(OUT)SMART THE PEER GROUP IN MARKET COMPARISON: BUILDING BUSINESS VALUATION MULTIPLES BY MACHINE LEARNING

Veronika Staňková[1]

[1] *Prague University of Economics and Business, Czech Republic*

EUROPEAN JOURNAL OF BUSINESS SCIENCE AND TECHNOLOGY

Volume 10 Issue 2
ISSN 2694-7161
www.ejobsat.com

ABSTRACT

Traditionally, market comparison requires identifying a peer group, which still poses unresolved practical difficulties today. This research seeks to provide valuable insights into the practicality, efficiency, and accuracy of machine learning in valuing a company. It employs a state-of-the-art machine learning technique, Gradient Boosting Decision Trees (GBDT), to predict the valuation multiple directly. A yearly dataset of U.S. public companies from 1980–2021 was used. The most common multiples (EV/EBITDA, EV/EBIT, P/E, and EV/Sales) were tested. The performance of GBDT was assessed against an industry-based method. GBDT consistently outperformed the alternative method with an average 24 percentage point decrease in the median average percentage error. The results support GBDT's potential as a supplementary tool in valuation practice.

KEY WORDS

market comparison method, Gradient Boosting Decision Trees, industry multiple, feature importance

JEL CODES

G12, G32

1 INTRODUCTION

The corporate value is the current net value of future payoffs, such as dividends or free cash flows. There are two principal strategies to approach the uncertainty of estimating a company's future payoffs. In a direct valuation, an appraiser estimates the future payoffs explicitly and in absolute terms, as opposed to a relative valuation, where the payoffs are derived from comparable firms in the capital market. In practice, the market comparison method is widely utilised due to its straightforward applicability, interpretability, and speed by which a valuation can be completed (Plenborg and Pimentel, 2016). A global survey performed by Pinto et al. (2019) among equity analysts demonstrated that most respondents (93%) use market comparison during the valuation process.

STAŇKOVÁ, Veronika. 2024. (Out)smart the Peer Group in Market Comparison: Building Business Valuation Multiples by Machine Learning. *European Journal of Business Science and Technology*, 10 (2): 156–172. ISSN 2694-7161, DOI 10.11118/ejobsat.2024.011.

The basic concept of the market comparison method comes down to the law of one price (Knudsen et al., 2017). In an efficient market, comparable firms (substitutes on the financial market) are expected to be priced similarly. Typically, market comparison involves four steps. First, a peer group with known firm values is selected. Second, a measure driving the value of the subject firm is identified (typically an accounting measure such as earnings, book value, or revenues). Third, a valuation multiple is built using the peer group from step one, with the firm value as the numerator and the value driver as the denominator. Finally, the valuation multiple is applied to the subject company's value driver to determine its firm value. This paper specifically focuses on how the benchmark is selected (steps one and three of the market comparison method), which is generally regarded as the critical aspect of the implementation (Plenborg and Pimentel, 2016).

Machine learning applications have gained prominence in various areas, including corporate finance (Sellhorn, 2020). Traditional methods often fail in performance compared to innovative machine learning applications (Ding et al., 2019; Alanis, 2022; Geertsema and Lu, 2023). The reason lies in the complexity, multicollinearity, and non-linearity of the financial relationships, making it challenging to capture the interactions using classical statistical models. On the contrary, the ability to handle such data is the advantage of machine learning. It benefits from a growing amount of data serving as input to models and advancements in computational techniques that facilitate the rapid development of machine learning and artificial intelligence models.

In this context, this study aims to explore the potential of state-of-the-art machine learning in enhancing market comparison valuations from the practitioners' point of view. For this objective, it applies a gradient boosting decision tree model to predict the most common valuation multiples (price-to-earnings, enterprise-value-to-EBIT, enterprise-value-to-EBITDA and enterprise-value-to-sales), broadly following the method by Geertsema and Lu (2023), who, however, primarily predicted a different set of valuation multiples, among other methodological differences. To investigate the superiority of machine learning prediction, the outcome is compared and interpreted in relation to an alternative model represented by the industry mean multiple in terms of the accuracy measured by error rates, R-squared and correlation, and the importance of input variables. The structure of this article is as follows. Section 2 presents an overview of the prior literature on the market comparison method and a brief theoretical introduction to the machine learning technique. Section 3 details the proposed research design and data used for the analysis. Section 4 summarises and discusses the results. Finally, Section 5 concludes with the study's key findings and discusses the contribution of this research compared to Geertsema and Lu (2023).

2 THEORETICAL FRAMEWORK

2.1 Market Comparison Method

In theory, it is generally accepted that the subject company and its peer group are comparable when their value drivers, especially profitability, risk, and growth rate, are alike (Damodaran, 2002). However, there is no consensus on how to ensure this in practice. Plenborg and Pimentel (2016) differentiated three approaches to peer group selection: (i) industry-based, (ii) fundamental, and (iii) innovative approaches.

The prevailing practice is that the peer group is often selected solely based on affiliation to the same industry. Previous academic research has accepted the industry-based approach to selecting comparable companies (Alford, 1992; Cheng and McNamara, 2000; Liu et al., 2002; and in terms of the presented performance of the alternative models also Geertsema and Lu, 2023) advocating that industry affiliation means sharing similar market characteristics and thus the expected future performance.

Representatives of the fundamental approach argue that sharing similar market characteristics does not stipulate comparability of economic fundamentals (Bhojraj and Lee, 2002; An et al., 2010). They point out limitations of the industry classification, such as the lack of a universal classification system. Bhojraj et al. (2003) conducted a study comparing four different established industry classification systems: SIC (Standard Industrial Classification), NAICS (North American Industry Classification System), GICS (Global Industry Classification Standard), and the Fama and French classification. Based on their research, the GICS classification provided the most accurate explanation of the cross-sectional variability in valuation multiples and other financial indicators (Bhojraj et al., 2003). Additionally, they highlight that the GICS matches the different classification systems only 56% of the time, increasing the importance of choosing the industry classification system. Another reason for the potential inaccuracy of traditional industry classification systems is that new business areas are created during the dynamic development of the economy (Hoberg and Phillips, 2010, 2016).

On the other hand, a handpicked peer group based on shared fundamentals increases the risks of potential bias in peer selection, as demonstrated by the research of De Franco et al. (2015). Therefore, researchers attempted to develop a systematic technique to mitigate the subjective judgment needed for the fundamental approach – such as the warranted multiple by Bhojraj and Lee (2002); and An et al. (2010) or the sum of rank differences by Knudsen et al. (2017). These techniques are primarily based on statistical methods, which usually require the input data to have specific statistical properties (such as homoscedasticity), which financial data often do not fulfil. Moreover, with an increasing number of (independently distributed) fundamentals considered for selection, the size of the intersection of the most comparable firms in all fundamentals decreases rapidly (Alford, 1992).

The third approach is a mix of innovative techniques, such as using the co-searches on the EDGAR website by the U.S. Securities and Exchange Commission by Lee et al. (2015) or the similarity of business descriptions in annual reports analysed by Hoberg and Phillips (2010, 2016) with the use of machine learning. Generally, these approaches profit from big data availability and attempt to offer an objective peer group selection process.

