Applied Machine Learning Algorithms for Intangible Asset Value Relevance

Songan Huang *

Department of Management, London School of Economics, London, UK

* Corresponding Author Email: s.huang56@lse.ac.uk

Abstract. In the era of the knowledge economy, investment and management of intangible assets are gradually receiving attention from companies. Through Ohlson's residual value model, scholars are able to build a bridge between a company's accounting information and its operating conditions. Many studies have confirmed that intangible assets have a significant positive relationship with companies' market capitalization and profit. However, past research has generally used traditional econometric models to explore the relationship between intangible assets and corporate value and net profit, using the correlation between accounting information and market value as a logical conduction. Machine learning methods have been widely used in business and economics research in recent years. Machine learning's ability to capture information in high-dimensional, non-numerical data and increase credibility through model training has partly compensated for the shortcomings of traditional methods. This paper will use machine learning methods (e.g., neural networks, decision trees) to research intangible asset value relevance further.

Keywords: Machine Learning, intangible asset, value.

1. Introduction

In the last decade, China's economy has developed rapidly, gradually entering an era of technology- and knowledge-intensive economies. Against such a backdrop, investment in and management of intangible assets is gradually being taken seriously by companies. Although intangible assets do not take physical form, they can contribute to a company's development and bring good economic returns [1]. Current research on the value relevance of intangible assets has mainly used traditional econometric models to study the relationship between accounting information and corporate value, with less intangible asset research applying machine learning methods. As machine learning has been shown to capture information in non-numerical, unstructured, missing value-filled and non-categorical data, there is a superior performance over traditional modelling [2,3]. Applying machine learning methods to intangible asset research promises to remove the disadvantages of traditional modelling methods that are more restrictive to data types and models that do not evolve, thereby exploring more realistic and explanatory theoretical models and reducing bias in results due to unknown data types not being included in the study [4]. On the other hand, however, some scholars have pointed out that machines are not comparable to human analysts in areas that require specific expertise [5].

This paper aims to compare and contrast traditional and machine-learning approaches. The paper's first half will explore intangible assets' value relevance using traditional econometric modelling methods. The second half uses machine learning methods. Validation is carried out by adding polynomial features, using KNN algorithms, neural network algorithms and many other ways to extend the explanatory power of the traditional linear model of intangibles to the firm's value.

2. The definition of intangible asset and characteristics of listed companies

2.1. The definition of intangible asset

Intangible assets, including patents, trade mark rights and so on, are identifiable non-monetary assets that a company controls. Its key features include lower marginal costs, higher sunk costs, monopoly and uncertainty of the value.

2.1.1 Trends in intangible assets, market capitalisation and net profit

This paper collects 5069 companies (including one with the same name, excluded) still on the A-share at the beginning of 2022. The *ST and ST companies are excluded, and the missing values are treated. The total number of intangible assets of A-share companies and their company values for each year are calculated and plotted. The trend of intangible assets of A-share companies in relation to market capitalisation and earnings is obtained, as shown in figure 1. As can be seen from the above chart, the holding of intangible assets by A-share listed companies has increased significantly since 2017. This reflects, on the one hand, that listed companies are gradually improving their disclosure of assets; on the other hand, it also reflects that listed companies are paying more and more attention to the value of intangible assets and improving their creation and management of them [6]. For companies, the number of intangible assets held can send positive information to investors, including the number of patents and development expenditures disclosed in financial reports, which can signal to the outside world that they are "strong in soft power" and facilitate their financing. Moreover, thanks to the transformation and upgrading strategy under the national policy, companies in many industries, such as high-tech industries and productive services, have rapidly grown.

