COMILLAS
UNIVERSIDAD PONTIFICIA

Facultad de Ciencias Económicas y Empresariales

ENHANCING THE ACCURACY OF THE EPS ESTIMATES CONSENSUS USING A META-MODEL

Autor: Carlota Iraizoz Sánchez
Clave: 201806593
Director: Alejandro Rodríguez Gallego

MADRID | June 2023

ABSTRACT

Corporate earnings per share (EPS) estimates are crucial financial metrics that provide valuable insights into a company's earnings potential and financial health. Indeed, many investors rely on EPS estimates to make informed investment decisions. Hence, it is not surprising that a relevant task of sell-side financial analysts is to forecast EPS with a view to providing investors with a reliable and clear picture of a company's financial soundness. In this context, the accuracy of EPS estimates is of particular importance and has therefore been a widely researched topic. More in detail, EPS estimates are usually computed using a consensus approach that aggregates individual forecasts to arrive at a median or average estimate. Since consensus methods are often simple, the aim of this research is to develop a meta-model to explore whether enhancing the accuracy of consensus EPS estimates is viable and worth exploring in further detail. This research concludes with an outline of next steps to be considered for the purpose of refining the proposed meta-model.

Key words: EPS, EPS estimates, analysts, accuracy, consensus, meta-model, wisdom of crowds

TABLE OF CONTENTS

TABLE OF ILLUSTRATIONS: FIGURES AND TABLES

GLOSSARY OF ACRONYMS AND ABBREVIATIONS

AAPL: Apple Inc.

ACN: Accenture Plc A

ADBE: Adobe Inc.

AVGO: Broadcom Inc.

CRM: Salesforce Inc.

CSCO: Cisco Systems Inc,

EMH: Efficient Market Hypothesis

EPS: Earnings per Share

LR: Linear Regression

MAE: Mean Absolute Error

ME: Mean Error

ML: Machine learning

MSFT: Microsoft Corp

NVDA: Nvidia Corp

ORCL: Oracle Corp

P/E: Price to Earnings

RF: Random forest

SVM: Support Vector Machine

TXN: Texas Instruments Inc.

WRDS: Wharton Research Data Services.

1. INTRODUCTION. OBJECTIVES AND THEME JUSTIFICATION

Earnings per share (EPS) is widely regarded as one of the most important financial metrics to determine a company's profitability and financial health. From investment decision-making to valuation modelling, the versatility of EPS estimates is astonishing and is thus considered a critical financial metric by investors, traders and managers. Since EPS estimates are widely used in financial markets, it is not surprising that one of the most relevant tasks of sell-side analysts is to forecast EPS. Provided the extent at which EPS estimates exert influence on business and investment decisions as a whole, ensuring their accuracy is of particular importance (more so now considering the current scenario of macroeconomic and global uncertainties). In this context, EPS estimates are currently computed using a fairly simple consensus approach that aggregates individual forecasts to arrive at a median or average estimate.

The aim of this research is to develop a meta-model to arrive at a better consensus methodology for EPS estimates. In other words, this paper will explore sophisticated aggregation approaches beyond that of simple mean consensus methods with a view to assessing whether enhancing the accuracy of consensus EPS estimates is viable and worth exploring in further detail. With a view to narrowing the scope of the research, the meta-model will be developed for the top 10 companies within the information technology sector of the S&P 500.

In order to proceed with the development of the meta-model, an empirical analysis will be carried out. In this context, the hypothesis to be tested is that the proposed meta-model will lead to lower forecast errors compared to baseline consensus approaches. This "model of models" is essentially an ensemble algorithm that will combine the predictions of multiple base models with a view to enhancing overall accuracy. Amongst the existing ensemble algorithms (bagging, boosting, random forests etc.), the most suitable method for the present research is stacking. In brief, a suite of heterogeneous predictive models will be trained using individual analysts' estimates as input data with a view to increasing the accuracy of consensus EPS estimates for each base model. Next, a meta-model will be trained to learn how to best combine the predictions of these base models using their predictions as input data. The thought process that motivated this decision was that accuracy can be greatly enhanced when models are based on multiple algorithms and data (Pavlyshenko, 2019).

Although the above methodology is of great value for this research, there are some limitations inherent in it. First, the topic at hand has not been widely explored since researchers have generally focused on individual estimates rather than on consensus (Kua, 2022). Since literature review is relatively scarce, acquiring a strong understanding of the field might be somewhat complex. Second, even though there are substantial amounts of easily accessible data regarding corporate EPS, more detailed data on analysts' estimates is challenging to locate. With a view to ensuring the quality and consistency of data throughout this research, the following databases have been accessed, with Wharton Research Data Services[1] being the main source of data:

- Refinitiv Workspace.
- S&P Global.
- World Bank Open Data.

Even though the complexity inherent in the present research is rather high, finding appropriate consensus approaches is of particular importance in the context of EPS. According to Islam et al. (2014), the vast majority of investors make their investment decisions based on EPS. Consequently, when listed companies publish their earnings reports and their results deviate from EPS estimates, stock price movements tend to happen (Jagliński, 2020). For this reason, extensive research was conducted on the accuracy of analysts' forecasts of corporate earnings per share. More in detail, researchers analyzed whether the accuracy of analysts' estimates were superior to time-series forecasts. After considerable research, it was concluded that analysts' estimates were indeed superior (Brown et al., 2008). In this framework, a topic that has been relatively unexplored is the accuracy of aggregation approaches that lead to a consensus estimate (Kua, 2022). Since companies "change their investment, employment and payout decisions to ensure that reported EPS meets of beats analyst consensus EPS estimates" (Almeida, 2018, p. 175), ensuring the accuracy of the latter is of particular importance.

[1] Wharton Research Data Services (WRDS) was used in preparing "Enhancing the Accuracy of the EPS Estimates Consensus using a Meta-model". This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

Within this context, this research attempts to contribute to the existing literature by proposing an alternative method of aggregating EPS estimates with a view to assessing whether enhancing their accuracy deserves further investigation.

This paper is structured in five main sections. The paper begins with a theoretical framework that reviews existing literature on both, the accuracy of EPS estimates and research regarding more sophisticated methods of aggregating EPS estimates. Apart from the literature review, this section provides the inquisitive reader with the necessary context for understanding the topic at hand. Having understood the importance of the present paper, the methodology applied in the research is explained in detail, shedding light on the techniques employed. Furthermore, the data needed to carry out this research is listed in the next section. All of the above leads to the central focus of the research, the analysis and development of the meta-model. The analysis will begin by exploring the effects of EPS surprises on stock prices as well as the forecast errors of individual analysts' estimates. The next step will involve proposing an alternative method of aggregating analyst estimates by training a "model of models" (using stacking as the ensemble learning model), whose results will then be compared against baseline methods. After evaluating results, an outline of next steps will be presented, and conclusions will be drawn.

2. THEORETICAL FRAMEWORK

2.1 Literature review

According to Timmermann (2018), forecasting in all fields of finance has proven to be a highly complex task. Following this trail of thought, EPS estimates have also been a subject of debate amongst researchers. Indeed, research on analysts' accuracy on corporate earnings per share traces back to Cragg and Malkiel (1968). Why so, one might ask?

As aforementioned, EPS is a highly valuable financial metric for investors and managers alike. On the one hand, EPS is the crucial catalyst for investors to making informed investment decisions, with the majority of them turning to this metric when managing and allocating their assets (Islam, et al., 2014). On the other hand, EPS estimates are fundamental for companies and managers because the latter make business decisions using EPS estimates as a benchmark (Almeida, 2018).

In light of the above, EPS estimates provide a "quantitative proxy for market sentiment" (Kua, 2022, p.3230) and thus, can derive in stock price movements when quarterly results deviate from market expectations (Jagliński, 2020). The rationale behind this statement is rooted in the Efficient-market hypothesis, a theory that suggests that asset prices fully reflect all available information in the market (Brecque, 2019). In this context, a market in which prices always reflect all available information is considered an "efficient" market (Fama, 1970, p.1). Elaborating on this theory, if public companies publish their quarterly earnings and these deviate from EPS estimates, stock prices must move according to the latest available information, which explains price deviations in this context. Another theory that might explain price deviations is the signaling theory. According to Prijanto et al. (2021, p. 75), the publication of information will signal investors in making investment decisions and will thus "trigger a market reaction in the form of stock price fluctuations". In the context of EPS, the announcement of quarterly results, which is essentially a source of information, will be seen as a better or worse signal according to market expectations and, after its analysis, a market reaction will be expected.

Considering the extent at which EPS estimates exert influence on financial markets, ensuring their accuracy is of particular importance. Within this framework, extensive research was conducted on whether EPS forecasts from timeseries models could beat analysts' estimates. Results were varied, with Cragg and Malkiel (1968) and Elton and Gruber (1972) arguing that analyst precision was no different from that of time series. However, most studies such as those carried out by Barefield and Comiskey (1975) and Brown et al. (2008) indicated that the precision of the analysts was much higher than the time series. In brief, academics have largely concluded that, even though analysts' estimates contain biases, they still outperform simple mechanical models such as timeseries (Kua, 2022).

While the accuracy and implications of analysts' EPS estimates have been widely researched, literature on more sophisticated methods of aggregation to compute an EPS consensus estimate with a lower forecast error is somewhat scarce. The relatively scarce literature on this subject is puzzling considering the importance of the EPS consensus estimates. In fact, as stated by Graham et al. (2005), 73.5% of Chief Financial Officers believe that consensus EPS estimate is most important performance target.

Building a more sophisticated aggregation model to reduce the forecast error when computing the consensus EPS estimate is a very compelling idea considering that, when "different models are based on different algorithms and data, one can receive essential gain in the accuracy" (Pavlyshenko, 2019, p.1). This theory has received further support by other authors such as Lemke and Gabrys (2010) and Yu et al. (2009). Even though there are multiple studies on meta-models to reduce forecast errors, there has not been extensive research in the context of earnings per share. Some authors such as Kua (2022) have explored the idea of developing iterative filtering algorithms for aggregating individual EPS estimates. Other authors such as Nilsson and Svensson (2019) developed weighted models for the same purpose. While Kua (2022) concluded that iterative filtering algorithms resulted in a lower forecast error, Nilsson and Svensson (2019) did not achieve so with weighted models.

As it can be observed, literature on more sophisticated aggregation methods for computing a consensus EPS estimate is relatively scarce. Having understood the importance of EPS estimates in financial markets, further research is need in this context with a view to exploring whether the accuracy of EPS consensus estimates can be enhanced.

2.2 Conceptual framework

Prior to developing the meta-model, this research will provide the inquisitive reader with an overview concerning EPS estimates so that one can gain the necessary context for understanding the topic at hand before diving into the complexity of the meta-model.

2.2.1 Interpreting EPS estimates

Amongst the multiple instruments used for measuring the success and profitability of companies, ratios analysis is one of the most commonly employed methods. Within this context, EPS plays a highly relevant role in determining share price and firm value (Islam et al., 2014).

Before elaborating further, one must comprehend the term “earnings per share”. In short, EPS represents the available amount of a company’s earnings, having subtracted taxes and preferred stock dividends, that is allocated to each outstanding share of common stock. This ratio can be easily calculated by dividing a company’s net income in a given period (commonly quarterly or annually) and dividing it by the total number of outstanding shares during that same period. Since the number of shares tend to fluctuate, this ratio is usually computed using a weighted average:

$$Earnings\ per\ share = \frac{Net\ Income - Dividends\ on\ preferred\ stock}{Average\ outstanding\ shares}$$

Having understood the meaning and math of EPS, the interpretation of this financial metric is more straightforward. Before elaborating further, it should be noted that, when classifying EPS as high (or low), one must compare this figure across similar companies within the same industry while also considering historical EPS trends for that company. Having made this point clear, a higher EPS denotes increased value because investors will be willing to pay more for a company’s shares if they believe profits will be greater than the share price and vice versa. In this context, it is worth mentioning that, even though a higher EPS indicates greater profitability, it does not provide insights into the reliability of the investment. For that matter, one must look into historical EPS. In brief, if a company’s EPS has been steadily increasing over time, it could be regarded as a more reliable investment that another one with a declining or fluctuating trend.

Even though the above may appear to be straightforward, the interpretation of EPS may differ depending on the type of EPS being analyzed due to differences in accounting principles (Jensen and Jones, 2020). In other words, there are many varieties of EPS, and depending on which one is selected, a stock may appear over or under-value. The 5 types of EPS are briefly described in Table 1.

Table 1: Types of EPS (Islam et al., 2014).

Type	**Description**
Reported EPS (or GAAP EPS)	Figure derived from GAAP accounting principles, which are reported in SEC filings.

Ongoing EPS	Figure calculated upon normalized (ongoing) net income. In other words, ongoing EPS excludes one-time events, such as unusual expenses. This type of EPS is used to find earnings from core operations and better forecast future EPS.
Pro Forma EPS (or Street earnings)	Figure that usually excludes some expenses or income that are otherwise included in GAAP EPS. For instance, if a company sold a division, it could exclude the expenses and income associated with it, thus allowing better comparison.
Headline EPS	Figure that is announced in the company's press release and featured in the media. It might sometimes coincide with pro forma EPS.
Cash EPS	Cash EPS is calculated differently from others in the sense that it is computed by dividing operating cash Flow by diluted shares outstanding. Cash EPS is said to be a "purer" number because operating cash flow cannot be as easily manipulated as net income.

Source: adapted from Islam et al. (2014).

After introducing the basics of EPS, understanding EPS estimates is not a complicated task. In short, EPS estimates are an analyst's forecast of a company's future EPS for a quarter or fiscal year. Continuing with the previous train of thought, a higher EPS estimate signals better future performance and greater value for the period analyzed. Since EPS estimates can be used in conjunction with the company's estimated P/E to derive an estimated future price for the company's stock (Clayman and Schwartz, 1994), EPS estimates are crucial in investment decisions and recommendations, stock valuation and business decisions as a whole.

Due to their importance in financial markets, one of the most important tasks of sell-side analysts is to forecast EPS. How they do so demands a great deal of work, resources and effort. Indeed, analysts factor multiple variables (from both internal and external sources) in their estimates:

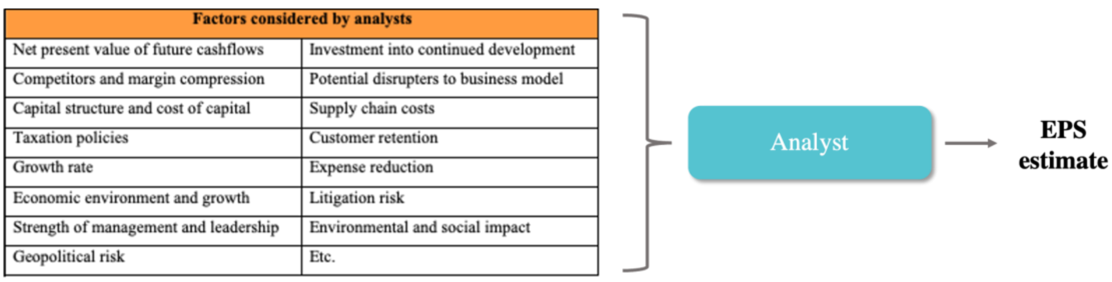
Table 2: Factors Considered by Analysts in their Estimates (Kua, 2022).

Factors considered by analysts	
Net present value of future cashflows	Investment into continued development
Competitors and margin compression	Potential disrupters to business model
Capital structure and cost of capital	Supply chain costs
Taxation policies	Customer retention
Growth rate	Expense reduction
Economic environment and growth	Litigation risk
Strength of management and leadership	Environmental and social impact
Geopolitical risk	Etc.

Source: adapted from Kua (2022).

Without delving too deeply into how analysts estimate EPS, analysts basically take multiple variables, both of quantitative and qualitative nature, into consideration and translate them into a "quantitative" earnings estimate (Kua, 2022).

Figure 1: Summary Process Estimating EPS.