2.2 Machine Learning Approach

Machine learning (ML) encompasses mathematical algorithms at the intersection of artificial intelligence and statistical models. ML detects patterns in structured and unstructured input data without being a priori given hypotheses about these mutual relationships (thus, a non-parametric model). Generally, these algorithms are used for clustering and dimensionality reduction on one hand (unsupervised ML) and for regression and classification tasks on the other hand (supervised ML). Predicting the valuation multiple is a regression task since the aim is to predict a continuous variable (target). Machine learning algorithms are data-driven models. Thus, they are distribution-free and do not require imposing any statistical properties on the input data, which is a significant benefit over any fundamental approach to peer group selection using statistical tools while maintaining objectivity. There is a growing body of machine learning applications in finance (Sellhorn, 2020), with many demonstrating excellent results.

In the market comparison method, emerging machine learning applications can achieve the needed trade-off between objectivity and flexibility due to their non-parametric nature. Our research references the work by Ding et al. (2019), Alanis (2022), and Geertsema and Lu (2023). Ding et al. (2019) tested ML on peer group selection to detect financial anomalies, and Alanis (2022) used it for regression of the beta factor, bypassing peer group selection. Both studies demonstrate the superiority of the machine learning approach to the industry-based approach. From a current point of view, the particular machine learning method they used – supported vector machine and random forest in Ding et al. (2019) and Alanis (2022), respectively – are somewhat overshadowed by more recent methods, particularly the Gradient boosting decision tree (GBDT) used by Geertsema and Lu (2023) to predict the valuation multiple for market comparison valuation and peer group identification.

GBDT is an ensemble model of decision trees widely used in many machine learning tasks. An ensemble model refers to a technique which builds and combines numerous simple models, sometimes called weak-learners (James et al., 2023), to achieve superior prediction accuracy compared to a single complex model. In the case of GBDT, the weak-learner models are single decision trees, and the particular ensemble technique is gradient boosting.

A decision tree is a hierarchical model that employs recursive binary splitting, well-suited to capturing complex and non-linear relationships within the data (James et al., 2023). At each decision node, where the tree branches divide, the target data are divided into two subsets to maximise homogeneity within each subset. This division is based on a selected input variable and its threshold value, leading to the highest homogeneity at each split.

Gradient boosting means that for each iteration, a new weak-learner (a single decision tree) is sequentially added to the existing ensemble of trees in a way that minimises the residual error (also called the negative gradient, hence the name of the model) of the previous iterations of the ensemble (James et al., 2023). In doing so, GBDT uses information from the previously grown trees and gradually learns the patterns in the input data with a lower tendency to learn the 'noise' among the data provided for learning (risk of overfitting). Each iteration is only responsible for an incremental improvement; that is, not even the last iteration of the weak-learner would represent the whole model.

The key benefits of GBDT are its accurate predictions, high efficiency, and relative ease of implementation. It requires very little to no pre-processing of the input data as it is robust to outliers in explanatory variables (features). Furthermore, it also implicitly handles missing data among the features. Lastly, the underlying tree-based approach makes it relatively easy to comprehend how the model was built. However, compared to single decision trees (the weak-learners), the higher accuracy of GBDT comes at the cost of aggravated interpretability, as it is not possible to summarise the GBDT model into a graphical representation of a decision tree because of the sequential nature of the ensemble. These considerations are important should practitioners adopt machine learning as a fully-fledged method.

2.3 Contribution and Aim

We would like to contribute to the pioneering literature on the ML approach to market comparison valuation, particularly the work of Geertsema and Lu (2023). They developed ML regression models for market-to-book value, enterprise-value-to-assets, and enterprise-value-to-sales valuation multiples. They then compared their performance with five valuation multiple estimation techniques derived from prior literature. They tested two industry-based approaches: (i) the harmonic mean of the SIC industry affiliation from Liu et al. (2002), and (ii) the mean of an alternative industry classification set up by Hoberg and Phillips (2010, 2016) based on similarities of business descriptions. The last three models represented the fundamental approach to market comparison valuation. These included (iii) a statistical regression model to estimate the so called warranted multiple first proposed by Bhojraj and Lee (2002). Finally, the studies by (iv) Bartram and Grinblatt (2018, as cited in Geertsema and Lu, 2023), and (v) Rhodes-Kropf et al. (2005, as cited in Geertsema and Lu, 2023), both estimated the equity value by a cross-sectional (i.e., fundamental) analysis directly, skipping the step of building the market multiple first or even selecting a peer group. The results of Geertsema and Lu (2023) unambiguously support the hypothesis about the superiority of ML valuation accuracy measured by out-of-sample valuation errors, R-squared values, and Pearson correlation coefficients. The ML models systematically generated more accurate valuations over the entire sample period (1980–2019) and across all firm subsamples.

Apart from the findings on ML's superiority, Geertsema and Lu's results can also be interpreted in relation to the peer group selection approach discussion described above. We find it noteworthy that none of the five traditional models can be identified as the best (non-ML) since the alternative models ranked differently in different performance metrics. Generally, the two industry-based models scored lower valuation errors. In contrast, the fundamental approach of Bhojraj and Lee (2002) scored better in correlation. These results lead us to conclude that the industry-based approach is not merely a lack of effort when building the valuation multiple. Instead, the industry-based approach has proven once again that in some contexts, it is well justified to apply it, as suggested by the unresolved academic discussion over the proper peer group selection process.

This study broadly adopts the methods used by Geertsema and Lu (2023) to test the performance of machine learning (ML) on market comparison valuation by utilising the ML algorithm and performance metrics. Nevertheless, we contribute to this line of research by altering some key assumptions relative to Geertsema and Lu (2023) to broaden the applicability of their conclusions and to add a more practitioner-oriented perspective. Specifically, our key modifications include changing the set of valuation multiples being predicted to those more commonly used and assessing the over/under-performance of GBDT using a previously untested benchmark method that also originates from practice. Further details of the comparison between this study and Geertsema and Lu's (2023) study are provided in the Discussion section, following the methodology described in the next section.

3 METHODOLOGY AND DATA

This study aims to investigate whether (and why) a machine learning approach can enhance market comparison valuation. Essentially the ML constructs the valuation multiples from fundamentals provided for learning, thus the proposed method represents the fundamental approach to market comparison valuation.

In line with the general evaluation framework of ML and the research of Geertsema and Lu (2023), we assessed the improvement of GBDT compared to an alternative method that provides the baseline for evaluation. To benchmark the GBDT, we selected a traditionally used industry-based approach to the market comparison method.

The following two subchapters elaborate in more detail on the GBDT model and the alternative model respectively.

3.1 Proposed Method: Gradient Boosting Decision Trees

This study utilises the LightGBM implementation of GBDT (gradient boosting decision trees) developed by Microsoft. GBDT is recognised as a state-of-the-art decision tree-based model, noted for its accuracy and ease of use (reasons are explained in the previous section). We opted for LightGBM to enhance comparability with Geertsema and Lu (2023), who also applied LightGBM.

The proposed method uses 52 fundamental variables (of a financial and nonfinancial nature, including the GICS industry classification) as the explanatory input variables (features). These were selected to be as comprehensive as possible and with consideration of the general availability of the data for most of the sample across all industries. GBDT then internally selects the most important features for the given task. Based on GBDT weighting, we picked a final collection of 25 features during the fine-tuning phase used in the final models and created 6 additional GICS industry-derived features during the modelling. The final decision on the size and selection of variables is made based on several analyses performed during the fine-tuning phase. These analyses include recursive feature elimination cross-validation on the number of features, feature importance, Spearman's correlations between the features, and an expert assessment (since not all combinations could be tested experimentally). We include an overview of variables and targets in the Annex including comments on usage and source and summary statistics (Tab. 4).