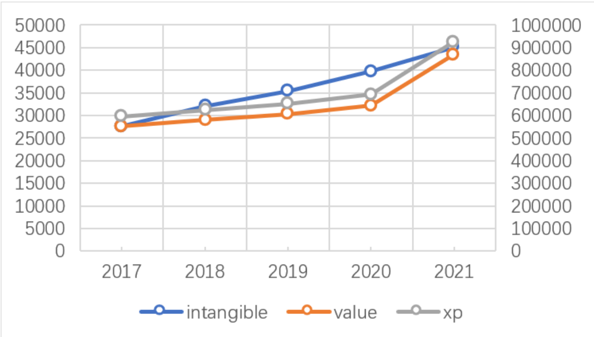

Fig 1. Intangible asset vs. value

Meanwhile, according to digure1, we can see that A-share enterprises' total intangible assets are broadly positively related to their market capitalisation and net profit. There is no doubt that intangible assets such as trademark rights and patented technologies are crucial to the development of enterprises, and an increase in the total amount of intangible assets is bound to positively impact the increase of market capitalisation and performance of enterprises [7]. We assume: There is a significant positive impact of intangible assets on the market capitalisation of A-share companies.

Table 1. Categories standard, companies

Industries	Number	Industries	Number
Power equipment	321	Machinery	503
Electronics	425	Pharmaceuticals and biology	462
Computer	328	Chemicals	378
Automotive	257		

Table 2. Number of companies average of intangible asset average of market value

Industries	number of companies	average of intangible asset		average of market value	
		2017	2021	2017	2021
Power equipment	321	0.0073	0.0946	101.3981	188.5004
Electronics	425	0.0100	0.0089	116.2764	187.7033
Computer	328	1.0987	1.7494	108.6254	129.0799
Automotive	257	5.3851	6.8206	148.4258	138.1768
Machinery	503	1.8799	2.2086	82.9714	76.8578
Pharmaceuticals and biology	462	1.8526	2.6789	129.2494	231.1603
Chemicals	378	2.2697	3.3544	80.2201	90.3094

Compared to physical assets, intangible assets such as patents and land use rights have been increasing their role in contributing to the development of enterprises. The more complex the business activities of an enterprise, the higher the proportion of intangible assets. The intangible assets of some high-tech enterprises are the core of that company's competitive strategy. And as mentioned earlier, listed companies are paying more and more attention to the creation and management of intangible assets. Thus, we assume that the contribution of intangible assets of A-share enterprises to their market capitalisation and operating performance is higher than that of tangible assets.

2.2. Industry classification of A-share companies

According to the SWS industries categories standard, companies are divided and shown in table 1. After excluding *ST and ST companies, the remaining 4900 companies belong to the following industries. As we can see from table1 and table2, A-share listed companies are mainly concentrated in seven major industries: Power equipment, Electronics, Computer, Automotive, Machinery, Pharmaceuticals and biology, Chemicals. In terms of growth rate, intangible assets vary from industry to industry, and the correspondence with enterprises' market capitalisation is unclear. Accordingly, we assume that the impact of intangible assets on the market capitalisation of A-share listed companies attributable to different industries varies and is significantly different (between industries).

3. Research methods

3.1. Ohlson model

In the late 1990s, Ohlson constructed the residual income model based on previous studies. The residual income is the accounting profit earned by the company minus the required return of the company. The value of the enterprise is the sum of all the required return and all the discounted residual income in the future.

The basic form of the Ohlson model is obtained as

$$P = y_0 + \sum_{t=1}^{\infty} \frac{x_t^a}{(1+r)^t} \tag{1}$$

The sum of the net book assets and the present value of future residual income is the value of the company's equity.

3.2. Model modification

This paper uses the company stock value model [8] as a basis with refer to the approach adopted by Ely & Waymire [9] and by Barth & Clinch [10]. Net assets (book value) are divided into intangible assets and tangible assets. The following model examines the relationship between intangible assets and share price.

$$P_{it} = \alpha_0 + \alpha_1 EPS_{it} + \alpha_2 TBV_{it} + \alpha_3 INTAN_{it} + \varepsilon_{it} \tag{2}$$

P_{it} is the closing price of the enterprise's share price on the last trading day in April of year t+1 (after compounding);EPS is Earnings per share for year t;TBV is the tangible net assets per share at the end of year t (tangible net assets = net assets - preferred stock - intangible assets);INTAN is the intangible asset per share at the end of year t; t denotes year; and i denotes company. Considering the difference in magnitudes, this paper multiply share price and EPS by the number of shares. The variables in the new equation are all company information disclosed in the financial statements and are equivalent in magnitude.

$$Value = \beta_0 + \beta_1 XP_{it} + \beta_2 OTAN_{it} + \beta_3 ITAN_{it} + \varepsilon_{it} \tag{3}$$

Value is company's market capitalization; $XP_{it}$is the net profit disclosed in the annual report of company i in year t; $OTAN_{it}$ net assets minus intangible assets of company i in year t, $ITAN_{it}$ is intangible asset of company i in year t.

