is further supported by a more recent study by Pinto et al. (2018) on valuation practices in global equity markets, in which they find that 93% of practitioners commonly use multiples form publicly traded peers, with a preference for P/E and EV/EBITDA. They also find that the second most popular method used is intrinsic valuation, for example 79% applied present discounted value models like the DCF. But relative valuation, particularly through trading multiples such as P/E and EV/EBITDA, is noted for its simplicity, market alignment, and adaptability across industries.

Eaton et al. (2021) provide a different perspective as they highlight how external incentives, like deal-driven fee structures, may lead to biased peer selection and overstated valuations. This undermines valuation accuracy, as the selection of peers may prioritise maximising the likelihood of securing the deal over ensuring objective comparability and precise valuations.

2.2. Systematic peer selection

Prioritising objectivity in identifying comparable peers optimised for valuation accuracy, systematic peer selection ensures consistency by employing predefined criteria to assess similarities between companies (Knudsen et al., 2017). The criteria used can be categorised into two dimensions: (i) financial profile and (ii) business profile. Financial profiles focus on quantifiable metrics that are comparable across all types of firms, aiming to capture characteristics such as growth, profitability, and risk (Bhojraj & Lee, 2002), with metrics as ROE and operating margin. In contrast, business profiles are inherently more qualitative, primarily assessed through industry classification.

These two dimensions interact in a complementary manner and typically involve using the business profile to delimit the sample of potential peers through industry belonging, after which financial characteristics are used to identify the most similar firms. While peers can also be matched independently based on either dimension alone, several studies have shown that valuation accuracy improves when peers are matched using both their financial and business profiles (Alford, 1992; Cheng & McNamara, 2000).

2.2.1. Multiple choices

An important choice when constructing a relative valuation model is the choice of multiple to use. Price-based multiples, such as the P/E and P/B multiple, consider the equity the side of a company, focusing on metrics relevant to shareholders. For example, as earlier stated, the P/E-multiple compares a company's market value of equity to its earnings, which are attributed to shareholders. In contrast, EV-multiples, such as EV/EBITDA, account for the valuation of the entire firm, incorporating both equity and debt holders to provide a holistic view of the company's capital structure. Moreover, the EV/EBIT multiple can be used to assess companies' operational efficiency (Berk & DeMarzo, 2016). Previous literature such as Knudsen et al. (2017) find that depending on the choice of multiple, the model accuracy varies. For example, the industry consumer discretionary had the highest accuracy when using the EV/EBIT multiple, while for the energy sector the P/E multiple had a higher accuracy.

Using multiples in relative valuation, for example P/E, one can look at (i) historical multiples, derived from past earnings data or (ii) forward-looking multiples based on, for example an analyst's or consensus forecasted earnings. Both perspectives are tools in financial analysis, providing insights into market expectations and the historical performance of a company (Berk & DeMarzo, 2017). When comparing the two and looking at their ability to capture the future valuation, Kim & Ritter (1999) show that forward-looking P/E multiples provide greater valuation accuracy compared to historical P/E multiples. Similarly, the market tends to focus on the future as Plenborg & Pimentel (2016) finds that the firm-values tend to be based on future expectations.

2.2.2. Peer group estimation

In formulating the focus company's estimated valuation from the peer group's multiples, three main approaches have emerged in previous literature: the arithmetic mean, harmonic mean and median. For the following formulas, *n* represents the number of peers.

$$Arithmetic\ mean = \frac{\sum_{i=1}^{n} Multiple_i}{n} \qquad Harmonic\ mean = \frac{n}{\sum_{i=1}^{n} \frac{1}{Multiple_i}}$$

$$Median = \ central\ position$$

Alford (1992) and Cheng & McNamara (2000) use the median and reasons that it is useful for estimation by avoiding extreme values in the P/E multiples. On the other hand, the arithmetic mean is more nuanced than the median. However, it is seen by Baker & Ruback (1999) to be a more "simple average" of a set of numbers compared to the harmonic mean which is more commonly used for rates or multiples (Baker & Rudbeck, 1999). Moreover, Baker & Rudbeck (1999) utilises Monte Carlo simulations to show that industry multiples by harmonic means are more accurate than value-weighted arithmetic means. Most prior literature uses the harmonic mean in formulating peer group estimates. For example, Liu et al. (2002), and Dittmann & Weiner (2005) find harmonic mean to be the one with the highest accuracy and particularly effective when accounting for outliers as it is not as impacted by extreme values. Lastly, Knudsen et al. (2017) find harmonics mean to have a superior performance supported by a sensitivity analysis.

2.2.3. Prediction accuracy

To evaluate the accuracy of a focus firm's estimated value, most previous literature studies a sample of public companies and compares the estimated value with the actual market valuation (e.g., Boatsman & Baskin, 1981; Alford, 1992; Cheng & McNamara, 2000; Liu et al., 2002; Lie & Lie, 2002). It is important to acknowledge a mismatch between the practical contexts in which relative valuation is often applied and the settings in which systematic peer selection methods are typically evaluated. For instance, Steffen (2023) highlight that valuations can vary significantly between public and private markets due to factors such as discount for lack of marketability, which attributes a 25-35% discount to private market companies compared to the publicly traded companies.

While the observable market price is commonly used as a benchmark due to the availability of data, Bhojraj & Lee (2002) introduce an alternative approach by measuring prediction errors relative to forecasted market values over a period of up to three years. They argued that market prices tend to converge toward intrinsic values over time. However, this view was challenged by Sloan (2002), who questions the reliability of using forecasted market prices as a benchmark, reasoning that if the current observable market price is not an accurate indicator of value, there is little reason to believe that future prices – or the prices of peers – would offer greater precision.

There are two main methods to evaluate the accuracy of a valuation estimate according to previous literature, the percentage error (PRE) and absolute percentage error (APRE). The PRE, included in prior literature (e.g., Lui et al., 2002; Lie & Lie, 2002), indicates whether the model tends to overestimate or underestimate the actual value. However, when looking at the accuracy of more than one observation, over- and underpredictions can offset each other, obscuring the accuracy of the valuation model. For this reason, APRE, is widely used in previous studies (e.g., Boatsman & Baskin, 1981; Alford, 1992; Cheng & McNamara, 2000; Lui et al., 2002; Lie & Lie, 2002; Knudsen et al., 2017), and is preferred because it provides a clearer and more objective assessment of the aggregated valuation accuracy of a model. The APRE measures the magnitude of the error between the predicted and actual value in absolute terms, i.e. irrespective of direction of estimation error, expressed as a percentage of the actual value.

$$PRE_t = \frac{V_t(M_v) - P_t}{P_t} \qquad \mathrm{APRE_t} = \left|\frac{\mathrm{V_t(M_v) - P_t}}{\mathrm{P_t}}\right|$$

2.3. Financial peer ranking methods

Over time, various methods have been developed to rank similarities of comparable companies based on their financial profiles. The distinctions among peer ranking methods lie in whether they assume a linear or non-linear relationship between financial characteristics of peers and the estimated valuation of the focus firm. Non-linear methods, dominate both in the number of tests conducted and in accuracy (Knudsen et al., 2017). These methods are preferred for their ability to capture the complexities of financial relationships more effectively than linear methods.

2.3.1. Linear financial peer ranking methods

Linear models establish a regression between valuation multiples and financial characteristic measures (Liu et al., 2002; Bhojraj & Lee, 2002). Bhojraj & Lee (2002) apply a regression model to analyse a sample of 1,498 U.S. firms from 1982 to 1998. Their study examines the relationship between valuation multiples – specifically EV/Sales and P/B – and eight financial metrics, including operating margin and ROE. By applying the focus firm’s financial data to the regression model, they introduced the concept of a "warranted multiple”.

$$Warranted\ Multiple_t = \alpha + \sum_{i=1}^{n} \beta_{i,t-1} \times X_{i,t} + \epsilon_t$$

- α: is the intercept of the regression.
- $\beta_{i,t-1}$**:** The regression coefficient for financial metric derived from the previous year.

- $X_{i,t}$: The value of the financial metric for the focus firm in the current year.
- ϵ_t: The error term in the regression.

The warranted multiples can be applied directly as an estimate of the focus firm's value. However, what sets Bhojraj & Lee (2002) apart from other linear methods is their innovative use of warranted multiples in an indirect peer selection approach. This method involves constructing a peer group by identifying firms with the most similar warranted multiples. The harmonic mean of the peer group’s multiples is subsequently used to estimate the valuation multiple for the focus firm.

The linearity of the approach stems from two aspects: (i) the use of a linear regression model to generate the coefficients; and (ii) the assumption in the direct estimation that changes in the focus firm’s financial characteristics have a proportional, linear effect on the warranted multiples and, consequently, the implied valuation.

Bhojraj & Lee (2002) argue that warranted multiples could effectively replace industry classifications, offering a more robust and objective framework for relative valuation. However, Sloan (2002) raises several critiques with their approach. He acknowledges the potential for warranted multiples to complement industry classifiers but points out that the study assumes a linear relationship between valuation multiples and their determinants, which is a simplification that will likely not hold in practice. The theoretical relationship between multiples and their drivers is not inherently linear, and such an assumption could limit the model's reliability. Also, Sloan questions the practical relevance of Bhojraj & Lee’s framework, suggesting that if practitioners rely on other factors, such as future earnings growth, when using valuation multiples, the contribution of the warranted multiples approach may be diminished.

2.3.2. Non-linear financial peer ranking methods

Most previous studies utilise a non-linear approach to analyse the relationship between valuation multiples and financial variables. This non-linearity implies that changes in the independent financial variables do not result in a proportional impact on the estimated valuation. Instead, the financial characteristics are primarily used to directly rank the closeness of firms to one another, rather than to establish a fixed relationship between these characteristics and the estimated valuation.

Boatsman & Baskin (1981) implement a non-linear univariate method for peer selection by identifying the closest firm based on ten-year average growth rate. Their findings demonstrated that selecting peers based on the growth rate yields higher accuracy than randomly selecting a company from the same industry, but their study is rather limited by its relatively small sample of 80 companies.

Building on this foundation, Alford (1992) introduces a multivariate financial approach to peer selection. The study explores two methods: (i) selecting the 2% of peers closest in terms of total assets and ROE respectively; and (ii) from these subsets, identifying the

14% of peers closest for each metric, effectively approximating the same number of firms as the 2% test (14% × 14% ≈ 2%). Alford finds that the second multivariate approach achieves significantly higher accuracy compared to the initial univariate methods applied individually.

Additionally, Alford (1992) combines financial profiles with business profile by incorporating industry classification. Specifically, the analysis evaluates the six closest firms based on total assets and ROE, with the potential sample of peers delimited by industry classification. The study concludes that the highest accuracy is achieved when combining financial and business profiles, particularly by using both industry classification and ROE. However, the improvement of relying solely on industry classification is not statistically significant. Cheng & McNamara (2000) expands Alford's work by finding that industry classification with ROE significantly outperforms using industry alone, particularly when evaluating P/E multiples.

Knudsen et al. (2017) further advances these methods by developing the Sum of Absolute Rank Differences (SARD) method, which introduces a multivariate element. In their study, they rank similarities among peers based on five financial metrics, although SARD imposes no limit on the number of metrics that can be used. The method utilises the Manhattan distance, calculated as the sum of the absolute differences in ordinal ranks across all financial variables. For example, if Company A is ranked 1st for ROE and 3rd for implied growth, while Company B is ranked 2nd and 5th respectively, the Manhattan distance is:

$$|1-2| + |3-5| = 1 + 2 = 3$$

The six peers with the lowest total SARD score are selected to form the peer group. Knudsen et al. (2017) observe that accuracy generally improves as more financial variables are incorporated, although results vary depending on the industry and the valuation multiple being examined. The authors argue that no single financial metric can capture all relevant information related to growth, profitability, or risk. Moreover, the SARD model outperforms the indirect linear model by Bhojraj & Lee (2002) in accuracy across all analysed multiples.

2.4. Business profile industry delimitation

As noted earlier, business profiles, determined through industry classification, have historically been used to define the sample of potential peers. Industry classifications offer a standardised framework for grouping companies based on qualitative business aspects such as revenue-generating activities or economic profiles. The dimension enables benchmarking across industries and higher valuation accuracy through peer selection (Alford, 1992; Cheng & McNamara, 2000; Knudsen et al., 2017).

2.4.1. Conventional industry classifiers

Conventional classifiers, which are often characterised by their wide scope and static nature, have often been used in previous studies. First established by the U.S. government in 1937, SIC have served as a primary industry classifier and have been widely used in prior literature (e.g., Alford, 1992; Cheng & McNamara, 2000; Lie & Lie, 2002; Dittmann & Weiner, 2005). It employs a four-digit code to categorise industries based on their primary revenue-generating activities. As seen in Appendix 1, Microsoft (which we use as an example for the structure of all classifiers, in Appendix 1-3) is assigned to the SIC industry 7372, with the first two digits representing a broad “major” industry group, while the latter two specify subcategories. Cheng & McNamara (2000) find that using SIC codes for peer selection significantly improves accuracy compared to the Market baseline. Additionally, peer selection based on three- or four-digit SIC codes demonstrates substantially higher accuracy than using one- or two-digit SIC codes, suggesting that more narrowly defined SIC classifications are better suited for peer selection. Moreover, using four-digit SIC code Cheng & McNamara find valuation accuracy to be higher in larger industries, suggesting that industry size acts as a proxy for factors like entry barriers and competition. For example, smaller industries may indicate less maturity, with high growth firms and unstable performance, making valuation more challenging.

Previous research identifies the proprietary Global Industry Classification Standard (GICS), developed by Morgan Stanley Capital International and Standard & Poor’s, as the most accurate classifier for capital markets research (Bhojraj et al., 2003). They find GICS to demonstrate significantly higher accuracy in explaining stock return comovements and cross-sectional variations in valuation multiples, than for example SIC. The GICS system, which is oriented towards financial markets, assigns firms individually based on how they are perceived by investment professionals. Similar to SIC, it employs a hierarchical code structure to categorise companies (see Appendix 1). Knudsen et al. (2017) also utilise the GICS classification to define business profile.

Despite their widespread use, traditional industry classification systems are not without limitations. Lee et al. (2015) highlight that these systems often act as broad benchmarks, lacking conceptual clarity. The definition of an "industry" and the process of identifying comparable peers frequently depend on subjective judgment. Moreover, traditional classifications struggle to account for the operational diversity of firms, as companies within a single industry group may have activities spanning over multiple industries.

2.4.2. Alternative industry classifiers

In response to the disadvantages of the conventional industry classifiers, alternative classifiers have been created that are often more dynamic. For example, Lee et al. (2015) explore peer selection based on co-search patterns on the SEC’s EDGAR database (containing SEC financial filings). They find that the search-based patterns have higher explanatory power than GICS when it comes to cross-sectional monthly returns, valuation multiples, financial ratios and fundamental characteristics. However, its practical use is limited as the classifier is not publicly available.

A publicly accessible and alternative approach to traditional industry classifications is TNIC, developed by Hoberg & Phillips (2016). TNIC is a dynamic system that relies on textual analysis of product descriptions within companies' mandatory SEC 10-K filings. By using natural language processing it evaluates the similarity of firms based on the overlap in language describing their products. This approach generates pairwise similarity scores, forming tailored networks where each company is linked to a unique set of peers, creating a firm-specific classification. In Appendix 2, the five most similar peers to Microsoft are presented over the years, showcasing TNIC's adaptability as the most similar peers and the number of identified peers change over time. TNIC's reliance on standardised and mandatory disclosures ensures a robust and detailed data source, with its dataset spanning from 1988 to 2021.

Empirical studies validate TNIC's effectiveness as a classifier in several areas. For example, Eaton et al. (2021) analyses over 3,900 M&A transactions from 1995 to 2017 and demonstrates TNIC's superior predictive power over SIC in terms of peers actually used by practitioners. Also, a master's thesis from the Stockholm School of Economics by Adebäck & Haqués (2022) finds the prediction accuracy to be significantly higher when using TNIC for peer selection compared to conventional industry classifiers.

