International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Predicting Market Capitalization of Large Global Companies using Ordinary Least Square, Feedforward and Bayesian Neural Network Models

Vasu Padmanabhan[1], Dr. S. Nirmala[2]

[1]Research Scholar, Department of Management, PSGRK College for Women, Bharathiar University, Coimbatore, India

[2]College Principal and Research Supervisor PSGRK College for Women, Bharathiar University, Coimbatore, India

Abstract: ***This study applies market factors, financial statement information, and economic parameters to predict market capitalization for large global companies listed in Forbes 2000 - 2019 across all sectors and 50+ countries.*** **Market capitalization** ***is a crucial financial metric that enables company valuation, financial resourcing, mergers and acquisitions, and benchmarking market value across companies. Along with revenue and earnings (net income), market capitalization provides a broader perspective on organizational financial performance and prospects. This prediction study employs multiple regression models, neural networks, and statistical validation techniques. Regression analysis identifies key factors influencing market capitalization at global and industry-country grouping levels. Furthermore, this research explores the impact of the COVID-19 pandemic on market capitalization predictability. The study applies a novel MAPE minimization technique to improve the prediction models' accuracy by selecting the most influential parameters. Our findings suggest that several financial variables, including dividends paid, investment cash flows, and equity, significantly contribute to market capitalization predictability by reducing auto-regressive models' MAPE by 11%. These findings can help financial managers in large companies optimize strategic actions.***

Keywords: Market Capitalization, Regression Models, Artificial Neural Networks, Bayesian Networks, Financial Forecasting

1. Introduction

The earnings, total revenue, and market value form the triangular aspect of the organization's financial performance. The earnings are commonly measured as net income or earnings per share. The market value is measured through price per share or market capitalization. In most valuation contexts, market capitalization is the leading metric. The size of a stock exchange is often gauged by the total market capitalization of the securities it lists. Market capitalization is also an economic indicator and shows the performance of how the top companies of any economy are valued and the dispersion of the values. For example, the US has a total market capitalization of 50.8 Trillion USD as of 1st Jan 2024, while the top 500 companies have a market capitalization of 42.1 Trillion USD, showing the importance of analysing the top companies for the predictability of financial performance. (US Stock Market Total Market Value (2024) | Siblis Research).

The market capitalization is a function of the number of shares outstanding and the price per share of a given company. The number of shares outstanding can undergo significant changes due to corporate actions such as share buybacks, the issuance of new shares, the distribution of bonus shares, stock splits, or mergers and acquisitions. However, these actions are much less frequent than the price volatility per share. Stock price changes reflect company fundamentals, including returns such as dividends and bonuses. The future performance expectations from internal and external factors also drive the changes in stock price. The technical analysis supports several methods to predict the stock performance. Profit maximization, portfolio diversification, and long-term investment strategies are some of the goals of stock price predictions.

On the other hand, market capitalization predictability supports many corporate and investment activities, including capital raising and allocation, mergers, and acquisitions. More importantly, the predictability of market capitalization enables companies to benchmark against peers and market indices, helping to evaluate the company's relative performance and derive strategic actions. In cross-national research, numerous studies have postulated a positive correlation between the evolution of stock markets and economic expansion. As quoted by Naresh Kumar (2011), Levine and Zervos discovered that the advancement of banking institutions and stock markets significantly contributes to economic growth. Furthermore, he also quotes that Henry (2000) (cross-ref) indicated a positive correlation between market liberalization and private investment. Naresh Kumar's (2011) study also establishes the relationship between market capitalization and economic growth in the Indian Context.

Hence, it becomes evident that market capitalization is an essential financial performance measure for the market economy. This article adds to the scarce body of knowledge on the predictability of market capitalization on a global scale with the following Objectives.

Develop prediction models at a global scale with key sector and country groupings.

In recent years, the increasing globalization of large companies and the interconnectedness of market economies between countries have shifted the focus of financial performance research beyond traditional country or industry sector boundaries. However, collecting and standardizing data globally poses intricacies, particularly in dealing with various currencies and country groupings. This study examines companies from the Forbes 2000 list in over 50 countries to predict market capitalization trends on a global scale. Despite these challenges, this comprehensive approach offers valuable insights into market dynamics and enables more accurate predictions in today's interconnected world.

Identify the significant financial variables that contribute to the predictability of market capitalization.

In its accurate representation, market capitalization encapsulates a company's total value, significantly influenced by its equity as reported in the balance sheet and the profits it generates, as detailed in the Income and cash flow statements. Economic parameters such as changes in Gross Domestic Product (GDP), inflation rate, and Fama and French Factors can also impact market capitalization predictability. Given the potentially high number of variables involved, it is crucial to identify the subset that significantly influences predictability and enhances accuracy. To address this challenge, we develop a simplified methodology called MAPE minimization technique for variable selection. By applying this approach, we aim to optimize model performance and ensure accurate market capitalization predictions for large global companies across various sectors and countries.

Study the potential of Machine Learning Neural Network Models to improve prediction accuracy.

Ordinary Least Square (OLS) Regressions applied are linear. This study also analyses the accuracy improvements through Feedforward Artificial Neural Networks (ANN) and confidence intervals through Bayesian Neural Networks (BNN). These networks possess multiple hidden layers with nonlinear activation functions that enable them to learn complex patterns in data, allowing them to handle nonlinear regression tasks more effectively than Ordinary Least Squares (OLS) regression models.

Study the prediction capability of the models during the pandemic and post-pandemic.

The COVID-19 pandemic, which struck in 2020, had a profound and unique impact on the revenue and net income of companies across the globe. (Vasu and Nirmala 2023-1, 2023-2). The impact of this event on the developed market capitalization prediction models needs to be studied. Given that economic conditions evolve, models developed using pre-pandemic data may no longer be applicable due to slower underlying systemic changes in the market environment. Consequently, assessing the post-pandemic prediction capability of these models is essential to ensure accurate financial forecasting and adapt to the evolving market landscape.

2. Literature Review

The comprehensive literature review shows that research focusing on the predictability of market capitalization needs to catch up to studies investigating returns or earnings. This observation aligns with Murugesan et al.'s (2016) consolidated review in their study on Determinants of Firm Performance: A Subjective Model, where they identified profitability performance as the most researched determinant of financial performance. This research builds upon our previous studies detailing revenue predictability (Vasu and Nirmala, 2023-1) and net income (Vasu and Nirmala, 2023-2) across global companies.

Despite the complexities involved in stock market prediction (non-linear, dynamic, stochastic, and unpredictable characteristics), it is essential to establish whether there exists a correlation between market capitalization change and stock price change. Preliminary analysis reveals a modest direct linear correlation (R^2 = 0.85%), as seen in Figure 1. To comprehensively understand existing research in this area, we will examine both market capitalization and share price predictability. It is crucial to acknowledge that stock market prediction constitutes a challenging endeavour due to its intricate nature (non-linear, dynamic, stochastic, and unpredictable characteristics).

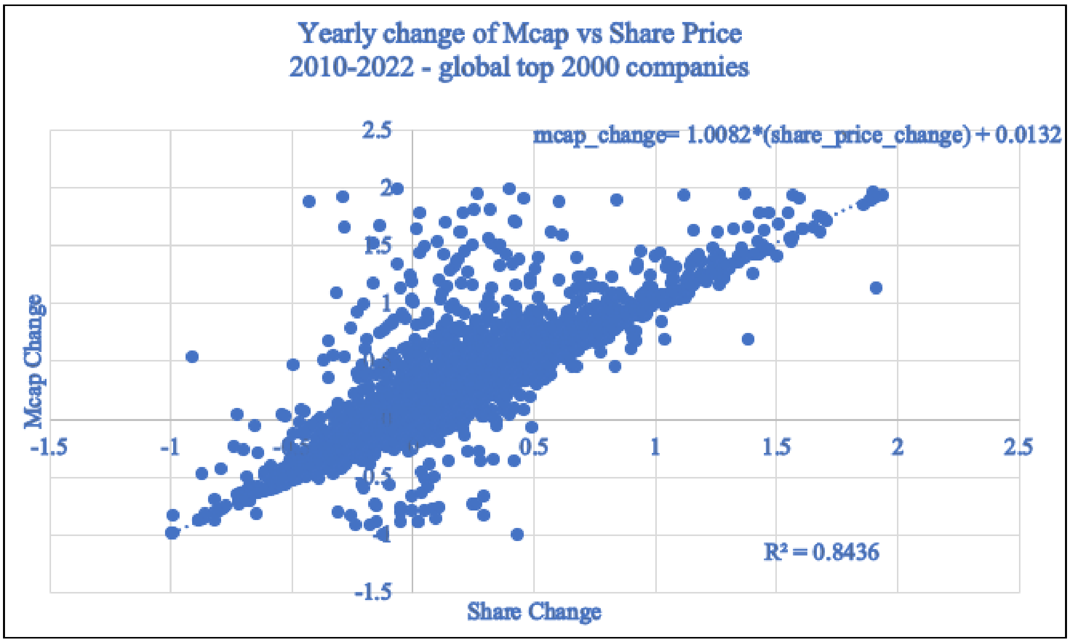

Figure 1: Market Capitalization Change versus Share Price Change

The key outcomes of our review are

Sock price predictions are commonly based on regressive models such as ARIMA (Autoregressive Integrated Moving Average), primarily designed to forecast short-term outcomes. These models are widely used due to their simplicity, interpretability, and ability to capture patterns in time series data

Nusrat Rouf et al.'s (2021) review shows that fundamental analysis, technical analysis, machine learning, and sentiment analysis approaches are applied in stock prediction research studies. In stock prediction methods, technical analysis represents a widely used approach for anticipating future price trends of equities. This approach examines historical stock price data through technical indicators to forecast future price movements. Technical indicators have gained widespread popularity in Stock Market Prediction (SMP) due to their ability to encapsulate trends effectively in time series data through their summative nature. Kapil Shirmal (2021) investigated the most effective ARIMA (Autoregressive Integrated Moving Average) model for predicting market capitalization using quarterly data from 21 Infrastructural Sector Companies listed in the S&P BSE-200 Index. ARIMA analysis was conducted in three categories: companies on an upward, linear, and downward trend. The selected models demonstrated adjusted R-squares ranging from 0.33-0.81, indicating substantial variation in predictability.

