https://lib.uliege.be https://matheo.uliege.be

Does the real economy still forecast technology stock returns in the United States listed companies? Evidence in an age of digital economy and crises

Auteur : Paschal, Alexi
Promoteur(s) : Hambuckers, Julien
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en ingénieur de gestion, à finalité spécialisée en Financial Engineering
Année académique : 2023-2024
URI/URL : http://hdl.handle.net/2268.2/19857

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

DOES THE REAL ECONOMY STILL FORECAST TECHNOLOGY STOCK RETURNS IN THE UNITED STATES LISTED COMPANIES? EVIDENCE IN AN AGE OF DIGITAL ECONOMY AND CRISES.

Jury:	Master thesis by
Supervisor:	**Alexi PASCHAL**
Julien HAMBUCKERS	With a view to obtaining the diploma of
Reader:	Master's Degree in Business Engineering
Romain CRUCIL	specialising in Financial Engineering
	Academic year 2023/2024

Ecole de Gestion de l'Université de Liège

1. Acknowledgements

I would like to express my deep gratitude and sincere thanks to HEC Liège for these five years of learning and enrichment from both a personal and intellectual point of view.

First of all, I would like to thank Mr Philippe Hubner for his guidance and help with RStudio.

I would also like to express my gratitude to Guofu Zhou for his enlightened advice and availability; his help was a major factor in the success of this work.

Finally, I would like to express my gratitude to my supervisor, Julien Hambuckers, for his support, advice, and scientific rigour, and to Romain Crucil for taking the time to read my thesis.

Finally, I would like to thank all those who have contributed in any way to the completion of this dissertation.

2. Table of contents

1. Acknowledgements 3
2. Table of contents 5
3. List of figures and tables 9
   3.1 List of figures 9
   3.1 List of tables 9
4. List of abbreviations - Glossary 11
   4.1 List of abbreviations 11
   4.2 Glossary 11
5. Introduction 13
   5.1 Presentation of this research paper 13
      5.1.1 What is the underlying goal of this work? 13
      5.1.2 What is the second underlying goal of this work? 14
   5.2 Setting the scene 14
      5.2.1 Forecasting stock returns 14
      5.2.2 What is the difference between in-sample and out-sample? 15
      5.2.3 Why focus the research on the US technology sector? 16
   5.3 Research Methodology 18
      5.3.1. Getting to know better the selected variables 18
      5.3.2. Explaining the choice of the Nasdaq100 against the S&P 500 19
      5.3.3. Methodology in brief 20
6. Step 1: Data collection and preprocessing 23
   6.1 Description of the data sources 23
   6.2 Data cleaning 23
      6.2.1 Data quality and integrity 23
      6.2.2 Data cleaning and variable transformation 23
7. Step 2: Descriptive analysis using Power BI 25
   7.1 Introduction 25
   7.2 Strengths of Power BI over Microsoft Excel 25
   7.3 Why choose Power BI? 25
   7.4 Final data 26
   7.5 Starting with Power BI 26
   7.6 Results of the descriptive analysis 27
      7.6.1 Scatter plots 27
      7.6.2. Correlation matrix 30
   7.7 Summary of the findings 31

8. Step 3: Methodology replication using Rstudio ... 33
8.1 Introduction ... 33
8.2 The choice of using RStudio instead of Matlab ... 33
8.3 Explanation of the methodology ... 34
8.3.1 The "Forecasts_quarterly_log_version" files ... 35
8.3.2 The "Forecasts_quarterly_version" files ... 37
8.3.3 The "nchoosek" file ... 37
8.3.4 The "nwest" file ... 38
8.3.5 The "perform_asset_allocation" file ... 38
8.3.6 The "zscore" file ... 38
8.4 Explanation of the three different out-of-sample periods ... 38
8.5 Explanation of the two scenarios ... 39
8.6 Explanation of the different models ... 39
8.6.1 The individual predictive regressions forecast ... 39
8.6.2 The Kitchen Sink forecast ... 40
8.6.3 The SIC forecast ... 40
8.6.4 The pool forecast ... 41
8.6.5 The diffusion index forecast ... 41
8.6.6 The sum-of-the-parts forecast ... 42
8.7 Presentation of the results ... 42
8.7.1 Results for the out-of-sample period: 2002-2022 ... 43
8.7.2 Results for the out-of-sample period: 2010-2022 ... 52
8.7.3 Results for the out-of-sample period: 2021-2022 ... 61
8.8 Discussion of findings ... 70
8.8.1 Regarding the individual predictors ... 70
8.8.2 About the combinations ... 71
8.8.3 About the models ... 72
8.8.4 Differences about the economic time ... 73
8.8.5 Differences between the unrestricted and the under C&T restrictions scenario ... 74
8.8.6 Summary of the main findings ... 74
8.9 Results comparison ... 75
9. Step 4: Machine learning analysis using Google Colab and Python ... 77
9.1 Introduction to machine learning ... 77
9.2 Introduction to Python ... 77
9.3 The choice of using Google Colab ... 78
9.4 Explanation of the methodology ... 78
9.5 Presentation of the results ... 80
9.6 Discussion of findings ... 82

10. Newly published article 85
  10.1 Introduction 85
  10.2 Main findings 85
  10.3 Comparison to the above-mentioned findings 86
11. Conclusions 87
  11.1 Summary of the findings sur R et sur le machine learning 87
  11.2 Implication and significance 88
  11.3 Limitations 89
  11.4 Future research 90
  11.5 Overall conclusion 90
12. Appendices 93
13. List of contacts 99
14. Bibliography and references 100
Executive Summary 106

3. List of figures and tables

3.1 List of figures

Figure 1 - Relative performance of the Nasdaq 100 against the S&P 500 in base 100 over the years, Refiniv Eikon 16
Figure 2 - Successful regions over ten years in attracting the world's (Thomas, 2023)........ 17
Figure 3 - Improvements in ability to create breakthrough innovation in core technologies over the next five years by region (Thomas, 2023) 17
Figure 4 - Correlation between "DFR" and "Equity Premium"........ 28
Figure 5 - Correlation between "log(DE)" and "Equity Premium" 28
Figure 6 - Correlation between "SVAR" and "Equity Premium 29
Figure 7 - Dates of U.S. recessions as inferred by GDP-based recession indicator (Dates of U.S. Recessions as Inferred by GDP-based Recession Indicator, 2024)........ 38

