RELAÇÕES INTERNACIONAIS NO MUNDO ATUAL

CENTRO UNIVERSITÁRIO CURITIBA - UNICURITIBA - VOLUME 6 - NÚMERO 40/2023 I e-6335 I ABRIL-JUNHO I CURITIBA/PARANÁ/BRASIL - PÁGINAS 1 A 20 - ISSN: 2316-2880

COMPARING THE METHODS OF PREDICTION AND BUSINESS COST ESTIMATION BASED ON INDUSTRY-SPECIFIC, CLUSTERING, AND REGRESSION MODELING

Vitaly Kovalev
Saint Petersburg State University – Russia
https://orcid.org/0000-0001-7696-3741
E-mail: v.v.kovalev@spbu.ru

Petr Koklev
Saint Petersburg State University – Russia
https://orcid.org/0000-0003-2594-7973
E-mail: koklevp@gmail.com

ABSTRACT

Background: Being the most common, the relative valuation method plays a special role in estimating the value of a business. Many studies consider various applications of multipliers. However, the study results are often contradictory. **Objective**: This article aims to determine the best method for assigning a fair value to the multiplier of the assessee company. **Methods**: Within the empirical study, the effectiveness of three forecasting methods (industry-specific, cluster, and regression) was compared. **Results**: Regression modeling is the most accurate approach and outperforms other methods in terms of *MAE* and R^2. The best multiplier is considered the one that can reach the maximum metrics for assessing the quality of models. The largest variance within the existing data set can be explained for multiples based on sales *P/Sales* and *EV/Sales*. Other issues were also solved in the course of the study. The best method for determining groups of peer companies has been determined. **Conclusion**: The proposed cluster approach is superior to the industry-specific approach. While comparing these approaches, the authors identify the best measure for calculating the typical value of multipliers within a group of peer companies. The simple average and median indicators were more accurate than the other calculation methods.

Keywords: Relative valuation; Multiples; Industry coefficients; Market multipliers; Business valuation; Company's value.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

COMPARAR OS MÉTODOS DE PREVISÃO E ESTIMATIVA DE CUSTOS EMPRESARIAIS BASEADOS EM MODELOS ESPECÍFICOS DA INDÚSTRIA, DE AGRUPAMENTO E DE REGRESSÃO

RESUMO

Antecedentes: Sendo o mais comum, o método de avaliação relativa desempenha um papel especial na estimativa do valor de um negócio. Muitos estudos consideram várias aplicações de multiplicadores. No entanto, os resultados dos estudos são muitas vezes contraditórios. **Objectivo:** Este artigo visa determinar o melhor método para atribuir um valor justo ao multiplicador da empresa arrendatária. **Métodos:** No âmbito do estudo empírico, foi comparada a eficácia de três métodos de previsão (específicos da indústria, agrupamento, e regressão). **Resultados**: A modelização da regressão é a abordagem mais precisa e supera outros métodos em termos de MAE e R^2. O melhor multiplicador é considerado aquele que pode atingir a métrica máxima para avaliar a qualidade dos modelos. A maior variação dentro do conjunto de dados existente pode ser explicada para múltiplos com base nas vendas P/Vendas e EV/Vendas. Outras questões foram também resolvidas no decurso do estudo. Foi determinado o melhor método para determinar grupos de empresas homólogas. **Conclusão:** A abordagem de agrupamento proposta é superior à abordagem específica da indústria. Ao comparar estas abordagens, os autores identificam a melhor medida para calcular o valor típico dos multiplicadores dentro de um grupo de empresas homólogas. A média simples e os indicadores medianos eram mais precisos do que os outros métodos de cálculo.

Palavras-chave: Avaliação relativa; Múltiplos; Coeficientes industriais; Multiplicadores de mercado; Avaliação empresarial; Valor da empresa.

1 INTRODUCTION

The use of multiples is the most common way to assess a company's value. Multipliers are used in many spheres, for example, in the reports and recommendations of stock market analysts, when an investment bank draws up conclusions on the placement price in IPO (DeAngelo, 1990) and M&A transactions. Finally, multipliers are used in regulations and when evaluating private companies in tax litigation (Jackson et al., 2013).

There is no consensus on the best practices for using the comparative approach among economists. In the first part of the article, we reviewed the existing literature and drew the following conclusion – the studies on the optimal use of coefficients are often contradictory. Thus, it is relevant to determine the best method for predicting a multiplier for the company being assessed.

The second part of the article discusses methodological aspects of the empirical study. Their main purpose is to find the best way to determine the coefficient of the company being assessed. As a result, we defined various characteristics of the three

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

compared forecasting methods and the used metrics for assessing the quality of models. In the third part of the article, we described the dataset used, the principles of constructing the test sample, as well as the results and discussion of the empirical study.

