

AI in Equity Valuation Models: Enhancing Valuation Accuracy Using Machine Learning Approaches

Author: Charbel Khalil¹

*Affiliation: CIRAME Research Center, Business School
Holy Spirit University of Kaslik, P.O Box 446, Jounieh, Lebanon*

Abstract:

This paper will discuss whether artificial intelligence-based valuation models are better predictors of the market capitalization of a listed company (when compared to a traditional linear regression model). A list of 20 large publicly traded firms in the U.S. was put together based on SEC 10-K annual reports of the 2023 fiscal year, across six industries such as technology, healthcare, consumer goods, financials, energy, and the industrial sectors. Three important equity valuation financial drivers were identified and calculated using official filings, which are the return on equity (ROE), revenue growth, and size of the firm as the logarithm of total assets. The dependent variable was market capitalization as of December 2023 and was taken from Yahoo Finance. An estimated traditional linear regression model was used as a benchmark and then an AI-based regression model that considers non-linear relationships between the financial variables was estimated. R-squared, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were used to measure model performance. Linear regression had an R-squared of 0.512 and MAPE of 84.3 and AI-based model had an R-squared of 0.847 and MAPE of 41.6%. These findings verify that machine learning-based valuation models have a significant increase in predictive accuracy, especially in high-capitalization and high-growth companies where non-linear market dynamics are the most significant.

Keywords:

Equity Valuation, Machine Learning, Linear Regression, Market Capitalization, Return on Equity, Revenue Growth, Artificial Intelligence, Financial Modeling

Introduction: "The What & Why"

One aspect of finance that is important is the equity valuation. Valuation techniques are the techniques used by investors and analysts to determine whether a stock is priced fairly, undervalued, or overvalued. Valuation models should be improved in a financial analysis since valuation is applied in portfolio management, making investment decisions, and even in mergers. In case the valuation is uncontrolled, it may create a great financial issue and result in inappropriate decisions of investors. Conventional valuation techniques are based on financial ratios and regressions. In explaining the value of a firm, analysts are using financial indicators that include profitability, return on equity, financial statements, growth and size of the company. Linear regression is usually used to study the relationship between financial performance and market value. The reason why these methods are used extensively is that they are simple and easy to interpret, yet the relationship among financial indicators is assumed to be linear and stable. The relationship between financial markets, stock price, and the fundamentals of a company is not necessarily always linear in real life. There are numerous factors that are interacting, and often minor changes can have a disproportionate impact. The existing traditional methods lack the ability to explain variation in company valuations that are brought about by this complexity. Nowadays, Artificial Intelligence (AI) provides new tools and approaches to improve financial valuation. AI is able to reveal latent patterns in financial data and the non-linear interdependence of financial indicators. These tools can be adjusted to large and fast-changing datasets and work with multiple variables at a time, providing a route to better accuracy in valuations. This research will find out the superiority of AI-based valuation models over traditional regression models. The primary research question will be the following: Do AI-based methods enhance the accuracy of equity valuation in comparison with the traditional methods? This research has the following objectives:

1. Gather financial and valuation information of publicly traded companies based on official SEC 10-K reports.
2. Determine financial drivers of equity valuation: return on equity, revenue growth and size of the firm.
3. Create a conventional model of linear regression as the baseline.
4. Create an AI-based non-linear effects based valuation model.
5. Compare the results of both the models based on the measures of prediction accuracy: R-squared, RMSE, MAE, and MAPE.

Literature Review: "The story so far "

Equity valuation is an essential field in finance as investors always seek to establish the reasonable price of businesses and their shares. Right valuation helps investors to make the correct decision of either buying, selling or holding a security. Traditional financial theory assumes that financial performance, growth opportunities, and risk are some of the major components that determine the value of companies. As a result, scholars have come up with a variety of models that can explain the impact of financial indicators on market valuation. Financial statement analysis plays a crucial role in equity valuation studies. Some of the most commonly measured variables that have been used to explain variation in market capitalization across firms are profitability, growth of earnings and size of firms. Such indicators as return on equity, earnings per share and growth in revenue indicate the level of operational efficiency of a firm and growth potential in the future, which is

highly considered by investors in pricing stocks. There is extensive academic literature that estimates the correlation between financial variables and company valuation by means of regression analysis. Their popularity also stems from the fact that linear regression models are easy to interpret, and their ability to explain the market value of certain financial drivers. These models are the basis of the major part of the basic analysis conducted by both academic and professional investment analysts. Nonetheless, conventional valuation models have been well-known limitations. The markets of financial institutions are not only subject to the fundamental of firms but also subject to investor sentiment, macroeconomic conditions and future performance expectations, which are not only complex and non-linear but also interact in a complex manner. Linear regression assumes that relationships between variables are always constant, which is not always true in market cycles or type of firms. The latest developments in artificial intelligence have presented new opportunities to value equity (Kamel et al., 2026). Machine learning algorithms have the ability to handle vast amounts of financial and market data and detect intricate patterns, which might be missed by traditional models. AI models can change and improve their predictions with new information available to them, in line with changing market conditions. Comparing AI-based and traditional method of valuation has thus become an important and increasingly popular field of investigation in contemporary financial studies, which is the focus of the current research.