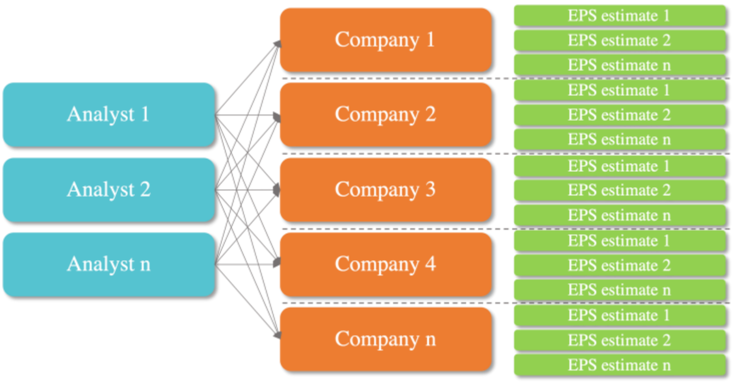


Source: own elaboration.

2.2.2 Consensus EPS estimates and "Wisdom of Crowds"

Since the broker industry is highly competitive, there are multiple broker firms (JP Morgan, Jefferies etc.) employing analysts to forecast EPS on the same company. Furthermore, analysts usually specialize in one industry and thus, provide EPS estimates for several companies. As a result, the number of analysts that cover a specific public company generally range from 1 to 50 (Jensen and Jones, 2020, p. 406). A diagram that better exemplifies the above statements is shown in Figure 2.

Figure 2: Analysts' EPS Estimates for Companies.



Source: own elaboration.

As aforementioned, despite these estimates being superior to simple mechanical models such as timeseries, they still contain biases. On the one hand, findings from authors such as Sedor (2002) argue that analysts have a tendency to make overly optimistic forecasts. On the other hand, some authors such as Hilary and Hsu (2013) suggest that analysts tend to generate estimates below the outcomes on purpose. Despite these contradictions, literature has concluded that analysts are biased and thus, so are their EPS estimates.

With a view to avoiding the bias and potential forecast errors that may arise when relying on a single analyst's estimate, a consensus estimate is computed. To summarize, the consensus EPS estimate aggregates various EPS estimates from multiple analysts and is the figure that is often used as a benchmark against which a company's performance is assessed and compared. Consequently, the consensus estimate is the figure that investors and managers generally turn to when making their investment and business decisions respectively.

Figure 3: Consensus Estimates.

Source: own elaboration.

The reasoning behind relying on the consensus EPS estimate rather than on a single analyst's estimate can be explained applying the "Wisdom of the Crowd" theory. This idea was first introduced by Galton (1907) after conducting an experiment in which individuals had to guess the weight of an ox. It is worth highlighting that, since they were competing for prizes, their judgements were not influenced by others. The wisdom of the crowd (the median guess of the group individuals) was accurate within 0.8% (Galton, 1907, p. 79). It has since been firmly acknowledged that aggregating predictions across individuals can be more accurate than that of a single expert. As stated by Yi et al. (2012, p. 452), the wisdom of the crowd relies on "being able to sift out the noise in individual judgments" by taking the average of individuals' predictions or guesses.

The conditions for this phenomenon to occur are outlined by Ray (2006). First of all, there must be diverse opinions or predictions (in this context, regarding EPS estimates). Second, individuals must calculate and express their guesses independently so that none of the predictions are influenced by others. Lastly, individuals must be able to benefit from their predictions and knowledge (as in the experiment). These three prerequisites do occur in the context of EPS since:

- Information is not uniform across analysts and hence, "two different analysts may come up with two different earnings estimates" (Kua, 2020, p. 3231).
- Internal processes for each analyst are un-observable (Kua, 2020, p.3231) and thus not easily replicable. Also, analysts don't benefit from making similar EPS estimates to others.
- Analysts make their investment recommendations based on their predictions. Hence, they do reap profits if their predictions are more accurate than not.

From the above, it can be inferred that the "wisdom of the crowd" can be applied to EPS estimates. In fact, "the consensus forecast is generally superior to forecasts of individual analysts" (Jensen and Jones, 2020, p.406), which explains why analysts' consensus EPS estimates are generally considered as a reliable proxy for the market's expectations by investors and managers.

3. METHODOLOGY

3.1 Main objectives

Recapitulating, the aim of this research is to develop a meta-model using stacking as the ensemble learning method with a view to arriving at a better consensus approach for EPS estimates. In brief, the proposed methodology will start by analyzing the importance of consensus EPS estimates in business decisions to then focus on developing a more sophisticated aggregation approach beyond that of simple mean consensus methods.

3.2 Scope of the research

With the purpose of limiting the scope of the research, only a subset of companies will be selected for the development of the meta-model. Since EPS is easier to forecast for larger companies due to the amount of publicly available information, companies will be selected from the S&P 500 and more specifically, from the "Information Technology" sector.

The reasoning behind this decision is that the latter sector accounts for approximately 27% of the market capitalization of the S&P 500. Furthermore, not only is the technology sector one of the most important in terms of market capitalization but also in terms of earnings weight (Gilmartin, 2023). The companies selected for this research (listed in Table 3) are the top 10 constituents of the S&P 500 "Information Technology".

Table 3: Top 10 Constituents of the Information Technology Sector S&P 500 (S&P Dow Jones Indices, 2023)

Constituent	Symbol	Sector
Apple Inc.	AAPL	Information Technology
Microsoft Corp	MSFT	Information Technology
Nvidia Corp	NVDA	Information Technology
Broadcom Inc	AVGO	Information Technology
Salesforce, Inc.	CRM	Information Technology
Cisco Systems Inc	CSCO	Information Technology
Accenture plc A	ACN	Information Technology
Adobe Inc.	ADBE	Information Technology
Texas Instruments Inc	TXN	Information Technology
Oracle Corp	ORCL	Information Technology

Source: own elaboration.

Regarding the time frames for data collection, the training of the meta-model requires significant amount of information. Consequently, this research will employ historical data spanning from 2000 to 2019. The rationale behind this decision is the fact that Covid-19 caused unprecedented disruptions in stock markets (Gherghina, 2023). In other words, the market volatility, policy interventions and worldwide instability had such a profound impact on stock markets that data spanning Covid-19 years is probably not representative of the underlying dynamics of EPS. As a result, the meta-model will only include pre-Covid-19 data to effectively capture the long-term trends and patterns of EPS estimates with a view to developing a more robust and reliable model.

Concerning individual analysts' EPS estimates, these will consist of 2-year-ahead EPS estimates since they are the "most prevalent forecasts of earnings issued by analysts" (Jung et al., 2017, p.434).

3.3 Methodology breakdown

Considering the complexity inherent in the research, providing basic guidelines regarding the selected methodology is the crucial catalyst to enabling the inquisitive reader to understand the underlying rationale behind the decision-making process for the design of the "model of models".

3.3.1 Proving the relevance of an enhanced consensus approach

Prior to diving into the development of the meta-model, a more quantitative context will be provided. Consequently, one of the main sub-sections of the "Analysis" section will be dedicated to examining the accuracy of consensus EPS estimates as well as the effect of EPS surprises and estimates' revisions on stock prices. In brief, the aim of this section is to further enlighten the rationale behind the purpose of the present research by providing relevant figures and data on the subject.

3.3.2 Collecting and choosing data for the meta-model

As aforementioned, meta-models need large amounts of data in order to properly identify patterns and make accurate predictions. In brief[2], the main data set[3] consists of individual

[2] A more thorough explanation of the data set collected for the training of the "model of models" is provided in section 4.
[3] Accessed through Wharton Research Data Services.

analysts' 2-year-ahead EPS estimates for each of the companies selected for the research. When referring to individual analysts, it is worth reiterating that the number of analysts that usually cover a public company (such as Apple Inc.) normally range from 1 to 50 (Jensen and Jones, 2020, p. 406). Clarifying this concept is important considering that their 2-year-ahead EPS estimates will be the ones collected for each of the companies and earnings season.

Given the quality of the referenced data, the training of meta-model should not require any further data. However, "when different models are based on different algorithms and data, on can receive essential gain in the accuracy" (Pavlyshenko, 2019, p. 1). In other words, the greater the diversity of the data, the better in terms of forecasting accuracy. Consequently, the present paper will also explore the idea of introducing macroeconomic factors into the training data.

3.3.3 Selecting the appropriate ensemble algorithm for the meta-model

In machine learning, two approaches outperform traditional algorithms: deep learning and ensemble methods (Mohammed and Kora, 2023). The rationale behind the ensemble methods is rooted in the Condorcet Jury Theorem, which was first introduced by Condorcet and Caritat (1781). In brief, this theory states that a pool of individuals has a greater chance of selecting the better of two alternatives in a context of uncertainty (neither of the two alternatives is preferred) than any single individual (Austen-Smith and Banks, 1996). As elucidated by Cunningham (2007, p. 1):

"If each voter has a probability p of being correct and the probability of a majority of voters being correct is P, then p > 0.5 implies P > p. In the limit, P approaches 1, for all p > 0.5, as the number of voters approaches infinity".

From this statement it can be inferred that the probability of being correct will increase as the ensemble grows in terms of size and diversity. The above rationale was later applied in the context of machine learning by introducing the concept of "ensemble learning methods". In summary, these are based on the assumption that the combination of various models can result in a more robust algorithm in terms of accuracy (Rocca, 2021). These base models are often regarded to as "unstable learners" (i.e., Neural Networks, Decision

Trees etc.) given that slight changes to the training data may result in different results and predictions (Cunningham, 2007).

But the question is, how do we combine these models? In brief, there are three main ensemble techniques, each of which are described in Table 4.

Table 4: Summary Ensemble Methods (Rocca, 2021).

Ensemble method	Description
Bagging	Trains a suite of homogeneous weak learners. The training is done independently from each other in parallel. Results are then combined using a deterministic averaging process.
Boosting	Considers a suite of homogeneous weak learners but trains them sequentially so that each base models depends on previous ones. Results are then combined using a deterministic strategy.
Stacking	Trains two or more heterogeneous weak learns and train them in parallel. Results are then combined by training a meta-model whose predictions are based on the different weak model predictions.

Source: own elaboration.

As it can be observed, stacking differs from boosting and bagging ensemble methods in two aspects (Rocca, 2021). First of all, stacking trains a series of heterogeneous weak learners while boosting and bagging consider only one type. Taking into account that each machine learning model makes different assumptions about the modeling task, it is often a good idea to train a range of models with different skills on the problem at hand. Secondly, stacking uses a meta-model to combine the predictions of the base models, unlike the other ensemble methods that use deterministic algorithms.

Since stacking combines the capabilities of a range of base models, it "usually performs better than all trained models" (Mohammed and Kora, 2023, p. 764). Indeed, using stacking methods to enhance modelling performance is not uncommon, with many authors such as Divina et al. (2018) and Qiu et al. (2014) applying stacking machine learning methods to forecast electric energy usage in different parts of the world. Other examples in the field of finance include Pavlyshenko (2019), who used a stacking approach to predict inventory demand based on historical sales data. Again, he concluded that staking machine learning methods can “improve the performance of predictive models” (Pavlyshenko, 2019, p. 257). Moreover, some authors such as Syarif et al. (2021) have even concluded with the superiority of stacking methods compared to bagging and boosting approaches.

Since the aim of this research is to arrive at a better consensus methodology for EPS estimates, using a stacking ensemble method seems the most suitable approach. By pursuing this course of action, a range of base models with different capabilities and skills on the dataset will be combined, thus reducing the time and resources spent on selecting a single algorithm for the problem at hand. In summary, time and resources will be used more efficiently by leveraging the strengths of a suite of base models and exploiting their potential to enhance modelling performance.

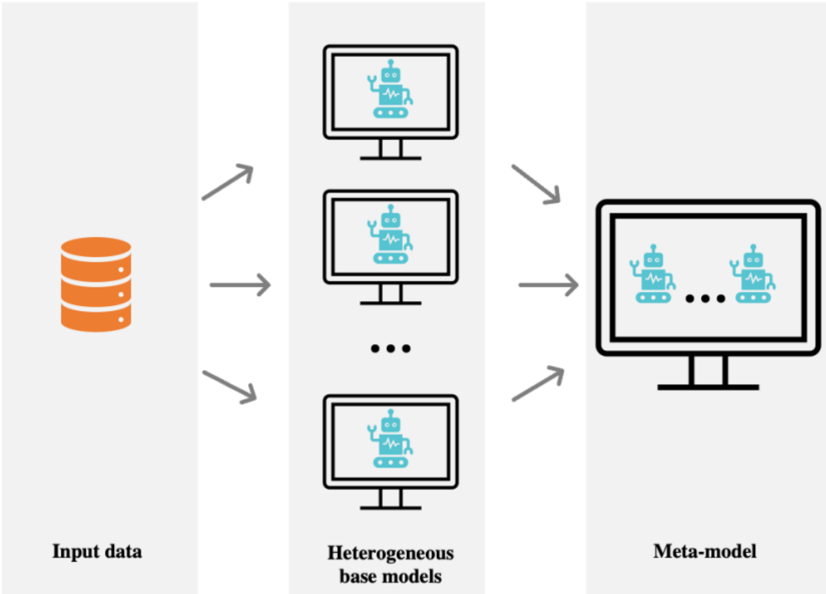
3.3.4 Stacking methodology

As aforementioned, stacking uses “a meta-learning algorithm to learn how to best combine the predictions from two or more base machine learning algorithms” (Brownlee, 2021, p. 379). With a view to providing the inquisitive reader with a brief overview of this ensemble algorithm, the present section will be dedicated to outlining the necessary steps involved in the development of the meta-model.

The first and most basic step when facing an algorithm is to clean and prepare the input data. This stage “deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data” (Rahm and Do, 2000, p.3). Nevertheless, the process of cleaning data in this research won’t be as time consuming given the quality of data available in the sources accessed (Wharton Research Data Services, Refinitv etc.).

Once the data has been properly prepared, the base models to be included in the ensemble algorithm must be chosen. In this context, it is worth mentioning that the architecture of a stacking algorithm involves two or more base models (level 0 models), whose predictions are then combined using a meta-model (level 1 model) (Mohammed and Kora, 2023).

Figure 4: Stacking models.



Source: own elaboration.

This step is highly relevant considering that choosing the appropriate number and range of level 0 models in the stacking ensemble is the crucial catalyst to generating better predictions (Mohammed and Kora, 2023). A guiding principle for selecting the base models to include in the stacking ensemble is that they should have diverse skills on the dataset so that their predictions (or errors) have low correlation. Since the aim of the meta-model is to improve the consensus EPS estimate of individual analysts, the base models will be based on predictive modeling algorithms such as:

- Linear regression.
- Decision trees.
- Gradient boosting machines.
- Neural networks.
- …

Having chosen the level 0 models[4], the next step will be to arrange the training and test sets. Since the meta-model will be dealing with time series, preserving the temporal order of the data is vital. For this reason, the variables will be sorted in ascendent order based

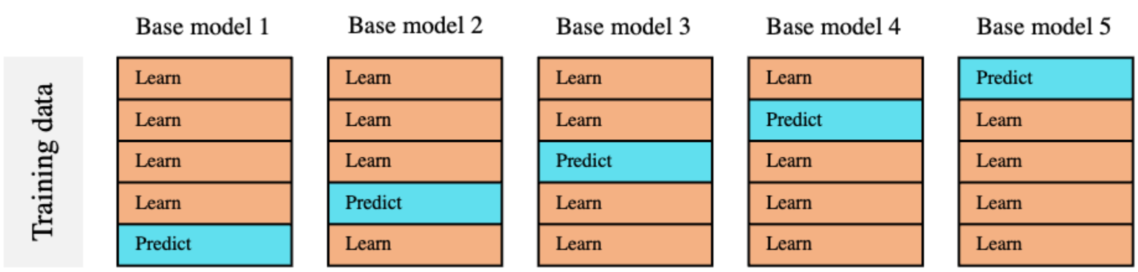
[4] The base models to be included in the stacking model as well as the reasoning behind their selection is further developed in following sections of the research.

on the dates to ensure the temporal order is preserved when splitting the dataset. For this meta-model, 80% of the values will be allocated to the train set and the remaining 20% to the test set.