3.1 List of tables

Table 1 - Composition of the Nasdaq 100 index per industry as of 03/28/2024 (Overview for XNDX, 2024)........ 20
Table 2 - Correlation matrix 31
Table 3 - Results under the C&T restrictions for the out-of-sample period 2002-2022 44
Table 4 - Summary of the six different models – C&T restrictions – Overall – 2002 to 2022........ 44
Table 5 - Results under the C&T restrictions for the out-of-sample period 2002-2022, during economic recession times........ 45
Table 6 - Summary of the six different models – C&T restriction – Recession – 2002 to 2022........ 46
Table 7 - Results under the C&T restrictions for the out-of-sample period 2002-2022, during economic expansion times........ 47
Table 8 - Summary of the six different models – C&T restrictions – Expansion – 2002 to 2022........ 47
Table 9 - Results under the unrestricted scenario for the out-of-sample period 2002-2022........ 48
Table 10 - Summary of the six different models – Unrestricted – Overall – 2002 to 2022........ 49
Table 11 - Results under the unrestricted scenario for the out-of-sample period 2002-2022, during economic recession times........ 50
Table 12 - Summary of the six different models – Unrestricted – Recession – 2002 to 2022........ 50
Table 13 - Results under the unrestricted scenario for the out-of-sample period 2002-2022, during economic expansion times........ 51
Table 14 - Summary of the six different models – Unrestricted – Expansion – 2002 to 2022........ 52
Table 15 - Results under the C&T restrictions for the out-of-sample period 2010-2022 53
Table 16 - Summary of the six different models – C&T restrictions – Overall – 2010 to 2022........ 53
Table 17 - Results under the C&T restrictions for the out-of-sample period 2010-2022, during economic recession times........ 54
Table 18 - Summary of the six different models – C&T restrictions – Recession – 2010 to 2022 55

Table 19 - Results under the C&T restrictions for the out-of-sample period 2010-2022, during economic expansion times 56
Table 20 - Summary of the six different models – C&T restrictions – Expansion – 2010 to 2022 56
Table 21 - Results under the unrestricted scenario for the out-of-sample period 2010-2022 57
Table 22 - Summary of the six different models – Unrestricted – Overall – 2010 to 2022 58
Table 23 - Results under the unrestricted scenario for the out-of-sample period 2010-2022, during economic recession times 59
Table 24 - Summary of the six different models – Unrestricted – Recession – 2010 to 2022 59
Table 25 - Results under the unrestricted scenario for the out-of-sample period 2010-2022, during economic expansion times 60
Table 26 - Summary of the six different models – Unrestricted – Expansion – 2010 to 2022 61
Table 27 - Results under the C&T restrictions for the out-of-sample period 2021-2022 62
Table 28 - Summary of the six different models – C&T restrictions – Overall – 2021 to 2022 62
Table 29 - Results under the C&T restrictions for the out-of-sample period 2021-2022, during economic recession times 63
Table 30 - Summary of the six different models – C&T restrictions – Recession – 2021 to 2022 64
Table 31 - Results under the C&T restrictions for the out-of-sample period 2021-2022, during economic expansion times 65
Table 32 - Summary of the six different models – C&T restrictions – Expansion – 2021 to 2022 65
Table 33 - Results under the unrestricted scenario for the out-of-sample period 2021- 2022 66
Table 34 - Summary of the six models – Unrestricted – Overall – 2021 to 2022 67
Table 35 - Results under the unrestricted scenario for the out-of-sample 2021-2022, during economic recession times 68
Table 36 - Summary of the six different models – Unrestricted – Recession – 2021 to 2022 68
Table 37 - Results under the unrestricted scenario for the out-of-sample period 2021-2022, during economic expansion times 69
Table 38 - Summary of the six different models – Unrestricted – Expansion – 2021 to 2022 70

4. List of abbreviations - Glossary

4.1 List of abbreviations

- RSZ = Rapach, Strauss and Zhou
- GW = Goyal and Welch
- GDP = Gross Domestic Product
- USA = United States of America
- NYSE = New York Stock Exchange
- S&P 500 index = Standard & Poor's 500 Composite Stock Price Index
- AI = Artificial Intelligence
- E12 = earnings per share
- D12 = dividend per share
- FRED = Federal Reserve Economic Data
- Log(DP) = Dividend-price ratio
- Log(DY) = Dividend yield ratio
- Log(EP) = Earnings-price ratio
- Log(DE) = Dividend-payout ratio
- SVAR = Stock variance
- BM = Book-to-market ratio
- NTIS = Net equity expansion
- TBL = Treasury bill rate
- LTY = Long-term yield
- LTR = Long-term return
- TMS = Term spread
- DFY = Default yield spread
- DFR = Default return spread
- INFL = Inflation
- DMSPE = Discount mean square prediction error
- SIC = Schwarz Information Criterion
- PCA = Principal Component Analysis
- SOP = sum-of-the-parts
- C&T = Campbell and Thomson

4.2 Glossary

- S&P 500 = The S&P 500 is a stock market index based on 500 large companies listed on stock exchanges in the United States (Contributeurs aux projets Wikimedia, 2024a)

- Nasdaq 100 = The Nasdaq-100 is a stock market index. It comprises the 100 largest non-financial companies listed on NASDAQ (Contributeurs aux projets Wikimedia, 2024)

5. Introduction

5.1 Presentation of this research paper

5.1.1 What is the underlying goal of this work?

Everyone wants to invest to make money and to be profitable but, is it a suitable time to invest in the stock market now? That is the question on everyone's mind, and the one to which everyone would like an answer. Research to predict stock market returns, or the equity premium, started more than one century ago. A good example is Dow and Selden (1920) who worked on the role of dividend ratios. Nowadays, a large panel of theoretical and normative literature exists to guide how investors should invest and on which variables they should rely on (Goyal & Welch, 2004).

Welch and Goyal (GW[1]) (2008) authored a paper with the following underlying idea: a real-world investor would not have access to information known only after the fact, whether for creating variables or for the regression coefficients across the entire sample. Moreover, Welch and Goyal (2008) identified bizarre results in the literature: different papers found out that different variables or methods could forecast the equity premium.