2 LITERATURE REVIEW

To determine the value to a company, we used the following three-stage procedure (Palepu et al., 2022):

1. Choosing the value driver, i.e. the denominator of a multiplier. For example, it could be revenue when using the EV/Sales multiplier.
2. Defining a group of comparable companies and calculating the average value of the M multiplier for this group.
3. Estimating the value of the target company V_i by multiplying its value driver by the typical value of the multiplier.

$$V_i = M * (Importance\ of\ a\ Company\ Value\ Driver\ i)$$

It is understood that the company being assessed "deserves" the same coefficient as the group of peer companies since it has similar characteristics.

There are many studies in which scholars try to determine the best ways to use market coefficients. Let us review the existing literature on the key aspects of applying the comparative approach.

The identification of peer companies. The key step in the application of multipliers is the formation of a group of comparable companies. There are many proven methods. In one of the early studies of 1981, J. Boatsman and E. Baskin (1981) concluded that selecting companies based on similar historical earnings growth rates can significantly reduce the forecast error for $\mathrm{P/E}$ if compared to random selection.

The classic works on choosing the best multiplier and methods for determining a group of similar companies were written by M.S. LeClair (1990), A.W. Alford (1992), R.P. Beatty, R.M. Susan, and R. Thompson (1999), S.C. Gilson, E.S. Hotchkiss, and R.S. Ruback (2000), and M. Kim and J.R. Ritter (1999).

The formation of a group of peer companies based on industry affiliation is the most common approach. Despite its simplicity, this approach is accurate. The use of industry affiliation allows applying different accounting practices to similar economic events. In

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

other words, it ensures the comparability of reporting by different companies. As a rule, companies in the same industry have similar accounting policies (Foster, 1986).

However, the industry-specific approach has its drawbacks. According to McKinsey researchers T. Koller, M. Goedhart and D. Wessels (2005), companies in the same industry do not necessarily have similar fundamental values that determine the value of an asset in the framework of DCF models. Companies in the same industry can have significant differences in the main drivers of asset values – expected growth rates of cash flows, return on equity, and risks. The rest of market ratio forecasts using the industry-specific method are often high and inferior to the other methods for determining a group of peer companies.

I. Dittmann and C. Weiner proposed to use return on assets (ROA) to identify peer companies. The authors proved that this approach was better than the choice based on industry affiliation and grouping in terms of balance sheet currency. The results are based on a representative sampling from 1993-2002, covering different countries and markets (Dittmann & Weiner, 2005).

A peculiar approach was developed by a group of scientists from the United States, who proposed using a co-search algorithm for queries within the traffic of the EDGAR system. EDGAR is database of financial statements of US stock market companies available on the website of the US Securities and Exchange Commission (SEC) (SEC, n.d.). Peer companies are those companies that appear in search queries of portal users in chronological order (Lee et al., 2015).

To more accurately determine peer companies, scientists also offer more complex methods. The regression approach can be regarded as the limiting case when a group of comparable companies can consist of all stock market enterprises. S. Bhojraj, C.M. Lee (2002), and A. Damadoran (2012) recommended using regression analysis to determine the fair value of the multiplier for the company being assessed. The difference between companies is controlled through independent variables that are proxy variables for the determinants of multipliers.

The best multiplier. Accounting income reflects past events and is not a good predictor of future earnings. It comes as no surprise that forward multipliers are more accurate. J. Liu, D. Nissim and J. Thomas (2002) showed that the P/E calculated using forecast earnings more accurately predicts the company's value if compared to the P/E based on the current earnings (trailing 12 months). This observation is statistically significant, constant over time, and applies to both US and EU capital markets (Schreiner & Spremann, 2007). Unfortunately, profit forecasts are not made for all

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

companies. Therefore, the conclusion that forward multipliers are better applies only to a specific group of companies. Profit forecasts are used for large-sized and mature companies monitored by a large number of analysts. To improve the accuracy of forecasts, the historical and forward values of coefficients can be used in combination (Yoo, 2006). To obtain the most generalized results, forward multipliers are not considered in this study.

Other scientific works on the accuracy of various market coefficients provide rather mixed results. Early studies (Liu et al., 2002) showed that the accuracy of earnings-based ratios outperforms the multipliers determining the book value of assets. The latter are more accurate than sales-based multipliers. However, a more recent study (Deng et al., 2012) indicates that sales-based multipliers contain a larger sample of companies and are more accurate.

Contrary to T. Copeland's well-known saying "Cash is the King" (Copeland et al., 2000), industry-based cash flow coefficients are the least preferred (Liu et al., 2007). This refers to an item in the cash flow statement rather than the free cash flow concept used in discounted cash flow models. This rule applies not only to the United States, but also to nine other regions: Australia, Canada, France, Germany, Hong Kong, Japan, South Africa, Taiwan, and the UK. J. Liu, D. Nissim, and J. Thomas explained it by the fact that profit measures were more representative drivers of value due to their ability to aggregate both cash flow and accrual transaction information that resulted in cash flows of a higher economic value. For example, credit sales are relevant to the value of a company. However, this transaction will only be shown in its income statement and will not be reflected in the cash flow statement. On the other hand, the purchase of inventory is a zero net present value transaction and is neutral to the company's value. Unlike cash flow statement items, profit measures remain the same after such a transaction and reflect this invariance.