Considering that more than one base model is included in the meta-model, further data splitting is required. With a view to arranging the train and test sets for the base models, k-fold cross-validation, which is one of the most common approaches when building a stacking model, will be employed. Examples of authors that have opted for this technique in the context of stacking include Chatzimparmpas et al. (2021), He et al. (2022) and Kalagotla et al. (2021). The rationale behind using a k-fold cross-validation technique is to reduce the risk of overfitting, one of the main issues of stacking techniques (Brownlee, 2021).

By using k-fold cross-validation, the base models will be trained on data that was split into K folds. In other words, level 0 models will be trained on all but one of the subsets (K-1 folds). The remaining fold for each of the base models will then be used for the predictions of the base models.

Figure 5: K-fold cross validation.

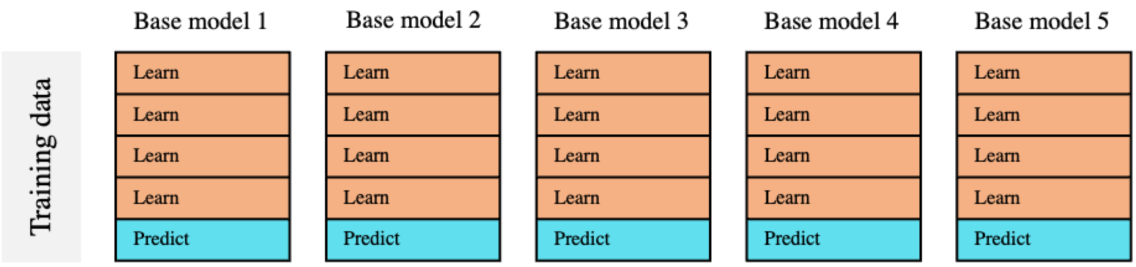


Source: own elaboration.

However, k-fold cross-validation normally splits the data randomly, thus ignoring the temporal dependency between observations (Shrivastava, 2020). Consequently, an alternative approach will also be tested, which will involve training the base models without using k-fold cross-validation. In brief, the two approaches to be applied in the development of the meta-model include:

- K-fold cross-validation: the base models will be trained using k-fold cross-validation in the train test as illustrated in Figure 5.
- Simplified approach: since the stacking algorithm must be trained using the predictions generated from the base models in the train set (not the test set considering these will be employed to make final predictions), the train set will be split as illustrated in Figure 6. In brief, the train set will be divided once again, allocating the first 80% of the data to the "training folds" and the remaining 20% to the "validation fold", thus preserving the temporal order of data[5].

Figure 6: simplified approach.

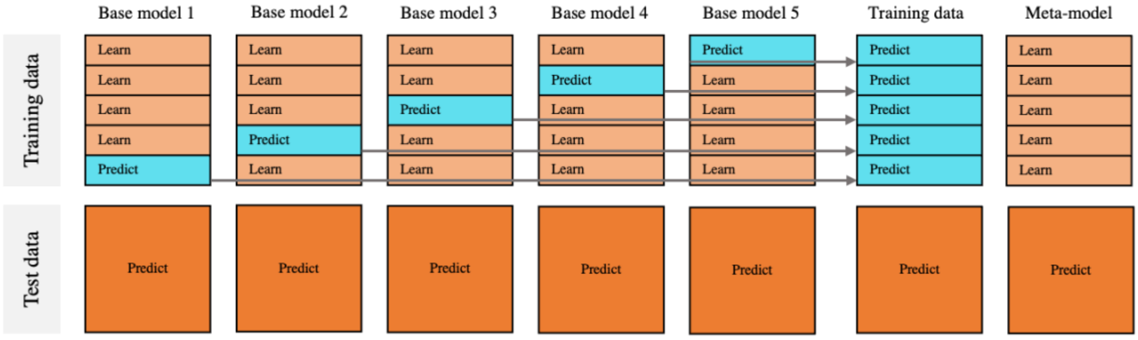


Source: own elaboration.

Assuming that the base models will be trained using k-fold cross-validation, the out-of-fold predictions of the base models will be combined along with the expected output (actual EPS values) to assemble the training data for the meta-model, which will learn how to combine the base models' predictions best (Mohammed and Kora, 2023). When training the meta-model, a linear regression will be used with the purpose of reducing the risk of overfitting. Once the meta-model has been trained, the test data will be used to generate predictions with the base models. Again, these predictions (along with the EPS actual values) will be fed to the meta-model to make final predictions. To better illustrate the process of building a stacking model, Figure 7 is presented[6].

[5] Only applicable if the "simplified" approach performs well.
[6] Only applicable if k-fold cross-validation proves to perform well.

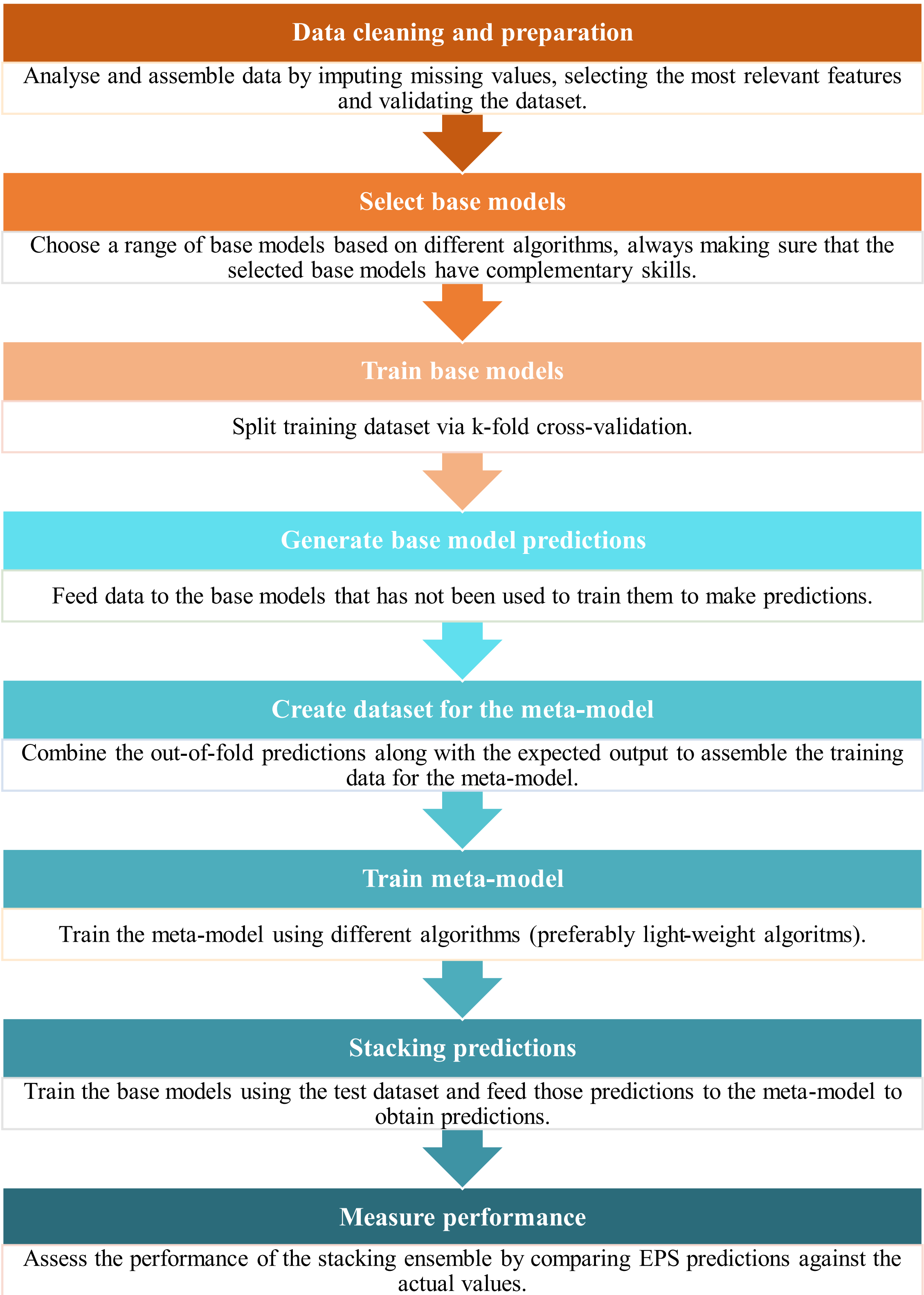
Figure 7: Stacking model process.



Source: own elaboration

The last step will be to measure the performance of the stacking ensemble by comparing EPS predictions against the actual values. The accuracy of the ensemble model will then be compared against the accuracy of consensus EPS estimates (which was calculated in previous sections) with a view to assessing whether building a better consensus methodology for EPS estimates is viable and worth exploring in further detail.

Figure 8: Theoretical Stacking methodology.



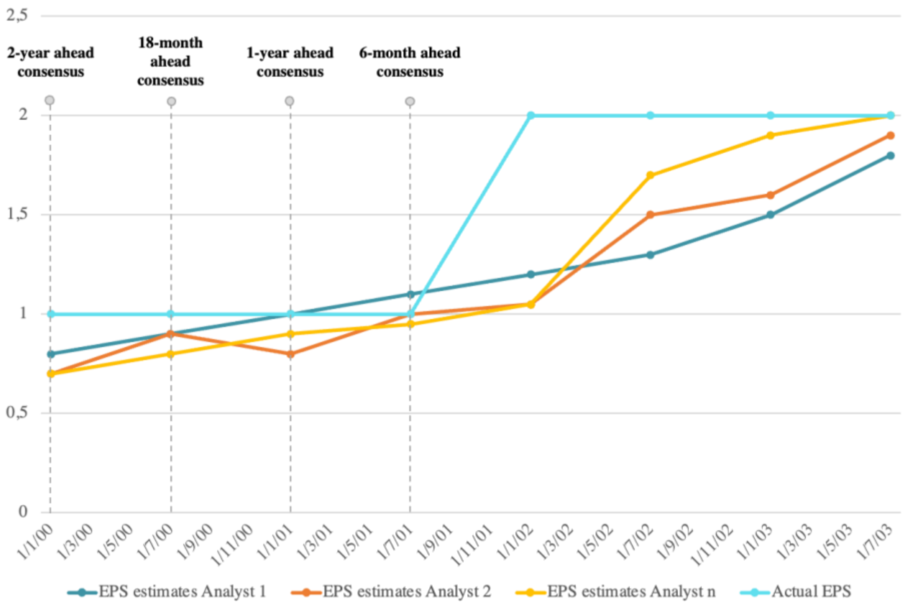
Source: own elaboration.

3.3.5 Selecting suitable base models

As aforementioned, selecting the "appropriate number of baseline models and the baseline models" is of particular importance when designing a stacking model (Mohammed and Kora, 2023, p. 764). In this context, level 0 models must have different skills on the dataset with a view to generating uncorrelated predictions and benefiting from their diversity.

Considering both, the vast array of machine learning algorithms and the different purposes they serve, choosing the most suitable algorithm for the task at hand is quite a complex task. The first step to choose the right algorithms for the proposed meta-model is to review the objectives and requirements of the research. As aforementioned, the purpose of the present paper is to develop a more sophisticated consensus approach beyond that of simple mean regarding EPS estimates. The methodology proposed to accomplish this objective involves training a set of base models to aggregate analysts' EPS estimates at different points in time by forecasting the EPS value of the annual earnings announcement date. This approach ensures that consensus EPS estimates will be computed taking into account the actual EPS, thus (potentially) enhancing the accuracy of the consensus.

Figure 9: Illustrative of the meta-model's variables.



Source: own elaboration.

Even though Figure 9 is only an illustrative of the variables that will serve as inputs for the dataset, it does provide a visual representation of the thought process behind the proposed methodology. In brief, analysts' estimates will be fed as inputs to the base models, which will be trained to predict actual EPS[7]. As a result, each EPS consensus will not be computed using the arithmetic mean but rather will consist of an EPS prediction using the individual analysts' estimates. This rationale is supposed to increase the accuracy of EPS consensus estimates.

Having clarified the objective and design of the meta-model, it is evident that the base models must consist of forecasting algorithms with a view to analyzing historical individual EPS estimates and making actual EPS predictions to enhance the accuracy of the consensus. For this purpose, the following supervised ML algorithms will be selected:

- Linear regression.
- SVM regression.
- Random Forest regression.
- KNN regression.

One of the main reasons behind choosing this set of predictive algorithms is the fact that they possess diverse and complementary skills, which increases the probability of generating uncorrelated predictions (imperative in the context of stacking). Without going into excessive detail, Table 5 provides a brief overview of the selected base models while highlighting the main differences between them.

Table 5: Overview of Base Models.

Supervised ML Algorithm	**Brief description**
Linear regression[8]	Studies the linear relationship between a dependent variable Y (in this case EPS values) and one or more independent variables (individual analysts' EPS estimates, macroeconomic factors etc.)

[7] Actual EPS refers to the value EPS of the earnings announcement date. This value represents the EPS that analysts attempt to estimate.
[8] Suitable due to the linearity between analysts' EPS estimates and actual EPS.

	(Schneider et al., 2010) to make predictions.
SVM regression	"Makes predictions from a geometric optimization problem that can be written as a convex quadratic optimization problem with linear constraints" (Martín Guareño, 2016, p.7). In other words, it identifies and selects the regressor hyperplane that best fits the training data.
Random Forest regression	Makes predictions by aggregating the predictions (majority vote or averaging) of the ensemble of regression trees generated by using bootstrap samples and random feature selection ().
KNN regression	Non-parametric technique that was originally used as a classification method but expanded to regression. In kNN regression, "the input consists of the k-closes training examples and the output is the average of the values of the k-nearest neighbors" (Hu et al., 2022, p.2).

Even though they are all regression algorithms, they use different techniques and skills to generate predictions, making them relatively uncorrelated base models. However, this set of algorithms will all face the same issue: the multicollinearity of the input data. Even though this issue is addressed in the following section, acknowledging that individual analysts' EPS estimates are highly correlated is vital for understanding the limitations that will be encountered when developing the meta-model.

Even though the performance of the base models will be probably affected by the multicollinearity inherent in the dataset, SVM and random forest regression might perform better than the others. On the one hand, "in the case of multivariate and mixture of distributions, SVM performs better than LR when high correlation structures are observed in the data" (Salazar et al., 2012, p.234). On the other hand, random forest regression algorithms "show high predictive accuracy and are applicable even in high-dimensional problems with highly correlated variables" (Strobl et al., 2008, p. 1). This conclusion was also drawn by Belgiu and Drăguţ (2016), who concluded that RF were quite successful at handling, not only high dimensional data, but also multicollinearity.

Even though PCA could potentially reduce the effects of multicollinearity, it will not be employed for two main reasons:

- Lack of high-dimensionality data: since the dataset is fairly manageable and the computational time is relatively low, PCA is not necessary concerning the reduction of complexity.
- Interpretation difficulty: PCA captures the underlying patterns, thus complicating the interpretation of variables. Considering that capturing which analysts are better at predicting EPS might improve the performance of the meta-model, PCA analysis may not be the most appropriate approach.

4. DATA

4.1 Data sources

Due to the aim and quantitative nature of the present paper, a significant amount of data has been collected from multiple and reliable data sources with a view to ensuring its quality. A list of the data sources employed as well as the information collected is provided in Table 6.

Table 6: List of data sources and data collection.