Missing observations were only addressed for the target variable by removing the particular firm-year observation. In contrast, we retained observations with missing input variables, as the LightGBM algorithm can handle them automatically. Moreover, However, the conclusions remain unchanged when testing the GBDT, even when all observations with any missing input variable are dropped. Due to extreme values (outliers) of the target variables in the retrieved dataset, the bottom and top 10% of the target variable distribution were trimmed. This process also eliminated any negative value of the target variable, aligning with the method of Liu et al. (2007). While the features were not explicitly treated for outliers or missing values, many potential outliers among the features were unintentionally removed due to the trimming of the targets. We also considered alternative outlier control approaches, as discussed in more detail in the section Model performance with alterations.

All absolute monetary values (among features only) were indexed to the price level of 2021 by the GDP Price Deflator reported by the World Bank for the USA. This adjustment ensures the values are comparable and prevents bias in the cross-sectional analysis due to inflation. Logarithmisation of the target variable was tested but eventually not used as it did not improve accuracy.

No other data manipulations were required, as the tree-based model is robust to outliers (among features) and multicollinearity and does not necessitate any standardisation.

During the fine-tuning of the model, we noticed a strong influence on the number of iteration rounds (num_boost_round hyperparameter) of the GBDT. Empirically, we selected 10,000 iterations. To avoid overfitting, we enabled an early stopping criterion, which stops training if the model does not improve for 10 consecutive rounds. The effect of the num_boost_round hyperparameter is discussed in the section Model performance with alterations.

The loss functions to optimise the model's internal parameters in the training phase included the root-mean-square deviation (RMSE) and the mean absolute percentage error (MAPE). Other hyperparameters of the GBDT were maintained at their default values.

3.2 Alternative Method: Mean GICS Industry Multiple

The alternative method serves as a benchmark for the proposed method. For this purpose, we selected a traditional industry-based approach to the market comparison method using the 6-digit GICS Industry level (GICS Industry Code). We opted for the GICS industry code for several reasons. Firstly, Bhojraj et al. (2003) claimed it to be the most suitable classification system and level for valuation. Therefore, it is, in our view, the best representative of the industry-based approach to the market comparison method. Secondly, we regard it as an addition to the five other alternative methods already tested by Geertsema and Lu (2023), of which two were industry-based approaches. Lastly, unlike Geertsema and Lu (2023), we included the same industrial affiliation (GICS) as a feature in the GBDT model to better highlight the effect of the GBDT. As a result, the GICS industry classification is utilised in both the proposed and alternative methods in this study.

The prediction of the alternative method for a given firm-year observation is then the aggregated industry multiple. It is computed as a mean of the individual companies' multiples affiliated within the same GICS industry in a particular year. We also explored other common function arithmetic and harmonic means, as discussed in more detail in the section on Model performance with alterations.

For the alternative model, the same adjusted dataset is used for the GBDT (including trimming outliers and missing observations among the target variables). Thus, the input for the proposed and alternative methods in terms of subjects, GICS industry affiliation and industry aggregated mean multiples is essentially the same.

3.3 Tested Valuation Multiples

Since the aim of this study is to provide relevant findings for both practitioners and academics, we focused on the high-profile multiples commonly used in appraisals: price to earnings (P/E); enterprise value to earnings before interest and tax (EV/EBIT); enterprise value to earnings before interest, tax, depreciation, and amortisation (EV/EBITDA); and enterprise value to sales (EV/Sales). See Plenborg and Pimentel (2016) for a review of academic discussion on the choice of the multiple.

We trained a separate GBDT model for each selected valuation multiple (target variable) for our analysis. We prioritised historical accounting data over forward-looking estimates to avoid a reduction in the sample size due to forecast data availability.

3.4 Data

Data are sourced from the Thomson Reuters Refinitiv Eikon database. The sample comprised the universe of all public companies headquartered in the USA traded from 1980 to 2021. The precise data retrieval setting is described in the Annex. Approximately 10 thousand unique companies meet the criteria.

To evaluate the proposed and alternative methods, we set aside a random split of 25% of the available observations (the remaining 75% were used to train the GBDT model and compute the industry multiple).

3.5 Performance Metrics

The performance of both the proposed and alternative models was assessed using several performance metrics comparing the actual (observed) target values with their predictions. The prediction of the proposed method results from the GBDT model trained for the given target. The prediction of the alternative model is the mean multiple of the train-set companies affiliated with the same GICS industry. For both methods, all performance evaluations are based on the test dataset (out-of-sample testing stipulated by ML theory; see James et al., 2023).

The accuracy was assessed primarily by the median absolute percentage error (MdAPE) and mean absolute percentage error (MAPE). To identify a systematic over- or underestimation (bias) of the target, the mean and median percentage errors are also reported (referred to as MPE and MdPE, respectively). The percentage error is a relative measure of how close the prediction is to the actual value of the target variable. Furthermore, for a more detailed comparison with the results of Geertsema and Lu (2023), we also present the R-squared (R^2) and Pearson correlation between the actual and predicted target values (ρ).

4 RESULTS

4.1 Model Performance

This study applies GBDT, an ML technique, to market comparison valuation to determine if it can achieve better accuracy of the valuation multiple than the alternative approach relying on the industrial classification (GICS), and to analyse the results in relation to previous studies.

Our assumption is that the ML would exhibit higher accuracy measured by error rates (MAPE, MdAPE, MPE, MdPE) and greater explanatory power measured by R-squared (R^2) and Pearson correlation (ρ).

The results in Tab. 1 indicate that the proposed (ML) model outperforms the alternative method in all metrics across all targeted valuation multiples. The following conclusions can be drawn:

- The errors of ML's market multiples are consistently smaller compared to the alternative model. The ML model's MdAPE (median absolute percentage error) is reduced by

Tab. 1: Performance measured by error rates, R-squared, and Pearson correlation for all models considered

	MdAPE	**MAPE**	**MdPE**	**MPE**	R^2	ρ
Proposed method, %						
EV/EBITDA (ML)	6%	17%	0%	6%	68%	82%
EV/EBIT (ML)	6%	16%	0%	6%	73%	86%
P/E (ML)	7%	18%	0%	7%	67%	82%
EV/Sales (ML)	8%	32%	0%	18%	80%	90%
Alternative method, %						
EV/EBITDA (GICS)	25%	32%	0%	7%	11%	39%
EV/EBIT (GICS)	26%	33%	0%	7%	14%	43%
P/E (GICS)	25%	33%	0%	7%	8%	37%
EV/Sales (GICS)	47%	77%	0%	36%	23%	52%
Improvement, p.p.						
EV/EBITDA	19	16	0	1	57	43
EV/EBIT	20	17	–	1	60	43
P/E	18	16	–	–	59	45
EV/Sales	39	44	–	17	57	37

Note: If the change in performance is a positive value, it is indicated as an improvement in percentage points (p.p.), "–" means no improvement, and "0" means an improvement rounded to 0.

an average of 24 p.p., and MAPE (mean absolute percentage error) is smaller by approximately 23 p.p.

- MdAPE is always smaller than MAPE. The distribution of prediction errors is positively skewed for both methods, indicating that most prediction errors are minor, with a few outlier errors. From a practical appraisal point of view, smaller errors are preferable to larger errors.
- MPE is always above zero. Positive values of MPE for both methods indicate a bias of predictions towards overvaluation. This inclination might be explained by the exclusion of negative targets from the dataset (see the Data and Methods section for reasoning). However, the MdPE remains zero, meaning the predictions' central tendency (the median) is unbiased.
- The superior precision of ML models is also evidenced by consistently higher coefficients of determination and correlation scores.
- Notably, the EV/Sales multiple consistently shows higher prediction errors across all metrics. This suggests that it is not a universally suitable valuation multiple for all industries and stages of a company's life cycle. This finding is consistent with Baker and Ruback (1999).