There are approaches based on the combination of several multipliers. For example, S. Penman (1998) suggested combining P/E and M/B with due regard to weight coefficients. The weighted combination of multipliers allows increasing the accuracy of estimates within a large sample.

In general, the results obtained are rather contradictory. While comparing the accuracy of the EV/EBITDA, EV/Sales, P/E, and P/S coefficients within the framework of the empirical study, we accumulate knowledge in determining the best multiplier.

The calculation of a typical multiplier for a group of peer companies. To assign a typical multiplier value for a group of peer companies, various measures of the central

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

tendency can be used: arithmetic mean, harmonic mean, geometric mean, median, mode, etc. M. Baker and R. Ruback (1999) supported the harmonic mean. According to the authors, the harmonic mean is always less than the arithmetic mean, therefore the cost estimate based on the arithmetic mean will be overestimated. In addition, there are studies whose results contradict this conclusion. Based on the sampling of US and EU companies, V. Herrmann and F. Richter (2003) concluded that the use of the median was preferable compared to the harmonic mean, which systematically underestimates the value of companies. T. Plenborg and R. Pimentel (2016) claimed that there was no clear evidence in favor of one or another approach to calculating the average.

The comparability of financial statements. Financial statements include some components that are irrelevant to cost: provision charges for impairment, litigation costs, and non-recurring costs. These components distort earnings data and lead to a biased valuation of business (Petersen & Plenborg, 2012). For this reason, it is believed that financial statements should be adjusted. In practice, the benefits of such amendments are rather small, and many assessors ignore this aspect.

The comparability of financial statements should not be neglected. Back in 1978, W. Beaver and D. Morse showed that variations in the *P/E* coefficient were conditioned by accounting policies, i.e. different interpretations of the same economic events (Beaver & Morse, 1978). P. Zarovin came to the same conclusion in 1990 (Zarowin, 1990). The difference in financial reporting standards at the international level does not allow expanding the sample of companies and deprives scientists and practitioners of the beneficial effects of the law of large numbers (Land & Lang, 2002). Differences in accounting policies lead to the fact that similar (different) companies will look different (similar). The work of S. Young and Y. Zeng (2015) demonstrated that improving the comparability of statements as part of the convergence of financial statements (mandatory adoption of IFRS) for 1997-2011 significantly increased the accuracy of estimates. It is believed that the minimum requirement for the use of multipliers is compliance by the companies with a single financial reporting standard. All the companies considered adhere to the *USGAAP* standards.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

3 METHODS

3.1 The considered methods for predicting multipliers

The main objective is to determine the most accurate method for assigning a fair value of the multiplier for the company being assessed. Auxiliary tasks are also to determine the most preferred multiplier, the method for calculating the typical value of the coefficient for a group of peer companies, as well as the optimal method for determining a group of peer enterprises.

The considered methods for predicting multipliers. To determine the optimal method for predicting multipliers, three methods are compared.

1. Determining a group of peer companies based on industry affiliation. The company being assessed is assigned a typical multiplier value based on data from companies in the industry to which the company belongs.

2. Determining a group of peer companies using clustering. Clusters are defined using a popular machine learning method, i.e. the *k-means* algorithm. In addition, the appendix discusses the results obtained through more advanced clustering algorithms.

3. Applying gradient boosting using *GBDT* decision trees for the regression modeling of the multiplier value.

When comparing the effectiveness of the first two strategies (industry-specific and clustering), we identified the best method for determining groups of peer companies, as well as the optimal measure of the central tendency. The comparison of the accuracy of all three strategies allowed us to determine the most accurate multiplier and the best method for assigning (forecasting) a fair value of the company being assessed. Let us consider each forecasting method.

3.1.1 The industry-specific approach

Within the framework of this method, the $\widehat{m}_i$ multiplier for the i assessed company is based on the typical value of the multiplier for companies from the C_i set that belong to the same industry as the i company.

$$\widehat{m}_i = avg(C_i)$$

The avg function calculates the average value of the multiplier for the C_i set of companies. Four methods of calculating the average are as follows: arithmetic mean,

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

median, geometric mean, and harmonic mean. For instance, the formulas for calculating the geometric and harmonic mean are presented below.

$$\widehat{m}_i = \sqrt[|C_i|]{\prod_{j:j \in C_i} m_j}$$

$$\widehat{m}_i = \frac{|C_i|}{\sum_{j:j \in C_i}(1/m_j)}$$

$|C_i|$ is a set of companies within the C_i set.