A subsequent study by Hoberg & Phillips (2023) introduces a new classifier, D2V which builds on TNIC but aligning more closely with traditional industry classifiers by defining industries. The D2V classifier creates 300 distinct industries derived from all product descriptions in SEC 10-K filings, each industry is represented by the most frequently occurring terms. For example, industry four could be labelled as the Cosmetic industry, characterised by words like "cosmetics" and "beauty", while industry 117 corresponds to the Financial industry, with terms such as "banks" and "depository".[2] Companies are assigned to one or more of these industries based on the similarity of their product descriptions to the defining industry terms. Appendix 3 presents Microsoft's classification under D2V, revealing that it is categorised into 15–25 industries depending on the year, with its primary industry (the one with the highest similarity score) varying across the years.

According to Hoberg & Phillips (2023) the D2V approach enhances the original TNIC by employing dense word vectors, which more effectively capture the nuances of word meanings and relationships. This allows D2V to combine the dynamic adaptability of TNIC with the predefined structure of traditional classifiers. Furthermore, they argue that D2V classification better encompasses firm similarity than other classifiers, as evidenced by its profitability prediction R^2 of 24%, surpassing the original TNIC's 19% R^2, which in itself was 46% more accurate than the SIC classifier.

[2] More detailed industry descriptions and the complete data library created by Hoberg & Phillips are accessible online via the Tuck School of Business at Dartmouth.

2.5. Peer group sizes

One methodological aspect that has received limited attention in prior research is how many peers to include in the peer group. Previous studies typically adhere to a fixed number of peers, offering little flexibility in this regard. However, in practice, the size of peer groups used for relative valuation in M&A transactions varies significantly. Eaton et al. (2021) find that peer group sizes ranges from one peer to as many as 48 peers, with an average of eight to ten peers per M&A transaction.

As already mentioned, one of the first systematic peer selection methods by Boatsman & Baskin (1981) conducted a study limited to a sample of 80 firms where each firm only needs at least one peer for their comparison. Alford (1992) raises critical questions about this as only selecting one peer to a prediction of the price leads to a higher standard error in comparison with larger peer group sizes. As such, Alford broadened the peer group size for the tests when using only industry delimitation to ensure that all groups contain at least six peers. He uses a stepwise approach across different SIC digit levels – from 4-digit down to 1-digit – until finding the criteria of least number of firms (LNOF) in the peer group, which results in a mean of between 16-20 peers in analysing the peer groups (Alford, 1992). This approach entails that the peer group size varies depending on which industry a firm is allocated to, and in which SIC digit hierarchy the focus firm has at least six potential peers. Cheng & McNamara (2000) further investigate the LNOF criteria and find that using at least six peers produces the highest accuracy, when testing a range of two to eleven number of peers.

When incorporating the financial profile into peer selection methods, previous research consistently maintains fixed peer group sizes. The financial profile's primary role is to rank firms by their relative similarity, enabling a clear cutoff point to exclude firms with insufficient similarity. This fixed peer group size approach ensures consistency and comparability of results across studies and focus firms by maintaining the same peer group size for all. Alford (1992) and Dittmann & Weiner (2005) base variables related to financial profile on a fixed number of peers of the 2% closest ROE, ROA and total assets in the sample (for Alford (1992), this was approximately 30 firms). Dittmann & Weiner (2005, p. 6) acknowledge a gap in the selection of peers, as they state that: "*Again the choice of 2% is arbitrary and, to our knowledge, has not been subject to a rigorous empirical study*.". Subsequently, Cheng & McNamara (2000) use a different approach of six peers for all groups, as they state that they are not aware of an optimal number of peers when using a fixed peer group size approach. Thus, they would rather have a fixed number of firms to assure that the difference in performance is not due to the size of the peer groups. Combining the business profile with the financial profile, all three studies above and Knudsen et al. (2017) adhere to a fixed peer group size of five to six peers.

Searching the sensitivity in the number of peers in fixed peer groups sizes, Knudsen et al. (2017) find higher accuracy in peer groups with between 6-16 peers, after testing groups of 4-100 peers. Moreover, they find that the valuation accuracy decreases linearly after

22 peers. Though, Knudsen et al. (2017) use six peers as default with the reason that it aligns with the convention in previous research referring Alford (1992).

2.6. Summary and contribution

Our review of previous literature highlights a widespread use of relative valuation among practitioners, particularly in private settings such as M&A transactions and IPOs. Systematic peer selection is based on the principle that comparable assets should have similar valuations and that the similarity of firms can be categorised into two dimensions, financial profile and business profile. Financial profiles are typically used to rank firms based on their relative similarity of firms, employing either univariate or multivariate peer ranking methods, which assumes a linear or non-linear relationship between valuation multiples and financial metrics. In contrast, business profiles have historically relied on conventional, static industry classifications to delimit the sample of potential peers. Combining these dimensions in peer selection generally yields the highest valuation accuracy. Furthermore, a notable area of consistency in prior research is the fixed number of peers in each peer group when combining both dimensions. Although, there is limited investigation on whether this assumption restricts the accuracy of systematic peer selection methods.

This study aims to contribute to the existing literature by comparing the performance of traditional classifiers, such as GICS and SIC, with novel, dynamic product-description-based classifiers, TNIC and D2V. Thus, we intend to answer the question of which industry classifier is most suitable for peer selection in relative valuation. While TNIC's performance has been previously explored in similar contexts, D2V, to the best of our knowledge, is completely novel in this setting. Given D2V's predefined industry structure, which aligns more closely with conventional classifiers than TNIC and its superior profitability prediction accuracy compared to TNIC and SIC, its inclusion offers a valuable opportunity to test in this context. Additionally, we present an idiosyncratic peer group size approach and test it against the traditional fixed peer group size approach. This will allow us to evaluate whether the rigid fixed peer group size assumption used in prior peer selection methods has limited their flexibility and overall effectiveness of the systematic peer selection.

In Table 2.1, we present an overview of the most influential studies on systematic peer selection for relative valuation. The table systematically presents information on the descriptive aspects of each study, along with their methodologies and key findings. Additionally, the final row includes a summary of our own study for comparison.

Table 2.1: Overview of previous literature

This table summarises the most relevant literature for this study, spanning from Boatsman & Baskin in the early 1980s to Knudsen et al. (2017), and culminating in our own analysis. It provides descriptive details, including the author, year of publication, journal, and sample used (unique firms refers samples where total observations are not presented). Furthermore, the table outlines the structure of each method, covering dimensions analysed, peer ranking methods, industry classifiers, dependent variables, the least number of firms (LNOF), peer group size conditions, and key findings.

Author (Year)	*Journal*	*Sample*	*Method dimensions*	*Peer ranking method*	*Industry classification*	*Dependent variable*	*LNOF*	*Peer group size condition*	*Findings*
Boatsman & Baskin (1981)	The Accounting Review	Unique U.S. firms: 80 between 1957-1976	Business & financial profile	Non-linear univariate method	COMPUSTAT	P/E (Historical)	One	Fixed peer group size	Industry classification and earnings growth had higher valuation accuracy
Alford (1992)	Journal of Accounting Research	Unique U.S. firms: 1,471-1,636 between 1978-1986	Business & financial profile	Non-linear univariate method	SIC (4 to 1 digit)	P/E (Historical & Forward)	Six	Fixed and variable peer group size	Industry classification, alone or combined with ROE had higher valuation accuracy
Cheng & McNamara (2000)	Review of Quantitative Finance and Accounting	U.S. firms: 30,310 between 1973-1992	Business & financial profile	Non-linear univariate method	SIC (4 to 1 digit)	P/E & P/B (Historical)	Six	Fixed and variable peer group size	Industry classification, alone and combined with ROE had higher accuracy
Bhojraj & Lee (2002)	Journal of Accounting Research	Unique U.S. firms: 741-1,498 between 1982-1998	Business & financial profile	Linear multivariate regression	SIC (2 digit)	EV/S & P/B (Historical)	Five	Variable peer group size	The warranted multiple provided a higher valuation accuracy compared to industry classification and size
Lie & Lie (2002)	Financial Analysts Journal	Global firms: 8,621 in 1998	Business profile	n.a.	SIC (3 digit)	EV/S, EV/B, EV/EBITDA, EV/EBIT, P/E (Historical & Forward)	Five	Variable peer group size	Forward P/E provided higher valuation accuracy
Dittmann & Weiner (2005)	Working Paper. Humboldt University Berlin	Global firms: 67,433 in 1992-2002	Business & financial profile	Non-linear univariate method	SIC (4 to 1 digit)	EV/EBIT (Historical)	Five	Fixed and variable peer group size	Selecting peers based on ROA or a combination of ROA and total assets delivers the higher valuation accuracy
Knudsen et al. (2017)	Financial Analysts Journal	U.S. firms: 12,350 between 1995-2014	Business & financial profile	Non-linear univariate method	GICS (6 digit)	EV/S, EV/EBIT, P/E & P/B (Forward)	Six	Fixed peer group size	SARD had higher accuracy than industry classification alone
Nilsson & Salomon-Sörensen (2024)	**Stockholm School of Economics Master Thesis**	**U.S. firms: 11,443-25,150 between 2010-2021**	**Business & financial profile**	**Non-linear univariate and multivariate method**	**SIC (4 digit), GICS (6 digit), TINC and D2V**	**P/E (Forward)**	**Six**	**Fixed and variable peer group size**	**TNIC has the highest accuracy out of the industry classifiers and idiosyncratic condition for peer group size had highest valuation accuracy**

3. Method and data

This section presents and discusses the methodology and data used in our study. We begin by outlining the dependent variable, followed by the peer group estimation, prediction error measurement method, and industry definitions applied throughout the study. Next, we provide a detailed overview of each systematic peer selection method, followed by our approach for determining the idiosyncratic optimal peer group size. The section concludes with a description of the data sample, including observation exclusions and descriptive valuation statistics. Our methodology draws significant inspiration from Knudsen et al. (2017), which represents the most recent advancement in systematic peer selection.

Throughout this section, we detail the formulation and calculation of the variables used in this study. Data for these variables have been sourced from three established databases: (i) COMPUSTAT, which provides reported financial data (denoted as ⟨*Datapoint*⟩); (ii) CRSP, which supplies market data (denoted as [*Datapoint*]); and (iii) I/B/E/S, a source for consensus analyst estimates (denoted as {*Datapoint*}).

3.1. Dependent variable

We have chosen to use the P/E multiple, specifically with forecasted earnings, i.e. forward looking P/E, as the dependent variable in our study. This decision is supported by the fact that practitioners favour earnings-based multiples like P/E ratio (Pinto et al., 2018) and market values are often based on future expectations (Plenborg & Pimentel, 2016). Furthermore, previous studies demonstrate the reliability of forward-looking multiples compared to the historical (Kim & Ritter, 1999; Liu et al., 2002). Knudsen et al. (2017)[3] suggest that different multiples are suited to different industries. However, we opted for the P/E ratio due to its broad applicability and general accuracy across most industries, and the investigation of multiple accuracy is not the focus of this study. Moreover, we used a one-year consensus analyst forecasts as the value driver in our study to maximise data coverage, prioritising it over a marginal improvement of valuation accuracy using two- and three-year forecasts shown by Liu et al. (2002).

Valuations have been gathered as of March 31 each year, following the approach of Knudsen et al. (2017), as this is typically when most listed companies release their annual reports. To account for firms with non-standard fiscal years, earnings estimates have been collected on a quarterly basis, incorporating the most recent forecast available prior to the valuation date, ensuring that the estimates accurately reflect the latest market expectations. This means forecasted earnings were calculated as the sum of the four closest quarters to the valuation date:

[3] Knudsen et al. (2017) do not explicitly state using forward multiples, but their reference to earnings estimates when discussing exclusions in valuation multiples suggests this approach.

$$Forward\ \frac{P_t}{E_{t+1}} = \frac{[\text{PRC}]_t \times\ [\text{SHROUT}]_t}{\sum_{i=1}^{4}\{NET\}_t^{(i)}}$$

3.2. Peer group estimation and prediction accuracy

The valuation of the focus firm was predicted using the harmonic mean of the peer group multiples, as it mitigates the influence of extreme outliers while incorporating all data points, resulting in a more balanced estimate. We argue that the harmonic mean combines the strengths of both the median – used by Alford (1992) and Cheng & McNamara (2000) – and the arithmetic mean, employed by Boatsman & Baskin (1981). Additionally, it is a method many previous studies favour (e.g., Baker & Ruback, 1999; Liu et al., 2002, Knudsen et al., 2017). Moreover, we limited our analysis to estimating the multiple, as taking the additional step of calculating the estimated market value by multiplying the multiple with the firm's value driver, does not provide any further insights into accuracy assessment. Using the harmonic mean, the estimate for the focus firm was thus calculated by dividing the number of peers, *n*, by the sum of the reciprocals of each peer's P/E, as shown below:

$$\overline{P/E}_t = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{P/E_i}\right)}$$

Consistent with prior research (e.g., Boatsman & Baskin, 1981; Alford, 1992; Cheng & McNamara, 2000; Knudsen et al, 2017) we used the absolute percentage error, APRE, to assess precision. While previous studies, such as Liu et al. (2002), also employ PRE to indicate the direction of the error, we did not present this measure, as over- and underpredictions can cancel out in aggregated results, making PRE only meaningful at the individual observation level. APRE, measuring only the magnitude of prediction error, was calculated dividing the absolute difference between the estimated multiple and the actual multiple, divided by the actual multiple. Values close to zero indicate high accuracy of the estimate:

$$APRE_t = \left|\frac{\overline{P/E}_t - P/E_t}{P/E_t}\right|$$

The APRE results will be presented through multiple metrics to evaluate the accuracy and precision of the peer selection method, including the mean of yearly medians, overall median, overall mean, and the mean of yearly means. These metrics enable comparisons with prior studies by Alford (1992), Cheng & McNamara (2000), and Knudsen et al. (2017), which utilise similar measures. While differences across these metrics are expected, each offers a distinct perspective on the accuracy of peer estimates. The median reflects the midpoint of accuracy, unaffected by outliers, while the mean captures aggregated accuracy, including outliers, and in both cases, lower values indicate higher accuracy. Mean of yearly medians and means[4] capture yearly consistency. Finally, dispersion in accuracy will be assessed using the interquartile range, which measures the spread of the middle 50% of the data. A lower interquartile range indicates tighter clustering around the central tendency, reflecting less variability and greater consistency.

$$\text{Median} = central\, APRE_i \qquad \text{Mean of yearly medians} = \frac{\sum_{t=1}^{y} \text{Median}\ APRE_t}{y}$$

$$\text{Mean} = \frac{\sum_{i=1}^{n} APRE_i}{n} \qquad \text{Mean of yearly means} = \frac{\sum_{t=1}^{y} \text{Mean}\ APRE_t}{y}$$

$$Interquartile\ range = Quartile\ 3 - Quartile\ 1$$

We employed the Wilcoxon signed-rank test to evaluate the statistical significance of differences across methods and classifiers. The Wilcoxon test is a non-parametric test, suitable for data that may not adhere to a normal distribution. Knudsen et al. (2017) demonstrated through a normality test that APRE deviated from the standard assumption of normal distribution, due to the presence of outliers, positive skewness, or heavy tails in the APRE. By not relying on the assumption of normality, the Wilcoxon test provides a robust and reliable method for assessing differences in accuracy, even in non-standard distributions.

The Wilcoxon test works by ranking the absolute differences between two methods (e.g., industry classifiers or peer group size approaches) for paired observations, in our case, the annual APRE from estimation of the focus firms' valuation. Positive or negative signs are assigned to the ranks based on the direction of the difference, i.e. which method performs better for a given observation. The sum of these signed ranks is then compared to the expected distribution of difference in APRE between test under the null hypothesis, which assumes no median difference between the methods. This approach enables the Wilcoxon test to detect shifts in the median difference while remaining robust to the influence of extreme outliers in the magnitude of APRE differences, as it considers only the absolute rank differences between the methods being tested. In Section 4.1., we provide practical examples from our results to illustrate how the Wilcoxon test output should be interpreted.