Fig. 1 analyses the performance of all models over time. For all valuation multiples, the MdAPE of the GICS industry models (grey lines) oscillated around a constant level. In contrast, the errors of the ML predictions (black lines) exhibited a decreasing trend, indicating that GBDT became more precise in recent predictions. This enhancement in GBDT's performance over time is likely attributable to the growing dataset available for prediction after the outlier removal, as illustrated by the clustered columns in Fig. 1. This reasoning is most evident in the high errors in the first years of the model period.

4.2 Feature Importance

Given the sequential nature of the GBDT, it is impossible to represent the model as a single decision tree. Each decision tree within the model serves the purpose of incrementally improving the error generated by the previous iteration runs. Only when combined do they constitute the final prediction of the target. The general solution is to view the importance of

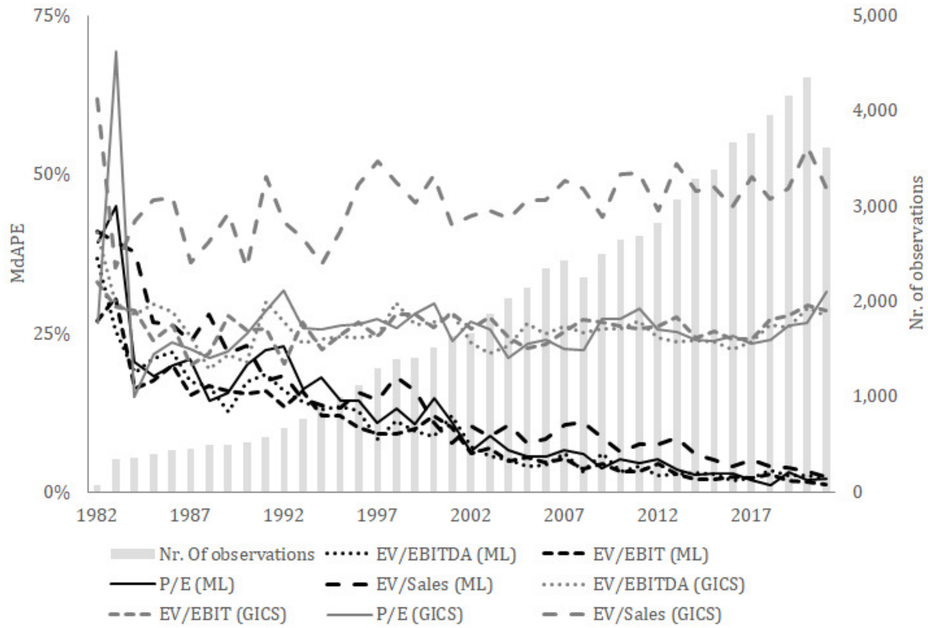

Fig. 1: MdAPE of the proposed method compared to the alternative method (primary axis) over time and number of observations (secondary axis) over time

each variable to illustrate the model's internal assumptions (James et al., 2023).

We present both the gain and split versions of the feature importance as computed by LightGBM during the modelling process. Gain importance measures the improvement in accuracy brought by a feature to the branches it is on, while split importance counts how frequently a feature is used to split data across all trees. We normalised both versions of the feature importance scores by dividing each score by the maximum score in the given target prediction model (separately for split and gain scores). This normalisation facilitates a more intuitive comparison across features. The following Fig. 2 shows both scores for each target (Panels B through E) and the gain score ranking across all scores (Panel A).

Fig. 2 reveals that the most significant feature contributing to increased accuracy is the mean industry multiple computed at the GICS Industry level (the same level used in the alternative industry multiple method). It ranks as the number one feature for all targets except for EV/Sales. Moreover, for EV/EBITDA and EV/EBIT, all other features contribute very little (less than 30%) when compared to the GICS Industry mean gain in importance. There are more significant features beyond the industry mean for the remaining targets. For P/E, an additional strongly contributing feature is the return on equity (ROE). For EV/Sales, the three most contributing features (in terms of gain importance) are Asset turnover, the mean GICS Industry multiple, and EBIT margin. The evidence of a more distributed importance across more key features with the industry mean being ranked only second explains why EV/Sales scored significantly higher errors in the alternative method, as presented in Tab. 1.

However, the low split importance of the mean GICS Industry multiple indicates that the GBDT model does not use it frequently – it actually ranks only in the 17th percentile or 27th out of 31 features in terms of the split importance scores summed across all targets. In contrast, the most frequent feature overall is the ratio of Property, plant, and equipment (PPE) net to PPE gross, indicating accounting obsolescence of the PPE (100th percentile).

We can summarise the findings on the industry multiple as follows: The gain in importance