In 2008, Welch and Goyal are quite categorical: they could not find a single variable that is a solid and robust help to a real-world investor or, in other words, to an investor that has no access to ex-post information. Following Welch and Goyal (2008), « most variables are just worse than the prevailing historical equity premium average as a predictor, and some even economically significantly so».

In 2009, Rapach, Strauss, and Zhou (RSZ[2]) wrote a paper based on the aforementioned statement. They combine forecasts from a variety of sources such as macroeconomic variables, financial variables, and technical indicators, to create a good forecast of the equity premium, without ex-post data. The authors find that combining these forecasts can significantly improve the accuracy of equity premium forecasts compared to using any single variable alone. Moreover, they indicate that simple combining methods typically outperform more complicated ones.

The paper also examined the links between these forecasts and the real economy. The authors find that the combination forecasts are closely related to a number of macroeconomic variables, such as real Gross Domestic Product (GDP[3]), real net cash flows and real profits. Moreover, they find that it is significantly correlated with growth rate in the three macroeconomic variables mentioned before.

This paper, written below, is an update of Rapach, Strauss, and Zhou (RSZ[4]) 's 2009 results, focusing on the tech stock in the United States (Rapach et al., 2009). The reasons for focusing on the US tech sector will be explained below.

Actually, it was felt appropriate to revisit RSZ's paper because conditions have changed considerably. Indeed, over the last twenty years, there have been three economic recessions: one in the early 2000s with the dot-com bubble, one in 2008 with the Great Recession and the subprime crisis, and one in

---

[1] List of abbreviations

[2] List of abbreviations

[3] List of abbreviations

[4] List of abbreviations

2020 with the covid-19 crisis (Goyal et al., 2021). It is therefore interesting to see how the predictor variables proposed by Goyal and Welch behave in this new sample (Welch & Goyal, 2008).

What is more, the last few years have seen an explosion in the use of digital technology in everyone's daily lives. This digitalisation has had a major impact on the investment and consumption habits of retailers, as described by Amundi (2023). So, it is therefore worth revisiting GW's paper (Welch & Goyal, 2008) to check whether these assumptions still hold.

In a recently published paper, GW themselves admit: "It does not indict papers or authors if findings no longer hold after publication. Except for tautologies, empirical external validity in the social sciences can never be taken for granted. This is even more the case when investors could actively attempt to profit from the findings in these papers (McLean and Pontiff (2016)). All empirical social science research deserves sceptical re-examination" (Goyal et al., 2021).

In the second part of this thesis, research was conducted in an attempt to optimise the proposed set of fourteen indicators.

5.1.2 What is the second underlying goal of this work?

To sum this part up, here are the two questions relating to this piece of writing:

1) Are the assumptions written in 2010 by RSZ still viable today, especially focusing on the technology sector in the United States of America (USA[5])?

2) Is it possible to optimise the set of variables in order to obtain more accurate results?

In this thesis, it will endeavour to meticulously address these two questions, using the foundation of current scientific literature as well as some calculations to evaluate the continued viability of RSZ's assumptions in today's US technology sector and to explore the optimization of variable sets for enhanced accuracy in results.

5.2 Setting the scene

5.2.1 Forecasting stock returns

For a long time, forecasting stock returns has been a major topic of discussion, both for finance workers and academics.

For practitioners in finance, asset allocation requires real-time forecasts of stock returns, and improved stock return forecasts hold the promise of enhancing investment performance (Rapach & Zhou, 2013).

Regarding academics, they are also very interested in stock return forecasts. Indeed, the capability to predict stock returns has significant implications for evaluating market efficiency.

Moreover, grasping how stock returns can be forecasted aids researchers in developing more accurate asset pricing models that align closely with empirical data. Understanding these dynamics is crucial for both theoretical aspects of financial economics (Rapach & Zhou, 2013).

---

[5] List of abbreviations

As highlighted in the RSZ's paper, it is possible to find in the literature plenty of economic variables that could be used as predictors of stock returns. Some of these are cited in the RSZ's paper, such as:

- Dividend-price [Dow (1920), Fama and French (1988, 1989)],
- Earnings-price [Campbell and Shiller (1988, 1998)],
- Book-to-market [Kothari and Shanken (1997), Pontiff and Schall (1998)],
- Nominal interest rates [Fama and Schwert (1977), Campbell (1987), Breen, Glosten, and Jagannathan (1989), Ang and Bekaert (2007)],
- Inflation rate [Nelson (1976), Fama and Schwert (1977), Campbell and Vuolteenaho (2004)],
- Term and default spreads [Campbell (1987), Fama and French (1989)],
- Corporate issuing activity [Baker and Wurgler (2000), Boudoukh, Michaely, Richardson, and Roberts (2007)],
- Consumption wealth ratio [Lettau and Ludvigson (2001)],
- Stock market volatility [Guo (2006)].

We can also find other indicators, such as the fourteen technical indicators in Neely, Rapach, Tu, and Zhou (2014), the short-stock interest holdings in Rapach, Ringgenberg, and Zhou (2016), aggregate accruals in Hirshleifer, Hou, and Teoh (2009), and fourth-quarter growth in personal consumption expenditures in Møller and Rangvid (2015).

It is clear that there is a wealth of scientific research on the subject. It is therefore of paramount importance, but the international community is unable to reach a consensus that is accepted by all. As outlined by Spiegel (2008), the conclusion of Goyal and Welch in 2008, started an issue and a debate within the scientific community. Indeed, Welch and Goyal (2008) stated that most variables are less effective than the average historical equity premium as predictors.

5.2.2 What is the difference between in-sample and out-sample?

To go a bit deeper on the Goyal and Welch statement, their study shows that most of the predictors from the literature do not consistently deliver out-of-sample forecasts of the U.S equity premium than a forecast based on the historical average (Rapach et al., 2009).

As explained on the website Experts (2023) *predictivethought,* on one hand, in-sample performance refers to the accuracy of a model on the same data it was trained on while, on the other hand, out-of-sample performance discusses the accuracy of a model on new and unseen data. Model overfitting and training set bias are phenomena that could impact out-of-sample performance without necessarily affecting in-sample performance.

Model overfitting is defined by Hawkins (2003) as the use of models that do not respect the principle of parsimony, which is, basically, the use of more than the necessary data. Indeed, to produce good modelling models, one should use the necessary data but nothing more.

Training set biases are defined by the website *towardsdatascience* as taking place when the data selected does not capture the entire real-world data (Kangralkar, 2023).