3.3. Industry classifiers

As has been previously mentioned our study aims to assess whether the alternative product-description-based classifiers, TNIC and D2V, can outperform conventional methods in peer selection for relative valuation. Given the differing structures of the classifiers, it is essential to harmonise them to ensure a meaningful comparison.

[4] Mean of yearly means has not be used in previous research.

3.3.1. Industry classification harmonisation

Similar to prior studies, industry classifiers have been applied to restrict the sample of potential peers to companies within the same industry. This approach makes it essential that all classifiers operate at comparable levels of granularity, as for example the conventional classifiers have different classification hierarchies. Alford (1992), Cheng & McNamara (2000) and Dittmann & Weiner (2005) use a stepwise approach to the hierarchy levels of classifiers to meet the LNOF peer group size requirement. However, we have opted not to adopt this approach, because both TNIC and D2V groups firms based on similarity above a specific threshold, aligning their granularity[5] with that of three-digit SIC. This alignment explains the "3" in the names TNIC-3 and D2V-TNIC-3. Both TNIC and D2V, along with the three-digit SIC code, classify approximately 2.05% of firm pairs as belonging in the same industry.

We have also looked at the GICS classifier, as it has previously been found to be the most accurate industry classifier for valuation purposes (Bhojraj et al., 2003). Consequently, we have identified that GICS industry level classification (six-digit hierarchy, ⟨gind⟩) best matches the 2% threshold, with a 2.59% granularity. Additionally, Knudsen et al. (2017) also looked at the six-digit GICS in their study.

Furthermore, the D2V classifier allows companies to be assigned to multiple industries within a given year. We thus analysed two potential approaches to using this classifier: (i) took into consideration all industries a firm is allocated to; and (ii) focused solely on the primary industry[6], resulting in a more limited set of potential peers. The subsets have been named D2V Full and D2V Primary respectively.

To gain an initial understanding of the relative performance of industry classifiers, we have conducted a preliminary test using each company's assigned industry group as its peer group, like previous studies (e.g., Alford, 1992; Cheng & McNamara, 2000; Dittmann & Weiner, 2002; Lie & Lie, 2002). For this test, industry harmonic mean (excluding the focus company) has served as estimates for the valuation of the company under investigation. While this approach is relatively simplistic – focusing solely on the business profile dimension of peer selection and omitting the financial profile – it provides an initial indication of the relative suitability of industry classifiers for systematic peer selection. Moreover, prior research shows that relying exclusively on industry classifiers often yields better results than considering financial characteristics alone (Alford, 1992; Cheng & McNamara, 2000; Knudsen et al., 2017).

[5] Defined as the probability that two random companies are assigned to the same industry within a given year.
[6] Defined as the industry with the highest similarity score for each company.

3.3.2. Expected performance of industry classification

As Hoberg & Phillips (2023) demonstrate that D2V outperforms both TNIC and SIC in profitability prediction accuracy, we aim to investigate its effectiveness in the context of valuation accuracy for relative valuation. While profitability is a key determinant of company valuation, suggesting D2V could also excel in valuation prediction, profitability is not the sole driver of value. Therefore, we do not have a definitive expectation regarding D2V's relative performance compared to the other classifiers.

For the other classifiers, we anticipate that TNIC will be the most accurate. Supporting this expectation, Eaton et al. (2021) identify TNIC as more effective than SIC in identifying peers used by investment banks in valuations for M&A transaction. Adebäck & Haqués (2022) also find in their comparative analysis that TNIC outperforms GICS in systematic peer selection for relative valuation. Moving forward, GICS is anticipated to rank second in accuracy. Research by Bhojraj et al. (2003) highlights its relative effectiveness among traditional industry classifiers in capital markets research, demonstrating it to be more accurate than SIC in that context. This rationale also underpins Knudsen et al.'s (2017) decision to use GICS as their industry classifier. Finally, SIC is likely to rank third as research (e.g., Alford, 1992; Cheng & McNamara, 2000) find it to produce higher valuation accuracy than the using the Market baseline, i.e. no industry classifier at all.

3.4. Peer ranking methods

Out of the peer ranking methods outlined in the literature review, we have chosen the non-linear approaches by Cheng & McNamara (2000) and SARD by Knudsen et al. (2017). We argue that these models are the most developed and they to some degree build on each other. The core principle is to identify comparable peers by ranking companies based on several key variables that reflect a company's growth, profitability and risk, and with the assumption that firms with similar characteristics should have similar valuation multiples (Knudsen et al., 2017). The non-linear methods assume there is not a linear relationship between the dependent variable, in our case forward looking P/E, and the various financial variables such as ROE.

We excluded the linear regression model proposed by Bhojraj & Lee (2002) due to its demonstrated limitations in accuracy when compared to non-linear approaches. This decision is supported by Knudsen et al. (2017), who use the linear model as a robustness test showing that the SARD method outperforms in all scenarios. The warranted multiples concept is also cumbersome to apply in practice, as it requires conducting a linear regression using financial data from the previous year and applying the resulting coefficients to current financials. This process introduces a lag in the model and relies on the assumption that the previous year's linear relationships remain valid for the subsequent valuation period. Liu et al. (2002) also highlight that multiples are appreciated for their simplicity and ease of communication, a characteristic that we argue is undermined by the complexity of a regression models in this context. In contrast, non-linear models enable direct comparison of financial characteristics without introducing a lag, making them both more efficient and easier to comprehend.

Furthermore, some industry classifiers result in small sample sizes for certain focus firms, which violates the assumptions of the central limit theorem, making it impossible to construct reliable regressions in such cases (Knudsen et al., 2017). Lastly, Sloan (2002) raises concerns about the robustness of valuation estimations produced by this model, further supporting our decision to exclude it. Consequently, we find no compelling reason to adopt the linear approach over non-linear models.

3.4.1. Non-linear univariate peer ranking method

Building on the peer ranking method introduced by Alford (1992) and later refined by Cheng & McNamara (2000), we integrated industry classification with one-year ROE. However, unlike Alford (1992) and Cheng & McNamara (2000) – who use historical P/E multiples – we have adopted a forward P/E multiple, as outlined in section 3.1. To align with this forward-looking approach, we have also modified the ROE formulation to use forecasted ROE. Using historical ROE alongside forward P/E introduces a mismatch in temporal alignment, potentially distorting results, especially since Cheng & McNamara (2000) says ROE serves as a proxy for growth. Moreover, we conducted a preliminary analysis comparing historical and forward-looking ROE that supports this adjustment, as it demonstrated a generally greater accuracy when forecasted ROE was employed in the test.

$$ROE_{t+1} = \frac{\text{Net income}_{\text{t+1}}}{Book\ value\ of\ equity_{t+1}} = \{Percentage\ ROE\}_{t+1}$$

It is also worth noting that Alford (1992) and Cheng & McNamara (2000) assessed several variations of non-linear peer selection methods, incorporating different combinations of industry classification, ROE, and total assets. However, only the ROE-based method was used in our study, as it produced the highest accuracy.

3.4.2. Non-linear multivariate peer ranking method

Knudsen et al. (2017) utilised the SARD method, which ranks relative peer similarity by calculating the sum of absolute rank differences across a predefined set of variables for each pair of companies. In this study we first ranked all individual companies in the sample to the potential peer set for each financial variable. Second, the SARD score was calculated for each individual company by summing the absolute differences in their ranks across all variables, providing a comprehensive measure of similarity that considers multiple dimensions simultaneously. An example of the SARD equation between company $_i$ and company $_j$ is presented below:

$$SARDi,j = \mid rX,i - rX,j \mid + \mid rY,i - rY,j \mid + \cdots + \mid rZ,i - rZ,j \mid$$

where:

- $SARD_{i,j}$ is the total score indicating how closely related company $_i$ is to company $_j$.
- $rX_{,i}$, $rY_{,i}$, and $rZ_{,i}$ are the ranks of company $_i$ based on variables X, Y, and Z, respectively.
- A lower SARD score suggests a closer match between the companies, implying they are more similar.
- The ranks are based on the potential peer base for the company in focus, which is delimited by the industry classifier.

Several independent variables were used as proxies for profitability, growth, and risk. Following the example of Knudsen et al. (2017), historical income statements items were calculated as the sum of the four preceding quarters, to accommodate companies with non-standard fiscal years, as discussed in Section 3.1. For balance sheet items, the method relied on the latest available data to the valuation date, ensuring consistency with the snapshot-based nature of these metrics:

- **ROE**: Knudsen et al. (2017) use historical ROE from the prior year. Unlike our adjustment to ROE in Cheng & McNamara's (2000) method, we have retained Knudsen et al.'s (2017) definition when using SARD. Their use of historical ROE alongside forward multiples supports our decision to retain their definition, as we also used a forward multiple. Moreover, although they do not explicitly justify their preference for historical over forward ROE, their method contrasts with Cheng & McNamara (2000), as they treat ROE as a proxy for risk and instead include explicit growth measures derived from forward-looking data.

$$ROE_t = \frac{\text{Net income}_t}{\text{Book value of equity}_t} = \frac{\sum_{q=t-3}^{t} \langle ibq \rangle_q}{\langle ceqq \rangle_t}$$

- **Implied growth**: The measure reflects the firm's expected growth potential, calculated using earnings growth between the first and second forecast years to reduce the impact of earnings normalisation. While growth is often measured using revenue, Knudsen et al. (2017) likely choose earnings growth to mitigate the effects of extreme revenue growth seen in sectors like technology, where companies often experience unprofitable and volatile expansion.

$$\text{Implied growth}_t = \frac{\text{EPS forecast}_{t+2}}{\text{EPS forecast}_{t+1}} = \frac{\{\text{EPS}\}_{t+2}}{\{\text{EPS}\}_{t+1}}$$

- **Operating profit margin (EBIT Margin)**: The EBIT margin is a profitability measure, defined as operating profit relative to sales and provides insight into a firm's operational efficiency.

$$EBIT\ \mathrm{M}argin_t = \frac{\mathrm{EBIT}_t}{\mathrm{Sales}_t} = \frac{\sum_{q=t-3}^{t} \langle oiadq \rangle_q}{\sum_{q=t-3}^{t} \langle saleq \rangle_q}$$

- **Size**: Proxied by market capitalisation as in previous studies (e.g., Alford, 1992; Dittmann & Weiner, 2005). Smaller companies tend to be less liquid and valued at lower multiples, making size an essential consideration according to Knudsen et al. (2017).

$$Market\ capitalisation = \ Stock\ price_t \times Shares\ outstanding_t$$
$$= [PRC]_t * [SHROUT]_t$$

- **Net debt over EBIT**: A leverage and risk indicator, calculated as long- and short-term debt adjusted for cash, divided by EBIT. According to Knudsen et al. (2017) this ratio effectively captures a company's payback capacity and is a core part of credit analysis.

$$\frac{Net\ Debt_t}{EBIT_t} =$$
$$\frac{(Long - term\ debt)_t + Current\ liabilities_t - (Cash\ and\ short - term\ investments)_t}{EBIT_t} =$$
$$\frac{\langle \mathrm{dlttq} \rangle_t \ + \langle \mathrm{dlcq} \rangle_t \ - \ \langle \mathrm{cheq} \rangle_t}{\sum_{q=t-3}^{t} \langle oiadq \rangle_q}$$

We have chosen not to alter the variables Knudsen et al. (2017) use to rank peers in the SARD model. However, we conducted preliminary tests adding three add-on measures from Hoberg & Phillips database reflecting: Scope of operations, Hoberg & Phillips (2023); Vertical integration, Frésard et al. (2019); and Market fluidity, Hoberg et al. (2014). However, since these variables did not improve the accuracy of the SARD model we chose not to continue investigating these variables further.

3.5. Idiosyncratic peer group size approach

One of the objectives of our study is to examine the impact of peer group size on the accuracy of peer selection. Previous research (Alford, 1992; Cheng & McNamara, 2000; Knudsen et al., 2017) consistently employed fixed peer group sizes for focus firms when using methods that looked at both business and financial profiles, with limited discussion on the implications of this approach. Since we know the value of the focus firm in our academic setting, we determined the optimal peer group size for each individual focus firm and generated a heuristic measure applicable in practical settings. By applying these heuristics, we calculated an idiosyncratic peer group size that accounts for the unique characteristics of the industry classifier structure and industry size.

Additionally, we have categorised focus companies into quartiles based on the total number of potential peers to account for variations in industry size across classifiers. Quartile one represents focus firms with the fewest potential peers, while quartile four includes those with the most. These quartiles effectively serve as a proxy for industry size, as the number of potential peers is constrained by industry boundaries. This segmentation acknowledges that industries of different sizes may require different optimal peer group sizes. In one table Cheng & McNamara (2000), similarly segmented industries into five portfolios, based on the number of firms, and found that larger industries achieve higher accuracy, though their analysis relies on a fixed peer group size.

3.5.1. Step 1: Optimal peer group size approach

The first step of our unique approach involved determining the lowest APRE, i.e. the optimal performance of the peer ranking method, for each focus firm. Using the peer ranking methodologies, we ranked each company's potential peers by financial similarity, based on either forward ROE, as in Cheng & McNamara (2000) or the five variables outlined for Knudsen et al. (2017). For each focus company, we calculated APRE iteratively, adding one peer at a time from the ranked list. For instance, if there were 100 potential peers for a given focus firm, we generated 100 APRE calculations, starting with only the most similar peer and progressively including all potential peers.

We then looked at the peer group with the highest accuracy and defined two heuristic measures for identifying the optimal peer group size:

- **Optimal peer count:** The number of peers in the peer group that yields the lowest APRE.
- **Optimal ratio:** The ratio of the optimal peer count to the total number of potential peers.

$$Optimal\ ratio = \frac{Optimal\ peer\ count}{Total\ number\ of\ potential\ peers\ available}$$

3.5.2. Step 2: Idiosyncratic peer group size approach

In the second step of our approach, we applied the heuristic measures in a practical setting, meaning where the value of the focus firm is unknown. This approach is referred to as the idiosyncratic peer group size approach, which can be calculated using either the optimal peer count or the optimal ratio.

- **Optimal peer count:** The specific peer count identified in Step 1 is applied to each quartile and classifier instead of the fixed peer group size previous studies use. For instance, if the optimal peer count for a given industry classifier and quartile is determined to be 10, all focus firms within that combination will have exactly the 10 most similar peers in their peer group.
- **Optimal ratio:** The number of peers is calculated by multiplying the optimal ratio with the total potential peers for each focus company. This approach allows for a more tailored peer selection process within each quartile, accommodating variations in the total number of potential peers among focus companies. For example,

applying a 10% optimal ratio means Company A with 100 total potential peers selects 10 most similar peers, while Company B with 200 total potential peers selects the 20 most similar. Moreover, the optimal ratio will differ depending on industry classifier and industry size, like optimal peer count.

The optimal ratio aligns conceptually with the peer selection approach used by Alford (1992) and Dittmann & Weiner (2005), who identified the 2% most similar firms as peers based on financial characteristics. However, unlike their fixed-percentage approach, the optimal ratio is adaptable across varying industry classifiers and industry sizes. In contrast, their method is not applied to industry-delimited samples, resulting in a fixed peer group size of approximately 30 firms. This adaptability makes the optimal ratio more context-sensitive and versatile. Furthermore, Dittmann & Weiner (2005) admits that their choice of 2% is largely arbitrary and lacks robust empirical justification, underscoring the importance of more systematic approaches like the optimal ratio.

3.5.3. Evaluation of the idiosyncratic peer approach and expected results

Each peer ranking method was evaluated under three peer group size criteria: (i) a fixed peer group size approach using six peers; (ii) optimal peer group size estimating the optimal peer count and optimal ratio; and (iii) the application of the heuristic value, optimal ratio, generating the idiosyncratic peer group size. We prioritised the optimal ratio over the optimal peer count as it creates a more tailored peer group size by adapting to each focus firm's total potential peers, whereas the optimal peer count assigns the same peer group size to all focus firms within the same quartile. The results from the first and third approaches have been compared using the Wilcoxon test to assess whether the idiosyncratic peer group size offers an improvement over the fixed peer group size.

