www.cya.unam.mx/index.php/cya

Contaduría y Administración 70 (3), 2025, e516

Beyond OLS; Exploring regression methods in value relevance model

Más allá del OLS; explorando métodos de regresión en el modelo de value relevance

Jefferson Ricardo do Amaral Melo[1,2*], Sirlei Lemes[1], Neirilaine Silva de Almeida[1]

[1]Federal University of Uberlândia, Brazil
[2]Federal University of Delta do Parnaíba, Brazil

Received March 13, 2024; accepted July 31, 2024
Available online February 19, 2025

Abstract

This paper aims to investigate the performance and reliability of different regression methods in a value relevance model. Ten regression methods, both linear and nonlinear, were investigated to assess the effects of outliers, sample size, and overfitting on the model's performance (R^2 and error). The findings revealed that the Ordinary Least Squares (OLS) method is susceptible to false positive results due to its high R^2 but is often affected by heteroscedasticity, is more sensitive to outliers, sample size, and overfitting compared to the other regression methods analyzed. Although OLS is widely used in research in the field, it may not adequately address research questions in value relevance. The research advocates for a broader use of advanced regression techniques, including machine learning, to enhance empirical studies in finance. This research offers important insights for regulators, academia, companies, investors, and other stakeholders on the use of value relevance studies for decision-making.

JEL Code: G10, M41, C10
Keywords: regression; ordinary least squares; machine learning; value relevance; overfitting

[*]Corresponding author.
E-mail address: jeffersonramelo@ufdpar.edu.br (J. R. Amaral Melo).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

http://dx.doi.org/10.22201/fca.24488410e.2025.5513

Resumen

Este artículo tiene como objetivo investigar el rendimiento y la confiabilidad de diferentes métodos de regresión en un modelo de value relevance. Se investigaron diez métodos de regresión, incluyendo tanto lineales como no lineales, para evaluar los efectos de los outliers, el tamaño de la muestra y el sobreajuste en el rendimiento del modelo (R^2 y error). Los hallazgos revelaron que el método de Mínimos Cuadrados Ordinarios (MCO) es susceptible a resultados falsos positivos debido a su alto R^2, pero a menudo se ve afectado por la heterocedasticidad, es más sensible a los outliers, al tamaño de la muestra y al overfitting en comparación con los otros métodos de regresión analizados. Aunque el MCO se utiliza ampliamente en la investigación en el campo, puede que no aborde adecuadamente las preguntas de investigación en value relevance. La investigación aboga por un uso más amplio de técnicas avanzadas de regresión, incluido machine learning, para mejorar los estudios empíricos en finanzas. Esta investigación proporciona ideas importantes para reguladores, académicos, empresas, inversores y otros interesados respecto al uso de estudios de value relevance para la toma de decisiones.

Código JEL: G10, M41, C10
Palabras clave: regresión; mínimos cuadrados ordinarios; machine learning; value relevance; overfitting

Introduction

Research in the field of accounting often faces challenges related to the estimation of regression model parameters, especially when using the Ordinary Least Squares (OLS) method. These challenges become clearer when dealing with outliers, which tend to influence the results, and heteroscedasticity of errors, which compromise the efficiency of estimators (Ohlson & Kim, 2015; Kalantonis et al., 2022). Furthermore, OLS assumes a linear relationship between variables, a condition often absent in reality.

These methodological complexities highlight the importance of a careful approach in Value Relevance (VR) studies of accounting information. Value relevance is a necessary field of research to understand how accounting information influences the value of companies (Barth et al., 2001), having a direct impact on the establishment of accounting standards and for investors.

The direct impact of VR research on the formulation of standards is supported in classic works such as those by Barth et al. (2001), as well as in recent works by Hoang et al. (2022), Amel-Zadeh et al. (2023), and Koonce et al. (2023). However, an inappropriate methodological choice can lead to false positives (Ohlson, 2022) and misguided decision-making by regulators and companies, potentially increasing costs in issuing standards as well as for companies in complying with them (Duarte et al., 2017). Other research perspectives argue that VR studies provide insights for investors to make decisions based on corporate information (Veltri & Silvestri et al., 2020; Albuquerque et al., 2023, and Barth et al., 2023).

Given the significance of VR studies for standard setters and investors, academia has been striving to enhance scientific rigor. There is a growing concern about the reliability of the employed research methods, especially on the use of the OLS method in VR studies. Criticisms have been raised about its suitability, encouraging the adoption of more robust and reliable methods to improve the quality and reliability of research in the area. Studies like those by Ohlson and Kim (2015) argue that estimation by Theil-Sen is more reliable and robust compared to OLS. The authors verified the consistency of coefficients over time and their adaptability to non-ideal conditions (presence of outliers and heteroscedasticity). Kalantonis et al. (2022) compared the performance of VR models of Ohlson with OLS in relation to Weighted Least Squares (WLS) and Neural Networks Regression (NNR). The authors concluded that, unlike other methods, OLS does not adequately address heteroscedasticity, which compromises the accuracy of coefficients and, consequently, increases the variability of estimates, compromising the reliability of research results that use this method.

Duarte et al. (2017) showed that Quantile Regression (QR) was more efficient and had a lower possibility of estimation errors (QR is less sensitive to outliers and heteroscedasticity) than linear regression by OLS. The main argument used by the authors is that VR studies in emerging markets, like Brazil, are more suitable with non-parametric methods, such as quantile regression, due to the limited corporate diversity and heterogeneity among companies.

Feltes et al. (2021) used only quantile regression in their VR study, arguing that quantile regression provides greater efficiency and a lower probability of estimation errors compared to OLS. Another advantage pointed out by the authors is that quantile regression does not require tests for heteroscedasticity, normality of residuals, and multicollinearity, as this method is not based on the mean, like OLS, but rather on the method of minimizing absolute errors. Costa (2022) used Bayesian Linear Regression (BLR) and OLS in his VR study, concluding that BLR inference provided higher R^2 values, in contrast to the low values found by OLS.

More recently, Barth et al. (2023) used OLS and the CART (Classification and Regression Trees) algorithm in a VR study. According to the authors, the advantages of CART over OLS lie in the fact that it does not make any assumptions about the distribution of the data, which can be linear or nonlinear, and moreover, it deals better with a larger set of variables in the model.

The issues already resolved in the presented studies are that WLS and quantile regression show better performance of estimators than OLS in VR studies. However, the methods of WLS and quantile regression, despite being more robust methods than OLS, are not widespread in the data science field, as they do not have the generalization power like machine learning algorithms (Abdou & Nasereddin, 2011; Loterman et al., 2012; Fang & Taylor, 2021; Hanauer & Kalsbach, 2023). According to Ohlson (2022), traditional statistics were developed in times of low computational capacity and when data was not available in large quantities. Recent advances in machine learning, as well as improvements in computing tools (more accessible software, reduction of hardware costs, increase in speed/storage), allow researchers in the accounting field to easily apply and implement "Big Data" techniques to address potentially new and important research questions (Siano & Wysocki, 2021).

Hanauer and Kalsbach (2023) explored various machine learning techniques including gradient boosting, neural networks, and random forest to predict stock returns in emerging markets. The study documented that machine learning methods proved to be economically and statistically superior to traditional linear models, in terms of generalization power, in the absence of data linearity.

This trend of arguing that some machine learning methods perform better than traditional statistics is documented in other works, such as those by Abdou and Nasereddin (2011), Loterman et al. (2012), Gu et al. (2020), Fang and Taylor (2021), Costa (2022), Chen et al. (2023), and Barth et al. (2023). However, the topic is still underexplored in terms of comparing the performance of these methods to answer research questions in VR.

Considering the presented scenario, this research aims to elucidate the following question: how can regression methods alternative to OLS be more suitable for value relevance studies? This study aims to assess, theoretically and empirically, various linear and nonlinear regression methods that can more reliably answer research questions in VR, considering data from the Brazilian capital market.