4. Comparison and analysis

4.1. Traditional model result

Through Wind and CSMAR databases, this paper selects companies that are still on the market in 2021 and finds their relevant data from 2017 to 2021, including company value, net profit, total intangible assets, net intangible assets, etc. Excluding ST* companies, this paper obtains a total of 4,935 companies' data for five years. Missing values are excluded. Linear regression result of all the companies: Standard errors in parentheses. As shown in Table 3.

Table 3. p<0.01, ** p<0.05, * p<0.1

VARIABLES	value	t	P>| t |
XP	5.208***	48.77	0.000
	(0.107)		
OTAN	0.173***	14.26	0.000
	(0.012)		
INTAN	0.869***	18.13	0.000
	(0.048)		
Constant	92.819***	32.23	0.000
	(2.880)		
Root MSE	378.35		
R-squared	0.702		

As can be seen, the regression coefficients for intangible assets, tangible assets and net income are all significant (at the 1% level). Moreover, the regression coefficient of intangible assets is much larger than that of tangible assets, which indicates that the market prices the intangible assets per unit of a firm much more than the tangible assets per unit. This paper notes that the coefficient ratio of intangible assets to tangible assets is approximately 5, which is a little smaller compared to the previous results. As shown in Table 4. There are several reasons for this. A-share listed companies belong to more industry classifications, and many belong to industries with low technological content, such as power equipment and electronics. The presence of these companies will reduce the assessment of the relevance of the value of intangible assets for the whole company. The top 100 companies listed in A-share market capitalization in 2021 are mainly concentrated in the food, petrochemical, and banking industries, which have huge market capitalization but low intangible asset content, which will make the regression equation more biased.

Intangible assets have highly significant regression coefficients in all six industries. This result indicates that the effect of intangible assets on the return of companies in different industries is significantly present. Also, some industries have negative regression coefficients, and some have positive coefficients with different directions and magnitudes. As a result, the role of intangible assets in influencing the market capitalization of listed companies in different industries is different and significantly different. This paper argues that the reason for the difference in coefficients is a significant difference in the factor composition of the production function between industries with higher technology content and those with lower technology content. The market reacts differently to intangible assets in different industries. It is also possible that this is due to the lack of transparency in the disclosure of intangible assets in some industries, and the disclosure of intangible assets is inflated. Pharmaceuticals and biology, as a high technology industry, has a negative coefficient of intangible assets. The possible reason for this is that the pharmaceutical industry has seen a significant increase in product sales due to the epidemic, driving up company share prices and market capitalization. In a relatively short period (five years), investment in intangible assets and technology R&D did not increase significantly, but net profit increased significantly, and investors' expectations of pharmaceutical stocks rose, overestimating the relationship between net profit and enterprise value and underestimating the relationship between intangible assets and enterprise value.

For listed companies in the basic chemical, machinery and equipment, and automotive industries, intangible assets have a greater impact on the company's market value than tangible assets. As shown in Table 5. According to the finding, industries such as machinery and equipment and automotive, which traditionally relied on fixed assets to produce products and increase profits from sales, now rely more on intangible assets. For companies, technological innovation and patent rights significantly impact a company's enterprise value. Companies in these industries should increase their investment in R&D.

Table 4. VARIABLES, Pharmaceuticals and biology, Chemicals, Machinery

VARIABLES	Pharmaceuticals and biology	Chemicals	Machinery
XP	21.776***	0.927***	8.468***
	(20.34)	(5.62)	(22.36)
OTAN	1.376***	1.606***	0.351***
	(6.48)	(21.45)	(4.10)
INTAN	-6.927***	3.297***	3.515***
	(-5.27)	(8.02)	(6.00)
Constant	29.814***	6.764**	31.100***
	(3.83)	(2.46)	(15.30)
Root RSE	246.67	80.513	77.046
R-squared	0.498	0.648	0.805