Inoue and Kilian (2005) states that: "it is widely accepted among applied researchers that in-sample tests of predictability have a tendency to indicate spuriously the existence of predictability when there is none.". In other words, they claim that in-sample tests are more likely to reject the null hypothesis of no predictability more frequently than is appropriate at the specified significance level.

In April 2009, Rapach et al. wrote: “Good examples are Bossaerts and Hillion (1999) who fail to find significant evidence of out-of-sample predictive ability in a collection of industrialised countries for a number of variables for 1990–1995, and Goyal and Welch (2008) find that the dividend-price ratio is not a robust out-of-sample predictor of the U.S. equity premium.”

As the goal is to update RSZ’s paper which is done using out-of-sample tests, this paper will only focus on out-of-sample test results.

5.2.3 Why focus the research on the US technology sector?

5.2.3.1 Focusing on the technology sector

Technology has seen rapid growth in recent decades, driven by a broad spectrum of digital advancements. From advanced computer systems, software, and mobile communications to digital platforms and robotics, these innovations are transforming markets as well as the business and work landscapes (Qureshi & Woo, 2022).

Recent developments in artificial intelligence, machine learning, cyber-physical systems, and the Internet of Things are propelling digital transformation even further. This new surge of innovations has the potential to elevate the digital revolution to unprecedented heights (Qureshi & Woo, 2022).

As seen during the macroeconomics lesson by Mr Jousten in the second year of the Bachelor, the Solow model is really interesting when speaking about the technology. Indeed, this course teaches that technological progress is the source of sustained growth, as it currently knows no limits (Jousten, 2021). Paul Romer (1989) says that: “Growth is driven by technological change that arises from intentional investments decisions made by profit maximising agents”.

It is therefore interesting to look closer at the technology sector because it drives economic growth, and it has experienced plenty of new advancements recently.

Moreover, looking at the chart below, it is easy to see that the Nasdaq 100[6] has outperformed the S&P 500[7] for decades. The orange curve stands for Nasdaq 100 while the green curve represents the S&P 500.

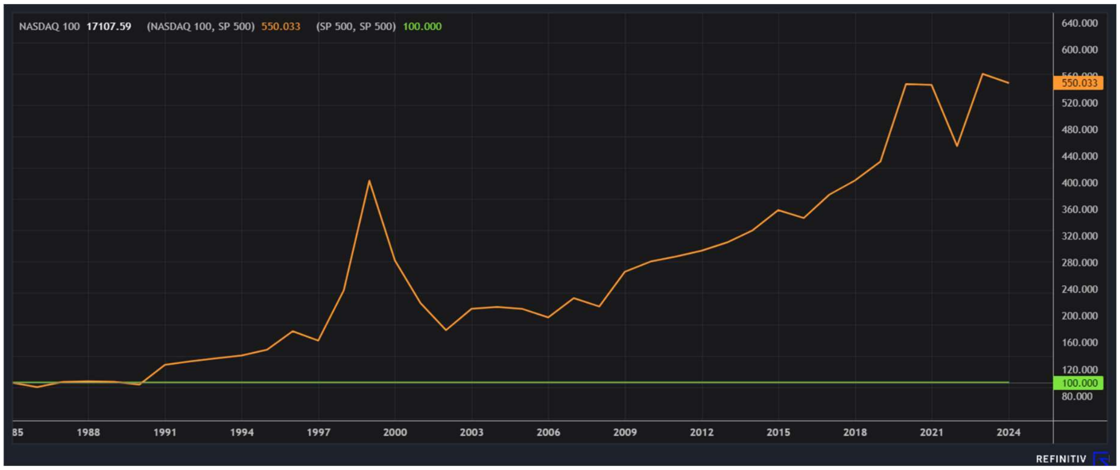


Figure 1 - Relative performance of the Nasdaq 100 against the S&P 500 in base 100 over the years, Refiniv Eikon

---

[6] Glossary

[7] List of abbreviations & Glossary

The Nasdaq 100 and the S&P 500 will be explained later, but this chart clearly shows the strength of the technology sector against the S&P 500. It is good to note that, in the RSZ's paper, the authors take the S&P 500 as a good indicator for the overall economy.

Finally, the technological sector is pretty interesting to analyse, because it is the sector of choice for disruptive technologies. Disruptive technology is defined as "an innovation that significantly alters the way that consumers, industries, or businesses operate." Moreover, "a disruptive technology sweeps away the systems or habits it replaces because it has attributes that are recognizably superior" (Smith, 2022). AI[8], e-commerce or even virtual reality are good examples of recent disruptive technologies that arose.

For all these reasons, it was decided to analyse the technology sector.

5.2.3.2 Focusing on the USA market

The U.S are seen as to have, technologically speaking, the strongest competitive position compared to the rest of the World. The below charts, published in 2023 by Chistopher A. Thomas show and quantified the advantage of the US (Thomas, 2023):

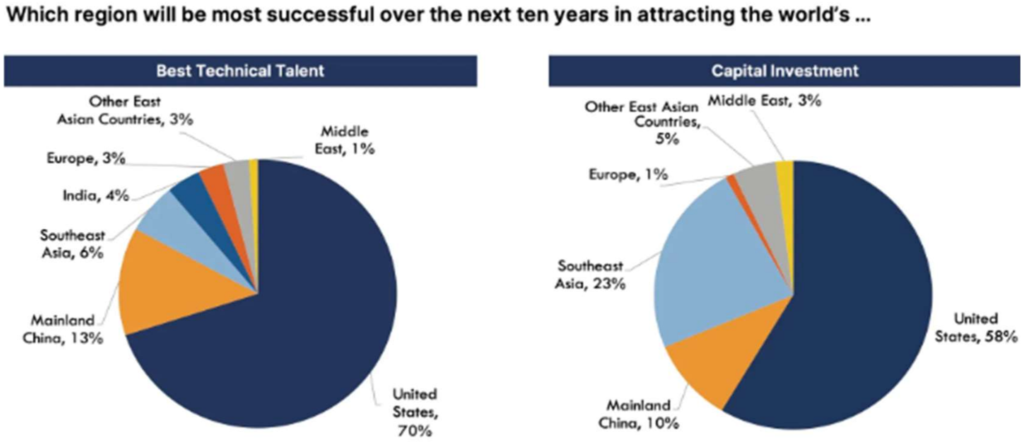


Figure 2 - Successful regions over ten years in attracting the world's (Thomas, 2023)