Hypothesis development

Value Relevance (VR) studies in accounting and finance represent an important research field to understand how accounting information influences company value and, thus, investigate how investors make decisions based on accounting information. One of the foundational milestones in this area is often attributed to Ball and Brown (1968), who used OLS to quantify the relationship between accounting information and the market value of company shares. This allowed for the empirical establishment of the relevance of accounting information to investors.

Since the seminal study by Ball and Brown (1968), most VR models have attempted to establish the relationship between accounting figures and company value using OLS. However, authors such as Ohlson and Kim (2015), Duarte et al. (2017), Feltes et al. (2021), and Ohlson (2022) point out limitations in this approach. They suggest that more robust regression techniques can infer the effect of accounting information on company value with greater reliability.

According to Ohlson (2022), the choice of OLS by researchers creates a chain problem that starts with the choice of method and extends to the process of submitting articles to journals. The author argues that it's unlikely for a reviewer's report to suggest the author use a different estimation technique than OLS, because all the literature referenced by the authors used this technique, giving no reason for the authors to do differently.

The debate on the reliability of VR studies has even encompassed works on the establishment of accounting standards. For Holthausen and Watts (2001), the inferences from these studies would not be sufficient for setting standards, as the literature on the topic only reports the existence of associations between accounting numbers and the value of shares, which have limited implications or consequences for the establishment of standards. Therefore, it would not be possible to make useful inferences from this type of study. In contrast, Barth et al. (2001) argued that VR research is designed to provide evidence for standard setters, as they can update their earlier beliefs about how accounting values are reflected in stock prices and, therefore, can be informative for their deliberations on accounting standards.

The debates between the works of Holthausen and Watts (2001) and Barth et al. (2001) show that this research is not only restricted to the academic environment and can have impacts on society in general. Thus, the need for methodological rigor in studies like VR cannot be denied.

In the search for statistical methods more robust than OLS, works such as those by Abdou and Nasereddin (2011), Loterman et al. (2012), Fang and Taylor (2021), Gu et al. (2020), Costa (2022), Chen et al. (2023), and Barth et al. (2023) defended the use of regression models alternative to OLS, robust regression, or regression with machine learning algorithms, which are more reliable than OLS.

Abdou and Nasereddin (2011) used Support Vector Regression (SVR), feedforward neural networks, and OLS in their study of hedge fund performance prediction. They argued that the absence of data normality and non-linearity, due to the high volatility of data characteristic of these funds' returns, could compromise the reliability of OLS. The results pointed to the superiority of SVM in terms of prediction accuracy compared to neural networks and OLS. According to the authors, SVM stands out for finding a global optimal point, while neural networks may find local optimal points, thanks to the Kernel function used in SVM.

Loterman et al. (2012), in studying credit risk prediction, employed one- and two-stage regression models, including OLS, beta regression, robust regression, ridge regression, spline regression, neural networks, and SVR. In total, 24 techniques, including individual techniques and their combinations, were used. They concluded that, given the data's tendency for non-linearity, SVM and neural networks surpassed traditional techniques. The two-stage models, which combined linear and non-linear approaches, also showed solid predictive power.

In the field of finance, Gu et al. (2020), Fang and Taylor (2021), and Chen et al. (2023) explored various machine learning techniques alongside traditional statistics for asset pricing. They justified these studies by showing that linear models have always been applied in this area, but the assumption of linearity

http://dx.doi.org/10.22201/fca.24488410e.2025.5513

does not stand for the true relationship between variables in this field of study, hence the need to use machine learning models to address the complexity of non-linear data.

Finally, Costa (2022) and Barth et al. (2023) used methods alternative to traditional statistics in their VR studies. Costa (2022) argued that BLR provided a better fit of models, as measured by R^2, compared to OLS. Barth et al. (2023), in turn, showed that CART is more efficient for non-linear data and when more variables are added to the model, compared to OLS. The evidence from these studies points to the direction that, under non-ideal data conditions, such as the presence of outliers, error heteroscedasticity, and non-linearity, the use of alternative techniques is more suitable than traditional statistics, represented by OLS, which supports the research hypothesis:

H1: Alternative methods to OLS are more reliable and perform better in a value relevance model.

Research design

The population consists of a total of 14,683 observations. After excluding companies with negative equity, companies in judicial recovery, and companies in the financial sector, which includes banks, financial intermediaries, insurance, and pension firms, as well as observations with missing values, the sample was composed of a total of 2,495 observations. The exclusion of companies with negative equity and those in recovery is justified by the distortion this could cause in the estimates, as they do not have profit recurrence, which would make it impossible to see the relationship with the company's market value. In turn, financial sector companies have their own accounting characteristics, where assets are liability guarantors, making this accounting standard different from other companies to be analyzed.

The data were extracted from the Economatica database, covering the period from 2010 to 2022. The start of the temporal cut coincides with the mandatory adoption of IFRS (International Financial Reporting Standards) in Brazil in 2010. According to Júnior et al. (2017), the adoption of IFRS had an impact on the relevance of accounting information in Brazil, an average increase of 61.92% in price variation when compared to the relevance of profits and equity before the adoption of IFRS. Thus, the beginning of the temporal cut is justified by the introduction of IFRS in 2010 and the possible impact of this on the relevance of accounting information in Brazil.

The econometric design follows one of the empirical and theoretical models used in the works of Ohlson and Kim (2015), who compared OLS with Theil-Sen and quantile regression and confirmed in Brazil by Duarte et al. (2017) when they related OLS to quantile regression.

$$MV_{t+1} = \alpha + \beta_1 E_t + \beta_2 NI_t + \varepsilon_t \quad (1)$$

Where: MV_{t+1} represents the market value at time t+1, which is 30 days after the annual financial statement delivery date, E_t represents equity at time t, NI_t is the net income at time t, and ε_t is the error. Initially, the models were estimated within the sample, with 13 cross-sections for each year, starting in 2010 and ending in 2022. The performance of the models estimated by OLS was compared with linear and nonlinear regression models discussed in the literature. The linear models include WLS (Loterman et al., 2012; Kalantonis et al., 2022) and BLR (Costa, 2022), and the nonlinear models used were Quantile Regression (Ohlson & Kim, 2015; Duarte et al., 2017; Feltes et al., 2021), CART (Barth et al., 2023), Neural Networks (Kalantonis et al., 2022; Hanauer & Kalsbach, 2023), Random Forest Regression (Hanauer & Kalsbach, 2023), Gradient Boosting Regression Trees (Hanauer & Kalsbach, 2023), and SVR (Loterman et al., 2012). The parameters of the methods are described in Table 1.

Table 1
Description of methods and hyperparameters used

Methods	Hyperparameters Used
Ordinary Least Squares with Variable Scaling (OLSS)	The data were normalized by the total assets at t-1. This scaling metric was used in the works of Ohlson and Kim (2015) and Duarte et al. (2017).
Weighted Least Squares (WLS)	The model used has robust weights based on the Huber loss function. This method was used in the works of Loterman et al. (2012) and Kalantonis et al. (2022).
Bayesian Linear Regression (BLR)	Calculates the posterior distribution of the coefficients, considering the prior distribution and the observed data. This approach helps to deal with uncertainty in the coefficients and provides a more robust estimate in scenarios where there is sparse data or multicollinearity (Costa, 2022).
Quantile Regression (QR)	Was estimated at the median, as per the works of Ohlson and Kim (2015) and Duarte et al. (2017).
Classification and Regression Tree (CART)	The hyperparameters of the method are following the work of Barth et al. (2023). A bagging fraction of 1.0 was employed, with 500 trees to ensure efficiency and at least 5 observations in each region, meaning a node will only be split if there are at least 5 samples in it.
Random Forest Regression (RFR) and Gradient Boosting Regression Trees (GBRT)	Both used 500 trees, with no defined depth. The superiority of this technique over traditional statistics was documented in the work of Hanauer and Kalsbach (2023).
Neural Network Regression (NNR)	The adopted parameters were a network architecture with 1 hidden layer with 30 neurons, making it a multilayer perceptron. The maximum number of iterations was 1,000 epochs, the activation function is ReLU (Rectified Linear Unit), and the learning rate is 0.001. The superiority of this technique over traditional statistics was documented in the work of Hanauer and Kalsbach (2023).
Support Vector Regression (SVR)	SVR is less sensitive to outliers than OLS, especially when non-linear kernels are used. The polynomial function was chosen for this work.