Table 5. t-statistics in parentheses *** p<0.01, ** p<0.05, * p<0.1

VARIABLES	Automotive	Power equipment	Electronics	Computer
XP	0.283	11.833***	7.257***	15.855***
	(0.67)	(13.07)	(15.10)	(23.25)
OTAN	0.896***	2.774***	1.100***	2.613***
	(13.56)	(12.52)	(14.94)	(15.20)
INTAN	3.793***	-14.378***	-0.817	-4.838***
	(12.93)	(-7.81)	(-1.34)	(-3.67)
Constant	32.662***	4.335	71.640***	22.864***
	(5.72)	(0.58)	(12.95)	(4.57)
Root MSE	165.5	218.82	180.21	137.08
R-squared	0.747	0.614	0.647	0.786

4.2. Applied Machine Learning algorithms

This paper aims to compare the traditional method with machine learning methods. Hence, the data set is still the same. The company accounting data is divided, and the market value is set as the target value. The number of Cross-validation is set to 10.

linear regression: Firstly, ordinary linear regression is used. The figure 2 below shows the learning curve and the fitted image of the ordinary linear regression. In the learning curve, the test metric is the negative mean squared error. As the number of samples in the training set gradually increases, the training set mean squared error decreases and gradually converges to the cross-validation score. Therefore, this paper uses 60% of the original dataset as the training set and 40% as the test set. The model is fitted to the training set, and the dependent variable is predicted based on the independent variables given in the test set. The prediction curve is shown in figure 3. These are the predicted and true values for the first 50 x's. As can be seen, the linear model is a good fit. The relationship between the firm's intangible assets, net profit and the firm's market capitalisation is accurately predicted, reflecting the association between the firm's accounting information and the firm's market capitalisation.

Adding polynomial features: Subsequently, an attempt is made to add polynomial features to the model. Since intangible assets, tangible assets and net income capture changes in market value, the higher dimensional information may better impact the model results. Polynomial features of order 1 to 10 are added, and the mean squared error is recorded for both the training and validation sets. As shown in figure 4, the MSE of the training set keeps decreasing after adding high-dimensional polynomial features, but the MSE in the validation set keeps increasing, representing that the model is overfitted in the training set. On the one hand, this reflects the validity of corporate accounting information and the validity of the choice of variables in this model. On the other hand, it implies that at least for the three independent variables selected in this paper, their high dimensional information is detrimental to the estimation of firm value. The ratio of the variables or other variables should be included in the follow-up to increase the information.