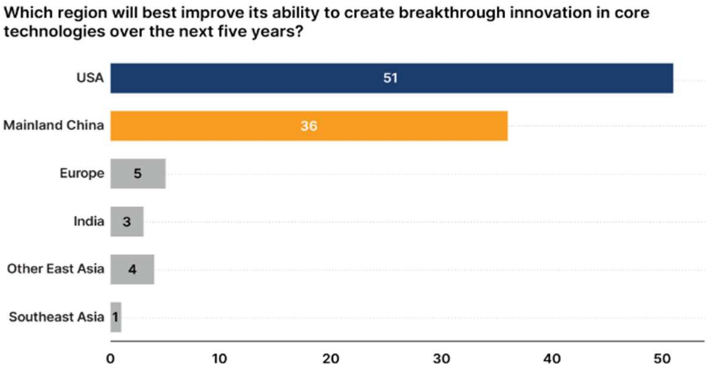


Figure 3 - Improvements in ability to create breakthrough innovation in core technologies over the next five years by region (Thomas, 2023)

Therefore, it is interesting to see how the technology sector behaves in the country that has the biggest competitor advantage in this sector itself.

---

[8] List of abbreviations

5.3 Research Methodology

5.3.1. Getting to know better the selected variables

In this section, a focus is made on the core reasoning and methodology that underpin the research of the paper in general.

The dataset, which comprises quarterly observations, serves as the cornerstone of the ensuing quantitative analyses. Comprising nineteen distinct variables, over a temporal span ranging from the inception of the Nasdaq 100 index in 1986 to the end of the year 2022. These variables, knowing that some are macroeconomic variables while other are microeconomic variables, are enumerated as follows:

- **Yyyy:** represents the date followed by the quarter (in a yyyyq format), serving as a temporal reference point for each observation.
- **Index:** captures the Nasdaq 100 price, this variable is fundamental to the analysis of technology stock performance.
- **D12:** represents the dividend per share over 12 months, D12 provides insights into the dividend yield of the Nasdaq 100 index.
- **E12:** signifies the earnings per share over 12 months, offering a key measure of corporate profitability.
- **B/m:** stands for the ratio of book value to market value for the Dow Jones Industrial Average, B/m, is an essential variable for assessing relative valuations.
- **Tbl:** signifies the interest rate on quarterly Treasury bonds on the secondary market, a critical indicator for assessing the yield curve's slope.
- **AAA:** reflects the yield on Moody's corporate bonds with an AAA- rating, AAA provides insights into the yield of high-quality bonds.
- **BAA:** represents the yield of Moody's corporate bonds with a rating of BAA-, BAA offers an indicator of yields for lower-rated corporate bonds.
- **Lty:** denotes the long-term yield of US government bonds, a benchmark for risk-free interest rates.
- **Ntis:** captures the moving sum of a 12-month rolling window of net issues by NYSE-listed stocks to total end-of-year market capitalization of NYSE stocks.
- **Rfree:** serves as the risk-free rate, crucial for assessing risk-adjusted returns.
- **Infl:** represents CPI inflation among all urban consumers, an important macroeconomic indicator.
- **Ltr:** signifies the long-term yield on government bonds, contributing to an understanding of interest rate trends.
- **Corpr:** capture long-term corporate bond yields, Corpr offers insights into corporate borrowing costs.
- **Svar:** or the variance of equities over the daily returns of the Nasdaq 100 index, provides a measure of stock market volatility.
- **Nasdaq 100 value-weighting return with dividends:** this variable measures the performance of the Nasdaq 100 index, taking into account the dividends paid by the constituent stocks. The value-weighting aspect means that the returns of each stock are weighted according to their market capitalization within the index. Including dividends offers a more comprehensive view of the total investment return, rather than just price changes.

- **Nasdaq 100 value-weighted return excluding dividends:** similar to the previous variable, this one also considers the performance of the Nasdaq 100 index in a value-weighted manner. However, it excludes dividends, focusing solely on the capital gains (or losses) from price changes of the stocks within the index.
- **NBER recession dummies:** this variable consists of indicators that identify periods classified as recessions by the National Bureau of Economic Research (NBER).
- **NBER recession dummies with peaks included:** this is a variant of the NBER recession dummies. Besides indicating recession periods, this variable also includes specific markers for the peaks of economic activity that precede recessions, as identified by the NBER. Including these peaks can provide additional insights into the timing and effects of economic transitions on markets or other economic variables.

It must be stressed that the main differences between this paper and the original RSZ's paper are the variables related to the index. Indeed, it will be explained below what the choice was to replace the S&P 500-related variables with Nasdaq 100-related variables.

All these variables are to be found in the excel file names:" dataset.xlsx".

5.3.2. Explaining the choice of the Nasdaq100 against the S&P 500

5.3.2.1. What is the S&P 500 index

As explained above, the major difference is the choice to use variables related to the Nasdaq 100 instead of the S&P 500 index.

Formally known as the Standard & Poor's 500 Composite Stock Price Index (S&P 500 index[9]), the S&P 500 index tracks the share prices of 500 of the largest public companies in the United-States (Tretina, 2023).

The S&P 500 is an index that monitors the stock prices of large-cap U.S. companies. It provides a snapshot of how the most significant U.S. stocks are performing, gaining, or losing value. As such, the S&P 500 is frequently used as a benchmark to gauge the general state of the stock market and, by extension, the overall U.S. economy (Tretina, 2023).

The idea behind using the S&P 500 index as a variable was, in the RSZ's paper, to replicate the full market.

5.3.2.2. What is the Nasdaq 100 index

The Nasdaq is the second-largest stock exchange in the world, behind the New York Stock Exchange (NYSE[10]).