3.1.2 Clustering

The final part of the process of assigning a multiplier to the company being assessed within the framework of clustering is identical to the industry-specific approach. However, this group of peer companies consists of companies in the same cluster rather than the industry. In a sense, clustering algorithms assign an artificial industry for companies based on financial reporting data. The method is to divide features into several non-overlapping subsets called clusters. These clusters will determine the "industry" of a particular company (Zura & Yu, 2016). To build clusters, the *k-means* method will be applied.

The *k-means* method is one of the simplest approaches to clustering introduced by H. Steinhaus in 1957 (Steinhaus, 1957). Let us divide the feature space into *K* clusters. In the first step, the algorithm randomly determines the position of K centroids $b_1, \ldots, b_K$ in each cluster. It is desirable that the centroids be evenly distributed throughout the feature space. Each centroid can be understood as the most typical representative of the corresponding cluster. Then the i company is assigned the cluster whose centroid is the closest in the feature space of the company.

$$Cluster_i = arg \min_{c \epsilon \{1,\ldots,K\}} \|X_i - b_c\|^2$$

The value of each K centroid is updated by calculating the average of the features for X_i companies belonging to the same cluster:

$$b_c = \frac{\sum_{i:C_i=c} X_i}{\sum_{i:C_i=c} 1}$$

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

The process of assigning companies to clusters and the subsequent updating of centroids takes place until the structure of these clusters ceases to change. The resulting model can be used to assign a cluster to any company. This article uses a slightly modified and more efficient version of the described algorithm – k-means++ proposed by D. Arthur and S. Vassilvitskii (2007). The graphical result of teaching the clustering model using the k-means method is shown in Figure 1. For visualization purposes, the feature space is reduced to three dimensions using principal component analysis (PCA). When training models, the PCA method will not be applied.

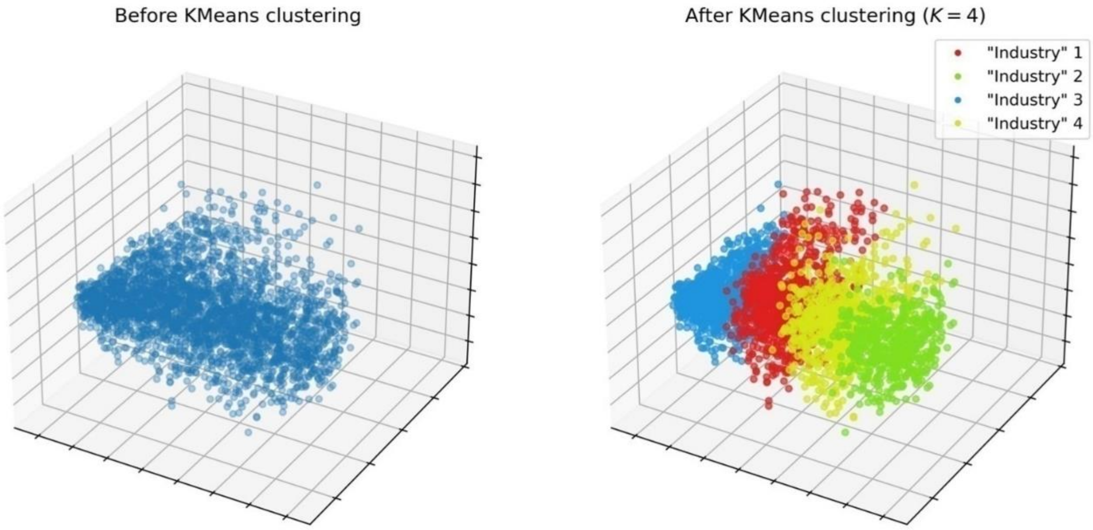

Figure 1. The process of teaching the clustering model using the *k-means* algorithm for the existing dataset (four clusters).

The number of K clusters is chosen in advance, which is a difficult task. When choosing clusters, it is necessary to be guided by theory or practical knowledge. A bad choice of K clusters can lead to poor-quality clusters. The first option is to use K=300, i.e. the approximate number of unique industries in the data set provided by the service financial data Alpha Vantage. AlphaVantage is a financial data API. All the company data used in the study is presented by this service (AlphaVantage, n.d.). We can also use the elbow method (Ketchen & Shook, 1996). This method determines the optimal K at the inflection point of the distortion function. Distortion is a widespread metric for evaluating the quality of clustering. Within the framework of this method, it is optimal to use 36 clusters. Thus, it is expedient to test two values of the number of K clusters ∊ {300.36}.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

3.1.3 Regression using gradient boosting

Multipliers can also be forecasted without defining a group of peer companies. Thus, the $\widehat{m}_i$ multiplier for a company will be modeled using the data of the company's financial statements.

$$\widehat{m}_i = h(x_i)$$

h is the x_i multiplier assignment function

This feature vector of the company being assessed contains data from the three main forms of financial statements for the previous eight quarters (Form 10-Q). Gradient boosting based on *GBDT* decision trees will be used for the regression modeling of multipliers.

