Ca' Foscari
University
of Venice

Redefining Financial Forecasting: A CAPE-Based Approach to S&P 500 Analysis and Portfolio Construction

Master's Degree in Economics and Finance

Marchesi Francesco

Matriculation number: 893466

Academic Year: 2022/2023

Supervisor: Ch. Prof. Stefano Tonellato

Abstract

Historically, the S&P 500 Index has been the object of numerous efforts by scholars and investment professionals seeking to deploy statistical and quantitative techniques in forecasting attempts. To this extent, a wide range of macroeconomic and financial variables have been studied to understand their potential influence on the Index's performance, primarily focusing on price-based fundamental and technical financial metrics.

This study diverges from the conventional approach by centring its analysis on the Cyclically-Adjusted Price-to-Earnings Ratio (CAPE), a concept made famous by Robert Shiller and John Campbell. Specifically, it implements linear regression models combined with ARIMA processes and the Newey–West estimator, to examine the extent to which behavioural and macroeconomic variables, such as investor sentiment and economic indicators, may carry explanatory power in forecasting CAPE fluctuations. Accordingly, this research argues that CAPE represents a more appropriate object of analysis rather than the raw Index price and explores the possibility of leveraging the evidence produced by statistical modelling to achieve superior portfolio returns.

Summary

1 Foundations and Context

1.1 Introduction

Over recent decades, global financial markets have undergone significant transformations, contending with economic crises and unparalleled market volatility. Accordingly, investors have been compelled to continuously refine and update investment methodologies, technological frameworks and operational processes. Herein, the last century has seen the introduction and development of numerous novel concepts, metrics, analytical tools and perspectives within the financial domain, aimed at dealing with and capitalising on diverse market inputs.

Among these, the Price-to-Earnings (P/E) ratio has captured the sustained attention of both academic critics and professional investors. The P/E ratio's prominence in financial analysis dates back to 1934 when Benjamin Graham and David Dodd first emphasized its meaningfulness and power in the inaugural edition of their seminal work, "Security Analysis" (Graham and Dodd (1934)). The implementation of portfolio strategies based on this basic yet insightful metric gave rise to the investment philosophy known as value investing. Two of the most famous proponents of this movement, which contributed to its rise to maximum peaks in the 20^{th} century, are Warren Buffet and Charlie Munger, the renowned investors behind the birth of Berkshire Hataway.

The P/E ratio is a rather simple fundamental[1] metric, defined as the ratio between the asset price and its earnings. Across the financial environment, there are several declinations regarding the terms of this metric, which can be applied; for example:

- Price: the price used to compute this metric may be chosen from either the spot (current) price of the asset, the price relative to the last fiscal year (FY), the price of the trailing twelve months (TTM) or the expected price for the next TTM;
- Earnings: in a similar fashion, earnings can be chosen to be either backwards-looking (i.e. using earnings from the last FY/TTM) or forward-looking (usually

[1]The term fundamental is used here in the context of highlighting the discrepancy between fundamental and technical financial analysis. Specifically, fundamental analysis refers to a financial analysis which takes into consideration the accounting and financial figures of a company (or index). Herein, fundamental metrics are based on tangible economic values, such as debt, equity, turnover, assets, liabilities and other similar financial statement elements. On the other hand, technical analysis primarily originates from the price chart of the company (or index) and goes on to elaborate on the price movements of the asset to derive graphical elements such as resistance and support zones, average daily ranges (ADRs) and different time horizons moving averages (MAs), among others.

the preferred pick from Wall Street financial analysts, reflecting forecasts of the company earnings for the future TTM).

The strategy of value investing generally stems from the empirical evidence that the P/E ratio may be interpreted as a proxy for potential speculation in the price of an asset. The reason behind this intuition is that, as it follows from the metric's definition, the P/E ratio measures how much are investors willing to pay for (i.e. how much are investors evaluating) one unit of earnings of the underlying company (this concept is largely explored in Chapter 2). Herein, stocks carrying significantly high P/E values are more likely to be overpriced, while the ones displaying a low P/E tend to be regarded as underpriced. This interpretation of the P/E ratio was first enounced in Graham (1949) and further developed by Warren Buffet. Most importantly, given the practical importance of this fundamental metric, the P/E ratio has been under the radar of several international academics for the whole 21^{st} century. As a result, this metric was tested and analysed by a multitude of studies, which in turn produced the positive externality of refining and extending the framework around the meaning and the impact of the P/E ratio on portfolio returns, laying the general foundations of this study.

One of the most notable works in this field is from Campbell and Shiller (1988); the authors extended the concept of the P/E Ratio and introduced a broader and smoother metric, called the Cyclically-Adjusted Price-to-Earning Ratio (CAPE)[2]. Differently from the classical P/E ratio, CAPE is computed using a moving average of the previous 10 years of inflation-adjusted earnings (i.e. real earnings) of the underlying asset. In so doing, this metric acquires a smoother and possibly more comprehensive declination.

The academic literature, which will be discussed in the subsequent section, has devoted considerable effort to the integration of these two metrics (P/E and CAPE) into statistical models used for forecasting future stock prices. These metrics are also often combined with other fundamental indicators to enhance the predictive accuracy of these models. A primary focus of these studies is represented by the S&P 500 Index, managed by the eponymous rating agency Standard & Poor's. However, two significant challenges are commonly encountered when developing statistical and forecasting models incorporating a range of fundamental metrics for predicting future prices of stock

[2]The CAPE ratio is also commonly referred to as P/E10 ratio or Shiler's P/E ratio.

indexes:

- Significant autocorrelation: the majority of fundamental metrics used by academic studies in this field contain the price of the index itself, but it also represents the dependent variable. This situation may (and often does) lead to complex bias and autocorrelation dynamics, which are harder to be dealt with given the contamination of the dependent variable into the models' regressors;
- As this study presents in depth (see Chapter 2, it is possible to empirically show that price levels are highly correlated with the earnings of the relative underlying assets. Nevertheless, Campbell and Shiller (2001) found that fundamental metrics have little capability of predicting earnings shifts, as well as similar dynamics related to the companies' business operations. Consequently, it appears counter-intuitive to continue pursuing using these metrics to make forecasts about raw price levels, instead of considering alternative approaches or methodologies.

To this extent, this work diverges from the conventional approach of using the CAPE ratio as a (potential) explanatory variable for the S&P 500 Index returns. Additionally, it aims to reverse this paradigm by focusing instead on forecasting future CAPE scenarios, specifically examining the extent to which behavioural and macroeconomic variables (such as investor sentiment and economic indicators) may play a role in predicting this metric's future fluctuations. Accordingly, this research argues that CAPE represents a more appropriate object of analysis rather than the raw Index price and explores the possibility of leveraging the evidence produced by statistical modelling to achieve higher portfolio returns through CAPE-based strategies.

This study is structured as follows: Chapter 1 sets the academic framework and defines the scope of the research. Chapter 2 investigates the financial data utilised and the variables considered. Chapter 3 applies statistical methodologies, particularly linear regression combined with ARIMA processes (ARIMAX) and heteroskedasticity and autocorrelation consistent error structures, as well as first differencing variables to deal with spurious regressions. Finally, Chapter 4 exploits the statistical models and empirical findings from Chapter 3 to formulate investment and portfolio management strategies.

1.2 Literature Review

Basu (1977) represents one of the pivotal publications in this field; the author has been the pioneer of the academic movement which backed the importance of P/E in the following decades. In particular, he investigated the possibility of exploiting P/E to achieve excess (i.e. above-par, risk-adjusted[3] portfolio returns. He concluded that even when adjusting for different systematic risk levels (represented by β) and for tax effects[4], low P/E portfolios systematically provided higher returns, compared to high P/E stocks selection. Basu's research further challenged the semi-strong form of the Efficient Market Hypothesis (EMH) by presenting evidence that investors could potentially exploit the P/E ratio to achieve excess returns. Additionally, his findings called into question the Capital Asset Pricing Model (CAPM), as higher returns observed on low P/E portfolios did not correspond with increased risk. Conversely, high P/E stocks did not exhibit the expected association with higher risk, suggesting a deviation from CAPM's foundational principle that would advocate for higher returns.

Dreman and Berry (1995) found that a stock's P/E level significantly impacts and leads to an asymmetrical market response to positive and negative news on the stock. High P/E stocks (which the author defines as "glamorous", underlining their higher relative price stemming from a boosted popular market demand) systematically delivered lower returns compared to low P/E stocks in case of positive news. At the same time, this selection of glamorous stocks also experienced harsher performances following negative news. Finally, the authors observed the significant presence of a mean reversion process[5] on stocks, especially impacting the highest and lowest quantiles (based on their P/E ratios). This effect appeared to extend over a period of up to five years following the initial news, irrespective of its positive or negative nature. This finding underscores

[3]In finance, the term excess returns refers to the instance of an investment earning a return higher than what the CAPM model would provide. Herein, excess returns are defined as above-par, risk-adjusted, since after controlling for the specific risk component of a portfolio (as CAPM does), the investment ranks above other comparable portfolios.

[4]Dividend taxation may carry a different tax rate with respect to capital gains. Most importantly, dividend taxation periodically reduces the cash flows received from the investor, thus hampering the re-investable amount and curbing the compounding effect of the investment itself. On the other hand, capital gains are only taxed when realized (i.e. at maturity), thus do not impact the effective investor's portfolio returns prior to that date

[5]The mean reversion process is a fundamental financial theory arguing that asset prices and historical returns eventually move back towards the mean or average level over time. In the context of the P/E ratio, mean reversion implies that if a stock's P/E ratio is significantly higher or lower than the industry average or its own historical average, it is expected to revert to its mean value. This reversion can occur due to adjustments in the stock price, changes in earnings, or a combination of both.

the persistent influence of mean reversion on stock performance, highlighting its role across various market segments delineated by P/E metrics.

Kane et al. (1996) developed a multivariate relationship model between P/E and a set of additional macroeconomic and fundamental variables tested as potential explanatory variables, including inflation, dividend yield (DY), and market volatility. The authors developed and tested this model in order to be able to evaluate the complex economic environment that was unravelling in the years prior to their publication. Indeed, they concluded that the high P/E levels of 1994 were not actually alarming when controlling for the variables mentioned above.

Campbell and Shiller (1988), which were previously cited as two of the most influential academics in this field, explored the predictive capabilities of both P/E and DY in forecasting long-term stock returns by regressing a combination of lagged earnings moving average (spanning 10 and 30 years) alongside growth in DY, against the price levels of the S&P500. Remarkably, the authors first introduced in this work the idea of extending the scope of the P/E ratio to more than a single year, by encompassing a broader earnings average. The ratio that Shiller and Campbell refer to as P/E10[6] in this work is the foundation of what will be introduced later in history as the CAPE ratio.

Another following work from Campbell and Shiller (2001) researched whether the predictive capabilities of fundamental metrics (as demonstrated in prior literature), stemmed from the effective ratios' ability to forecast increases (and decreases) in future production levels, earnings and dividends. Contrary to expectations, the authors determined that fundamental metrics, such as the P/E ratio, exhibit limited effectiveness in predicting "organic" trends within the market. However, this study showed that fundamental metrics do possess substantial predictive value for the price component of these ratios. This finding suggests that, while such metrics may not directly reveal earnings and production level movements, they are instrumental in forecasting future

[6] Literature uses figures next to elements in fundamental ratios in order to indicate whether the correspondent element should be taken as the spot value or as the average of a number of years, as indeed indicated by the figure. In this context, P/E10 indicates the ratio between the stock/index price and the average of the last 10 years of earnings. Similarly, P/E30 suggests the same interpretation, using an average of the last 30 years of earnings instead. Some studies also rely on the average as well for the price component of this ratio; in this case, the relative indication would appear next to the price factor at the nominator (e.g. P10/E10). This notation is found as well in following academic works.

price adjustments.

Fisher and Statman (2000) investigated the relationship between investor sentiment and stock market returns. This study analysed whether investor sentiment, as measured through various surveys and indicators, was able to provide significant predictions of future stock returns. The paper also examined the correlation between changes in sentiment among multiple groups of investors and stock market performance, emphasizing the diverse impacts of sentiment across different investor types. This research provided an important and pioneeristic insight into an investment approach based on behavioural finance factors.

Similarly, W. Y. Lee et al. (2002) explored the impact of investor sentiment on stock market volatility and excess returns. Utilizing the Investors' Intelligence sentiment index, the study employed a generalized autoregressive conditional heteroskedasticity model to assess how changes in sentiment affected market returns and volatility. This study concluded that sentiment, particularly shifts in it, systematically influenced market risk and return dynamics, underlining the impact of investor psychology in shaping market outcomes.

A following work by Fisher and Statman (2003) focused instead on the relationship between consumer confidence and stock market performance. This research primarily examined how consumer confidence, measured through the University of Michigan and the Conference Board surveys, was able to predict stock returns. The study found a generally negative relationship between high consumer confidence and subsequent stock returns, especially for the Nasdaq Index and small-cap stocks. Additionally, the paper explored the interplay between consumer confidence and investor sentiment, highlighting that changes in consumer confidence correlate with shifts in investor sentiment, particularly among individual investors. Herein, this research contributed to understanding the predictive power of consumer sentiment on market dynamics and extended the framework around behavioural finance.

Chahine and Choudhry (2004) investigated the effectiveness of value versus growth investment strategies in the European markets. Specifically, this study assessed the Price-to-Earnings Growth (PEG) ratio's role in these strategies, finding that low PEG ratio portfolios generally outperformed those with high PEG ratios, indicating a preference for value strategies over growth strategies within the European context. These results are consistent with the literature aforementioned.

Weigand and Irons (2005) analysed the impact of high P/E and P/E10 (CAPE) ratios during different historical periods on subsequent stock performance. Specifically, the authors discriminated between high valuations due to rapid price increases (without corresponding earnings growth) and those resulting from temporary earnings recessions. The findings indicated that high P/E levels, when caused by a temporary dip in earnings, did not necessarily lead to poor subsequent returns, as earnings often rebounded to their prior levels in the following years. Conversely, significant price increases that were not supported by a comparable growth in earnings consistently resulted in adverse outcomes for investors, irrespective of whether the perspective was short-term, medium-term, or long-term.

Aga and Kocaman (2006) tested relevant propositions concerning the effects of inflation and the P/E ratio on the Istanbul Stock Exchange (ISE). Their analysis revealed that neither the Industrial Production Index (IPI) nor the Consumer Price Index (CPI) provided adequate explanatory power for stock movements, encompassing both returns and volatility. To address this, they employed an EGARCH(1,1) model, which effectively captured the asymmetry observed in the ISE, and tested the explanatory power of the P/E ratio in this framework. Finally, they found that this metric was able to successfully provide insights about future ISE movements.

Huang et al. (2007) explored the strategy of leveraging the P/E ratio in order to build portfolios that would yield above-average returns, drawing on the mean-reversion process aforementioned. Their study, focusing on a 3-year holding period, revealed that stocks within the top decile by P/E ratio (which the authors again define as "glamour stocks") exhibited a significant tendency to deliver poor returns. Conversely, stocks in the bottom decile by P/E ratio (referred to as "value stocks") demonstrated notably superior returns. This distinction underscored the potential for mean reversion as a basis for investment strategy.

Aras and Yilmaz (2008) analysed short- and mid-term stock return predictability in 12 emerging markets, using a set of explanatory variables that included the P/E ratio, DY and the Market-to-Book (MtB) ratio. The study revealed that these variables possess significant forecasting power for returns over a one-year period, with the MtB ratio emerging as the most powerful predictor, followed closely by DY and P/E.

Akintoye (2008) critically revised the EMH in light of the paradigms of behavioural finance. In particular, this research found that, indeed, cognitive biases and emotions significantly influence investor behaviour, leading to market anomalies which should not exist under EMH. The study finally examines empirical research on market efficiency, the impact of accounting information and the effects of behavioural factors on financial markets, highlighting the complexity and interplay of rational and irrational behaviours in market dynamics.

Kelly et al. (2008) examined the performance of low P/E ratio stocks in the Australian market, aiming to test the viability of a low P/E-based trading strategy. This study extended the research from Basu (1977), effectively finding that low P/E stocks provided significant excess risk-adjusted returns, further challenging the semi-strong form of the EMH.

Angelini et al. (2013) introduced a dynamic model for predicting stock index returns, structured around three primary components: momentum, fundamental and driving forces. Following a similar analytical approach, the authors further substantiated the predictive efficacy of the CAPE ratio, particularly highlighting its utility in long-term forecasting. This work emphasizes the value of integrating diverse market indicators to enhance the accuracy of return predictions, with CAPE standing out for its significant long-term predictive power.

Bathia and Bredin (2012) conducted an in-depth analysis on the influence of different investor sentiments, proxied by market returns of G7 countries, on both value and growth stocks, categorized according to their price-to-book value (P/BV) ratio. The study revealed that sentiment from investor surveys had a significant impact on aggregate market returns. In contrast, the Put-Call Ratio (PCR) provided mixed and inconsistent findings, with its effects dissipating rapidly. Herein, the authors' overarching observation was that high investor sentiment showed a consistent correlation with lower future returns, while, on the other hand, low sentiment was frequently associated with higher future returns. This research underscores the predictive value of investor sentiment on market performance.

Smales (2017) examined the impact of investor sentiment, particularly fear, on stock market returns. Utilizing a range of sentiment proxies, including the VIX as a measure of investor fear, the study showed a significant relationship between sentiment and returns. It highlighted that sentiment, especially fear, plays a more influential role during recessions, affecting returns across various firm sizes and industries. The research underscored the complexity of investor sentiment's role in financial markets and its variable impact depending on economic conditions.

Baek and I. Lee (2018) investigated the potential presence of structural changes in S&P 500 historical P/E. Employing interesting techniques, such as a cumulative sum control chart and the Bai-Perron algorithm, the authors identified multiple structural breakpoints in P/E ratios over a 142-year period, finding that these structural changes significantly influenced long-term returns. This research challenges the traditional mean-reverting view of P/E ratios by showing that structural shifts can asymmetrically affect future returns, depending on whether the P/E period is high or low.

Allahyaribeik et al. (2020) used a Bohemian quantum approach[7] to analyse a bucket of Tehran's Stock Market companies and investigate the joint behaviour of P/E and price returns. By extending the quantum potential concept, the research reveals that both P/E ratios and price returns are confined within specific domains, influenced by underlying quantum potentials. This innovative approach facilitates a deeper understanding of market dynamics, concluding that quantum mechanics can offer valuable insights into financial market fluctuations and the relationship between critical financial indicators.

Campisi and Muzzioli (2020) explored the influence of investor sentiment on financial market dynamics. Focusing on two groups of investors following a value investing approach, encompassing different perceptions of fundamental value, the paper introduced a sentiment index to model trading decisions and analyse complex market scenarios, including fear and greed dynamics. The study demonstrates that investors' sentiment, measured through the sentiment index, significantly impacts market volatility and stock return asymmetry. The paper's insights contribute to understanding how investor psychology and market sentiment drive financial market fluctuations and highlight the importance of investor sentiment as a key factor in financial market behaviour.

Kenourgios et al. (2022) analysed the forecasting capabilities of CAPE and CAPE5[8] on the Greek Stock Market. Through regression models assessing 1, 3, 5, and 10-year real returns against fundamental variables, such as P/E, CAPE5, CAPE and

[7]In the context of this paper, the "Bohemian Quantum Approach" refers to a formulation of quantum mechanics that includes particle positions along with wave functions. In the context of this paper, the Bohemian mechanic principles are used to analyse financial markets, exploiting the concept of "quantum potential" to model and predict stock market behaviours, such as the P/E ratio and price return dynamics.

[8]CAPE5 is a variations of CAPE fundamental metric, built by employing the moving average of the last 5 years of inflation-adjusted earnings, rather than the 10-years horizon used by CAPE

P/BV, the study reveals a departure from traditional findings. Specifically, the authors indicate that the P/E ratio and P/BV have limited predictive value for future returns. Conversely, CAPE and CAPE5 display significant forecasting power for mid- and long-term market returns, suggesting their superior utility in predicting market movements in the Greek context.

Wang et al. (2021) extended the framework of behavioural financial analysis to encompass both developed and emerging markets, utilizing the Consumer Confidence Index ("CCI") to explore its impact on market returns. This research found that CCI displayed a significant impact on market returns, even after controlling for potential outside economic indicators that could have affected CCI. Additionally, the research underscored the importance of a broader set of classical behavioural metrics, including IDV, UAI, and MAS[9], as well as intelligence and education factors like IQ and literacy levels, in explaining the divergent responses to comparable macroeconomic situations between developed and emerging markets.

[9]IDV (Individualism), UAI (Uncertainty Avoidance Index), and MAS (Masculinity) are dimensions from Hofstede's cultural dimensions theory, which provides insights into the influence of culture on values in the workplace and society. Individualism (IDV) measures the degree of interdependence a society maintains among its members. Uncertainty Avoidance Index (UAI) assesses a society's tolerance for uncertainty and ambiguity. Masculinity (MAS) reflects the distribution of emotional roles between the genders. These metrics are used to understand behavioral differences across cultures, influencing consumer behavior, market trends, and investment decisions.

2 S&P 500 Index Analysis and Long-Term CAPE Trends

2.1 S&P 500 History and Data Methodology

The S&P 500 is one of the most famous index funds, designed by Standard & Poor's to track the performance of 500 of the largest and most important companies based in the United States. Specifically, as of February 2024, the Index covered approximately 80% of the total available market capitalization (*S&P 500® Overview* (2024)). It was officially established in March 1957 (Valetkevitch (2013)), yet data is available extending back to 1871. This retrospective extension of the S&P 500 Index data was made possible through a collaborative effort between Standard & Poor's and the economist Alfred Cowles. Standard & Poor's describes the extension of the Index to periods predating its official inception in 1957 as "hypothetical back-tested", since it employs the S&P 500 methodology that was in effect at the time of the Index launch (*S&P 500® Overview* (2024)). Herein, the company exploited this backward interpolation in order to reflect a hypothetical performance of this index extended back to 1923[10]. In a similar fashion, Alfred Cowles and associates (Cowles (1939)) broadened the S&P 500 time horizon back to the year 1871.

The S&P 500 is widely accepted in the finance world as one of the main benchmark indexes for investment comparison and it is arguably the main one in the field of equity investments[11]. For this reason, as mentioned in the introduction, it has been the centre of attention of academics and institutional investors for over a century, particularly representing the object of statistical and forecasting modelling using broad and diverse sets of macroeconomic, behavioural, fundamental and technical variables. An interesting publication in this field, named "The Little Book of Common Sense Investing" (Bogle (2017)), applied an alternative point of view to decompose the S&P 500 Index

[10] Standard & Poor's disclaims that this methodology may be potentially affected by survivor or look-ahead bias, in addition to a multitude of financial risks that could not have been accounted for when building the Index hypothetical performance prior to 1957. For this reason, financial data implying ad-hoc, backward-looking construction, are not to be taken as references for actual returns. Nevertheless, the scope of this study is not affected by the facts stated, since the use of the S&P 500 Index as a benchmark for investment comparison made in Chapter 4 only encompasses data from July 2001.

[11] Given its importance, the S&P 500 Index is also commonly referred to as "the market". This study may refer to it accordingly.

performance. The author, late founder of The Vanguard Group[12], argued that returns from the S&P 500 Index[13] could be split into two main categories:

- An organic component, representing the earnings growth of the underlying assets (thus, aggregate earnings of the 500 companies underlying the S&P 500);
- A speculative component, stemming from investors' behaviour and other sets of factors which may be able to influence it. In this context (and from now on, in this study) the speculative component is simply defined as the difference between the S&P 500 Index returns and the relative organic component (earnings growth).

This work attempts to test the accuracy of the framework laid out in Bogle (2017), as well as implementing an alternative view based on the CAPE ratio to provide a more extensive analysis of the S&P 500. Specifically, this chapter aims to elaborate on the Index's performance in order to weigh the influence of the organic and speculative components on the S&P 500's historical returns. Accordingly, this project assesses to which extent the Index's performance was fuelled by actual earnings growth of underlying companies and, on the other hand, which part of it was instead driven by speculative market dynamics, proxied by the CAPE ratio.

The dataframe used for this purpose is curated by Robert Shiller, Sterling Professor of Economics at Yale University and recipient of the Nobel Prize in Economic Sciences. This dataset is widely used in the academic environment and it is publicly available on the author's website, alongside several other insightful collections of historical data. Specifically, the dataframe used is called: "U.S. Stock Markets 1871-Present and CAPE Ratio".