In depth, the Nasdaq 100 is an index launched in 1985, which is composed of 100 of the largest and most actively traded stocks listed on the Nasdaq stock exchange.
This index features a diverse mix of both domestic and international firms across multiple sectors such as (Tretina, 2023a):

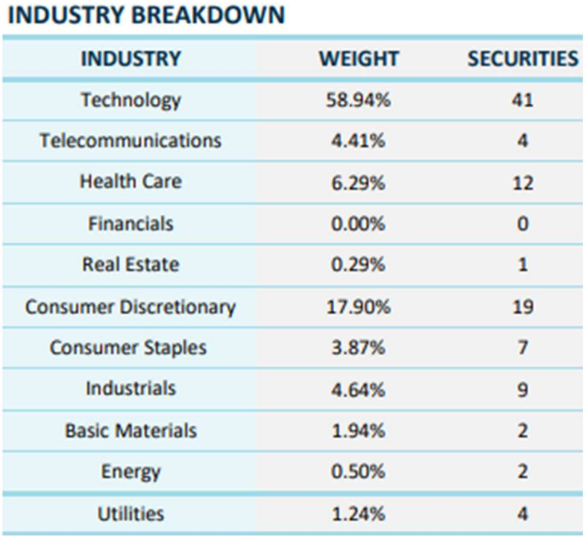
---

[9] List of abbreviations

[10] List of abbreviations

- Basic materials
- Consumer goods and services
- Healthcare
- Industrial
- Technology
- Telecommunications
- Utilities

In the factsheet provided by the Nasdaq Index Services team, the technology industry weighs 59% of the whole Nasdaq weight. Some even say that the Nasdaq 100 index is used as a barometer of the health of the technology sector (Overview for XNDX, 2024).



INDUSTRY BREAKDOWN

INDUSTRY	WEIGHT	SECURITIES
Technology	58.94%	41
Telecommunications	4.41%	4
Health Care	6.29%	12
Financials	0.00%	0
Real Estate	0.29%	1
Consumer Discretionary	17.90%	19
Consumer Staples	3.87%	7
Industrials	4.64%	9
Basic Materials	1.94%	2
Energy	0.50%	2
Utilities	1.24%	4

Table 1 - Composition of the Nasdaq 100 index per industry as of 03/28/2024 (Overview for XNDX, 2024)

This paper has been constructed using the Nasdaq 100 index because it is a good compromise between focusing on the technology sector while not neglecting the rest of the economy. Indeed, it was also possible to find other technology indexes, such as the NXDT (Nasdaq 100 technology Sector Index) , launched in 2008 or the NYTECH (NYSE Technology Index), launched in 2003 but this would reduce the sample of available data, which would have a negative effect on the representativeness of this study. What is more, in both cases, it would be too recent to have the effects of the dot-com crisis.

5.3.3. Methodology in brief

This paper reproduces to the best as it can the paper written in 2009. It says that plenty of the variables proposed in the existing literature predict well in-sample but do poorly out-of-sample.

Moreover, the paper choses multiple variables to evaluate whether or not they can predict the equity premium well or not. As said earlier by Spiegel (2008), predicting the equity premium has been an interesting topic for years, both for academics and for the people in the finance industry.

The paper presents their results, where they focused on the coefficient of determination, the p-value, and the utility function. This allows them to conclude that a combination of different single variables make a better and relevant prediction out-of-sample.

This paper has the same methodology and purposes, only focusing on the technology sector for the U.S. listed companies. Moreover, after the computations explained above, there will be a section that tries to optimise the set of variables, using some machine learning techniques. Back in time, when the authors drafted their papers, machine learning techniques were not as developed as they are now, and they could bring some relevant information to this topic.

To do so, this paper contains four distinct steps:

1) Data collection and preprocessing.
2) Descriptive analysis using Power BI.
3) Replication of computations made for the main study using RStudio.
4) Machine learning analysis using Google Colab and Python.

Let's dive into the following steps now.

6. Step 1: Data collection and preprocessing

6.1 Description of the data sources

The temporal scope of the dataset is carefully delineated, covering a span from the inception of the Nasdaq 100 index in 1986, commencing with the first quarter, denoted as "19861," and culminating with the conclusion of the year 2022. This period is a period of great interest because it includes an era characterised by the advent of the digital economy (Amundi, 2023), as well as some economic crises that have punctuated this era (Goyal et al., 2021). Moreover, the entire study is conducted at a quarterly time frame, in consonance with the original study, facilitating temporal comparability.

The data collection process for this research endeavour adheres to rigorous academic standards to ensure the reliability and validity of the dataset. The primary source of data for this study is derived from Amit Goyal's meticulously compiled database[11] (*Amit Goyal*, n.d.). This database, which has been maintained and updated until the conclusion of the year 2022, forms the foundation upon which the subsequent analyses are constructed. As said previously, these variables encompass a spectrum of economic indicators, encompassing elements such as the bond return, bond yields, and other key financial metrics.

However, for certain variables such as the Nasdaq 100 index, earnings per share (E12[12]), and dividend per share (D12[13]), external sources were consulted. The Nasdaq 100 index data was retrieved from the Federal Reserve Economic Data (FRED[14]), while E12 and D12 data were extracted from GuruFocus. The decision to include data from these external sources was informed by the consistency and verifiability of the information. Moreover, there was a need to retrieve the recession information. These data were found on the FRED as well.

6.2 Data cleaning

6.2.1 Data quality and integrity

To maintain data quality and integrity, an assessment of the dataset was conducted. This assessment encompassed an examination of potential issues such as data completeness, accuracy, and consistency.

Any breach or anomalies detected during were addressed. Missing data, where encountered, was either imputed using suitable methodologies such as regression to ensure that the dataset remained reliable and internally consistent.

6.2.2 Data cleaning and variable transformation

Data cleaning commenced with a comprehensive examination of the dataset to detect and address any irregularities or discrepancies that might compromise data integrity. An initial step involved identifying and rectifying missing values. In instances where outliers were detected, they were scrutinised for accuracy, and when deemed valid, were retained in the dataset with appropriate notation.

---

[11] https://docs.google.com/spreadsheets/d/1bM7vCWd3WOt95Sf9qjLPZjoiafgF_8EG/edit#gid=407859737

[12] List of abbreviations

[13] List of abbreviations

[14] List of abbreviations

Another step was the variable transformation, which was executed in order to make the variables suitable for the analysis. Notably, the E12 variable contained data gaps during the period spanning from 19861 to 20054. To address this issue, a temporally coherent imputation methodology was employed, predicated on relevant historical data. In this case, a regression analysis has been done with the Nasdaq 100. E12 has been estimated with the value of the Nasdaq 100 multiplied by the constant. To this, the variable X has been added and the standard error removed to the results so approximate E12.