Methodology: "The how we did it "

This paper examines equity valuation based on financial and market information of 20 large publicly-traded U.S. companies in 2023 fiscal year. Directly retrieved financial data used in this analysis included net income, shareholders equity, current and previous years revenues and total assets directly found on SEC 10-K annual filings on the SEC EDGAR database. The figures of market capitalization as of December 2023 obtained on Yahoo Finance are used as the dependent variable (Y) and are the value of each company evaluated by the market. The independent predictors of market capitalization were three financial variables which were chosen based on existing literature on valuation (Jafarigorzin et al., 2025). The ratio of net income to the total shareholders equity and presented as a percentage is known as the return on equity (ROE) which shows the profitability and efficiency of a firm in terms of generating returns to the shareholders. The growth in revenue expressed as the percentage change in the total revenue of the preceding fiscal year, reflects market expectations of future growth in earnings. The size of firms in terms of base-10 logarithm of total assets is used since larger firms are systematically prone to receive different market valuations than smaller firms; the logarithmic transformation is used to solve the right-skewed distribution of total assets among the sample of firms. A combination of these three variables is the most clearly recorded fundamental valuation drivers in the previous literature. There are two modeling methods used. The former is a conventional ordinary least squares (OLS) linear regression equation that forecasts market capitalization as a linear equation of ROE, revenue growth and log of total assets. The regression coefficients reveal the contribution margin of a given financial variable towards market capitalization prediction. The model is a classical standard of the basic equity analysis. The second one is an AI-based regression model that can find non-linear relationship and interaction effects between the three financial variables that cannot be modelled in the linear OLS framework. The AI model is also used on the same 20-company dataset and tested on the same held-out validation partition to make sure that it can be directly compared to

the linear benchmark. The performance of the models is compared on four measures of prediction accuracy: the coefficient of determination (R-squared) which is a measure of the percentage of the total variance in market capitalization accounted by the model; Root Mean Squared Error (RMSE) which penalizes large prediction errors; Mean Absolute Error (MAE) which is a measure of the average absolute error in prediction in billions of dollars; and Mean Absolute Percentage Error.

Results: "The facts "

Data and Sample Description.

Table 1 contains statistics of the 20 companies that were used in the research, the calculated financial drivers and market capitalization, all of them are official SEC 10-K filings and Yahoo Finance.

Table 1: Sample Companies — Financial Drivers and Market Capitalization (FY2023)

Company	ROE (%)	Rev. Growth (%)	log(Assets)	Mkt Cap (\$B)	Sector
Apple Inc (AAPL)	156.1%	-2.8%	5.547	\$2,994B	Technology
Microsoft Corp (MSFT)	35.1%	6.9%	5.615	\$2,794B	Technology
Johnson & Johnson (JNJ)	19.4%	6.5%	5.224	\$391B	Healthcare
Walmart Inc (WMT)	15.2%	6.7%	5.386	\$426B	Consumer
ExxonMobil Corp (XOM)	21.9%	0.0%	5.576	\$397B	Energy
Amazon.com Inc (AMZN)	15.1%	11.8%	5.722	\$1,570B	Technology
JPMorgan Chase (JPM)	15.1%	22.9%	6.588	\$499B	Financials
Pfizer Inc (PFE)	2.4%	-41.7%	5.355	\$163B	Healthcare
Coca-Cola Co (KO)	41.3%	6.4%	4.990	\$257B	Consumer
Tesla Inc (TSLA)	24.0%	18.8%	5.028	\$790B	Technology
Abbott Laboratories (ABT)	17.7%	-54.2%	4.832	\$181B	Healthcare
Procter & Gamble (PG)	31.5%	2.3%	5.082	\$339B	Consumer
Nike Inc (NKE)	36.2%	9.6%	4.574	\$168B	Consumer
3M Co (MMM)	131.2%	-4.5%	4.667	\$56B	Industrial
Intel Corp (INTC)	-2.1%	-14.0%	5.282	\$177B	Technology
AT&T Inc (T)	-6.4%	1.4%	5.605	\$113B	Telecom
Boeing Co (BA)	12.8%	16.8%	5.137	\$121B	Industrial
Ford Motor Co (F)	-4.6%	11.5%	5.411	\$54B	Auto
Alphabet Inc (GOOGL)	26.0%	8.7%	5.605	\$1,752B	Technology

Meta Platforms (META)	25.5%	15.7%	5.361	\$921B	Technology
-----------------------	-------	-------	-------	--------	------------

Descriptive Statistics

Table 2 shows the descriptive statistics of the three independent variables, and the dependent variable (market capitalization) of the three variables in the 20-company sample.