We expect that our analysis will reveal variations in the optimal number of peers and the optimal ratio across different industry classifiers, driven by differences in the structure of the industry classifiers and the sizes of the industries. Additionally, we expect the optimal peer count to vary across quartiles within each classifier, as the total number of potential peers can influence prediction accuracy positively. By reductio ad absurdum, a company with exactly six potential peers is statistically less likely to share financial characteristics with its peer group, compared to a company with 1,000 potential peers, ceteris paribus. Due to space limitations in this thesis[7], we do not formalise these expectations as hypotheses, but we will instead focus on high-level observations and discussions to highlight key trends.

Finally, we hypothesise that the idiosyncratic peer group size approach will significantly outperform the fixed peer group size approach, as it better captures the unique characteristics of the industries the focus firms are allocated into.

[7] The Wilcoxon statistical test is a pairwise test, which generates extensive data when presented in tables. To manage space constraints, we have prioritised including other key data points in the tables instead.

H_0: There is no significant difference between the idiosyncratic and fixed peer group size approaches, or the fixed peer group size approach performs better.

H_1: The idiosyncratic peer group size approach significantly outperforms the fixed peer group size approach.

3.6. Empirical data

The study has utilised data from the COMPUSTAT, CRSP, and I/B/E/S databases, along with the industry classifications Hoberg & Phillips database, requiring certain adjustments and exclusions. The analysis has focused on the period from 2010 to 2021, with the start year selected to avoid the effects of the 2008–2009 financial crisis and the end year determined by the coverage limit of the Hoberg & Phillips database. Our dataset has further been constrained by this database, as it provides only approximately 50,000 observations during this period. Additionally, the industry classifiers developed by Hoberg & Phillips are available only for companies filing annual reports with the SEC, restricting the study to U.S. firms. By contrast, the conventional SIC and GICS classifiers, due to their static nature, are less constrained and allow for a more comprehensive dataset.

3.6.1. Exclusion and adjustments

To ensure representativeness, our sample underwent extensive cleaning to address missing or inaccurate data across databases. Additionally, we winsorized the valuation data to mitigate the impact of outliers, targeting valuation levels consistent with the descriptive statistics Knudsen et al. (2017) report. Since our focus is not on comparing methods peer ranking, we did not aim to analyse an identical sample between tests. Instead, we prioritised constructing the most comprehensive sample possible for each individual test while ensuring consistency within each approach for a given test. Details of specific data exclusions and their effects are provided in Appendix 4.

3.6.2. Descriptive statistics

Below, we present the descriptive statistics for Market capitalisation and the dependent variable, Forecasted P/E, for the years 2010–2021 and aggregated for the entire study period. Valuations align closely with those used by Knudsen et al. (2017), who reports a mean Market capitalisation of USD 8,327 million and a mean forecasted P/E of 27.6 for the period 1995–2014.

Table 3.1: Descriptive valuation statistics

The table presents arithmetic Mean, Median and Interquartile for valuation metrics Market capitalisation and Forecasted P/E. The Market capitalisation, gathered at 31st of March each year. Forecasted P/E is the Market capitalisation divided by the one-year forecasted earnings.

	Number of	**Market capitalisation (USDm)**			**Forecasted P/E**		
Year	observations	Mean	Median	Interquartile	Mean	Median	Interquartile
2010	1,921	5,610,948	973,868	2,924,581	22.4	16.8	11.5
2011	2,040	6,328,293	1,212,843	3,606,894	21.7	16.3	10.8
2012	2,050	6,558,844	1,214,701	3,536,010	20.1	14.8	9.5
2013	2,048	7,257,922	1,366,262	3,968,204	20.9	15.7	9.5
2014	2,181	8,352,032	1,685,324	4,753,831	23.8	17.8	11.4
2015	2,188	9,095,563	1,843,247	5,271,903	23.1	17.8	11.4
2016	2,114	8,816,110	1,648,603	4,706,536	21.2	16.4	11.4
2017	2,134	10,154,165	1,972,331	5,604,862	23.7	18.7	11.4
2018	2,171	11,037,443	2,049,265	6,020,974	22.9	16.9	12.4
2019	2,147	11,881,684	2,113,078	6,611,057	21.4	15.4	13.6
2020	2,063	10,671,656	1,480,482	5,226,015	18.4	11.9	13.5
2021	2,093	17,008,592	3,045,220	8,921,711	27.2	19.7	19.8
Overall	**25,150**	**10,655,860**	**1,770,522**	**5,430,179**	**22.1**	**16.5**	**12.8**

4. Results and discussion

In the following section, we systematically present the results derived from the methods outlined in the Section 3. Our aim is to address our research question regarding the most suitable industry classifier for peer selection in relative valuation and to examine the potentially idiosyncratic nature of optimal peer group sizes. Following the approach of Cheng & McNamara (2000), we divided the sample into quartiles based on the total number of potential peers to explore the impact of industry size on valuation accuracy. Results are presented for the full sample and by quartiles in all tables, except Table 4.1 as we use this test only to analyse the relative accuracy of the industry classifiers. Examining accuracy by industry size offers deeper insights into the potential idiosyncratic nature of peer selection, an aspect that has received limited attention in previous research.

The results are presented using multiple metrics to capture different aspects of peer estimation accuracy. These include the arithmetic mean of yearly medians and means for the entire sample, as well as the overall median and mean, providing a comprehensive assessment of the accuracy of the peer selection methods. Additionally, the interquartile range is used to gauge accuracy dispersion. It is important to note that the interquartile range is presented only for the whole sample, as calculating it for each quartile would not be meaningful[8]. Collectively, these metrics reveal various aspects of the method's accuracy, offering a comprehensive view of the effectiveness of each industry classifier in peer selection. All tables are presented in the end of each section to optimise space[9] and to

[8] The interquartile range of a quartile would effectively measure a subset of an already segmented dataset, providing limited additional insight.

[9] The tables each take up two landscape-oriented pages, with the exception of Table 4.1.

easier compare the results. Moreover, we have also categorised our observations and discussions into *Relative accuracy of industry classifiers*, *Impact of industry size* and *Evaluation of peer group size approaches*¸ to enhance clarity and make it easier for the readers to follow the discussions throughout the text.

4.1. Peer group construction using only industry classification

The first test evaluates accuracy by relying solely on the industry classifier harmonic mean of multiples to estimate the focus firm's value. Results are presented in Table 4.1, and only address our industry classification research question regarding the most accurate industry classifier.

Relative accuracy of industry classifiers

The findings support previous research, showing a significant increase in accuracy when using an industry classifier compared to a Market baseline, as all classifiers showed significantly lower prediction error seen in Panel B and C. Among the classifiers, TNIC emerged as the most accurate industry classifier across all measures, followed by D2V Primary. Notably, only these two Hoberg & Phillips (2016; 2023) industry classifiers outperformed some of the others significantly according to the Wilcoxon test in Panel C. This outcome aligns with our expectations to some extent concerning the relative performance of the classifiers.

Moreover, the results highlight the importance of considering multiple metrics when assessing accuracy. For instance, GICS appeared to perform worse than SIC in terms of median APRE. However, the Wilcoxon test revealed no significant difference in central tendency between GICS and D2V Primary, and at the same D2V Primary significantly outperformed SIC. Entailing that in this case, relying solely on median APRE could lead to the misleading conclusion that GICS underperforms SIC, whereas the Wilcoxon test provided evidence that challenged this inference, and potentially indicates the opposite.

Furthermore, the Wilcoxon signed-rank test is often interpreted as a measure of median differences, but it goes beyond this by examining the consistency and direction of differences across paired observations. In this study, each pair consisted of the APRE values for a single focus firm and year, comparing one reference classifier to the classifier under consideration. Instead of simply comparing median APREs, the test assessed whether the comparison classifier consistently produced lower APREs than the reference classifier across these pairs. By ranking the magnitude and sign of differences for each pair, the test identified whether there was a statistically significant trend in one direction.

Additionally, we conducted the test in both directions. Values below the empty diagonal in Panel C reflect comparisons where the column classifier serves as the reference classifier, while values above the diagonal use the row classifier as the reference. For example, in the first column, all classifiers demonstrated significantly lower APREs than the Market, as shown by the negative signs and asterisks below the diagonal. Conversely, in the top row, where Market is the classifier under consideration, the results indicated that it had significantly higher APREs than the other classifiers. The empty diagonal represents self-comparisons, which yield no difference.

Table 4.1: Industry classification peer group

The table displays prediction errors for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC. Results are provided for the entire sample. The dataset analyses 25,150 company-years from the period 2010 to 2021. Panel A presents descriptive statistics of number of observations analysed. Panel B shows the prediction error in terms of Absolute Percentage Error (APRE) across different metrics: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and Interquartile range. The number in the parenthesis indicate the rank of each classifier's performance relative to the other classifiers. Panel C provides the statistical test on pairwise differences in central tendencies using the Wilcoxon signed-rank test, looking at prediction error between industry classifiers, where ***denotes significance at the 1% level, **denotes significance at the 5% level, *denotes significance at the 10% level.

	Market	GICS	SIC	D2V Full	D2V Primary	TNIC
Panel A: Descriptive statistics						
Observations	**25,150**	**25,150**	**25,150**	**25,150**	**25,150**	**25,150**
Panel B: APRE per industry classifier						
Mean of yearly medians	**0.283 (6)**	**0.235 (4)**	**0.230 (3)**	**0.255 (5)**	**0.225 (2)**	**0.195 (1)**
Median	**0.280 (6)**	**0.233 (4)**	**0.229 (3)**	**0.253 (5)**	**0.224 (2)**	**0.193 (1)**
Mean of yearly means	**0.339 (6)**	**0.290 (3)**	**0.293(4)**	**0.312 (5)**	**0.284 (2)**	**0.253 (1)**
Mean	**0.339 (6)**	**0.291 (3)**	**0.295 (4)**	**0.312 (5)**	**0.281 (2)**	**0.251 (1)**
Interquartile range	**0.373 (6)**	**0.318 (3)**	**0.327 (4)**	**0.346 (5)**	**0.302 (2)**	**0.301 (1)**
Panel C: Wilcoxon test - industry classifier						
Market		+***	+***	+***	+***	+***
GICS	-***		-	-	-	+***
SIC	-***	+		-	+*	+***
D2V Full	-***	+	+		+***	+***
D2V Primary	-***	+	-*	-***		+***
TNIC	-***	-***	-***	-***	-***	

4.2. Non-linear univariate peer ranking method

This section applies the peer ranking method by Cheng & McNamara (2000), where peer groups have been constructed by selecting the closest peers in terms of forward ROE within the same industry, as outlined in Section 3.4.1. We present three tables: (i) Table 4.2 evaluates peer selection accuracy across industry classifiers using six peers in all peer groups; (ii) Table 4.3 summarises results for the optimal peer group size approach, including maximum potential accuracy, optimal peer count, and optimal ratio (see Section 3.5.1.); and (iii) Table 4.4 presents the idiosyncratic peer group size approach, applying the optimal ratio to each quartile (see Section 3.5.2.).

4.2.1. Fixed peer group size approach – non-linear univariate method

Due to the non-standardised samples across tests (as harmonisation risks observation loss, discussed in Section 3.6.1.), direct comparisons of accuracy between tests are not appropriate. However, comparing accuracy between classifiers within the same test provides valuable insights, and general trends across tests can also be analysed.

An initial observation from the results is the considerable differences between the mean and median values within each quartile across all classifiers. The mean was consistently 65–95% higher than the median, suggesting either a positively skewed distribution or that the difference is primarily driven by outliers inflating the mean. This aligns with Knudsen et al.'s (2017) argument that the distribution of prediction error is not normally distributed. Consequently, their decision to use the non-parametric Wilcoxon significance test, rather than a parametric t-test, appears prudent.

Relative accuracy of industry classifiers

Notably, the Market classifier exhibited the lowest accuracy, reinforcing prior findings that industry-specific classifiers enhance peer selection precision. Among the other industry classifiers, TNIC consistently achieved the highest accuracy across all metrics, echoing earlier findings by Adebäck & Haqués (2022). GICS ranked second in this test, with D2V Primary and SIC closely following. The differences among these three classifiers were minimal and the D2V Primary only had significantly lower prediction error than D2V Full shown in Panel C.

Moreover, both D2V Full and Market, performed worse than all other classifiers. This likely results from an excessive number of potential peers (e.g., up to 1,569 for D2V Full compared to 241 for D2V Primary). The effect of this is a reduced relevance in the business profile dimension, leading to lower overall accuracy as the delimitation is not precise enough. Notably, D2V Primary is a refined subset of D2V Full, meaning a larger pool of potential peers and entailing D2V Full finds peers with more similar financial profile, ceteris paribus. Similarly, the Market, with no business profile restrictions, consistently finds peers with the closest financial profiles yet performed the worse than all others. This outcome highlights a critical trade-off between financial profile and business profile similarity, with optimal peer selection achieved by balancing both dimensions. Moreover, the findings are supported by Cheng & McNamara's (2000) observation that the more precise four-digit SIC classifications achieved significantly higher accuracy compared to the broader one-digit SIC classifications.

Impact of industry size

Another key observation from the results is that APRE measures were consistently lower in quartile four than quartile one, indicating that accuracy generally improves as the industry sizes increases (with the exception of quartiles two and three, that mostly had lower accuracy than quartile one). This finding is congruent with the analysis from Cheng & McNamara (2000) that found that focus companies allocated to industries with large number of firms tend to have higher valuation accuracy within the same industry classification hierarchy. Moreover, they propose that the size of the industries might be illuminative of qualitative industry characteristics such as barriers to entry, competition and degree of maturity. Where for example a smaller industry would be less mature than a larger industry and thus experiencing more volatility.

Cheng & McNamara's (2000) explanation appears reasonable, though it is difficult to verify empirically due to the challenge of quantifying qualitative industry characteristics. While their explanation highlights the importance of qualitative business profile characteristics, the financial profile aspects also need to be factored in. As noted earlier, the Market classifier always identifies the closest peers in terms of financial characteristics. Similarly, focus companies in larger industries should benefit from a broader pool of potential peers, as this increases the likelihood of identifying financially similar ones. Although this could be examined by analysing financial similarity differences between peer groups across quartiles, such an analysis is beyond the scope of this study.

4.2.2. Optimal peer group size approach – non-linear univariate method

Table 4.3 presents the lowest potential APRE using the peer ranking method by Cheng & McNamara (2000) in Panel A. The optimal peer count is presented in Panel B and the optimal ratio Panel C. These results are derived from calculating the optimal number of peers for each individual focus firm, outlined in Section 3.5.1.

The accuracy metrics are not the primary focus of this section, as generating them in a practical setting would require prior knowledge of the focus firm's valuation, which would negate the purpose of relative valuation. Thus, significance tests are omitted. The main aim is instead to generate heuristic measures for the subsequent idiosyncratic peer group size approach.

Relative accuracy of industry classifiers

The first observation aligns with previous findings: TNIC had the lowest potential APRE indicated by Panel B, followed by GICS, SIC, D2V Primary, D2V Full, and finally the Market. This ranking reinforces the importance of the balance between financial and business profile observed in the preceding section, as the classifiers with the largest potential peer pools, i.e. D2V Full and the Market, produce the lowest accuracy.

Peer group size observations

A similar pattern as in Section 4.2.1. emerges across quartiles, with accuracy generally increasing with the size of the industry. However, unlike the preceding approach, where quartiles two and three often displayed lower accuracy than quartile one, in this test accuracy improvements consistently increased with the size of the industry.

Panel C highlights that the mean optimal peer count often aligned closely with six peers, the criterion used in previous fixed peer group size approach. Combined with the trends in Panel B, where larger industries exhibited consistently higher accuracy, these results suggest that the mean might be the most relevant measure for determining the appropriate optimal peer count to be used in the idiosyncratic approach[10]. Additionally, the two findings indicate that companies in quartiles two and three are disproportionately disadvantaged by the fixed six peer group size approach. This observation reinforces the necessity of adapting peer group size based on industry-specific conditions, rather than relying on a fixed number of peers.