The recent advances in Artificial Intelligence have fuelled research to enhance stock prediction accuracy.

Ranjan Kumar et al. (2021) developed a finely tuned Support Vector Regression (SVR) model based on time series data for stock predictions. They employed grid search techniques to optimize model performance and choose the kernel function with its optimized parameters. Non-financial statement data, including up-to-daily and up-to-monthly returns, cumulative monthly returns, volatility, and associated risk, served as input variables in this study. By leveraging advanced machine learning techniques like SVR, researchers are pushing the boundaries of stock prediction accuracy, capturing intricate relationships within complex financial data.

Measures such as mean absolute percentage error (MAPE), R-square, and root mean square error (RMSE) are commonly employed to assess the accuracy of stock price or market capitalization prediction models.

MAPE provides a more straightforward and scaled accuracy measurement as demonstrated by Ranjan Kumar et al. (2021) in their application of both MAPE and RMSE as accuracy measures. Similarly, in Kapil Shirmal's (2021) study on ARIMA models for market capitalization prediction, models were selected based on criteria including Adjusted R-square, DW Statistic, RMSE, AIC, and SC, with the model having the smallest values in these criteria considered as the best fit.

Identifying and selecting the most significant variables to predict outcomes is often necessary because the system cannot accommodate extensive arrays of potential variables.

Researchers typically aim to examine as many input variables as possible based on their Pearson correlation or assumed causal relationships (Viktor, 2022). However, determining the optimal set of regressors that maximizes prediction accuracy necessitates generating an exponential number of trial models. For example, Viktor's study (2022) generated all 2k possible sets of k regressors and identified the best sets based on varying subset sizes ranging from 1 to 10 regressors. Some methods, such as Laaso and Ridge regressions, may help but have limitations.

The market value measured in market capitalization is researched in limited studies and outside the major developed countries.

Previous research on market capitalization predictability often relies on financial ratios (Viktor, 2022). These ratios are derived from fundamental drivers such as dividends. Researchers select independent variables based on their understanding of cause-and-effect relationships (Al-Afeef & Mohammad, 2020). For instance, Al-Afeef's study used factors like the Number of Transactions, Turnover Ratio, Earnings Per Share, Dividend Yield Ratio, Price Book value ratio, and Price Earnings Ratio. A regression study by Viktor (2022) to determine the influence Covid pandemic considered Net Profit Margin, Total Debt to Enterprise Value, Asset Turnover, Total Equity to Total Assets, Price to Book Value, Quick Ratio, Current Ratio, Interest Coverage Ratio, Gross Margin, Cash Operating to Total Assets, Cash Operating to Total Liabilities, Cash Investing to Total Assets, Cash Investing to Total Liabilities and Asset Turnover for prediction of market capitalization. Similarly, the study of Pavone (2019) demonstrates a positive relationship between market capitalization and Price/Earnings Ratio, Operating income/Turnover per share and Working Capital per Share and a negative relationship between market capitalization and ROE, ROA, and Earnings Yield for Italian companies.

Some of the research aimed to understand the coefficient of regression rather than predict with respect to time.

The prediction or forecasting models follow various techniques and timelines based on the author's logical reasoning. For example, Gregg, Ronnie, and Sheridan (2016) followed Fama and French (1992) (cross-ref) and formed all input variables at the end of June in year t, using fiscal year t-1 accounting information and analyst estimates from June of year t. For valuation ratios such as Price/Book Value, the authors used market equity from December of year t-1.

Valuation models, such as Asset-Based Valuation Models and the Gordon Growth Model (DDM), can also provide valuable input variables for market value estimation (Viktor, 2022). However, it is essential to note that these models are not predictive but rather contemporary regressive, meaning they calculate intrinsic values based on current or historical data rather than future forecasts.

The research studies we examined are specific to one geography.

Limitations such as data availability, data transformation complexities, and study context may restrict the research scope in market capitalization prediction studies. For instance, the study of Gregg, Ronnie, and Sheridan (2016) focused on NYSE, AMEX, and NASDAQ stocks listed in both the Centre for Research in Security Prices (CRSP) return files and the Compustat annual industrial files from 1982 through 2014

Paper ID: SR24908221149 DOI: https://dx.doi.org/10.21275/SR24908221149

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

The literature review thus clearly validates our study goals. Therefore, our study's critical components are anchored on the global scale, a large number of input financial parameters, and the development of multiple computer-based prediction models.

3. Materials and Methods

Our study involved an extensive analysis of financial data from 1953 large companies featured in Forbes' Global 2000 list between 2000 and 2019, where we could reliably fetch the financial data. We primarily sourced this data from company websites and the Morningstar Database, which provides standardized formats. Our predictive modelling was based on data from 2010 to 2016, while predictions for future performance were generated from 2017 to 2021.

While examining global trends offers significant insights, it's crucial to consider both the industry-specific nuances of individual companies and the capital market maturity of their respective host countries for a comprehensive understanding. The data depicted in Figure 2 undergoes further analysis to establish sector-capital market groupings, which we refer to as groupings in this article.

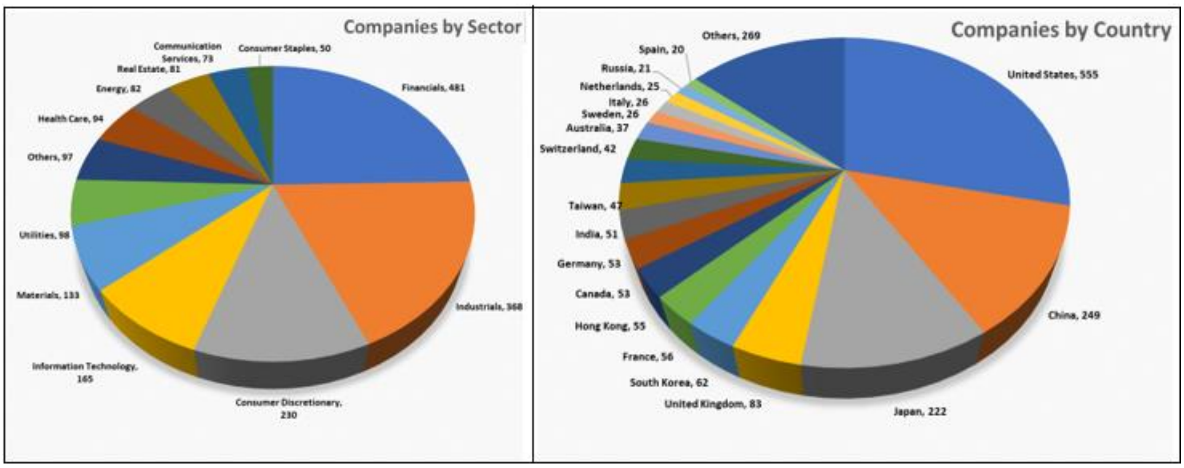

Figure 2: Distribution of selected companies across sectors and companies

For this research, we identified three major sectors subjected to group analysis: Financials, Consumer Discretionary, and Industrials.

- **Financials**: This sector encompasses institutions such as banks, insurance companies, real estate investment trusts (REITs), investment funds, and other financial organizations.
- **Consumer Discretionary**: This sector includes businesses that manufacture or offer non-essential goods or services. Examples include automobiles, hotels, luxury goods, and other consumer-oriented industries.
- **Industrials:** This sector comprises businesses like aerospace, defence, machinery, construction, engineering, transportation, and other industrial enterprises.

The other dimension for classification is the country grouping based on market maturity as given by Kenneth and French. The list is given in the table 1 below.

Table 1: Country grouping

Top Economy	Developed Countries	Developing Countries
United States	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, United Kingdom	Argentina, Bahrain, Brazil, Chile, China, Colombia, Egypt, Hungary, India, Indonesia, Israel, Kazakhstan, Kenya, Kuwait, Malaysia, Mexico, Morocco, Nigeria, Peru, Philippines, Poland, Qatar, Russia, Saudi Arabia, South Africa, South Korea, Taiwan, Thailand, Turkey, United Arab Emirates, Vietnam

The models are developed from 44 input variables chosen from income statements, cash flow statements, balance sheets, selected financial ratios, and market and economic factors. The same is listed in the table 2 below:

Table 2: Input variables.