In light of the purposes of this research, as can be seen by downloading the dataset mentioned above, the raw data provided by Robert Shiller is quite inconvenient for a few reasons: (i) column names are spread among multiple rows, (ii) there are a lot of empty columns in between useful variables and (iii) real (inflation-adjusted) data is computed on last observation's CPI (which is September 2023, as the date of writing); however, since quarterly (Q3) financial information was not available yet at that time, last observations are incomplete. For these reasons, it is necessary to transform data

[12] As of February 2024, The Vanguard Group retained approximately $7.7 billion of assets under management; it was the largest global provider of index funds and the second largest provider of ETFs, only behind BlackRock.

[13] This approach can be applied to whatever stock index or, alternatively, to individual companies.

adequately before handling it. Specifically, in order to address the 3 issues mentioned above:

- Column names are collapsed (and renamed when necessary) in one row only, so they will present the correct form in order to be taken as column labels;
- Empty, unused or redundant columns are deleted;
- Real price, earnings and dividends are adjusted to take June 2023 CPI as the base figure;
- 2023 data is trimmed at the end of June 2023, to avoid confusion potentially stemming from unused and incomplete data.

Finally, for the purpose of handling the final dataframe, as described and adjusted, this work exploits the R environment. The following snippet of code offers a preview of how the comprehensive dataframe first presents, alongside a general summary of the main variables explored by this study.

```
print(df, n = 2, width = Inf)
```

```
      Date  S&P500    Dividend   Earnings   CPI  Long Int. Rate (GS10)
 1881-01-01    6.19        0.27       0.49  9.42                  3.70
 1881-02-01    6.17        0.27       0.48  9.51                  3.69

 Real Price  Real Dividend Real  Total Return Price  Real Earnings Real
     200.50                8.58              370.94               15.74
     197.86                8.66              367.38               15.45

 TR Scaled Earnings        CAPE       10yr Ann.Stock.Real.Ret
              29.11       18.47                      0.045353
              28.68       18.15                      0.046774

 10yr Ann.Bond.Real.Ret             10yr Ann.Excess.Real.Ret
               0.056468                            -0.011115
               0.056199                            -0.009425
 #1,708 more rows
```

```
       Date                  S&P500            Dividend
  Min.   :1881-01-01   Min.    :   3.810     Min.    : 0.180
  1st Qu.:1916-08-08   1st Qu.:    8.685     1st Qu.: 0.470
  Median :1952-03-16   Median :  24.745      Median : 1.430
  Mean   :1952-03-16   Mean    : 395.898     Mean    : 7.988
  3rd Qu.:1987-10-24   3rd Qu.: 268.825      3rd Qu.: 8.748
  Max.   :2023-06-01   Max.    :4674.770     Max.    :68.710

      Earnings              CPI          Long Int. Rate     Real Price
   Min.    :  0.16     Min.    :  6.28   Min.    : 0.620   Min.    : 107.0
   1st Qu.:  0.67      1st Qu.: 10.95    1st Qu.: 3.120    1st Qu.: 222.1
   Median :  2.51      Median : 26.45    Median : 3.695    Median : 357.6
   Mean    : 18.79     Mean    : 68.73   Mean    : 4.469   Mean    : 778.1
   3rd Qu.: 16.10      3rd Qu.:115.38    3rd Qu.: 5.000    3rd Qu.: 843.2
   Max.    :197.91     Max.    :305.11   Max.    :15.320   Max.    :5123.5

   Real Dividend           CAPE
   Min.    : 7.59      Min.    : 4.78
   1st Qu.:11.16       1st Qu.:11.95
   Median :17.39       Median :16.47
   Mean    :21.07      Mean    :17.38
   3rd Qu.:26.33       3rd Qu.:21.10
   Max.    :70.98      Max.    :44.20
```

2.2 Monthly Data Overview

The dataframe described above provides observations collected on a monthly basis, which is notably convenient for refined granularity and precision, and addresses multiple insightful fields of information. The main 3 variables taken into greater consideration by this Chapter are represented by the S&P 500 Index price, earnings (S&P 500 companies aggregate earnings level) and dividends (S&P 500 companies aggregate dividends level).

In order to provide a general overview of these different time series, it is useful to graphically present the 3 main variables in all of their main declinations, specifically: (i) nominal (standard) levels, (ii) real (inflation-adjusted) levels and (iii) logarithmic levels.

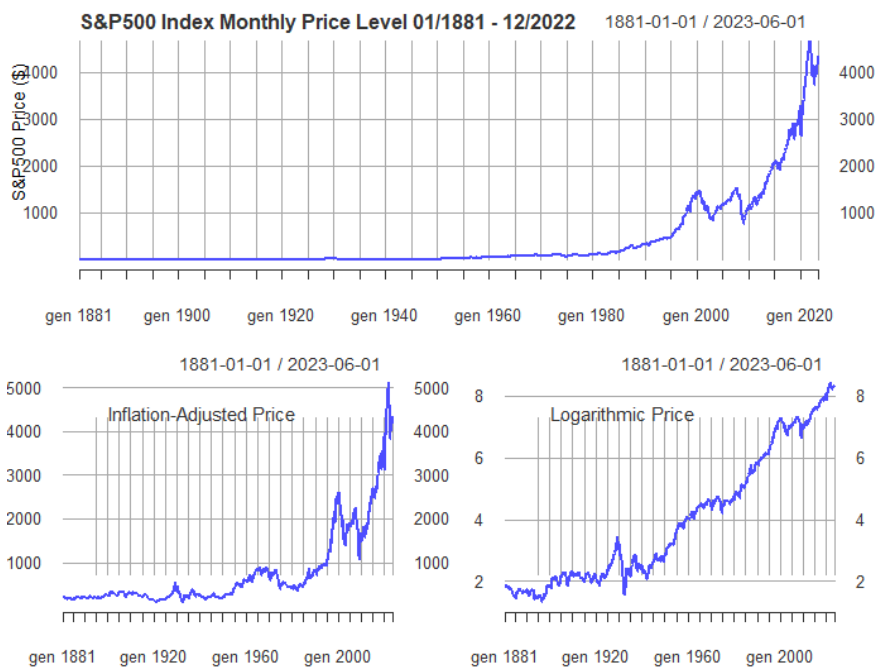

Figure 1: S&P 500 Monthly Returns

Figure 1 displays the graphs relative to the previously mentioned elements, pertaining to monthly price levels of the S&P 500 Index from January 1881 to June 2023.

- The top plot illustrates the nominal price, which, upon initial observation, exhibits a more pronounced curvature for the periods encompassing the last 50 years. This is due to the large extension of the dataset in terms of the historical period covered, as well as the very different magnitude of the latest price levels, compared to the first century of data;

- The plot in the bottom-left corner presents the real prices of the S&P 500, offering a neat visual representation of the significant impact of inflation on the movements of the Index.
- Lastly, the bottom-right plot illustrates the logarithmic trajectory of prices. From this perspective, it becomes evident that the notable spikes observed in the other graphs are attributable to the compounding effect[14] of returns rather than an improved performance in recent years.

In a similar fashion, Figure 2 focuses on the S&P 500 earnings level, providing a measure of the aggregate earnings of the 500 companies underlying the Index.

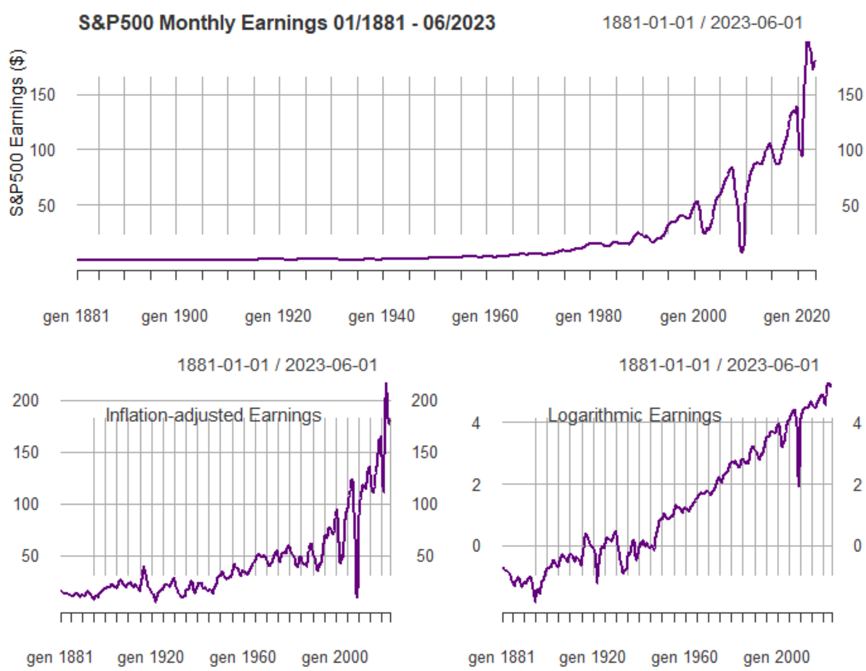

Figure 2: S&P 500 Monthly Earnings

It is quite intuitive to notice large similarities between these plots and the ones displayed in Figure 1. However, at a deeper level, earnings appear to present two

[14]In a financial investment framework, the compounding effect refers to the process by which both capital gains and interest accumulate additional returns over time, in addition to the principal invested. This phenomenon occurs as investment earnings are reinvested to generate their own share of returns. Mathematically, compounding is the result of the exponential growth of the initial principal amount of an investment, stemming from the accumulation of reinvested interest or dividends over multiple periods.

distinctive characteristics, in contrast with price levels. On the one hand, earnings seem to report more accentuated drawdowns in recession periods, as can be seen right after the year 2000, in the occurrence of the burst of the doc.com bubble and the subprime mortgage crisis (2007-2010). This observation is supported by the logarithmic plot (bottom-right corner), which underlines notable sharp declines not present with the same magnitude in the S&P 500 logarithmic price level chart. On the other hand, earnings also seem to follow a smoother trajectory, both in nominal and real terms, as they produce an overall significantly more linear logarithmic plot.

Figure 1 and Figure 2 already seem to indicate a notable correlation between the S&P 500 earnings and the Index's price. As remarked by Bogle (2017), stock assets' prices (especially on a long-term basis) are largely driven by their "organic" component in the long run. This appears to be evident by analysing the plots described above and this study will try to display the quantitative arguments backing this graphical intuition. Nevertheless, it seems to be also true that stock prices do not forcefully follow this process on a short- and mid-term[15] basis. Indeed, under shortened horizons, a broad set of outside variables referred to as speculative factors impact and alter the Index's performance, which does not have sufficient time to re-align with the organic component.

Finally, Figure 3 pictures insightful information regarding the dividends level of the S&P 500. Once again, the dividends trajectory appears to be very similar to the dynamics displayed in the Index's price and earnings, as previously discussed. One thing that captures the attention is a steep increase in the inflation-adjusted value of real dividends following the 2008 financial crisis, as can be seen in the bottom-left corner plot. Dividends represent a fundamental aspect of value investing and DY, as mentioned in the first chapter, has been tested as a potential predictor of future index returns, sometimes showing satisfactory results. Nevertheless, dividends will not constitute a fundamental factor in this work, as they do not play a significant role in the context of the CAPE ratio nor are influenced by behavioural factors (since investors' behaviour does not affect companies' dividend payout policies). For these reasons, dividends will not enter this specific field of research.

[15]Generally speaking, in finance, the expression short-term refers to periods of up to 1 year, while a mid-term horizon encompasses from 1 to 3/5 years. However, these references do not have a largely agreed definition but rather depend on the time horizon under scrutiny.

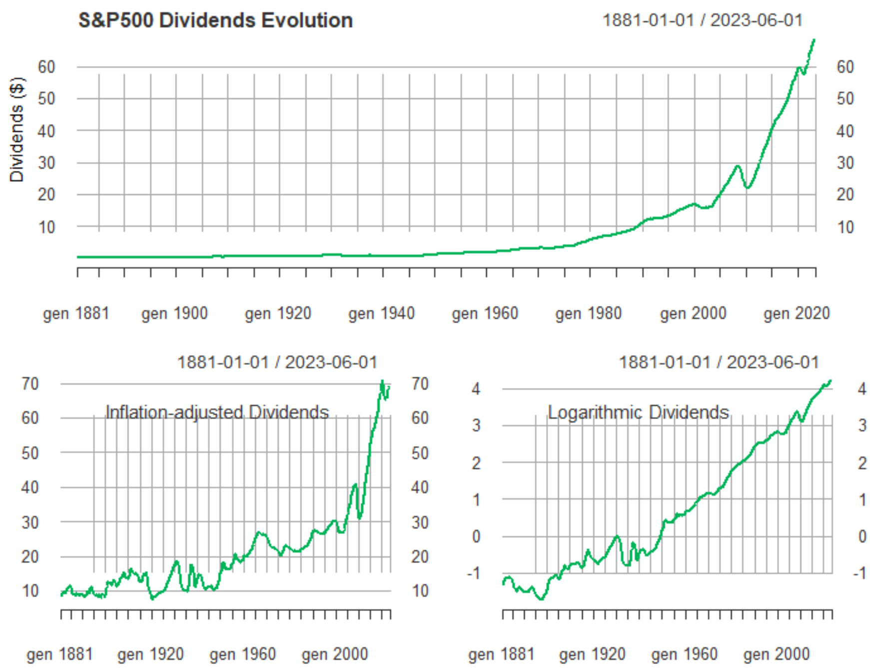

Figure 3: S&P 500 Monthly Dividends

2.3 On The Impact of CAPE

In the preceding section, the study examined historical dynamics affecting the S&P 500's price, earnings and dividends. As mentioned, the first two components are of particular interest and warrant a detailed joint analysis. In this framework, the CAPE ratio becomes particularly relevant and represents a fundamental tool to explore the interplay between the Index's price and earnings movements.

Figure 4 provides a visual representation of the CAPE ratio levels historical time series, boxplot and frequency distribution. A few consideration arises from these different graphs:

- As can be noticed from the plot displayed in the top row, the S&P 500's CAPE ratio trajectory indicates that valuations have not been stationary over time, suggesting a dynamic interplay between organic factors and speculative variables;
- The pronounced peak in the CAPE ratio around the year 2000 appears to have initiated a relatively consistent uptrend, persisting approximately over the past decade and establishing a new resistance level. This breakout[16] event encompasses

[16]In financial technical analysis, a breakout denotes the scenario where a security's price surpasses

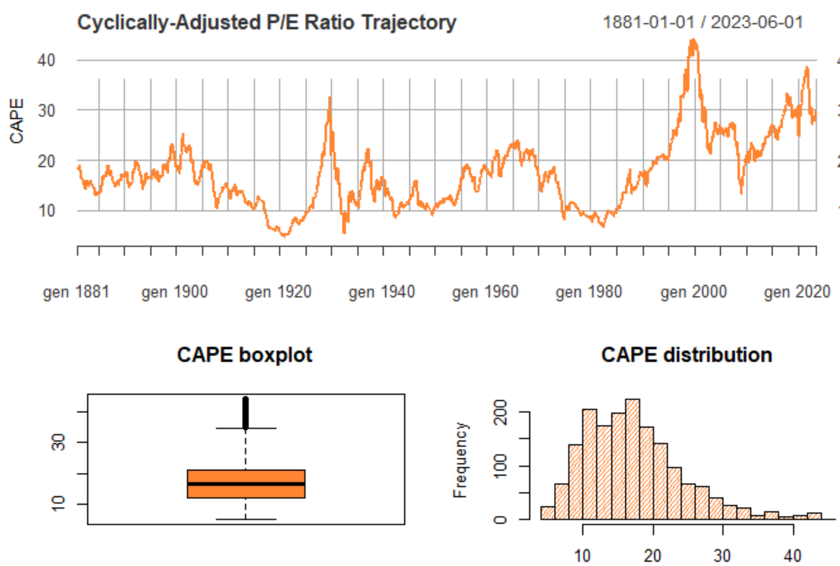

Figure 4: S&P 500 Cyclically-Adjusted P/E Ratio

the latest structural change in the CAPE ratio outlined in Baek and I. Lee (2018) (which the authors refer to as P/E10).

- The box plot depicted in the lower left quadrant of the figure serves as a compact representation of the CAPE ratio's distribution, picturing the median and the interquartile ("IQR") range of the dataset. As can be noticed, this plot presents clear evidence of the high recent values being the primary outliers, as observations are notably above the historical average levels;
- The histogram pictured in the bottom-right corner further illustrates the empirical density of observed values and reaffirms that, for the majority of history, the CAPE ratio showed values comprised from 10 to 20. Moreover, the distribution appears to be right-skewed, as evidenced by a longer tail extending towards the higher end of the CAPE ratio spectrum. This skewness is coherent with a concentration of lower CAPE ratios and fewer instances of recent significantly high ratios.

Despite it may be pictured as a simple descriptive ratio, CAPE has a deep intrinsic power on investment returns. First of all, considering the nature of this metric, CAPE

established support or resistance levels, indicating a strong market inclination to trade the asset at new price points, either higher or lower, depending on whether the breakout is upward or downward.

may be regarded as a proxy for speculative market dynamics. This characteristic stems from the fact that, by definition, this ratio measures how much are investors willing to pay for (i.e. how much are investors valuing) one unit of earnings of the underlying asset. Herein, the further CAPE values are from historic averages, the more probable it is that the market entered a speculation-dominated territory. On the one hand, a high CAPE may indicate extended market overpricing, implying an increased risk of a potential sudden price decline. On the other hand, low CAPE values may represent a signal of the economy entering into a generalised recession, with investors not finding companies' growth appealing anymore and hampering the market's development. It follows from this reasoning that, as further developed in the cited literature (Kane et al. (1996), Weigand and Irons (2005)), the CAPE (P/E) ratio may be regarded as a powerful tool to evaluate the overall speculative dynamic of the market, assessing whether stocks are approaching under- or overpricing territories.

In this context, the financial critic tried to argue that, in some circumstances, high/low CAPE (P/E) value may be justified by the relative expectations on future stock/index performance (as mentioned by Graham (1949), among others). Herein, a low CAPE ratio may incorporate a curbed future outlook, reflecting the impact that limited earnings growth would have on re-balancing the CAPE ratio back to standard (higher) levels. A similar argument can be made for some circumstances leading to high CAPE values, which may be seen as enhanced expectations of future earnings. If positive forecasts are realised, higher earnings levels are likely to lower the CAPE ratio over time.

Shifting towards a pragmatic context, following the significant considerations aforementioned, CAPE may also be regarded as a "financial leverage" for portfolio returns. Specifically, this leverage effect is provided by the impact of speculative forces (embedded in the numerator of this metric) on the earnings of the underlying companies (at the denominator). The meaning of leverage in this circumstance refers to the fact that the price and earnings may grow/decline at different paces. It follows that, while an investor may theoretically expect his investment to grow at a rate close to the underlying company(-ies) growth/decline, this does not always happen (especially in the short- and mid-term, as previously mentioned). Remarkably:

- In periods of upward CAPE movements, an increase in companies' earnings would have a more-than-proportional impact on portfolio returns and a decrease in earnings will not be matched by a decrease in the portfolio's value of the same size. In these situations, the CAPE ratio provides investors with a positive leverage effect;

- On the other hand, a CAPE downtrend supplies negative leverage, as investors earn limited returns for positive earnings news and harsher performances in case of earnings declines.

These two conclusions logically draw on the fact that CAPE is defined as the ratio between the price of the Index and the (moving average of) inflation-adjusted earnings of underlying companies. This CAPE fundamental principle, which was referred to as financial leverage, constitutes the basis of the portfolio strategy developed in Chapter 4. Specifically, investment strategies will be developed in order to exploit the process described above, such that portfolio returns would be able to take advantage of the positive leverage and maintain a neutral state in circumstances of negative financial leverage.

2.4 Comparing S&P 500 Cumulative Logarithmic Returns vs Organic Growth

The previous sections introduced the general framework and the overall theoretical foundations around the CAPE ratio. It follows that a more detailed analysis is needed to investigate different historical periods and spot possible trends, as well as to best comprehend the impact that CAPE (and the leverage effect previously described) may have had on past S&P 500 returns. Specifically, this work will analyse the aspects aforementioned under 2 different time perspectives, focusing on:

- Exploring cumulative returns for different time horizons;
- Computing average decade returns to asses macro-trends.

This section will address the first point, while the following will take care of the latter.

Cumulative returns are a common approach in financial analysis, particularly in the context of evaluating two (or more) investment performances over time. The simple cumulative return over a period measures the total percentage change in the value of an investment, accounting for the compounding of returns. Given a sequence of prices

$p_0, p_1, \ldots, p_n$, where p_i is the price at time i, the simple return R_i from time $i-1$ to i can be expressed as:

$$R_i = \frac{p_i - p_{i-1}}{p_{i-1}}$$

The simple cumulative return CR over the period from 0 to n is then computed by multiplying the individual period returns plus one, and then subtracting one to find the overall percentage change:

$$CR = \left(\prod_{i=1}^{n}(1+R_i)\right) - 1 = \frac{p_n}{p_0} - 1$$

Simple cumulative returns show how the initial investment value grows over time, considering the compounding effect of returns. They provide a straightforward measure of the total return, reflecting both capital gains and losses over the investment period.

However, it is a best practice to use a logarithmic scale, rather than a nominal one, for data clarity's sake. The main reason for this lies in the difference between S&P 500 Index price (for example, as of the first observation in 1881-01-01, the price is equal to \$6.19) and earnings (\$0.49). It follows that the same increase in absolute terms will lead to a higher relative (percentage) increase for the latter, and vice versa. This dynamic would depict misleading data and translate into erroneous interpretations. The cumulative logarithmic return over a period is calculated by summing the logarithmic returns of individual periods. Given a sequence of prices $p_0, p_1, \ldots, p_n$, where p_i is the price at time i, the logarithmic return r_i from time $i-1$ to i is given by:

$$r_i = \ln\left(\frac{p_i}{p_{i-1}}\right)$$

Hence, the cumulative logarithmic return R over the period from 0 to n is:

$$R = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} \ln\left(\frac{p_i}{p_{i-1}}\right) = \ln\left(\frac{p_n}{p_0}\right)$$

This formula emphasizes the additive property of logarithmic returns, making them particularly suitable for analyzing the performance of investments over time, as it directly relates to the total percentage change from the initial to the final period. Herein, this study exploits this methodology to compare the S&P 500 returns with the growth of the relative organic component. In so doing, this analysis will show the trajectory difference between these two factors and, by construction, the resulting cumulative difference will represent the impact of speculative market dynamics (i.e. CAPE uptrend/downtrend periods, in which the market yields higher/lower returns compared to earnings growth).

The first reference point taken into consideration to compute cumulative logarithmic returns starts with the year 1900. Herein, Figure 5 pictures the trajectory of the S&P 500 Index total returns compared to its relative companies' aggregate earnings over more than a century. Additionally, this plot highlights the correlation in the movement, as well as the divergences between market performance and fundamental corporate profitability.

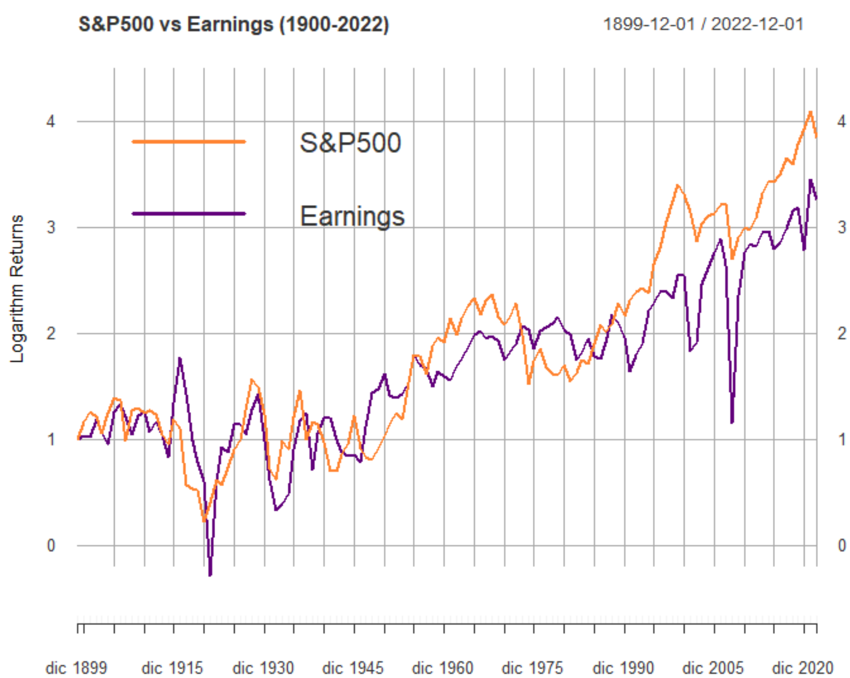

Figure 5: Total vs. Earnings Cumulative Logarithmic Returns

The orange line, representing the S&P 500 index, and the purple line, denoting earnings, exhibit periods of confluence and disparity, reflecting the dynamic interplay between investor sentiment and corporate financial health. Notably, the relative steepness of each line indicates the rate of growth, with steeper ascents suggesting more rapid increases in returns/earnings. On a deeper level, the graph illustrates that while the two measures tend to move in tandem over the long term, there are intervals where the S&P 500 index either outpaces or lags behind the growth in earnings, potentially indicating market overvaluation or undervaluation trends, respectively.