Source: Own elaboration.

http://dx.doi.org/10.22201/fca.24488410e.2025.5513

Except for OLS, WLS, and quantile regression, all other methods were performed using 10-fold cross-validation as they are machine learning methods. This involves dividing the dataset randomly into 10 different groups, where in each test round, one of these groups is used to test the model, while the other nine are used for training. This is repeated 10 times, so that each group is used as a test set once. The overall accuracy of the model is evaluated by calculating the average of the accuracies obtained in all 10 test rounds. This technique helps to estimate the model's performance in a more robust manner (Kuzey et al., 2014; Barth et al., 2023).

To ensure the robustness of the estimates obtained using the Ordinary Least Squares (OLS) method, an estimation-test-estimation process was implemented. All methods were subjected to sensitivity tests regarding outliers. Initially, all methods were estimated without outlier adjustments and then with their adjustment. To treat the outliers, 1% winsorization was adopted, using the same parameters as the research by Ohlson and Kim (2015), Duarte et al. (2017), and Barth et al. (2023). In addition to the influence of outliers on the performance of the methods, the presence of heteroscedasticity in them was also verified.

To address heteroscedasticity in the OLS, two approaches were adopted. First, the variables were normalized by scaling them to the total assets of t-1 (OLSS), following the procedures adopted by Ohlson and Kim (2015) and Duarte et al. (2017). Then, the WLS regression was applied, according to the methods used by Loterman et al. (2012) and Kalantonis et al. (2022). Heteroscedasticity was identified using White's test.

OLS also pointed to the non-normality of the residuals and did not show any serious problems with multicollinearity as verified by the VIF (Variance Inflation Factor) being less than 10. According to Zhang et al. (2022), if the VIF is greater than 10, it indicates that the model has serious multicollinearity issues and needs to be corrected.

Since OLS did not present problems with multicollinearity, it did not make sense to use regularization to correct these issues, such as Ridge regression, Lasso regression, or Elastic Net regression.

The performance of the models was measured by indicators common to all of them. It was concluded through the works of Ohlson and Kim (2015), Duarte et al. (2017), and Barth et al. (2023) that the indicators would be R^2 and Error. R^2 provides a measure of the quality of the fit of the methods, both in linear models and in nonlinear models. This performance measure shows how much a model explains the variation of the dependent variable, which can vary between 0 and 1, being a measure of the higher, the better. The error measure used was the median absolute error, calculated by the absolute difference between the actual and predicted values. Then, the median of these differences was calculated and divided by the average value of the response variable. This measure is represented by the percentage of the average

http://dx.doi.org/10.22201/fca.24488410e.2025.5513

value of y. The interpretation is that the smaller this measure, the smaller the error in relation to the actual values.

For sensitivity analysis, panel data was used instead of cross-sectional data, thus making it possible to evaluate whether the sample size impacts the performance of the models. The performance of the models was also checked within and outside the sample, to ascertain the models' generalization power and the presence of overfitting.

Results

Table 2 presents the results of the descriptive statistics. The observations total 2,495 companies. Through the net profit variable, it is noticeable that companies that reported losses were not segregated. A Notable aspect of Table 2 is the minimum, maximum, and quartile values of the variables, which, along with the standard deviation, show the dispersion of these data around the median and mean. Figure 1 provides a more detailed observation of these characteristics.

Table 2
Descriptive statistics

Statistics	Market Value	Equity	Net Profit	Total Assets
Mean	10 894 864	6 512 127	737 366	16 114 325
SD	35 533 261	26 262 430	6 038 277	64 002 439
Min	1 917	177	-44 212 187	30
25%	492 697	479 315	4 729	1 024 780
50%	2 383 101	1 341 127	100 526	3 301 169
75%	7 608 375	3 952 657	457 993	10 135 865
Max	559 359 926	387 329 000	188 328 000	987 419 000
Obs	2 495	2 495	2 495	2 495

Source: Own elaboration.

Figure 1 shows that the total asset variable exhibited the greatest dispersion around the median, followed by the market value. The figure allows the observation of the extreme values for these variables. According to Kalantonis et al. (2022), data not within the range [Q1 - 1.5 * IQR, Q3 + 1.5 * IQR], where Q1 is the first quartile, Q3 is the third quartile, and IQR is the interquartile range between Q1 and Q3, are considered outliers. To prevent outliers from influencing the results, the models were estimated with and without outlier treatment, and the influence of these values on model performance was seen.

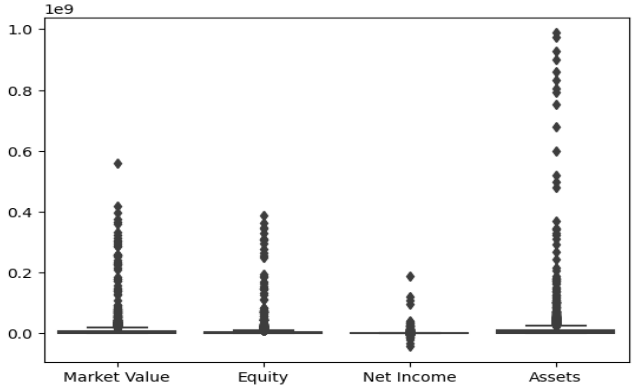

Figure 1. Box plot of the variables used.
Source: Own elaboration.

Table 3 presents the correlation between the model variables. The correlation allows us to understand the direction and strength of relationships between variables. To choose the proper type of correlation, a normality test of the data was first conducted. The Shapiro-Wilk and Kolmogorov-Smirnov tests indicated that the data do not follow a normal distribution. Therefore, Spearman's correlation was adopted, which is more suitable when the data do not have a normal distribution and the variables are continuous or discrete.

From Table 3, it is observable that market value has a positive, strong, and meaningful relationship with equity and total assets. This relationship is weaker with net profit. According to Kalantonis et al. (2022), when there are indications of a linear relationship between variables, as indicated in Table 3, linear regression models are appropriate. Otherwise, there would be no reason to use linear regression models.

Table 3
Spearman's Correlation

	Market Value	Equity	Net Profit
Equity	0.87*		
Net Profit	0.69*	0.61*	
Total Assets	0.82*	0.92*	0.53*

Note. *Statistical significance at the 5% level.
Source: Own elaboration.

Table 4 shows the results of the 13 cross-sections from 2010 to 2022, highlighting the performance of R^2 and the median absolute error of the 10 explored methods. In these models, outliers were not treated. Through OLS, it was possible to observe, not only from this table but in all of them, that heteroscedasticity identified by the White test, is a common problem. The presence of heteroscedasticity in the residuals can lead to various problems in the estimates, ranging from inefficiency of the coefficient estimates, inappropriate significance tests, distorted confidence intervals, and even bias in the estimates of standard errors, which invalidates their results. Heteroscedasticity means that the variance of the residuals is not constant in relation to the independent variables, violating one of the basic assumptions of OLS, which assumes homoscedasticity (constant variance of residuals).

As already mentioned, to correct the problems with heteroscedasticity in the model that adopts OLS, the models were re-estimated using the total assets of t-1 to scale the variables (OLSS). This procedure did not solve the heteroscedasticity issues. Therefore, to address the heteroscedasticity problems, Weighted Least Squares (WLS) regression was used. Seven more types of regression, both linear and nonlinear, were evaluated to find those that demonstrate better performance, higher R², and lower median absolute error, and that are robust to problems of heteroscedasticity.