Boosting is an ensemble meta-algorithm of machine learning that significantly increases the efficiency of simple methods. Boosting can be used within any machine learning algorithm. However, the choice of decision trees as the base model is the most common. Boosting combines the forecasts of multiple decision trees. The forecasts obtained with multiple decision trees are more stable. Initially, boosting was applied only to classification (Freund, 1995). Later D. Friedman and his colleagues proposed a regression method (Friedman, 2001; Friedman et al., 2000).

The gradient boosting model will be trained using the default set of hyperparameters within the *CatBoost* library, without searching for the optimal set using *RandomSearch* (Bergstra & Bengio, 2012) and cross-validation. CatBoost is a library for training decision tree models based on gradient boosting developed by PJSC Yandex (CatBoost, n.d.).

3.2 Evaluating the quality of models

The comparative effectiveness of methods will be evaluated using the R^2 index calculated for a separate test set. In addition, most potential readers have an idea about the values of common multipliers. Therefore, the mean absolute forecast error *(MAE)* (Tarkhanova et al., 2020) will be used as an auxiliary quality assessment metric.

$$MAE_{oos} = \frac{\sum_{\in S_{test}} |\widehat{m}_i - m_i|}{|S_{test}|}$$

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

The best multiplier forecasting method (industry-specific, clustering, or regression) is the one for which R^2(*MAE)* is the highest (lowest). In addition, the most preferable multiplier is the one for which the best value of the model quality assessment metric is achieved. Finally, comparing the accuracy of the industry-specific and clustering approaches helps identify the best way to calculate the average for a group of peers, as well as the best method for identifying groups of peers.

The appendix presents the results of advanced clustering algorithms. Due to the high complexity and lack of incremental efficiency compared to the simpler *k-means* algorithm, the methodological aspects of advanced algorithms are beyond the scope of this study.

4 EMPIRICAL STUDY AS EXEMPLIFIED BY THE NASDAQ AND NYSE COMPANIES

4.1 Data

As independent variables for clustering and regression algorithms, we used data from the three main forms of financial statements for the previous eight quarters. The financial reporting data and industry affiliation of companies are provided by *Alpha Vantage*. We also applied the following data pre-processing procedures: iterative imputation for missing data (most companies have missing values for some items in their financial statements) and quantile transformation to standardize independent variables. The listed data pre-processing steps are necessary because the *k-means* clustering algorithm can only be applied when all variables have a similar order of value and dispersion. For industry-specific and regression methods, imputation of missing data and standardization are not required. The industry-specific approach for determining groups of peer companies does not imply the use of financial reporting data. Decision trees are invariant to data transformations. In addition, the CatBoost machine learning library offers native support for working with missing data, without the need for imputation.

Furthermore, observations with extreme multiplier values (with a value of more than 100) and negative values were excluded from the data set. The final sample consists of six data sets obtained at different points in time. The data were collected quarterly from March 2021 to August 2022.

Considering the split of 85 vs. 15%, the total training and testing sampling includes

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

16,000 and 3,000 companies, respectively. The sampling size is indicative since the data was pre-processed for each multiplier. The exact data on the sampling size are presented in Table 1. For the *EV/Sales* multiplier, the final sampling is larger compared to the *P/E* coefficient since a significant number of companies should be excluded for the latter coefficient due to the negative value of net income. On the contrary, revenue is always positive (Bezdudnaya et al., 2018).

Table 1. The sampling for the analyzed multipliers between March 2021 and August 2022

Multiplier	Training data size (number of companies)	Testing data size (number of companies)
$EV/EBITDA$	12,872	2,276
$EV/Sales$	16,558	2,925
P/E	11,964	2,115
$P/Sales$	18,749	3,311

The data for each quarter is divided into training and test sets. Thus, there are several conservative estimates of forecast accuracy for each method. In practice, the test set consists of only one company, i.e. the assessed one and the data from all other companies can be used to develop a model.

For each of the six quarters (March 2021 – August 2022), the results of forecast accuracy will be obtained for each forecasting method, multiplier, and central tendency measure. Except for the regression approach that does not define groups of peer companies and does not calculate the typical value of the multiplier. The overall performance of various methods will be assessed based on the average quality assessment metric (R^2 and *MAE*) during six quarters.

4.2 Results and discussion

The results for the industry-specific method are shown in Table 2. Within the framework of this forecasting method, the largest share of variance was explained using multipliers based on the *P/Sales* and *EV/Sales* sales: 15.61 and 14.95%, respectively. These results can be achieved using the simple mean. From the viewpoint of the mean absolute error, the use of the simple mean is no longer effective (Ksenofontova et al., 2017). Since the typical range of values for each multiplier is different, MAE can only be used to compare the performance of approaches for a given multiplier rather than comparisons between multipliers. For both the geometric mean and the median, the absolute forecast errors turn out to be lower. The median slightly exceeds the geometric mean.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Table 2. The forecasting accuracy of multipliers of test sample companies for the industry-specific approach