This analysis reveals that the S&P 500 Index's total cumulative logarithmic returns from 1900 to the present have outpaced the earnings growth of the underlying companies. Specifically, the Index's price has increased by approximately 59% more than the earnings, as retrieved by looking at the current level of cumulative logarithmic returns.

```
last(from.1900$Tot.cumlogreturns)-last(from.1900$Earnings.cumlogreturns)
```

```
 [1] 0.5912372
```

This indicates that the CAPE ratio has increased accordingly, providing investors with returns exceeding the actual earnings growth of the companies within the same time frame. Nevertheless, it is possible to notice also that under the historical horizon analysed, the trajectories of the two time series intersected multiple times. This fact underscores the impact of speculative market dynamics on short- and mid-term returns (Bogle (2017)).

In order to expand the range of this analysis, carried out through cumulative logarithmic returns, it is of great interest to repeat the same process for different windows of time. Herein, the second illustration (Figure 6) narrows the focus to the time horizon starting from 1957, which (recalling from Chapter 2) represents the official inception year of the S&P 500 Index as it is known today. This reference point is particularly useful in removing potential bias and instrumental inaccuracies potentially, stemming from the backward computation of the Index prior to its launch date.

Cumulative logarithmic returns from 1957 to the present show comparable dynamics with respect to the previous plot (Figure 5). However, under this time frame, the movements of the two time series become more pronounced, highlighting periods of occasional divergence. Specifically, Figure 6 vividly pictures intervals of volatility, particularly during economic downturns, where earnings temporarily plummeted (as visualized by the sharp decline in the purple line) while the S&P 500's price exhibited a more muted response. This divergence can be indicative of the market's anticipatory nature, combined with the presence of non-earnings-related factors influencing investor sentiment (i.e. speculative dynamics). Conversely, the rapid ascents observed in both lines (e.g. after the subprime mortgage crisis) suggest robust economic expansion phases

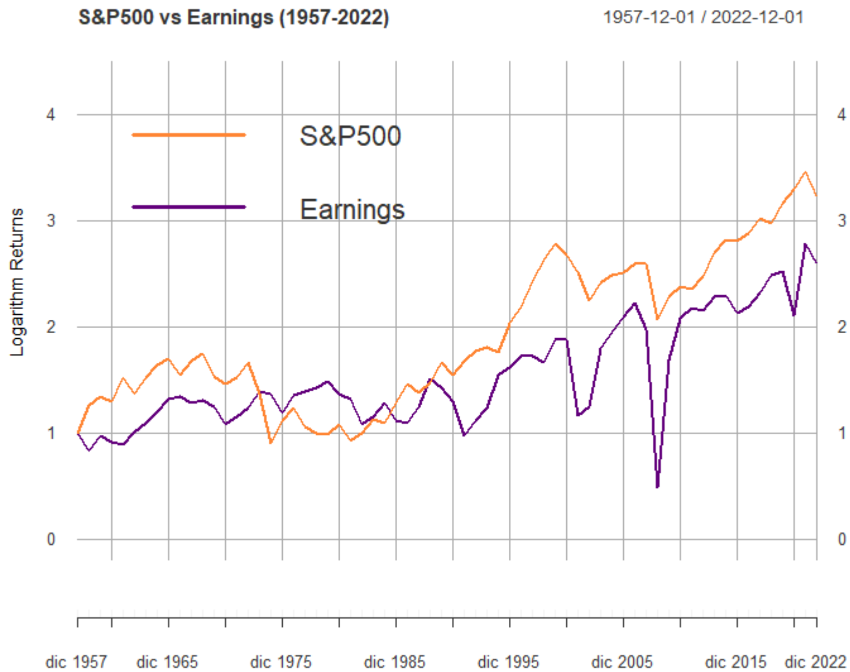

Figure 6: Total vs. Earnings Cumulative Logarithmic Returns

where market valuations and earnings ascend in tandem. It is interesting to note that, under this time frame, the S&P 500 grew ≈ 64% more than the underlying companies' earnings, implying that, on average, the Index provided an additional investment return of around 1% for the period ranging from 1957 to 2022[17].

To conclude this analysis, a final insightful focus is provided on a more recent and restricted time period, ranging from the year 2000 to the present day. Figure 7 underlines more closely both the S&P 500 Index's price and earnings reactions to the burst of the dot.com bubble at the beginning of the century, followed as well by the 2008 financial crisis. Remarkably, this plot presents a contrasting trend compared to earlier periods. Indeed, as can be noticed from Figure 7 , in the 21^{st} century the earnings of S&P 500 companies have exceeded the growth in the Index's price by a margin of ≈ 27%. Recalling the financial leverage effect exposed in earlier sections, this downward CAPE trajectory translates into lower portfolio returns for investors, with respect to relative earnings growth.

[17]It will be shown in the next section that this value is coherent with the average yearly impact of speculative returns throughout the whole dataset analysed.

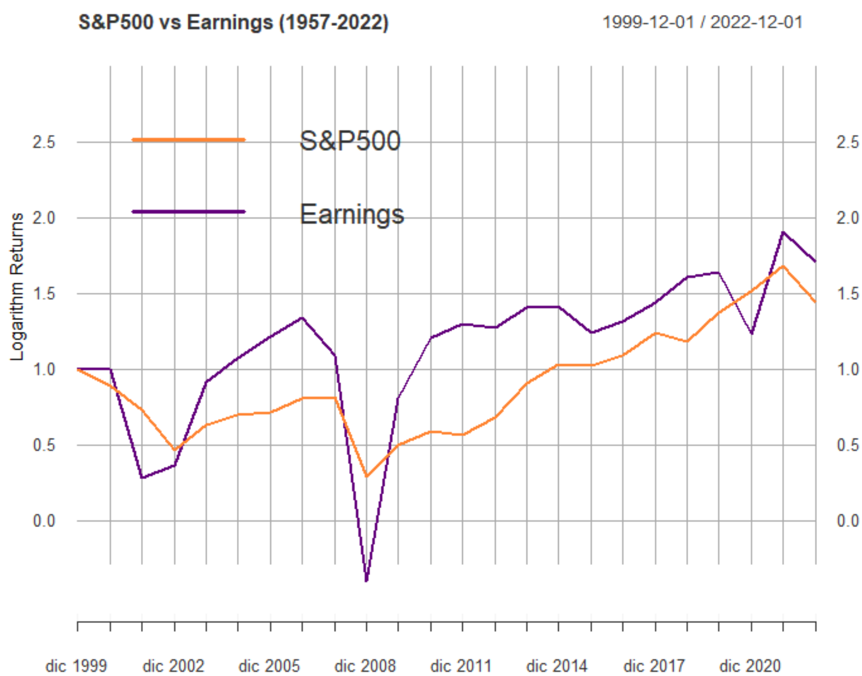

Figure 7: Total vs. Earnings Cumulative Logarithmic Returns

2.5 Decadal Returns Analysis

The preceding sections of the paper primarily focused on the analysis of monthly and annual data. To bring this chapter to a more comprehensive conclusion, it is insightful to extend the scope to longer time frames. Therefore, this section will segment returns into distinct decades, ranging from 1890 until present. This approach facilitates a clear and systematic presentation of average market returns, as well as picturing the evolution of the CAPE ratio for each decade. Additionally, this perspective aims to enhance the depth of the analysis presented in this Chapter and to unveil patterns and trends emerging over these historical intervals. The code provided below explains how the dataframe containing decadal data (**decade.dft**) was built, drawing from yearly data (**year.dft**). Specifically, it is worth noting that the following results represent the averages of yearly logarithmic returns.

```
decade.dft <- year.dft[-1, ] %>%
  group_by(Decade) %>%
  summarise(Total.logreturns = mean(Total.logreturns) %>%
                 multiply_by(100) %>% round(digits = 2),
```

```
    Earning.logreturns = mean(Earning.logreturns) %>%
        multiply_by(100) %>% round(digits = 2),
    Speculative.logreturns = mean(Speculative.logreturns) %>%
        multiply_by(100) %>% round(digits = 2))
```

Figure 8 displays a bar chart offering an extended overview of the S&P 500 performance in every decade. Specifically, each bar represents the average yearly Index logarithmic returns. Notice that, as for the cumulative returns approach, data is presented through a logarithmic scale, as to deal with the same potential issues aforementioned.

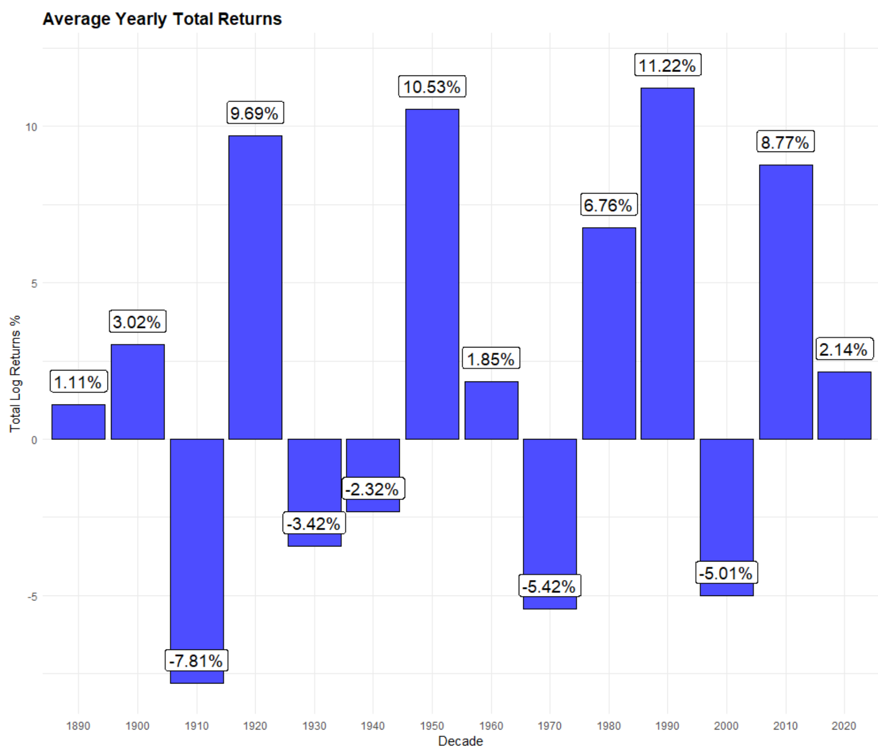

Figure 8: Decadal S&P 500 Total Returns

This analysis of the S&P 500 returns constitutes the basis for assessing decadal organic and speculative components. Herein, average yearly logarithmic earnings growth can be computed and graphically presented (Figure 9) in a similar fashion, to picture the impact of underlying companies' profitability growth against average market total returns.

As already noticed throughout the cumulative logarithmic returns analysis, there are some rather significant discrepancies between total returns and earnings growth. This overall effect has been referred to as speculative returns. To this extent, Figure 10

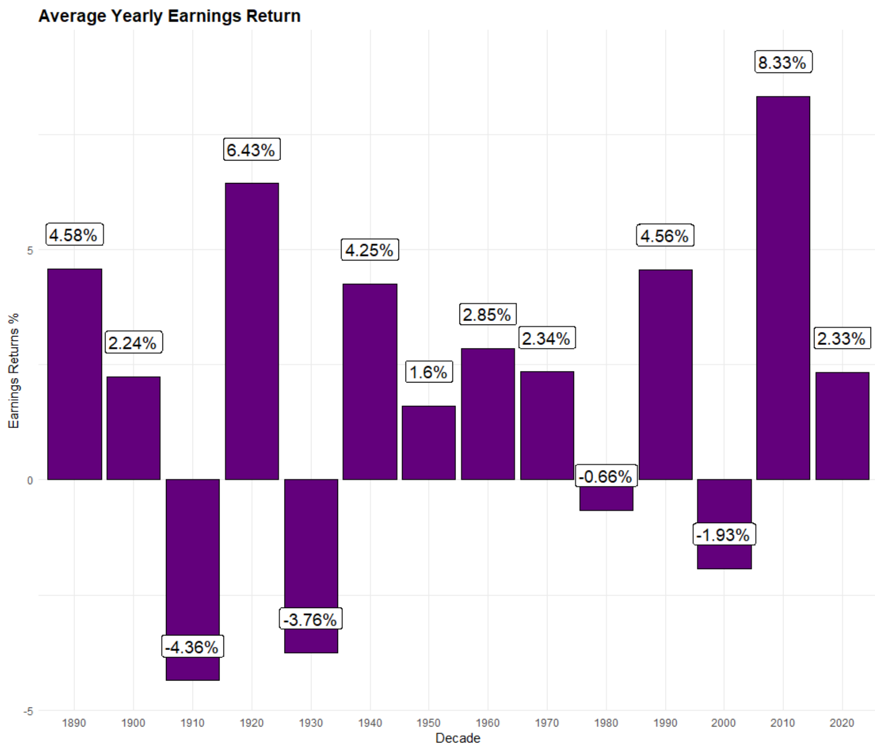

Figure 9: Decadal S&P 500 Earnings Growth

provides a comprehensive representation of the actual impact of speculative dynamics on the S&P 500 Index returns, displaying the impact that the "financial leverage" embedded in the nature of CAPE delivered across different decades.

This graph provides utterly important pieces of evidence in portraying the discrepancy between S&P 500 returns and relative earnings growth, emphasising the different magnitude of speculative returns. Surprisingly, overall speculative returns appear to cancel each other out, in aggregate terms. As computed in the code below, from the period of time ranging from 1890 to 2022, speculative returns reported an average yearly logarithmic value of 0.165%.

```
mean(decade.dft$Speculative.logreturns) %>% paste("%", sep = "")

 [1] "0.165%"
```

This fact is coherent with the arguments exposed in Bogle (2017), stating that, under a long-term time horizon, total market returns ultimately align with the organic component. On the other hand, Figure 10 likewise proves that short- and mid-term

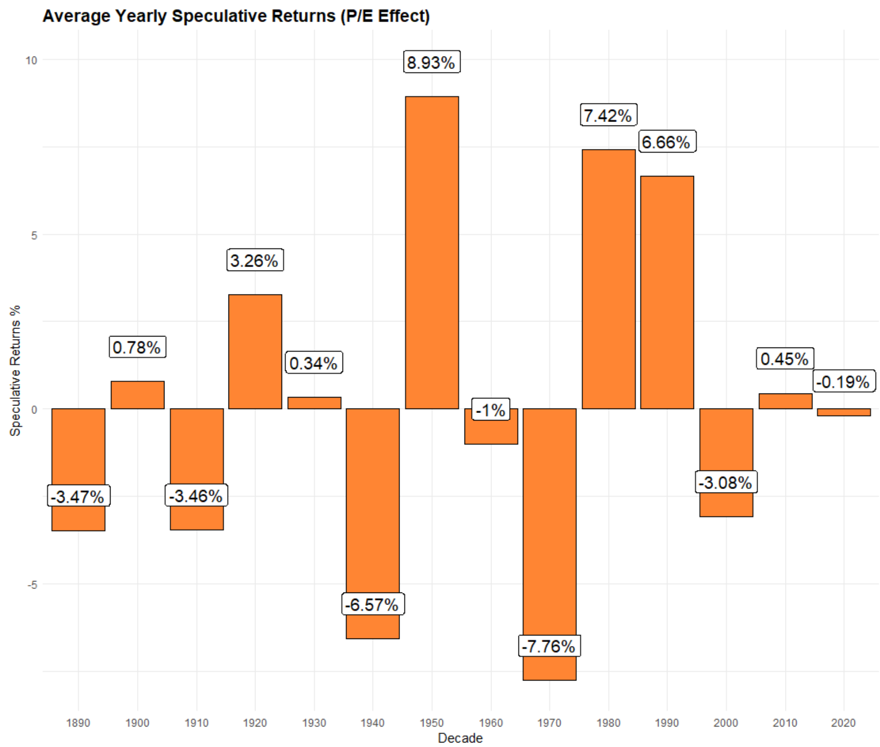

Figure 10: Decadal S&P 500 Speculative Returns

investments are highly exposed to a set of non-companies-centred factors, providing a significant positive or negative leverage to portfolio returns. Additionally, this piece of evidence also constitutes a compelling argument backing the academic literature (e.g. Dreman and Berry (1995) and Huang et al. (2007)) advocating for the presence of tangible mean reversion processes in the context of the CAPE (P/E) ratio.

The 3 plots commented on above (Figures 8, 9 and 10) provided an insightful and comprehensive overview of the various dynamics occurring at different historical stages, highlighting a notable discrepancy between S&P 500 returns and the earnings growth of underlying assets. The pertinent conclusion is encapsulated in Figure 10, which accurately depicts the impact of CAPE leverage across decades, from 1890 to the current 2020 decade.

To bring this chapter to a valuable conclusion, Figure 11 enhances the interpretation of CAPE's box plot previously discussed in this study (see Figure 4). This expanded analysis demonstrates and reaffirms considerable temporal fluctuation and heterogeneity in market valuations, with median CAPE values exhibiting significant variability, as well as pronounced elevation during the 2000s. Notably, decades of compressed CAPE

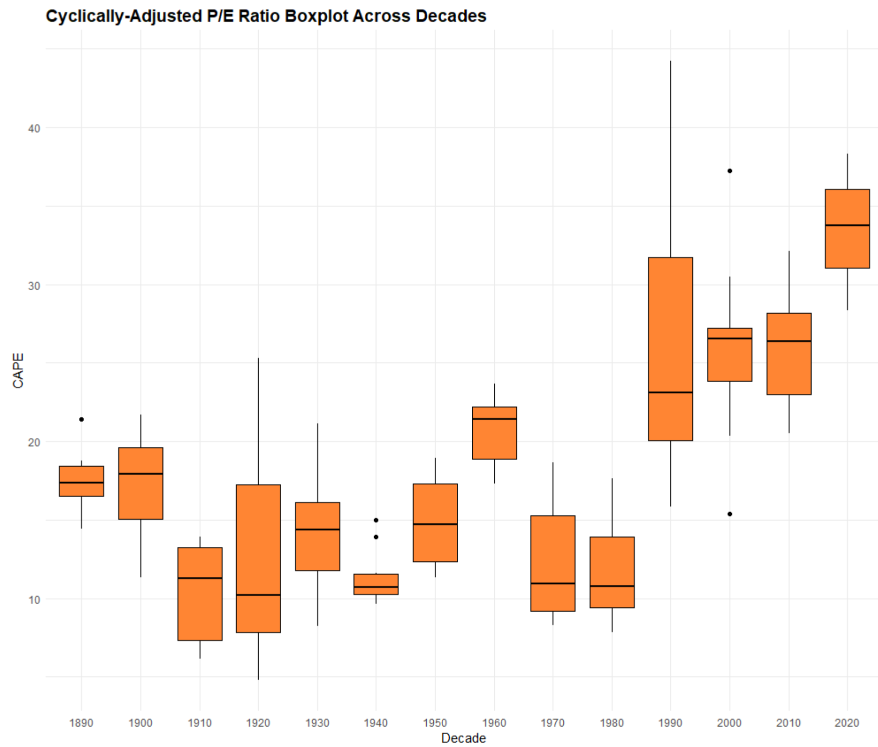

Figure 11: Decadal S&P 500 Speculative Returns

shifts alternate with periods of great volatility, with upward movements appearing to provide the most fluctuations[18]. Additionally, the presence of outliers in the data further underscores periods of exceptional valuation deviations.

Following the overall conclusion discussed in this section, Chapter 3 will assess whether, using a different set of statistical models, it is viable to forecast future swings in the CAPE ratio and, subsequently, predict the impact of speculative returns on the total S&P 500 returns. Accordingly, Chapter4 attempts to incorporate the evidence produced in an investment strategy, exploiting the financial leverage effect provided by positive speculative returns, as previously discussed.

[18]This piece of evidence can be interpreted by Figure 11 interquartile ranges. Specifically, 1920s and 1990s seemingly experienced the most pronounced period of enhanced market valuation, as demonstrated by the extended upper-quartile section of corresponding boxplots.

3 Financial Modelling - CAPE Forecasts

3.1 Introducing Behavioural Variables

This chapter aims to test the viability of building a statistical model capable of successfully forecasting future CAPE ratio value movements. Accordingly, this study argues that recalling the interpretation of this metric as a proxy for speculative market dynamics, investors' sentiments play a notable role in influencing CAPE levels and directly impact the S&P 500 returns. Declinations of this proposition were widely tested in the academic literature presented (Fisher and Statman (2000) and 2003, W. Y. Lee et al. (2002), Bathia and Bredin (2012), Smales (2017), Campisi and Muzzioli (2020), Akintoye (2008), Wang et al. (2021)), which found consistent evidence of fear and greed components of investors' sentiment having a tangible repercussion on the forward economic outlook.

It is worth providing a clear perimeter of the types of variables tested by other studies and explaining the choice of the parameters used in this work, as it could be a potential source of ambiguity. Literature has recognised embedded behavioural components in a multitude of metrics, such as the PCR (Bathia and Bredin (2012)), as reflecting the proportion between short-side versus long-side investors, as well as CAPE and P/E themselves (Dreman and Berry (1995), Weigand and Irons (2005)), as measures of the potential over- and underappreciation of the market. These variables could be referred to as "indirect" behavioural metrics since they do not represent a direct measure of sentiment- and behavioural-related factors, but rather a secondary impact they have on variables relating to different underlying components. Diverging from this approach, this study mainly focuses on behavioural-centred measures, primarily encompassing indexes and surveys tracking different sides of investors' sentiment. Herein, the dataframe pertinent to confidence and valuation surveys, as will be explained below, is maintained once again by Robert Shiller, the same source of the S&P 500 dataset used in Chapter 2. Indeed, he is also the director of Yale University's Investor Behavior Project, which he claims being "the longest-running effort to measure investor confidence and related investor attitudes"Yale School of Management, International Center for Finance (2024), with observations dating back to 1989[19].

[19]To this extent, it is worth noting that, even though the dataset regarding individual and institutional investors' confidence and valuation indexes does extend back to 1989, it does so with varying observations' frequency. Indeed, observations used to be collected biannually (in April and October of every year) until July 2001, which represents the beginning of the monthly collection of observations.