Table 4
Performance of methods based on R² and median absolute error (without outlier adjustment) - cross-sectional data

Year/Method		Model: $MV_t = \alpha + \beta_1 E_t + \beta_2 NI_t + \varepsilon_t$									
		OLS	OLSS	WLS	QR	BLR	SVR	CART	RFR	GBR	NNR
2010	R^2	0.963	0.574	0.961	0.709	0.669	-0.215	0.489	0.523	0.472	0.649
	Erro	0.199	0.391	0.120	0.057	0.360	0.342	0.142	0.135	0.172	0.125
2011	R^2	0.809	0.473	0.805	0.548	0.339	-0.137	0.452	-3.020	0.575	0.429
	Erro	0.342	0.388	0.212	0.101	0.461	0.261	0.146	0.335	0.164	0.127
2012	R^2	0.807	0.262	0.790	0.510	-1.858	-0.086	0.579	0.588	0.513	0.264
	Erro	0.335	0.254	0.213	0.109	0.528	0.358	0.161	0.172	0.171	0.174
2013	R^2	0.636	0.892	0.623	0.418	-0.329	-0.086	0.434	0.429	-0.008	0.072
	Erro	0.407	0.373	0.222	0.113	0.692	0.373	0.205	0.215	0.253	0.129
2014	R^2	0.777	0.373	0.755	0.496	-0.143	-0.171	0.051	0.135	0.085	0.416
	Erro	0.174	0.470	0.134	0.089	0.259	0.228	0.149	0.149	0.165	0.194
2015	R^2	0.593	0.260	0.579	0.445	-0.262	-0.126	0.566	0.529	0.401	0.084
	Erro	0.194	0.534	0.122	0.097	0.339	0.321	0.205	0.196	0.225	0.231
2016	R^2	0.708	0.246	0.686	0.515	-0.339	-0.143	-0.427	-1.375	-0.911	0.020
	Erro	0.207	0.449	0.139	0.096	0.358	0.388	0.162	0.165	0.198	0.104
2017	R^2	0.752	0.329	0.747	0.590	0.513	-0.157	0.026	-0.250	-4.698	0.684
	Erro	0.154	0.422	0.121	0.078	0.282	0.385	0.148	0.143	0.160	0.159
2018	R^2	0.811	0.378	0.811	0.544	-0.583	-0.164	-1.679	-1.525	-3.010	-0.988
	Erro	0.254	0.428	0.150	0.065	0.369	0.309	0.205	0.178	0.180	0.092
2019	R^2	0.744	0.217	0.742	0.417	0.302	-0.183	0.194	0.215	-0.092	0.247
	Erro	0.407	0.506	0.240	0.070	0.503	0.271	0.202	0.189	0.169	0.128
2020	R^2	0.745	1.000	0.736	0.395	0.119	-0.151	0.324	0.379	-0.995	0.427
	Erro	0.304	0.026	0.157	0.086	0.396	0.273	0.196	0.197	0.205	0.123
2021	R^2	0.870	0.653	0.868	0.546	0.553	-0.180	0.245	0.188	-0.156	0.569
	Erro	0.319	0.411	0.183	0.081	0.420	0.270	0.157	0.161	0.175	0.117
2022	R^2	0.770	0.223	0.768	0.511	0.523	-0.125	0.130	0.390	0.071	0.523
	Erro	0.233	0.392	0.134	0.082	0.324	0.249	0.158	0.158	0.128	0.101

Source: Own elaboration.

http://dx.doi.org/10.22201/fca.24488410e.2025.5513

Table 5 shows the results of statistical inference with outlier treatment, for which 1% winsorization was used. It can be observed that even with outlier treatment, the problems associated with violations of the basic assumptions of OLS were not resolved. Regarding OLSS, in almost all the years there were problems with heteroscedasticity, except in 2019 and 2022.

Table 5
Performance of methods based on R² and median absolute error (with outlier adjustment) - cross-sectional data

Year/Method		Model: $MV_t = \alpha + \beta_1 E_t + \beta_2 NI_t + \varepsilon_t$									
		OLS	OLSS	WLS	QR	BLR	SVR	CART	RFR	GBR	NNR
2010	R^2	0.947	0.592	0.945	0.690	0.665	-0.227	0.503	0.490	0.215	0.354
	Erro	0.218	0.386	0.131	0.060	0.369	0.345	0.143	0.136	0.166	0.130
2011	R^2	0.790	0.489	0.784	0.581	0.577	-0.147	0.476	0.462	0.543	0.543
	Erro	0.388	0.380	0.234	0.053	0.505	0.263	0.146	0.141	0.183	0.183
2012	R^2	0.886	0.256	0.873	0.531	0.098	-0.133	0.579	0.609	0.504	0.309
	Erro	0.175	0.437	0.125	0.103	0.326	0.358	0.161	0.172	0.165	0.147
2013	R^2	0.895	0.735	0.888	0.546	0.374	-0.136	0.417	0.437	0.020	0.409
	Erro	0.147	0.399	0.124	0.125	0.234	0.374	0.207	0.214	0.225	0.185
2014	R^2	0.747	0.386	0.743	0.472	0.391	-0.184	-0.052	-0.125	-1.035	0.426
	Erro	0.162	0.452	0.119	0.102	0.226	0.231	0.155	0.156	0.185	0.161
2015	R^2	0.772	0.259	0.765	0.465	0.530	-0.137	0.568	0.508	0.312	0.492
	Erro	0.144	0.510	0.122	0.119	0.191	0.324	0.208	0.201	0.197	0.185
2016	R^2	0.783	0.250	0.780	0.512	0.266	-0.155	-1.193	-0.110	-1.926	0.020
	Erro	0.185	0.449	0.132	0.090	0.301	0.392	0.165	0.166	0.199	0.104
2017	R^2	0.802	0.332	0.792	0.5439	0.405	-0.161	0.132	0.033	-1.592	0.515
	Erro	0.128	0.418	0.113	0.100	0.243	0.387	0.149	0.144	0.144	0.188
2018	R^2	0.778	0.286	0.779	0.528	-0.704	-0.166	-1.274	-1.445	-2.959	-0.328
	Erro	0.222	0.430	0.134	0.082	0.313	0.311	0.207	0.181	0.192	0.094
2019	R^2	0.797	0.216	0.796	0.495	0.444	-0.188	0.212	0.226	-0.114	0.344
	Erro	0.253	0.500	0.141	0.086	0.332	0.271	0.202	0.191	0.191	0.118
2020	R^2	0.621	0.745	0.619	0.388	0.368	-0.158	0.383	0.407	-0.819	0.409
	Erro	0.234	0.522	0.146	0.097	0.304	0.278	0.201	0.202	0.214	0.145
2021	R^2	0.770	0.358	0.769	0.479	0.313	-0.194	0.286	0.205	-0.114	0.363
	Erro	0.109	0.393	0.104	0.080	0.137	0.275	0.161	0.168	0.173	0.136
2022	R^2	0.747	0.212	0.748	0.508	0.385	-0.138	0.208	0.292	0.340	0.477
	Erro	0.103	0.393	0.102	0.106	0.128	0.255	0.161	0.161	0.129	0.109

Source: Own elaboration.

Table 6 summarizes the results from Tables 4 and 5 through the ranking of the performance of the 10 methods. In this table, the average R² is placed in descending order and the average median absolute error in ascending order. The averages of these two metrics were calculated from the cross-section results from the years 2010 to 2022. To rank the models, scores from 0 to 10 were assigned, with 10 being the method with the highest R² or lowest median absolute error, and 0 the method with the lowest R² or highest median absolute error. For example, if a method has the highest R² (score 10) and the lowest error (score 10), the final score will be 10 + 10 = 20. Scores were assigned to all methods and then ranked in descending order of the sum of the scores.

http://dx.doi.org/10.22201/fca.24488410e.2025.5513

It can be noted (Table 6) that OLS is ranked among the methods with the four highest errors, even though it presents the highest R^2, with or without outlier adjustment, which points to a certain inconsistency in this method. Although OLS can have a high R^2, violations of assumptions may affect the validity of statistical inference. This means that while the model may fit well to the data, the coefficient estimates might be unstable, significance tests might be invalid, and predictions might not be reliable.

Additionally, Table 6 shows that QR, WLS, NNR, and CART are the regression methods that performed best. The treatment of outliers in general did not show significant changes in the ranking of the methods, except for OLS, which moved from 5th place to 3rd with outlier treatment. It is important to highlight that the outlier treatment improved the average R^2 of the 10 methods, which went from 0.181 to 0.276, and the median absolute error decreased from 0.233 to 0.211.