Data source	Data collected	Use
S&P Dow Jones Indices (2023)	Top 10 constituents in terms of market capitalization of the S&P 500 "Information Technology" sector.[9]	Companies whose EPS estimates will be included as input data in the meta-model.
World Bank Open Data (2023)	US historical inflation, GDP and real interest rates.	Analyze whether to include macroeconomic

[9] Data from the rest of data sources has been collected for each of the companies selected for the research.

		variables as input data for the meta-model.
Refinitiv Workplace (n.d.)	Historical prices of the subset of companies selected for the research.	Analyze effects of earnings surprises and EPS revisions on prices.
Wharton Research Data Services (n.d.)	Historical EPS consensus estimates when earnings were published.	Analyze effects of earnings surprises on stock prices.
	Historical EPS.	Target value for the predictions.
	Historical revisions of EPS consensus.	Analyze effects of EPS revisions on prices.
	Historical analysts' EPS estimates.	Input data for the meta-model.

Source: own elaboration.

4.2 Data preprocessing

While all collected data is relevant for the present research, "historical analysts' EPS estimates" are the crucial catalyst to building the stacking ensemble considering that they will employed as input data for the meta-model. Consequently, more emphasis must be placed on exploring and preparing this dataset.

As aforementioned, the main dataset consists of individual analysts' 2-year-ahead EPS estimates for each of the companies selected for the research spanning from 2000 to 2019. These estimates are "collected each day as they are released by analysts" (Dai, 2020, slide 7). It is worth noting here that I/B/E/S tracks "street" or "pro forma earnings". Recalling the types of EPS, "pro forma" refer to earnings that exclude some expenses or income to allow better comparison.

The original dataset obtained from WRDS was comprised a total of 27 variables. However, this dataset has been reduced to 6 variables, which are listed and briefly explained in Table 7.

Table 7: Variables input data for the meta-model.

Variable	Description
Official Ticker Symbol	Unique combination of letters used to identify a publicly traded company.
Analyst Code	Unique identifier assigned to individual analysts within a financial institution.
Estimate Value	EPS estimate of the analyst at the time the forecast was reported.
Announce Date, SAS Format	Date in which the EPS forecast was reported.
Announce date of the Actual, from the Actuals File	Date for which the EPS was forecasted.
Actual Value, from the Detail Actuals File	Actual value of the EPS for the date EPS estimates were forecasted.

Source: own elaboration.

With a view to providing the inquisitive reader with an overview of the pre-processed dataset as well as the variables to be considered, a data sample will be shown in Table 8 using AAPL as an example.

Table 8: Pre-processed data.

Symbol	Analyst Code	Estimate Value	Announce Date, SAS Format	Announce date of the Actual	Actual Value
AAPL	70648	0,0241	2000-10-19	2002-10-16	0,0059
AAPL	10014	0,0187	2000-10-19	2002-10-16	0,0059
AAPL	92159	0,0259	2000-10-19	2002-10-16	0,0059
AAPL	70648	0,0089	2000-12-06	2002-10-16	0,0059
AAPL	92159	0,0196	2000-12-06	2002-10-16	0,0059
AAPL	10014	0,0045	2000-12-06	2002-10-16	0,0059
AAPL	1047	0,0286	2001-01-12	2002-10-16	0,0059

Source: own elaboration.

Even though the original dataset contains all required information, it needs to be transformed into a format to be properly processed by the meta-model. The main issues regarding the format of the original dataset are:

- The number of individual analysts vary across the time frame. For instance, an analyst started estimating EPS from APPL in 2002 but stopped in 2005, which results in missing values.
- Analysts don't generate the same number of estimates for a specific announce date. For example, analyst with code 10014 might make 5 predictions for 2002 EPS announcement but analyst with code 1047 might generate 20 predictions. This lack of standardization results, once again, in missing values.
- Analysts' estimates, although close in time, are not generally reported in the same dates, another standardization issue which might lead to further missing values.

Taking the above into account, the original dataset for the meta-model was transformed into the following format:

Table 9: Processed Data.

Official Ticker Symbol	Announce Date Forecasts	Announce date of the Actual, from the Actuals File	Actual Value, from the Detail Actuals File	40709	9834	Days until Announce-ment Date
AAPL	1999-10-18	2001-10-17	-0,0048	0,0272	0,0346	730
AAPL	2000-03-11	2001-10-17	-0,0048	0,0357	0,0362	585
AAPL	2000-10-17	2001-10-17	-0,0048	0,0357	0,0295	365
AAPL	2001-04-17	2001-10-17	-0,0048	0,0357	0,0295	183
AAPL	2001-09-17	2001-10-17	-0,0048	0,0357	0,0295	30
AAPL	2000-10-16	2002-10-16	0,0059	0,02	0,016645	730
AAPL	2001-03-10	2002-10-16	0,0059	0,0137	0,014665	585
AAPL	2001-10-16	2002-10-16	0,0059	0,0109	0,01231	365
AAPL	2002-04-16	2002-10-16	0,0059	0,0109	0,01231	183
AAPL	2002-09-16	2002-10-16	0,0059	0,0109	0,01231	30

Source: own elaboration.

As it can be observed, for each annual EPS announcement date from the "Announce date of the Actual" 5 different dates were computed for the "Announce Date Forecasts":

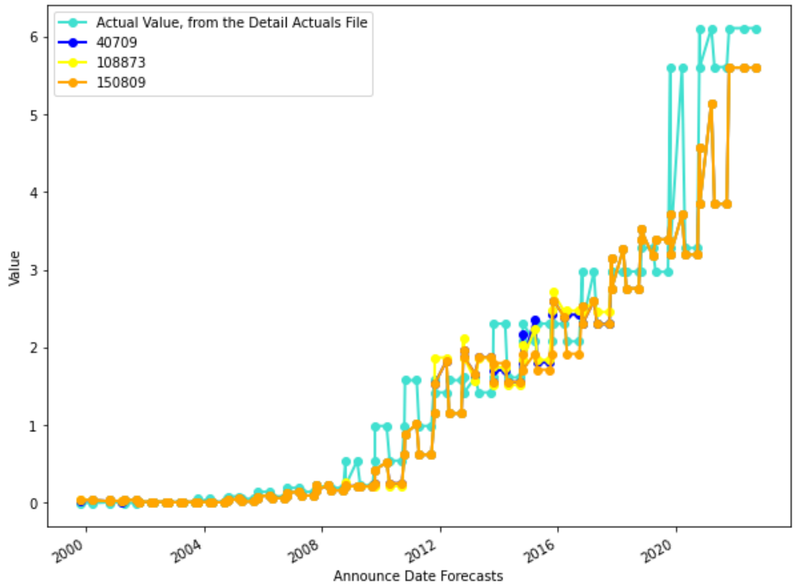
- 2-year ahead.
- 18-month ahead.
- 1-year ahead.

- 6-month ahead.
- 1-month ahead.

As aforementioned, one of the main issues of the original dataset was the fact that analysts don't record their EPS estimates at the same time. Consequently, the "Announce Date Forecasts" column was computed. The purpose of those dates is to serve as a common point in time for recording analysts' EPS estimates. This was achieved by selecting, for each analyst, the EPS estimate[10] that was recorded closest to the corresponding date in the "Announce Date Forecasts".

Since analysts don't usually generate predictions for 19 years straight for the same company, there were several missing values. For instance, analyst with code 40709 only made EPS predictions for AAPL from 2000 to 2003. Hence, from 2003 onwards only missing values recorded for the latter analyst. With a view to tackling this issue, missing values were filled using the EPS estimate mean average across analysts for the corresponding date. A visual representation of the processed data can be observed in Figure 11 (only includes data from 3 analysts from AAPL).

Figure 10: AAPL Processed Data.



Source: own elaboration.

[10] EPS estimate must have been recorded for the same "Announce date of the Actual".

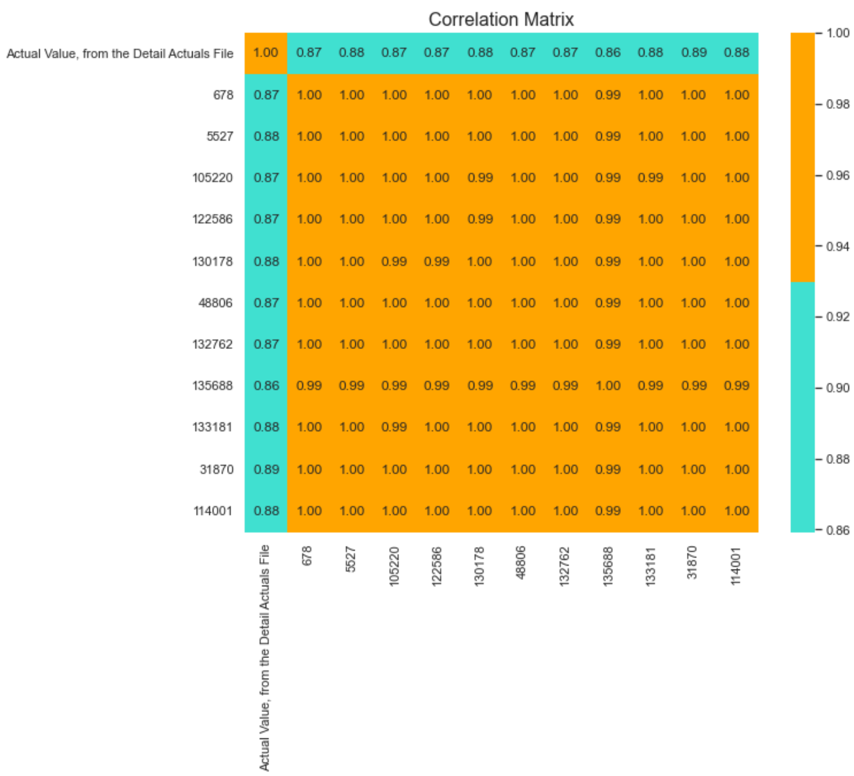
Last, the column “Days until Announcement Date” was included with a view to capturing the relationship between both, the “Announce Date Forecasts” and “Announcement Date of the Actuals” in the predictions.

However, after processing and cleaning the original dataset, individual analysts’ EPS estimates showed multicollinearity due to:

- All EPS estimates are aimed at predicting the same EPS value, thus leading to high correlation.
- This correlation is further enhanced to due how the dataset had to be processed in order to fill missing values.

In order to show the multicollinearity of data, the following correlation matrix was plotted using NVDA as the case study. For the sake of simplicity only 11 individual analysts were included.

Figure 11: Correlation Matrix NVDA.



Source: own elaboration.

Since the individual analyst variables exhibit significant multicollinearity between them, exploring additional variables that are not as highly correlated with a view to incorporating them in the meta-model is of particular importance. The rationale behind including other variables is the fact that they can help mitigate the effects of multicollinearity and improve the overall quality of the model.

As previously shown in Figure 1, EPS estimates are influenced by a combination of both, intrinsic company and macroeconomic factors. Consequently, the following variables will be incorporated in the meta-model:

- Gross Domestic Product (GDP): since GDP represents overall economic activity, it generally indicates a healthy economy. Hence, a country with high GDP might suggest higher revenues for companies, which can positively influence EPS.
- Interest rate: lower interest rates mean fewer borrowing costs for companies, which might result in lower interest expenses and enhanced growth, again potentially impacting EPS.
- Inflation rate: although inflation impacts companies in many different ways, high inflation leads to an increase in the cost of inputs, therefore diminishing companies' margins and (potentially) EPS.

Since the companies of interest for this research are based in the United States, these variables will also represent U.S. information. For further detail, the exploratory analysis conducted on the macroeconomic factors is included in the appendix section.