In summary, the cleaning of this dataset has been done with the aim of ensuring that all the data are assembled in order to have a solid and consistent foundation to start the calculations, which will be explained in the next step. The objective of this data cleaning process was to ensure data completeness, accuracy, and consistency before starting the computations.

7. Step 2: Descriptive analysis using Power BI

7.1 Introduction

Per *researchmethod,* descriptive analysis is centred on summarising and explaining raw data to make it easily understandable. This analytical approach offers insights into past events by analysing historical data to uncover patterns, trends, and key observations. It frequently employs visual tools to present data in an easily interpretable format, enhancing the understanding of what has occurred (Hassan, 2023).

Originally, Microsoft Excel was initially used to analyse all data for the research. However, due to the large volume of data and the need to aggregate it for meaningful insights, Excel proved inadequate. A search was conducted to find a more suitable tool, informed by various articles and the "Business Analytics" course taught by Professor Michael Schyns (Schyns, 2022). After investigation, "Power BI" was identified as a fitting solution. This will be developed in the next part.

7.2 Strengths of Power BI over Microsoft Excel

Power BI, an online service by Microsoft, empowers users to visualise and represent data. This platform allows for seamless sharing of reports and dashboards within departments, companies, and even to the general public (Singh & Jadhav, 2022). Recently, it has gained significant traction in the fields of artificial intelligence, data visualisation, and data analysis (Krishnan et al., 2017).

A primary functional advantage of Power BI over Microsoft Excel lies in its ability to clearly visualise data. Power BI offers a range of visualisation tools, from simple figures to gradient-coloured maps (Krishnan et al., 2017). What distinguishes Power BI is its ability to create dynamic, rather than static, charts, which lets users observe trends and changes over time. When a specific element is selected, the rest of the report adjusts automatically to highlight information relevant to that selection. Additionally, Power BI offers a feature to explore data at varying levels of aggregation, which is very insightful (Mercurio & Merrill, 2021).

7.3 Why choose Power BI?

Power BI was chosen for its ability to accept Excel sheets as input sources, simplifying the transfer of data from initial Excel files to Power BI. Singh & Jadhav (2022) explain that this tool can transform small amounts of raw Excel data into shareable analytical reports and dashboards, offering unparalleled interaction with Excel ("Why Power BI - Features and Benefits | Microsoft Power BI," n.d.).

Furthermore, this software provides real-time visual representations of collected data by autonomously and instantly updating information (Patil et al., 2022). This feature offers a significant time-saving advantage.

Finally, data analysis involves building datasets, examining them, cleaning them, removing undefined values and outliers, and converting them into usable results (Singh & Jadhav, 2022). Power BI includes "Power Query Editor," a powerful tool that records all actions required to ingest, transform, delete, and load data, enabling automated repetition of these steps for data updates. Updates to data sources automatically refresh all Power BI views (Capris et al., 2023).

Capris et al. (2023) note that Power BI's functionality extends beyond data visualisation and can enhance decision-making by integrating data analysis. They suggest that Power BI could be applied at higher data analysis levels using machine learning.

To summarise, Power BI is a robust data visualisation tool that enhances the analysis of linked data, allowing users to ask the right questions for actionable insights (Mercurio & Merrill, 2021). Though many tools exist for data analysis, Power BI is the most popular method to learn the basics of data analysis (Singh & Jadhav, 2022), making it the ideal choice for this project phase.

7.4 Final data

Based on the variables explained at the point 5.3.1 new variables have been computed through Excel. Here is the list from the RSZ's paper : "

- Dividend-price ratio, **log(DP)**: difference between the log of dividends paid on the Nasdaq 100 index and log of stock prices (Nasdaq 100 index), where dividends are measured using a one-year moving sum.
- Dividend yield, **log(DY)**: difference between the log of dividends and log of lagged stock prices.
- Earnings-price ratio, **log(EP)**: difference between the log of earnings on the Nasdaq 100 index and log of stock prices, where earnings are measured using a one-year moving sum.
- Dividend-payout ratio, **log(DE)**: difference between the log of dividends and log of earnings.
- Stock variance, **SVAR**: sum of squared daily returns on the Nasdaq 100 index.
- Book-to-market ratio, **BM**: ratio of book value to market value for the Dow Jones Industrial Average.
- Net equity expansion, **NTIS**: ratio of twelve-month moving sums of net issues by NYSE listed stocks to total end-of-year market capitalization of NYSE stocks.
- Treasury bill rate, **TBL**: interest rate on a 3-month Treasury bill (secondary market).
- Long-term yield, **LTY**: long-term government bond yield.
- Long-term return, **LTR**: return on long-term government bonds.
- Term spread, **TMS**: difference between the long-term yield and Treasury bill rate.
- Default yield spread, **DFY**: difference between BAA- and AAA-rated corporate bond yields.
- Default return spread, **DFR**: difference between long-term corporate bond and long-term government bond returns.
- Inflation, **INFL**: calculated from the CPI (all urban consumers); following Welch and Goyal (2008), since inflation rate data are released in the following month, we use $x_{i,t-1}$ in (1) for inflation. "

These variables[15] are the final individual predictors that are going to be evaluated in this research and need to be computed in order to have a solid basis for comparison with the paper we update.

7.5 Starting with Power BI

Attached to this document, it is possible to find a file regarding Power BI. All these analyses have been done by first importing the *'PowerBI_fin.xlsx'*, which stores all the final data explained at the point explained above.

---

[15] All of them are referenced in the list of abbreviations.

7.6 Results of the descriptive analysis

7.6.1 Scatter plots

Firstly, in order to visualise the data globally, it was decided to create a scatter plot between each variable and Y. According to *Atlassian:* "a scatter plot uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables" (Atlassian, n.d.).

According to *Atlassian,* a scatter plot is valuable for spotting different patterns in data. It allows for the grouping of data points by observing how closely certain sets cluster together. Additionally, scatter plots can reveal unexpected gaps in the data or the presence of outliers. This capability is particularly useful for segmenting the data into distinct categories, such as in the creation of user personas. Therefore, concerning the distribution of points:

- scattered points indicate high variability in the data.
- clusters of points could indicate sub-populations or specific behaviours in the data.