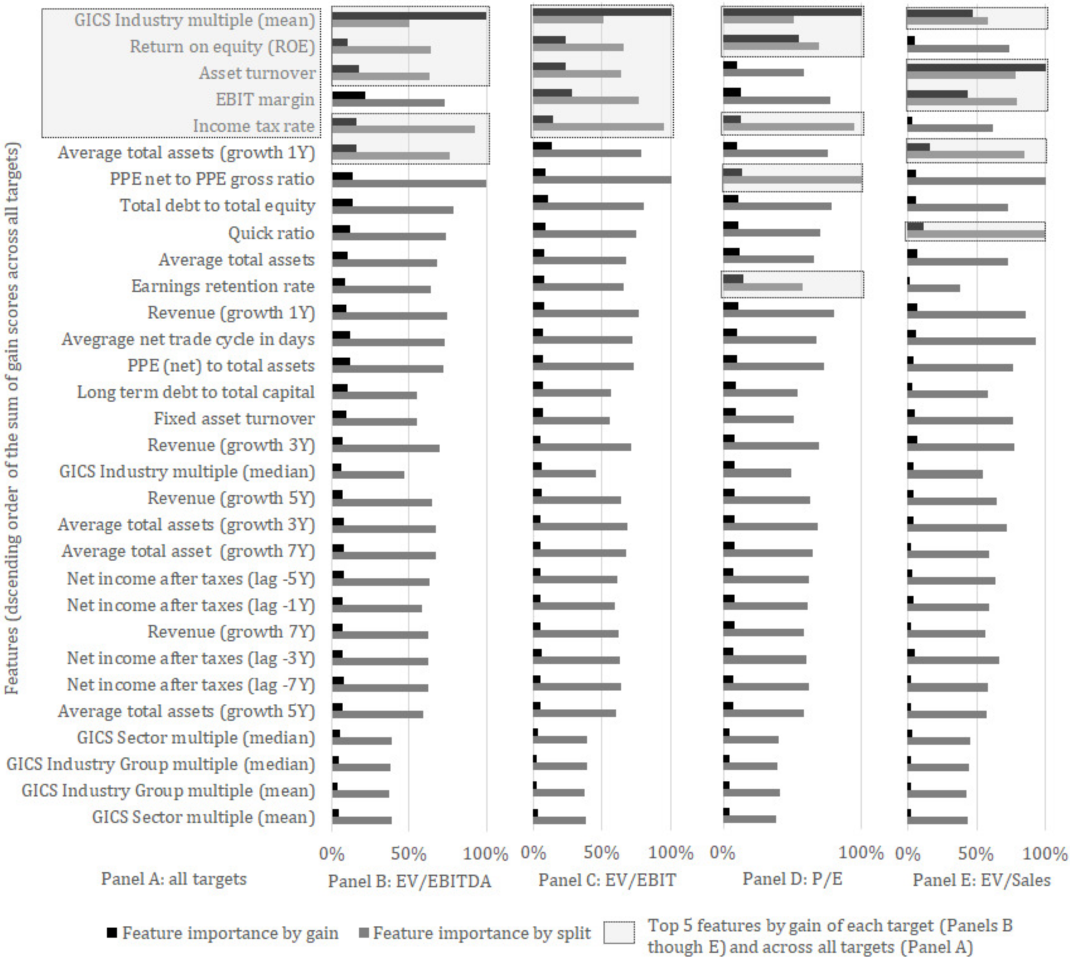

Fig. 2: Feature importance by gain and split for all target variables, expressed as a percentage of the maximum score for the given target. Labels in Panel A are in descending order based on the sum of the absolute gain importance scores across all targets. The top 5 most important features by gain score are highlighted with dotted squares for each target (Panels B through E) and across all targets (Panel A).

suggests that the industry multiple captures a significant portion of the target's variability (except EV/Sales), consistent with the common practice of leveraging the industry affiliation for benchmarking and the research supporting the use of industry multiples discussed above. Furthermore, while the industry multiple is crucial for GBDT to make a prediction, it likely serves as a starting point. It is evidently less utilised during the later iteration runs when the GBDT seeks finer interconnections within the data and refines the prediction. This nuanced data exploration is the added value of GBDT, contributing to its superior accuracy.

Given that the mean GICS Industry multiple is leveraged by the GBDT, all other GICS multiples are ranked as relatively insignificant both in gain and split importance. Any other industry level would merely duplicate the same information as the GICS Industry, which represents the most detailed level allowed in the model. Additionally, we note the GBDT's preference for the mean aggregation. For a discussion of mean versus median aggregation, please refer to the following section, Construction of the industry multiple.

4.3 Model Performance with Alterations

4.3.1 Alternative Outlier Control Methods

The control for outliers is the sole data cleansing procedure we implemented. We analysed the sensitivity of the performance metrics to changes in the definition of an outlier as the bottom and the top 1%, 5%, 10% (the primary configuration in this study), and 20% of the target variable.

As depicted in Panel A of Fig. 3, prediction errors produced by GBDT (black lines) consistently outperform GICS industry multiples when subjected to varying degrees of outlier control. Our findings indicate that the extreme values of the target variable are difficult to predict by both the ML and the GICS industry multiples (grey lines). By implementing stringent outlier controls to ensure data homogeneity the error is reduced substantially. However, when data size is excessively pruned – evident at the 20% data exclusion threshold – industry multiples exhibit a rise in error rate. This implies a critical threshold below which the data set no longer provides sufficient observations to construct a representative industry multiple. On the contrary, GBDT exhibits a lower sensitivity to such (subjective) data sample manipulation.

The pronounced advantage of GBDT is further supported by the Pearson correlation coefficients presented in Panel B of Fig. 3. The disparity between the ML approach and the conventional industry methodology is particularly pronounced at the intermediate data exclusion thresholds (5% and 10%). At the 20% threshold, where data homogeneity is highest, the correlation decreases slightly for GBDT (even as the prediction error continues to improve). In contrast, for industry multiples, the trend is reversed (simultaneous improvement of the correlation coefficient and worsening prediction error). This inverse relationship may indicate an overly homogeneous data set that offers limited learning potential due to its lack of diversity.

4.3.2 Number of Iteration Rounds

The number of boosting iterations represents the number of trees and is an important hyperparameter of the GBDT (num_boost_round). Panels C and D of Fig. 3 compare outcomes for 100 (the default setting for GBDT), 1,000, 10,000 (the primary configuration in this study), and 100,000 iterations. Through empirical analysis, we determined the optimal number of iterations to be 10,000, as it provides an optimal trade-off between accuracy and computational efficiency. The early stopping criterion is satisfied shortly after the 100,000th iteration.

4.3.3 Construction of the Industry Multiple

The industry multiple (serving as the prediction of the alternative model and one of the features in the proposed model) is commonly aggregated using a median or arithmetic mean function for its simplicity of interpretation. Often, the median is preferred by practitioners thanks to its robustness to outliers. However, some academic research advocates for the use of harmonic mean (Baker and Ruback, 1999; Liu et al., 2002). We tested all three aggregation methods for the construction of the industry multiple.

Our analysis suggests that the aggregation method does not affect the conclusion of the superiority of the ML models. When comparing the aggregation functions for industry multiples, we observed a trend where the medians produce the lowest values of median errors (MdAPE, and MdPE), whereas the harmonic means result in the lowest mean errors (MAPE and MPE), and the arithmetic means deliver the highest explained variance (R^2) and correlation coefficient (ρ). There is no single optimal method; different aggregation functions perform better under different criteria.

The highest correlation was likely decisive for the GBDT model (where all aggregation methods were considered), as indicated by the strong feature importance score, particularly for the mean industry multiple (Fig. 2). These findings suggest that the mean works well in conjunction with other features as explanatory variables. However, the median or harmonic

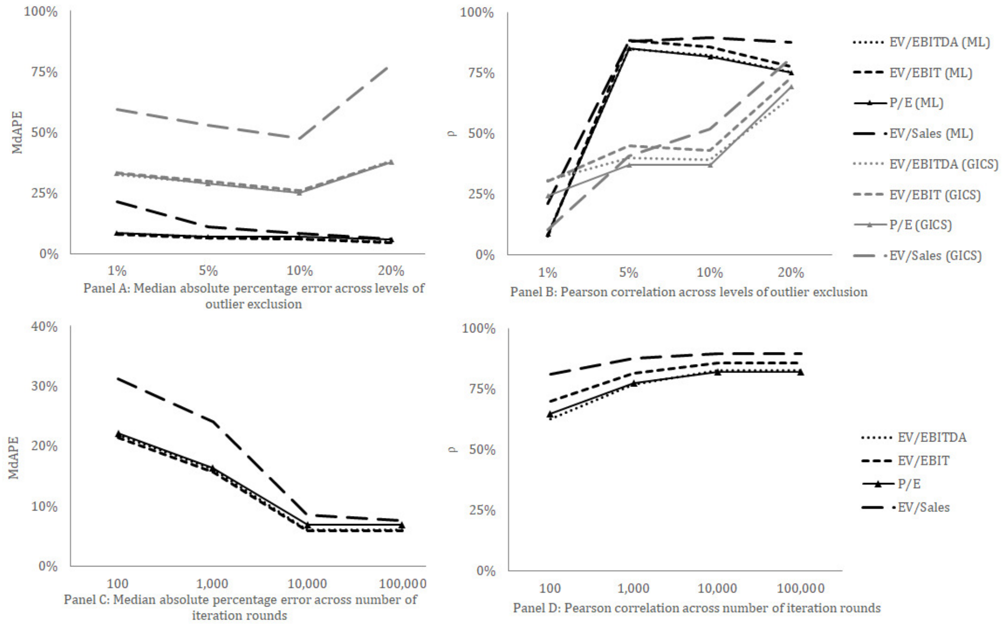

Fig. 3: Panel A and B: MdAPE and Pearson correlation under 1%, 5%, 10% and 20% trimming rule for outlier control. Panel C and D: MdAPE and Pearson correlation under increasing number of iterations rounds hyperparameter (sensitivity illustrated in Panels C and B is not relevant for GICS industry multiple).