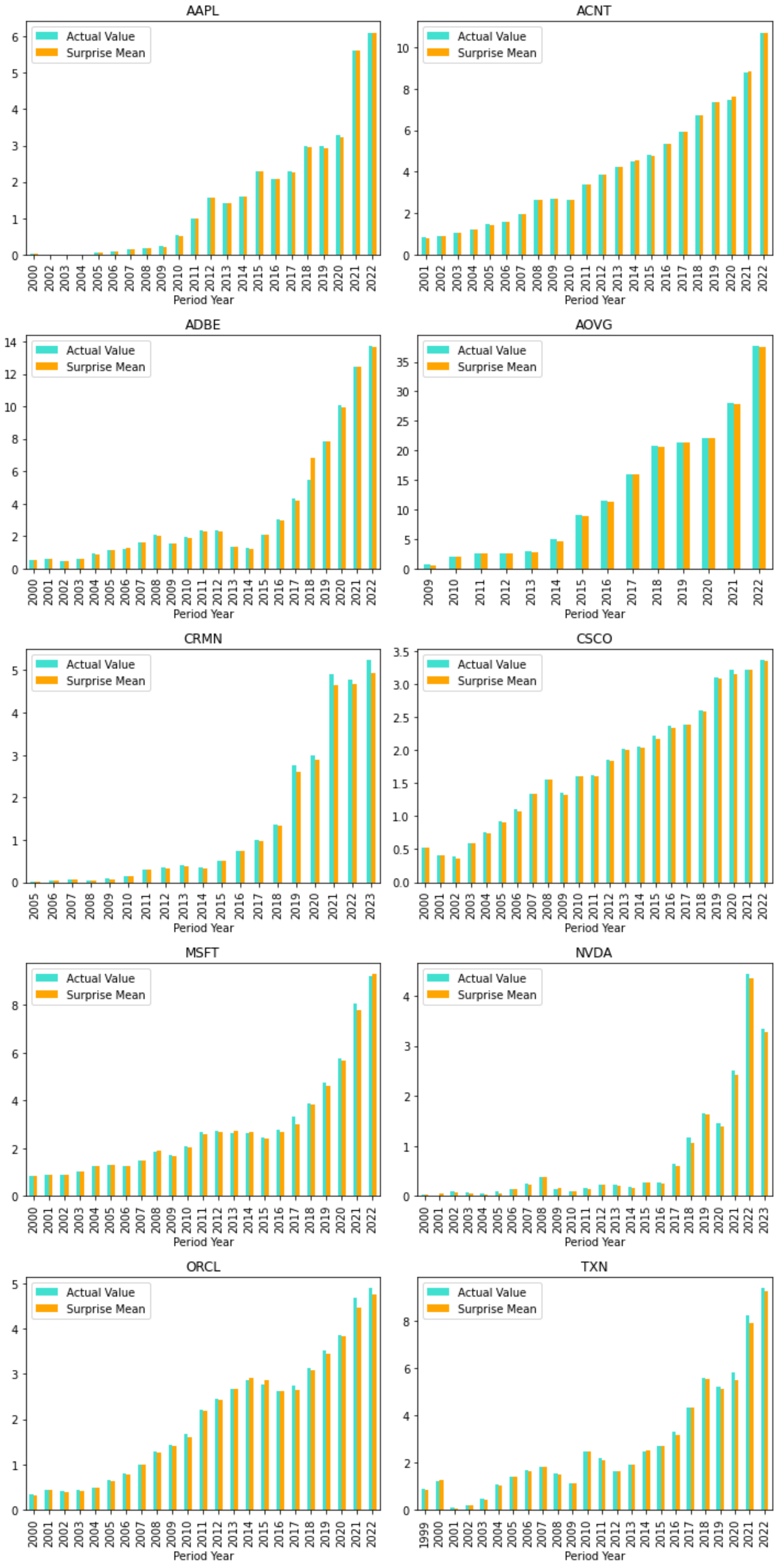
5. ANALYSIS

5.1 Relevance of the meta-model

5.1.1 Accuracy of consensus EPS estimates

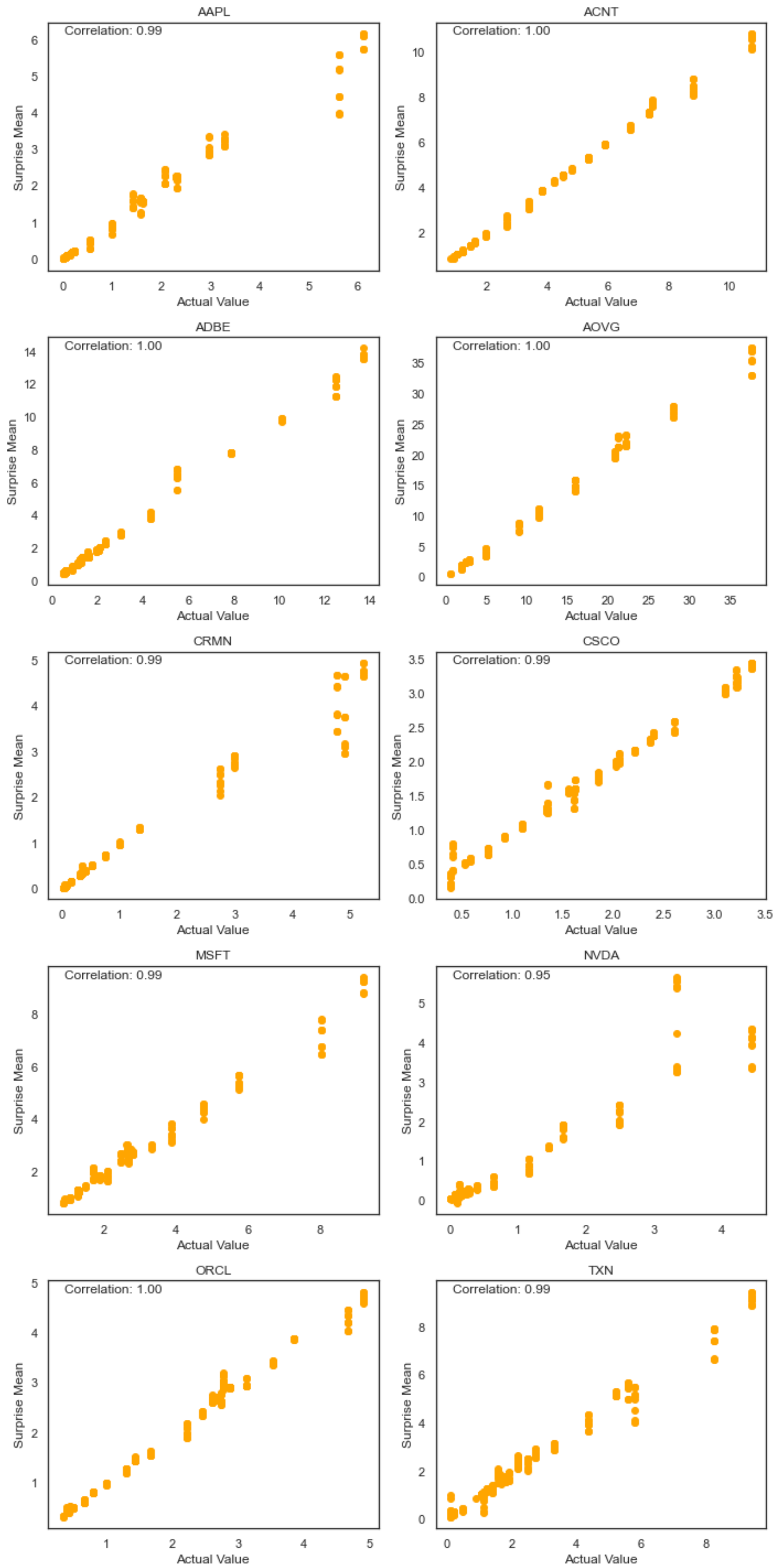
From the above, it is evident that consensus EPS estimates play a crucial role in financial stakeholders' minds and actions. Therefore, ensuring their accuracy is a matter of more than passing interest. With a view to assessing their accuracy, historical EPS (actual values) from the companies selected for this research is compared to the historical consensus estimate at the time of the earnings announcement in Figure 12.

Figure 12: Actual Values EPS against Consensus Estimates.



Source: own elaboration.

Figure 13: Scatter plot Consensus Mean and Actual Value EPS.



Source: own elaboration.

At first glance, it is obvious that the consensus EPS mean tends to deviate from the actual EPS reported by companies, denoting forecast errors in both analysts' and consensus estimates. Since these deviations can be observed throughout the 23-year period analyzed across the companies selected for this research, it can be inferred that analysts' forecast errors regarding EPS have a tendency to occur.

Despite forecast errors, Figure 13 shows the evidently high correlation between consensus EPS estimates and actual EPS value. This is quite straightforward considering that EPS estimates (both consensus and individual) all attempt to track and predict actual EPS, which explains the aforementioned multicollinearity of the dataset.

With a view to calculating the forecast error inherent in consensus EPS estimates, the mean error (ME) was calculated using the following formula:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)$$

In the above formula, y_i refers to the actual values of EPS while f_i represents the historical EPS consensus estimates. Even though this statistic might be biased due to the "offsetting effect of positive and negative forecast errors" (Rybnik, 2022), it sheds light on whether analysts are over or underestimating EPS.

It is worth noting here that individual analysts often make revisions to their EPS estimates (further explained in detail in following sections). Considering that the "surprise mean" represents the EPS consensus estimate at the time of the earnings announcement date, it also represents the last EPS revision for the considered period. Consequently, the "surprise mean" contains less biases and forecast errors than the original consensus EPS estimate for each forecast period.

With a view to properly analyzing consensus EPS estimates' accuracy, the present section will compute and compare the 1-year-ahead EPS consensus estimates and "surprise mean" forecast errors.

Table 10: Mean Error of "surprise mean" and 1-year ahead consensus estimates.

I/B/E/S Ticker Symbol	1-year ahead consensus Mean Error	"Surprise Mean" Mean Error
AAPL	0,116152	0,0148159

ACNT	0,123182	0,0019673
ADBE	0,112391	-0,0275557
AOVG	0,942857	0,1104864
CRMN	0,242224	0,0545879
CSCO	0,023913	0,0167709
MSFT	0,176087	0,0448465
NVDA	0,01719	0,0238125
ORCL	0,056957	0,0346074
TXN	0,586739	0,0617654

Source: own elaboration.

At first glance, it is obvious that 1-year ahead consensus EPS estimates are far less accurate than the "surprise mean" consensus. This is not surprising considering that the "surprise mean" consensus has been adjusted for all available information in the market and is thus more accurate. Another insight that was drawn from Table 10, which is similar to the conclusions that Hilary and Hsu (2013) derived from their studies, is the fact that consensus EPS estimates tend to be lower than the actual values of EPS.

Although the ME provides insights into how EPS is over or underestimated by analysts, it does conceal forecasting inaccuracies. With the purpose of tackling the inaccuracy of the ME, the mean absolute error (MAE) was calculated using the following formula:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i|$$

Advantages of this statistic include the "cancellation of errors with opposite signs" (Rybnik, 2022), making the MAE more reliable than the ME. Results are shown in Table 11.

Table 11: MAE 1-year ahead and "surprise mean" consensus.

Official Ticker Symbol	1-year ahead consensus estimate MAE	"Surprise mean" MAE
AAPL	0,22110	0,0160705
ACN	0,16682	0,0202964
ADBE	0,20196	0,0887678
AVGO	1,32143	0,1112807
CRM	0,26538	0,0549763
CSCO	0,11696	0,0169370
MSFT	0,33609	0,0672500
NVDA	0,24838	0,0296258

ORCL	0,12957	0,0472222
TXN	0,91196	0,0675779

Source: own elaboration.

Again, the 1-year-ahead consensus EPS estimate shows greater inaccuracies compared to the “surprise mean”. As it can be observed, forecast errors range from 0,0160705 for Apple to 0,1112807 for Broadcom Inc., with the average mean error for all companies analyzed being 0,052. Since EPS for these companies currently ranges from 3,5 to 35, forecasts errors appear to be relatively low.

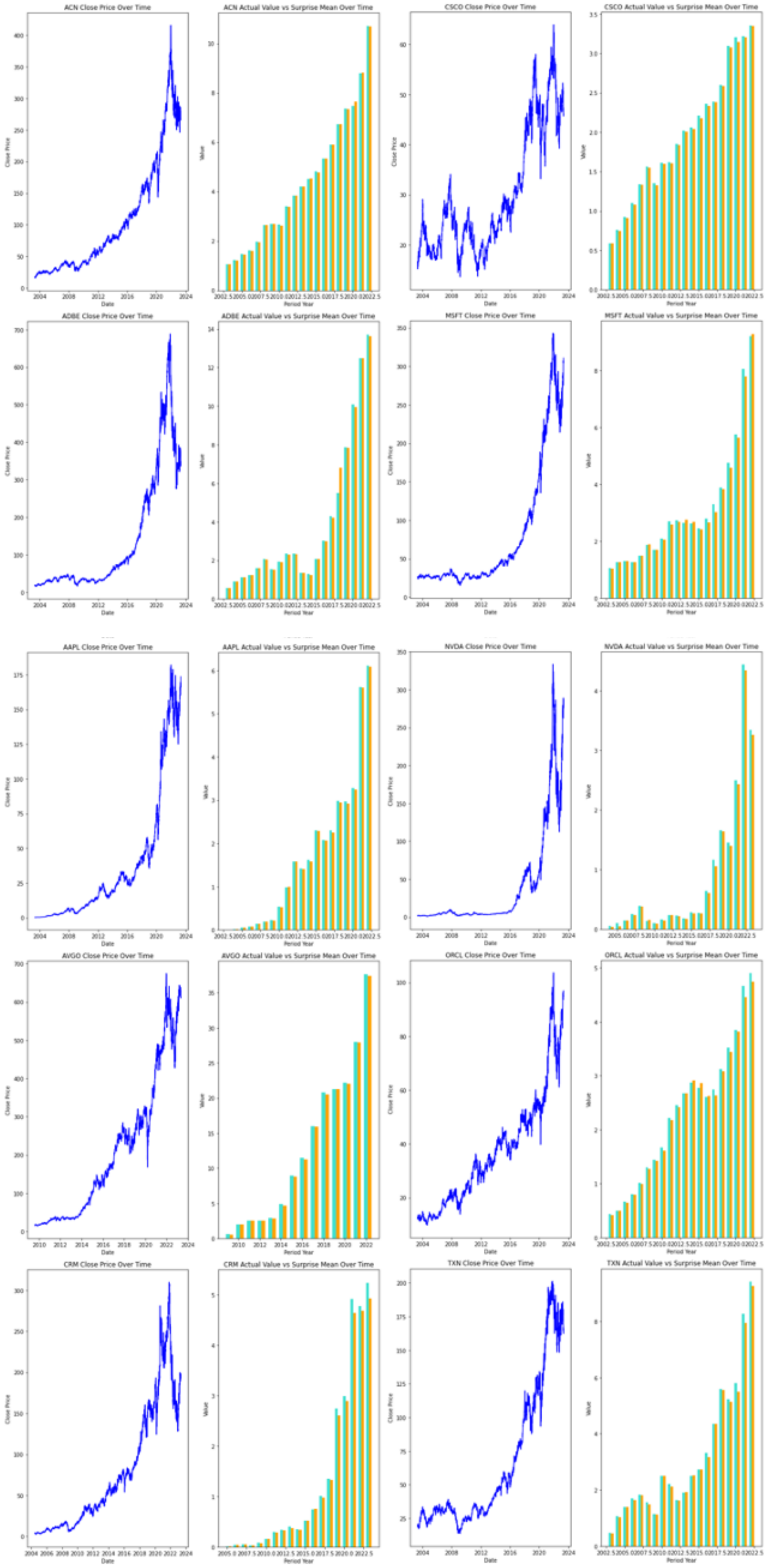
Considering that these estimates are generally computed using simple mean consensus methods, there is still room for improvement regarding the accuracy of consensus methods (particularly for consensus EPS estimates that are furthest away in time from the EPS announcement date).

5.1.2 Effect of surprises in stock prices

Since investors form their market expectations and make their investment decisions based on EPS consensus estimates, how do they react when the latter deviate from actual EPS reported by companies? According to the signaling theory, which was touched upon in previous sections, investors will re-consider their investment decisions based on the signals they receive from the publication of information. In the context of EPS, investors perceive better signals when companies report a higher-than-expected EPS and vice versa. In other words, a higher EPS might create a positive impression of the company in the minds of investors, thus leading to greater demand and ultimately, higher prices. However, a lower-than-expected EPS might depress stock prices.

Regardless of the type of signals investors receive, a market reaction is expected in the form of stock price fluctuations (Prijanto et al., 2021). With the purpose of analyzing how deviations affect stock prices, both reported EPS and consensus EPS estimates were plotted alongside stock prices in Figure 14 for each of the companies selected for this research.

Figure 14: Reactions in the Stock Price to EPS Deviations.



Source: own elaboration.

At first glance, it is obvious that there is a positive correlation between EPS and stock price over time. The strong correlation between EPS (reported and consensus estimates) and stock price can be observed in Figure 15:

Figure 15: Scatter Plot Stock Prices & Actual and Consensus EPS.

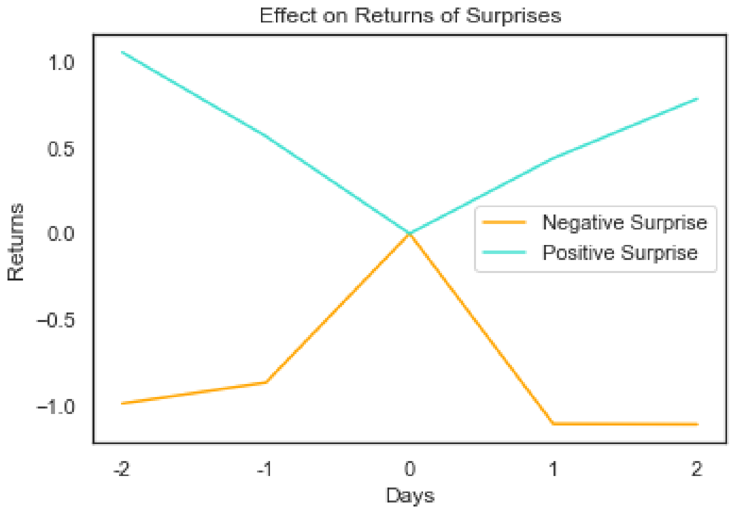
Source: own elaboration.

However, more detail is needed in order to properly analyze the relationship between these two metrics. With this purpose in mind, the evolution of returns based on the EPS surprise has been analyzed for different periods of time prior and since the EPS announcement date. More in detail, the evolution of returns has been classified into "positive" and "negative" surprises. In brief, Figure 16 represents the aggregate mean of daily changes in stock returns in two instances:

- "Positive surprise": actual EPS surpassed the expected (or consensus) EPS.
- "Negative surprise": actual EPS fell below the expected (or consensus) EPS.

For further guidance, the values of the x-axis represent the days prior and since the EPS announcement date, where 0 is the earnings announcement date. Following this line of thought, -2 represents the returns two days prior and so on. On a last note, Figure 16 aggregates information across all companies selected for this research.

Figure 16: Effects of EPS Surprises on returns.



Source: own elaboration.

As it can be inferred from Figure 17, when investors expect the EPS to underperform, stock returns are generally negative, and more so when EPS actually misses the mark (and vice versa).

The reasoning behind the change in returns can be explained using the "Efficient Market Hypothesis", which, as aforementioned, explains that prices reflect all information available information. However, many studies suggest that market information does not spread instantaneously because there are "many statistically significant lagged correlations" (Basnarkov et al., 2020, p.2). Nevertheless, Fama (1970) explored whether stock prices adjust efficiently newly available information (such as annual earnings announcement) and concluded that information did spread rapidly, thus affecting prices instantly and confirming the "Efficient Market Hypothesis". This can be observed in Figure 16, where stock prices (and hence returns) adjust automatically.