Income Statements	Cash Flow and Balance sheet	Ratios and Financial Market and Economic Factors
Comprehensive Income/Losses, Cost of Goods Sold (COGS), Depreciation, Irregular Income/Expenses, Irregular income, Net Income, Net Income (Yes Flag), Net Property, Plant, and Equipment (PPE), Non-Interest Expenses, Operating Income, Provision and Impairment for Loan Losses and Credit Risk, Provision for Income Tax, Reserves/Accumulated, Retained Earnings/Accumulated Deficit, Selling, General and Administrative Expenses (SGA), Total Contractual Obligations	Cash Dividend, Cash Flow from Financing Activities, Cash Flow from Investing Activities, Cash and Cash Equivalents, Change in Cash, Change in Operating Capital, Debt Repayment, Inventories, Net Intangible Assets, Operating Cash Flow, Purchase/Sale of Investments, Total Assets, Total Current Assets, Total Current Liabilities, Total Equity, Total Liabilities	Current Ratio, Debt to Equity Ratio, Net Profit Margin, Price to Earnings Ratio, Return on Assets, Return on Equity Investment (Conservative Minus Aggressive – CMA), Market Risk (Market Factor), Momentum Factor, Profitability (Robust Minus Weak – RMW), Size (Small Minus Big - SMB), Value (High Minus Low – HML) Gross Domestic Product and Inflation

4. Models developed and applied in this research

The models utilized in this research can be categorized into two primary classes: Linear Models and Neural Network Models. Linear models employ a technique widely known as Ordinary Least Squares (OLS) regression to establish the relationship between various input variables and financial outcomes. This method aims to determine how these variables influence each other through a simple multi-variate linear equation. Moreover, OLS is a foundation for inferential statistics, such as hypothesis testing and constructing confidence intervals for the coefficients. These inferences provide valuable insights into the significance and reliability of the identified relationships between input variables and financial results.

AR Model:
This is a linear auto-regressive model that uses the market capitalization from the previous year as the sole input variable.

$$Y_t = \beta_0 + \beta_1 \cdot X_{t-1} + \varepsilon_t \quad \ldots\ldots\ldots\ldots (1)$$

ARX Model:
This is a linear auto-regressive model with external inputs, utilizing both the previous year's financial results and a set of variables from previous year as input variables.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 X_{1,t-1} + \beta_3 X_{2,t-1} + \cdots + \beta_n X_{n,t-1} + \epsilon_t \quad \ldots\ldots\ldots\ldots (2)$$

REG MODEL: This is a linear model that is not auto-regressive and relies solely on external inputs from a collection of variables from previous year.

$$Y_t = \alpha + \beta_1 X_{1,t-1} + \beta_2 X_{2,t-1} + \cdots + \beta_n X_{n,t-1} + \epsilon_t \quad \ldots\ldots\ldots\ldots (3)$$

- Y_t denotes the dependent variable's value at time t.
- α stands for the model's intercept
- $\beta_1, \beta_2, \ldots, \beta_n$ are the model's coefficients.
- Y_{t-1} indicates the dependent variable's value at time t−1, reflecting a one-year lag.
- $X_{1,t-1}, X_{2,t-1}, \ldots, X_{n,t-1}$ symbolizes the exogenous variables from the preceding year.
- ϵ_t represents the error term for time t

While Ordinary Least Squares (OLS) models capture linear connections between financial outcomes and input variables, neural network models can uncover and encapsulate nonlinear and segmented relationships through their weights and biases, possibly enhancing predictive accuracy.

AR_ANN: This is an auto-regressive model based on Artificial Neural Network Back Propagation (ANN-BP) that exclusively utilizes the financial outcome from the previous year as its input variable

$$y_t = f(\mathbf{w} \cdot y_{t-1} + b) \quad \ldots\ldots\ldots\ldots (4)$$

ARX_ANN: This is an auto-regressive model using Artificial Neural Network Auto Regressive Back Propagation (ANN-BP) that incorporates external inputs. It utilizes the financial results from the previous year and a collection of variables as its inputs.

$$y_t = f(\mathbf{w} \cdot y_{t-1} + \mathbf{V} \cdot \mathbf{X}_{t-1} + b) \quad \ldots\ldots\ldots\ldots (5)$$

REG_ANN: This is a non-auto regressive model based on Artificial Neural Network Auto Regressive Back Propagation (ANN-BP) that solely uses external inputs from a set of variables from previous years.

$$y_t = f(\mathbf{V} \cdot \mathbf{X}_{t-1} + b) \quad \ldots\ldots\ldots\ldots (6)$$

- y_{t-1} denotes the market capitalization value observed in the preceding year.
- $\mathbf{w}$ symbolizes the weight associated with the observation from the last year.
- $\mathbf{X}_{t-1}$ is the vector of exogenous predictors or variables at time t-1
- $\mathbf{V}$ represents the matrix of weights for the exogenous variables.
- b is the bias for each of the nodes
- $f(\cdot)$ is a nonlinear activation function, which could be sigmoid, ReLU, or tanh. In this study we used tanh as the activation function.

Neural network models operate with static weights for making predictions once the training phase is completed, whereas Bayesian models consider these weights as probabilistic variables governed by specific distributions. This approach enables the models to determine the standard deviation and overall distribution for each weight. This structure hence is able to predict the aveage and confidence intervals for each prediction point.

AR_BNN: This is a Bayesian Neural Network model with auto-regressive properties, utilizing solely the financial results from the previous year as its input variable.

$$\mu_t, \sigma_t = f(W \cdot Y_{t-1} + b) \quad \ldots\ldots\ldots\ldots (7)$$

ARX_BNN: This is an auto-regressive Bayesian Neural Network (ARX_BNN) model that incorporates external inputs, employing both the financial results from the prior year and additional parameters as input variables.

$$\mu_t, \sigma_t = f(W \cdot Y_{t-1} + V \cdot X_{t-1} + b) \quad \ldots\ldots (8)$$

REG_BNN: This is a Bayesian Neural Network (REG_BNN) model without auto-regression, utilizing only exogenous inputs from a collection of variables from previous years.

$$\mu_t, \sigma_t = f(W \cdot X_{t-1} + b) \quad \ldots\ldots\ldots\ldots (9)$$

Paper ID: SR24908221149 DOI: https://dx.doi.org/10.21275/SR24908221149

- μt is the forecasted mean, and σt is the forecasted standard deviation at time t. $\mathbf{Y}_{t-1}$ represents the t-1 value of the market capitalization
- $\mathbf{b}$ signifies the vector of biases, which are also considered as random variables.
- $\mathbf{X}_{t-1}$ is the vector of exogenous predictors or variables at time t-1
- $\mathbf{W}$ and $\mathbf{V}$ represent the weight matrix, with each individual weight w_{ij} being treated as a random variable *that* has its unique prior distribution.
- $f(\cdot)$ is a nonlinear activation function that is applied to the linear mix of weights and historical observations..
- ϵ_t is the error term, potentially modelled with a distribution to account for the data's noise.
- $\mathbf{X}_{t-1}$ indicates the vector of additional predictors or variables at time t-1

The Random Walk model posits that the current year's outcome serves as the foundation for predicting the following year's outcome, supplemented by a random factor. To accommodate this element of randomness, we incorporated random values within a standard deviation of the actual outcome. Nonetheless, this model consistently exhibited lower predictive accuracy compared to other models and will not be elaborately covered in this article. Our attention will instead be directed towards models that have shown greater efficacy in forecasting financial outcomes.

4.1 Accuracy Measurements:

From the literature review and our examination, we determine that Mean Absolute Percentage Error (MAPE) is the most suitable accuracy measurement for assessing predictability. MAPE assesses the prediction's error relative to the expected outcome's actual value. Furthermore, it offers a uniform benchmarking standard by scaling the measurement against the actual outcome. This measure enables us to compare models' predictive performance more effectively and consistently across financial data sets. The formula for MAPE is shown below:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{y_t - y_{t(Model)}}{y_t} \right| \ldots\ldots\ldots\ldots\ldots (10)$$

where $y_{t(Model)}$ is the predicted value of market capitalization of companies for a predicted year. y_t is the actual market capitalization.

In most places, we have also used percentages where the MAPE, as the above formula, is multiplied by 100.

In addition, we also used several other accuracy measures such as adjusted R-square, mean absolute error, and root mean square error.

In addition, we employed the Wald test to assess the equality of the slope and intercept between the actual and predicted outcomes of an Ordinary Least Squares (OLS) regression model. For a perfect fit, the slope should equal 1, signifying that each unit of the actual outcome corresponds to the same unit in the predicted outcome. Meanwhile, the intercept should be zero, indicating no constant error in the model when predicting financial outcomes from actual data.
We compute Wald Static as follows using Python scripts:

$$W = \frac{(\hat{\beta} - \beta_0)^2}{\text{Var}(\hat{\beta})}$$

-Where:

$\hat{\beta}$ is the estimated coefficient from the model.