Table 6
Ranking of methods based on Average Performance of R^2 and error – 2010 to 2022 - cross-sectional data

Panel 1: Sample without outlier adjustment						
Score	Method Ranked	Average R^2 2010 a 2022	Method Ranked	Average Error 2010 a 2022	Final Ranked Method	Final Score
10	OLS	0.768	QR	0.086	QR	18
9	WLS	0.759	NNR	0.139	WLS	17
8	QR	0.511	WLS	0.165	NNR	15
7	OLSS	0.407	CART	0.172	OLS	14
6	NNR	0.261	GBRT	0.182	CART	12
5	CART	0.106	RFR	0.184	OLSS	8
4	BLR	-0.038	OLS	0.271	RFR	7
3	SVR	-0.148	SVR	0.310	GBRT	7
2	RFR	-0.215	BLR	0.407	BLR	6
1	GBRT	-0.596	OLSS	0.418	SVR	6
Mean		0.181		0.233		
Panel 2: Sample with outlier adjustment						
Score	Method Ranked	Average R^2 2010 a 2022	Method Ranked	Average Error 2010 a 2022	Final Ranked Method	Final Score
10	OLS	0.782	QR	0.095	QR	19
9	QR	0.778	WLS	0.133	WLS	16
8	OLSS	0.504	NNR	0.146	OLS	14
7	WLS	0.377	RFR	0.175	NNR	13
6	CART	0.332	CART	0.177	CART	12
5	NNR	0.287	GBRT	0.183	RFR	11
4	RFR	0.229	OLS	0.188	OLSS	9
3	SVR	0.125	BLR	0.270	GBRT	6
2	BLR	-0.158	SVR	0.310	BLR	5
1	GBRT	-0.487	OLSS	0.440	SVR	5
Mean		0.276		0.211		

Source: Own elaboration.

Sensitivity analysis

Sample size and performance

Methods such as neural networks involve learning a large number of parameters, which requires a considerable number of iterations to identify ideal values. In situations where the dataset is small, conducting a large number of iterations can lead to overfitting problems, resulting in a model with low interpretability and limitations in the ability to make predictions outside the scope of the available data (Kalantonis et al., 2022). The cross-sections from Tables 4 and 5 presented observations ranging from 169 to 246 individuals, which could impair the performance of the methods due to the limitation of sample size.

To observe the effect of sample size, the study's cross-sections were grouped into time intervals, thus testing the data in a pooled panel type. In the first analysis, three-time cuts were made, with four or five years in each cut, from 2010 to 2013, 2014 to 2017, and 2018 to 2022, each with observations of 706, 702, and 1087, respectively. In the second analysis, two cuts were made, establishing six or seven years in each cut, from 2010 to 2016 and 2017 to 2022, accounting for 1,229 and 1,266 observations, respectively, and finally, all the years were observed at once.

The established cuts were grouped in a pooled panel format (aggregation). This type of analysis does not seek to capture the individual effect as in fixed effects or consider the effects of individuals as random (random effects). Table 7 shows the results of these analyses without and with outlier adjustments, respectively.

Table 7
Performance of methods based on R² and median absolute error (with and without outlier adjustment) - pooled panel data

Method	Statistic	Panel 1: without outliers adjustment Model: $MV_t = \alpha + \beta_1 E_t + \beta_2 NI_t + \varepsilon_t$									
		OLS	OLSS	WLS	QR	BLR	SVR	CART	RFR	GBR	NNR
2010 a 2013	R^2	0.803	0.665	0.801	0.520	0.700	-0.043	0.824	0.828	0.828	0.724
	Error	0.330	0.399	0.197	0.095	0.370	0.228	0.100	0.101	0.105	0.095
2014 a 2017	R^2	0.656	0.287	0.654	0.477	0.086	-0.080	0.323	0.325	0.537	0.542
	Error	0.215	0.462	0.112	0.087	0.260	0.227	0.113	0.112	0.116	0.120
2018 a 2022	R^2	0.733	1.000	0.729	0.458	0.597	-0.052	0.570	0.551	0.473	0.545
	Error	0.314	0.146	0.165	0.083	0.343	0.224	0.140	0.144	0.116	0.094
2010 a 2016	R^2	0.678	0.598	0.675	0.483	0.616	-0.015	0.831	0.835	0.812	0.607
	Error	0.337	0.450	0.169	0.093	0.377	0.226	0.105	0.108	0.100	0.083
2017 a 2022	R^2	0.720	1.000	0.716	0.467	0.593	-0.041	0.480	0.445	0.424	0.549
	Error	0.312	0.163	0.164	0.079	0.329	0.221	0.137	0.141	0.125	0.089
All	R^2	0.673	1.000	0.669	0.445	0.569	0.058	0.643	0.640	0.559	0.471
	Error	0.369	0.283	0.182	0.086	0.384	0.179	0.097	0.098	0.099	0.078
		Panel 2: with outliers adjustment Model: $MV_t = \alpha + \beta_1 E_t + \beta_2 NI_t + \varepsilon_t$									
2010 a 2013	R^2	0.896	0.489	0.893	0.549	0.725	-0.067	0.844	0.844	0.838	0.768

	Error	0.152	0.392	0.137	0.103	0.160	0.240	0.106	0.108	0.111	0.115
2014 a 2017	R^2	0.785	0.289	0.782	0.508	0.605	-0.101	0.362	0.335	0.258	0.610
	Error	0.166	0.448	0.120	0.104	0.188	0.242	0.123	0.120	0.129	0.115
2018 a 2022	R^2	0.743	0.537	0.741	0.461	0.649	-0.089	0.605	0.590	0.543	0.653
	Error	0.165	0.557	0.113	0.094	0.172	0.237	0.145	0.150	0.129	0.106
2010 a 2016	R^2	0.838	0.433	0.835	0.539	0.747	-0.050	0.872	0.868	0.841	0.782
	Error	0.143	0.441	0.120	0.100	0.155	0.242	0.113	0.112	0.109	0.104
2017 a 2022	R^2	0.751	0.524	0.751	0.462	0.651	-0.086	0.572	0.542	0.351	0.629
	Error	0.150	0.560	0.110	0.095	0.164	0.237	0.142	0.144	0.126	0.116
All	R^2	0.731	0.482	0.728	0.441	0.661	-0.086	0.692	0.674	0.606	0.670
	Error	0.156	0.567	0.123	0.096	0.162	0.19	0.105	0.104	0.107	0.100

Source: Own elaboration.

Tables 8 and 9 summarize the results from Table 7. On the overall average performance of these methods, there were significant improvements in the models' performance in terms of R^2 and median absolute error. There is a clear trend between the increase in R^2 and the decrease in error with sample size. When the model was estimated with data ranging from 169 to 246 observations, the average R^2 performance was 0.181 (without outlier adjustment) (Table 8) and 0.276 (with outlier adjustment) (Table 9), with larger time cuts, the average R^2 increased to 0.535, 0.553, 0.571 without outlier adjustments (Table 8) and 0.554, 0.592, and 0.513 with outlier adjustments (Table 9).

The median absolute error decreased from 0.233 (without outlier adjustment) (Table 8) to 0.211 (with outlier adjustment) (Table 9) for samples ranging between 169 to 246 observations. When seeing larger samples, the errors were 0.187, 0.204, and 0.192 without outlier adjustment (Table 9) and 0.174, 0.174, and 0.181 with outlier adjustment (Table 9).