Multiplier	Method for calculating the average value	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	Geometric	-8.75%	9.03
	Harmonic	-43.63%	10.84
	Median	-2.18%	8.54
	Simple	2.75%	9.31
$EV/Sales$	Geometric	6.86%	4.09
	Harmonic	-12.68%	4.68
	Median	7.75%	**4.00**
	Simple	**14.95%**	4.63
P/E	Geometric	6.12%	10.69
	Harmonic	-13.86%	11.68
	Median	7.90%	10.33
	Simple	11.70%	11.08
$P/Sales$	Geometric	10.17%	3.74
	Harmonic	-6.62%	4.15
	Median	10.98%	**3.68**
	Simple	**15.61%**	4.25

When using the harmonic mean, the p-square is in the negative zone for each multiplier and the average forecast error is relatively high. Thus, the following conclusions are valid for the industry-specific method:

– The most accurate multiples are $P/Sales$ and $EV/Sales$;

– The best p-squared value can be achieved using a simple average;

– The median can be used to minimize average absolute forecast errors.

Table 3 shows the results of clustering using *the k-means* algorithm for 300 clusters. The efficiency of 300 clusters (K=300) turned out to be higher compared to 36 (K=36) (Appendix 1). Like within the industry-specific approach, the largest p-squared value is achieved by applying based multipliers based on sales and together with the simple mean.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Table 3. The forecasting accuracy of multipliers of test sample companies for the *k-means* clustering $(K = 300)$

Multiplier	Method for calculating the average value	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	Geometric	-0.17%	8.57
	Harmonic	-24.83%	9.62
	Median	-0.40%	8.30
	Simple	7.46%	9.07
$EV/Sales$	Geometric	11.35%	3.98
	Harmonic	-6.64%	4.35
	Median	14.57%	**3.89**
	Simple	**21.60%**	4.40
P/E	Geometric	-1.23%	10.73
	Harmonic	-20.00%	11.63
	Median	-0.71%	10.46
	Simple	4.19%	11.12
$P/Sales$	Geometric	18.85%	3.59
	Harmonic	3.39%	3.87
	Median	17.99%	**3.53**
	Simple	**26.34%**	3.96

The consideration of both methods based on the definition of groups of peer companies leads to the same conclusions, making them more generalized and reliable. Comparing the accuracy of methods with each other indicates the superiority of the clustering approach. The creation of artificial industries increases the maximum p-square: 26.34 vs. 15.61%. The mean absolute error of forecasts for most combinations "multiplier – method of calculating the average value" also turns out to be preferable for *k-means* clustering. Thus, the proposed approach has a high economic value. We can draw the main conclusions:

– The clustering approach for determining groups of peer companies is superior to the industry-specific one;

– The preferred measures of central tendency are the simple mean (to maximize R^2) and the median (to minimize *MAE*).

Let us consider the most complex method, i.e. a regression model based on the gradient boosting of decision trees (*GBDT*). Test R^2 and *MAE* for the regression model are shown in Table 4. As noted, the regression approach does not require the identification of groups of peer companies and the subsequent calculation of their typical value. For this reason, there is no column "Method for calculating the average value" in Table 4.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Table 4. The forecasting accuracy of multipliers of test sample companies for the *GBDT* gradient boosting regression model

Multiplier	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	26.49%	2.04
$EV/Sales$	**46.25%**	**0.79**
P/E	28.03%	2.52
$P/Sales$	**51.81%**	**0.80**

It is worth mentioning completely different values of the model quality metrics. *GBDT* outperforms both methods mentioned above. For each multiplier, gradient boosting significantly outperforms both industry-specific and clustering methods. For the *P/Sales* ratio, the coefficient of determination is 51.81%, which is almost twice bigger than the next best option (26.34% for the clustering approach). In addition, the average absolute forecast errors for regression are 3-5 times smaller than those for methods based on the identification of groups of peer companies.

5 CONCLUSION

The use of multipliers is the most common way of estimating the value of a business entity. However, the existing literature revealed a lack of consensus among researchers on the best ways to use coefficients. The empirical research determines the optimal application of multipliers within the key aspects of their use.

Within the framework of methods based on the definition of peer companies, the proposed clustering approach based on the *k-means* algorithm is superior in accuracy to the traditional, industry-specific approach. However, both methods are suboptimal. Regression modeling using the gradient boosting of decision trees is superior to both clustering and industry-specific approaches. Due to the ability to obtain a larger sample, sales-based multipliers (*EV/Sales and P/Sales*) not only provide more general results but also achieve greater accuracy in cost estimates than profit multipliers. The optimal measures of the central tendency in calculating the typical value of the multiplier for a group of peer companies are the simple mean (p-squared maximization) and the median (*MAE* minimization). The consideration of a larger sample, as well as other capital markets within the framework of the proposed methodology, will determine the generalization of the results obtained in the study.

REFERENCES

Alpha Vantage (n.d.). Free Stock APIs in JSON & Excel. Alpha Vantage. https://www.alphavantage.co/: (Accessed date: July 15, 2022).