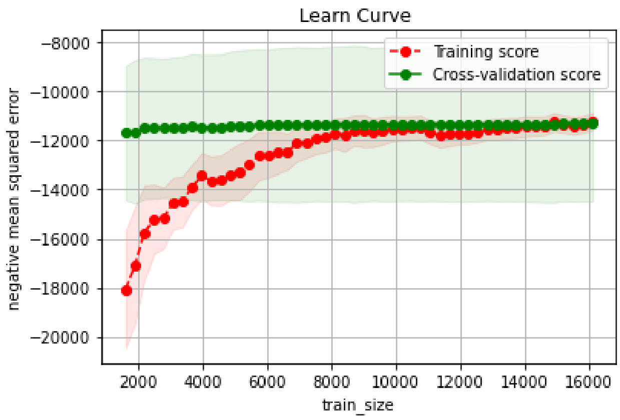

Fig 2. MSE with different train size

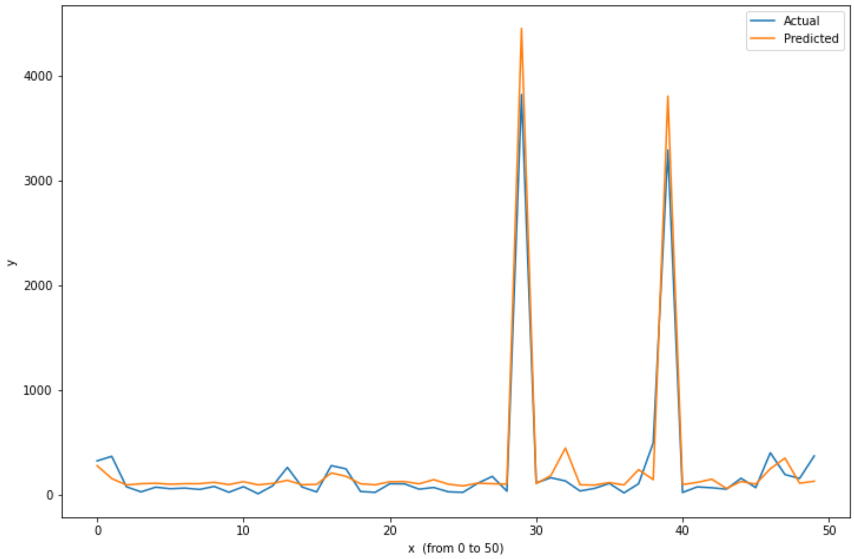

Fig 3. predict value vs. actual values

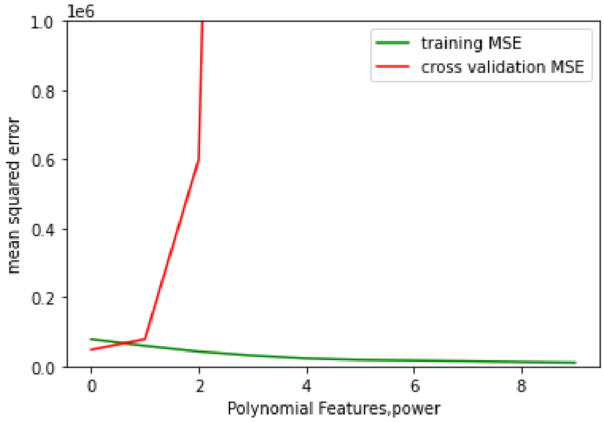

Fig 4. MSE for different polynomial

K-nearest neighbors algorithm (k-NN): The KNN model is actually a technical implementation of the common intuition that things with similar characteristics tend to be similar. It can be used in classification problems and regression problems. In regression problems, the algorithm derives a reasonable approximation to the output by counting the points near the chosen x-value at a distance. The algorithm's prediction of the model can be changed by setting the number of nearest neighbours. As with the classifier example, setting a higher value of k helps us to avoid overfitting, although the predictive power of the edges may start to be lost. Here k is set to 3, 5, 10, etc, and the average negative mean squared error is recorded. As can be seen, after setting k to 20, the model fit is actually relatively good. In practice, one can choose k=20, which gives an MSE of 10515, an improvement of about 10% over linear regression. KNN is easy to implement and handles non-linear problems well. Its problem is a lack of interpretability. because asking which features have the greatest effect does not really make sense. It is a trade-off between interpretability and model fit.

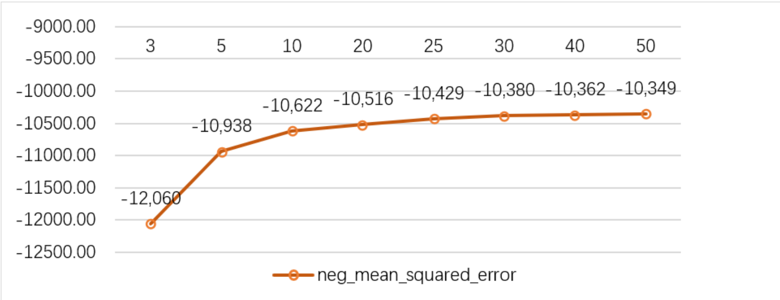

Fig 5. KNN algorithm

Neural Network: A neural network is a mathematical or computational model that mimics the structure and function of biological neural networks and is used to estimate or approximate functions. This paper uses fewer variables and therefore uses a neural network with one hidden layer. As shown in Figure 5. The layer type is dense, and the activation function is “relu”. Different numbers of neurons are set, and the letters indicate different combinations of neurons. The loss is visualized through the tensorboard. The loss metric is mean squared error. As shown in Table 6.