Table 8
Ranking of methods based on average performance of R^2 and Error (without outlier adjustment) - pooled panel data

Score	Method	Average R^2 2010/13 2014/17 2018/22	Method	Average Erro 2010/13 2014/17 2018/22	Method	Average R^2 2010/16 2017/22	Method	Average Erro 2010/16 2017/22	Method	Average R^2 All	Method	Average Erro All	Method Ranked Final	Final Score
10	1	0.731	4	0.088	1	0.699	4	0.086	1	0.673	10	0.078	CART	44
9	3	0.728	10	0.103	3	0.695	10	0.086	3	0.669	4	0.086	WLS	42
8	9	0.613	9	0.112	7	0.655	9	0.112	7	0.643	7	0.097	NNR	42
7	10	0.604	7	0.117	8	0.640	7	0.121	8	0.640	8	0.098	GBRT	4139
6	7	0.572	8	0.119	9	0.618	8	0.124	5	0.569	9	0.099	QR	38
5	8	0.568	3	0.158	5	0.604	3	0.166	9	0.559	6	0.179	RFR	38
4	4	0.485	6	0.226	2	0.598	6	0.223	10	0.471	3	0.182	OLS	36
3	2	0.476	1	0.286	10	0.578	1	0.324	4	0.445	1	0.369	BLR	17
2	5	0.461	5	0.324	4	0.475	5	0.353	6	0.058	5	0.384	SVR	16
1	6	-0.058	2	0.430	6	-0.028	2	0.450	2	*	2	*	OLSS	11
Mean		0.535		0.187		0.553		0.204		0.571		0.192		

Note. 1 – OLS, 2 – OLSS: 3 – WLS, 4 – QR, 5 – BLR, 6 – SVR, 7 – CART, 8 – RFR, 9 – GBRT e 10 – NNR. *Overfitting problems, R^2 equal to 1.000
Source: Own elaboration.

Based on the results presented in Tables 8 and 9, it can be concluded that the sample size and the adjustment of outliers influence the performance of regression models, whether linear or nonlinear. Larger sample sizes benefit the methods by capturing complex patterns in the relationships between variables and avoiding overfitting of the data.

Outliers are data points that can exert a disproportionate influence on the outcomes of regression analysis. This can lead to inaccurate coefficient estimates and significant deviations between observed and predicted values. Winsorization eliminates or drastically reduces the influence of these outliers, resulting in a model that better fits the rest of the data.

Table 9
Ranking of methods based on average performance of R² and Error (with outlier adjustment) - pooled panel data.

Score	Method	Average R^2 2010/13 2014/17 2018/22	Method	Average Error 2010/13 2014/17 2018/22	Method	Average R^2 2010/16 2017/22	Method	Average Error 2010/16 2017/22	Method	Average R^2 All	Method	Average Erro All	Method Ranked Final	Final Score
10	1	0.808	4	0.100	1	0.794	4	0.097	1	0.731	10	0.096	WLS	45
9	3	0.805	10	0.112	3	0.793	10	0.110	3	0.692	9	0.100	NNR	44
8	10	0.677	9	0.123	7	0.722	3	0.115	4	0.674	8	0.104	OLS	39
7	5	0.660	3	0.123	10	0.705	9	0.117	5	0.670	7	0.105	QR	38
6	7	0.604	7	0.124	8	0.705	7	0.127	6	0.661	6	0.107	CART	38
5	8	0.590	8	0.126	5	0.699	8	0.128	7	0.606	5	0.123	GBRT	34
4	9	0.546	1	0.161	9	0.596	1	0.146	2	0.482	4	0.156	RFR	32
3	4	0.506	5	0.173	4	0.500	5	0.159	8	0.441	3	0.162	BLR	29
2	2	0.438	6	0.239	2	0.478	6	0.239	9	0.018	2	0.567	SVR	15
1	6	-0.086	2	0.465	6	-0.068	2	0.500	10	-0.086	1	0.667	OLSS	12
Mean		0.554		0.174		0.592		0.174		0.513		0.181		

Note. 1 – OLS, 2 – OLSS, 3 – WLS, 4 – QR, 5 – BLR, 6 – SVR, 7 – CART, 8 – RFR, 9 – GBRT e 10 – NNR.
Source: Own elaboration.

It's worth noting that quantile regression loses its ranking when the sample size is changed; in small samples, this method was superior to others. When the sample size increases, it loses its position in the ranking, settling in 4th or 5th place. In larger samples, methods such as RNN and WLS have a better performance rate, higher R², and lower error, compared to quantile regression, when there is a change in sample size.

Overfitting analysis-in-sample and out-of-sample R

The performance of R^2 was checked by evaluating its in-sample and out-of-sample performance. This approach prevents overfitting (Barth et al., 2023) and allows evaluating the model's predictive ability (Ohlson & Kim, 2015). According to Barth et al. (2023), an out-of-sample R^2 that is lower than an in-sample R^2 indicates overfitting, i.e., the training data fits very well but lacks the capacity to predict new data.

To assess overfitting, 10-fold cross-validation was used instead of splitting the data into training and test sets. This approach is more robust and avoids selection bias. This method was utilized in the studies by Kuzey et al. (2014) and Barth et al. (2023). As the average performance of the models was better with outlier treatment, as revealed in Table 6, these models were estimated with outliers treated by 1% winsorization.

To proceed with the overfitting analysis, the difference between in-sample (IS) R^2 and out-of-sample (OS) R^2 was calculated. To rank the methods, the average of the differences between the R^2s from 2010 to 2022 was first calculated. After this procedure, the average was calculated, and all methods were ranked in ascending order, based on the average of the R^2 differences (IS and OS).

The results from Table 10 revealed that WLS and RNN were the models that performed best out-of-sample, with average differences of 0.422 and 0.475, respectively. OLS and CART follow, respectively, with 0.541 and 0.731. It is important to note that OLS and CART were among the top five methods in Table 6. However, the out-of-sample results may reveal problems with overfitting for more limited samples.

Table 10
In-sample (IS) and out-of-sample (OS) R^2 - cross-sectional data

Year/Method		Model: $MV_t = \alpha + \beta_1 E_t + \beta_2 NI_t + \varepsilon_t$								
		OLS	OLSS	WLS	BLR	SVR	CART	RFR	GBR	NNR
2010	IS	0.947	0.580	0.946	0.947	-0.043	0.951	0.965	0.999	0.927
	OS	0.667	-1.667	0.662	0.665	-0.227	0.503	0.49	0.215	0.354
	Dif	0.280	2.247	0.284	0.282	0.184	0.448	0.475	0.784	0.573
2011	IS	0.794	0.495	0.785	0.794	-0.047	0.954	0.957	0.999	0.780
	OS	0.557	0.331	0.388	0.577	-0.147	0.476	0.462	0.543	0.199
	Dif	0.237	0.164	0.397	0.217	0.100	0.478	0.495	0.456	0.581
2012	IS	0.889	0.277	0.876	0.889	-0.043	0.943	0.968	0.999	0.887
	OS	0.089	-0.363	0.417	0.098	-0.133	0.610	0.609	0.504	0.309
	Dif	0.800	0.64	0.459	0.791	0.090	0.333	0.359	0.495	0.578
2013	IS	0.896	0.724	0.888	0.896	-0.049	0.921	0.94	0.999	0.911
	OS	0.371	0.162	0.473	0.374	-0.136	0.417	0.437	0.020	0.409
	Dif	0.525	0.562	0.415	0.522	0.087	0.504	0.503	0.979	0.502
2014	IS	0.760	0.392	0.753	0.760	-0.069	0.83	0.918	0.999	0.788
	OS	0.380	0.234	0.449	0.391	-0.184	-0.052	-0.215	-1.035	0.426
	Dif	0.380	0.158	0.304	0.369	0.115	0.882	1.133	2.034	0.362