β_0 is the hypothesized value of the coefficient under the null hypothesis (0 for the intercept and 1 for the slope).

$\text{Var}(\hat{\beta})$ is the variance of the estimated coefficient.

- The Wald statistic (W) follows a chi-square distribution with one degree of freedom under the null hypothesis, where the coefficient is equal to zero for the intercept and 1 for the slope
- If the calculated Wald statistic is greater than the critical value from the chi-square distribution, the null hypothesis is rejected, and the coefficient is not equal to zero for the intercept and 1 for the slope

4.2 Statistical tests:

The accuracy measures, excluding the Wald test, are absolute values used for model comparison based on their predictions' deviation from actual market capitalization values. However, these measures do not evaluate the existence of statistical differences in the mean, variance, or uniformity of underlying distributions between predicted and actual market capitalization values. To address this limitation, we employed statistical tests specifically designed for non-normal populations: the Mood Median test, the Kruskal-Wallis test, and the Kolmogorov-Smirnov (KS) test.

4.3 Methodology Summary

Figure 3, depicted below, provides a summary of the methodology, including the variables for input and output, data transformation processes, grouping of data, the predictive models that were developed and utilized, the statistical analyses conducted, and the metrics used to evaluate accuracy.

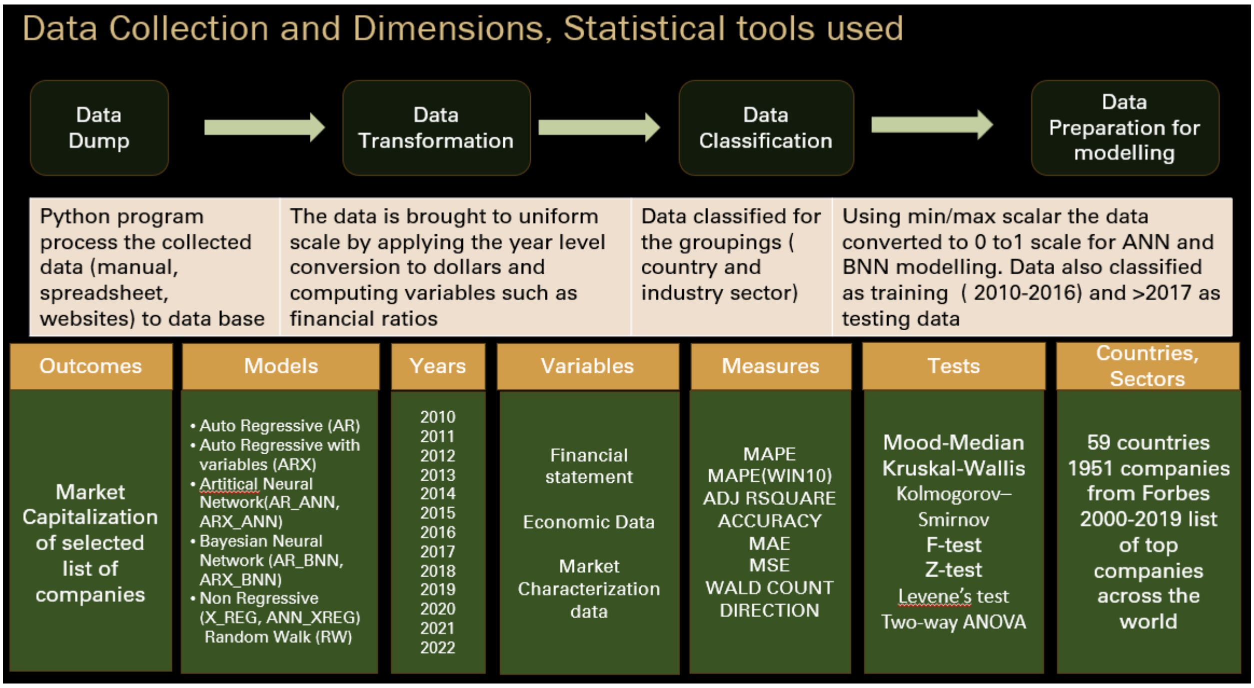

Figure 3: Data collection and Dimensions

The market capitalization data for the companies selected across the world shows a skewed distribution with a long tail. The Q-Q plot (year 2018) as seen in Figure 4 and the Anderson-Darling test show the data is not normal. Clearly 2021 Market capitalization has significantly higher than the other year ends. According to data from Factset as of Dec. 23, 2021, Apple Inc., Microsoft Corporation, Alphabet Inc., Amazon.com, Inc., Tesla, Inc., and Facebook, Inc., collectively added approximately $2.9 trillion to their market capitalizations. The US total market capitalization went up from 42 TUSD to 50 TUSD in 2021. The resurgence of optimism, primarily driven by the accelerated distribution of COVID-19 vaccines, played a pivotal role in revitalizing investor confidence. This renewed sense of hope was further bolstered by the phased reopening of numerous economies across the globe, which, in turn, catalyzed a revival in consumer demand. Concurrently, the persistence of a low-interest rate environment, established by central banks to mitigate the economic fallout of the pandemic, served to lower the cost of borrowing. This financial landscape not only facilitated increased spending and investment but also enhanced the attractiveness of equities relative to fixed-income securities. Additionally, the implementation of a substantial fiscal stimulus package in the United States acted as a critical economic lever, injecting liquidity and stability into the market. Collectively, these factors were instrumental in damping down market volatility and propelling stock markets to unprecedented peaks in 2021. (Source: FY 2021 Market Highlights v3.pdf (world-exchanges.org)). It is also critical to note that the economic environment changed significantly during 2022 and 2023 as the inflation rates started rising and GDP also slipped in many developed economies. The Market capitalization fell back to 40 TUSD in 2022. However, this aspect is out of our research scope window, which goes only till 2021.

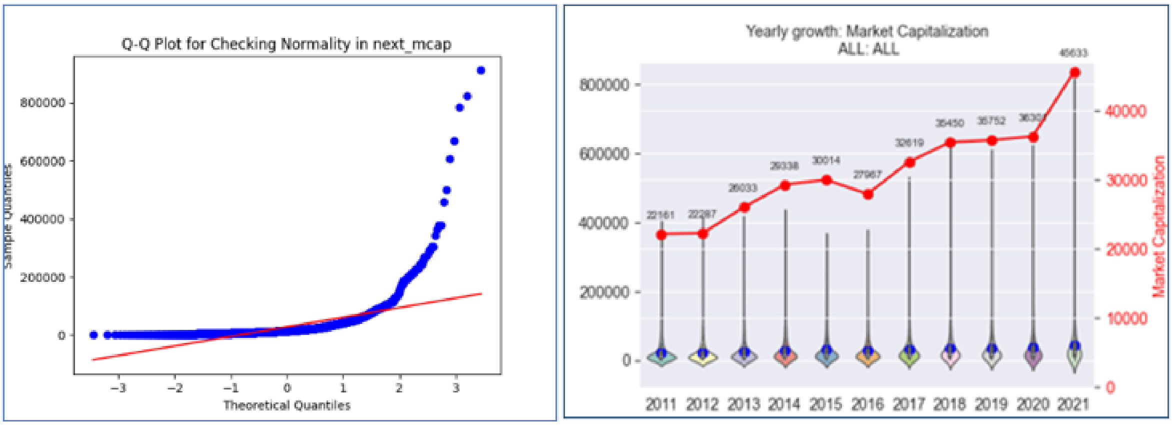

Figure 4: Normality Graph and Year on Year Market Capitalization growth of global companies

5. Model Runs, Analysis and Results

Marko and Lucas (2022) highlight that forecasting performance metrics is a critical component of empirical accounting research, differing from descriptive or explanatory studies. Forecasting efforts aim to identify the predictive power of certain input variables for a specific outcome variable. Such research involves multiple iterations and varying conditions to develop the most effective predictive model relevant to the question (Ding et al., 2020). This approach, originating in information systems, adheres to the tenets of design science research (Hevner, March, Park, & Ram, 2004). Additionally, it pertains to management research in terms of financial outcome predictions, organizational goal setting, and strategy formulation. This methodology is also valuable for understanding potential peer growth and navigating competitive landscapes effectively. Our earlier research also explained these concepts on predictability of Revenue and net income (Vasu and Nirmala 2023-1, 2023-2).

Market capitalization provides insights into company value, investor sentiment, and the overall economic trajectory, which is at the core of this predictability study. However, the number of combinations for modeling increases factorially as the number of input variables grows, leading to an unsustainably high number of models to test. Pearson correlation coefficients offer some support in optimizing the number of input variables; however, they do not help to identify the combination of input variables to develop models with minimal MAPE values. Traditional techniques like factor analysis and cluster analysis fail to identify the precise input variables and their level of impact. Similarly, the Lasso Regression and Ridge Regression Models could not identify input variables' marginal but significant influence over market capitalization with the trials we conducted. To address this challenge, we introduce a novel approach called the "MAPE minimization technique," leveraging the computing power available much more easily today. By employing this method, we aim to identify the most suitable set of input variables for each grouping and outcome to achieve minimal MAPE values.