Tab. 2: Results of the alternations of the alternative method using the median, (arithmetic) mean and harmonic mean for industry aggregation (GICS Industry level)

	MdAPE	MAPE	MdPE	MPE	R^2	ρ
EV/EBITDA median	**25%**	32%	**0%**	7%	11%	39%
EV/EBITDA mean	27%	36%	9%	16%	**16%**	**41%**
EV/EBITDA harmonic mean	25%	**30%**	−4%	**1%**	9%	40%
EV/EBIT median	**26%**	33%	**0%**	7%	14%	43%
EV/EBIT mean	28%	38%	9%	18%	**19%**	**45%**
EV/EBIT harmonic mean	26%	**31%**	−4%	**1%**	12%	44%
P/E median	**25%**	33%	**0%**	7%	8%	37%
P/E mean	28%	38%	10%	18%	**13%**	**39%**
P/E harmonic mean	26%	**31%**	−4%	**1%**	6%	37%
EV/Sales median	**47%**	77%	**0%**	36%	23%	52%
EV/Sales mean	57%	114%	33%	88%	**28%**	**54%**
EV/Sales harmonic mean	48%	**59%**	−19%	**3%**	8%	49%

Note: The best aggregation function is highlighted for each performance metric and valuation multiple. For the alternative method the median aggregation was used for comparability with the proposed method.

mean aggregation is a more precise choice for the industry multiple method. We chose median aggregation mainly because it is the more common choice among practitioners due to its intuitive explicability.

Additionally, we tested the sensitivity of the level of industry classification used for the prediction. Aggregating all companies together (i.e., disregarding the different industries) represents the most straightforward possible valuation approach. This global multiple method can be viewed as a baseline for evaluating any valuation model. We then compared the global multiple method with the industry multiple method to contextualise the usefulness of the alternative method. The improvement from the simple global multiple to the industry multiple method (consistently using the second most detailed level of GICS classification) amounts to, on average, a 9 p.p. lower MdAPE and MAPE, which corresponds to a 21% (and 13%) improvement of the MdAPE (and MAPE, respectively,) relative to the global multiple. This significant improvement supports the common practice of using industry affiliation for benchmarking and the research supporting the use of industry multiples discussed above.

5 DISCUSSION AND CONCLUSIONS

5.1 Discussion

This research is closely related to the work of Geertsema and Lu (2023), further abbreviated as GL. In this section, we first summarise the aspects shared between these two studies. We list the key modifications representing our approach's contributing novelty, and finally, we discuss the differences in results, including identifying further minor differences in research designs.

This study and the GL study share the following:

- The objective is to enhance the market comparison method by implementing a machine learning algorithm (the GBDT).
- Research design where the ML model is benchmarked against the traditional alternative model(s) (however, the alternative model selection differs, as discussed below).
- A comparable dataset in terms of subjects (namely U.S. public companies of the past four decades; however, the frequency of the dataset differs, as discussed below).

The key modifications to GL's approach introduced by this study include:

- The valuation multiples assumed in the analyses differ. In this study, we predicted P/E, EV/EBIT, EV/EBITDA, and EV/Sales, which we selected because they represent the most common valuation multiples. In contrast, GL prioritised other selection criteria (availability and nonnegativity), and tested the market-to-book, enterprise-value-to-asset, and EV/Sales valuation multiples.
- To extend the variety of benchmark methods that serve to evaluate the performance of the proposed GBDT model, we introduced the mean industry multiple under the 6-digit GICS Industry level as the alternative method. GL tested five other alternative models, one of which represented an industry-based approach using an established industry affiliation system (SIC) and harmonic mean aggregation, while another used an alternative industry classification set up by Hoberg and Phillips (2010, 2016).

Tab. 3 compares the results of GL with results from this study for the proposed method (Panel A) and the alternative method (Panel B). In the first lines of each panel, in the squared brackets, we present the results as intervals ranging from the minimum value to the maximum value of the given metrics across all the targeted multiples. Next, we show the result of EV/Sales separately, which is the only target common to both studies under consideration.

The results in Tab. 3 indicate that both studies converge on the superiority of GBDT models over any alternative tested. A comparison of this study and the GL's (specifically the EV/Sales target) indicates that this study scores improved accuracy (lower errors). We attribute this improvement to the following factors (minor differences in the research design):

- The sample of GL consists of firm-month data (using quarterly accounting data). This study, however, utilises firm-year (with fiscal ends annual data), which likely contains less noise.
- GL trimmed the 10th percentile of selected accounting variables (book value, total assets value and sales) to exclude small firms. However, this study trimmed the lower

Tab. 3: Results of the proposed and alternative methods in this and the GL study. In Panel B, we emphasise the alternative method from the GL study which is the closest to this study (i.e., an industry-based approach using an established industry affiliation system). Squared brackets represent the interval of min. and max. values across all multiples.

	MdAPE	MAPE	MdPE	MPE	R^2	ρ
Panel A: Proposed method, %						
This study (ML), all multiples	[6%; 8%]	[16%; 32%]	[0%; 0%]	[6%; 18%]	[67%; 80%]	[82%; 90%]
GL study (ML), all multiples	[29%; 32%]	[48%; 59%]	[0%; 1%]	[19%; 28%]	[54%; 80%]	[89%; 90%]
EV/Sales (this study: ML)	8%	32%	0%	18%	80%	90%
EV/Sales (GL: ML)	32%	59%	1%	28%	80%	89%
Panel B: Alternative method, %						
This study (GICS, medians), all multiples	[25%; 47%]	[32%; 77%]	[0%; 0%]	[7%; 36%]	[8%; 23%]	[37%; 52%]
GL study (SIC, harm. mean), all multiples	[42%; 64%]	[58%; 129%]	[−35%; −15%]	[0%; 12%]	[−3%; 2%]	[38%; 64%]
GL study (all methods), all multiples	[40%; 86%]	[58%; 384%]	[−35%; 60%]	[−158%; 130%]	[−157%; 83%]	[38%; 86%]
EV/Sales (this study: GICS, median)	47%	77%	0%	36%	23%	52%
EV/Sales (this study: GICS, harm. mean)	48%	59%	−19%	3%	8%	49%
EV/Sales (GL: SIC, harm. mean)	64%	129%	−35%	0%	−1%	38%

and the upper 10% of target variables to exclude outliers of the target variable. As demonstrated above, a reasonably cleaner dataset lead to better accuracy.

- The industry boundaries likely differ between studies as GL applied the Fama-French 49 industries classification for the ML method. We rather used the GICS (i.e. same as for the alternative approach). According to Bhojraj et al. (2003) the GICS is the most suitable industry classification for valuation purposes.

Other methodological differences, which appear not to impact the accuracy, include the following.

- GL sample amounted to 1.8 mil. observations. This study used 93k observations, which were sufficient for the training of outperforming models.
- GL included 97 features, while this study uses 31. This aligns with GL's findings that a high number of variables does not significantly enhance precision. They reported the accuracy improved only by 7 p.p. when they increased the number of input variables from 10 to 97 (Geertsema and Lu, 2023, pp. 348–349).
- To improve the information quality of the data, GL performed a logarithmical transformation of the target variables and commented that it is one of the main sources of improved accuracy of the GBDT (hence the selection of the targeted valuation multiples). Whereas this study inflated the absolute values to 2021 price level.