	IS	0.777	0.269	0.769	0.777	-0.074	0.862	0.915	0.973	0.794
2015	OS	0.527	0.145	0.570	0.530	-0.137	0.568	0.508	0.312	0.492
	Dif	0.250	0.124	0.199	0.247	0.063	0.294	0.407	0.661	0.302
	IS	0.801	0.261	0.794	0.801	-0.068	0.887	0.938	0.999	0.836
2016	OS	0.252	-0.138	0.394	0.266	-0.155	-0.193	-0.110	-1.926	0.301
	Dif	0.549	0.399	0.4	0.535	0.087	1.080	1.048	2.925	0.535
	IS	0.817	0.341	0.804	0.817	-0.054	0.894	0.938	0.999	0.813
2017	OS	0.390	0.239	0.554	0.405	-0.161	0.132	0.033	-1.592	0.515
	Dif	0.247	0.102	0.25	0.412	0.107	0.762	0.905	2.591	0.298
	IS	0.783	0.292	0.781	0.783	-0.065	0.886	0.933	0.998	0.785
2018	OS	-1.356	-0.022	-0.676	-0.704	-0.166	-1.274	-1.445	-2.959	-0.328
	Dif	2.139	0.314	1.457	1.487	0.101	2.160	2.378	3.957	1.113
	IS	0.798	0.228	0.794	0.798	-0.099	0.890	0.938	0.999	0.800
2019	OS	0.437	-0.752	0.462	0.444	-0.188	0.212	0.226	-0.114	0.344
	Dif	0.361	0.980	0.332	0.354	0.089	0.678	0.712	1.113	0.456
	IS	0.624	0.940	0.619	0.624	-0.100	0.919	0.945	0.999	0.611
2020	OS	0.361	-20.257	0.383	0.368	-0.158	0.383	0.407	-0.819	0.409
	Dif	0.263	21.197	0.236	0.256	0.058	0.536	0.538	1.818	0.202
	IS	0.768	0.365	0.766	0.768	-0.093	0.947	0.969	0.999	0.767
2021	OS	0.310	0.168	0.363	0.313	-0.194	0.286	0.205	-0.114	0.363
	Dif	0.458	0.197	0.403	0.455	0.101	0.661	0.764	1.113	0.404
	IS	0.750	0.221	0.749	0.75	-0.078	0.897	0.951	0.999	0.751
2022	OS	0.380	0.187	0.399	0.385	-0.138	0.208	0.292	0.34	0.477
	Dif	0.370	0.034	0.35	0.365	0.06	0.689	0.659	0.659	0.274

Source: Own elaboration.

To evaluate the effect of sample size, the data from Table 11 were placed in a pooled panel type, following the same procedure as the tests in Table 7.

Table 11
In-sample (IS) and out-of-sample (OS) R^2 - pooled panel data

Year/Method		Model: $\mathrm{MV_t} = \alpha + \beta_1 \mathrm{E_t} + \beta_2 \mathrm{NI_t} + \varepsilon_t$								
		OLS	OLSS	WLS	BLR	SVR	CART	RFR	GBR	NNR
2010	IS	0.897	0.489	0.894	0.897	-0.038	0.985	0.988	0.996	0.903
a 2013	OS	0.725	0.398	0.740	0.725	-0.067	0.844	0.844	0.838	0.768
(1)	Dif	0.172	0.091	0.154	0.172	0.029	0.141	0.144	0.158	0.135
2014	IS	0.785	0.319	0.782	0.785	-0.062	0.920	0.950	0.983	0.788
a 2017	OS	0.604	0.287	0.612	0.605	-0.101	0.362	0.335	0.258	0.610
(2)	Dif	0.181	0.032	0.170	0.180	0.039	0.558	0.615	0.725	0.178
2018	IS	0.742	0.320	0.741	0.742	-0.078	0.933	0.960	0.993	0.383
a 2022	OS	0.649	0.301	0.647	0.649	-0.089	0.605	0.590	0.543	0.345
(3)	Dif	0.093	0.019	0.094	0.093	0.011	0.328	0.370	0.450	0.038
Mean (1),	IS	0.808	0.376	0.805	0.808	-0.059	0.946	0.966	0.990	0.691
(2) e (3)	OS	0.659	0.328	0.666	0.659	-0.085	0.603	0.589	0.546	0.574
	Dif	0.148	0.047	0.139	0.148	0.026	0.342	0.376	0.444	0.117
2010	IS	0.839	0.358	0.835	0.839	-0.033	0.977	0.883	0.913	0.338
a 2016	OS	0.747	0.340	0.756	0.747	-0.050	0.872	0.220	0.112	0.310
(4)	Dif	0.092	0.018	0.079	0.092	0.017	0.105	0.663	0.801	0.028
2017	IS	0.751	0.307	0.750	0.751	-0.070	0.928	0.960	0.928	0.464
a 2022	OS	0.650	0.294	0.649	0.651	-0.086	0.572	0.542	0.267	0.429

(5)	Dif	0.101	0.013	0.101	0.100	0.016	0.356	0.418	0.661	0.035
Mean (4) e (5)	IS	0.799	0.347	0.796	0.799	-0.054	0.950	0.936	0.943	0.497
	OS	0.685	0.320	0.690	0.685	-0.073	0.682	0.450	0.308	0.437
	Dif	0.113	0.026	0.106	0.113	0.019	0.267	0.485	0.635	0.060
All (2010 a 2022)	IS	0.767	0.323	0.766	0.767	-0.029	0.939	0.883	0.895	0.325
	OS	0.745	0.301	0.745	0.745	-0.033	0.781	0.204	0.105	0.307
	Dif	0.022	0.022	0.021	0.022	0.004	0.158	0.679	0.790	0.018

Source: Own elaboration.

The results revealed that NNR and WLS rank highest when the sample size is increased in all configurations. The results from Tables 10 and 11 showed that these two methods consistently exhibited less overfitting regardless of the sample size when compared to other methods.

Discussion

The results of this study align with the research by Kalantonis et al. (2022) regarding the use of OLS, WLS, and ANN. The authors conducted a VR study with panel data from companies listed on the Athens stock exchange. Their research highlights the limited sample size, which can impact the performance of neural networks, as revealed in the results of this study. Furthermore, in line with this work, the authors argued that the most suitable method was Weighted Least Squares, but only when there is a linear relationship between the variables. According to the authors, this method provides analysts with precise indications of the effect of independent variables on the dependent variable, even with a limited sample.

The performance of quantile regression in this research also aligns with the work of Ohlson and Kim (2015). The authors found that median quantile regression had a better fit compared to OLS. They highlighted that quantile regression and OLS are indistinguishable under ideal conditions, but in the presence of outliers and homoscedastic errors, OLS performs poorly compared to quantile regression.

Regarding the performance of OLS compared to other regression methods, the results achieved here showed that OLS presented unreliable results, mainly due to problems with heteroscedasticity, outliers, high error, and overfitting. In line with these results, Ohlson and Kim (2015) argued that OLS would generally perform poorly compared to any reasonably robust estimation method.

It is important to highlight that in the findings of this research, OLS always showed the highest R^2 among the analyzed models. However, this result may indicate overfitting, as was shown in the results of this study. Barth et al. (2023) also observed this effect when they identified a high R^2 in in-sample estimates and a low R^2 in out-of-sample estimates, denoting an overfitting situation. According to the authors, this is where CART surpassed OLS. They compared the in-sample and out-of-sample R^2 of CART and OLS. The R^2 of CART exceeded the R^2 of OLS out-of-sample relative to in-sample in all cases, indicating the superiority of CART both in performance and in overcoming overfitting problems.

Barth et al. (2023) found that the use of CART as a non-parametric approach reduces the risk of underestimating explanatory power. This is because CART does not require specifying the functional form of the relationship, allowing the importance of accounting items to be more fully revealed. The authors add that flexible research estimation methods are becoming increasingly necessary to obtain reliable inferences about the link between accounting information and equity values.

From the results of this research and the results of earlier studies on the topic, the sample size influences the performance of models, especially if it is from a machine learning algorithm. Overfitting is a problem that must be seen since a high R^2 may mean excessive adjustment to training data but with low generalization power for out-of-sample data. Furthermore, nonlinear regression models are better suited to the data, as they do not assume data linearity like OLS.

Considerations

The main findings of the research revealed that despite OLS presenting a high R^2, it has a high error rate, is more sensitive to outliers, is recurrently affected by heteroscedasticity and overfitting, making the use of OLS less reliable and with a high probability of false positive results compared to alternative regression methods. Moreover, OLS demands a linear relationship between the model variables, a condition often absent. Alternative methods such as neural network regression, quantile regression, and WLS are more robust to outliers, are not affected by the heteroscedasticity problem, and are less subject to overfitting.