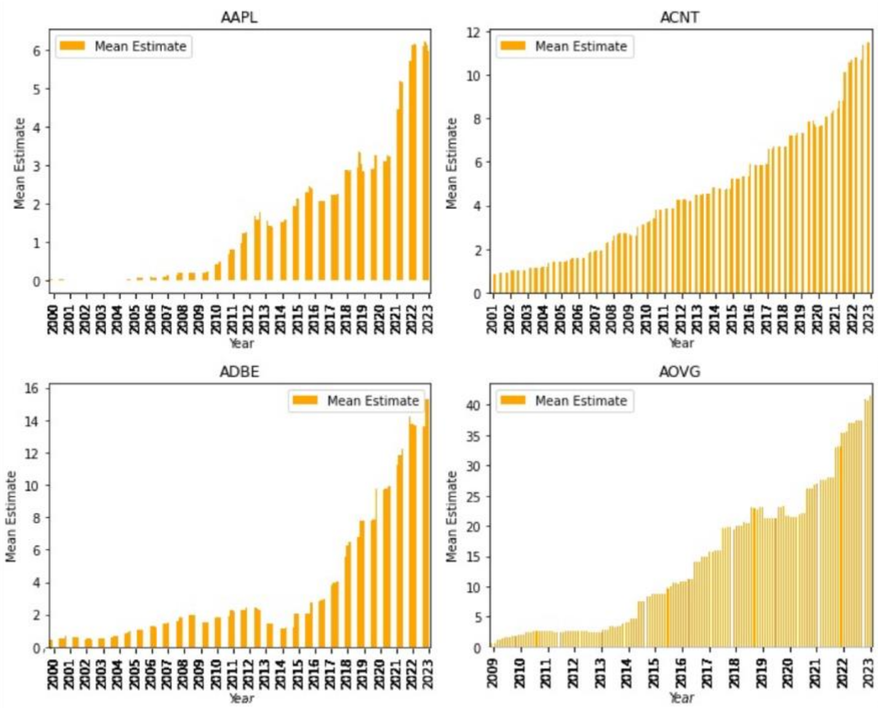
Acknowledging that EPS deviations affect stock prices is enough to highlight the importance of increasing the accuracy of consensus EPS estimates.

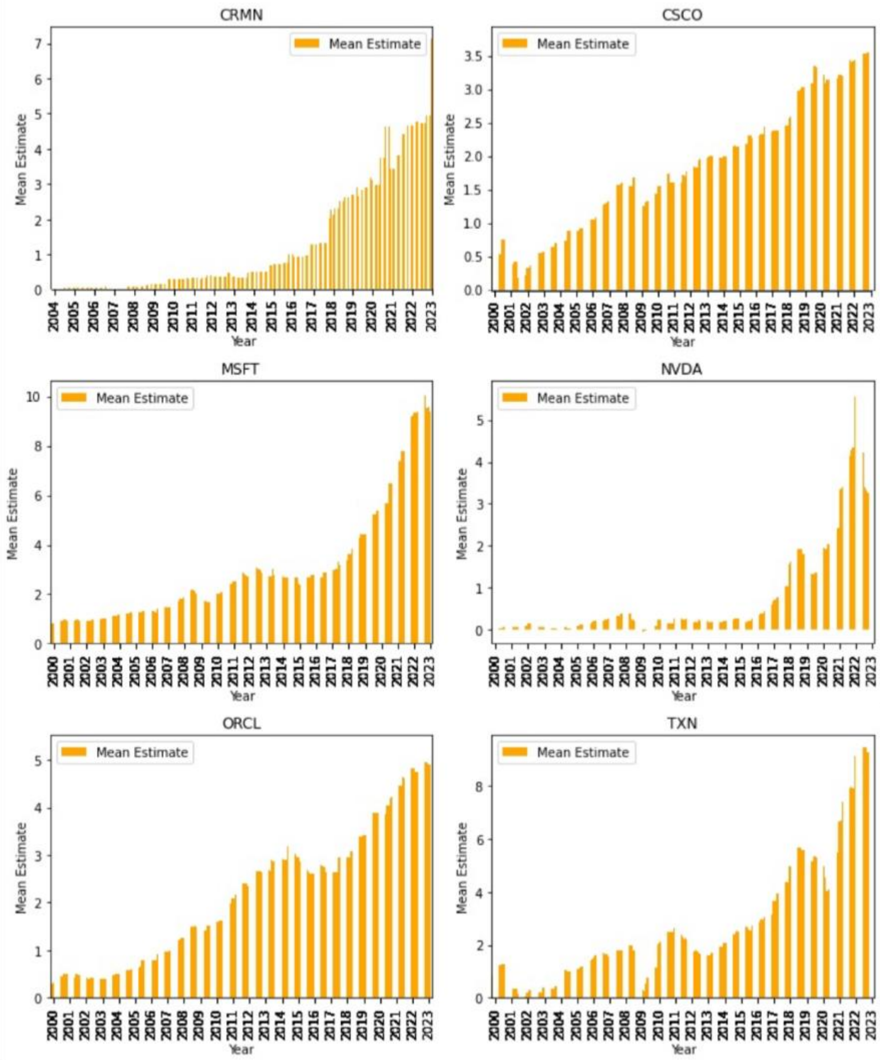
5.1.3 Effect of revisions in stock prices

Even though the analysis has focused on stand-alone consensus EPS estimates, revisions to EPS estimates are "of equal or greater importance because they reflect changes in expectations of a firm's future performance" (Jung et al., 2017, p. 434).

Analysts generally make revisions to their EPS estimates in order to adjust to changes in the market, the company's financial performance or other factors that may influence EPS. For instance, if a company announces the launch of a new product and an analyst believes that, due to this launch, the company will experience an increase in earnings in the coming year, the analyst may proceed to revise its EPS estimate upward. Alternatively, if the same analyst expects a company to encounter challenges regarding its financial performance, its EPS estimate might be revised downwards to reflect this change in earnings expectation. In the end, investors use these revisions to stay up to date regarding a company's expectations and make more informed investment decisions. Figure 17 shows the number of EPS revisions in terms of consensus estimates over time. It is worth mentioning that the below charts represent 1-year-ahead consensus EPS estimates.

Figure 17: Consensus EPS Revisions Over Time.

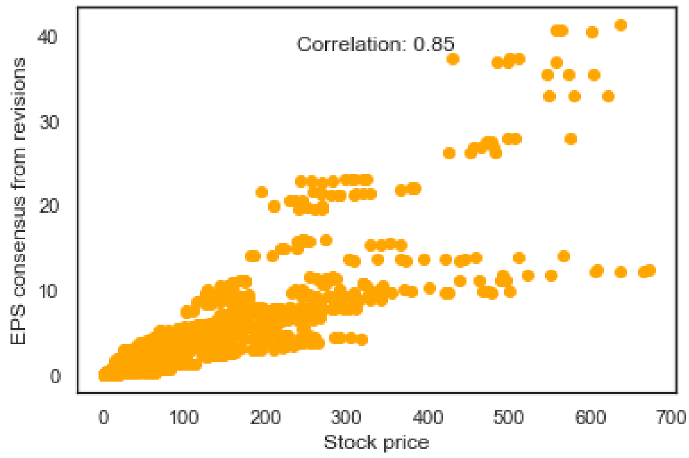




Source: own elaboration.

Considering that EPS revisions also guide investment decisions it seems logical to think that they also influence stock prices. In fact, authors such as Jung et al. (2017, p. 434) have stated that revisions are highly correlated with "contemporaneous changes in stock prices". With the aim of assessing the relationship between EPS revisions and stock price, the correlation between the two was calculated (aggregated correlation for all companies).

Figure 18: Scatter Plot Stock Prices & EPS Revisions.



Source: own elaboration.

As it can be observed, stock prices and EPS revisions show strong correlation suggesting that stock prices tend to move in the same direction as EPS revisions over the same time period. This conclusion is consistent with findings of other authors such as Ivković and Jegadeesh (2004), Copeland et al. (2014) and Jung et al. (2017), amongst many others. Having demonstrated with enough evidence that EPS estimates in all its shapes and forms influence business decisions as a whole as well as stock prices, the need to improve the accuracy of EPS consensus has become more self-evident.

5.2 Development of meta-model

All these findings lead us to the conclusion that ensuring the accuracy of consensus EPS estimates from the time they are first recorded until the earnings announcement is the crucial catalyst to enabling investors to make informed business decisions. Consequently, the aim of the meta-model is to explore whether the accuracy of consensus estimates can be enhanced. It is worth mentioning here that attempting to beat the accuracy of individual analysts' EPS estimates is out of scope considering the resources and time available for this research.

5.2.1 Base models performance

The first step involved in the development of the meta-model involves arranging the train and test sets. As aforementioned, considering that the meta-model deals with time series, preserving the temporal order of data is the crucial catalyst to enhancing of the performance of the model. Consequently, variables will be sorted in ascendent order based on the “Announce Date Forecasts”. Having done so, the target variable (Actual Value EPS) will be separated from the rest of features and then, the dataset will be split allocating 80% to the train set and the remaining 20% to the test set[11].

```
# Divide dataset into target and features.
features = df.iloc[:, 1:]
target = df.iloc[:, 0]
# Split the dataset into training and testing sets
train_size = int(len(df) * 0.8)
X_train, X_test = features[:train_size], features[train_size:]
y_train, y_test = target[:train_size], target[train_size:]
```

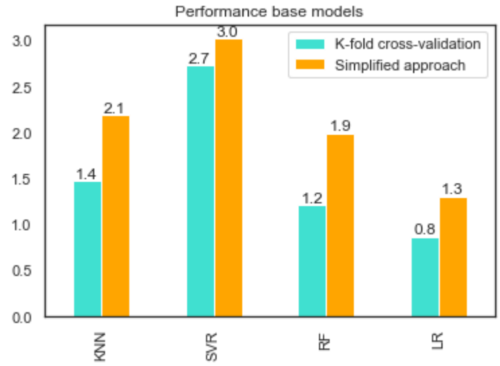
Having arranged the train and test sets, the base models will be trained using two different methods[12] with a view to selecting the approach that enhances the performance of the base models the most.

- K-fold cross-validation: a k-fold of 5 will be applied to the train set to train all 4 base models.
- Simplified approach: due to both, the low dimensionality of data and the limitations of k-fold cross-validation when dealing with time series, each of the base models will also be trained using the same data from the train set.

After training the base models, with and without k-fold cross-validations, predictions were generated on the test set with a view to assessing and comparing the performance of both approaches.

[11] Done to each of the companies using a loop in Python.
[12] Explained in detail in the methodology section.

Figure 19: Base models' MAE.

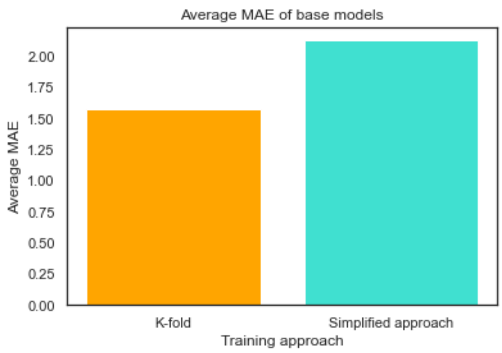


Source: own elaboration.

As it can be inferred from Figure 19, the base models seem to show consistent performance across both approaches, with SVM regression showing the highest MAE and linear regression the lowest. At present, the SVM regression will not be excluded from the set of base models for the purpose of generating as many uncorrelated predictions as possible for the stacking algorithm.

Another insight derived from the above figure is the fact that the base models might underperform when k-fold cross-validation is not applied. Reasons that explain why using k-fold cross-validation mat lead to better results despite not preserving the temporal order of data include the fact that the train data in the simplified approach was split "manually", not efficiently like in k-fold cross-validations. In other words, k-fold cross-validation allows the base models to be trained and evaluated on different subsets of data, thus making an efficient use of the available data, something vital in this research due to the low dimensionality of data.

Figure 20: Average MAE of base models in the training set.



Source: own elaboration.

Even though none of the approaches produce show strong performance, the meta-model will be trained using k-fold cross-validation since it shows lower predictive errors. It is worth mentioning here that, in order to reproduce the previous results, a random seed of 15 will be set.

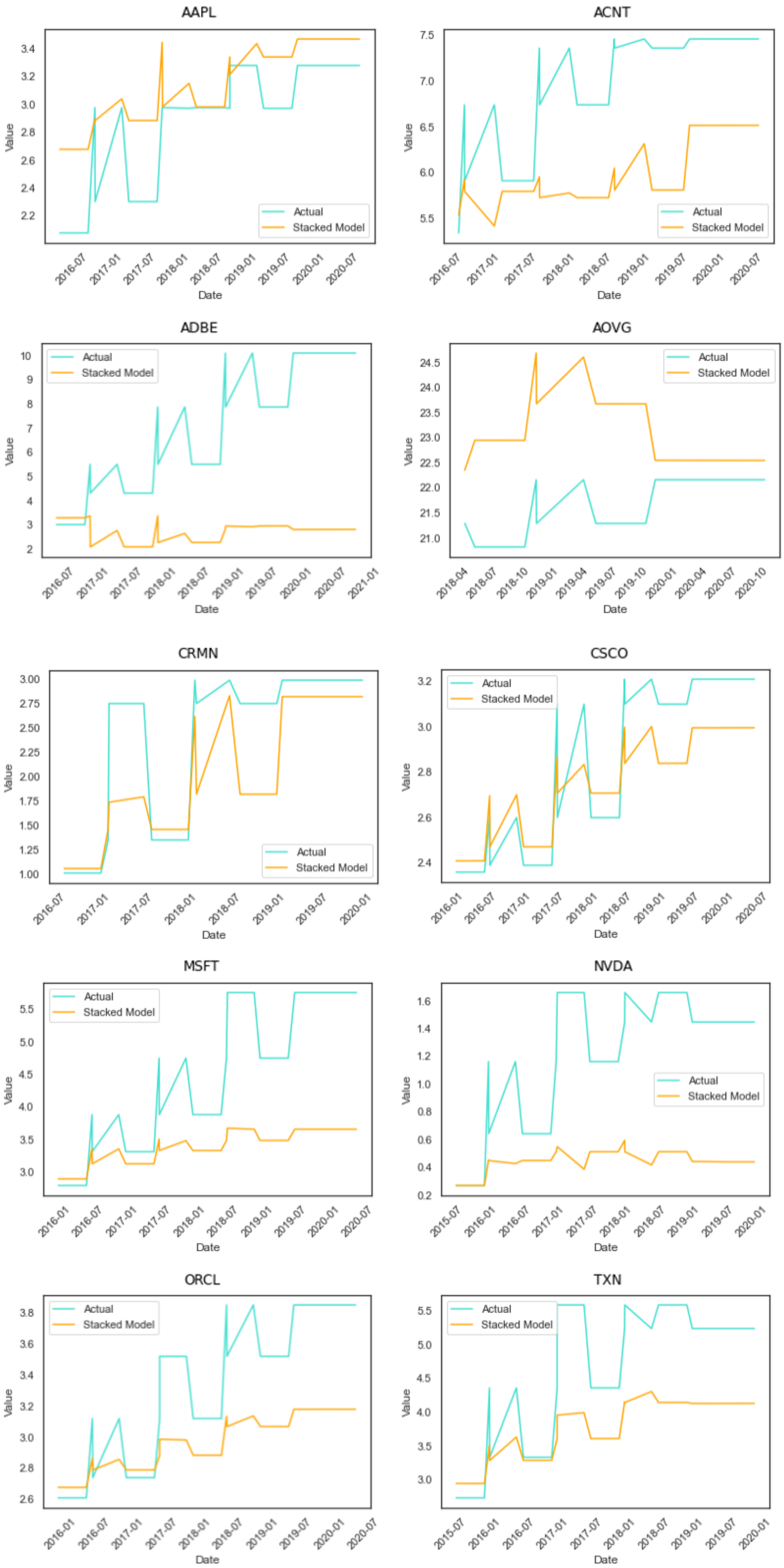
5.2.2 Consensus approach for the meta-model

For the sake of simplicity, the meta-model will be trained using k-fold cross-validation. Even though this is explained in further detail in the methodology section, a brief overview will be provided again with a view to guiding the inquisitive reader.

- Selection of the base models to then define a StackingRegressor (stacking algorithm) using the Scikit Learn's module. How the StackingRegressor is defined is fairly simple as it only requires setting the base estimators, a final estimator and the number of cross folds. As aforementioned, the stacking algorithm will consist of a linear regression with the purpose of reducing the risk of overfitting. Moreover, k-fold cross-validation will be set to 5 as done with the base models.
- The stacked algorithm is trained upon the predictions generated from the base estimators using the validation fold.
- The meta-model generates predictions using the predictions made by the base models in the test set.
- Since MAE has been the performance metric employed throughout the research, the performance of the meta-model will also be evaluated using this metric.