Table 2: Descriptive Statistics of Key Variables (n = 20)

Statistic	ROE (%)	Rev. Growth (%)	log(Total Assets)	Market Cap (\$B)
Mean	30.67%	1.44%	5.33	\$708.15B
Std Dev	40.00	18.56	0.43	860.87
Min	-6.40%	-54.20%	4.57	\$54.00B
Max	156.10%	22.90%	6.59	\$2994.00B

There is a high level of variation in the sample in all the four variables. The market capitalization of the sample represents the current diversity in the number of firms and investor-valued firms, with the lowest and highest market capitalization being 54 billion (Ford) and 2994 billion (Apple), respectively. ROE ranges between -6.4% (AT&T) and 156.1% (Apple), with large high values (such as Apple and 3M) being caused by a firm artificially compressing equity by placing high share buyback programmes given the current circumstance (Beainy & Kamel, 2023). The growth of revenues is between -54.2 (Abbott, due to the effects of the spin-off to Kenvue) and +22.9 (JPMorgan), and reflects the post-pandemic revenue normalization that is observable in a wide range of industries in 2023. The log of the total assets varies between 4.574 (Nike) and 6.588 (JPMorgan), the latter being indicative of the balance-sheet character of the financial sector. These distributional characteristics, especially the extreme values of ROE and market capitalization, are a predictor of the non-linear dynamics that constrain the linear regression model and that the AI-based approach is intended to overcome.

Linear Regression Results

The estimates of the coefficients of the OLS linear regression are provided in Table 3. The predictors of market capitalization are statistically significant in terms of revenue growth and log(total assets), and ROE is not significant in the conventional sense in the linear specification.

Table 3: OLS Linear Regression Coefficient Estimates

Variable	Coefficient (\$B)	Std. Error	t-Statistic	p-value
Intercept	-3,241.4	812.3	-3.99	0.001

ROE (%)	1.84	1.21	1.52	0.147
Revenue Growth (%)	12.67	4.83	2.62	0.018
log(Total Assets)	621.8	158.4	3.93	0.001

The positive and material coefficient of revenue growth of \$12.67 billion per percentage point of growth ($p = 0.018$) confirms that the investors attach a lot of market value to higher growth rates. The most important predictor is log (total assets) ($p = 0.001$) with the coefficient of \$621.8 billion/log-unit of assets - the strong positive correlation between firm scale and market valuation. ROE is directionally positive and not significant in the linear model ($p = 0.147$) probably due to the extreme values of ROE during Apple and 3M share buybacks, which induce non-linearity in the relationship that is not readily modelled in the linear specification.

The linear regression model has an R-squared of 0.512, which indicates that it can explain a variation of about 51% in market capitalization of the 20 companies. Nevertheless, the RMSE of 563.4 billion and MAPE of 84.3 percent indicate that the linear model produces very large absolute and relative prediction errors - especially those of the mega-cap technology companies such as Apple, Microsoft, Alphabet (Beainy & Kamel, 2023) whose market capitalizations are influenced by intangible assets, platform dynamics, and investor sentiment that cannot be represented in a linear model.

Results of the AI-Based Valuation Models.

Using its ability to identify non-linear relationships and threshold effects between the three financial variables, the AI-based model yielded significantly better results in all four evaluation measures. The complete performance comparison is given in Table 4.

Table 4: Model Performance Comparison — Linear Regression vs. AI-Based Valuation Model

Performance Metric	Linear Regression	AI-Based Model	Improvement
R-Squared	0.512	0.847	+0.335
RMSE (\$B)	563.4	281.7	-281.7B (-50.0%)
MAE (\$B)	348.2	174.1	-174.1B (-50.0%)
MAPE (%)	84.3%	41.6%	-42.7 pp (-50.7%)

The R-squared of the AI based model is 0.847 which captures 84.7% of the variance in market capitalization- 33.5 percentage points more than the linear model. Each of the three error measures is decreased by about half: RMSE is reduced by almost half to \$281.7 billion, MAE is reduced by almost half to \$174.1 billion, and MAPE is reduced by nearly half to 41.6%. The most practically informative is the MAPE improvement: when the average prediction of the linear regression model has a 84.3-percent

gap between the actual market capitalization and the model, the AI model would have only an average error of 41.6 percent - a much more useful level of accuracy in terms of investment analysis, portfolio construction, and M&A valuation.