Let us denote the predicted outcome as F(X), a function of multiple input variables X1, X2, X3, ..., Xn, such that,

$F(X) = F(X_1, X_2, \ldots, X_n)$.

Based on the above, the MAPE equation (10) can be written as follows:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \frac{|Y_i - F(X_{1i}, X_{2i}, \ldots, X_{ni})|}{Y_i} \quad \ldots. (11)$$

This formula shows that MAPE is a function of the actual incomes Y_i and the values of the variables $X_{1i}, X_{2i}, \ldots, X_{ni}$ for each instance i. The MAPE provides an average measure of how accurate the outcome prediction model is, taking into account the variability in both the actual outcome and the values of the predictive variables.

For each of the Xij, an error component can be attached called eij and incorporating the same in the MAPE formula,

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - F(X_{1i}+e_{1i}, X_{2i}+e_{2i}, \ldots, X_{ni}+e_{ni}, X_{(n+1)i}+e_{(n+1)i})}{Y_i}\right| \quad \ldots\ldots. (12)$$

Reducing MAPE through the addition of variable:

If it is assumed that adding the new variable $X(n+1)i$ v could potentially reduce the MAPE by a certain factor 'd' due to the valuable information it provides, even after considering the error $e(n+1)I$ the inequality can be expressed as:

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - F(X_{1i}+e_{1i}, \ldots, X_{ni}+e_{ni}, X_{(n+1)i}+e_{(n+1)i})}{Y_i}\right| < (1-d)\cdot MAPE_{\text{original}} \quad \text{---- (13)}$$

Where:

- 'd' is the proportional reduction in MAPE due to the new variable $X(n+1)I$ considering its error $e(n+1)i$
- $e(n+1)I$ is the error term for the new variable, which is sufficiently small to ensure that the addition of $X(n+1)$ results in a lower overall MAPE.

In this case, 'd' would be a positive number between 0 and 1, representing the percentage by which $MAPE$original is reduced. This reflects the idea that the information gain from the new variable outweighs the additional noise it introduces, leading to an improvement in prediction accuracy as measured by the MAPE.

Increasing MAPE through addition of variable:

If it is assumed that the new variable $X(n+1)I$ could potentially reduce the MAPE by a certain factor d, if there were no error associated with it, but the error $e(n+1)i$ is so high that it not only negates this reduction but actually increases the MAPE, we can write:

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - F(X_{1i}+e_{1i}, \ldots, X_{ni}+e_{ni}, X_{(n+1)i}+e_{(n+1)i})}{Y_i}\right| > (1-d)\cdot MAPE_{\text{original}} \quad \text{---(14)}$$

Where:

- 'd' is the proportional reduction in MAPE that would occur due to $X_{(n+1)i}$ in the absence of $e_{(n+1)i}$
- $e_{(n+1)i}$ is the error term for the new variable, which is large enough to ensure that $MAPE_{\text{new}}$ is greater than $MAPE_{\text{original}}$ after accounting for the reduction factor d.

This inequality shows that the increased error due to $e(n+1)i$ not only compensates for the potential improvement from adding $X(n+1)i$ but actually results in a higher MAPE overall. Applying this concept, progressively adding the variable that potentially will have maximum reducing potential among the variables not added in the equation, the overall MAPE starts reducing and, after a point, starts increasing back, leading to a U-shaped curve. Suppose the curve represents a parabolic equation, FP. In that case, when dFP (the first order

Paper ID: SR24908221149 DOI: https://dx.doi.org/10.21275/SR24908221149

differential) becomes zero, it gives the number and list of variables participating in the minimum MAPE equation.

The total number of regression models based on combination is nC1+nC2+nC3+ … nCn, totaling two power n. The total number of regression models that will be created using the MAPE minimization technique is (n)+(n-1)+(n-2)….+2 +1, which is n* (n+1) / 2, a manageable set of regression equations for identification of the input variables leading to lowest MAPE. The lowest MAPE occurs when the condition, as illustrated by inequality 13, moves to inequality 14.

The models represented by equation 2 and 3 are optimized using MAPE minimization technique. **MAPE minimization and optimizing input variables for predicting Market capitalization**

In the MAPE minimization method, we systematically identify the optimal list of input variables through a set of regression trials. Initially, each input variables are regressed against Market Capitalization, and the variable resulting in the minimum Mean Absolute Percentage Error (MAPE) is selected. For Autoregressive eXogenous (ARX) models, we regress the t-1 year value of the Market Capitalization with other input variables individually. Following this, we add the remaining variables successively to the list based on their ability to minimize MAPE. Initially, as variables are added, the overall MAPE decreases; however, when added variables lead to higher error levels, the overall MAPE increases, forming a U-shaped curve. The minimum of this U-curve represents the lowest MAPE, and the corresponding variables can be considered the optimal set of input variables for the given outcome. The subsequent graphs illustrate MAPE minimization curves for both Auto regressive and Non-auto regressive models, grouped by country and industry classifications. Depending on the intensity of influence versus the error component added, the MAPE minimization curve shape may vary as shown in the graph.

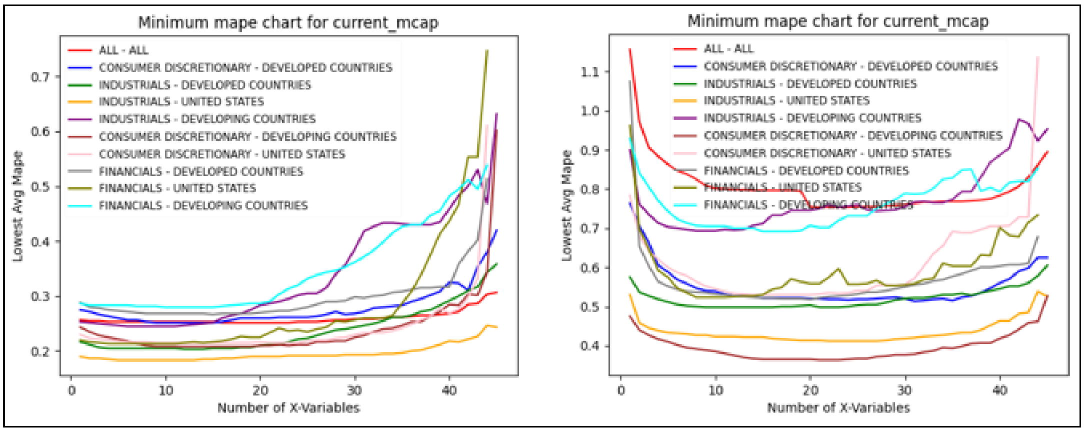

Figure 5: Mape Minimization Chart for ARX and REG Models, respectively, from left.

=While the ARX models have close MAPE values until the 20 input variables and then sharply rising MAPE values, the REG models give much clearer U-shaped curves, as seen in the right-side graph of Figure 5. The industrials of the United States show the lowest levels of MAPE.

We evaluate the accuracy and significance of the chosen combination of input variables and select the combination that yields minimum MAPE. Furthermore, these select input variables are employed in parameter optimization trials for Neural Models to optimize their performance.

The optimized trials show (Table 3) overall MAPE improvements to 11 % due to the exogenous variables in ARX_BNN (Bayesian) models. The ARX and ARX_BNN models have lower MAPE (Complete and 2.5% total winzorized values). The ARX models also show the highest Wald count, showing that the slope and intercept of actual versus predicted market capitalization are closer to 1 and 0, respectively, than the AR or ARX_BNN models. ARX models also perform better than the AR and ARX_BNN models by having higher accuracy, as seen through MASE, RMSE, and Adjusted R-Square.

Table 3: Key Accuracy Results for Market Capitalization – Auto-Regressive Models

Industry	Countries	MAPE			Wald Count			2.5% WINSORIZED MAPE		
		AR	ARX	BNN	AR	ARX	BNN	AR	ARX	BNN
ALL	ALL	0.260	0.252	0.255	4	3	2	0.197	0.190	0.192
CONSUMER DIS	DEVELOPED	0.290	0.252	0.240	6	8	6	0.228	0.207	0.193
	DEVELOPING	0.258	0.208	0.260	8	10	10	0.232	0.192	0.223
	UNITED STATES	0.247	0.215	0.217	7	9	5	0.202	0.177	0.182
FINANCIALS	DEVELOPED	0.300	0.267	0.222	5	8	7	0.275	0.243	0.202
	DEVELOPING	0.288	0.280	0.248	5	6	3	0.223	0.217	0.193
	UNITED STATES	0.223	0.213	0.205	8	7	9	0.195	0.192	0.182
INDUSTRIALS	DEVELOPED	0.220	0.203	0.220	6	6	4	0.193	0.183	0.193
	DEVELOPING	0.275	0.245	0.237	8	9	8	0.237	0.212	0.207
	UNITED STATES	0.203	0.183	0.183	7	6	5	0.165	0.152	0.153
AVERAGE		**0.257**	**0.232**	**0.229**	**6.4**	**7.2**	**5.9**	**0.215**	**0.196**	**0.192**