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.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Arthur, D., & Vassilvitskii, S. (2007). k-Means++: The advantages of careful seeding. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete algorithms* (pp. 1027-1035). Philadelphia, PA: Society for Industrial and Applied Mathematics.

Baker, M., & Ruback, R.S. (1999). *Estimating industry multiples*. Working Paper. Harvard University, Cambridge, MA.

Beatty, R.P., Susan, R.M., & Thompson, R. (1999). The method of comparables and tax court valuations of private firms: An empirical investigation. *Accounting Horizons, 13*(3), 177-199.

Beaver, W., & Morse, D. (1978). What determines price-earnings ratios? *Financial Analysts Journal, 34*(4), 65-76.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research, 13*(10), 281-305.

Bezdudnaya, A.G., Ksenofontova, T.Y., Razumovsky, V.M., Zinchik, N., & Iudin, D.S. (2018). Evaluation of youth competence in the field of sustainable development: Lifecycle approach. *Espacios, 39*(21), 5.

Bhojraj, S., & Lee, C.M. (2002). Who is my peer? A valuation-based approach to the selection of comparable firms. *Journal of accounting research, 40*(2), 407-439.

Boatsman, J.R., & Baskin, E.F. (1981). Asset valuation with incomplete markets. *Accounting Review, 56*(1), 38-53.

CatBoost (n.d.). CatBoost - state-of-the-art open-source gradient boosting library with categorical features support. https://catboost.ai/ (Accessed date: July 18, 2022).

Copeland, T., Koller, T., & Murrin, J. (2000). *Valuation: Measuring and managing the value of companies*. 3rd ed. New York, NY: John Wiley & Sons, Inc.

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset*. John Wiley & Sons.

DeAngelo, L.E. (1990). Equity valuation and corporate control. *The Accounting Review, 65*(1), 93-112.

Deng, M., Easton, P.D., & Yeo, J. (2012). Another look at equity and enterprise valuation based on multiples. Social Science Research Network. http://dx.doi.org/10.2139/ssrn.1462794

Dittmann, I., & Weiner, C. (2005). Selecting comparables for the valuation of European firms. Working Paper. Collaborative Research Center, Humboldt University, Berlin.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Foster, G. (1986). *Financial statement analysis*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation, 121*(2), 256-285.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics, 28*(2), 337-407.

Gilson, S.C., Hotchkiss, E.S., & Ruback, R.S. (2000). Valuation of bankrupt firms. *The Review of Financial Studies, 13*(1), 43-74.

Herrmann, V., & Richter, F. (2003). Pricing with performance-controlled multiples. *Schmalenbach Business Review, 55*(3), 194-219.

Jackson, M., Pippin, S., & Wong, J.A. (2013). Asset and business valuation in estate tax cases: The role of the courts. *The Journal of the American Taxation Association, 35*(2), 121-134.

Ketchen, D.J., & Shook, C.L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441-458.

Kim, M., & Ritter, J.R. (1999). Valuing ipos. *Journal of financial economics, 53*(3), 409-437.

Koller, T., Goedhart, M., & Wessels, D. (2005). The right role for multiples in valuation. *McKinsey on Finance, 15*, 7-11.

Ksenofontova, T.Y., Bezdudnaya, A.G., & Kadyrova, O.V. (2017). Basic problems of interregional differentiation in Russia and innovative and reproduction prerequisites to overcome them. *International Journal of Applied Business and Economic Research, 15*(8), 1-10.

Land, J., & Lang, M.H. (2002). Empirical evidence on the evolution of international earnings. *The Accounting Review, 77*(1), 115-133.

LeClair, M.S. (1990). Valuing the closely-held corporation: The validity and performance of established valuation procedures. *Accounting Horizons, 9*, 31-42.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Lee, C.M., Ma, P., & Wang, C.C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics, 116*(2), 410-431.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research, 40*(1), 135-172.

Liu, J., Nissim, D., & Thomas, J. (2007). Is cash flow king in valuations? *Financial Analysts Journal, 63*(92), 56-68.

Palepu, K.G., Healy, P.M., & Peek, E. (2022). *Business analysis and valuation*. IFRS edition. Cengage Learning EMEA.

Penman, S.H. (1998). Combining earnings and book value in equity valuation. *Contemporary Accounting Research, 15*(3), 291-324.

Petersen, C.V., & Plenborg, T. (2012). *Financial statement analysis: Valuation, credit analysis and executive compensation*. Harlow: Pearson Longman.

Plenborg, T., & Pimentel, R.C. (2016). Best practices in applying multiples for valuation purposes. *The Journal of Private Equity, 19*(3), 55-64.

Schreiner, A., & Spremann, K. (2007). Multiples and their valuation accuracy in European equity markets. Social Science Research Network. https://doi.org/10.2139/ssrn.957352

SEC (n.d.) EDGAR—Search and Access. https://www.sec.gov/edgar/search-and-access/ (Accessed date: July 19, 2022).