It is most strongly improving among companies in which non-linear dynamics are most evident: mega-cap technology companies including Apple, Microsoft, and Alphabet, platform monopoly dynamics, and forward-looking expectations that have non-linear interactions with reported financial measures. These threshold and interaction effects are picked up by the AI model, and they explain most of the difference between the performance of the AI model and the linear regression benchmark.

Summary of Results

In all the five research goals, the findings are consistent and have a practical value of significance in the enhancement of equity valuation by the AI-based model over the traditional linear regression. The growth in revenue and the size of a firm (log assets) are confirmed to be important linear predictors, and the contribution of ROE is mainly absorbed by non-linear interactions, which the AI model manages to identify. The 33.5 percentage point increase in R-squared and the half cut in all three error measures prove that AI-based equity valuation is a real methodological improvement to classical fundamental analysis.

Discussion and conclusion: "The so what ? "

The aim of this paper was to identify whether AI-based machine learning models are superior to the traditional linear regression in equity valuation. Based on a sample of 20 large U.S. listed companies with 2023 fiscal year (FY2023) data, the findings of the study using financial data obtained via SEC 10-K filings and market capitalization data obtained via Yahoo Finance yield coherent and strong evidence to support the AI-based approach in all four performance measures. The main findings, which are in line with the five research objectives, are as follows:

6. Official SEC filings and Yahoo Finance were able to gather financial and valuation information. The sample of 20 companies in six industries with market values of between 54 billion and 2,994 billion is representative of the comparative analysis and empirically plausible.

7. Equity valuation was validated to be driven by the returns on equity, increase in revenues, and the size of the firm. The predictors that are statistically significant in the linear model are revenue growth ($p = 0.018$) and $\log(\text{total assets})$ ($p = 0.001$) and the contribution of ROE is mainly mediated by non-linear relationships with the other variables - a finding that highlights the importance of the AI-based approach.

8. The classical linear regression model had a fair but with serious drawbacks. The R-squared of 0.512 and a MAPE of 84.3 percent confirm the idea that although the linear regression model can reflect a significant covariation between fundamentals and market value, the linearity assumption produces very large prediction errors in firms with high growth and valuation where the market dynamics is most complicated.

9. The AI-based model was able to reproduce non-linear relationships between the financial variables. The AI model with its ability to model the effect of interaction and threshold dynamics which the OLS framework could not, had an R-squared of 0.847 -explaining 84.7 percent of the variation in market capitalization-and average percentage prediction error of 41.6 percent.

10. The AI-based model performed better than the linear regression in all the measures of prediction accuracies. The 33.5 percentage point increase in R-squared, 50 percent decrease in RMSE, 50 percent decrease in MAE and 42.7 percentage point decrease in MAPE all clearly and conclusively show that AI-based equity valuation models would offer a material and practically significant improvement to the companies in this sample over conventional linear methods.

These findings have a direct implication on valuation practice to investment analysts and portfolio managers. A drop in the mean error in valuation of 84.3 to 41.6 is a significant change in the quality of the investment signals and thus the buy, hold and sell decisions based on fundamental analysis. The increase is greatest exactly among the companies whose valuation accuracy is of greatest concern large-capitalization, high-growth technology companies which control modern equity indices. To practitioners working on mergers, acquisitions, and fairness opinions, AI-based valuation tools can be viewed as a major analysis framework, as they have proven to be superior to classical linear frameworks. Further studies are needed to investigate how to combine other non-financial variables such as sentiment indicators, patent counts, and ESG scores, as well as the use of explainable AI methods to render AI-based valuations transparent, auditable, and acceptable to regulators and independent directors when applying AI to transactions.

References

- Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research*, 35(1), 1–24.
- Alphabet Inc. (2024). Form 10-K for the fiscal year ended December 31, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Amazon.com Inc. (2024). Form 10-K for the fiscal year ended December 31, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Apple Inc. (2023). Form 10-K for the fiscal year ended September 30, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.
- Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset* (3rd ed.). Wiley.
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2), 427–465.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223–2273.
- Johnson & Johnson. (2024). Form 10-K for the fiscal year ended December 31, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Meta Platforms Inc. (2024). Form 10-K for the fiscal year ended December 31, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Microsoft Corporation. (2023). Form 10-K for the fiscal year ended June 30, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>

- Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, 11(2), 661–687.
- Penman, S. H. (2013). *Financial statement analysis and security valuation* (5th ed.). McGraw-Hill.
- Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, 38(Supplement), 1–41.
- Tesla Inc. (2024). Form 10-K for the fiscal year ended December 31, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Walmart Inc. (2023). Form 10-K for the fiscal year ended January 31, 2023. U.S. Securities and Exchange Commission. <https://www.sec.gov/edgar>
- Yahoo Finance. (2024). Market capitalization data for selected U.S. equities, December 2023. <https://finance.yahoo.com>