The snippet of code provided below aims to present an overview of the dataset of variables used by this study, which encompasses behavioural surveys retrieved from the dataset mentioned above and sentiment-related indexes (CCI, BCI, VIX), in addition to a general macroeconomic variable (UNRATE).

```
summary(regressiondata)

      Date                  Real Price          PE               CAPE3      
 Min.   :2001-07-01   Min.   :1086   Min.   : 13.50   Min.   :11.50  
 1st Qu.:2006-12-24   1st Qu.:1794   1st Qu.: 18.15   1st Qu.:20.79  
 Median :2012-06-16   Median :2129   Median : 22.03   Median :23.66  
 Mean   :2012-06-16   Mean   :2505   Mean   : 25.51   Mean   :23.71  
 3rd Qu.:2017-12-08   3rd Qu.:3161   3rd Qu.: 24.90   3rd Qu.:26.43  
 Max.   :2023-06-01   Max.   :5123   Max.   :123.73   Max.   :33.97  

     CAPE5              CAPE        Confidence.inst  Confidence.ind 
 Min.   :11.38   Min.   :13.32   Min.   :61.54    Min.   :58.16  
 1st Qu.:21.76   1st Qu.:23.13   1st Qu.:72.72    1st Qu.:68.51  
 Median :24.71   Median :26.19   Median :76.70    Median :75.00  
 Mean   :24.17   Mean   :26.16   Mean   :76.51    Mean   :75.28  
 3rd Qu.:26.42   3rd Qu.:28.94   3rd Qu.:80.57    3rd Qu.:81.86  
 Max.   :35.34   Max.   :38.58   Max.   :92.59    Max.   :95.62  

 Valuation.inst  Valuation.ind       UNRATE             CCI        
 Min.   :36.71   Min.   :28.57   Min.   : 3.400   Min.   : 96.13  
 1st Qu.:57.68   1st Qu.:47.95   1st Qu.: 4.575   1st Qu.: 98.52  
 Median :66.28   Median :59.67   Median : 5.400   Median : 99.73  
 Mean   :65.40   Mean   :56.99   Mean   : 5.926   Mean   : 99.57  
 3rd Qu.:74.86   3rd Qu.:66.90   3rd Qu.: 6.900   3rd Qu.:100.88  
 Max.   :88.76   Max.   :81.82   Max.   :14.700   Max.   :101.64  

      BCI               VIX       
 Min.   : 95.77   Min.   : 9.51  
 1st Qu.: 99.40   1st Qu.:13.99  
 Median :100.05   Median :17.79  
 Mean   :100.03   Mean   :20.01  
 3rd Qu.:100.80   3rd Qu.:23.81  
 Max.   :102.12   Max.   :59.89  
```

In order to provide a better explanation of the variables included, it is worth commenting on the most important aspects of the dataframe.

Herein, to guarantee consistency in the data, observations preceding June 2001 have been discarded.

- **Real price**: as in previous sections, it is the inflation-adjusted S&P 500 Index price
- **PE**: spot price-to-earnings ratio, computed using earnings of the correspondent fiscal year instead of a broader moving average of the same variable;
- **CAPE3** and **CAPE5**: represent variations of the CAPE ratio, computed as a moving average of the past 3 and 5 years of real earnings of the underlying companies, respectively;
- **Confidence.inst** and **Confidence.ind**: report the percentage of institutional and individual respondents (respectively) who expected an increase in the Dow Jones Industrial Average Index[20] for the following year;
- **Valuation.inst** and **Valuation.ind**: similarly, show the percentage of institutional and individual respondents (respectively) who thought that current stock prices (at the time of the survey) were either low or correctly valued;
- **UNRATE**: represents the unemployment rate for the U.S., sourced from the FRED (Federal Reserve Economic Data). Notably, the unemployment rate is also strictly tied to a profound social aspect, besides being a widely used macroeconomic variable. For this reason; it may be considered an indirect indicator of a more generalised population sentiment;
- **CCI** and **BCI**: Consumer Confidence Index and Business Confidence Index are significant economic indicators that represent the overall confidence levels of consumers and businesses within an economy[21], respectively. These two variables are sourced from the OECD (Organization for Economic Co-operation and Development);
- **VIX**: the Implied Volatility Index, calculated and published by the Chicago Board Options Exchange (CBOE), measures the stock market's expectation of volatility based on S&P 500 index options, over the next 30-day period (Chicago Board

[20] The Dow Jones Industrial Average ("DJIA", also known as "the Dow") is the oldest US stock market index. It is one of the 2 most followed indexes, alongside the S&P 500. The Dow encompasses 30 of the largest companies based in the United States.

[21] CCI measures how optimistic or pessimistic consumers are regarding their expected financial situation, as well as their country's economic situation in the short term. Similarly, BCI reflects the level of optimism or pessimism that business executives feel about the prospects of their companies and the overall economy. High index values indicate an increased optimism, and vice versa

Options Exchange (2024)). This index was notably proposed as an explanatory variable in Smales (2017), as a proxy for investor fear[22].

At the beginning of this chapter, it was mentioned that the scope of this work was to test whether behavioural variables carried any explanatory power in forecasting future n-step forward CAPE values. Specifically, this chapter tests this proposition for a 1-month ahead CAPE forecast. The dataset has been split into two distinct parts, as best practice to create convenient training and test sets. To this extent, the scientific literature does not provide clear indications relative to the proportion of observations to be attributed to either of these two sets. Toleva (2021) indicates that a 70/30 split particularly improved issues of classification accuracy in the context of the data frame analysed. On the other hand, Joseph (2022) suggests a ratio (between training set and test set) of $\sqrt{p}$:1, where p represents the number of statistically significant parameters of the linear regression model built on the dataset of interest. However, both studies highlight that there is not a one-fits-all measure and that this proportion should be adjusted according to the size of the dataset under consideration. Herein, a common rule of thumb in the field of machine learning and statistical modelling suggests a split of 80/20. Given the not-so-large dimension of the dataset presented (264 observations), this study decides to allocate the first 80% of the dataframe for model training purposes and the latest 20% of observations to the test set.

3.2 Linear Regression Models with Behavioural Variables

In order to test the relevancy of the behavioural variables aforementioned on one-month ahead values of the CAPE ratio, this study employs a linear regression model, defined as follows:

$$\begin{aligned} CAPE_t = \beta_0 &+ \beta_1 \times Confidence.inst_{t-1} + \beta_2 \times Valuation.inst_{t-1} \\ &+ \beta_3 \times Confidence.ind_{t-1} + \beta_4 \times Valuation.ind_{t-1} + \beta_5 \times VIX_{t-1} \\ &+ \beta_6 \times BCI_{t-1} + \beta_7 \times CCI_{t-1} + \beta_8 \times UNRATE_{t-1} + \epsilon_t \end{aligned}$$

Notably, all of the explanatory variables are lagged for one period, with respect to the independent variable (**CAPE**). In so doing, this process effectively tests whether the

[22]Indeed, VIX is often referred to as "fear index". The reason behind this is that this index is derived from the implied volatility underlying price inputs of S&P 500 index options. Hence, high values of the VIX indicate increased uncertainty in the market, as well as investors fearing significant changes in the S&P 500 Index levels.

set of information explored in the preceding section may be able to explain forward movements of the CAPE ratio, to any extent. Herein, *model*1 is defined from the formula provided above. Running the linear regression testing via R yields the following results:

```
Call:
lm(formula = f.CAPE ~ Confidence.inst + Valuation.inst + Confidence.ind
   + Valuation.ind + VIX + BCI + CCI + UNRATE, data = trset)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0821 -0.7697 -0.0668  0.6372  5.5348

Coefficients:
                 Estimate | Std. Error | t value | Pr(>|t|)
------------------------------------------------------------
(Intercept)     -59.00315 |   18.05916 | -3.267 |   0.00128 **
Confidence.inst  -0.05341 |    0.02252 | -2.372 |   0.01864 *
Valuation.inst   -0.07018 |    0.01722 | -4.075 |  6.63e-05 ***
Confidence.ind    0.07241 |    0.02053 |  3.527 |   0.00052 ***
Valuation.ind    -0.04499 |    0.02163 | -2.080 |   0.03877 *
VIX              -0.09568 |    0.01746 | -5.481 |  1.26e-07 ***
BCI               0.73163 |    0.12130 |  6.032 |  7.67e-09 ***
CCI               0.24801 |    0.17001 |  1.459 |   0.14617
UNRATE           -1.00955 |    0.09210 |-10.961 |   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.314 on 201 degrees of freedom
Multiple R-squared:  0.8849,Adjusted R-squared:  0.8803
F-statistic: 193.2 on 8 and 201 DF,  p-value: < 2.2e-16
```

As can be noticed from the output of the linear regression provided above, *model*1 appears to show a notable fit, displaying an *adjusted* R^2 of 88.03%, paired with the statistical significance of all variables but one (namely: CCI). However, in order to make any type of significant inference, the model first needs to be evaluated and validated to ensure its reliability and accuracy in making predictions. This phase, known as model diagnostic, involves a series of steps designed to assess the model's performance and identify any potential issues that could compromise its effectiveness. Notably, model diagnostic largely involves the examination of residual plots, in order to spot patterns that might suggest violations of the model's assumptions. Additionally, the diagnostic process also requires the use of statistical tests and criteria to evaluate the model's overall fit and the significance of individual regressors. Overall, model diagnostic is primarily focused on spotting 3 potential statistical issues, notably: non-stationarity, heteroskedasticity and autocorrelation.

- **Non-stationarity**: arises when the statistical characteristics of the residuals change over time. Stationarity of residuals is a fundamental assumption in regression analysis and, specifically, encompasses a constant mean, constant variance and no autocorrelation. The presence of non-stationarity in residuals poses several issues for regression analysis, as it can lead to biased or inconsistent parameter estimates, undermining the interpretability and reliability of the model. Moreover, non-stationarity can also impair the model's predictive accuracy, in case of trends/patterns that the model is not able to account for;

- **Heteroskedasticity**: occurs when the variance of the error terms of a regression model is not constant across all levels of the independent variables, violating one of the key assumptions of ordinary least squares (OLS) regression. This condition may lead to inefficient estimates and affect the reliability of hypothesis tests due to incorrect standard errors;

- **Autocorrelation**: also known as serial correlation, refers to the correlation of a time series with its own past (and future) values. Herein, it quantifies the degree to which current values of the series are related to previous observations over different time lags. Autocorrelation may lead to biased statistical inferences if not properly accounted for. and models assuming independent errors can be inappropriate if they show autocorrelated residuals. Autocorrelation may indicate model inadequacies, such as the omission of important predictors or the presence of a more complex dynamic structure not captured by the model.

As mentioned above, model diagnostic combines both graphical and quantitative elements to spot either of the potential issues presented. Herein, Figure 12 exploits the *checkresiduals* functions from the $forecast$ package to provide a comprehensive overview of $model1$ residuals, picturing the relative trajectory, autocorrelation function (ACF[23]) and distribution function.

[23]In the context of model diagnostics, the ACF is a tool used to analyze the correlation between observations of a time series separated by various time lags. ACF plot is a crucial graphical tool for

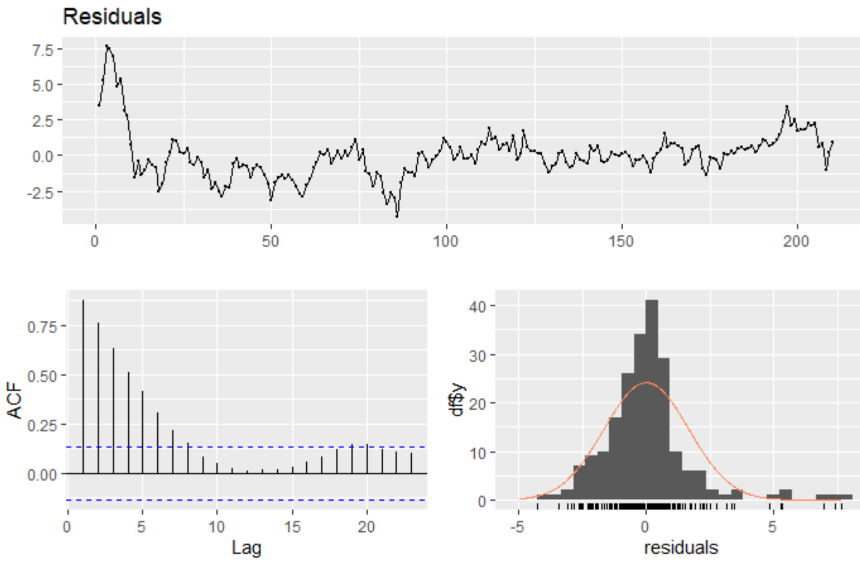

Figure 12: Model Diagnostic

At first glance, the plot in the bottom-left corner (the ACF of residuals from $model1$) seems to show evidence of autocorrelation. This inconvenience can be graphically spotted since autocorrelation between residuals at different lags of time decreases slowly and shows notable values (graphically represented by the bars' height) up to 8 lags. A common statistical approach to test for the presence of autocorrelation in a dataset is represented by the Ljung-Box test (LB). The Ljung-Box test (Ljung and G. E. P. Box (1978)) is a statistical approach designed to evaluate the presence of autocorrelation in the residuals of a time series regression model, at several (k) lags. Specifically, the LB test statistic (Q) is computed as follows (as outline in Ruppert and Matteson (2015)):

$$Q = n(n+2)\sum_{l=1}^{k}\frac{\hat{\rho}_l^2}{n-l}$$

where:

- n is the sample size,
- $\hat{\rho}_l$ is the sample autocorrelation at lag l,
- k is the number of lags being tested.

Under the null hypothesis of no autocorrelation up to the k^{th} lag, the test statistic

identifying the presence of autocorrelation.

Q follows a chi-squared (χ^2) distribution with $k-p$ degrees of freedom (where p indicates the number of predictors in the model excluding the intercept). Confirming the graphical intuition stemming from Figure 12, running the LB test on $model1$ residuals yields a relative p-value $< 2.2e^{-16}$, confirming autocorrelation suspects[24].

```
Box.test(model2$residuals, fitdf = 4, lag=10, type="Ljung")
```

```
Box-Ljung test
data:  model2$residuals
X-squared = 337.26, df = 6, p-value < 2.2e-16
```

Moreover, the plot on the top row of the same model diagnostic tool (Figure 12) may hint at (limited) heteroskedasticity phenomenons (especially in the first observations of the dataset). This intuition can be graphically noticed from the large downward variations seemingly characterising the initial part of the residuals, while the second half seems to be steadier, implying different volatility values. Several statistical tests can be used in order to assess heteroskedasticity in a given dataset; one of the most famous (and the one this study exploits) is the Breusch-Pagan (BP) test. The BP test is a statistical procedure used to detect the presence of heteroscedasticity in the residuals of a regression model[25]. Specifically, it tests the null hypothesis of constant variance in the model's residuals (homoskedasticity), against the alternative hypothesis of the variance being dependent on one or more independent variables, indicating heteroskedasticity issues. A statistically significant result from the BP test indicates evidence of heteroskedasticity, suggesting that the model's residuals' variance changes across levels of the independent variables. As shown in the code below, $model1$ exhibits a p-value for the BP test of $1.889e^{-10}$ (largely statistically significant at the 0.05 significance level) confirming the suspects of heteroskedasticity graphically hinted by Figure 12.

[24] In this study, the significance level is set at 0.05.

[25] The test involves first estimating the original regression model and obtaining the squared residuals. Then, a new auxiliary regression is performed where these squared residuals are regressed on the original independent variables (and potentially their transformations or additional variables). The test statistic is derived from this auxiliary regression, typically following a chi-square distribution, which is then used to determine whether the null hypothesis of homoscedasticity can be rejected Johnston and DiNardo (1997).

```
Studentized Breusch-Pagan test
data:  model1
BP = 61.996, df = 8, p-value = 1.889e-10
```

The diagnostic process relative to *model*1 assessed the presence of autocorrelation and heteroskedasticity in the residuals, thus providing evidence of a flawed regression model. To this extent, model diagnostic is of fundamental importance, as it suggests that *model*1 cannot be trusted to directly produce forecasts about future CAPE values and make inferences about the model's parameters.

Financial data often exhibits autocorrelation; hence, this finding is not surprising. There are several ways to deal with these inconveniences and adjust *model*1 in order to fix the statistical issues arising from these two circumstances. Specifically, the next sections will address these points in two different ways:

- Exploiting robust standard errors: heteroskedasticity and autocorrelation consistent (HAC)[26] estimators are used to mitigate the impact of autocorrelation and heteroskedasticity in the error terms of a regression model (Ruppert and Matteson (2015)). Specifically, this study relies on the Newey-West covariance matrix, which extends the concept of robust standard errors by adjusting the covariance matrix of the model's coefficients to account for potential autocorrelation, as well as possible non-constant variance in the error terms;

- Estimating a linear regression model with ARIMA errors: this statistical process combines the traditional linear regression framework with an ARIMA (AutoRegressive Integrated Moving Average) model for the error terms.

It is worth noticing that the methodologies mentioned above and followed by this study to adjust the model's error structure are not the only viable options for proceeding in such occurrences. For example, the presence of either non-stationarity, heteroskedasticity or autocorrelation in residuals may first indicate the absence of an explanatory variable which could capture (i.e. is responsible for) otherwise unexplained volatility movements in the observations. Herein, one way of proceeding is testing additional vari-

[26]Traditional ordinary least squares (OLS) regression assumes that the error terms are independently and identically distributed with constant variance. In this context, HAC estimators are used to adjust standard errors for the presence of heteroskedasticity and autocorrelation within the error terms. Specifically, they provide a method to correct the covariance matrix of the regression coefficients, ensuring that hypothesis tests and confidence intervals remain valid even when the classical assumptions are violated.

ables in order to check whether the implemented model works better and presents more adequate properties than the preceding one. It is also worth mentioning that another interesting way of handling statistical issues in a linear regression model's residuals is represented by implementing a dynamic regression process instead of HAC estimators (Baillie et al. (2022)). However, as the next section and Chapter 4 show, the two methodologies implemented by this study provided satisfactory results; thus, these latter options were not pursued as well.

3.3 Dealing with Autocorrelation and Heteroskedasticity: Newey-West Robust Covariance Matrix

The Newey-West (NW) estimator (Newey and West (1987)) provides a way to calculate consistent covariance matrices in the presence of both serial correlation and heteroskedasticity in a model's residuals. The implementation of the Newey-West robust covariance matrix is a standard practice and it is particularly pertinent in the financial environment, given the propensity for financial time series to display autocorrelation. Specifically, the Newey-West estimator draws from the aggregate of squared residuals in order to mitigate autocorrelation, as follows:

$$\hat{\Omega}_{NW} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \mathbf{x}_t \mathbf{x}_t' + \frac{1}{T}\sum_{l=1}^{L} w_l \left(\sum_{t=l+1}^{T} \hat{u}_t \hat{u}_{t-l} (\mathbf{x}_t \mathbf{x}_{t-l}' + \mathbf{x}_{t-l} \mathbf{x}_t') \right)$$

where:

- $\hat{\Omega}_{NW}$ is the Newey-West adjusted covariance matrix.
- T is the total number of observations.
- $\hat{u}_t$ is the residual at time t.
- $\mathbf{x}_t$ is the vector of regressors (including any constant term) at time t.
- L is the lag length, chosen by the researcher.
- w_l is a weight function for lag l, which typically decreases as l increases to reduce the influence of distant observations[27].

[27] A common choice is the Bartlett weight: $w_l = 1 - \frac{l}{L+1}$.

While the econometric process behind the computation of the NW estimator may be complex and extensive, its implementation is straightforward in the R environment. The function $NeweyWest$ from the *sandwich* package directly estimates the NW robust (HAC) covariance matrix. Herein, $model2$ is derived by applying the Newey-West test on the former $model1$ specification. The resulting NW estimator can then be used as a reference to create a new model ($model2$), whose variance-covariance matrix[28] will work under the assumption to be HAC. The code provided below shows how it is possible to do so, specifying the base model and modifying its error terms with the NW estimator.

```
coeftest(model1, vcov. = NeweyWest(model1, adjust = T, verbose = T,
                                   diagnostics = T, prewhite = F))

 t test of coefficients:

                   Estimate | Std. Error | t value |  Pr(>|t|)
 -------------------------------------------------------------
 (Intercept)     -59.003148 |  50.854302 | -1.1602 |  0.247328
 Confidence.inst  -0.053407 |   0.043306 | -1.2333 |  0.218920
 Valuation.inst   -0.070179 |   0.035791 | -1.9608 |  0.051285 .
 Confidence.ind    0.072414 |   0.067487 |  1.0730 |  0.284557
 Valuation.ind    -0.044995 |   0.040560 | -1.1093 |  0.268610
 VIX              -0.095677 |   0.030367 | -3.1507 |  0.001877 **
 BCI               0.731634 |   0.321724 |  2.2741 |  0.024016 *
 CCI               0.248009 |   0.293155 |  0.8460 |  0.398558
 UNRATE           -1.009548 |   0.143526 | -7.0339 | 3.083e-11 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Applying the Newey-West estimator shows that the only variables still retaining statistical significance at the 95% confidence level are VIX, BCI and $UNRATE$. All of the other regressors appear to be disregarded by this test, implying that their presence in the regression model may be linked to the error term presenting either autocorrelation or heteroskedasticity issues. For this reason, the non-statistically significant variables need to be removed from the model specification by re-estimating the model. Herein, the output of the final regression, consistent with the NW estimator, is represented in

[28] Note that the Newey-West estimator does not need to be directly incorporated in the original model. Indeed, inference about future predictions and interpretation of parameters does not depend on the error term. However, HAC standard errors provided by the NW estimator need to be used under the statistical inference framework (e.g. hypothesis testing, constructing confidence intervals).

the code below[29].

```
Call:
lm(formula = f.CAPE ~ VIX + BCI + UNRATE, data = trset)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2608 -0.8419 -0.0503  0.6034  7.6544

Coefficients:
               Estimate | Std.  Error | t value | Pr(>|t|)
-----------------------------------------------------------
(Intercept)   -82.51187 |    13.80782 |  -5.976 | 9.94e-09 ***
VIX            -0.08621 |     0.01984 |  -4.345 | 2.19e-05 ***
BCI             1.17599 |     0.13631 |   8.627 | 1.70e-15 ***
UNRTE          -1.40748 |     0.07218 | -19.499 |  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.666 on 206 degrees of freedom
Multiple R-squared:  0.8104,Adjusted R-squared:  0.8076
F-statistic: 293.5 on 3 and 206 DF,  p-value: < 2.2e-16
```

Since this model is now consistent with the NW estimator, autocorrelation and heteroskedasticity can be considered successfully addressed; $model2$ is then defined as:

$$CAPE_t = \beta_0 + \beta_1 \times VIX_{t-1} + \beta_2 \times BCI_{t-1} + \beta_3 \times UNRATE_{t-1} + \epsilon_t$$

It is then possible to start making inferences about important characteristics provided by $model2$. As previously mentioned, the adjusted R^2 usually represent one of the most important aspects of a statistical model, as it assesses the goodness of the model's fit on the training set used to construct it. In this case, the *adjusted* R^2 reports a value of 80.76%, which represents a largely remarkable result and provides evidence of the notable explanatory power of behaviour-linked factors (VIX, BCI), as well as the macroeconomic and social component ($UNRATE$). Another instance worth noting is represented by the coefficients displayed by the three variables:

- The negative VIX coefficient is coherent with the academic findings presented

[29] Running the NW test on this set of regressors provides that all of the independent variables are statistically significant at the 0.05 level.

in Chapter 1, underlining that increased uncertainty in future market outlooks (implied by increased volatility level) is negatively correlated with forward movements of the CAPE ratio, and vice versa;

- A significantly high BCI coefficient (1.17599) indicates that one-month ahead CAPE values appear to be directly correlated with shifts in the level of business executives' sentiment about the prospects of their companies and the overall economy. This finding produces evidence that a generalised optimism by business directors may not only indicate an increase in their companies' (aggregated) earnings but also a more-than-proportional response from the market;
- $UNRATE$ may be interpreted according to what was mentioned in the previous section: indeed, the notable $UNRATE$ negative coefficient (-1.40748) signals that increased unemployment is deeply correlated with a decline in the CAPE ratio. This result shows that economic recessions, proxied by rising unemployment rates, tend to hamper the equity market (in this case, recall, represented by the S&P 500) more than companies' earnings. This intuition (derived from the negative correlation between $UNRATE$ and CAPE) underlines the behavioural and sentiment component implied by this metric, as its impact is ultimately perceived more by market prices than by the relative organic component.

After having assessed the reliability of $model2$ through the application of the NW estimator and having made inferences about the meaning of the resulting coefficient, it is finally possible to exploit the regression model to try and forecast future expectations of one-month ahead CAPE values. For this purpose, the model's parameters are applied to the set of explanatory variables previously excluded from the training set (**trset**), hence stored in the test set (**teset**). This process enables the forecasting of one-step ahead expected CAPE values through $model2$. Figure 13 displays the result of the forecasting process, comparing one-month ahead predictions with the realised values of the one-month forward CAPE ratio.

It is possible to notice that the forecasts provided by this study seem to systematically underestimate the actual forward CAPE values. Even though this situation may generate some confusion, this is actually not surprising for a notable reason. In this study, $model2$ is estimated through a linear regression of a fixed and limited training set, encompassing observations from 07/2001 to 01/2019. This implies that the dataframe used to train the model is based on data which may potentially become outdated as

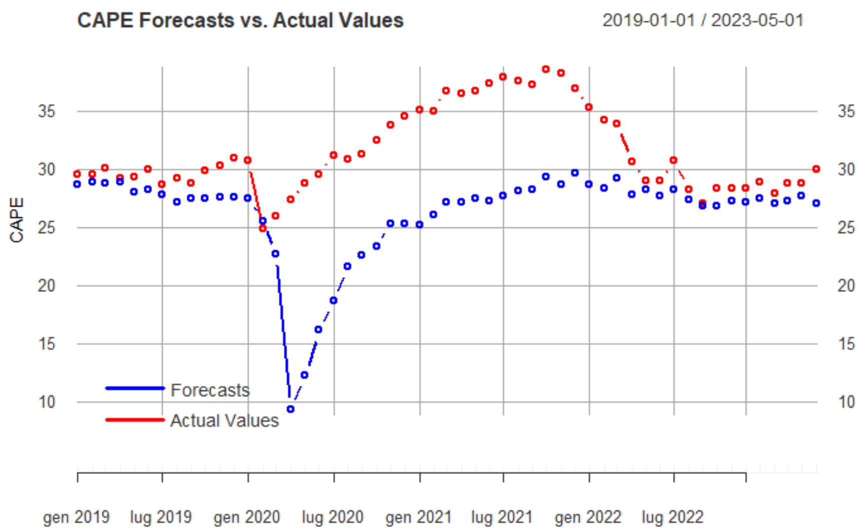

Figure 13: Test Set CAPE Forecasting

time passes. Additionally, the test set unfortunately retains observations related to the COVID-19 pandemic, which coincide (in Figure 13) with the time period of the furthest discrepancy between the model's estimates and realised values. To this extent, the model may require to be fine-tuned more dynamically, in order to encompass updated information every specified period of time and to adjust the model's parameters accordingly.

Nevertheless, and more importantly, the two issues acknowledged above do not influence the final scope of this study and, consequently, should not be the centre of additional adjustment efforts. The reason behind this is that recalling what mentioned in Chapter 1 and 2, this work aims to develop a statistical model to predict future movements of the CAPE ratio and to produce superior portfolio returns from there. Specifically, as Chapter 4 explains in depth, the latter objective does not involve accurate predictions of the nominal value of the CAPE ratio but rather draws on the direction of this metric's movements. Herein, even though *model*2 appears to have a negative bias which compromises the model's accuracy on the y-axis, it works well in predicting upward and downward CAPE shifts, as noted from the R^2 of the model (referring to backwards estimates), as well as graphically shown in Figure 13 (for future predictions).

For the reason mentioned above, running standard accuracy tests may result in misleading and counterproductive inferences. Indeed, computing main statistical accuracy metrics yields the results provided in the following snippet of code.