The main takeaway from this comparison is that GBDT consistently outperforms various alternative methods of market comparison valuation in independent studies.

Nevertheless, it is essential to acknowledge the limitations and challenges associated with this approach. The primary concern is the general perception of machine learning as a 'black box', which can hinder its acceptance and integration into valuation best practice. In spite of the fact that GBDT algorithm is based on decision trees, which are well interpretable, the higher complexity of GBDT comes at the price of losing a portion of this transparency and representability. Additionally, the scope of the data used in our study is restricted to publicly traded U.S. companies with positive market multiple values, which may limit the generalisability of our findings. Future research addressing these limitations (such as extending the geographical scope, and/or including private companies) could further validate the broader applicability.

Practically, we advocate for a balanced view where machine learning is seen as a supplementary, rather than a substitutive, tool to the common methodologies like the industry multiple method and multiples built from a handpicked peer group.

5.2 Conclusions

This article covers the topic of applying machine learning to market comparison valuation method. We aim to provide both academic and practitioner audiences with a perspective on the benefits of integrating machine learning within the traditional valuation framework.

We employed the GBDT (gradient boosting decision tree) model for estimating market multiples. The prediction accuracy was compared with the traditional approach of the mean industry multiple method, which represents common traditional practice. The research was conducted on a dataset of U.S. public companies from 1980 to 2021. We tested a set of high-profile multiples commonly used in appraisals: P/E, EV/EBIT, EV/EBITDA, and EV/Sales.

We found that GBDT significantly enhances the accuracy of market comparison valuations with an average decrease of 24 p.p. in the median average percentage error, and it does not rely on human judgment when developing the valuation multiples. The error reduction of GBDT is attributable to its ability to handle complex financial data and computational efficiency. The results support GBDT's potential as a supplementary tool in valuation practice.

Additionally, we provided insights into market comparison method discovered by the GBDT during the prediction process, which could benefit both academia and practitioners. Firstly, from the feature importance analysis we note that the industry multiple ranked as the most important source of accuracy gain (most pronounced by EV/EBIT and EV/EBITDA), confirming the importance of the industry classification in current practice, thus supporting the industry-based approach to market comparison. Secondly, from testing various aggregation functions for industry-based approaches (median, arithmetic mean and harmonic mean), we drew a conclusion that there is no single optimal method, contrary to the pursuit in the literature for a universally superior approach; different aggregation functions perform better under different criteria.

Upon comparative analysis with related research, we conclude that GBDT consistently outperforms traditional market comparison valuation methods across independent studies. Furthermore, we want to point out that even datasets of reasonable size – both in terms of the number of features and observations – can yield very good results.

6 ACKNOWLEDGEMENTS

This paper is processed as an output of the research project "Use of machine learning in business valuation" registered by the Internal Grant Agency of the Faculty of Finance and Accounting at Prague University of Economics and Business under the registration number F1/54/2020.

7 REFERENCES

Alanis, E. 2022. Forecasting Betas with Random Forests. *Applied Economics Letters*, 29 (12), 1134–1138. DOI: 10.1080/13504851.2021.1912278.

Alford, A. W. 1992. The Effect of the Set of Comparable Firms on the Accuracy of the Price-Earnings Valuation Method. *Journal of Accounting Research*, 30 (1), 94–108. DOI: 10.2307/2491093.

An, J., Bhojraj, S. and Ng, D. T. 2010. Warranted Multiples and Future Returns. *Journal of Accounting, Auditing & Finance*, 25 (2), 143–169. DOI: 10.1177/0148558X1002500201.

Baker, M. and Ruback, R. S. 1999. *Estimating Industry Multiples*. Harvard Business School.

Bartram, S. M. and Grinblatt, M. 2018. Agnostic Fundamental Analysis Works. *Journal of Financial Economics*, 128 (1), 125–147. DOI: 10.1016/j.jfineco.2016.11.008.

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

Bhojraj, S., Lee, C. M. C. and 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. DOI: 10.1046/j.1475-679X.2003.00122.x.

Cheng, C. S. A. and 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. DOI: 10.1023/A:1012050524545.

Damodaran, A. 2002. *Valuation Multiples: First Principles* [online]. Available at: `https://pages.stern.nyu.edu/~adamodar/pdfiles/papers/multiples.pdf`.

De Franco, G., Hope, O.-K. and Larocque, S. 2015. Analysts' Choice of Peer Companies. *Review of Accounting Studies*, 20 (1), 82–109. DOI: 10.1007/s11142-014-9294-7.

Ding, K., Peng, X. and Wang, Y. 2019. A Machine Learning-Based Peer Selection Method with Financial Ratios. *Accounting Horizons*, 33 (3), 75–87. DOI: 10.2308/acch-52454.

Geertsema, P. and Lu, H. 2023. Relative Valuation with Machine Learning. *Journal of Accounting Research*, 61 (1), 329–376. DOI: 10.1111/1475-679X.12464.

Hoberg, G. and Phillips, G. 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *The Review of Financial Studies*, 23 (10), 3773–3811. DOI: 10.1093/rfs/hhq053.

Hoberg, G. and Phillips, G. 2016. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124 (5), 1423–1465. DOI: 10.1086/688176.

James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J. E. 2023. *An Introduction to Statistical Learning: With Applications in Python*. Springer.

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

Lee, C. M. C., Ma, P. and Wang, C. C. Y. 2015. Search-Based Peer Firms: Aggregating Investor Perceptions through Internet Co-Searches. *Journal of Financial Economics*, 116 (2), 410–431. DOI: 10.1016/j.jfineco.2015.02.003.

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

Liu, J., Nissim, D. and Thomas, J. 2007. Is Cash Flow King in Valuations? *Financial Analysts Journal*, 63 (2), 56–68. DOI: 10.2469/faj.v63.n2.4522.

Pinto, J. E., Robinson, T. R. and Stowe, J. D. 2019. Equity Valuation: A Survey of Professional Practice. *Review of Financial Economics*, 37 (2), 219–233. DOI: 10.1002/rfe.1040.

Plenborg, T. and Pimentel, R. C. 2016. Best Practices in Applying Multiples for Valuation Purposes. *The Journal of Private Equity*, 19 (3), 55–64. DOI: 10.3905/jpe.2016.19.3.055.

Rhodes-Kropf, M., Robinson, D. T. and Viswanathan, S. 2005. Valuation Waves and Merger Activity: The Empirical Evidence. *Journal of Financial Economics*, 77 (3), 561–603. DOI: 10.1016/j.jfineco.2004.06.015.

Sellhorn, T. 2020. Machine Learning und empirische Rechnungslegungsforschung: Einige Erkenntnisse und offene Fragen. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 72 (1), 49–69. DOI: 10.1007/s41471-020-00086-1.

8 ANNEX

Data Retrieval Definition

The data analysed in the study were downloaded from the Thomson Reuters Refinitiv Eikon database: `https://eikon.refinitiv.com/login` under the license agreement granted to the Prague University of Economics and Business. The sample comprised the universe of all public companies headquartered in the USA. The sample was further refined as follows: Active and Inactive companies, Primary Issues only, Instrument Type limited to the Fully Paid Ordinary Shares and the Ordinary Shares. The sample period spans fiscal year ends from 1980 through 2021. There are 10 thousand companies meeting the above-mentioned criteria and 73k–93k firm-year observations (depending on the target variable) as of the date of the data retrieval in April 2022).