Industry	Countries	MASE			RMSE			Adj R _Square		
		AR	ARX	BNN	AR	ARX	BNN	AR	ARX	BNN
ALL	ALL	6.1	5.9	6.2	17.5	17.2	17.9	0.93	0.93	0.92
CONSUMER DIS	DEVELOPED	4.8	4.8	4.9	8.7	8.6	9.8	0.95	0.95	0.93
	DEVELOPING	5.1	4.5	5.1	9.1	8.5	9.3	0.84	0.86	0.79
	UNITED STATES	6.8	6.4	7.8	29.8	28.2	35.6	0.92	0.92	0.9
FINANCIALS	DEVELOPED	4.2	4.3	4.1	7.7	7.9	7.6	0.96	0.95	0.96
	DEVELOPING	4.5	4.6	5.4	13.7	14.0	16.2	0.9	0.91	0.85
	UNITED STATES	7.2	6.9	7.4	17.5	16.8	18.9	0.96	0.97	0.96
INDUSTRIALS	DEVELOPED	3.7	3.6	4.1	7.2	7.2	8.2	0.96	0.96	0.95
	DEVELOPING	3.1	3.0	3.0	5.7	5.7	5.6	0.9	0.9	0.9
	UNITED STATES	5.3	5.0	5.2	11.0	10.5	10.4	0.91	0.92	0.92
AVERAGE		**5.1**	**4.9**	**5.3**	**12.8**	**12.5**	**14.0**	**0.923**	**0.927**	**0.908**

On the other hand, REG Models (Non-Regressive Exogenous input variables), have clearly higher MAPE and lower accuracy. But the models are critical to have better understanding of the influence of input variables on the predictability of Market Capitalization. Table 4 shows the MAPE, Winzorized MAPE, and Wald Count for each of the industry and country groupings.

Table 4: Key Accuracy results for Market Capitalization – Non-Auto Regressive Models (only Exogenous Variables)

INDUSTRY	COUNTRIES	MAPE			Wald Count			2.5% WINSORIZED MAPE		
		BNN	REG	ANN	BNN	REG	ANN	BNN	REG	ANN
ALL	ALL	0.947	0.753	0.638	4	3	2	0.692	0.592	0.510
CONSUMER DIS	DEVELOPED	0.705	0.513	0.500	6	8	6	0.583	0.453	0.445
	DEVELOPING	0.462	0.363	0.433	8	10	10	0.395	0.338	0.403
	UNITED STATES	0.993	0.530	0.803	7	9	5	0.747	0.457	0.643
FINANCIALS	DEVELOPED	0.588	0.518	0.423	5	8	7	0.527	0.463	0.387
	DEVELOPING	0.993	0.692	0.930	5	6	3	0.845	0.582	0.783
	UNITED STATES	0.698	0.523	0.662	8	7	9	0.468	0.452	0.537
INDUSTRIALS	DEVELOPED	0.757	0.498	0.480	6	6	4	0.648	0.457	0.427
	DEVELOPING	1.012	0.693	1.020	8	9	8	0.790	0.567	0.822
	UNITED STATES	0.458	0.412	0.433	7	6	5	0.405	0.368	0.380
AVERAGE		**0.761**	**0.550**	**0.632**	**6.4**	**7.2**	**5.9**	**0.610**	**0.473**	**0.534**

For the global set of data, with the group name "ALL -ALL" indicating all sectors and all group countries, table 5 gives the details of X_REG and ARX regression coefficients and the corresponding p values in the order of Pearson correlation. The variables that show significant influence (p values < 0.05) in both models are highlighted.

Table 5: Regression coefficients of selected variables in ARX, REG (OLS) models

VARIABLE	ARX	X_REG	PEARSON	ARX_P	X_REG_P
Market Capitalization	1.0703		0.962	0	
Pretax Income	**0.3963**	**5.0422**	**0.694**	**0**	**0**
Equity		0.3799	0.576		0
Intangible		0.3354	0.452		0
Total Assets		-0.0655	0.255		0.001
Total Liabilities		0.0468	0.215		0.02
Current Liabilities		0.0118	0.193		0.353
Inventories		-0.3859	0.171		0
Cash	**-0.0098**		**0.167**	**0.05**	
ROA	**4685.365**		**0.147**	**0.002**	
Non-Cash Adjusted Income	-0.0543	-1.3345	0.061	0.196	0
Debt Repayment	**-0.1573**	**0.9644**	**0.05**	**0**	**0**
FF_SMB	152.7144		0.031	0.442	
Dividend		0.0088	0.023		0.214
Current Ratio	**135.4776**	**711.4855**	**0.017**	**0.006**	**0**
ROE	6.1026		0.012	0.884	
GDP	**-160.873**		**0.005**	**0**	
Change in Cash	**0.0542**		**0.004**	**0.004**	
Net Margin		-9.4415	-0.001		0.566
Inflation	25.078		-0.002	0.572	
Cash Flow - Finance	**0.1458**	**-0.3919**	**-0.003**	**0**	**0**
Debt to Equity Ratio	0.7387	6.1328	-0.01	0.713	0.16
FF_RMW		-5443.61	-0.016		0
Irregular Income		-2.1019	-0.07		0
Cash Dividend		-5.757	-0.122		0
Sales Investment	**-0.0352**	**-0.2428**	**-0.154**	**0.063**	**0**

Cash Flow - Investment		-0.073	-0.246		0.227
SGA		-0.9157	-0.495		0

The ARX models utilize a minimum set of additional variables. Some variables, such as Cash Flow-Finance, Cash Flow-Investing, Change in Cash, Country GDP, Comprehensive Income, Revenue, and Debt Payment are used more frequently than the others. The detailed list is given in the annexure tables. On the other hand, the more frequent input variables in the equation of optimized X_REG models include Cash Flow-Investing, Comprehensive Income, Equity, Debt Repayment, Dividend, Intangible, Inventories, Investment Sale, Irregular Income, Net Investment, Noncash adjusted Income, Sales and General Administrative expenses.

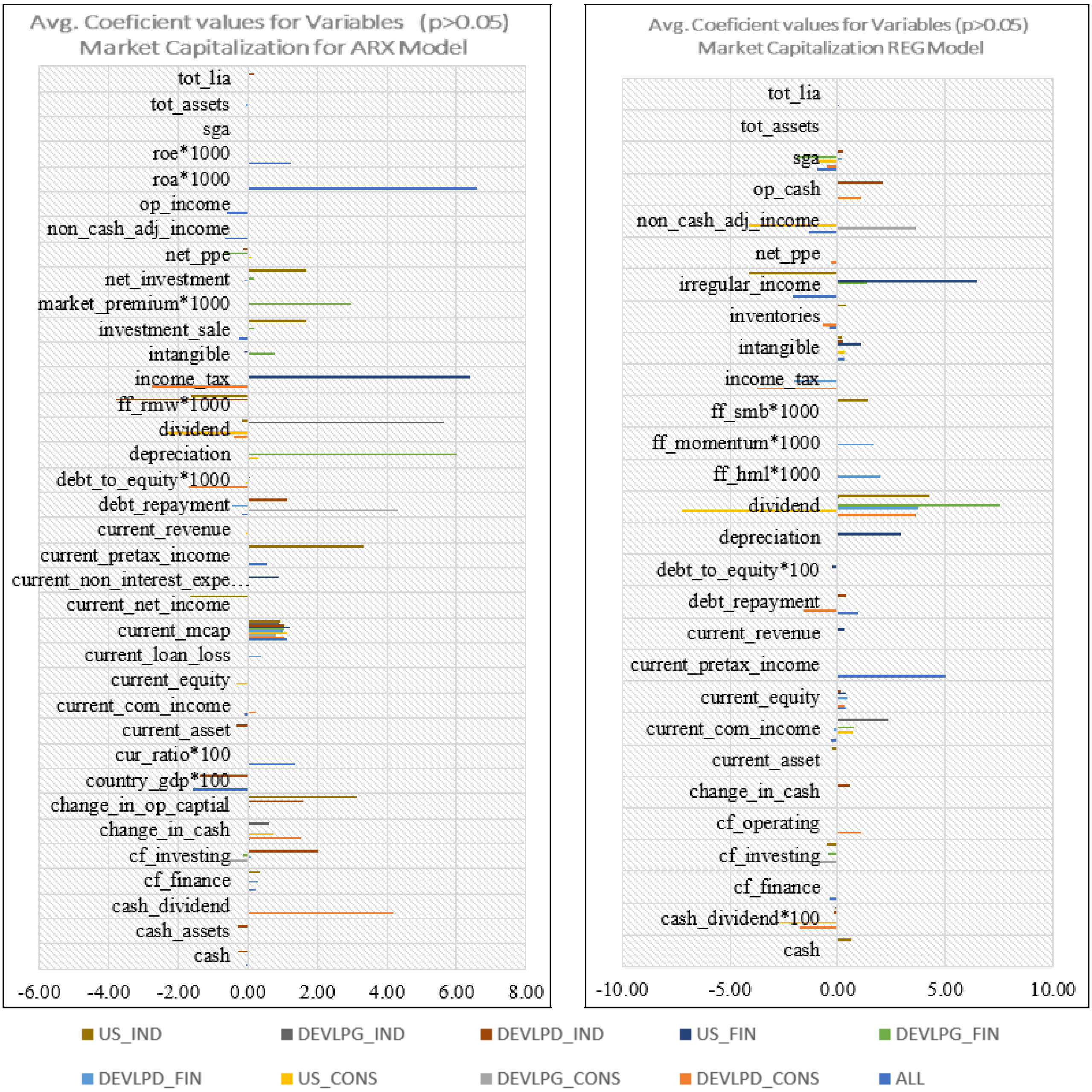

Figure 6: Coefficients of significant variables across the groupings. (All input values are for the year t for the predicted outcome of market capitalization t+1

A visual representation of the ARX and REG model outcomes, as coefficients of input variables is shown in the figure 6. This covers all the 10 groupings. All the variables in the represented regressive equation is shown irrespective of the p-value. The graph shows that dividend and cashflow Investing appear in the equation of most of the models.