A sample of the code employed to train, validate and test the meta-model using k-fold cross-validation is shown in the appendix section. It is worth mentioning that the code is inspired by Lim (2022).

Figure 21: predicted EPS consensus estimates versus actual values.



Source: own elaboration.

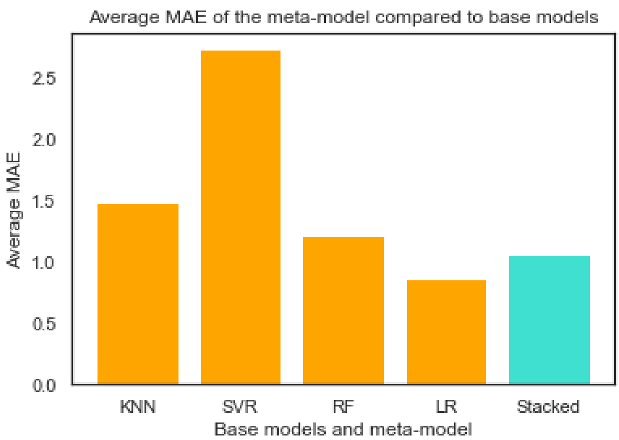
As it can be observed, the predicted EPS consensus estimates seem to have captured the patterns and behavior of EPS actual values overall. However, even though predicted EPS consensus estimates seem to generally track EPS value closely for the first years of the test set, they seem to divert from then onwards. Reasons that might explain this decrease in the performance of the meta-model across companies as time passes might include overfitting caused by the low dimensionality of data.

Another valuable insight derived from Figure 21 is the fact that predicted consensus EPS estimates tend to be lower than the actual values of EPS. This insight is consistent with previously mentioned findings that, similar to Hilary and Hsu (2013), seem to indicate that analysts generally underestimate EPS.

5.2.3 Results

After generating the predictions, the performance of the meta-model across the selected companies was evaluated. Figure 22 shows the average MAE of the base models compared to the meta-model.

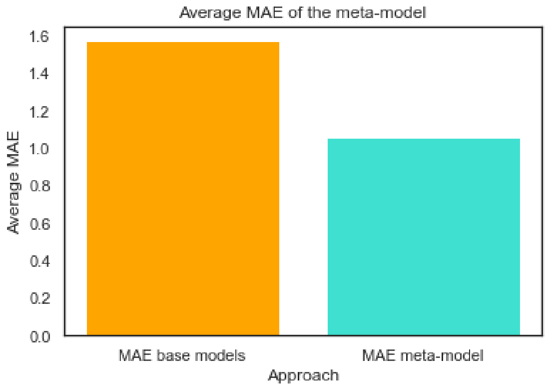
Figure 22: Average MAE of the meta-model compared to base models.



Source: own elaboration.

Across companies, the average MAE of the stacking algorithm was 1.05, which has been improved compared to the individual performance of each of base models (except linear regression). In brief, the meta-model has proven successful since it has performed better that all trained models on average, showing that, when different models are trained on different data, "once can receive essential gain the accuracy" (Pavlyshenko, 2019, p.1).

Figure 23: Improvement MAE compared to base models.



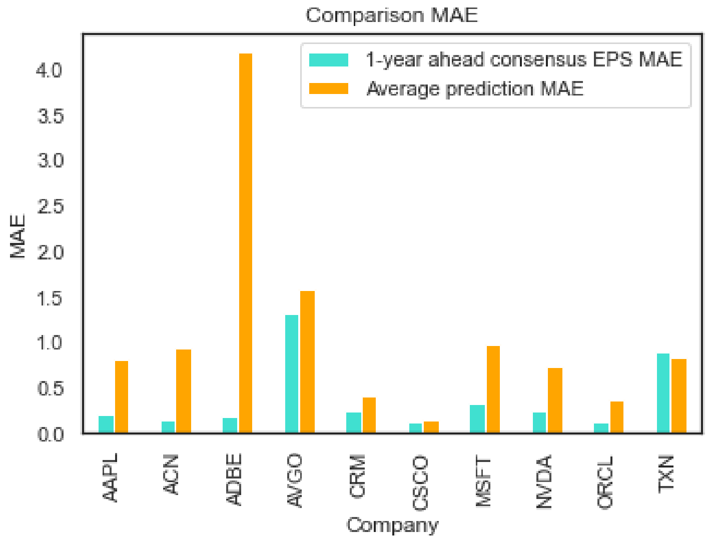
Source: own elaboration.

It is worth noting that the MAE represents the absolute difference between the predicted consensus EPS estimates and the actual EPS values that estimates attempt to forecast. However, the aim of this research is to build a meta-model to (potentially) arrive at a better consensus approach for EPS estimates. Consequently, the MAE of the meta-model must be compared against the previously calculated 1-year ahead consensus estimate MAE (Table 11) with a view to assessing whether the meta-model has accomplished the aforementioned objective. The 1-year ahead consensus estimate is probably the best reference for comparison considering that EPS estimates for the meta-model were recorded for 5 different dates and that the 1-year ahead consensus EPS estimate was the middle date:

- 2-year ahead.
- 18-month ahead.
- 1-year ahead.
- 6-month ahead.
- 1-month ahead.

Consequently, the 1-year ahead consensus estimate MAE might be the best proxy to evaluate the MAE of the predictions of the meta-model and determine whether the latter has succeeded at enhancing the accuracy of consensus estimates.

Figure 24: Comparison MAE.



Source: own elaboration.

As it can be inferred from both figures, the MAE of the meta-model is significantly higher than the MAE of consensus EPS estimates, indicating that the simple mean consensus method is more accurate than the proposed meta-model. Recapitulating, the hypothesis to be tested is that the proposed meta-model will lead to lower forecast errors compared to baseline consensus approaches. Nevertheless, the results of the analysis lead to the rejection of the hypothesis.

A reason that might explain why the meta-model has not proven useful is rooted in the semi-strong form of the Efficient Market hypothesis. As aforementioned, an "efficient" market can be defined as a market in which prices always reflect all publicly available information (Fama, 1970, p.1). It is important to clarify that "all publicly available information" does not include private or insider information. Taking this assumption into consideration, the semi-strong form of the EMH implies that no investor can consistently outperform the market by taking advantage or trading on any public information.

This point of view can also be applied to the present research. Again, the objective of this paper was to develop a more complex aggregating approach for EPS estimates, which involved the development of a meta-model. However, results clearly showed that the "traditional" consensus EPS estimates were more accurate than the consensus computed by the meta-model. What differentiated this approach to other attempts made by researchers was the fact that the consensus EPS estimates were computed by capturing the underlying patterns and behaviors of EPS. Since prices reflect all publicly available information (according to the semi-strong form of the EMH), consensus EPS estimates have also been computed using “all publicly available information” and thus can’t be consistently beaten by the meta-model. Following this train of thought, the meta-model only outperformed consensus EPS estimates once (the meta-model’s MAE was lower in the case of TXN) (Figure 24).

The semi-strong form of the EMH also explains why incorporating macroeconomic variables (such as GDP, inflation and interest rate) have not enhanced the performance of the meta-model. The rationale behind this statement is that these factors had already been included in the estimates computed by individual analysts.

Not only do individual analysts’ estimates account for all available information in the market but also, there are some technical and functional limitations as to the proposed meta-model that have affected its performance:

- Even though there are substantial amounts of easily accessible data regarding corporate EPS, more detailed data on analysts’ estimates is challenging to locate.
- There is high multicollinearity inherent in the dataset due to the high correlations between the predictor variables (individual analysts’ estimates), thus negatively affecting the performance of the meta-model.
- The low dimensionality of the dataset (due to missing values, differences in forecasting dates and number of analysts covering a company) may have led to overfitting.
- As stated by multiple authors, such as Sedor (2002) and Hilary and Hsu (2013), individual analysts’ estimates already contain biases and forecast errors, which might have affected the meta-model further.
- K-fold cross-validation splits the data randomly, thus ignoring the temporal order of the dataset.

The above set of limitations have likely affected the performance of the meta-model. However, it is worth noting that the performance issue probably stems from the base models rather than from the stacking algorithm itself considering that the latter has succeeded at aggregating the predictions of the base models, (the average MAE of the meta-model is lower than the MAE of the weak learners). Since ensuring the performance of the base models is the crucial catalyst to generating better predictions, it is not surprising that the meta-model has fallen short of expectations.

5.3 Recommended next steps

The present section will be dedicated to providing an outline of the next steps to be followed with the purpose of improving the present research. In line with the main objectives of this paper, further actions will be listed regarding both, the refinement of the meta-model and other potential methods of aggregating EPS.

As aforementioned, the meta-model has not achieved the objective of enhancing the accuracy of consensus EPS estimates, hence suggesting that further research needs to be conducted in order to come up with a successful ensemble method.

Concerning the ensemble method, one of its main limitations was the fact that the performance of the base models was not satisfactory. Considering that improving the performance of the base models is the crucial catalyst to developing a more accurate meta-model, a series of potential refinements will be outlined.

First, a wider range of algorithms must be trained with a view to identifying the most suitable number and type of base models. Other predictive algorithms that could be explored include neural network regression, ridge regression or XGBoost regression, amongst many others. In addition, hyperparameter tuning should be performed with a view to identifying the optimal combination that improves the performance of the base models.

Secondly, applying PCA to the dataset should be explored with the purpose of tackling another limitation of the meta-model: the high multicollinearity of the predictor variables. Another course of action that could potentially reduce the correlation between the input features would be to include other micro or macroeconomic variables that might affect EPS.

Thirdly, the TimeSeriesSplit function should be employed to train the base models instead of k-fold cross-validation with a view to preserving the temporal order of the data, that was otherwise omitted when using k-fold cross-validation.

Due to the importance of improving the accuracy of methods for aggregating EPS estimates, building upon the present research could prove valuable. However, if the above actions don't contribute to the accuracy of the meta-model, other aggregating approaches should be explored.

For instance, another promising approach involves performing a weighted average in which more importance was given to better performing analysts. This approach has already been tested, and with success, by Thomson Reuters and is called StarMine SmartEstimates. This method delivers more accurate consensus EPS estimates by placing "more weight on recent forecasts and top-rated analysts" (Frame, 2019, p.34).

Another approach that might be worth exploring in further detail includes using iterative filtering algorithms, as Kua (2022) proposed, which resulted in lower forecast errors compared to the simple mean.

Regardless, building more robust methods of aggregating analysts' EPS estimates is of particular importance to allow investors to make informed business decisions. Consequently, building on the present work (or exploring other consensus approaches) would prove highly valuable.

6. CONCLUSIONS

EPS estimates exert influence on business decisions and the economy as a whole. From investors making their investment decisions based on expected EPS to companies being pressured to beat their earnings estimates, consensus EPS estimates are, without a shadow of a doubt, a vital financial metric for multiple stakeholders. Consequently, ensuring the accuracy of EPS estimates is the crucial catalyst to helping investors and managers alike to make informed investment and business decisions. This is of particular importance in today's context due to the uncertainty surrounding the global economy.

Nevertheless, predictions are inevitably associated with forecast errors, and more so considering that analysts' estimates contain several biases due to the behavioral characteristics inherent to humans. These forecast errors definitely don't go unnoticed, with stock prices fluctuating when there are surprise earnings, thus affecting returns and decisions of investors and managers respectively. Therefore, proposing alternative methods to aggregate EPS estimates that are robust to the biases inherent in them is extremely relevant.

Due to the resources and time available for this research, beating individual analysts' forecasts is out of scope. What this research attempts to achieve, nonetheless, is to enhance the accuracy of consensus EPS estimates using the "wisdom of the crowd" theory as the underlying rationale for the meta-model. Even though the stacking algorithm did succeed at improving the overall accuracy of the base models, it failed at enhancing the accuracy of baseline consensus approaches.

Even though the proposed meta-model did not achieve the main objective of the present research, it probably shed light on the importance of developing more robust aggregation approaches given the influence they exert on business decisions and the economy as a whole. Following this train of thought, this research has hopefully provided the inquisitive investors with some ideas regarding the development of potential aggregation approaches beyond that of simple mean consensus so that the present research is further developed (successfully).

However, researchers must become aware of the scarcity of data available on individual analysts' EPS estimates. In brief, the problem lies in the lack of sources available to gather comprehensive data on individual estimates, with Thomson Reuters probably being the primary and only source available. Consequently, researchers that are not granted access to the I/B/E/S Estimates database might encounter multiple limitations when accessing individual analysts' EPS estimates.

Even though developing more robust aggregation approaches might pose a challenging task due to the limited accessibility of data, this endeavor has the potential to be highly valuable for multiple financial stakeholders. Nevertheless, the present research is not only relevant in the context of EPS estimates but could also be applied to other fields of finance considering that analysts don't only make estimations of EPS. In fact, analysts often estimate other financial metrics such as revenue and profitability ratios amongst many other key performance metrics. Consequently, the underlying rationale of the proposed meta-model could also be applied in other contexts.

In conclusion, this research is the crucial catalyst to highlighting the importance of developing more accurate and comprehensive methods of aggregating EPS that are robust to behavioral biases inherent in individual analysts' EPS estimates, and more so considering the relatively scarce literature on this subject. On a final note, this research will hopefully serve as a foundation for other researchers to build on and explore other aggregation approaches further.

7. BIBLIOGRAPHY

Almeida, H. (2018). Is it time to get rid of earnings-per-share (EPS)? The Review of Corporate Finance Studies, 8(1), 174–206. https://doi.org/10.1093/rcfs/cfy010

Austen-Smith, D., & Banks, J. S. (1996). Information aggregation, rationality, and the Condorcet jury theorem. *American Political Science Review*, *90*(1), 34–45. https://doi.org/10.2307/2082796

Barefield, R. M., & Comiskey, E. E. (1975). The accuracy of analysts' forecasts of earnings per share. *Journal of Business Research*, *3*(3), 241–252. https://doi.org/10.1016/0148-2963(75)90025-9

Basnarkov, L., Stojkoski, V., Utkovski, Z., & Kocarev, L. (2020). Lead–lag relationships in foreign exchange markets. *Physica A: Statistical Mechanics and Its Applications*, *539*, 122986. https://doi.org/10.1016/j.physa.2019.122986

Belgiu, M., & Drăguţ, L. (2016). Random Forest in remote sensing: A review of applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Brecque, C. (2019, July 31). *Forecasting earning surprises with machine learning*. Medium. Retrieved April 23, 2023, from https://towardsdatascience.com/forecasting-earning-surprises-with-machine-learning-68b2f2318936

Brown, L. D., Hagerman, R., Griffin, P. A., & Zmijewski, M. (2008, April 18). *Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings*. SSRN. Retrieved April 23, 2023, from https://papers.ssrn.com/abstract=1121425

Brownlee, J. (2021). *Ensemble Learning Algorithms With Python* (1.1). Machine Learning Mastery.

Chatzimparmpas, A., Martins, R. M., Kucher, K., & Kerren, A. (2021). Empirical study: Visual analytics for comparing stacking to blending ensemble learning. *2021 23rd*

International Conference on Control Systems and Computer Science (CSCS). https://doi.org/10.1109/cscs52396.2021.00008

Clayman, M. R., & Schwartz, R. A. (1994). Falling in Love again—analysts' estimates and Reality. *Financial Analysts Journal*, *50*(5), 66–68. https://doi.org/10.2469/faj.v50.n5.66

Condorcet, M. J., & Caritat, M. D. (1781). Sur les elections par scrutiny. Histoire de l'Academie Royale des Sciences, 31-34.