```
accuracy(pred, teset$f.CAPE)

                 ME     RMSE      MAE      MPE     MAPE
Test set 5.273902 6.993208 5.301739 16.07537 16.18754
```

Specifically, each accuracy metric presented refers to different aspects of the statistical model tested. The following paragraph aims to provide a clear overview of the pieces of information provided by these statistics. Given a set of observations y_t and corresponding forecasts $\hat{y}_t$, where T represents the number of observations of the dataset, the forecast accuracy metrics are defined as follows.

$$\text{Mean Error (ME)} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)$$

The ME is the arithmetic average of the residuals, thus the differences between predicted and observed values. It serves as an indicator of systematic bias in the predictions; a mean error of zero implies no bias, whereas a non-zero mean error indicates a tendency of the model to either overestimate or underestimate the observed values systematically. In this model, a mean error of approximately 5.27 suggests a tendency to overestimation (positive bias)-

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$$

The RMSE is a widely used measure of the average magnitude of predictive errors, particularly sensitive to large errors and providing an estimate of the standard deviation of the unexplained variance.

$$\text{Mean Absolute Error (MAE)} = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t|$$

The MAE represents the average of the absolute values of the residuals.

$$\text{Mean Percentage Error (MPE)} = \frac{100\%}{T}\sum_{t=1}^{T}\left(\frac{y_t - \hat{y}_t}{y_t}\right)$$

The MPE is the average of the percentage errors. It can reveal the tendency of the model to systematically overestimate or underestimate the observed values in percentage terms. An MPE of approximately 16.08, indicates that the average forecast is about 16.08% lower than the actual value

$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{100\%}{T} \sum_{t=1}^{T} \left(\frac{|y_t - \hat{y}_t|}{y_t} \right)$$

Finally, the Mean Absolute Percentage Error (MAPE) represents the mean of the absolute values of the individual percentage errors. This metric expresses the average error as a percentage of the actual values, which provides an intuitive understanding of the prediction accuracy in relative terms. It is particularly useful when comparing models across different scales. For this model, the MAPE is approximately 16.19, meaning that the average absolute error is 16.19% of the actual values.

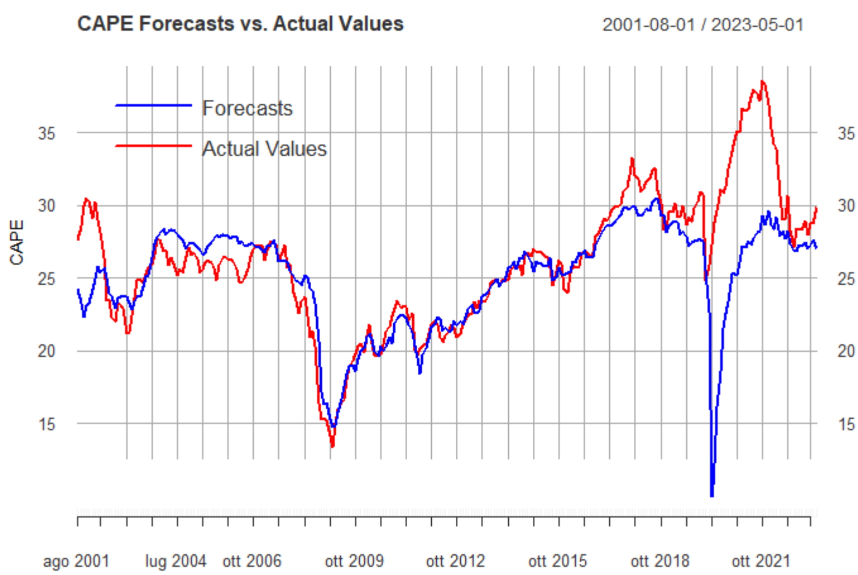

Figure 14: CAPE Forecasting on the complete dataset

As previously mentioned, these figures are not utterly relevant in the context of this study, since the scope of the model is to best capture CAPE trends and shifts, rather than precise values. As a last step, $model2$, needs to be applied to the whole dataset of observations available, in order to estimate the one-month ahead CAPE predictions that will constitute the basis for developing the investment strategies explored in Chapter 4.

Herein, Figure 14 shows the estimated one-step ahead CAPE predictions against actual values for the whole length of the dataframe.

For reference, the code below reports the same accuracy snippet previously mentioned for the application of $model2$ on the test set alone.

```
accuracy(pred2, regressiondata$f.CAPE)

                ME     RMSE      MAE      MPE     MAPE
 Test set 1.063409 3.474622 1.958478 2.888244 6.818769
```

To the extent of providing a clearer measure of accuracy for this model, one viable and more coherent solution could be represented by the Pearson correlation coefficient, which is not as influenced by vertical gaps as the accuracy metrics described above. The Pearson correlation coefficient (r) is computed as follows:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where:

- n is the number of data points;
- $\sum_{i=1}^{n} xy$ is the sum of the product of paired scores;
- $\sum_{i=1}^{n} x$ and Σy are the sums of the X scores and Y scores, respectively;
- $\sum_{i=1}^{n} x^2$ and Σy^2 are the sums of the squared X scores and Y scores, respectively.

Let x be the time series composed of the forecasts from $model2$ and y the time series referring to realised CAPE observations. Computing the r coefficient for $model2$ yields a value of 0.7014 confirming a significant correlation between the model's predictions and actual values from the test set.

3.4 Linear Regression Models with ARIMA Errors (ARIMAX)

In the previous section, autocorrelation found in residuals from $model2$ was dealt with by exploiting a HAC variance-covariance matrix, specifically provided by the Newey-West estimator. An alternative interesting process in this field (as explored by Hyndman and Athanasopoulos (2018) and Maggina (2011)) is represented by linear regression models with ARIMA (AutoRegressive Integrated Moving Average[30]) errors. This statistical methodology is a convenient approach when dealing with financial data particularly, as it draws on residuals from the linear regression model exhibiting autocorrelation, thus violating OLS usual assumption of independent errors. Herein, $model3$ is constructed starting from the same linear regression of the former model:

$$\begin{aligned} CAPE_t = \beta_0 &+ \beta_1 \times Confidence.inst_{t-1} + \beta_2 \times Valuation.inst_{t-1} \\ &+ \beta_3 \times Confidence.ind_{t-1} + \beta_4 \times Valuation.ind_{t-1} + \beta_5 \times VIX_{t-1} \\ &+ \beta_6 \times BCI_{t-1} + \beta_7 \times CCI_{t-1} + \beta_8 \times UNRATE_{t-1} + \epsilon_t \end{aligned}$$

However, in this case, the error term ϵ_t follows an ARIMA(p, d, q) process:

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d \epsilon_t = (1 + \sum_{j=1}^{q} \theta_j L^j)\eta_t$$

which is equivalent to

$$\phi(L)\nabla^d(L)(y_t - x_t'\beta) = \theta(L)\epsilon_t$$

where $y_t = CAPE_t$ and x_t' is the vector containing all of the model's regressors at time t.

This approach extends linear regression analysis by explicitly modelling the error structure to capture autocorrelation and non-stationarity within the residuals. In essence, after fitting a linear regression to explain the relationship between independent variables and the dependent variable, any residual autocorrelation is addressed by applying an ARIMA model to the residuals. This particular statistical approach is a further extension of the ARMA (AutoRegressive Moving Average) model with exogenous variables explored in G. Box et al. (2015), which literal sense indicates an ARMA model encompassing external variables (i.e. the independent variables of the regression model). For this reason, this specification is also commonly referred to as ARMAX.

[30] An AutoRegressive Integrated Moving Average (ARIMA) process is a forecasting model combining autoregressive features (i.e. where future values are assumed to be a linear function of past observations) with moving average components (i.e. where future values are modelled as a function of past errors). Additionally, the integration component is incorporated to make the data stationary, removing potential trends and seasonality. The model is specified by three parameters: p (autoregressive order), d (degree of differencing needed to make the series stationary), and q (moving average order), collectively represented as ARIMA(p,d,q).

Throughout this section, as well as in the following chapters, this study uses the term ARIMAX to refer to a linear regression model with ARIMA errors (as also in Hyndman and Athanasopoulos (2018)), indeed.

There are two main benefits stemming from applying an ARIMA process to model the error term of a linear regression:

- Enhanced accuracy of the parameter estimates of the regression model;
- Adjusted standard errors enable more reliable statistical inferences.

In the R environment, ARIMAX can be implemented directly through the function *auto.arima*, whose term *xreg* allows to specify the linear regression to be fitted first. To this extent, one important precaution is to transpose all of the regressors into a matrix, as the argument *xreg* only works with matrixes, indeed. It is also important to note that, through the *auto.arima* function, different potential models are automatically compared and selected by means of the Akaike or Bayesian Information Criterion[31] (AIC/BIC) and estimates are computed via maximum likelihood estimation, assuming a normal distribution. Let L being the maximized value of the likelihood function for the estimated model, k the number of estimated parameters in the model and n the number of observations in the dataset. Then, the two ICs are defined as follows:

$$\text{AIC} = -2\ln(L) + 2k$$

$$\text{BIC} = -2\ln(L) + k\ln(n)$$

In this study, both ICs provided the same final model. Herein, AIC is reported in the code provided below (as well as in the following sections).

[31] In the context of statistical modelling, information criteria (IC) are measures used to evaluate and balance the goodness of the fit of a model and the number of parameters included. In so doing, ICs contribute to the specification of the most appropriate final model.

```
reg <- trset %>% select(Confidence.inst, Confidence.ind, Valuation.ind,
                        Valuation.inst, CCI, BCI, UNRATE,
                        VIX) %>% as.matrix()
model2 <- auto.arima(trset$f.CAPE, max.p = 5, max.q = 5, stepwise = F,
                     stationary = F, seasonal = F, ic = 'aic',
                     test = 'adf', xreg = reg)
coeftest(model2)
```

```
 z test of coefficients:

                     Estimate | Std. Error | z value |  Pr(>|z|)
 --------------------------------------------------------------
 ar1               -0.5505264 |  0.3778928 | -1.4568 |  0.145163
 ar2                0.7467707 |  0.3127463 |  2.3878 |  0.016950 *
 ar3                0.7259450 |  0.1253425 |  5.7917 | 6.968e-09 ***
 ma1                1.5228601 |  0.5122810 |  2.9727 |  0.002952 **
 ma2                0.6781697 |  0.2395630 |  2.8309 |  0.004642 **
 intercept        -81.9633786 | 31.8478870 | -2.5736 |  0.010065 *
 Confidence.inst   -0.0386294 |  0.0313679 | -1.2315 |  0.218138
 Confidence.ind    -0.0048480 |  0.0257074 | -0.1886 |  0.850419
 Valuation.ind      0.0033978 |  0.0232667 |  0.1460 |  0.883892
 Valuation.inst     0.0206193 |  0.0267022 |  0.7722 |  0.439999
 CCI                0.4305970 |  0.2452919 |  1.7554 |  0.079183 .
 BCI                0.6735022 |  0.2255836 |  2.9856 |  0.002830 **
 UNRATE             0.0994235 |  0.3633560 |  0.2736 |  0.784372
 VIX               -0.0609112 |  0.0176317 | -3.4546 |  0.000551 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Herein, the snippet of code provided above creates a linear regression model with an ARIMA error structure ($model3$). Notably, the regressors selected are the same tested for both two of the previous regression models. Coefficients from $model3$ are also shown in the snippet of code provided below[32]. Similar to the former model, the z-test of coefficients shows that the only statistically significant parameters from $model3$ appear to be BCI and VIX, while the $UNRATE$ variable encompassed in $model2$ seems to be left out. This is not surprising nor constitutes evidence against the finding of the previous model, since this latter $model3$ introduced an ARIMA structure for the error term and thus may yield different results. Accordingly to what was noted above, $model3$ needs to be re-estimated in order to only retain the variables that showed

[32]Since the function *auto.arima* does not work properly with the standard *summary* command, the code exploits the $coeftest$ function from the *lmtest* package.

statistical significance, hence the VIX and BCI indexes. Running the same process presented above leads to estimating an improved ARIMAX, exhibiting the following parameters:

```
z test of coefficients:

             Estimate | Std. Error | z value |  Pr(>|z|)
--------------------------------------------------------
ar1          1.002778 |  0.076557  | 13.0984 | < 2.2e-16 ***
ar2         -0.012943 |  0.109619  | -0.1181 |  0.906011
ar3         -0.014727 |  0.074212  | -0.1984 |  0.842700
intercept -44.205537 | 22.370223  | -1.9761 |  0.048145 *
BCI          0.712808 |  0.222591  |  3.2023 |  0.001363 **
VIX         -0.061925 |  0.013005  | -4.7615 | 1.921e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As expected, all the linear regression parameters now appear to be statistically significant at the 0.05 level. However, it can be noted from the output of the coefficients provided above that only the first parameter of the autoregressive component (ϕ_1) of the ARIMA(3,0,0) process fitted seem to be significant. Conversely, ϕ_2 and ϕ_3 does not provide evidence to significantly differ from 0, as they exhibit high p-values. Recall that this output is obtained through the *auto.arima* function, which automatically searches for an appropriate ARIMA(p,d,q) fit. However, in this case, the model needs to be fine-tuned by explicitly specifying the order of the ARIMA process to be used to model the linear regression's residuals. In a similar fashion to the path followed previously, $model3$ needs to be re-estimated once again, in order to fit an ARIMA(1,0,0) process[33] to the error term of the regression. The code below provides the steps implemented to address the issues aforementioned, from the adjustments needed to the fitting of the desired ARIMA process to the re-estimation of the model.

[33] Meaning, an AR(1) process.

```
model2 <- Arima(trset$f.CAPE, order=c(1,0,0), xreg=reg)
coeftest(model2)
```

```
z test of coefficients:

            Estimate | Std. Error | z value |  Pr(>|z|)
-------------------------------------------------------
ar1         0.976352 |   0.014136 | 69.0698 | < 2.2e-16 ***
intercept -44.500908 |  21.129299 | -2.1061 | 0.0351936 *
BCI         0.717141 |   0.209982 |  3.4153 | 0.0006372 ***
VIX        -0.063350 |   0.011637 | -5.4439 | 5.212e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notably, the two parameters pertaining to the linear regression part of this model report similar coefficients to the ones estimated through $model2$. Specifically, BCI reports a positive correlation (close to 1 as for the previous model) and VIX reports a negative coefficient (once again, very close to the VIX coefficient for $model2$). The intuition provided by this model is consequently consistent with the findings previously expressed for $model2$.

Recall from the previous section that, for a statistical model to be validated, it is not enough to assess whether the relative coefficients carry any statistical significance. In a similar fashion, $model3$ needs to be analysed through the model diagnostic process, which may encompass both graphical and quantitative tools. Recall also that this statistical approach primarily aims to establish whether the residuals from the model under scrutiny may be affected by occurrences of non-stationarity, heteroskedasticity or autocorrelation. Either of these issues may signal an unreliable model and unable to correctly capture the data's nature, thus suggesting necessary adjustments. Nevertheless, ARIMAX diagnostic requires a slightly different process compared to the approach followed to validate $model2$. This is strictly correlated to the model's nature: since the residuals originating from the linear regression have been modelled through an ARIMA process, model diagnostic is consequently performed on the residuals stemming from the application of the ARIMA itself. In this context, the *tsdiag*[34] function helps to display the three most important graphical representations needed to make inferences about the model's residuals, which are the standardized residuals, ACF and the Ljung-Box

[34]The *tsdiag* function is a useful and common diagnostic tool which mainly works with ARIMA processes and cannot be applied to standard linear regressions. Hence, it was not used for previous model diagnostics.

test statistics plots. Results are pictured in Figure 15.

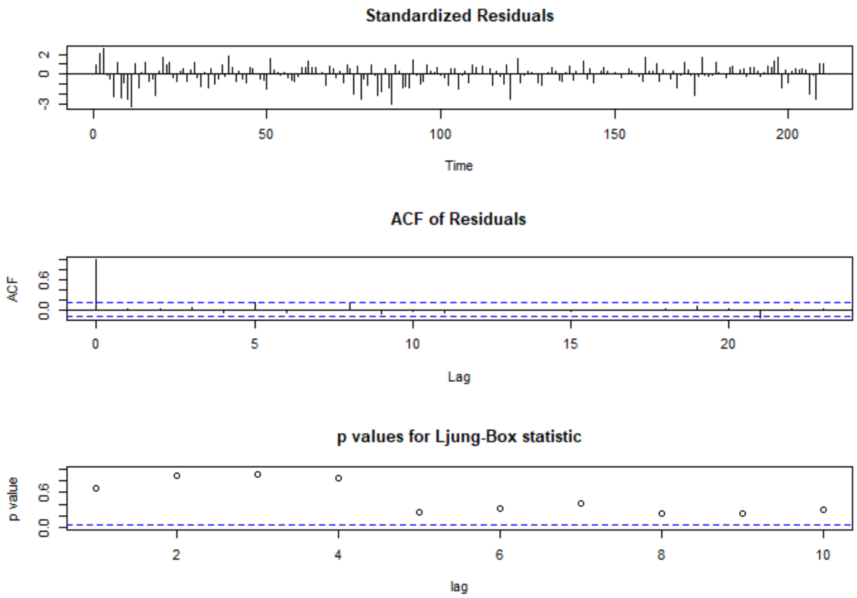

Figure 15: Diagnostic relative to *model*3 ARIMAX

These three plots contain significant pieces of information worth commenting on:

- For a model to exhibit a notable and acceptable fit, the standardized residuals graph (in the top row) should ideally resemble a white noise process, thus fluctuating randomly around zero with constant variance. In this case, the residuals appear to be stationary (they do not display any trends), suggesting that the model has captured the underlying process adequately. Additionally, variance also seems to be constant over time, with an exception made for a few initial observations. However, these occurrences are very limited in number and they are typical of financial data. Hence, they do not constitute any threats to the homoskedasticity assumption of *model*3;

- The ACF (in the middle row) exhibits values that are largely within acceptable bounds (represented by the blue dotted lines). The fact that autocorrelation (always equal to 1 at lag 0) drops to near-zero values from lag 1 onwards presents strong evidence against the presence of autocorrelation in the error term, implying that the ARIMA component of *model*3 correctly modelled the residuals of the linear regression;

- Finally, the bottom plot displays the p-values for the Ljung-Box test at various lags. This test checks for autocorrelation in the residuals at different lags, collectively. If the p-values are above the typical significance level of 0.05 (represented once again by the blue dotted line), the null hypothesis of no autocorrelation at those lags cannot be rejected. It follows vividly from the relative plot that all p-values are above the statistical significance level mentioned, further supporting the absence of autocorrelation in $model3$ residuals;

- Some characteristics from the residuals analysed (especially the trajectory pictured in the top row plot) may hint at the threat of the presence of a spurious regression. This phenomenon may imply that two seemingly related time series exhibit a notable correlation owing to chance or external factors, rather than genuine causation. The next section will address this potential issue by re-modelling non-stationary explanatory variables, as well as the independent variable, through first-differentiation.

Even though the *tsdiag* function already provided satisfactory results, it is possible to expand the quantitative side of the Ljung-Box test presented in the bottom plot of Figure 15. Specifically, this further step may be of interest in order to check the impact of the declining p-values displayed in the aforementioned plot starting at lag 5. Herein, the *Box.test* function implements this test (which in the R snippet is referred to as the Box-Ljung test); results are reported in the code below, showing a p-value of 0.0706 (above the 0.05 level and coherent to the conclusion of no autocorrelation in $model3$ residuals).

```
Box.test(model3$residuals, fitdf = 4, lag=10, type="Ljung")

 Box-Ljung test:
 data:  model3$residuals
 X-squared = 11.636, df = 6, p-value = 0.0706
```

Finally, from the evidence presented in Figure 15 and through the Box-Ljung test, as well as from the considerations laid out in the points above, it follows that $model3$ is correctly validated and can be exploited to make predictions about future CAPE values. Herein, $model3$ is defined as a linear regression model with parameters BCI and VIX, paired with an ARIMA(1,0,0) error term. The linear regression backing this model is then as follows:

$$CAPE_t = \beta_0 + \beta_1 \times BCI_{t-1} + \beta_2 \times VIX_{t-1} + \phi_1 \epsilon_{t-1} + \eta_t$$

At this point, it is possible to exploit this model in order to assess its accuracy in the forecasting process. Recall that this is done by applying $model3$ to the set of explanatory variables stored in the test set, and then comparing the output with the realised CAPE values (left out from the training set as well). To this extent, it is worth mentioning that the *predict* function used in the forecasting process with $model2$ is not designed to support ARIMA processes, hence it cannot be exploited to generate predictions with $model3$. In this situation, it is instead convenient to exploit the $forecast$ function, from the eponymous $forecast$ package. This process is implemented through the code provided below and estimates one-step ahead CAPE values through $model3$ ARIMAX. In a similar fashion to the initial model fitting, the $forecast$ function draws on a matrix containing the regressors' observations excluded from the training set.

```
newdata <- teset %>% select(BCI, VIX) %>% as.matrix()
pred3 <- forecast::forecast(model3, xreg = newdata)
```

A useful characteristic of the $forecast$ function applied to ARIMAX models is that it additionally provides upper and lower bounds for different confidence intervals (under residuals normality assumption, which will be tested below). Herein, Figure 16 pictures the output of the forecasting process, showing predictions' mean alongside 95% and 80% confidence interval ranges. The overlying black line represents instead the actual realised CAPE values, drawn from the test set.

Overall, $model3$ appears to exhibit notable predictive power against forward CAPE, up to the unravelling of the COVID-19 pandemic shortly after observation 220[35]. Specifically, the model still produces satisfactory results in light of the sudden CAPE drop at the beginning of the year 2020; on the other hand, it does not seem able to keep up with the large upside movement shown starting from observation 230 (August 2020). Nevertheless, this situation does not indicate a deteriorating fit of the model but rather underlines the huge rebound the S&P 500 had following the first pandemic period. Specifically, this period saw a large increase in the Index's price, not backed by a pro-

[35]The number of observations expressed in Figure 16 refers to the whole set of observations. Herein, predictions (and the test set) start with the 212^{th} observation from the *regressiondata* dataframe, corresponding to January 2019, and end in June 2023.