The research also reveals the necessity for researchers to initially verify which function best explains the relationship between the variables. If the relationships are nonlinear, nonlinear regression methods should be applied. If the relationships are linear, then one can continue with the use of linear or nonlinear regressions.

It should be noted that the sample size influences the performance of the methods, which is a frequent problem in data from emerging markets like Brazil. With restricted data, quantile regression performs well. However, as the sample size increases, methods like neural networks surpass this type of method. It is recommended to use panel data instead of cross-section to obtain more observations.

The choice of method depends on the type of VR study being researched. If the goal is to see the estimated coefficients, then White Box regression methods, where it is possible to verify the relationships between the variables, should be used. In this condition, if the relationship between the variables is linear, WLS can be used. For nonlinear relationships, quantile regression is recommended. If the goal is to evaluate only the R^2, black box models can also be used. It is recommended to adopt, in addition to the models already mentioned, neural networks, which showed satisfactory performance in the tests conducted.

OLS is an especially useful method and easy to interpret the coefficients under ideal conditions, such as the presence of linearity, absence of outliers, and without heteroscedasticity issues. However, these conditions are rare in real data. Under non-ideal conditions, the use of alternative regression methods, as already mentioned, is recommended.

The relevance of VR research transcends the academic sphere, directly influencing the formulation of standards and investor decisions. This reality amplifies the need for such studies to adopt rigorous and reliable research methods. A robust methodology is crucial to ensure that the answers provided to research questions are appropriate and beneficial for society as a whole. Thus, the need for continuous improvement in research approaches is emphasized.

For future VR research, it is recommended to always assess alternative regression methods and evaluate the methods' performance. If using OLS, another method should be used to compare the efficiency of the estimated coefficients. Another key factor is to check for overfitting issues, which seems to be a common problem that biases results in restricted data due to the low number of observations, as observed for data from the Brazilian capital market. Additionally, it is encouraged to replicate this study beyond the VR model.

The conclusions of this work are limited to the specific conditions of the Brazilian capital market data and value relevance model.

Acknowledgements

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support through grant number 88887.809291/2023-00, which made the development of this work possible.

Data availability

The study's data is available at jeffersonramelo/Paper-pos-doc: scripts, dados e outros materiais do pos doc (github.com)

References

Albuquerque, F., Velez, A. R., & Pinto, V. (2023). What is more relevant? A comparison of cash flows indicators versus profit or loss from listed European companies. Cogent Business & Management, 10(3), 2251214. https://doi.org/10.1080/23311975.2023.2251214

Abdou, K., & Nasereddin, M. (2011). The persistence of hedge fund strategies in different economic periods: A support vector machine approach. Journal of Derivatives & Hedge Funds, 17(1), 2–15. https://doi.org/10.1057/jdhf.2011.1

Amel-Zadeh, A., Glaum, M., & Sellhorn, T. (2023). Empirical goodwill research: insights, issues, and implications for standard setting and future research. European Accounting Review, 32(2), 415-4. https://doi.org/10.1080/09638180.2021.198385446

Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. Journal of Accounting Research, 6(2), 159. https://doi.org/10.2307/2490232

Barth, M. E., Beaver, W. H., & Landsman, W. R. (2001). The relevance of the value relevance literature for financial accounting standard setting: another view. Journal of Accounting and Economics, 31(1–3), 77–104. https://doi.org/10.1016/S0165-4101(01)00019-2

Barth, M. E., Li, K., & McClure, C. G. (2023). Evolution in value relevance of Accounting Information. The Accounting Review, 98(1), 1–28. https://doi.org/10.2308/TAR-2019-0521

Chen, L., Pelger, M., & Zhu, J. (2023). Deep learning in asset pricing. Management Science. https://doi.org/10.1287/mnsc.2023.4695

Costa, T. B. (2022). Evolução do value relevance: uma abordagem Bayesiana [Tese de doutorado]. Universidade Federal de Uberlândia. https://doi.org/10.14393/ufu.te.2022.185

Duarte, F. C. de L., Girão, L. F. de A. P., & Paulo, E. (2017). Avaliando Modelos Lineares de value relevance: Eles Captam o que Deveriam Captar? Revista de Administração Contemporânea, 21(spe), 110–134. https://doi.org/10.1590/1982-7849rac2017160202

Fang, M., & Taylor, S. (2021). A machine learning based asset pricing factor model comparison on anomaly portfolios. Economics Letters, 204. https://doi.org/10.1016/j.econlet.2021.109919

Feltes, T., Vicente, E. F. R., & Ribeiro, A. M. (2021). Relevância do goodwill e períodos de recessão: evidências no mercado de capitais brasileiro. Revista Ambiente Contábil – Universidade Federal Do Rio Grande Do Norte – ISSN 2176-9036, 13(2), 63–79. https://doi.org/10.21680/2176-9036.2021v13n2ID20386

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Hanauer, M. X., & Kalsbach, T. (2023). Machine learning and the cross-section of emerging market stock returns. Emerging Markets Review, 55, 101022. https://doi.org/10.1016/j.ememar.2023.101022

Hoang, K., Luo, Y., & Salterio, S. E. (2022). Evidence-Informed Audit Standard Setting: Exploring Evidence Use and Knowledge Transfer. Contemporary Accounting Research, 39(4), 2243-2283. https://doi.org/10.1111/1911-3846.12795

Júnior, A. C. B., Ott, E., Coelho, D. M., Menegat, L. A., & Webber, F. L. (2017). Persistência e value relevance dos resultados contábeis com a adoção das IFRS: um estudo em empresas de países do GLENIF. Enfoque: Reflexão Contábil, 36(3), 33-51. https://doi.org/10.4025/enfoque.v36i3.31820

Kalantonis, P., Delegkos, A. E., Sotirchou, E., & Papagrigoriou, A. (2022). Modern business development and financial reporting: exploring the effect of corporate governance on the value relevance of accounting information—evidence from the Greek listed firms. Operational Research, 22(3), 2879–2897. https://doi.org/10.1007/s12351-021-00637-2

Koonce, L., Mongold, C., Quaid, L., & White, B. J. (2023). Experimental research on standard-setting issues in financial reporting. Accounting, Organizations and Society, 101509. https://dx.doi.org/10.2139/ssrn.3893693

Kuzey, C., Uyar, A., & Delen, D. (2014). The impact of multinationality on firm value: A comparative analysis of machine learning techniques. Decision Support Systems, 59(1), 127–142. https://doi.org/10.1016/j.dss.2013.11.001

Leone, A. J., Minutti-Meza, M., & Wasley, C. E. (2019). Influential observations and inference in accounting research. The Accounting Review, 94(6), 337-364. https://doi.org/10.2308/accr-52396

Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. International Journal of Forecasting, 28(1), 161–170. https://doi.org/10.1016/j.ijforecast.2011.01.006

Ohlson, J. A. (2022). Researchers' data analysis choices: an excess of false positives? Review of Accounting Studies, 27(2), 649–667. https://dx.doi.org/10.2139/ssrn.3089571

Ohlson, J. A., & Kim, S. (2015). Linear valuation without OLS: the Theil-Sen estimation approach. Review of Accounting Studies, 20(1), 395–435. https://doi.org/10.1007/s11142-014-9300-0

Siano, F., & Wysocki, P. (2021). Transfer learning and textual analysis of accounting disclosures: Applying big data methods to small (er) datasets. Accounting Horizons, 35(3), 217-244. https://doi.org/10.2308/HORIZONS-19-161

Veltri, S., & Silvestri, A. (2020). The value relevance of corporate financial and nonfinancial information provided by the integrated report: A systematic review. Business Strategy and the Environment, 29(8), 3038-3054. https://doi.org/10.1002/bse.2556

Zhang, X., Wang, Z., Zhong, X., Yang, S., & Siddik, A. B. (2022). Do green banking activities improve the banks' environmental performance? The mediating effect of green financing. Sustainability, 14(2), 989. https://doi.org/10.3390/su151511717