Two minor observations can also be made: First, the optimal peer count increased with industry size across all classifiers (except for quartile two in D2V Full), while the optimal ratio decreased. This suggests that the optimal peer count does not scale proportionally with industry size. Second, there was a substantial difference between the mean and median optimal peer count in quartiles three and four, indicating the presence of outliers within these subsets that significantly inflate the mean peer count.

4.2.3. Idiosyncratic peer group size approach – non-linear univariate method

In the idiosyncratic peer group size approach, we applied the mean of yearly medians for the optimal ratio within each quartile. The mean of yearly median was chosen because when applying the mean measure the peer group sizes are to be considered too high[11], and the yearly median ensures consistency over time. Results are presented in Table 4.7, with descriptive statistics in Panel A, APRE per industry classifier in Panel B, and Wilcoxon test results when comparing peer group size approaches and industry classifiers in Panels C and D, respectively.

Relative accuracy of industry classifiers

The results presented in Table 4.4 revealed minor shifts in the accuracy rankings of industry classifiers. Most notably, the D2V Primary classifier outperformed SIC to become the third most accurate classifier, with this result being statistically significant, unlike in the fixed peer group test. However, these findings suggest that no classifier gains significantly disproportionate advantages from the fixed peer group size approach.

Evaluation of peer group size approaches

The most striking findings arise when comparing the accuracy from idiosyncratic peer group size test with the fixed peer group size test. While the fixed group test exhibited slightly lower overall median APRE, the Wilcoxon test in Panel C demonstrated that the idiosyncratic approach consistently outperformed the fixed group approach across all classifiers.

Furthermore, the mean measure revealed a consistent 20–30% improvement in performance across all classifiers when using the idiosyncratic approach. Interestingly, this improvement remained relatively constant across all quartiles. This may seem paradoxical,

[10] As noted in Section 3.5.3, we chose to apply the optimal ratio in the idiosyncratic peer group size approach rather than the optimal peer count, as it is more adaptable to each focus firm.
[11] Using the mean optimal ratio for D2V Full results in a peer group size of 315 peers for focus firms with the most potential peers (1,569 × 20.1% = 315.4), compared to 28 peers when using the mean of yearly medians (1,569 × 1.8% = 28.2).

as the first quartile in the idiosyncratic approach aligned most closely with the fixed six-peer group size criteria (as seen in Panel A), suggesting less potential for improvement over the fixed peer test. Additionally, the idiosyncratic test had much lower difference between the median and mean measures. From 65–95% difference in the fixed peer test to 21–52% with the idiosyncratic approach, entailing an improvement of between 40-65%. Lastly, the interquartile range decreased by 0–12% across all industry classifiers and sizes. These findings highlight the idiosyncratic approach's ability to better accommodate outliers, resulting in a denser and potentially more normally distributed prediction error sample.

Tabel 4.2: Non-linear univariate peer ranking method, with six peers

The table displays prediction errors for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC. Results are provided for the entire sample as well as segmented by quartiles, based on the total number of potential peers available. The dataset analyses 11,582 company-years from the period 2010 to 2021, drawing peers a broader pool of 20,383 potential company-year with valid valuation and variable data. Panel A display the descriptive statistics, Observations analysed and Maximum potential peers. Panel B shows the prediction error in terms of Absolute Percentage Error (APRE) across different metrics: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and Interquartile range. In the "All" column, numbers in parentheses indicate the rank of each classifier's performance relative to the others. Panel C provides the statistical test on pairwise differences in central tendencies using the Wilcoxon signed-rank test for the full sample (i.e. not the quartiles), looking at prediction error between industry classifiers, where ***denotes significance at the 1% level, **denotes significance at the 5% level, *denotes significance at the 10% level.

	Market	GICS					SIC				
	All	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics											
Observations	**11,582**	**11,582**	2,895	2,897	2,896	2,894	**11,582**	2,896	2,895	2,895	2,896
Max potential peer	**2,188**	**255**	24	63	104	255	**238**	16	46	174	238
Panel B: APRE per industry classifier											
Mean of yearly medians	**0.335 (6)**	**0.253 (2)**	0.290	0.275	0.294	0.185	**0.260 (4)**	0.249	0.268	0.399	0.188
Median	**0.337 (6)**	**0.251 (2)**	0.282	0.277	0.293	0.175	**0.262 (4)**	0.250	0.264	0.372	0.192
Mean of yearly means	**0.601 (6)**	**0.453 (2)**	0.532	0.478	0.491	0.307	**0.481 (3)**	0.439	0.468	0.678	0.358
Mean	**0.603 (6)**	**0.453 (2)**	0.533	0.483	0.490	0.306	**0.483 (3)**	0.441	0.465	0.645	0.379
Interquartile range	**0.451 (6)**	**0.375 (2)**	n.m	n.m	n.m	n.m	**0.396 (4)**	n.m	n.m	n.m	n.m
Panel C: Wilcoxon test - industry classifier											
Market		+***					+***				
GICS	-***						**-**				
SIC	-***	+									
D2V Full	-***	+					+				
D2V Primary	-***	+					**-**				
TNIC	-***	-***					-***				

	D2V Full					D2V Primary					TNIC				
	All	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics															
Observations	**11,582**	2,894	2,896	2,896	2,896	**11,582**	2,895	2,895	2,896	2,896	**11,582**	2,896	2,895	2,896	2,895
Max potential peer	**1,569**	364	593	750	1,569	**241**	11	25	53	241	**368**	20	51	188	368
Panel B: APRE per industry classifier															
Mean of yearly medians	**0.318 (5)**	0.340	0.396	0.298	0.271	**0.255 (3)**	0.279	0.289	0.296	0.182	**0.211 (1)**	0.225	0.244	0.243	0.158
Median	**0.318 (5)**	0.332	0.388	0.296	0.273	**0.253 (3)**	0.280	0.287	0.288	0.183	**0.209 (1)**	0.225	0.243	0.236	0.156
Mean of yearly means	**0.572 (5)**	0.588	0.768	0.511	0.446	**0.483 (4)**	0.531	0.535	0.549	0.333	**0.379 (1)**	0.386	0.437	0.422	0.274
Mean	**0.574 (5)**	0.581	0.752	0.507	0.455	**0.483 (3)**	0.529	0.543	0.528	0.333	**0.380 (1)**	0.391	0.441	0.422	0.265
Interquartile range	**0.437 (5)**	n.m	n.m	n.m	n.m	**0.381 (3)**	n.m	n.m	n.m	n.m	**0.312 (1)**	n.m	n.m	n.m	n.m
Panel C: Wilcoxon test - industry classifier															
Market	+***					**+*** **					**+*** **				
GICS	**-**					**-**					**+*** **				
SIC	**-**					+					**+*** **				
D2V Full						**+*** **					**+*** **				
D2V Primary	**-*** **										+***				
TNIC	-***					-***									

Tabel 4.3: Optimal table non-linear univariate peer ranking method

The table displays the minimum prediction errors, optimal peer ratio and optimal number of peers for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC, using the Sum of absolute rank difference peer ranking method. Results are provided for the entire sample as well as segmented by quartiles, based on the total number of potential peers available. The dataset analyses 11,582 company-years from the period 2010 to 2021, drawing peers a broader pool of 20,383 potential company-year with valid valuation and variable data. Panel A display the descriptive statistics, Observations analysed and Maximum potential peers. Panel B shows the minimum prediction error in terms of Absolute Percentage Error (APRE). In the "All" column, numbers in parentheses indicate the rank of each classifier's performance relative to the others. Panel C shows the number of peers with the lowest APRE. Panel D present the optimal ratio, which is the ratio of number of potentials which generate the lowest prediction error to the potential number of peers. Panel B, C and D is presented across: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and interquartile range.

	Market	GICS					SIC				
	All	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics											
Observations	**11,582**	**11,582**	2,895	2,897	2,896	2,894	**11,582**	2,896	2,895	2,895	2,896
Max potential peer	**2,188**	**255**	24	63	104	255	**238**	16	46	174	238
Panel B: APRE per industry classifier											
Mean of yearly medians	**0.084 (6)**	**0.067 (2)**	0.088	0.077	0.078	0.033	**0.073 (3)**	0.089	0.078	0.102	0.041
Median	**0.080 (6)**	**0.067 (2)**	0.087	0.076	0.077	0.034	**0.073 (3)**	0.089	0.073	0.096	0.042
Mean of yearly means	**0.203 (6)**	**0.172 (2)**	0.212	0.178	0.182	0.115	**0.179 (3)**	0.197	0.182	0.222	0.125
Mean	**0.203 (6)**	**0.171 (2)**	0.210	0.179	0.182	0.115	**0.179 (3)**	0.197	0.181	0.211	0.129
Interquartile range	**0.317 (6)**	**0.237 (2)**	n.m.	n.m.	n.m.	n.m.	**0.247 (4)**	n.m.	n.m.	n.m.	n.m.
Panel C: Optimal number of peers											
Mean of yearly medians	**12.4**	**6.8**	4.7	6.8	8	14.5	**6.2**	3.7	6.5	8.5	13.8
Median	**12**	**7**	5	7	8	12	**6**	4	6	9	10
Mean of yearly means	**311.9**	**23.9**	6.4	13	21.8	54.6	**22.6**	4.8	10.6	28.3	47.8
Mean	**311.7**	**24**	6.4	13.1	21.8	54.7	**22.7**	4.8	10.6	29.8	45.5
Interquartile range	**179.8**	**19**	n.m.	n.m.	n.m.	n.m.	**16**	n.m.	n.m.	n.m.	n.m.
Panel D: Optimal ratio											
Mean of yearly medians	**0.006**	**0.162**	0.330	0.180	0.108	0.077	**0.18**	0.403	0.236	0.086	0.068
Median	**0.006**	**0.161**	0.333	0.172	0.109	0.062	**0.171**	0.40	0.231	0.087	0.05
Mean of yearly means	**0.161**	**0.325**	0.427	0.327	0.281	0.268	**0.334**	0.478	0.368	0.259	0.235
Mean	**0.16**	**0.325**	0.426	0.327	0.281	0.267	**0.333**	0.479	0.367	0.265	0.223
Interquartile range	**0.092**	**0.53**	n.m.	n.m.	n.m.	n.m.	**0.554**	n.m.	n.m.	n.m.	n.m.

	D2V Full					D2V Primary					TNIC				
	All	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics															
Observations	**11,582**	2,894	2,896	2,896	2,896	**11,582**	2,895	2,895	2,896	2,896	**11,582**	2,896	2,895	2,896	2,895
Max potential peer	**1,569**	364	593	750	1,569	**241**	11	25	53	241	**368**	20	51	188	368
Panel B: APRE per industry classifier															
Mean of yearly medians	**0.077 (5)**	0.08	0.126	0.067	0.056	**0.075 (4)**	0.102	0.095	0.079	0.039	**0.065 (1)**	0.089	0.072	0.074	0.035
Median	**0.075 (5)**	0.078	0.121	0.064	0.053	**0.074 (4)**	0.102	0.094	0.067	0.040	**0.065 (1)**	0.088	0.073	0.073	0.034
Mean of yearly means	**0.186 (5)**	0.195	0.223	0.168	0.165	**0.184 (4)**	0.217	0.208	0.190	0.124	**0.163 (1)**	0.195	0.175	0.173	0.110
Mean	**0.187 (5)**	0.193	0.220	0.167	0.167	**0.184 (4)**	0.217	0.209	0.185	0.124	**0.164 (1)**	0.196	0.177	0.173	0.109
Interquartile range	**0.283 (5)**	n.m.	n.m.	n.m.	n.m.	**0.242 (3)**	n.m.	n.m.	n.m.	n.m.	**0.229 (1)**	n.m.	n.m.	n.m.	n.m.
Panel C: Optimal number of peers															
Mean of yearly medians	**11.4**	10.1	9.2	14.7	15.5	**5.3**	3.4	5.3	7.4	13.8	**6.5**	4	7.4	9.1	14.9
Median	**11**	10	9	13	14	**5**	4	5	7	10	**6**	4	7	9	12
Mean of yearly means	**113.9**	52.2	89.5	141.3	175.1	**17**	4	6.9	12.7	46.7	**27.3**	5.3	11.8	26.5	65.8
Mean	**114.6**	52.7	91	142.4	172.4	**17.1**	4	6.8	12.9	44.5	**27.3**	5.3	11.8	26.8	65.5
Interquartile range	**104**	n.m.	n.m.	n.m.	n.m.	**11**	n.m.	n.m.	n.m.	n.m.	**18**	n.m.	n.m.	n.m.	n.m.
Panel D: Optimal ratio															
Mean of yearly medians	**0.025**	0.043	0.019	0.022	0.018	**0.261**	0.460	0.328	0.199	0.084	**0.190**	0.406	0.225	0.103	0.057
Median	**0.024**	0.042	0.019	0.019	0.017	**0.250**	0.444	0.333	0.200	0.059	**0.188**	0.400	0.225	0.105	0.046
Mean of yearly means	**0.206**	0.218	0.187	0.209	0.206	**0.382**	0.508	0.425	0.348	0.258	**0.346**	0.479	0.361	0.288	0.254
Mean	**0.205**	0.219	0.190	0.210	0.201	**0.381**	0.507	0.423	0.349	0.244	**0.345**	0.479	0.361	0.290	0.250
Interquartile range	**0.239**	n.m.	n.m.	n.m.	n.m.	**0.587**	n.m.	n.m.	n.m.	n.m.	**0.580**	n.m.	n.m.	n.m.	n.m.

Tabel 4.4: Non-linear univariate peer ranking method, idiosyncratic optimal ratio

The table displays prediction errors for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC, utilising the optimal ratios from the optimal table for each industry classifier and quartile, identifying a unique number of potential peers for each investigated company. Results are provided for the entire sample as well as segmented by quartiles, based on the number of potential peers available. The dataset analyses 11,582 company-years from the period 2010 to 2021, drawing peers a broader pool of 20,383 potential company-year with valid valuation and variable data. Panel A display the descriptive statistics, Observations analysed, Maximum potential peers, Mean of yearly medians of optimal ratio (the same as the optimal table for each quartile) and Mean of yearly medians peer count. Panel B shows the prediction error in terms of Absolute Percentage Error (APRE) across different metrics: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and Interquartile range. In the "All" column, numbers in parentheses indicate the rank of each classifier's performance relative to the others. Panel C and D provides the statistical test on pairwise differences in central tendencies using the Wilcoxon signed-rank test for the full sample (i.e. not the quartiles). Panel C looks at the prediction error between the peer ranking method using six peers, and the idiosyncratic test, while Panel D compare the prediction error of the industry classifiers in the idiosyncratic test. Where ***denotes significance at the 1% level, **denotes significance at the 5% level, *denotes significance at the 10% level.