Steinhaus, H. (1957). Sur la division des corps matériels en parties. *Bulletin L'Académie Polonaise des Science, 4*, 801-804.

Tarkhanova, N.P., Kosheleva, T.N., Vasilchikov, A.V., & Ksenofontova, T.Y. (2020). Public-private partnership as an instrument for regional entrepreneurial development. *Journal of Advanced Research in Dynamical and Control Systems, 12*(S1), 544-548.
Yoo, Y.K. (2006). The valuation accuracy of equity valuation using a combination of multiples. *Review of Accounting and Finance, 5*(2), 108-123.

Young, S., & Zeng, Y. (2015). Accounting comparability and the accuracy of peer-based valuation models. *The Accounting Review, 90*(6), 2571-2601.

Zarowin, P. (1990). What determines earnings-price ratios: Revisited. *Journal of Accounting, Auditing & Finance, 5*(3), 439-457.

Zura, K., & Yu, W. (2016). Statistical industry classification. *Journal of Risk & Control, 3*(1), 17-65.

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

APPENDIX 1.
Results For Alternative Clustering Methods

In addition to *k-means*, we also obtained results for more advanced clustering algorithms. Despite the complexity, alternative clustering algorithms did not improve the accuracy of a company valuation.

Table 5. The forecasting accuracy of multipliers for companies in the k-means clustering test sample (K=36)

Multiplier	Method for calculating the average value	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	Geometric	-8.23%	9.83
	Harmonic	-47.21%	11.81
	Median	-1.53%	9.62
	Simple	5.98%	10.38
$EV/Sales$	Geometric	4.12%	4.49
	Harmonic	-19.54%	5.08
	Median	6.35%	4.45
	Simple	18.18%	5.07
P/E	Geometric	-1.85%	12.24
	Harmonic	-27.89%	13.53
	Median	1.30%	12.12
	Simple	7.56%	12.74
$P/Sales$	Geometric	7.55%	4.00
	Harmonic	-8.70%	4.41
	Median	8.41%	3.97
	Simple	17.40%	4.54

Table 6. The forecasting accuracy of multipliers of test sample companies for the *Affinity propagation* clustering method (Tarkhanova et al., 2020)

Multiplier	Method for calculating the average value	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	Geometric	-5.79%	9.53
	Harmonic	-38.18%	11.03
	Median	-2.11%	9.34
	Simple	-2.11%	10.16
$EV/Sales$	Geometric	-2.18	4.66
	Harmonic	-24.95%	5.50
	Median	-1.06%	4.63
	Simple	11.06%	5.42
P/E	Geometric	-0.85%	12.13
	Harmonic	-25.40%	13.32
	Median	2.08%	11.97
	Simple	7.54%	12.63
$P/Sales$	Geometric	16.17%	3.84
	Harmonic	-2.50%	4.23
	Median	16.91%	3.80
	Simple	26.46%	4.30

Relações Internacionais do Mundo Atual Unicuritiba.
[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]

Table 7. The accuracy of forecasting multipliers of test sample companies for the *BIRCH* clustering method (Bezdudnaya et al., 2018)

Multiplier	Method for calculating the average value	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	Geometric	-18.96%	10.54
	Harmonic	-104.06%	16.22
	Median	-6.07%	10.18
	Simple	1.22%	10.86
$EV/Sales$	Geometric	-9.48%	4.97
	Harmonic	-32.17%	5.90
	Median	-9.60%	4.96
	Simple	5.39%	5.86
P/E	Geometric	-10.32%	13.28
	Harmonic	-44.06%	14.97
	Median	-10.00%	13.24
	Simple	0.03%	14.15
$P/Sales$	Geometric	-5.41%	4.41
	Harmonic	-19.11%	4.97
	Median	-5.16%	4.40
	Simple	4.15%	5.23

Table 8. The forecasting accuracy of multipliers of test sample companies for the *Meanshift* clustering method (Ksenofontova et al., 2017)

Multiplier	Method for calculating the average value	Average R^2_{oos}	Average MAE_{oos}
$EV/EBITDA$	Geometric	-16.09%	10.49
	Harmonic	-103.07%	16.15
	Median	-5.53%	10.13
	Simple	1.24%	10.84
$EV/Sales$	Geometric	-11.20%	4.96
	Harmonic	-31.39%	5.77
	Median	-11.22%	4.94
	Simple	1.82%	5.89
P/E	Geometric	-10.14%	13.25
	Harmonic	-45.13%	15.04
	Median	-9.98%	13.21
	Simple	-0.26%	14.10
$P/Sales$	Geometric	-8.14%	4.44
	Harmonic	-20.44%	4.96
	Median	-7.34%	4.43
	Simple	1.16%	5.35

[Received/Recebido: Julho 21, 2022; Accepted/Aceito Novembro 10, 2022]