Table 6. Network parameters

Layer(type)	activation type	a	b	c	d	e
input layer (Dense)	relu	1	1	1	3	2
hidden layer (Dense)	relu	3	6	10	6	2
output layer (Dense)	relu	1	1	1	1	1

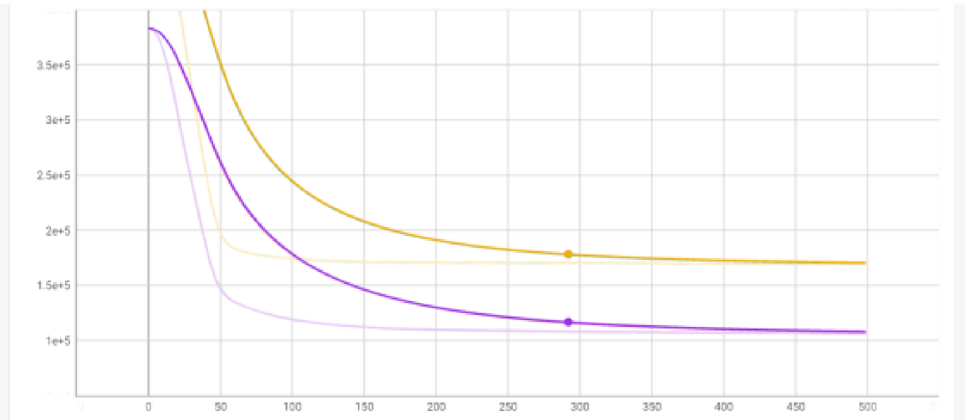

Fig 6. b vs. c

Figure 6 shows the loss curves for combinations b and c. As the number of training sessions increases, the loss decreases, and there is no significant difference between the two sets of curves. figure7 shows the comparison of combinations a and b. Combination a does not have a loss drop and is poorly trained. The input layer is adjusted, and combinations b and e in figure 8 are similar (rate of descent and final result). Similar results were obtained for combinations d and a. The above combinations imply that a higher number of neurons is not better and needs to be matched to the actual dataset. Secondly, the training results may not be good when the number of neurons varies too much from layer to layer.