	Market	GICS					SIC				
	All	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics											
Observations	**11,582**	**11,582**	2,895	2,897	2,896	2,894	**11,582**	2,896	2,895	2,895	2,896
Max potential peer	**2,188**	**255**	24	63	104	255	**238**	16	46	174	238
Mean of yearly medians of optimal ratio	**0.006**	**0.150**	0.330	0.180	0.108	0.077	**0.161**	0.403	0.236	0.086	0.068
Mean of yearly median peer count	**12.1**	**8.3**	5.8	7.4	8.8	17.6	**8.3**	4.8	7.3	12.4	15.1
Panel B: APRE per industry classifier											
Mean of yearly medians	**0.328 (6)**	**0.243 (2)**	0.274	0.277	0.286	0.172	**0.260 (4)**	0.249	0.268	0.389	0.181
Median	**0.327 (6)**	**0.243 (2)**	0.269	0.275	0.282	0.165	**0.259 (4)**	0.249	0.264	0.370	0.183
Mean of yearly means	**0.424 (6)**	**0.346 (2)**	0.407	0.355	0.375	0.248	**0.370 (4)**	0.351	0.359	0.500	0.276
Mean	**0.425 (6)**	**0.345 (2)**	0.404	0.358	0.376	0.245	**0.370 (4)**	0.352	0.359	0.483	0.287
Interquartile range	**0.450 (6)**	**0.352 (2)**	n.m	n.m	n.m	n.m	**0.373 (4)**	n.m	n.m	n.m	n.m
Panel C: Wilcoxon test - fixed and idiosyncratic											
ROE + Industry with fixed peer group size	+***	+***					+***				
Panel D: Wilcoxon test - industry classifier											
Market		+***					+***				
GICS	-***						-				
SIC	-***	+									
D2V Full	-***	+					+				
D2V Primary	-***	+					-***				
TNIC	-***	-***					-***				

	D2V Full					D2V Primary					TNIC				
	All	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics															
Observations	**11,582**	2,894	2,896	2,896	2,896	**11,582**	2,895	2,895	2,896	2,896	**11,582**	2,896	2,895	2,896	2,895
Max potential peer	**1,569**	364	593	750	1,569	**241**	11	25	53	241	**368**	20	51	188	368
Mean of yearly medians of optimal ratio	**0.020**	0.043	0.019	0.022	0.018	**0.253**	0.460	0.328	0.199	0.084	**0.154**	0.406	0.225	0.103	0.057
Mean of yearly median peer count	**13.4**	8.7	10.2	14.9	25.1	**6.5**	4.3	5.6	7.8	17.8	**8.7**	5.5	7.9	9.3	15.3
Panel B: APRE per industry classifier															
Mean of yearly medians	**0.292 (5)**	0.313	0.359	0.268	0.251	**0.252 (3)**	0.284	0.291	0.299	0.172	**0.214 (1)**	0.217	0.254	0.257	0.167
Median	**0.292 (5)**	0.307	0.351	0.267	0.248	**0.249 (3)**	0.280	0.292	0.283	0.171	**0.214 (1)**	0.214	0.252	0.255	0.159
Mean of yearly means	**0.384 (5)**	0.413	0.434	0.349	0.356	**0.353 (3)**	0.381	0.394	0.391	0.257	**0.288 (1)**	0.282	0.316	0.327	0.230
Mean	**0.385 (5)**	0.408	0.429	0.348	0.355	**0.353 (3)**	0.380	0.397	0.382	0.253	**0.288 (1)**	0.285	0.318	0.327	0.223
Interquartile range	**0.385 (5)**	n.m	n.m	n.m	n.m	**0.361 (3)**	n.m	n.m	n.m	n.m	**0.305 (1)**	n.m	n.m	n.m	n.m
Panel C: Wilcoxon test - fixed and idiosyncratic															
ROE + Industry with fixed peer group size	**+*** **					**+*** **					**+*** **				
Panel D: Wilcoxon test - industry classifier															
Market	**+*** **					**+*** **					**+*** **				
GICS	**-**					**-**					**+*** **				
SIC	**-**					**+*** **					**+*** **				
D2V Full						**+*** **					**+*** **				
D2V Primary	**-*** **										**+*** **				
TNIC	**-*** **					**-*** **									

4.3. Non-linear multivariate peer ranking method

We present three results tables from the test using the SARD peer ranking method introduced by Knudsen et al. (2017). Like the structure outlined in Section 4.2., the tables include: (i) results for the fixed peer group size approach (presented in table 4.5), (ii) results for the optimal peer group size approach (presented in Table 4.6), and (iii) the idiosyncratic peer group size approach (presented in Table 4.7). The following analysis looks at the robustness of the findings from the previous sections, expands on the discussion, and provides validation for the conclusions reached later.

4.3.1. Fixed peer group size approach – non-linear multivariate method

Relative accuracy of industry classifiers

Consistent with the findings in Section 4.2. and previous research, employing an industry classifier significantly improves accuracy, as the Market classifier performed significantly worse than all other industry classifiers. But the results regarding the best-performing industry classifier were not consistent with our previous findings. No single classifier consistently outperformed the others across all measures or with high statistical significance. In the aggregated accuracy measures in Panel B, SIC generally performed the best, but according to the Wilcoxon test in Panel C, its accuracy was not significantly higher compared to the other classifiers. TNIC on the other hand demonstrated significantly higher accuracy than both D2V classifiers and GICS, despite its relatively poor accuracy according to the aggregated accuracy measures, i.e. ranging from relative rank of 2-3.

Thus, examining the SARD method with a fixed six-peer group size revealed that the results are somewhat inconclusive regarding which industry classifier is superior. Relying solely on either the mean or median measures could lead to conclusions that are not fully supported by the broader results measures. Consequently, the finding from Section 4.2. that TNIC is the most appropriate classifier for peer selection is not entirely robust when accounting for different peer ranking methods and peer group size approaches.

Impact of industry size

Consistent with previous findings, notable differences in accuracy were observed across industry sizes. Quartile four consistently exhibited the highest accuracy across all measures, reinforcing the impact of industry size characteristics on the valuation accuracy. These results further validate the robustness of the findings from Section 4.2.

4.3.2. Optimal peer group size approach – non-linear multivariate method

The results when combining the SARD peer ranking method with optimal peer group size approach are presented in Table 4.6. Panel A presents the APRE, the optimal peer count is shown in Panel B, and the optimal ratio is displayed in Panel C.

Relative accuracy of industry classifiers

The results using the optimal peer group size approach revealed a significant shift in the performance of industry classifiers. D2V Full achieved the highest accuracy across all tests, followed closely by Market and TNIC. This contrasted with earlier findings where Market and D2V Full consistently underperformed compared to other classifiers and failed to achieve the lowest potential APRE, despite theoretically having the highest financial similarity. Moreover, we observed opposite results using the optimal peer group size approach when employing the peer ranking method by Cheng & McNamara (2000). However, as noted in Section 4.2.2., the purpose of the optimal peer group size approach is not to compare the relative accuracy of classifiers, as the optimal accuracy cannot be calculated in a practical setting.

Impact of industry size

The results for the optimal peer count and optimal ratio align with the findings presented in Table 4.3. The mean optimal peer count for quartile one often approximated six peers. However, substantial differences in peer counts were observed across quartiles. Moreover, Knudsen et al. (2017) found a linear decrease in valuation accuracy when peer groups exceeded 22 peers, with the optimal range identified as 6–16 peers. They used a fixed number of peers in their analysis, making direct comparisons to our findings difficult. On average, the optimal peer count was below 22 peers both in terms of mean and median (with the exception of quartile four for Market and D2V Full), hence Knudsen et al.'s (2017) finding that in general the peer group size should not exceed 22 peers is supported by our results.

Additionally, the results reinforce the observation in Section 4.2.2. that there is a significant difference between the mean and median optimal peer count for quartiles three and four.

4.3.3. Idiosyncratic peer group size approach – non-linear multivariate method

As in the previous idiosyncratic test in Section 4.2.3., the mean of yearly medians for the optimal ratio was applied to generate the idiosyncratic peer group sizes. Table 4.7 presents the results, including descriptive statistics in Panel A, prediction error in Panel B, and Wilcoxon tests in Panel C and D.

Relative accuracy of industry classifiers

The idiosyncratic approach led to notable changes in the relative performance of industry classifiers. Specifically, TNIC demonstrated a substantial improvement compared to the fixed peer test, achieving significantly higher accuracy than all classifiers except SIC. GICS also exhibited improved relative accuracy, while SIC experienced a decline in overall relative performance. These results suggest that the fixed six-peer group size criteria in the SARD method particularly favoured SIC, and this relative advantage diminished with the idiosyncratic peer group size approach. Interestingly, this observation contrasts with the finding that relative performance remained largely unchanged when using Cheng & McNamara's (2000) peer ranking method.

Looking at all the results from the different tests in Section 4.1., 4.2., and 4.3. in combination, TNIC stands out as the most accurate classifier, all things considered. The anomaly is the fixed peer group size test using SARD, where SIC overall performed with the highest accuracy. GICS, SIC and D2V Primary perform in general quite similar in most regards and Market and D2V Full consistently performs worst. Consequently, while the idiosyncratic approach accounts for aspects like industry classifier structure, this factor appears to be of less importance in determining overall accuracy.

Evaluation of peer group size approaches

The comparison between the fixed peer approach and the idiosyncratic approach revealed consistent trends across both Cheng & McNamara (2000) and Knudsen et al. (2017), as Table 4.7 shows:

- The idiosyncratic test produced significantly more accurate results according to the Wilcoxon test in Panel C.
- Accuracy increased by 20–30% measured by the mean.
- The difference between the mean and median decreased by 50–60% and interquartile range decreased 3-8%.

Hence, we can reject the null hypothesis formulated in Section 3.5.3. and conclude that the idiosyncratic approach outperforms the fixed peer group size approach.

Furthermore, all industry classifiers except D2V Full, the mean of yearly median peer count across all quartiles falls below the 16-peer threshold identified by Knudsen et al. (2017) as the upper limit for preferable peer group sizes. Comparing these results to the idiosyncratic test using the method by Cheng & McNamara (2000) revealed similar patterns, with only quartile four for GICS and D2V Primary slightly exceeding this threshold. These findings suggest that applying the optimal ratio in practice aligns reasonably well with the rule-of-thumb range proposed by Knudsen et al. (2017). Additionally, Eaton et al. (2021) observed that investment banks typically use eight to ten peers, with a practical range spanning from one to 48 peers, which further supports the broader flexibility exhibited by the idiosyncratic approach.

Tabel 4.5: Non-linear multivariate peer ranking method, with six peers

The table displays prediction errors for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC. Results are provided for the entire sample as well as segmented by quartiles, based on the total number of potential peers available. The dataset analyses 11,443 company-years from the period 2010 to 2021, drawing peers a broader pool of 21,901 potential company-year with valid valuation and variable data. Panel A display the descriptive statistics, Observations analysed and Maximum potential peers. Panel B shows the prediction error in terms of Absolute Percentage Error (APRE) across different metrics: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and Interquartile range. In the “All” column, numbers in parentheses indicate the rank of each classifier’s performance relative to the other classifiers. Panel C provides the statistical test on pairwise differences in central tendencies using the Wilcoxon signed-rank test for the full sample (i.e. not the quartiles), looking at prediction error between industry classifiers, where ***denotes significance at the 1% level, **denotes significance at the 5% level, *denotes significance at the 10% level.

	Market	GICS					SIC				
	All	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics											
Observations	**11,443**	**11,443**	2,861	2,860	2,859	2,863	**11,443**	2,865	2,859	2,856	2,863
Max potential peer	**2,188**	**250**	26	68	240	250	**233**	18	71	188	233
Panel B: APRE per industry classifier											
Mean of yearly medians	**0.283 (6)**	**0.234 (1)**	0.296	0.244	0.273	0.164	**0.235 (2)**	0.260	0.261	0.289	0.168
Median	**0.282 (6)**	**0.234 (1)**	0.296	0.239	0.271	0.162	**0.234 (1)**	0.258	0.260	0.287	0.159
Mean of yearly means	**0.508 (6)**	**0.459 (3)**	0.585	0.464	0.467	0.322	**0.444 (1)**	0.491	0.473	0.523	0.288
Mean	**0.507 (6)**	**0.458 (3)**	0.582	0.462	0.467	0.321	**0.443 (1)**	0.490	0.473	0.522	0.286
Interquartile range	**0.418 (6)**	**0.374 (2)**	n.m	n.m	n.m	n.m	**0.356 (1)**	n.m	n.m	n.m	n.m
Panel C: Wilcoxon test - industry classifier											
Market		+***					+***				
GICS	-***						-				
SIC	-***	+									
D2V Full	-***	+					+				
D2V Primary	-***	+					+				
TNIC	-***	-*					+				

	D2V Full					D2V Primary					TNIC				
	All	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics															
Observations	**11,443**	2,860	2,863	2,857	2,863	**11,443**	2,864	2,858	2,858	2,863	**11,443**	2,858	2,859	2,856	2,860
Max potential peer	**1,581**	350	592	715	1,581	**243**	12	25	79	243	**363**	21	55	240	363
Panel B: APRE per industry classifier															
Mean of yearly medians	**0.268 (5)**	0.297	0.335	0.255	0.208	**0.244 (4)**	0.289	0.309	0.256	0.162	**0.235 (2)**	0.278	0.282	0.261	0.159
Median	**0.268 (5)**	0.292	0.339	0.255	0.204	**0.243 (4)**	0.289	0.302	0.253	0.161	**0.235 (3)**	0.276	0.284	0.263	0.155
Mean of yearly means	**0.490 (5)**	0.540	0.610	0.453	0.359	**0.487 (4)**	0.564	0.587	0.487	0.309	**0.457 (2)**	0.500	0.535	0.492	0.301
Mean	**0.489 (5)**	0.538	0.610	0.451	0.358	**0.486 (4)**	0.563	0.587	0.485	0.308	**0.456 (2)**	0.497	0.534	0.491	0.300
Interquartile range	**0.399 (5)**	n.m.	n.m.	n.m.	n.m.	**0.391 (4)**	n.m.	n.m.	n.m.	n.m.	**0.375 (3)**	n.m.	n.m.	n.m.	n.m.
Panel C: Wilcoxon test - industry classifier															
Market	+***					+***					+***				
GICS	**-**					-					**+***				
SIC	-					-					-				
D2V Full						**+***					**+***				
D2V Primary	**-***										**+***				
TNIC	-***					-***									

Tabel 4.6: Optimal table non-linear multivariate peer ranking method

The table displays the minimum prediction errors, optimal peer ratio and optimal number of peers for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC, using the SARD ranking method. Results are provided for the entire sample as well as segmented by quartiles, based on the total number of potential peers available. The dataset analyses 11,443 company-years from the period 2010 to 2021, drawing peers a broader pool of 21,901 potential company-year with valid valuation and variable data. Panel A display the descriptive statistics, Observations analysed and Maximum potential peers. Panel B shows the minimum prediction error in terms of Absolute Percentage Error (APRE) across different metrics. In the "All" column, numbers in parentheses indicate the rank of each classifier's performance relative to the others. Panel C shows the number of peers with the lowest APRE. Panel D present the optimal ratio, which is the ratio of number of potentials which generate the lowest prediction error to the potential number of peers. Panel B, C and D is presented across: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and Interquartile range.