The influence of Fama and French factors, including HML (High Minus Low), Momentum, RMW (Robust Minus Weak), and SMB (Small Minus Big), on market capitalization is prominently observed in developed countries, including the United States. Notably, at least one of these factors makes a significant contribution to the regression equation.

The regression equations for REG models of global data (ALL-ALL) show that developed countries and the United States have 10 significant variables while developing countries show 5 significant variables. This distinction underscores the advanced level of financial reporting maturity, facilitating greater predictability within mature markets, particularly the financial sector.

In each of the 10 groupings, it is noted that at least one variable possesses a coefficient in the regression equation with a p-value less than 0.05, leading to the rejection of the null hypothesis (Ho) that the variable does not significantly affect predictability. The details of coefficients and p-value are given in the annexure in Tables 7, 8, and 9.

While OLS models, as described in equations 1, 2 and 3 are efficient in quantifying the relationship between variables when assumptions of linearity hold true, Neural Networks , as described in equations 4, 5 and 6 , (NN) can model complex nonlinear relationships between input variables and the predicted outcome. By incorporating nonlinear activation functions within NN models, we can effectively address potential nonlinearity within our data while simultaneously handling both categorical and continuous variables.

Although Feed Forward NN models (Artificial Neural Networks or ANN models) generate deterministic outcomes based on input variables, Bayesian Neural Network (BNN) models offer additional advantages. BNN models, as described in equations 7, 8 and 9 possess two heads: one for averaging predictions and another for standard deviations. This unique feature enables them to supply confidence levels for each set of input variables at any given alpha level, shifting from deterministic to stochastic outcomes. By incorporating these models in our research, we aim to boost prediction accuracy further.

Visual representation of the predicted vs actual market capitalization is a convenient way for inferences. Figure 7 illustrates the Bayesian Model (ARX_BNN) for the global data (ALL-ALL).

For the years 2016, 2018, and 2019, the slope of the actual vs. predicted outcome is greater than 0.9, with an adjusted r-square also greater than 0.9. This strongly supports the hypothesis that the global model can predict market capitalization.

Bayesian Neural Network (BNN) charts (Figure 7) demonstrate individual data point confidence interval limits and population-level confidence intervals. However, as market capitalization increases, confidence intervals become significantly wider in proportion to the value. Financial markets were well supported in the COVID-19 year of 2020 as part of protecting the economy. The Market capitalization further improved significantly in 2021. This can be seen in the 2020 graph with a significantly lower slope. The model could not fully absorb this variation from input variables, leading to a higher Mean Absolute Percentage Error (MAPE). This is also confirmed by the World Federation of Exchanges data. The WFE data shows that the world's market capitalization increased from 83 Trillion USD in 2019 to 88.4 Trillion USD in 2020. However, it sharply rose in 2021 to 118.5 TUSD.

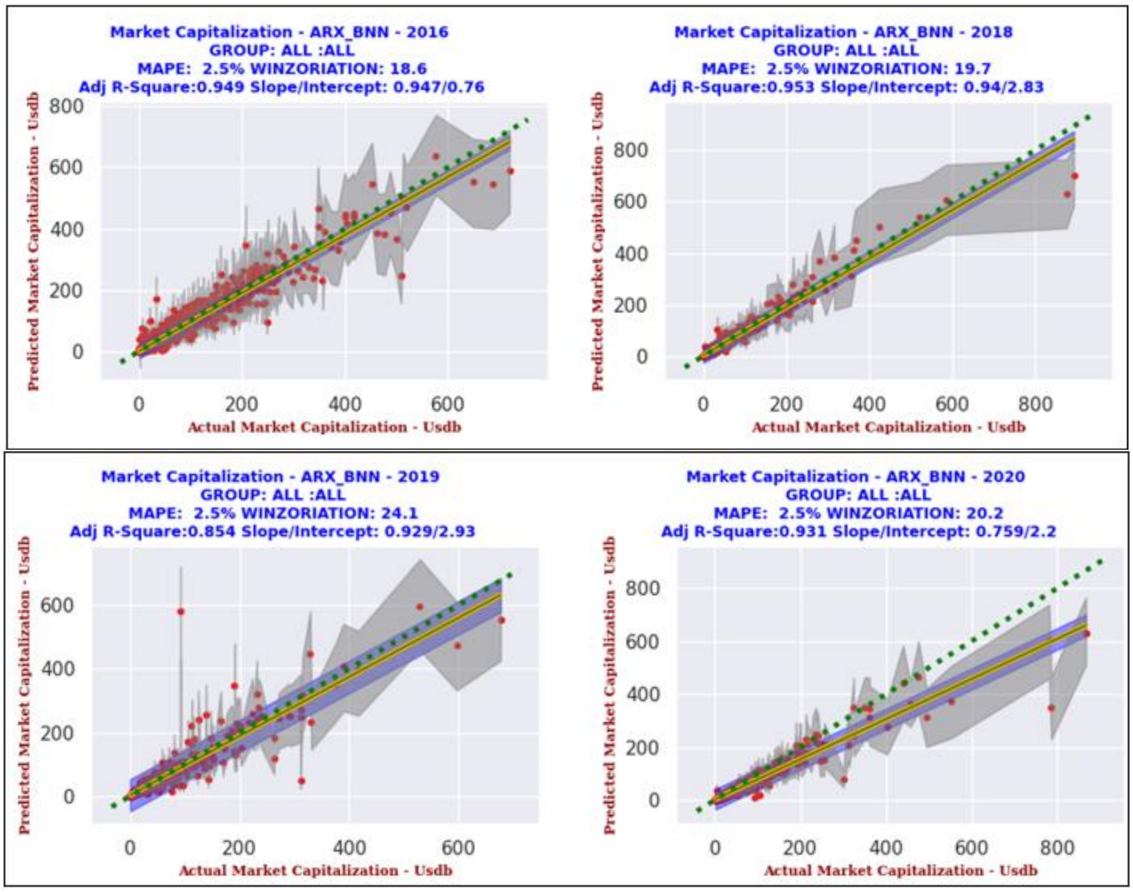

Figure 7: X-Y chart view of predicted vs actual market capitalization for different models

While visual analysis offers an understanding of the workings of the prediction models, the statistical evaluations using the Kruskal-Wallis test, Mood's median test, and Kolmogorov-Smirnov test, as presented in Tables 6, indicate that the ARX models yield a smaller number of test outcomes with p-values below the significance threshold (alpha = 0.05). H0 is rejected in 14 tests in the ARX Models and 24 tests in the AR model. In all the cases, the ARX_BNN model best performs (0nly in 10 tests, the null hypothesis is rejected). As expected, REG models have much higher rejection rates. These results infer the superiority of ARX models over AR models and further the higher predictability exhibited by Bayesian Neural Networks.

Table 6: By Year and Model, Hypothesis rejection for statistic to be the same.(H0 rejected)

Year	AR	ARX	ARX_ANN	ARX_BNN	X_REG BNN	X_REG	X_REG ANN	Total Rejections
2016	3	1	5	2	17	11	17	56
2017	0	0	0	0	6	3	10	19
2018	5	3	3	2	7	4	8	32
2019	5	3	3	2	12	6	11	42
2020	3	2	3	3	15	17	13	56
2021	0	0	0	0	16	15	21	52
All Years	8	5	3	1	16	9	17	59
Total Rejections	24	14	17	10	89	65	97	316

The year-wise comparison indicates that in 2020, there were more rejections of the hypothesis than in other years in the out-of-sample testing. This increase in rejections is obvious as the Market capitalization significantly increased in 2021, impacting the predictability of the 2020 data over 2021. Due to the larger data set in ALL years and 2010-2016, the number of rejections is higher. In 2020, market capitalization was not impacted to the extent of net income or even revenue as investor confidence bounced back after the COVID academic break due to the government's pre-emptive actions to infuse money into the economy, such as PPP in the United States. The lower and comparable rejections for 2021 predicting the market capitalization in 2022 show the robustness of the models.

For individual country groupings and sectors, the ARX model was defined with variables selected to match or exceed the accuracy measurements of AR models using the MAPE minimization method. (Refer to Table 6). The higher degree of reduction of 97.5% MAPE compared to the overall MAPE shows each grouping's long tail of higher market capitalization. The United States' 97.5% MAPE is 34% better than that of developed and developing countries, again showing business and economic maturity.