Based on discussions with Reuter's representatives, Reuter's definition of the universe of public companies excludes companies that were public during the sample period but have been delisted since. Status inactive relates solely to companies which seized to issue financials whatsoever. Thus, delisted and yet active companies were not included in the sample.

Tab. 4: Summary statistics of the variables considered in this study. All variables are displayed after pre-processing as described in the Proposed method section (features are displayed after pre-processing for the EV/Sales multiple). N refers to the number of observations. Columns p10–p90 show the respective percentiles

Item	Usage	Source	N	Mean	p10	p50	p90
Enterprise value to EBIT (EV/EBIT)	used as target and for a computed feature	TR.EVEBIT	62,121	14.86	7.40	12.71	25.93
Enterprise value to EBITDA (EV/EBITDA)	used as target and for a computed feature	TR.HistEnterpriseValueEBITDA	65,882	10.85	5.61	9.59	18.23
Enterprise value to sales (EV/Sales)	used as target and for a computed feature	R.HistEnterpriseValueRevenue	74,411	3.37	0.61	1.88	8.74
Price to earnings (P/E)	used as target and for a computed feature	TR.HistPE	58,072	20.30	10.27	17.49	34.97
Features pro EV/Sales:							
Asset turnover	used as feature	TR.AssetTurnover	73,426	1.07	0.21	0.82	1.92
Average net trade cycle in days	used as feature	TR.AvgNetTradeCycleDays	72,202	−47.74	−23.82	60.17	186.33
Average total asset (growth 7Y)	used as feature	computed	50,116	267.48	−0.36	0.55	4.74
Average total assets (in mil.)	used as feature and for a computed feature	TR.TotAssetsPeriodToPeriodAvg	73,665	7.32	0.02	0.65	11.43
Average total assets (growth 1Y)	used as feature	computed	72,277	3.40	−0.12	0.06	0.51
Average total assets (growth 3Y)	used as feature	computed	64,843	55.73	−0.26	0.21	1.79
Average total assets (growth 5Y)	used as feature	computed	57,262	129.57	−0.33	0.38	3.18
Current ratio	not used	TR.CurrentRatio	68,011	2.93	0.71	1.90	5.24
Earnings retention rate	used as feature	TR.EarningsRetentionRate	50,751	0.35	0.08	0.94	1.00
EBIT margin	used as feature	TR.EBITMarginPercent	74,411	−21.23	−45.45	8.45	28.26
EBITDA margin	not used	TR.EBITDAMarginPercent	74,411	−12.81	−35.86	13.03	39.62
Employees	not used	TR.EmployeesscalYrEndPeriodToPeriodAvg	59,763	11,508	51	1,605	22,916
Equity (in mil.)	not used; used for a computed feature	TR.TotalEquityAndMinorityInterest	74,299	2.29	0.00	0.29	4.23
Equity (growth 1Y)	not used	computed	73,466	11.18	−0.33	0.06	0.86
Equity (growth 3Y)	not used	computed	68,878	62.31	−0.56	0.22	2.89
Equity (growth 5Y)	not used	computed	61,364	93.44	−0.64	0.37	4.77
Equity (growth 7Y)	not used	computed	53,996	69.96	−0.67	0.53	6.76
Fixed asset turnover	used as feature	TR.FixedAssetTurnover	70,818	57.84	0.62	5.37	28.90
GICS Industry Group EV/Sales multiple (mean)	used as feature	computed	73,142	3.36	1.49	2.69	6.47
GICS Industry Group EV/Sales multiple (median)	used as feature	computed	73,142	2.48	0.99	1.86	4.87
GICS Industry EV/Sales multiple (mean)	used as feature	computed	73,102	3.35	1.23	2.71	6.25
GICS Industry EV/Sales multiple (median)	used as feature	computed	73,102	2.58	0.92	1.88	5.08
GICS Subindustry EV/Sales multiple (mean)	not used	computed	72,911	3.35	1.07	2.65	6.77
GICS Subindustry EV/Sales multiple (median)	not used	computed	72,911	2.67	0.85	1.88	5.33
GICS Sector EV/Sales multiple (mean)	used as feature	computed	73,147	3.36	1.66	2.93	5.86
GICS Sector EV/Sales multiple (median)	used as feature	computed	73,147	2.38	1.04	1.88	4.49
Income tax rate	used as feature	TR.IncomeTaxRatePct	50,547	23.62	0.00	31.58	42.02
Long term debt to total capital	used as feature	TR.LTDebtToTtlCapitalPct	72,192	38.80	0.00	24.14	68.17
Net income after taxes (in mil.)	used as feature	TR.NetIncomeAfterTaxes	74,406	0.32	−0.04	0.02	0.59
Net income after taxes (lag −1Y) (in mil.)	used as feature	computed	73,807	0.29	−0.04	0.02	0.54
Net income after taxes (lag −3Y) (in mil.)	used as feature	computed	70,883	0.27	−0.03	0.01	0.50
Net income after taxes (lag −5Y) (in mil.)	used as feature	computed	64,820	0.26	−0.03	0.01	0.47
Net income after taxes (lag −7Y) (in mil.)	used as feature	computed	57,578	0.25	−0.03	0.01	0.46
PPE (gross) (in mil.)	not used	TR.PptyPlantEqpmtTtlGross	67,294	3.10	0.00	0.23	5.34
PPE (net) (in mil.)	used for a computed feature	TR.PropertyPlantEquipmentTotalNe	72,288	1.74	0.00	0.10	3.04
PPE (net) to total assets	used as feature	computed	71,589	0.27	0.03	0.17	0.70
PPE net to PPE gross ratio	used as feature	computed	67,035	0.00	0.28	0.53	0.82
PPE net (growth 1Y)	not used	computed	70,885	0.00	−0.20	0.03	0.71
PPE net (growth 3Y)	not used	computed	65,752	0.00	−0.36	0.21	2.71
PPE net (growth 5Y)	not used	computed	58,394	0.00	−0.45	0.37	4.60
PPE net (growth 7Y)	not used	computed	51,394	0.00	−0.51	0.51	6.59
Quick ratio	used as feature	TR.QuickRatio	68,011	2.41	0.48	1.40	4.41
Return on assets (ROA)	not used	TR.ROATotalAssetsPercent	73,664	−18.59	−32.20	3.67	14.78
Return on equity (ROE)	used as feature	TR.ReturnonAvgTotEqtyPctNetIncomeBeforeExtraItems	68,236	−13.40	−45.63	9.36	30.82
Revenue (in mil.)	used for a computed feature	TR.TotalRevenue OR TR.BankTotalRevenue	74,411	4.17	0.01	0.50	7.91
Revenue (growth 1Y)	used as feature	computed	73,147	14.39	−0.17	0.07	0.56
Revenue (growth 3Y)	used as feature	computed	69,235	114.92	−0.26	0.21	1.90
Revenue (growth 5Y)	used as feature	computed	63,195	203.48	−0.30	0.37	3.64
Revenue (growth 7Y)	used as feature	computed	56,050	168.36	−0.33	0.51	5.50
Total debt to total equity	used as feature	TR.TtlDebtToTtlEquityPct	68,571	235.34	0.00	43.36	224.67
Headquarters country	not used	TR.HeadquartersCountry	74,411	n/a	n/a	n/a	n/a
Research and development expense (in ths.)	not used	TR.ResearchAndDevelopment	29,726	258.17	0.83	23.32	340.26

AUTHOR'S ADDRESS

Veronika Staňková, Prague University of Economics and Business, W. Churchill sq. 4, 130 67 Prague, Czech Republic, e-mail: xstav31@vse.cz