	Market	GICS					SIC				
	All	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics											
Observations	**11,443**	**11,443**	2,861	2,860	2,859	2,863	**11,443**	2,865	2,859	2,856	2,863
Max potential peer	**2,188**	**250**	26	68	240	250	**233**	18	71	188	233
Panel B: APRE per industry classifier											
Mean of yearly medians	**0.056 (2)**	**0.058 (4)**	0.088	0.062	0.058	0.034	**0.575 (3)**	0.081	0.065	0.057	0.033
Median	**0.056 (2)**	**0.057 (3)**	0.086	0.062	0.058	0.033	**0.056 (2)**	0.082	0.067	0.056	0.031
Mean of yearly means	**0.171 (5)**	**0.163 (4)**	0.213	0.161	0.162	0.116	**0.160 (3)**	0.192	0.175	0.167	0.108
Mean	**0.171 (5)**	**0.162 (4)**	0.212	0.161	0.161	0.115	**0.161 (3)**	0.193	0.175	0.167	0.108
Interquartile range	**0.249 (5)**	**0.213 (2)**	n.m	n.m	n.m	n.m	**0.214 (3)**	n.m	n.m	n.m	n.m
Panel C: Optimal number of peers											
Mean of yearly medians	**11**	**5.9**	4.1	6.1	7.6	9.1	**5.3**	3.9	5.7	6.7	8.7
Median	**11**	**6**	4	6	8	9	**5**	4	6	7	8
Mean of yearly means	**311.2**	**19.8**	6.2	12.2	20.2	40.4	**18.8**	4.8	9.8	21.2	39.3
Mean	**311.9**	**19.8**	6.2	12.2	20.3	40.4	**18.8**	4.7	9.8	21.2	39.3
Interquartile range	**140**	**16**	n.m	n.m	n.m	n.m	**13**	n.m	n.m	n.m	n.m
Panel D: Optimal ratio											
Mean of yearly medians	**0.005**	**0.133**	0.29	0.15	0.10	0.05	**0.150**	0.40	0.21	0.07	0.04
Median	**0.005**	**0.133**	0.29	0.15	0.10	0.05	**0.151**	0.40	0.21	0.07	0.04
Mean of yearly means	**0.157**	**0.29**	0.40	0.29	0.27	0.20	**0.304**	0.47	0.34	0.20	0.20
Mean	**0.156**	**0.291**	0.40	0.29	0.27	0.20	**0.304**	0.47	0.34	0.20	0.20
Interquartile range	**0.071**	**0.427**	n.m	n.m	n.m	n.m	**0.460**	n.m	n.m	n.m	n.m

	D2V Full					D2V Primary					TNIC				
	All	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics															
Observations	**11,443**	2,860	2,863	2,857	2,863	**11,443**	2,864	2,858	2,858	2,863	**11,443**	2,858	2,859	2,856	2,860
Max potential peer	**1,581**	350	592	715	1,581	**243**	12	25	79	243	**363**	21	55	240	363
Panel B: APRE per industry classifier															
Mean of yearly medians	**0.034 (1)**	0.043	0.051	0.027	0.018	**0.069 (5)**	0.110	0.092	0.054	0.037	**0.056 (2)**	0.091	0.062	0.045	0.032
Median	**0.033 (1)**	0.042	0.046	0.027	0.019	**0.068 (4)**	0.108	0.093	0.054	0.035	**0.056 (2)**	0.088	0.060	0.550	0.032
Mean of yearly means	**0.146 (1)**	0.161	0.167	0.134	0.123	**0.177 (6)**	0.220	0.206	0.167	0.115	**0.159 (2)**	0.196	0.170	0.161	0.110
Mean	**0.146 (1)**	0.160	0.167	0.134	0.122	**0.177 (6)**	0.220	0.206	0.167	0.115	**0.159 (2)**	0.196	0.169	0.160	0.110
Interquartile range	**0.210 (1)**	n.m.	n.m.	n.m.	n.m.	**0.235 (4)**	n.m.	n.m.	n.m.	n.m.	**0.214 (3)**	n.m.	n.m.	n.m.	n.m.
Panel C: Optimal number of peers															
Mean of yearly medians	**12.1**	10.7	11.2	13.5	17.6	**4.9**	3.2	4.5	6.0	8.6	**5.5**	3.9	5.5	6.9	9.4
Median	**12**	10	11	13	18	**5**	3	5	6	8	**6**	4	5	7	9
Mean of yearly means	**320.5**	99	206.9	413.7	762.4	**14.4**	3.9	6.1	11.1	36.5	**21.1**	5.1	10	19.8	49.5
Mean	**319.0**	99.2	207.8	413.9	767.1	**14.4**	3.9	6.1	11.1	36.4	**21.1**	5.2	10	19.9	49.5
Interquartile range	**198**	n.m.	n.m.	n.m.	n.m.	**9**	n.m.	n.m.	n.m.	n.m.	**14**	n.m.	n.m.	n.m.	n.m.
Panel D: Optimal ratio															
Mean of yearly medians	**0.023**	0.037	0.023	0.018	0.017	**0.232**	0.433	0.313	0.169	0.061	**0.145**	0.358	0.177	0.084	0.037
Median	**0.022**	0.036	0.022	0.017	0.016	**0.227**	0.429	0.308	0.162	0.056	**0.143**	0.357	0.174	0.082	0.035
Mean of yearly means	**0.160**	0.171	0.151	0.152	0.164	**0.359**	0.496	0.409	0.314	0.216	**0.296**	0.443	0.313	0.238	0.191
Mean	**0.160**	0.171	0.152	0.152	0.164	**0.359**	0.496	0.409	0.314	0.216	**0.296**	0.442	0.312	0.237	0.190
Interquartile range	**0.121**	n.m.	n.m.	n.m.	n.m.	**0.540**	n.m.	n.m.	n.m.	n.m.	**0.461**	n.m.	n.m.	n.m.	n.m.

Tabel 4.7: Non-linear multivariate peer ranking method, idiosyncratic optimal ratio

The table displays prediction errors for various industry classifiers: Market (considering the entire market as a single industry), GICS, SIC, D2V (both full and primary industry classifications), and TNIC, utilising the optimal ratios from the optimal table for each industry classifier and quartile, identifying a unique number of potential peers for each investigated company. Results are provided for the entire sample as well as segmented by quartiles, based on the number of potential peers available. The dataset analyses 11,443 company-years from the period 2010 to 2021, drawing peers a broader pool of 21,901 potential company-year with valid valuation and variable data. Panel A display the descriptive statistics, Observations analysed, Maximum potential peers, Mean of yearly medians of optimal ratio (the same as the optimal table for each quartile) and Mean of yearly medians peer count. Panel B shows the prediction error in terms of Absolute Percentage Error (APRE) across different metrics: Mean of yearly medians, Median, Mean of yearly means, Overall Mean, and Interquartile range. In the "All" column, numbers in parentheses indicate the rank of each classifier's performance relative to the others. Panel C and D provides the statistical test on pairwise differences in central tendencies using the Wilcoxon signed-rank test for the full sample (i.e. not the quartiles). Panel C looks at the prediction error between the peer ranking method using six peers, and the idiosyncratic test, while Panel D compare the prediction error of the industry classifiers in the idiosyncratic test. Where ***denotes significance at the 1% level, **denotes significance at the 5% level, *denotes significance at the 10% level.

	Market	GICS					SIC				
	All	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics											
Observations	**11,443**	**11,443**	2,861	2,860	2,859	2,863	**11,443**	2,865	2,859	2,856	2,863
Max potential peer	**2,188**	**250**	26	68	240	250	**233**	18	71	188	233
Mean of yearly medians of optimal ratio	**0.05**	**0.136**	0.289	0.150	0.101	0.048	**0.180**	0.401	0.207	0.068	0.044
Mean of yearly median peer count	**10.5**	**7.0**	5.1	6.8	7.9	11.5	**7.0**	4.4	6.5	9.5	8.8
Panel B: APRE per industry classifier											
Mean of yearly medians	**0.285 (6)**	**0.231 (1)**	0.287	0.246	0.268	0.156	**0.242 (3)**	0.247	0.267	0.324	0.164
Median	**0.283 (6)**	**0.230 (1)**	0.286	0.244	0.272	0.151	**0.242 (4)**	0.250	0.265	0.322	0.152
Mean of yearly means	**0.390 (6)**	**0.335 (2)**	0.414	0.337	0.355	0.233	**0.340 (3)**	0.346	0.355	0.418	0.242
Mean	**0.390 (6)**	**0.334 (2)**	0.413	0.336	0.355	0.233	**0.340 (3)**	0.346	0.354	0.418	0.240
Interquartile range	**0.407 (6)**	**0.350 (2)**	n.m	n.m	n.m	n.m	**0.356 (3)**	n.m	n.m	n.m	n.m
Panel C: Wilcoxon test - fixed and idiosyncratic											
SARD + Industry with fixed peer group size	+***	+***					+***				
Panel D: Wilcoxon test - industry classifier											
Market		+***					+***				
GICS	-***						-				
SIC	-***	+									
D2V Full	-***	+					+				
D2V Primary	-***	+					-				
TNIC	-***	-*					+				

	D2V Full					D2V Primary					TNIC				
	All	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4	**All**	Q1	Q2	Q3	Q4
Panel A: Descriptive statistics															
Observations	**11,443**	2,860	2,863	2,857	2,863	**11,443**	2,864	2,858	2,858	2,863	**11,443**	2,858	2,859	2,856	2,860
Max potential peer	**1,581**	350	592	715	1,581	**243**	12	25	79	243	**363**	21	55	240	363
Mean of yearly medians of optimal ratio	**0.024**	0.037	0.023	0.018	0.017	**0.272**	0.433	0.313	0.169	0.061	**0.154**	0.406	0.225	0.103	0.057
Mean of yearly median peer count	**12.0**	9.4	11.2	12.1	24.5	**5.0**	4.0	4.9	6.8	13.3	**9.0**	5.5	7.9	9.3	15.3
Panel B: APRE per industry classifier															
Mean of yearly medians	**0.266 (5)**	0.301	0.365	0.245	0.195	**0.244 (4)**	0.283	0.306	0.261	0.162	**0.235 (2)**	0.273	0.281	0.270	0.154
Median	**0.267 (5)**	0.301	0.365	0.245	0.191	**0.241 (3)**	0.276	0.305	0.257	0.156	**0.234 (2)**	0.270	0.276	0.267	0.151
Mean of yearly means	**0.350 (5)**	0.386	0.418	0.320	0.276	**0.348 (4)**	0.389	0.397	0.357	0.248	**0.330 (1)**	0.365	0.372	0.362	0.223
Mean	**0.350 (5)**	0.385	0.418	0.319	0.276	**0.347 (4)**	0.388	0.396	0.356	0.248	**0.330 (1)**	0.364	0.371	0.362	0.222
Interquartile range	**0.388 (5)**	n.m.	n.m.	n.m.	n.m.	**0.362 (4)**	n.m.	n.m.	n.m.	n.m.	**0.344 (1)**	n.m.	n.m.	n.m.	n.m.
Panel C: Wilcoxon test - fixed and idiosyncratic															
SARD + Industry with fixed peer group size	+***					+***					+***				
Panel D: Wilcoxon test - industry classifier															
Market	+***					+***					+***				
GICS	-					-					-				
SIC	-					+					+***				
D2V Full						+**					+***				
D2V Primary	-**										+***				
TNIC	-***					-***									

4.4. Industry classifier robustness over time

This section presents results that evaluate the robustness of findings related to the relative accuracy of industry classifiers. The idiosyncratic peer group size approach has been applied consistently throughout, as it was the most accurate approach across all scenarios. Accuracy is measured in terms of mean APRE, as this metric showed the greatest improvement over the fixed peer group size approach.

As outlined in Section 4.2., the TNIC industry classifier emerged as the most accurate across all tests, a conclusion further supported by Figure 1 and 2. These figures demonstrate that TNIC consistently outperformed other classifiers over time. In contrast, the Market classifier showed the highest prediction errors throughout the period, reaffirming earlier trends observed in prior sections. Notably, Figure 2 revealed that the relative differences between classifiers are less pronounced when using the SARD peer ranking method, meaning the gaps in accuracy between classifiers were narrower compared to the results in Figure 1.

A notable observation in Figures 1 and 2 is the spike in prediction errors during the years 2020 and 2021, coinciding with the financial turmoil caused by the COVID-19 pandemic. Although this phenomenon is outside the scope of this study and cannot be fully explained with the available data, it is interesting that prediction errors peaked across all classifiers during this period.

Figure 4.1: Non-linear univariate using idiosyncratic peer group size approach

Present the mean APRE for industry classifiers when using idiosyncratic peer group size approach and the peer ranking method by Cheng & McNamara (2000).

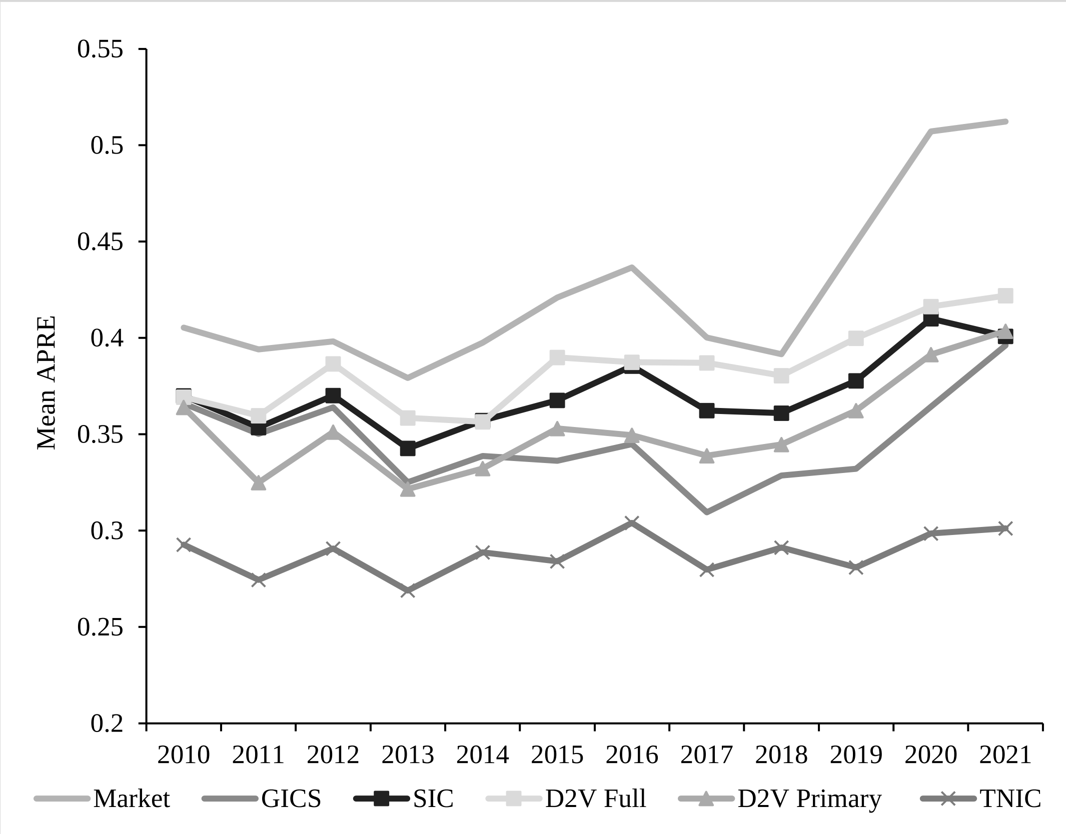
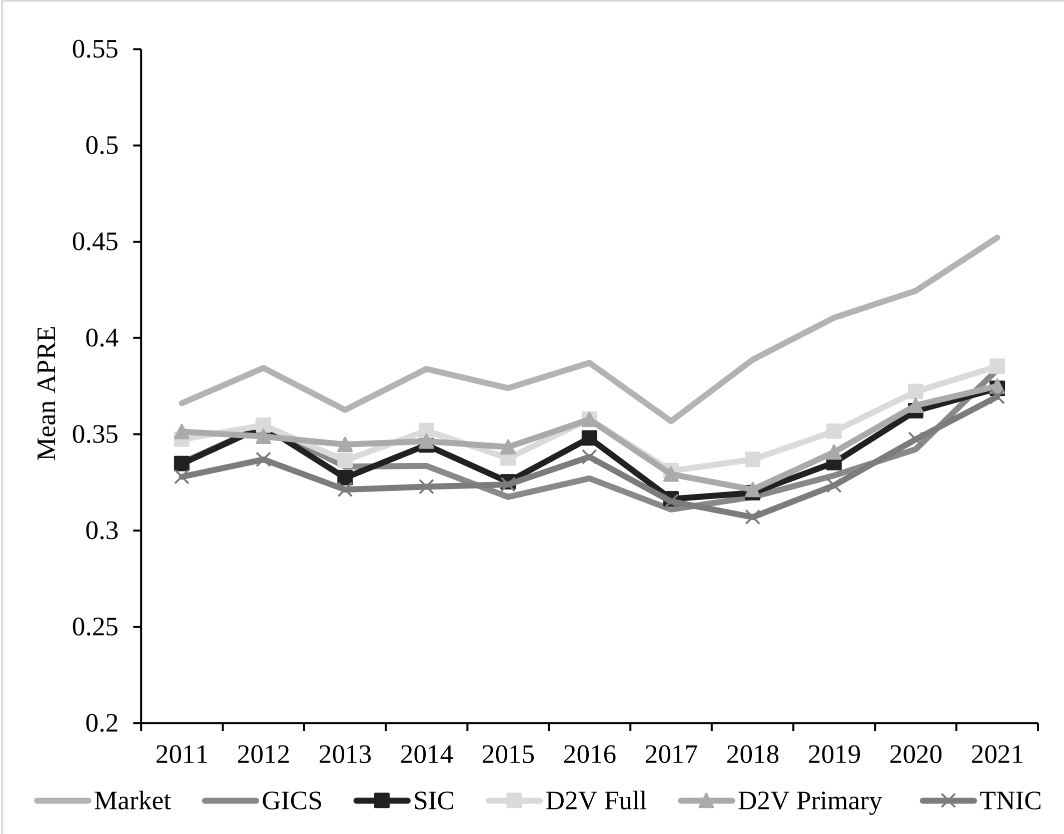


Figure 4.2: Non-linear multivariate using idiosyncratic peer group size approach

Present the mean APRE for industry classifiers when using idiosyncratic peer group size approach and the peer ranking method by Knudsen et al. (2017).