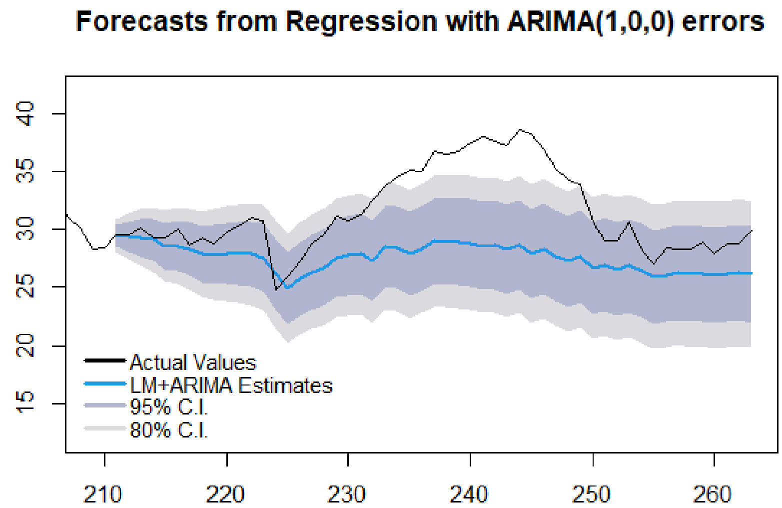

Figure 16: *model*3 ARIMAX one-step ahead forecasts

portional rebound in earnings levels. This context created a rare environment in which the price's steep upward run was not supported adequately by *model*3 parameters. Figure 17 provides evidence of this interpretation, noting that the S&P 500 almost doubled its spot P/E in a period of time shorter than one year.[36] In the subsequent year, the CAPE ratio for the S&P 500 index started to deflate back to the range of forecasts originated by *model*3. Recall that also *model*2 produced looser predictions relative to this period (Figure 13). Unfortunately, financial data encompasses extreme events that do not always find explanations in statistical terms; the COVID-19 pandemic represented one such period in which financial markets exhibited out-of-frame behaviours.

The impact of some inconsistent and extreme data also reflects in the distribution of *model*3 residuals, as pictured by the Q-Q plot[37] provided in Figure 18. The assumption of normality used to compute the model seems to hold for the majority of the dataset, with the exception of the lower quantile. This circumstance may signal the presence of

[36]This study refers to spot P/E as the P/E ratio computed with each month's price and the earnings of the trailing twelve months, both in real terms.

[37]The quantile-quantile (Q-Q) plot serves as a graphical tool to evaluate the goodness of fit of a dataset to a specific distribution. Herein, the calculated quantiles from the data under scrutiny are plotted against the theoretical quantiles of the distribution which is assumed to be followed by the dataset. Alignment along a straight line in the plot suggests conformity with the assumed distribution, while deviations indicate potential disparities.

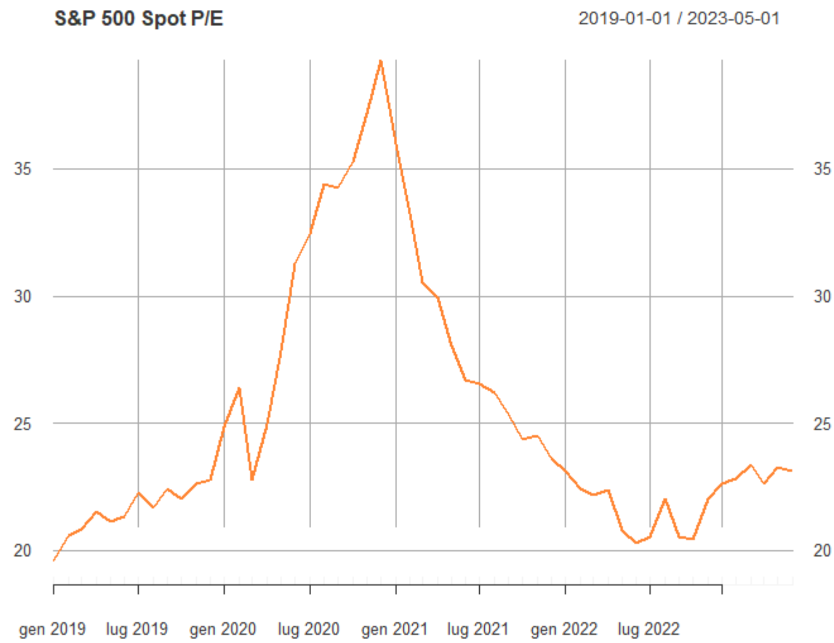

Figure 17: S&P 500 spot P/E levels provide a tangible explanation about the CAPE sudden increase not followed adequately by $model3$

slightly left-skewed residuals. In this context, it is worth commenting on the distribution of the residuals shown by $model3$. As mentioned, they appear to be slightly left-skewed, meaning that the first quantile seems to depart from the assumption of normality in the model's residuals implied in linear regression. Nevertheless, it is important to note that this circumstance does not invalidate the model and, more specifically, its predictive power. The assumption of normality in linear regression is fundamental to the theoretical framework of inferential statistics, encompassing hypothesis testing and the establishment of confidence intervals. However, the Central Limit Theorem provides assurance that, with a sufficiently large sample size, the sampling distribution of the mean will approximate a normal distribution, irrespective of the population distribution's shape. Consequently, despite observed skewness in the residuals, the model retains robustness for prediction purposes. Moreover, it is acknowledged that financial data frequently deviates from idealized assumptions. In this instance, the marginal skewness observed in the residuals was determined to have a negligible impact on the model's predictive validity (as proved in Chapter 4). Consequently, the model was deemed suitable for generating dependable predictions.

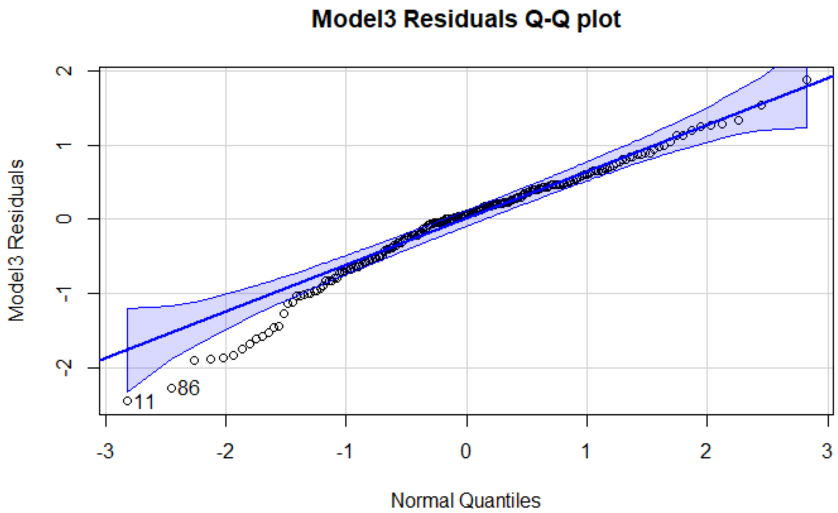

Figure 18: Quantile-Quantile plot relative to *model*3 residuals

The snippet of code provided below displays the output of the *accuracy* function used also for *model*2, in order to be able to compare the level of accuracy of the two models. However, it is worth mentioning again that the scope of this study does not rely on forward CAPE values accurate estimates, but it is rather centred on capturing this metric's upward and downward trends. As Chapter 4 shows, *model*2 and *model*3 exhibit significant results in this context.

```
accuracy(pred3, teset$f.CAPE)
                       ME      RMSE       MAE        MPE      MAPE
Training set -0.02676369 0.7213908 0.5450894 -0.2143404  2.286704
Test set      3.92146015 4.9545338 3.9707267 11.6118638 11.810382
```

To conclude this chapter, *model*3 is used to produce one-month ahead predictions of CAPE value for the whole dataset. Similarly to *model*2 estimates, this forecast set will be used to test different portfolio strategies in Chapter 4. Figure 19 pictures the results of this fitting process. It is apparent that *model*3 forecasts exhibit a less accurate solution with respect to predictions provided by *model*2 (previously portrayed in Figure 14). However, recall once again that the scope of this study is not to be able to forecast precise CAPE values, but rather to test these statistical models' ability to predict upward/downward CAPE movements and potentially exploit them in order to achieve superior portfolio returns. Chapter 4 will test the capabilities of the two models

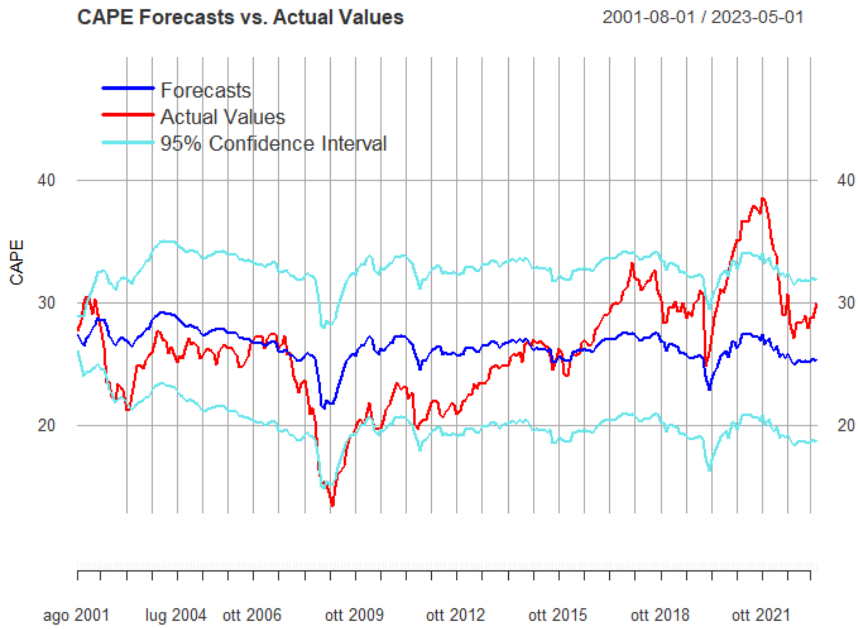

Figure 19: *model*4 one-month ahead estimates on the entire dataset

provided to this extent.

3.5 Dealing with Spurious Linear Regressions

In the previous section, it was mentioned that *model*3 may pose threats of being a spurious regression. In time series analysis, this concept relates to the fact that a set of variables may exhibit a notable correlation with the time series object of analysis, but this explanatory power may stem from a casual relationship or external factors (i.e. not included in the model) but correlated with both the dependent and the independent variable(s). This issue also draws from the fact that financial time series are often nonstationary or they tend to exhibit long-term autocorrelation. Thus, regressing a nonstationary series upon another may lead to substantial autocorrelation issues in the model's residuals (Ruppert and Matteson (2015)).

In order to address this situation and prove the validity of the findings explored by *model*3 (ARIMAX), this section implements a variation of the linear regression model with ARIMA errors previously presented. Specifically, the regressors suspected of not being stationary (all of the variables, with the exception of VIX and $UNRATE$) are first differentiated. Similarly, the dependent variable (formerly represented by the one-step ahead CAPE) has to be adjusted accordingly as well. To this extent, as a shared best practice used when dealing with financial time series, forward CAPE values are transposed to a logarithmic series and then first differentiated[38]. The approach of converting a (nominal) time series using a logarithm function is ”probably the most widely used transformation in data analysis”, as mentioned in Ruppert and Matteson (2015). In this context, the use of logarithmic returns (instead of first difference) for the independent variable provides convenient advantages for statistical inference purposes, among which:

- The variance of the dataset is stabilised, addressing the issue of heteroskedasticity often observed in raw financial time series and posing threats to the homoskedasticity assumption in linear regression modelling (with both OLS or MLE estimators);
- The impact of extreme values and outliers (which may represent a notable issue in the dataset under scrutiny, since it encompasses data from the COVID-19 pandemic, as previously mentioned) is mitigated, limiting the potential distortion of the results of regression analysis and its robustness.

Following these specific transformations, $model4$ can be constructed in a similar fashion to $model3$, thus by implementing a linear regression model with an ARIMA error structure. The snippet of code provided below displays the model's coefficients computed from this process; notably, the ”$d.$” preceding the name of a variable means that it has been first differentiated. Consistently, $dl.CAPE$ implies that the variable represents the first difference of the logarithmic series of the CAPE values. In so doing, it may be also considered to represent the logarithmic CAPE change.

[38]Note that, the process of first differentiating a logarithmic times series implies computing the logarithmic returns pertinent to that series. This stems from the fact that $\ln(p_t) - \ln(p_{t-1}) = \ln(p_t/p_{t-1})$, which is the definition of logarithmic returns (also outlined in Chapter 2).

```
z test of coefficients:

                     Estimate | Std. Error | z value |  Pr(>|z|)
----------------------------------------------------------------
ar1                -0.13112527 | 0.06935428 | -1.8907 |   0.05867 .
ar2                -0.18354658 | 0.07507994 | -2.4447 |   0.01450 *
ar3                -0.04381990 | 0.07344324 | -0.5966 |   0.55074
ar4                 0.03997135 | 0.07423110 |  0.5385 |   0.59025
ar5                 0.15049641 | 0.07162039 |  2.1013 |   0.03561 *
d.Confidence.inst -0.00165206 | 0.00083833 | -1.9706 |   0.04876 *
d.Confidence.ind   0.00182947 | 0.00098807 |  1.8516 |   0.06409 .
d.Valuation.ind    0.00073166 | 0.00085091 |  0.8599 |   0.38987
d.Valuation.inst   0.00118756 | 0.00086563 |  1.3719 |   0.17009
d.CCI               0.03994402 | 0.00741829 |  5.3845 | 7.263e-08 ***
d.BCI               0.04242642 | 0.00674994 |  6.2854 | 3.269e-10 ***
UNRATE              0.00470948 | 0.00077754 |  6.0569 | 1.388e-09 ***
VIX                -0.00156809 | 0.00024874 | -6.3041 | 2.898e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As for all of the other models explored in this study, it is important to notice that only a few regressors appear to be statistically significant (at the 0.05 level), namely: *d.Confidence.inst*, *d.CCI*, *d.BCI*, $UNRATE$ and VIX. However, re-estimating the model only encompassing significant variables returns a high p-value for *d.Confidence.inst*, which then has to be discarded[39]. Hence, the final parameters of *model*4 are *d.CCI*, *d.BCI*, $UNRATE$ and VIX (as shown in the snippet of code below). Notably, the three latter variables are the same which previously resulted as effectively significant in the computation of *model*3. On the other hand, the first-differentiated series from the Consumer Confidence Index seems to play a significant role in this regression, as well.

[39] This step is not reported for redundancy and clarity's sake.

```
z test of coefficients:

          Estimate | Std. Error | z value |  Pr(>|z|)
-----------------------------------------------------
ar1    -0.44327990 | 0.08551742 | -5.1835 | 2.178e-07 ***
ar2    -0.94557912 | 0.05891800 |-16.0491 | < 2.2e-16 ***
ar3    -0.20362079 | 0.07500647 | -2.7147 |  0.006633 **
ma1     0.36453934 | 0.05794597 |  6.2910 | 3.154e-10 ***
ma2     0.87824820 | 0.07805547 | 11.2516 | < 2.2e-16 ***
d.CCI   0.04227228 | 0.00812679 |  5.2016 | 1.976e-07 ***
d.BCI   0.04019823 | 0.00728808 |  5.5156 | 3.476e-08 ***
VIX    -0.00156046 | 0.00024186 | -6.4519 | 1.105e-10 ***
UNRATE  0.00465035 | 0.00075666 |  6.1459 | 7.949e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the output presented above, it emerges that all the coefficients tied to the ARIMA process are significant this time. Since *model*4 does not need any additional adjustment efforts up to this point, it can be defined as a linear regression model with an ARIMA(3,0,2) error term, as follows:

$$\begin{aligned}\text{dl.CAPE}_t = {} & \beta_1 \times \text{d.CCI}_{t-1} + \beta_2 \times \text{d.BCI}_{t-1} + \beta_3 \times \text{VIX}_{t-1} \\ & + \beta_1 \times \text{UNRATE}_{t-1} + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \phi_3 \epsilon_{t-3} \\ & - \theta_1 a_{t-1} - \theta_2 a_{t-2} + \eta_t\end{aligned}$$

Overall, this result provides evidence in favour of the reliability of the former model, since they share most of the significant variables and an autoregressive component of the same order. The usual model diagnostic is performed on *model*4 and pictured in Figure 20.

Results displayed in the diagnostic plots are worth commenting on, emphasising the common points of interest analysed in the preceding sections as well:

- Standardized residuals appear to be stationary with zero mean, with a few instances of large values (e.g. in the initial part of the observations). The same reasoning made for former models applies, meaning that financial time series may exhibit such extreme values in a few occurrences;
- ACF provides strong evidence of no autocorrelation in *model*4 residuals, as for

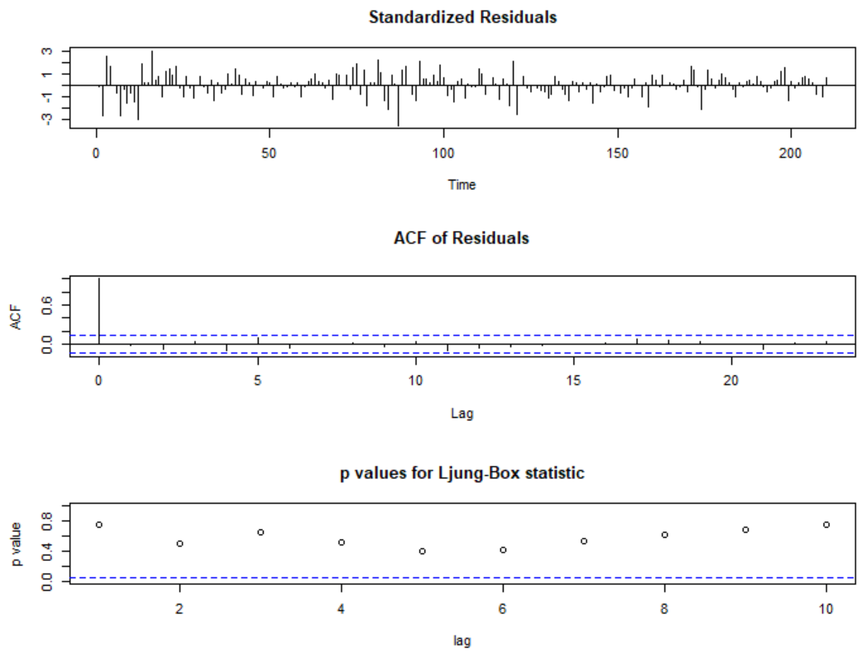

Figure 20: Diagnostic relative to *model*4 ARIMAX

lags > 0 autocorrelation values do not cross the statistically significant threshold (represented by the dotted blu line);

- Similarly, the Ljung-Box test's plot displays p-values largely above the statistical significance level for all of the lag considered, supporting the hypothesis of uncorrelated residuals.

The usual quantitative implementation of the Ljung-Box test (as follows) confirms the graphical intuition stemming from the plot mentioned above, returning a p-value for the test statistic of 0.3422 (again, largely above the 0.05 significance level.

```
Box.test(model4$residuals, fitdf = 4, lag=10, type="Ljung")


 Box-Ljung test
 data:  model4$residuals
 X-squared = 6.7747, df = 6, p-value = 0.3422
```

Hence, *model*4 is considered correctly validated from a statistical point of view. Remarkably, considering the models tested until this point, it also appears to be the most sensible one, in strict terms of forecasts' accuracy. Notably, Figure 21 displays a

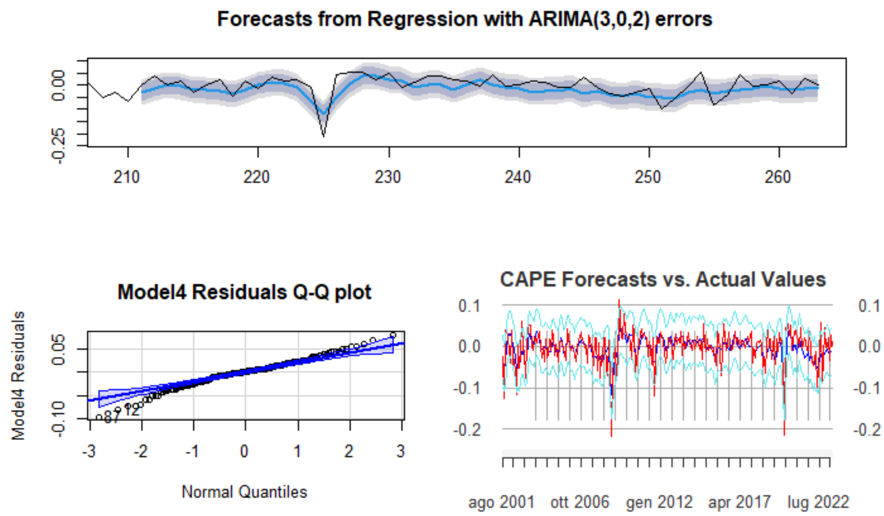

Figure 21: Comprehensive overview of *model*4 performance and analytics

comprehensive picture of the three points of focus on which previous analysis similarly lingered:

- The top row plot outlines the one-month ahead predictions for the $dl.CAPE$ variable (meaning, the forward CAPE logarithmic returns) on the test set. Notably, as for the estimates provided by *model*3, the shaded areas surrounding the main forecast line represent the 95% and 80% confidence intervals, respectively, assuming normality. It appears from this graph that *model*4 is able to accurately pick up the sharp decline corresponding to the beginning of the COVID-19 pandemic, as well as the subsequent rebound to initial logarithmic levels. This feature of *model*4 is not shared with the former models, which conversely displayed a higher discrepancy in this circumstance;

- The Q-Q plot (shown in the bottom-left corner) displays similar overall results with respect to *model*3. Specifically, once again the first quantile's dynamic appears to be hinting at a slightly left-skewed distribution, while the normality assumption seems to well capture the majority of residuals. The same comments made on *model*3 Q-Q plot apply;

- Finally, the plot pictured in the bottom-right corner shows the estimates, computed through *model*4, for the whole dataset, with the relative 95% and 80% confidence intervals (light blue lines).

4 CAPE-Based Portfolio Strategies

4.1 Theoretical Strategy Basis

Academic literature, as presented in Chapter 1, extensively argued that low-P/E ratio portfolios systematically seemed to deliver superior returns, compared to high-P/E investments, even after controlling for the level of risk carried by these different strategies. Evidence of the tendency of low-P/E stocks to outperform high-P/E ones emerges vividly from the conclusions drawn in Basu (1977), Chahine and Choudhry (2004), Weigand and Irons (2005), Huang et al. (2007), Aga and Kocaman (2006) and Kelly et al. (2008). In this context, P/E and CAPE were also broadly tested as potential explanatory variables for future economic outlooks, as represented by the S&P 500 and other foreign equity market indexes (Kane et al. (1996), Campbell and Shiller (1988) and 2001, Aras and Yilmaz (2008), Angelini et al. (2013), Allahyaribeik et al. (2020), Kenourgios et al. (2022)).

Focusing instead on a different framework, a notable part of the literature mentioned explored the influence that investors' behaviour (intended as sentiment, emotional reactions, hopes and fears) appears to exercise on market returns (Fisher and Statman (2000) and 2003, W. Y. Lee et al. (2002), Akintoye (2008), Bathia and Bredin (2012), Smales (2017), Campisi and Muzzioli (2020), Wang et al. (2021)). Notably, the considerations provided by the authors in this field of research are coherent with the results outlined in Chapter 3, where both $model2$ and $model3$ found that behavioural variables retain significant predictive capability against forward CAPE values.