6. Conclusions and Further Study

Our research findings suggest that the auto-regressive market capitalization model, with the previous year's market capitalization, effectively predicts the market capitalization of a given year. Including additional exogenous variables derived from financial statements, macroeconomic indicators, and capital market factors enhances the prediction models' accuracy on average by 11%. Furthermore, Bayesian models offer promising alternatives by enhancing prediction accuracy and supplying confidence intervals for individual predictions. The greater number of significant variables contributing to predictability in advanced economies reflects the maturity of markets in these regions. Among the 44 input variables selected, Cash Flow-Investing, Comprehensive Income, Equity, Debt Repayment, Dividend, Intangible Assets, Inventories, Investment Sales, Irregular Income, Net Investment, Noncash Adjusted Income, Sales, and General Administrative Expenses have demonstrated influence over market capitalization predictability. Our newly developed MAPE minimization method can identify these significant variables while minimizing Mean Absolute Percentage Error (MAPE) to optimize predictive accuracy. As in Table 4, the United States exhibits the lowest combined Mean Absolute Percentage Error (MAPE) of 0.366. Developed countries also display a similar MAPE of 0.382, while developing countries have a higher MAPE of 0.437. These observations further highlight the maturity of capital market-based economies. Regarding industry-level predictability, the consumer sector shows a higher MAPE of 0.54, whereas industrials demonstrate the lowest MAPE of 0.4, and financials have a MAPE of 0.433. The main value of this research is finding the key exogenous variables using MAPE minimization techniques and predicting the market capitalization with a linear equation using the coefficients of all exogenous variables for each country-industry grouping from the annexure tables. Stock prediction can easily be done using the predicted market capitalization.

In conclusion, this study demonstrates that the autoregressive model, coupled with external financial and economic variables, significantly improves market capitalization predictability for large global companies. Bayesian Neural Networks, in particular, offer enhanced accuracy and provide confidence intervals, making them a valuable tool for financial forecasting. These findings are essential for corporate finance professionals looking to optimize decision-making and strategic planning. Future research should explore industry-specific factors to refine prediction models

References

[1] Al-Afeef, Mohammad Abdel Mohsen (2020). Factors affecting market capitalization: A practical study case 1978-2019. International Journal of Scientific and Technology Research 9.3 (2020): 7049-7053.

[2] Dash, R.K., Nguyen, T.N., Cengiz, K., & Sharma, A. (2021). Fine-tuned support vector regression model for stock predictions. *Neural Computing and Applications, 35*, 23295-23309.

[3] Ding, K., Lev, B., Peng, X., Sun, T., & Vasarhelyi, M. A. (2020). Machine learning improves accounting estimates: Evidence from insurance payments. Review of Accounting Studies, 25(3), 1098-1134. https://doi.org/10.1007 /s11142-020-09546-9

[4] Fisher, Gregg S. and Shah, Ronnie and Titman, Sheridan (2016), Analyst Long-Term Growth Forecasts, Accounting Fundamentals and Stock Returns, Available at SSRN: https://ssrn.com/abstract=2759332 or http://dx.doi.org/10.2139/ssrn.2759332, April 4, 2016.

[5] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105. https://doi.org/10.2307/25148625

[6] Kumar, Naresh & Rai (2011). Use of Mathematical Models to Forecast Market Capitalisation and Economic Growth in India. 10.13140/RG.2.1.2109.1365.

[7] Kureljusic, Marko and Reisch, Lucas. (2022). Revenue forecasting for European capital market-oriented firms: A comparative prediction study between financial analysts and machine learning models. Corporate Ownership & Control, 19(2), 159–178. https://doi.org/10.22495/cocv19i2art13. Available at SSRN: https://ssrn.com/abstract=4057309

[8] Murugesan Selvam, JayapalGayathri, Vinayagamoorthi Vasanth, Kasilingam Lingaraja, Sigo Marxiaoli, (2016). Determinants of Firm Performance: A Subjective Model, International Journal of Social Science Studies, Vol. 4, No.7, July 2016.

[9] Pavone, P. (2019). Market Capitalization and Financial Variables: Evidence from Italian Listed Companies. International Journal of Academic Research Business and Social Sciences, 9(3), 1356–1371.

[10] Rouf, N.; Malik, M.B.; Arif, T.; Sharma, S.; Singh, S.; Aich, S.; Kim, H.-C (2021). Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. Electronics 2021, 10, 2717. https://doi.org/10.3390/electronics10212717

[11] SÖRMAN, VIKTOR ELDÉUS (2022), Multiple Linear Regression of the market capitalization through financial ratios of listed large cap companies on Stockholm Stock Exchange during COVID-19, Kungliga Tekniska högskolan, Stockholm, Sweden 2022. https://www.diva-portal.org/smash/get/diva2:1756988/FULLTEXT01.pdf

[12] Vasu Padmanabhan, Dr S. Nirmala (2023). "Impact of Covid pandemic and its recovery phase on the net income of large companies and its predictability (A global analysis)." Boletin de Literatura Oral-The Literary Journal 10.1 (2023): 608-622.

[13] Vasu Padmanabhan, Dr S. Nirmala (2023). "PREDICTABILITY OF TOTAL REVENUE IN GLOBAL LARGE COMPANIES FROM FINANCIAL AND ECONOMIC PARAMETERS." International Development Planning Review 22.2 (2023): 862-879.

Annexure

The OLS Regression Results for country groupings and Sectors

Table 7: The long table of coefficients for the Consumer Discretionary Sector for the country groupings

	CONSUMER DISCRETIONARY					
	DEVELOPED		DEVELOPING		UNITED STATES	
Input Variables	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Cash			0.28	0.39		
Cash Dividend	1.99	0.20				
Cashflow Investing			-1.23	0.48		
Change in Cash	0.59	0.14			5.88	0.09
Change in Operating Capital	0.38	0.22	0.40	0.42		
GDP			370.75	0.34		
Current Ratio			1575.84	0.54		
Com Income	0.04	0.61	-0.14	0.73		
Equity					-0.12	0.33
Market Capitalization	1.10	0.00	0.86	0.00	1.23	0.00
Revenue			0.07	0.43	-0.16	0.22
Debt Repayment			0.48	0.50	-0.36	0.24
Debt to Equity	-78.37	0.49	325.67	0.55	-150.61	0.35
Depreciation					1.46	0.36
FF- Momentum			403.73	0.87		
FF-RMW			-70.24	0.47	-219.80	0.66
FF-SMB					2744.53	0.32
Income Tax	-0.15	0.31	6.63	0.47		
Investment Sale	-2.11	0.46				
FF-Market Premium			280.39	0.81		
Net Investment	2.00	0.47				
Net PPE					0.21	0.33
ROE	-709.71	0.81				
Total Assets			-0.02	0.43		

Table 8: The long table of coefficients for Financials Sector for the country groupings

	FINANCIALS					
	DEVELOPED		DEVELOPING		UNITED STATES	
Input Variables	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
CF_Finance	0.02	0.43				
CF_Investing	-0.03	0.33	0.29	0.31		
CF_Operating	`		0.22	0.33		
Change in Operating Capital	-0.01	0.39				
Inflation			-63.13	0.63		
Current Com Income	0.03	0.31			-0.06	0.33
Loan Loss	0.32	0.30				
Market Capitalization	1.05	0.00	0.97	0.00	1.15	0.00
Non Interest Expense	-0.01	0.44			0.48	0.31
Debt Repayment	-0.11	0.32	-0.12	0.22		
Debt to Equity	18.88	0.43	-86.71	0.45	36.09	0.54
Depreciation			12.07	0.25		
FF_CMA			3073.68	0.40		
Income Tax					2.47	0.26
Intangible			0.10	0.45	0.06	0.21
Investment Sale			-0.10	0.37		
Market Premium			498.08	0.60		
Net Investment	0.02	0.47	-0.10	0.37	-0.06	0.11
Net PPE			0.22	0.30	0.12	0.42
SGA	-0.01	0.29				
Total Assets	-0.05	0.23				
Total Liabilities	0.05	0.26				

Table 9: The long table of coefficients for the Industrials Sector for the country groupings

	INDUSTRIALS					
	DEVELOPED		DEVELOPING		UNITED STATES	
Input Variables	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Cash	0.03	0.43				
CF_Finance					0.25	0.27
CF_Investing	0.18	0.38				
Change in Cash			0.31	0.47		
Change in Operating Capital	0.34	0.40			1.36	0.20
Contract Obligation			0.02	0.42		
GDP	-60.72	0.44				
Current Ratio			-245.84	0.28		
Market Capitalization	1.06	0.00	1.02	0.00	0.97	0.00
Net Income					-2.71	0.23
Pretax Income					3.68	0.22
Debt Payment	0.50	0.25				
Debt to Equity					-0.91	0.45
Dividend			-1.50	0.24	0.91	0.18
FF_RMW	-776.91	0.34	-852.83	0.51	-2200.33	0.19
Investment Sale			0.26	0.61	-8.01	0.46
Net Investment					7.23	0.46
ROA			9313.79	0.14		
ROE			623.72	0.85	1666.68	0.52
Total Liabilities	0.02	0.20				