5. Conclusion and practical implications

By applying non-linear peer ranking methods to U.S. companies, we investigated the relative accuracy of conventional and alternative industry classifiers for peer selection used for relative valuation. Additionally, we challenged the prevailing criterion in prior studies of using a fixed peer group size by introducing a novel idiosyncratic peer group size approach. This approach contributes to the existing literature by accommodating differences in both industry classifier structure and industry size. We argue that while systematic peer selection can enhance the accuracy of relative valuation, it must better account for the unique context of the focus firm – a perspective that has been underexplored in previous research.

Relative accuracy of industry classifiers

Our findings on the most accurate industry classifier are not entirely conclusive. However, the alternative classifier TNIC consistently outperformed the others in most tests, typically followed by either the conventional GICS and SIC or the alternative D2V Primary classifier. Hence, we do not find definitive support for the idea that GICS is more suitable for peer selection than SIC. Contrasting the findings by Bhojraj et al. (2003), who argued that GICS is a more effective industry classification system for capital market research. Their study did not specifically address peer selection for relative valuation, which can explain the difference in conclusions. Additionally, D2V does not appear to be significantly more suitable for peer selection, contrary to the findings of Hoberg & Phillips (2023), which highlighted its superiority in predicting profitability compared to SIC and TNIC.

Consistent with previous findings (Alford, 1992; Cheng & McNamara, 2000), the Market classifier consistently exhibited the lowest accuracy. This result highlights the importance of balancing financial and business profile characteristics to improve valuation accuracy. This notion is further supported by the relatively higher accuracy of D2V Primary compared to D2V Full, as D2V Primary applies stricter industry delimitation, while D2V Full finds peers with greater financial similarity.

The strong performance of TNIC, particularly when using the non-linear univariate peer ranking method by Cheng & McNamara (2000), indicates that dynamic product description classifiers can significantly improve relative valuation accuracy. Given the novelty and limited development of alternative classifiers, there is considerable potential for further improvement in this area. Future research could explore utilising the similarity scores, in D2V and TNIC, as weights in peer selection. However, we were unable to find any suitable implementation of this in the peer ranking methods we utilised.

From a practical perspective, these findings advocate for more flexible approaches to company grouping. TNIC's adaptable structure validates the use of dynamic classifiers over rigid, predefined classifiers like GICS and SIC. Although Hoberg & Phillips' systems may be unsuitable for direct practical applications due to their academic focus and constrained data scope, they highlight the potential of language-driven methods. For instance, proprietary solutions such as AI-based grouping using scraped website data or product descriptions could offer a more versatile and practical alternative to traditional classification systems when used in practice.

Evaluation of peer group size approaches

The more experimental aspect of our study – exploring the idiosyncratic nature of peer group sizes – produced the most impactful results. Our idiosyncratic approach improved accuracy by 20–30% compared to the fixed peer group size approach, with the improvement being statistically significant. The size of the industry to which the focus firm belongs appears to be a crucial factor to consider, while the industry classifier structure appeared less critical, as there was minimal variation in classifier rankings across different peer group size approaches.

Our findings highlight a critical gap in prior research, which often relied on arbitrary assumptions about appropriate peer group size, as noted by Dittmann & Weiner (2005). However, we believe further refinement of this approach could yield even greater insights and improvements. For instance, future research could explore what the underlying reasons are for industry size being a strong determinant of the optimal peer group size. As inspiration, Cheng & McNamara (2000) argued the industry size could be a proxy for qualitative factors, such as industry maturity. Moreover, the robustness of these findings should be tested across different time periods and samples to ensure their broader applicability and reliability.

Finally, the proposed method offers potential for practical application. If the heuristic measures used in the idiosyncratic peer group size approach prove robust over time and across different samples, generating the idiosyncratic peer group size is entirely feasible in a practical setting. Nonetheless, even if these heuristics are not completely robust, our findings demonstrate that peer selection ought not to be completely a rigid science but rather be more adaptable. The idiosyncratic peer group size approach underscores the importance of accommodating flexible criteria, such as variable peer group sizes, instead of relying on fixed, predetermined rules.

6. Literature

6.1. Published articles

Adebäck, H., Haqués, D. (2022). *Systematic Peer Selection: A Comparative Study of Methods Used in Relative Valuation*. Master's Thesis. Stockholm School of Economics.

Akin, I., Akin, M. (2024). *Behavioral finance impacts on US stock market volatility: an analysis of market anomalies*. Behavioural Public Policy: 1-25.

Alford, A.W. (1992). *The Effect of Set of Comparable Firms on the Accuracy of Price Earnings*. Journal of Accounting Research, 30: 94-108.

Baker, M., Ruback, R.S. (1999). *Estimating Industry Multiples*. Harvard Working Paper.

Bhojraj, S., Lee, C.M. (2002). *Why is My Peer? A Valuation-Based Approach to the Selection of Comparable Firms*. Journal of Accounting Research, 40 (2): 407-439.

Bhojraj, S., Lee, C.M., Oler, D.K. (2003). *What's My Line? A comparison of Industry Classification Schemes for Capital Market Research*. Journal of Accounting Research, 41(5): 745-774.

Boatsman, J.R., Baskin, E.F. (1981). *Asset Valuation with Incomplete Markets*. The Accounting Review, 56 (1): 38-53.

Bradshaw, M.T. (2001). T*he Use of Target Prices to Justify Sell-Side Analysts' Stock Recommendations*. Working Paper. Harvard University.

Cheng, C.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.

Dittmann, C., Weiner, I. (2005). *Selecting Comparables for the Valuation of European Firms*. Working Paper. Humboldt University Berlin.

Eaton, G.W., Guo, F., Liu, T., Officer, M.S. (2021). *Peer Selection and Valuation in Mergers and Acquisitions.* Journal of Financial Economics, 146: 230-255.

Fama, E.F. (1991). *Efficient Capital Markets: II*. Journal of Finance, 46 (5): 1,575-1,617.

Fama, E.F., French, K.R. (1992). *The Cross-Section of Expected Stock Returns*. Journal of Finance, 47(2): 427-465.

Forte, G., Gianfrate, G., Rossie, E. (2020). *Does relative valuation work for banks?*. Global Finance Journal, Vol 44.

Frésard, L., Hoberg, G., Phillips, G. (2019). *Innovation Activities and Integration through Vertical Acquisitions*. The Review of Financial Studies, 33 (7): 2,937-2,976.

Hoberg, G., Phillips, G. (2016). *Text-based Network Industries and Endogenous Product Differentiation.* Journal of Political Economy, 124 (5): 1,423-1,465.

Hoberg, G., Phillips, G. (2023). *Scope, Scale and Concentration: The 21st Century Firm.* Forthcoming at The Journal of Finance.

Hoberg, G., Phillips, G., Prabhala, N. (2014). *Product Market Threats, Payouts and Financial Flexibility*. The Journal of Finance, 69 (1): 293-324.

Kennedy, D.W., Cameron, J., Wu, P P-Y., Mengersen, K., Cordeiro, A.R. (2021). *Peer groups for organisational learning: Clustering with practical constraints*. PloS one, 16 (6).

Kim, M., Ritter, J.R. (1999). *Valuing IPOs*. Journal of Financial Economics, 53: 409-437.

Knudsen, J.O., Kold, S., Plenborg, T. (2017). *Stick to the Fundamentals and Discover Your Peers*. Financial Analysts Journal, 73 (3): 85-105.

Lee, 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.

Lie, E., Lie, H.J. (2002). *Multiples Used to Estimate Corporate Value*. Financial Analysts Journal, 58 (2).

Liu, J., Nissim, D., Thomas, J. (2002). *Equity Valuation Using Multiples*. Journal of Accounting Research, 40 (1): 135-172.

Malkiel, B.G. (2003). *The Efficient Market Hypothesis and Its Critics*. Journal of Economic Perspectives, 17 (1): 59-82.

Miller, M., Modigliani, F. (1961). *Dividend Policy, Growth, and the Valuation of Shares.* The Journal of Business, 34: 411.

Pandey, P., Bandhu, K.C. (2022). *A credit risk assessment on borrowers classification using optimized decision tree and KNN with bayesian optimization*. International Journal of Information Technology, 14 (7): 3679-3689.

Pinto, J.E., Robinson, T.R., Stowe, J.D. (2018). *Equity Valuation: A survey of professional practice*. Review of Financial Economics, 19 (2): 219-233.

Plenborg, T., Pimentel, R.C. (2016). *Best Practices in Applying Multiples for Valuation Purposes.* The Journal of Private Equity, 19 (3): 55-64.

Roosenboom, P. (2007). *How Do Underwriters Value Public Offerings? An Empirical Analysis of the French IPO Market*. Contemporary Accounting Research, 24 (4): 1,217-1,243.

Skogsvik, K., Skogsvik, S. (2009). *P/E-ratios in relative valuation: a mission impossible?* Investment Management and Financial Innovations, 5 (4): 237-248.

Sloan, R.G. (2002). *Discussion of Who Is My Peer? A Valuation-Based Approach to the Selection of Comparable Firms.* Journal of Accounting Research, 40 (2): 441-444.

Steffen, B.J. (1992). *The Evolution of the Discount for Lack of Marketability*. American Bankruptcy Institute journal, 42 (3): 14-60.

Theising, E., Wied, D., Ziggel, D. (2023). *Reference class selection in similarity-based forecasting of corporate sales growth.* Journal of forecasting, 42 (5): 1,069-1,085.

Zhang, X. (2024). *The Impact of Macroeconomic Factors on Stock Valuation*. Highlights in Business Economics and Management, 41: 513-519.

6.2. Books

Berk, J., DeMarzo, P. (2017). *Corporate Finance, Global Edition*. United States: Pearson Education Limited

Damodaran, A. (2002). *Investment Valuation*. United States: John Wiley Sons

Koller, T., Goedhart, M., Wessels D. (2020). *Valuation: Measuring and Managing the Value of Companies*, 7th edition. United States: John Wiley & Sons

7. Appendix

Appendix 1: Microsoft example for SIC and GICS

Appendix 1: Conventional classifiers for Microsoft

Presents the conventional industry classifiers, SIC and GICS, as applied to Microsoft in this study. The "Industry Code" and "Industry Name" columns display Microsoft's classification within each system, with the hierarchy levels separated by a "-". Both classifiers feature four hierarchical levels, but SIC advances by one digit per level, whereas GICS progresses by two digits.

Classifier	Industry code	Industry name
SIC	7-3-7-2	Services - Business Services - Computer Programming, Data Processing, And Computer Related Services - Prepackaged Software
GICS	45-10-30-20	Information Technology - Software & Services - Software - Systems Software

Appendix 2: Microsoft example for TNIC

Appendix 2: TNIC classifier for Microsoft

Presents the TNIC classifier for Microsoft during the years studied, 2010 to 2021. It includes the number of peers identified each year and lists the top five most similar peers based on product similarity.

Year	Number of peers	1th closest peer	2th closest peer	3th closest peer	4th closest peer	5th closest peer
2010	96	Adobe	Red Hat	Citrix Systems	Oracle	McAfee
2011	122	Citrix Systems	Red Hat	Oracle	Adobe	Versant Corp
2012	106	Citrix Systems	Red Hat	Adobe	Polycom	Monotype Imaging
2013	99	Red Hat	Citrix Systems	Adobe	Polycom	Insight Enterprises
2014	106	Citrix Systems	Red Hat	Adobe	CDW Corporations	Smith Micro Software
2015	190	Citrix Systems	Jive Software	Polycom	Monotype Imaging	Red Hat
2016	150	Red Hat	Jive Software	Adobe	BroadSoft	Citrix Systems
2017	169	Oracle	Citrix Systems	Adobe	Splunk	Monotype Imaging
2018	150	Talend	Citrix Systems	Cisco Systems	Nutanix	Oracle
2019	142	Nutanix	Cisco Systems	Citrix Systems	Oracle	Insight Enterprises
2020	125	Adobe	Cisco Systems	VMware	Oracle	Splunk
2021	144	VMware	Adobe	Cisco Systems	Oracle	Nutanix

Appendix 3: Microsoft example for D2V

Table 7.3: D2V classifier for Microsoft

Presents the D2V classifier for Microsoft during the years studied, 2010 to 2021. It includes the number of industries allocated to Microsoft, the primary industry code (representing the industry with the highest similarity based on product descriptions), and the ten words that best describe the primary industry.

Year	Number of industries allocated to	Primary industry code	Ten word vector for primary industry
2010	21	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2011	15	294	billboards, billboard, oaaa, outdoor, posters, beautification, lithographed, highways, transit, subways
2012	16	294	billboards, billboard, oaaa, outdoor, posters, beautification, lithographed, highways, transit, subways
2013	16	294	billboards, billboard, oaaa, outdoor, posters, beautification, lithographed, highways, transit, subways
2014	15	294	billboards, billboard, oaaa, outdoor, posters, beautification, lithographed, highways, transit, subways
2015	19	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2016	20	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2017	21	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2018	21	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2019	22	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2020	23	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes
2021	25	106	microcomputer, resellers, microcomputers, peripherals, reseller, networking, multivendor, gtsi, merisel, hayes

Appendix 4: Data treatment

Treatment	Cheng & McNamara (2000)	Knudsen et al. (2017)
Intersection of company-years for all industry classifiers, between 2010-2021	47,969	47,969
Less missing or faulty valuation data from I/B/E/S and CRSP database	(20,025)	(20,025)
Less winsorisation 5% on each tail	(2,794)	(2,794)
Total sample of company-years with valid and representable valuations	**25,150**	**25,150**
Less missing data for independent variables	(4,767)	(3,249)
Observations with valid valuations and independent variables	**20,383**	**21,901**
Less company-years with less than six valid peers in all classifiers	(8,801)	(10,458)
Sample company-years with valid data and adequate peer count	**11,582**	**11,443**

Appendix 5: Usage of AI in thesis writing

During the writing of this thesis, we utilised ChatGPT 4o as a supportive tool. Its primary use was in generating and refining Python code for data cleaning, adjustment, and testing. Specifically, ChatGPT was instrumental in originating initial code outlines and iteratively assisting in resolving minor coding issues and ensuring consistency across tests. This significantly reduced the time required for development of the code, while also enhancing code quality by serving as a brainstorming tool for addressing problems in the code. However, there were instances where ChatGPT produced incorrect or non-functional code. To mitigate this, we rigorously double-checked all outputs and validated results manually. For example, when calculating market capitalisation, we manually verified the results for over twenty companies to ensure accuracy.

Additionally, ChatGPT was used for spelling and grammar checks, contributing to the overall quality of the text. However, we avoided using its direct phrasing due to its tendency toward imprecision, verbosity, and occasional factual inaccuracies. Instead, it was used to suggest phrasing alternatives and synonyms, which were then carefully reviewed and adapted to fit the precise tone and clarity required for the thesis.

For advanced concepts, such as the Manhattan distance, ChatGPT was used alongside other sources to provide explanations and clarify the concept for the authors. This approach was similar to consulting resources like Wikipedia or academic references for initial understanding.

In summary, ChatGPT played a supportive role in improving both the coding and writing processes of the thesis. Its usage was carefully managed to limit potential pitfalls, such as factual hallucinations or overly generalised responses, by keeping prompts.