Encompassing both of the points of interest presented above, this study proposes to further strengthen the framework around the CAPE ratio by incorporating into one theory the two concepts outlined by the academic literature. To this extent, the next sections exploit $model2$, $model3$ and $model4$, as constructed in Chapter 3 around investors' behaviour-centred variables, to implement three different investment and portfolio management strategies, specifically:

- **Long CAPE**: this strategy implies entering into a long position in one unit of the S&P 500 Index for the upcoming month, where the model used ($model2$/ $model3$/ $model4$) forecasts an upward CAPE movement for that period. If, conversely, predictions indicate a decline in the CAPE ratio value for the following month, the investor will maintain a neutral position, thus exiting the market if he was invested in the previous period or not opening any new positions in case he was not already invested. This methodology aims to exploit the positive financial leverage effect provided by CAPE uptrends, whose effects were explored in Chapter 2;

- **Long/Short**: this strategy builds on the Long CAPE approach mentioned above, replicating the same process in the case of positive forward CAPE expectations. However, the fundamental variation provided by this approach is that, where the forecasting model predicts a downward shift of the CAPE ratio for the following month, the investor will enter into a short position, thus short-selling one unit of the S&P 500 Index (instead of remaining "outside" of the market, as in the Long CAPE strategy). This strategy tests the extent to which negative CAPE expectations may be exploited to bet against the market and achieve even higher returns;

- **Leveraged Long**: as the name suggests, this approach aims to replicate the first strategy, but exploits a leveraged (long) position when the forecasting model predicts a larger increase in the CAPE ratio. Herein, where predictions expect this metric to rise at a monthly rate of $> 2\%$, the investor will leverage its position[40] for 25% of its value. It is worth noting that while a leveraged position may increase returns, it also increases the risk of significant losses of the same magnitude. For this reason, if the forecasting models chosen would not prove accurate in spotting large potential movements, this strategy could in turn provide worse results than the Long CAPE one.

Overall, implementing these portfolio management strategies on *model*2, *model*3 and *model*4 predictions has the double scope of (i) testing whether these models are capable of efficiently capturing CAPE trends and (ii) assessing the extent to which investment methodologies based on the CAPE ratio's expected movements may be used in practice to achieve superior portfolio returns. This latter point of interest is remarkable for one particular reason. To the best of the author's knowledge, this is the

[40] There are several ways to implement a leveraged position in the stock market. The most common approaches are represented by the use of financial derivatives, borrowing or leveraged exchange-traded funds (ETFs). Derivatives like options and futures allow investors to control large amounts of stocks with relatively small capital outflows. Borrowing to invest, commonly referred to as "buying on margin", involves using funds borrowed from a broker to purchase more securities than what the investor could do with his own funds. Leveraged ETFs are financial vehicles built to deliver multiples (e.g. $2x$, $3x$) of the daily performance of the index they track. This study does not discriminate on the means to achieve a leveraged position, since it does not impact the final result of the strategy.

first academic study testing the viability of exploiting movements in CAPE valuations to enhance a portfolio's profitability. Several works (mentioned in the literature presented in Chapter 1 and in this section) based different investing strategies on P/E and CAPE absolute values. Herein, the approach which this study referred to as value investing is based on findings showing that low-CAPE stock portfolios appeared to systematically outperform high-CAPE ones, which instead seemed to deliver below-average returns. In this context, this research argues that the most important underlying factor is not the nominal value of the CAPE ratio, but rather the direction of this metric's trajectory in the subsequent investment period. This chapter explores the argument according to which overall low CAPE values were not the actual strength of value investing strategies per se, but they only indicated an increased probability that CAPE valuations would reverse to their historical (higher) mean, thus following an upward movement for the subsequent periods. The same argument may be similarly used to disregard the other side of the theory, under which high-CAPE values are systematically tied with below-par returns. Indeed, this study argues that high CAPE values may only warn about an enhanced probability of a decline in this metric, but do not constitute a negative stock factor per se.

To compare the performance of the aforementioned strategies, the relative investment benchmark selected is a "buy and hold" strategy in the S&P 500, consisting of entering into one long position at the start ($t = 0$) and never adjusting the portfolio exposure again. The following two sections will implement the strategies discussed above using $model2$ (linear regression model with Newey-West estimator for robust standard error), $model3$ (linear regression model with ARIMA error terms) and $model4$ (first-differentiated ARIMAX based on logarithmic CAPE series), respectively. All of the portfolio simulations presented are based on the one-month ahead predictions made on the whole dataset of data available, as mentioned in Chapter 3. Finally, it is important to acknowledge that the three portfolio strategies investigated in this study do not incorporate the adverse impact of trading costs, short-selling fees and taxes on investment returns. This omission is attributed to the significant variability of these factors across investors' capital, jurisdiction of residence, brokerage specifications and other pertinent variables. Nevertheless, to offer a comprehensive understanding of the potential influence exerted by these components on the efficiency of the strategies proposed, a simulation assessing their effects will be presented in the concluding section.

4.2 Investing Strategies with Behavioural CAPE-based Model

The three investment strategies, outlined in the previous section, are based on the discriminant of whether one-month ahead predictions of the CAPE ratio are increasing (upward) or decreasing (downward). Herein, incorporating a dummy variable (DV) in the methodology tested represents a useful statistical approach, enabling the control of categorical factors and the assessment of their impact on the dependent variable. A dummy variable (or indicator variable) is generally a binary factor whose only scope is to discriminate between the presence or the absence of some categorical effect that may be expected to shift the outcome of a statistical model. In this context, the categorical factor to which the dummy variable should refer to is the presence of positive CAPE forecasts (for which DV will take a value of 1) or not ($DV = 0$). Herein, the term DV is computed as follows:

$$DV_t = \begin{cases} 1 & \text{if } C\hat{A}PE_t - C\hat{A}PE_{t-1} \geq 0 \\ 0 & \text{if } C\hat{A}PE_t - C\hat{A}PE_{t-1} < 0 \end{cases}$$

where $C\hat{A}PE_t$ and $C\hat{A}PE_{t-1}$ represent the one-month ahead predictions of CAPE values made at time t and $t-1$, respectively.

The snippet of code provided below presents the step necessary to implement the DV into the dataframe containing the final estimates used for the portfolio simulation (**sim**)[41].

```
sim$DV <- c(NA, ifelse(diff(sim$pred.CAPE)<0,0,1))
```

After implementing the dummy variable through the process described above, it is possible to compute cumulative logarithmic returns to compare the performances of the two investments. Recall that Chapter 2 outlined the mathematical process behind the computation of both simple and logarithmic cumulative returns. Also recall that, in that circumstance, the latter methodology resulted in a more appropriate solution. Nevertheless, the strategy tested here is built on the same underlying portfolio (constituted by the S&P 500 Index) of the benchmark referred to. For this reason, there is no need to use logarithmic returns, as these two investments are directly comparable.

[41] It is worth mentioning that the first value of the DV column is put as an NA. The reason for this is the fact that DV, by definition, is computed through estimate differences. Thus, at $t = 0$, only one estimate is available and it is not possible to assess whether it implies an upward or downward movement.

In order to include the dummy variable, simple cumulative returns for the Long CAPE strategy are computed as follows. Let ptf_t be the value of the portfolio at time t. Similarly to Chapter 2, the notional amount invested at time $t = 0$ is chosen to be 1, while for $1 \leq t \leq k-1$ (where k is the number of observations in the whole dataset, including both training and test sets[42]) cumulative returns are computed recursively:

$$ptf_t = \begin{cases} 1 & \text{if } t = 0 \\ ptf_{t-1} \times (1 + \text{S\&P 500 return}_t \times DV_t) & \text{if } 1 \leq t \leq k-1 \end{cases}$$

Cumulative returns are computed similarly for the S&P 500 buy-and-hold (B&H) benchmark, leaving out the DV term since it only refers to the strategy tested. Figure 22 pictures the overall performances of the Long CAPE strategy on the S&P 500 versus a buy-and-hold approach on the same index.

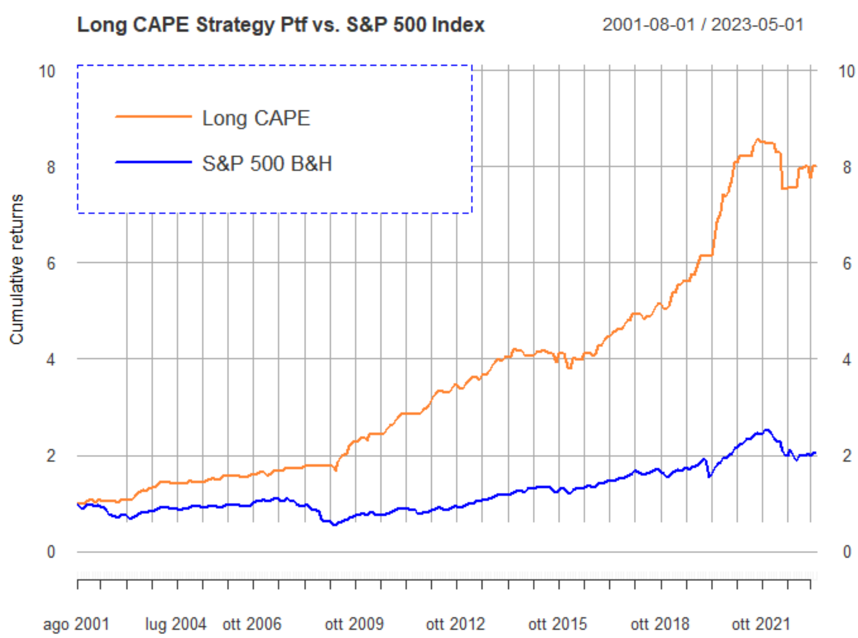

Figure 22: Long CAPE strategy vs. S&P 500 buy-and-hold

It appears vividly from the plot that the Long CAPE investing approach yielded remarkable results. In the 12 years ending as of the last observation's date of the

[42]As mentioned throughout Chapter 3, the models explored have been constructed on the training set (comprising n observations) and tested on the test set (with m observations). Finally, portfolio simulations are implemented exploiting the entire range of data available, meaning $k = n + m$

dataset (July 2001 - June 2023), the notional amount invested in the S&P 500 with a buy-and-hold strategy (represented by the blue line) grew to a portfolio value of 2.1, implying a total cumulative growth[43] of $\approx$ 110% (or an average yearly return[44] of $\approx$ 6.4%). On the other hand, the Long CAPE strategy yielded a total cumulative growth of $\approx$ 701% (or $\approx$ 18.93% in yearly terms), with the simulated portfolio growing up to a value of 8.01. Hence, the Long CAPE portfolio provides significant evidence of both (i) *model*2 capabilities in forecasting CAPE upward shifts and (ii) the viability of exploiting a portfolio management strategy based on this ratio to achieve superior returns. Notably from Figure 22, the enhanced performance of the tested strategy particularly reflects on *model*2 accuracy in predicting two harsh drawdowns which hit the market during the subprime mortgage crisis and the COVID-19 pandemic, avoiding significant losses.

The second strategy discussed was referred to as Long/Short, underlining its focus on exploring both sides of the CAPE ratio leverage effect. Similarly to the previous approach, a DV is introduced in the model. However, its definition is adjusted accordingly to the scope of this strategy, as follows:

$$DV_t = \begin{cases} 1 & \text{if } C\hat{A}PE_t - C\hat{A}PE_{t-1} \geq 0 \\ -1 & \text{if } C\hat{A}PE_t - C\hat{A}PE_{t-1} < 0 \end{cases}$$

The equations outlined above provide that in case of a lower expected CAPE for the subsequent month, the investor will enter into a short position (-1) on the underlying portfolio, thus earning a capital gain if the S&P 500 declines, and vice versa. In so doing, the DV successfully allows the investor to capture the potential of both expected CAPE uptrend and downtrend movements. Figure 23 pictures the results of this strategy.

Remarkably, the Long/Short approach appears to largely enhance the investing application of *model*2, providing an astonishing cumulative return of $\approx$ 2390% in the near 12 years of the period under scrutiny (translating into an average annual return of $\approx$ 30.7%). Notably, declines in the strategy trajectory (represented by the purple line) underline periods in which *model*2 was not able to correctly forecast the direction of the CAPE ratio for the subsequent month, in either direction. Conversely, an increasing

[43] From the simple cumulative return formula expressed in Chapter 2, it follows that: $\mathrm{CR} = 1 + \prod_{i=1}^{n}(1 + \mathrm{Return}_i) = \mathrm{ptf}_i - 1$.

[44] Cumulative returns can be transposed in yearly terms by computing their geometrical average, as follows: Average $\mathrm{AR} = (1 + \mathrm{CR})^{\frac{1}{\text{n of years}}} - 1$.

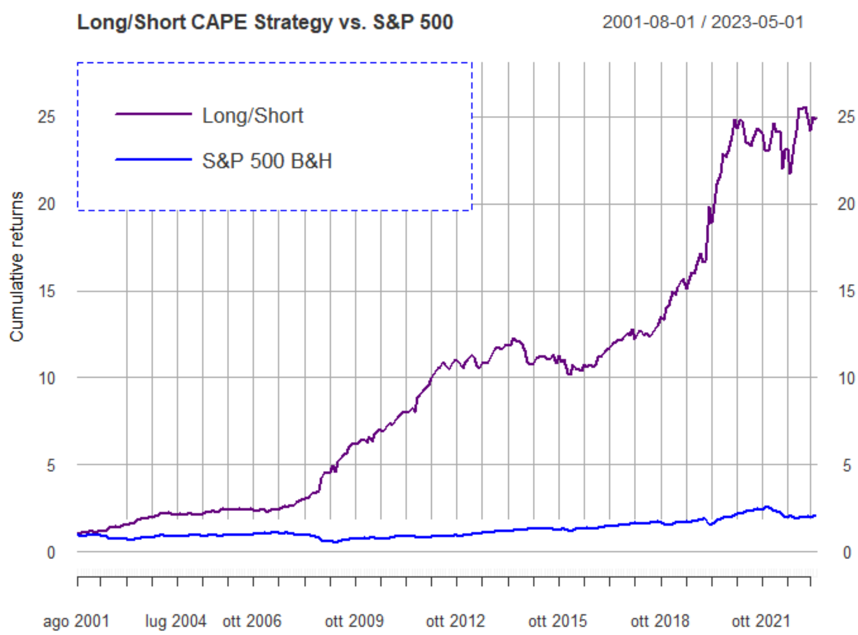

Figure 23: Long/Short strategy vs. S&P 500 buy-and-hold

trajectory implies the model successfully predicting upward/downward CAPE shifts.

Finally, the last strategy implemented using forecasts from $model2$ is the Leveraged Long. This approach aims to test whether above-average CAPE shift expectations may be exploited to produce additional portfolio returns, with respect to the Long CAPE strategy. Herein, this methodology also explores whether $model2$ is able to capture the magnitude of CAPE trends. The dummy variable needs to be adjusted accordingly again, in order to specify the periods in which to leverage the position. To this extent, define a variable $c.CAPE$ computing the percentage change in expected CAPE values, such that:

$$c.CAPE_t = \frac{C\hat{A}PE_t - C\hat{A}PE_{t-1}}{C\hat{A}PE_{t-1}}$$

Hence, DV will be defined as follows:

$$DV_t = \begin{cases} 0 & \text{if } c.CAPE_t < 0 \\ 1 & \text{if } 0 \leq c.CAPE_t < 0.025 \\ 1.25 & \text{if } c.CAPE_t \geq 0.025 \end{cases}$$

Stemming from the equations outlining the dummy variable, in months where $model2$

forecasts an increase in the CAPE ratio of 25%, the return earned by the portfolio will be leveraged by a factor of 1.25. If, on the other hand, this circumstance is not verified, the portfolio will similarly follow a Long CAPE approach. Figure 24 pictures the result of this strategy.

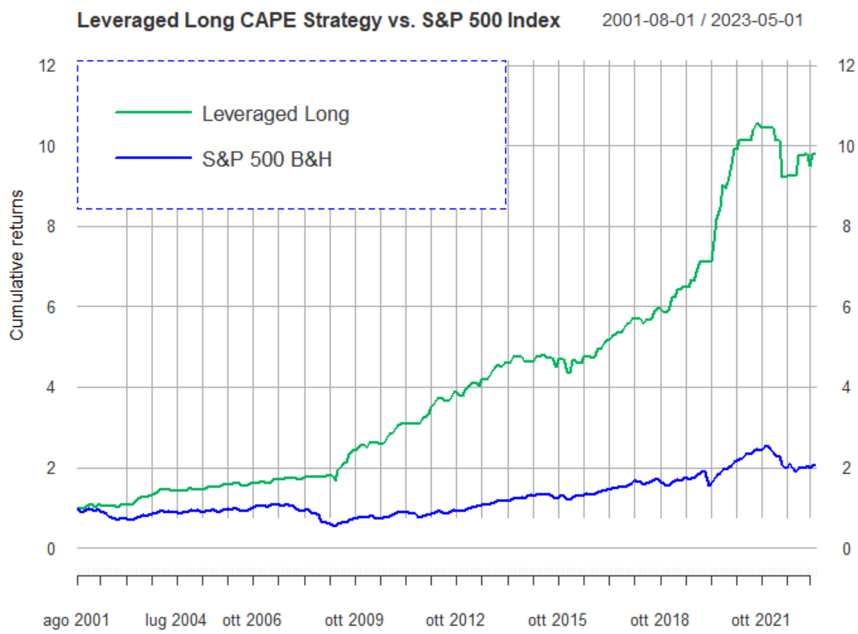

Figure 24: Leveraged Long strategy vs. S&P 500 buy-and-hold

The Leveraged Long strategy appears to perform better than Long CAPE, yielding a total cumulative return of $\approx 880\%$ (corresponding to an average return of $\approx 20.9\%$ in annual terms). This fact provides evidence of *model*2 capabilities to capture significant jumps in forward CAPE valuations and the viability of exploiting these above-average movements to achieve higher portfolio returns.

To bring this section to a comprehensive conclusion, it is useful to provide a joint overview of all three strategies explored. Herein, Figure 25 picture a clear overview of the comparison between the different performances explored in this section.

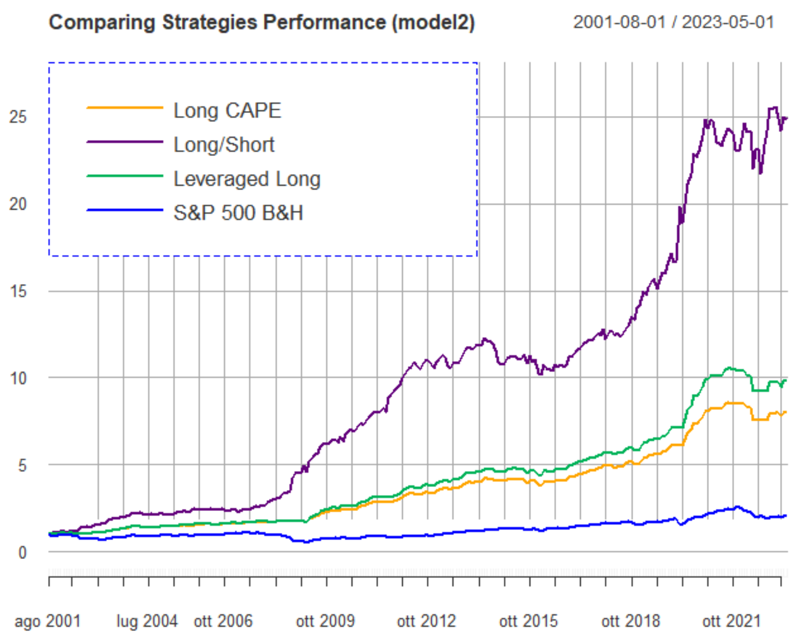

Figure 25: CAPE-based strategies (*model*2) vs. S&P 500 buy-and-hold

4.3 Investing Strategies with ARIMA-errors Model

The same three portfolio strategies (Long CAPE, Long/Short and Leveraged Long) described in the previous section may be implemented using predictions generated from *model*3. The equations used to make predictions and to compute the appropriate *DV* for each strategy are the same as those explored for *model*2, hence they are not repeated here. Additionally, since the previous section already detailed the process of implementing the three different approaches mentioned, this part will not provide a separate focus for each strategy. Instead, Figure 26 provides a comprehensive overview of the performances of the three strategies, this time based on *model*3 estimates.

Notably, there are some similarities, as well as a few significant differences (with respect to what was shown in Figure 25), which are worth commenting on:

- Overall, all three strategies proposed appear to largely and consistently outperform the S&P 500 Index, with similar trajectories compared to the same strategies applied through *model*2. This is a quite remarkable result considering that a buy-and-hold strategy in such a diversified index is often argued to be one of the most efficient investing strategies providing steady returns;

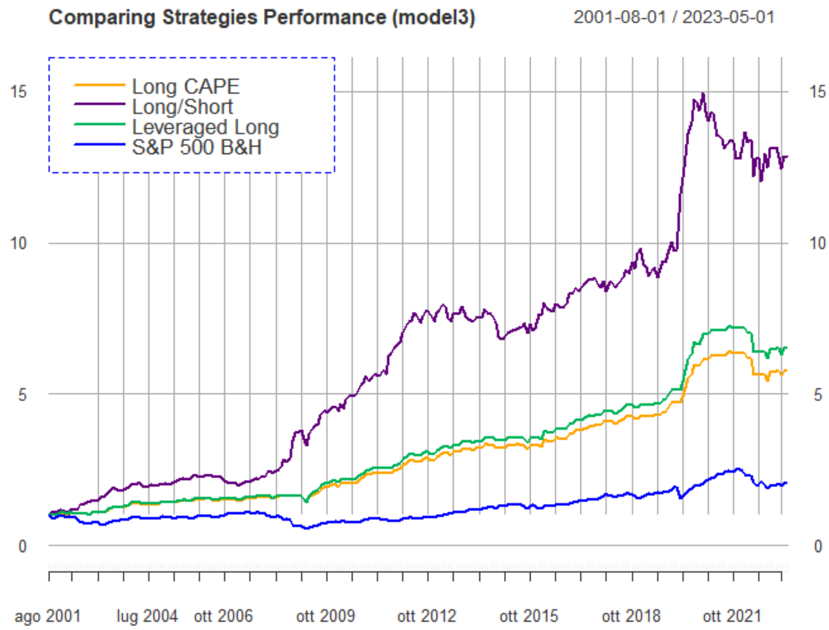

Figure 26: CAPE-based strategies (*model*3) vs. S&P 500 buy-and-hold

- The Long CAPE strategy yielded a $\approx$ 476% cumulative return in the 12 years of simulated performance, transposing to an average annual return of $\approx$ 15.7%. This represents a considerably worse performance compared to the same approach implemented with *model*2 (which provided an annual return of 18.9%), confirming the points of interest touched before;

- The Long/Short approach appears to be once again the most profitable, by a large margin. In this simulation, the portfolio following this strategy achieved a total cumulative return of $\approx$ 1181% (or $\approx$ 23.7% in annual terms). Similarly to Long CAPE, this result appears to be inferior to the performance achieved through the same strategy based on *model*2 (which yielded an average annual return of 30.7%);

- Finally, the Leveraged Long approach provides a slightly enhanced performance with respect to the Long CAPE strategy, also for this model. Figure 26 shows that Leveraged Long (represented by the green line) has yielded a higher return consistently for the whole investment period. However, also note that this strategy outperforms Long CAPE only by a thin margin, while for the majority of the time, the two methodologies were almost interchangeable. This intuition suggests that large CAPE shifts do not occur often and cannot be relied on to achieve a significantly superior portfolio management strategy Hence, Leveraged Long achieved a cumulative return of $\approx$ 553% (or $\approx$ 16.9% in annual terms) which, while still being slightly superior to the Long CAPE strategy, is considerably lower than the same approach backed by $model2$ ($\approx$ 20.9% annual return).