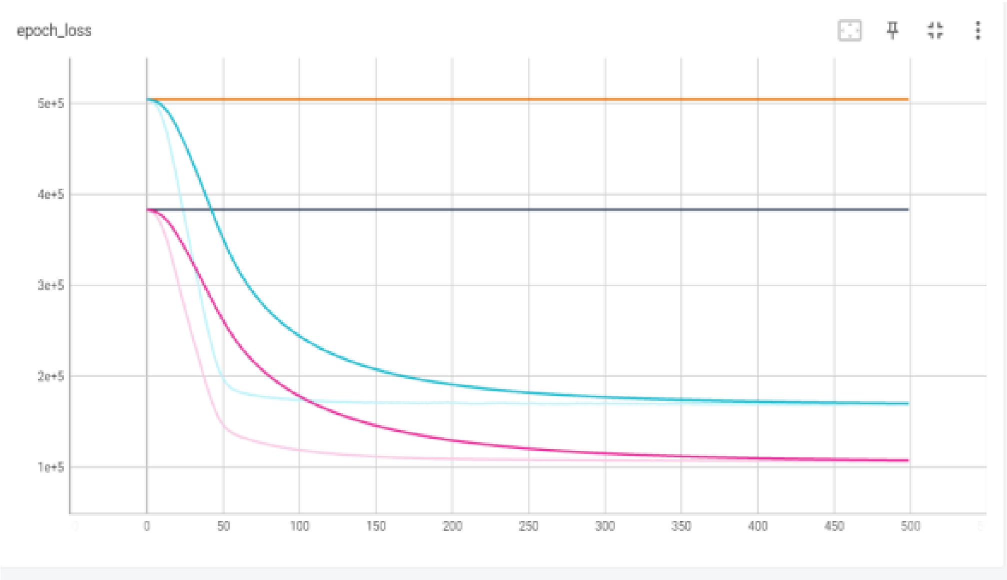

Fig 7. a vs. b

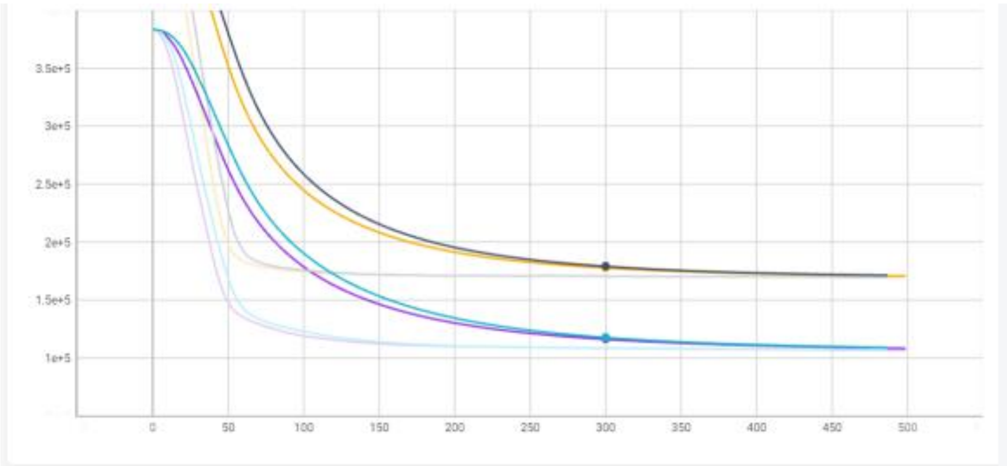

Fig 8. b vs. e

5. Conclusion

This paper uses traditional econometric models and several machine learning algorithms to investigate the relevance of intangibles to firm value. In the econometric model section, the paper confirms the significant impact of intangible assets on firm value and discusses the variability of the impact by industry. In the machine learning algorithms section, this paper demonstrates that the independent variables in the revised model already capture changes in firm value well and that high-dimensional information does not make the model any better. The KNN algorithms can reduce the mean squared error of the model, but this comes at the cost of interpretability. In fact, the results of the KNN and neural network algorithms cannot be intuitively perceived and depicted. For the model in this paper, the results of linear regression are undoubtedly better: the addition of only a small mean squared error greatly increases the interpretability compared to non-linear methods. This does not mean that the results of the machine learning algorithm are not superior. First, the model in this paper is about corporate accounting information and market value, and accounting information itself is associated with market value. In other words, this is the main area of application for linear regression.

It is already worthwhile to marvel at the superior performance of non-linear methods such as KNN. Secondly, the small number of independent variables selected in this paper and the more intuitive data are not conducive to the superiority of non-linear algorithms in terms of information capture. For example, this paper could increase the ratio between variables in the model and include other variables that have a relationship with the value of the firm.

References

[1] Oliveira L, Rodrigues L L, Craig R. Intangible assets and value relevance: Evidence from the Portuguese stock exchange[J]. The British Accounting Review, 2010, 42(4): 241-252.

[2] Duan Y, Goodell J W, Li H, et al. Assessing machine learning for forecasting economic risk: Evidence from an expanded Chinese financial information set[J]. Finance Research Letters, 2022, 46: 102273.

[3] Makarius E E, Mukherjee D, Fox J D, et al. Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization[J]. Journal of Business Research, 2020, 120: 262-273.

[4] Goodell J W, Kumar S, Lim W M, et al. Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis[J]. Journal of Behavioral and Experimental Finance, 2021, 32: 100577.

[5] Cao S, Jiang W, Wang J L, et al. From man vs. machine to man+ machine: The art and AI of stock analyses[R]. National Bureau of Economic Research, 2021.

[6] Nichita M E. Intangible assets–insights from a literature review[J]. Journal of Accounting and Management Information Systems, 2019, 18(2): 224-261.

[7] Ritter A, Wells P. Identifiable intangible asset disclosures, stock prices and future earnings[J]. Accounting & Finance, 2006, 46(5): 843-863.

[8] Olveira L, Rodrigues L, Craig R. Intangible assets and value relevance: Evidence from the Portugues stock exchange[J]. The British Accounting Review, 2010, 42(4): 241-252.

[9] Ely K, Waymire G. Intangible assets and stock prices in the pre-SEC era[J]. Journal of Accounting Research, 1999, 37: 17-44.

[10] Barth M E, Clinch G. Revalued financial, tangible, and intangible assets: Associations with share prices and non-market-based value estimates[J]. Journal of accounting research, 1998, 36: 199-233.