Copeland, T., Dolgoff, A., & Moel, A. (2004). The role of expectations in explaining the cross-section of stock returns. *Review of Accounting Studies*, *9*(2/3), 149–188. https://doi.org/10.1023/b:rast.0000028184.06279.57

Cragg, J. G., & Malkiel, B. G. (1968). The consensus and accuracy of some predictions of the growth of corporate earnings. *The Journal of Finance*, *23*(1), 67–84. https://doi.org/10.1111/j.1540-6261.1968.tb02998.x

Cunningham, P. (2007). (tech.). *Ensemble Techniques*. Retrieved May 9, 2023, from https://UCDCSI_technicalreport_ensembletechniques

Dai, R. (2020). I/B/E/S @WRDS 101: Introduction and Research guide [PowerPoint Slides]. Wharton Research Data Services. SlidesWharton

Divina, F., Gilson, A., Goméz-Vela, F., García Torres, M., & Torres, J. (2018). Stacking Ensemble Learning for short-term electricity consumption forecasting. Energies, 11(4), 949. https://doi.org/10.3390/en11040949

Elton, E. J., & Gruber, M. J. (1972). Earnings estimates and the accuracy of Expectational Data. *Management Science*, *18*(8). https://doi.org/10.1287/mnsc.18.8.b409

Fama, E. F. (1970). Efficient Capital Markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383. https://doi.org/10.2307/2325486

Frame, S. J., Tu, R., Martin, J. M., & Berding, J. M. (2019). The value of publicly available predicted earnings surprises. *The Journal of Finance and Data Science*, *5*(1), 33–47. https://doi.org/10.1016/j.jfds.2018.10.004

Gherghina, Ș. (2023). The impact of covid-19 on financial markets and the real economy. *Economies*, *11*(4), 107. https://doi.org/10.3390/economies11040107

Gilmartin, B. (2023, February 27). *S&P 500 earnings: Can Tech Sector Earnings growth stay positive in 2023?* Seeking Alpha. Retrieved April 30, 2023, from https://seekingalpha.com/article/4582276-sp500-earnings-tech-sector-earnings-growth-2023

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of Corporate Financial Reporting. *Journal of Accounting and Economics*, *40*(1-3), 3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

He, M., Gu, H., & Xue, J. (2022). Log interpretation for lithofacies classification with a robust learning model using stacked generalization. *Journal of Petroleum Science and Engineering*, *214*, 110541. https://doi.org/10.1016/j.petrol.2022.110541

Hilary, G., & Hsu, C. (2013). Analyst forecast consistency. *The Journal of Finance*, *68*(1), 271–297. https://doi.org/10.1111/j.1540-6261.2012.01800.x

Howell, E. (2023, January 10). *How to correctly perform cross-validation for Time Series*. Medium. https://towardsdatascience.com/how-to-correctly-perform-cross-validation-for-time-series-b083b869e42c

Hu, X., Wang, J., Wang, L., & Yu, K. (2022). K-nearest neighbor estimation of functional nonparametric regression model under NA samples. *Axioms*, *11*(3), 102. https://doi.org/10.3390/axioms11030102

Islam, R., Tonmoy, T., Choudhury, K., & Mahmood, A. (2014). How Earning Per Share (EPS) Affects on Share Price and Firm Value. *European Journal of Business and Management*, *6*(17), 97–108.

Ivković, Z., & Jegadeesh, N. (2004). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics*, *73*(3), 433–463. https://doi.org/10.1016/j.jfineco.2004.03.002

Jagliński, P. (2020). The impact of quarterly earnings announcements on stock prices. *Financial Sciences*, *25*(2-3), 24–40. https://doi.org/10.15611/fins.2020.2.02

Jensen, G. R., & Jones, C. P. (2020). *Investments: Analysis and management*. Wiley.

Jung, M. J., Keeley, J. H., & Ronen, J. (2017). The predictability of analyst forecast revisions. *Journal of Accounting, Auditing & Finance*, *34*(3), 434–457. https://doi.org/10.1177/0148558x17722710

Kalagotla, S. K., Gangashetty, S. V., & Giridhar, K. (2021). A novel stacking technique for prediction of diabetes. *Computers in Biology and Medicine*, *135*, 104554. https://doi.org/10.1016/j.compbiomed.2021.104554

Koller, T., Raj, R., & Saxena, A. (2013, January 1). *Avoiding the consensus-earnings trap*. McKinsey & Company. Retrieved April 23, 2023, from https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/avoiding-the-consensus-earnings-trap

Kua, K. (2022). Novel methods for aggregating analyst estimates. *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 3230–3235. https://doi.org/10.1109/icde53745.2022.00305

Lemke, C., & Gabrys, B. (2010). Meta-learning for time series forecasting and Forecast Combination. *Neurocomputing*, *73*(10-12), 2006–2016. https://doi.org/10.1016/j.neucom.2009.09.020

Lim, Y. (2022, April 5). *Stacked ensembles - improving model performance on a higher level*. Medium. https://towardsdatascience.com/stacked-ensembles-improving-model-performance-on-a-higher-level-99ffc4ea5523

Martín Guareño, J. J. (2016). *Support Vector Regression: propiedades y aplicaciones* (thesis).

Mohammed, A., & Kora, R. (2023). A comprehensive review on Ensemble Deep Learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Nilsson, J., & Svensson, P. (2019). *Accuracy of Analysts' Earnings Estimates* (thesis). Retrieved April 23, 2023, from https://gupea.ub.gu.se/handle/2077/60999.

Pavlyshenko, B. (2019). Machine-learning models for sales time series forecasting. *Data*, *4*(1), 15. https://doi.org/10.3390/data4010015

Prijanto, B., Pulung, R. F., & Sari, A. R. (2021). The influence of net profit margin on stock price with earnings per share (EPS) as moderating variables. *Journal of Economics, Finance and Accounting Studies*, *3*(2), 74–80. https://doi.org/10.32996/jefas.2021.3.2.8

Qiu, X., Zhang, L., Ren, Y., Suganthan, P., & Amaratunga, G. (2014). Ensemble deep learning for regression and time series forecasting. *2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)*. https://doi.org/10.1109/ciel.2014.7015739

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. . *IEEE Data Eng. Bull.*, *23*(4), 3–13.

Ray, R. (2006). Prediction markets and the financial "Wisdom of crowds". *Journal of Behavioral Finance*, *7*(1), 2–4. https://doi.org/10.1207/s15427579jpfm0701_1

Refinitiv (n.d.). IBES Academic [EPS Estimates][Data set]. Retrieved April 23, 2023, from Wharton Research Data Services https://wrds-web.wharton.upenn.edu/

Refinitiv Workspace (n.d.). [Price History][Data set]. Retrieved April 23, 2023, from https://workspace.refinitiv.com/web

Rocca, J. (2021, March 21). *Ensemble methods: Bagging, boosting and stacking*. Medium. https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205

Rybnik, R. (2022, December 26). *Measuring forecast accuracy: Complete guide with examples in Python*. Medium. Retrieved May 6, 2023, from https://fischerbach.medium.com/forecast-evaluation-statistics-with-examples-in-python-6b540ef751c2

S&P Dow Jones Indices. (2023, March). *S&P 500 Information Technology*. S&P Dow Jones Indices. Retrieved April 23, 2023, from https://www.spglobal.com/spdji/en/indices/equity/sp-500-information-technology-sector/#overview

Salazar, D. A., Vélez, J. I., & Salazar, J. C. (2012). Comparison between svm and logistic regression: which one is better to discriminate? *Revista Colombiana de Estadística*, *35*(2), 223–237.

Sedor, L. M. (2002). An explanation for unintentional optimism in analysts' earnings forecasts. *The Accounting Review*, *77*(4), 731–753. https://doi.org/10.2308/accr.2002.77.4.731

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis. *Deutsches Ärzteblatt International*. https://doi.org/10.3238/arztebl.2010.0776

Shrivastava, S. (2020, January 17). *Cross validation in Time Series*. Medium. https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1). https://doi.org/10.1186/1471-2105-9-307

Syarif, I., Zaluska, E., Prugel-Bennett, A., & Wills, G. (2012). Application of bagging, boosting and stacking to intrusion detection. *Machine Learning and Data Mining in Pattern Recognition*, 593–602. https://doi.org/10.1007/978-3-642-31537-4_46

Timmermann, A. (2018). Forecasting methods in finance. *Annual Review of Financial Economics*, *10*(1), 449–479. https://doi.org/10.1146/annurev-financial-110217-022713

World Bank Open Data. (2023). *GDP (current US$) - united states*. World Bank Open Data. Retrieved April 23, 2023, from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=US

World Bank Open Data. (2023). *Inflation, consumer prices (annual %) - united states*. World Bank Open Data. Retrieved April 23, 2023, from https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?locations=US

World Bank Open Data. (2023). *Real interest rate (%) - united states*. World Bank Open Data. Retrieved April 23, 2023, from https://data.worldbank.org/indicator/FR.INR.RINR?locations=US

Yi, S. K., Steyvers, M., Lee, M. D., & Dry, M. J. (2012). The wisdom of the crowd in Combinatorial Problems. *Cognitive Science*, *36*(3), 452–470. https://doi.org/10.1111/j.1551-6709.2011.01223.x

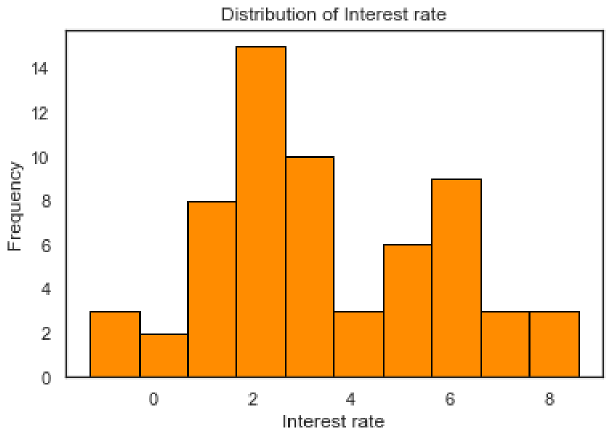
Yu, L., Wang, S., & Lai, K. K. (2009). A neural-network-based nonlinear metamodeling approach to Financial Time Series forecasting. *Applied Soft Computing*, *9*(2), 563–574. https://doi.org/10.1016/j.asoc.2008.08.001

Zhang, W., Cao, Q., & Schniederjans, M. J. (2004). Neural network earnings per share forecasting models: A Comparative Analysis of Alternative Methods. *Decision Sciences*, *35*(2), 205–237. https://doi.org/10.1111/j.00117315.2004.02674.x

8. APPENDIX

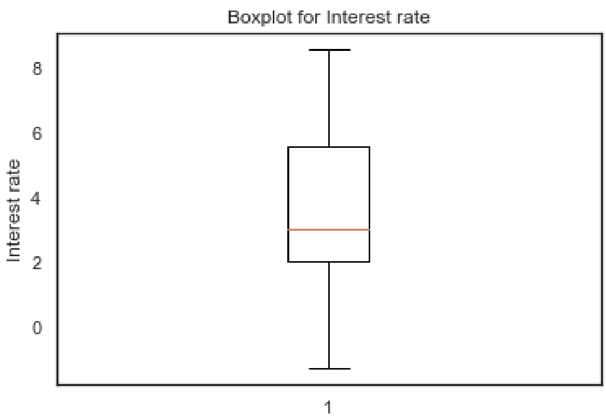
8.1 Exploratory analysis

Figure 25: Distribution Interest Rate.



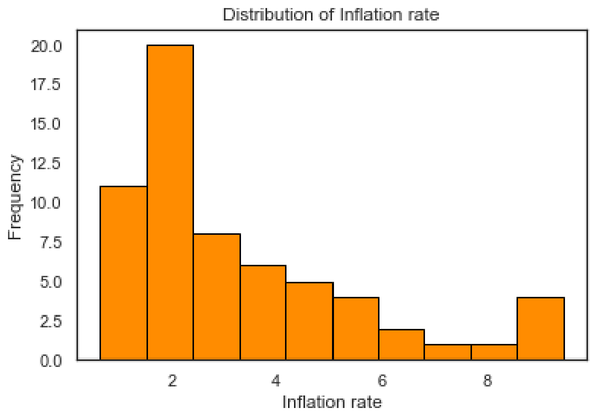
Source: own elaboration.

Figure 26: Boxplot Interest Rate.



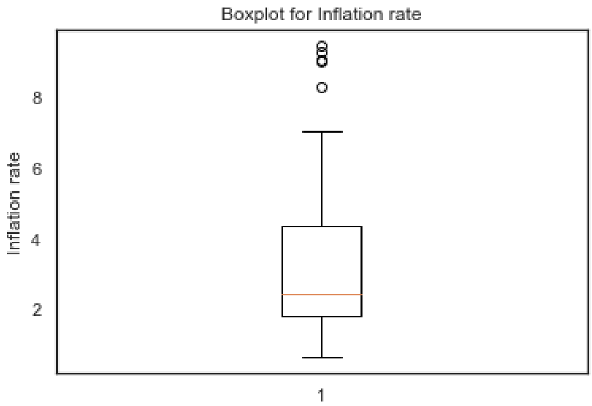
Source: own elaboration.

Figure 27: Distribution Inflation Rate.



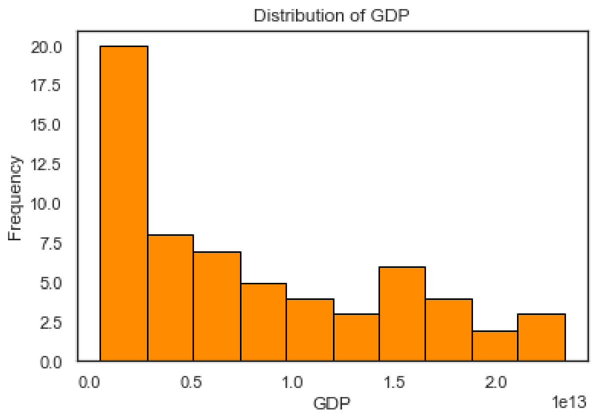
Source: own elaboration.

Figure 28: Boxplot Inflation Rate.



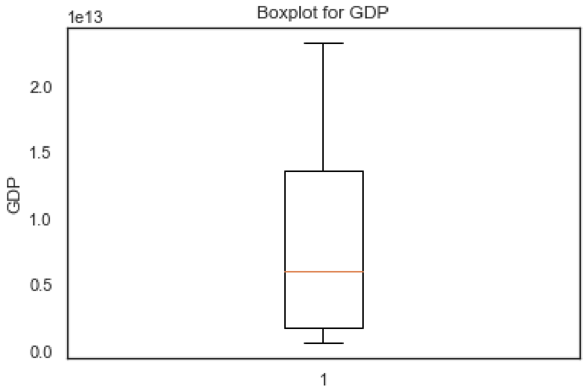
Source: own elaboration.

Figure 29: Distribution US GDP.



Source: own elaboration.

Figure 30: Boxplot US GDP.



Source: own elaboration.

8.2 Code sample

```
# Define a StackingRegressor.
basemodels = [
    ('KNN', KNeighborsRegressor()),
    ('SVR',SVR()),
    ('Random Forest',RandomForestRegressor()),
    ('Linear Regression',LinearRegression()),
    ]
stacked = StackingRegressor(
    estimators = basemodels,
    final_estimator = LinearRegression(), cv=5)

# Create a loop to develop a meta-model for each of the companies selected for this research.
dataframes=[metaAAPL,metaACN,metaADBE,metaAVGO,metaCRM,metaCSCO,metaMSFT,metaNVDA,metaORCL,metaTXN]
for df in dataframes:

# Split the data intro train and test set.
    features = df.iloc[:, 1:]
    target = df.iloc[:, 0]
    train_size = int(len(df) * 0.8)
```

```
    X_train, X_test = features[:train_size], features[train_size:]
    y_train, y_test = target[:train_size], target[train_size:]
    random.seed(15)

# Train the stacking algorithm.
    stacked.fit(X_train, y_train)

# Generate predictions.
    stacked_prediction = stacked.predict(X_test)

# Evaluate performance.
    stacked_mae = mean_absolute_error(y_test, stacked_prediction)
```