4.4 Investing Strategies with First-Differenced-Logarithmic ARIMAX Model

Bringing this chapter to a comprehensive conclusion, the three investment strategies previously outlined are tested for estimates stemming from $model4$ as well. Note that, since this model encompasses logarithmic CAPE variations as the independent variable (rather than the nominal CAPE itself), the construction of DV has to be adjusted to reflect this factor accordingly. Recall that the Long CAPE strategy discriminates whether the expected one-month ahead CAPE value is higher than the one pertaining to the current month. In $model4$, this circumstance is addressed by requiring that logarithmic variations should be > 0. Specifically:

$$DV_t = \begin{cases} 1 & \text{if } dl.C\hat{A}PE_t \geq 0 \\ 0 & \text{if } dl.C\hat{A}PE_t < 0 \end{cases}$$

where $dl.C\hat{A}PE_t$ represents one-month ahead predictions of logarithmic CAPE movements. The second strategy tested (Long/Short) requires a similar approach, but recall that DV should reflect a value of -1 in case of negative expectations, in order to capture the short-selling performance of the portfolio when forecasting a decline in CAPE. For this reason, the dummy variable relative to Long/Short is defined as follows:

$$DV_t = \begin{cases} 1 & \text{if } dl.C\hat{A}PE_t \geq 0 \\ -1 & \text{if } dl.C\hat{A}PE_t < 0 \end{cases}$$

Lastly, the Leveraged Long strategy requires one additional consideration, before computing the final DV. This is due to the fact that $model4$, as mentioned, produces forecasts on the logarithmic change in CAPE value, the factor $c.CAPE$ exploited for previous models needs to be adjusted accordingly. Since, in this case:

$$dl.C\hat{A}PE_t = \ln(CA\hat{P}E_t) - \ln(CA\hat{P}E_{t-1}) = \ln\left(\frac{CA\hat{P}E_t}{CA\hat{P}E_{t-1}}\right)$$

and $c.CAPE$ represent the percentage change in CAPE value, defined as

$$c.CAPE_t = \frac{CA\hat{P}E_t - CA\hat{P}E_{t-1}}{CA\hat{P}E_{t-1}} = \frac{CA\hat{P}E_t}{CA\hat{P}E_{t-1}} - 1$$

then it must follow that

$$c.CAPE_t = \exp(dl.C\hat{A}PE_t) - 1$$

From here, the DV for the Leveraged Long strategy applied through $model4$ estimates may be computed similarly to what was mentioned for previous models, specifically:

$$DV_t = \begin{cases} 0 & \text{if } c.CAPE_t < 0 \\ 1 & \text{if } 0 \leq c.CAPE_t < 0.025 \\ 1.25 & \text{if } c.CAPE_t \geq 0.025 \end{cases}$$

The reiterating portfolio value is then computed encompassing the different dummy variables explored for the three different strategies, as follows:

$$ptf_t = \begin{cases} 1 & \text{if } t = 0 \\ ptf_{t-1} \times (1 + \text{S\&P 500 return}_t \times DV_t) & \text{if } 1 \leq t \leq n-1 \end{cases}$$

Similarly to the previous section, Figure 27 displays an overall picture of the performance achieved by the three strategies, applied through estimates from $model4$.

As mentioned in the last section of Chapter 3, this model appeared to be the most precise in terms of accuracy of capturing CAPE shifts. Herein, Figure 27 provides clear evidence of this fact, noting that:

- The portfolio based on the Long CAPE strategy achieved a cumulative return of $\approx 970\%$ in the period under scrutiny, corresponding to an average annual return of $\approx 21\%$. Note that this result is considerably higher than the same approach implemented through the former models. Also, it appears from the plot that for

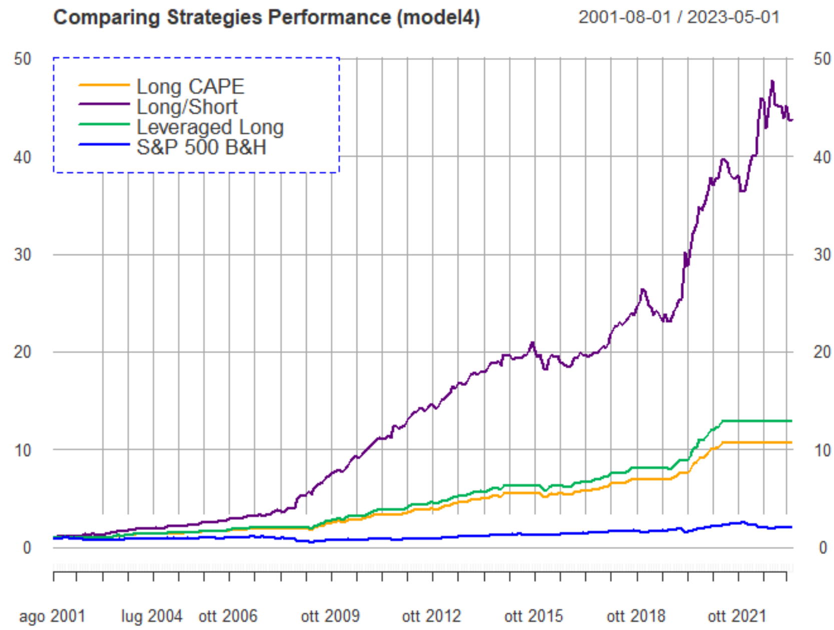

Figure 27: CAPE-based strategies (*model*4) vs. S&P 500 buy-and-hold

the period starting in April 2021, this strategy displays a flat trajectory. This is due to the fact that, for the corresponding period, *model*4 appears to predict a consistently downward CAPE trend;

- Once again, Long/Short retains its title of most profitable investment strategy, achieving an astonishing cumulative return of ≈ 4278% (or ≈ 36.96% in annual terms). This result represents the best performance by any strategy applied through any of the models analysed in this study, by a notably large margin;

- Finally, the Leveraged Long strategy confirms the intuition mentioned throughout previous sections regarding the low reliability of such a strategy to achieve a significantly better performance than the base Long CAPE approach. Specifically, such a portfolio returned a cumulative return of ≈ 1187%, transposing to an average annual return of ≈ 23.7%. Similarly to the two other strategies, this performance is still superior to the one achieved by exploiting *model*2 and *model*3.

To recapitulate the most important point of interest outlined in this chapter, it is possible to conclude that all the three strategies presented, implemented through the three models constructed in Chapter 3, achieved notable results in the context of providing different portfolio management methodologies able to outperform the benchmark of a buy-and-hold investment in the S&P 500 Index. Specifically, this study provided significant evidence in favour of a Long/Short strategy systematically displaying the highest profitability, outperforming the other two strategies by a large margin. This finding is especially remarkable considering that, as mentioned at the beginning of this Chapter, this study has a pioneeristic role in this field, as CAPE movements were not explicitly taken into consideration in any of the previous academic literature, to the best of the author's knowledge.

5 Final Thoughts and Conclusion

This study aimed to redefine the academic framework surrounding the field of financial forecasting through the Cyclically-Adjusted Price-to-Earnings (CAPE) ratio. Specifically, it was argued that employing this metric as an explanatory variable to predict future price levels constituted a rather inconvenient and outdated approach. The reason behind this intuition is underlined by the empirical analysis relative to the S&P 500 Index presented in Chapter 2, which provided evidence about how the Index's price level is ultimately largely driven by earnings growth (or decline) of the underlying companies. This factor's impact on the S&P 500 returns was labelled as the Index's organic component. Additionally, this study introduced the concept of speculative component as well, measured as the difference between total S&P 500 returns and earnings growth. Herein, it was concluded that speculative market dynamics largely affect the Index's price in the short- and mid-term, while extending the analysis on a longer horizon provides evidence that the speculative component tends to average out, implying that returns are ultimately fuelled by organic companies' growth.

Following the empirical intuitions aforementioned, this study argued that rather than using the CAPE (or P/E) ratio as a potential explanatory variable, it would have been of greatest interest attempting to shift its role as the centre of attention of a statistical analysis. Specifically, the scope behind this process stemmed from the finding that, where the S&P 500 price struggled to align with underlying companies' earnings, the Index generated higher/lower returns (as represented by the speculative component previously introduced).

Herein, this research tested and successfully assessed the impact of variables related to investors' behaviour and population's sentiment on the CAPE ratio, outlining the extent to which this metric appears to be notably correlated with optimism/pessimism dynamics implied by the U.S. economy. These variables showed significant predictive capabilities for one-month-ahead CAPE movements. Hence, Chapter 3 found that the main behavioural variables displaying notable correlation with forward CAPE values are mainly represented by the Business Confidence Index and the Implied Volatility Index (which were found to be statistically significant in all three models analysed). Additionally, the Unemployment Rate and the Consumer Confidence Index also displayed significant predictive capabilities. Specifically, it is possible to provide a summary on the process behind each model, as well as the different final parameters retained:

- *model*2 was built as a linear regression exploiting a HAC robust error structure. This model encompassed the Newey-West estimator in order to compute an adjusted variance-covariance matrix which could address the evidence of residuals autocorrelation and heteroskedasticity found through the model diagnostic process. The final form of the model was concluded to be the following:

	Intercept	VIX	BCI	UNRATE
Coefficient	-82.512	-0.086	1.176	-1.407
Std. Error	13.808	0.02	0.136	0.072

Table 1: *model*2 parameters

- On the other hand, *model*3 exploited the common approach to financial time series modelling of incorporating an ARIMA process in the regression's error structure. Herein, this model resulted in a linear regression model with an ARIMA(1,0,0) error term. Hence, *model*3 parameters were proven to be the following:

	Intercept	VIX	BCI	ar1
Coefficient	-44.501	-0.063	0.717	0.976
Std. Error	21.129	0.012	0.21	0.014

Table 2: *model*3 parameters

- The last model analysed (*model*4) exploited first differentiation for non-stationary variables, exploring the correlation with logarithmic CAPE shifts. Herein, it was found that this model, stemming from its higher complexity, was also the one which more accurately seemed to capture the CAPE trajectory. Specifically, *model*4 was built as follows:

	d.CCI	d.BCI	VIX	UNRATE	
Coeff.	0.042	0.04	-0.002	0.005	
Std. Error	0.008	0.007	0.0002	0.0008	

	ar1	ar2	ar3	ma1	ma2
Coeff.	-0.443	-0.946	-0.204	0.365	0.878
Std. Error	0.086	0.059	0.075	0.058	0.078

Table 3: *model*4 parameters

Finally. Chapter 4 tested three different investing strategies, based on different methodologies of managing a portfolio built around one unit of the S&P 500 Index. Each strategy (Long CAPE, Long/Short and Leveraged Long) was designed to test different extents to which was possible to exploit CAPE movements to achieve higher portfolio returns, with respect to a standard buy-and-hold strategy on one unit of the same index. This concept represents what was defined in Chapter 2 as the financial leverage effect provided by the CAPE ratio. Specifically, one-month ahead CAPE estimates derived from $model2$, $model3$ and $model4$ were used to evaluate whether CAPE was expected to move upward or downward in the following month and to adjust the portfolio according to the different strategies tested.

Remarkably, this study found that all three strategies proposed, applied through all three models, systematically provided significantly superior returns, compared to the benchmark aforementioned. Notably, the Long/Short approach yielded the best results for each model specification, providing an average annual return of approximately 30.7%, 23.7% and 37% (for $model2$, $model3$ and $model4$, respectively), against the notably inferior performance of the S&P 500 buy-and-hold portfolio, whose average annual return assessed on a much lower 6.4%). This enhanced performance stems from (i) the ability of the constructed models to forecast upward/downward shifts in the CAPE ratio for subsequent months (even though, as mentioned in Chapter 3, they did not seem to provide much accuracy in terms of nominal values, with the exception of $model4$) and from (ii) the tested viability of the CAPE ratio to provide a leverage effect, by providing more- or less-than-proportional returns compared to earnings growth of the companies underlying the portfolio.

Recall that all the simulations stemming from the CAPE-based portfolio management approaches proposed were implemented without taking into consideration the negative impacts of transaction costs, brokerage fees, tax effects and other similar factors which could hamper the cumulative investment returns, for the reasons aforementioned. Nevertheless, it is possible to provide evidence of these strategies proving effective and profitable (with respect to the usual S&P 500 buy-and-hold benchmark), even accounting for the potential impact of these negative components. To this extent, Figure 28 pictures the performance of the three approaches proposed in this study, based on one-month ahead expectations forecasted by $model4$, assuming that the investor would incur a 2% transaction fee every time he adjusts the portfolio's exposure.

Notably, all three strategies still appear to outperform the S&P 500 buy-and-hold investment, with Long/short similarly representing the most profitable methodology

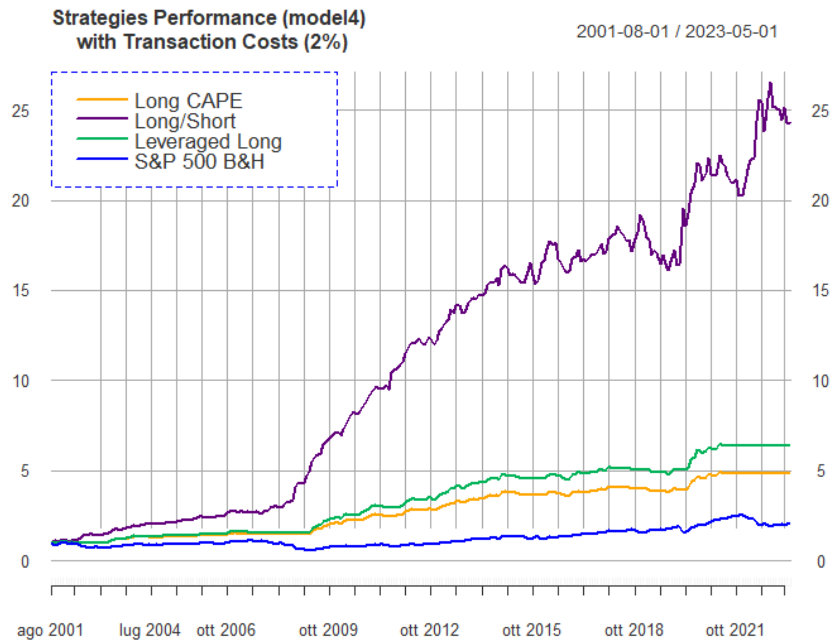

Figure 28: CAPE-based strategies ($model4$) with 2% transaction costs

by a considerable margin. Overall, the effect of a 2% transaction fee has the impact of approximately halving the cumulative return achieved by the portfolio management approaches explored. Specifically:

- Long CAPE displays a total cumulative return of $\approx$ 387% (compared to $\approx$ 970% achieved assuming the absence of transaction fees). Hence, this strategy resulted in being the most hampered by the occurrence of investment costs, reducing the final cumulative value of the portfolio by $\approx$ 60%;

- The presence of investment costs impacted Long/Short by decreasing its cumulative returns to $\approx$ 2325% (from $\approx$ 4278%). However, even accounting for potential negative factors, this strategy still significantly outperformed the S&P 500 buy-and-hold portfolio, providing evidence in favour of the potential practical application of this study's finding in real investment scenarios;

- Lastly, Leveraged Long's cumulative performance declined from $\approx$ 1187% to $\approx$ 538%;

- Conversely, it is important to note that the performance of the S&P 500 buy-and-hold strategy remains unchanged compared to previous representations, since the definition of the benchmark in question requires that the portfolio is never adjusted throughout the investment period analysed.

This study shed light on an alternative view relative to the role of the CAPE ratio in investment strategies, exploring portfolio simulations on the set of data encompassing the S&P 500 Index, from August 2001 to May 2023[45], paired with the selected behavioural variables extensively described throughout this work. To conclude, it is worth mentioning as a disclaimer that the implementation of the methodologies discussed may vary depending on the current condition of financial markets and the investor's brokerage account (e.g. short-selling and leveraged positions may not be allowed or the margin required may change significantly depending on the current interest rates level), as well as on the tax rate applied in different jurisdictions, among other similar factors. Additionally, the Long/Short and Leveraged Long approaches require an increased risk exposure, by definition. Hence, this implies that the strategies explored in this study may result attractive and efficient only to a selected group of professional investors. As a final remark, also note that this work did not argue nor test that the propositions mentioned applied to other countries, stock indexes or equity market instruments.

[45]The original dataset, as described at the beginning of Chapter 3, included observations from July 2001 to June 2023. However, due to the computation of variables such as forward CAPE and differentiated series, the final simulations have been carried out on the horizon mentioned in this paragraph.

References

Aga, Mehmet and B. Kocaman (Jan. 2006). "An empirical investigation of the relationship between inflation, P/E ratios and stock price behaviors using a new series called index-20 for Istanbul Stock Exchange". In: *International Research Journal of Finance and Economics* 6, pp. 133–165.

Akintoye, Ishola (Nov. 2008). "Efficient Market Hypothesis and Behavioural Finance: A Review of Literature". In: *European Journal of Social Sciences* 7(2).

Allahyaribeik, Neda et al. (2020). "The impact of P/E ratio and price return on the stock market Bohmian quantum potential approach". In: vol. 5(18), pp. 55–62.

Angelini, Natascia et al. (Mar. 2013). *Value matters: Predictability of Stock Index Returns*. arXiv: `1204.5055 [q-fin.GN]`.

Aras, Güler and Mustafa Kemal Yilmaz (2008). "PRICE-EARNINGS RATIO, DIVIDEND YIELD, AND MARKET-TO-BOOK RATIO TO PREDICT RETURN ON STOCK MARKET: EVIDENCE FROM THE EMERGING MARKETS". In: *Journal of Global Business and Technology* 4, p. 18. URL: `https://api.semanticscholar.org/CorpusID:150642952`.

Baek, Chung and Ingyu Lee (May 2018). "U.S. stock market P/E ratios, structural breaks, and long-term stock returns". In: *Journal of Business Economics and Management* 19, pp. 110–123. DOI: `10.3846/16111699.2017.1409263`.

Baillie, Richard et al. (Mar. 2022). "On Robust Inference in Time Series Regression". In: URL: `https://ar5iv.labs.arxiv.org/html/2203.04080`.

Basu, S. (1977). "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis". In: *The Journal of Finance* 32(3), pp. 663–682. ISSN: 00221082, 15406261. URL: `http://www.jstor.org/stable/2326304` (visited on 02/01/2024).

Bathia, Deven and Don Bredin (Jan. 2012). "An examination of investor sentiment effect on G7 stock market returns". In: *European Journal of Finance - EUR J FINANC* 19. DOI: `10.1080/1351847X.2011.636834`.

Bogle, John C. (2017). *The Little Book of Common Sense Investing: The Only Way to Guarantee Your Fair Share of Stock Market Returns*. 2nd. John Wiley & Sons, p. 307.

Box, G.E.P. et al. (2015). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9781118674925. URL: `https://books.google.it/books?id=rNt5CgAAQBAJ`.

Campbell, John and Robert Shiller (1988). *Stock Prices, Earnings and Expected Dividends*. NBER Working Papers 2511. National Bureau of Economic Research, Inc. URL: https://EconPapers.repec.org/RePEc:nbr:nberwo:2511.

Campbell, John and Robert Shiller (2001). *Valuation Ratios and the Long-Run Stock Market Outlook: An Update*. NBER Working Papers 8221. National Bureau of Economic Research, Inc. URL: https://EconPapers.repec.org/RePEc:nbr:nberwo:8221.

Campisi, Giovanni and Silvia Muzzioli (Sept. 2020). "Investor sentiment and trading behavior". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30(9). ISSN: 1054-1500. DOI: https://doi.org/10.1063/5.0011636.

Chahine, Salim and Taufiq Choudhry (Feb. 2004). "Price-to-Earning Growth Ratio and Value Vs. Growth Based Strategies: Some European Evidences". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.498306.

Chicago Board Options Exchange (2024). *VIX Index*. https://www.cboe.com/tradable_products/vix/. Accessed: 2024-02-05.

Cowles, Alfred (Jan. 1939). *Common-Stock Indexes*. 2nd ed. Available at: https://cowles.yale.edu/research/cfm-32-common-stock-indexes-2nd-ed. Cowles Foundation for Research in Economics at Yale University: New Haven, CT.

Dreman, David N. and Michael A. Berry (1995). "Overreaction, Underreaction, and the Low-P/E Effect". In: *Financial Analysts Journal* 51, pp. 21–30. URL: https://api.semanticscholar.org/CorpusID:154678403.

Fisher, Kenneth L. and Meir Statman (2000). "Investor Sentiment and Stock Returns". In: *Financial Analysts Journal* 56(2), pp. 16–23. ISSN: 0015198X. URL: http://www.jstor.org/stable/4480229 (visited on 02/01/2024).

Fisher, Kenneth L. and Meir Statman (2003). *Consumer Confidence and Stock Returns*. DOI: 10.3905/jpm.2003.319925.

Graham, Benjamin (1949). *The Intelligent Investor*. Harper & Brothers: New York.

Graham, Benjamin and David L. Dodd (1934). *Security Analysis*. First edition. Whittlesey House, McGraw-Hill Book Company, Inc.: New York.

Huang, Ying, Chia-Hui Tsai, and Carl R. Chen (Apr. 2007). "Expected P/E, Residual P/E, and Stock Return Reversal: Time-Varying Fundamentals or Investor Overreaction?" In: *International Journal of Business and Economics* 6(1), pp. 11–28.

Hyndman, Rob J and George Athanasopoulos (2018). *Forecasting: Principles and Practice*. ISBN: 9780987507112.

Johnston, J. and J. DiNardo (1997). *Econometric Methods*. 4th. McGraw-Hill. ISBN: 0079131212.

Joseph, V. Roshan (2022). "Optimal Ratio for Data Splitting". In: *arXiv preprint arXiv:2202.03326*.

Kane, Alex, Alan J. Marcus, and Jaesun Noh (1996). "The P/E Multiple and Market Volatility". In: *Financial Analysts Journal* 52(4), pp. 16–24. ISSN: 0015198X. URL: http://www.jstor.org/stable/4479930 (visited on 02/01/2024).

Kelly, Simone, Jenna McClean, and Ray McNamara (Aug. 2008). "The Low P/E Effect and Abnormal Returns for Australian Industrial Firms". In: *Business papers*. DOI: 10.2139/ssrn.1254643.

Kenourgios, Dimitris, Spyros Papathanasiou, and Anastasia Bampili (Sept. 2022). "On the predictive power of CAPE or Shiller's PE ratio: the case of the Greek stock market". In: *Operational Research* 22. DOI: 10.1007/s12351-021-00658-x.

Lee, Wayne Y., Christine X. Jiang, and Daniel C. Indro (2002). "Stock market volatility, excess returns, and the role of investor sentiment". In: *Journal of Banking & Finance* 26(12), pp. 2277–2299. URL: https://EconPapers.repec.org/RePEc:eee:jbfina:v:26:y:2002:i:12:p:2277-2299.

Ljung, G. M. and G. E. P. Box (Aug. 1978). "On a measure of lack of fit in time series models". In: *Biometrika* 65(2), pp. 297–303. ISSN: 0006-3444. DOI: 10.1093/biomet/65.2.297. eprint: https://academic.oup.com/biomet/article-pdf/65/2/297/649058/65-2-297.pdf. URL: https://doi.org/10.1093/biomet/65.2.297.

Maggina, Anastasia (2011). "MBAR Models: A Test of ARIMAX Modelling". In: *Review of Pacific Basin Financial Markets and Policies* 14(02), pp. 347–366. DOI: 10.1142/S0219091511002299. eprint: https://doi.org/10.1142/S0219091511002299. URL: https://doi.org/10.1142/S0219091511002299.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55(3), pp. 703–708. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1913610 (visited on 02/06/2024).

Ruppert, David and David S. Matteson (2015). *Statistics and Data Analysis for Financial Engineering*. 2nd. Springer. DOI: 10.1007/978-1-4939-2614-5.

S&P 500® Overview (2024). https://www.spglobal.com/spdji/en/indices/equity/sp-500/\#overview. Accessed: 2024-02-02.

Smales, Lee (Jan. 2017). “The importance of fear: investor sentiment and stock market returns”. In: *Applied Economics* 49, pp. 1–27. DOI: 10.1080/00036846.2016.1259754.

Toleva, Borislava (May 2021). “The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems”. In: *Business Systems Research Journal* 12. DOI: 10.2478/bsrj-2021-0015.

Valetkevitch, Caroline (May 2013). “Key dates and milestones in the S&P 500’s history”. In: *Reuters*. Accessed: 2024-02-02. URL: https://www.reuters.com/article/idUSBRE9450WL/.

Wang, Wenzhao, Chen Su, and Darren Duxbury (2021). “Investor sentiment and stock returns: Global evidence”. In: *Journal of Empirical Finance* 63, pp. 365–391. ISSN: 0927-5398. DOI: https://doi.org/10.1016/j.jempfin.2021.07.010. URL: https://www.sciencedirect.com/science/article/pii/S092753982100058X.

Weigand, Robert and Robert Irons (Mar. 2005). “The Market P/E Ratio: Stock Returns, Earnings, and Mean Reversion”. In: *The Journal of Portfolio Management*, pp. 87–101.

Yale School of Management, International Center for Finance (2024). *Stock Market Confidence Indices*. https://som.yale.edu/centers/international-center-for-finance/data/stock-market-confidence-indices. Accessed: 2024-02-05.