Next, according to *Atlassian*, it is typical to include a trend line. This line represents the best mathematical fit to the data and offers further insight into the strength of the relationship between the variables. It also helps identify any outliers that might influence the trend line's calculation. In other words, concerning the direction of the trend line:

- a positive trend indicates that the values are increasing simultaneously: if the value of X increases, the value of Y also increases.
- a negative trend indicates that the values act inversely: if the value of X increases, the value of Y decreases, and vice versa.

Thirdly, according to *Atlassian*, scatter plots are frequently used to identify correlational relationships between variables. Typically, the plots help predict the vertical value for a given horizontal value. The variable on the horizontal axis is often referred to as the independent variable, while the variable on the vertical axis is called the dependent variable. The relationships between these variables can be characterised in various ways: they may be positive or negative, strong or weak, and linear or nonlinear:

- if the change in the X variable is proportional to the change in the Y variable, then the trend line is linear.
- if the change in variable X involves a change in variable Y that is not proportional, and varies at diverse levels of the variable, then the trend line is not linear.

In this way, certain behaviours have already been identified between the different variables and the target variable, Y.

A graph was generated for each variable in order to analyse how it affected the reference variable, Y. Three different cases were chosen for analysis. The other graphs can be found in Appendix from 1 to 11.

- Relationship between DFR and Equity Premium:

First of all, it may be noted that the data are grouped together. This means that there is less variability between the data. In addition, the curve shows a positive trend between the two variables, meaning that an increase in DFR also results in an increase in the Equity Premium variable.

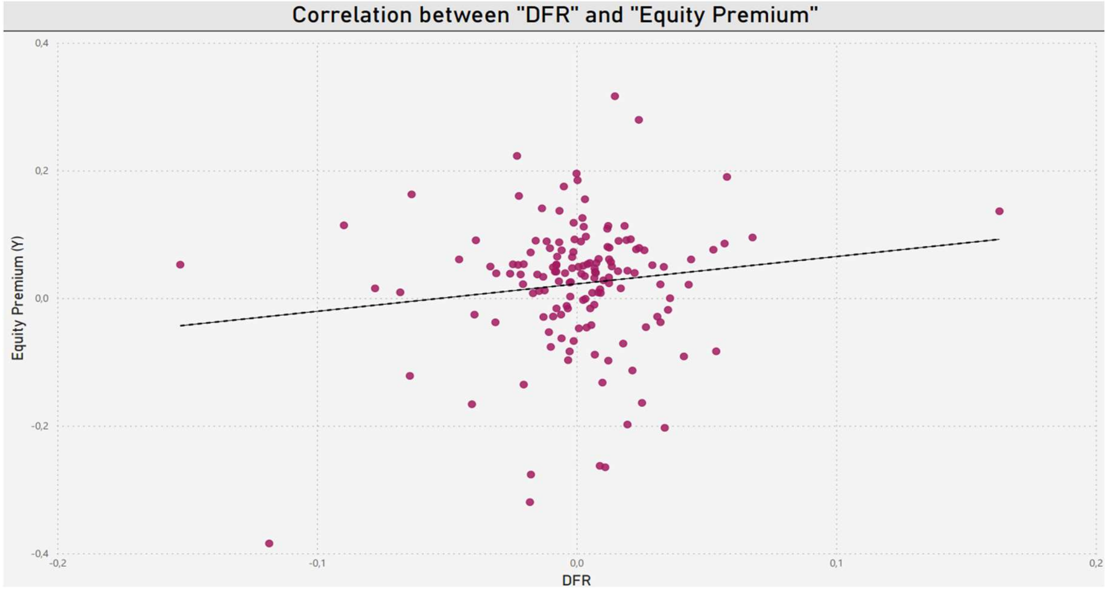


Figure 4 - Correlation between "DFR" and "Equity Premium"

- Relationship between log(DE) and Equity Premium :

The graph shows that the data points are scattered. The trend line is relatively flat, indicating little direct relationship between the log(DE) and Equity Premium variables.

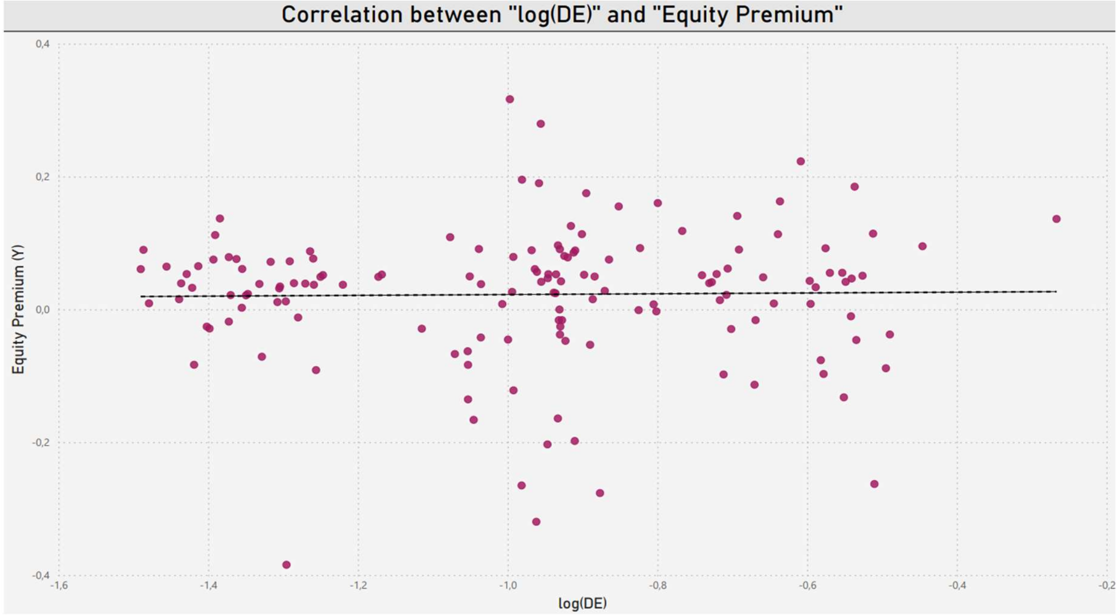


Figure 5 - Correlation between "log(DE)" and "Equity Premium"

- Relationship between SVAR and Equity Premium:

There is a clear clustering of most of the data towards the left. The variability between these values is therefore lower. The trend line shows a clearly negative relationship between the two variables. However, once the extreme values are removed, the relationship becomes much more neutral.

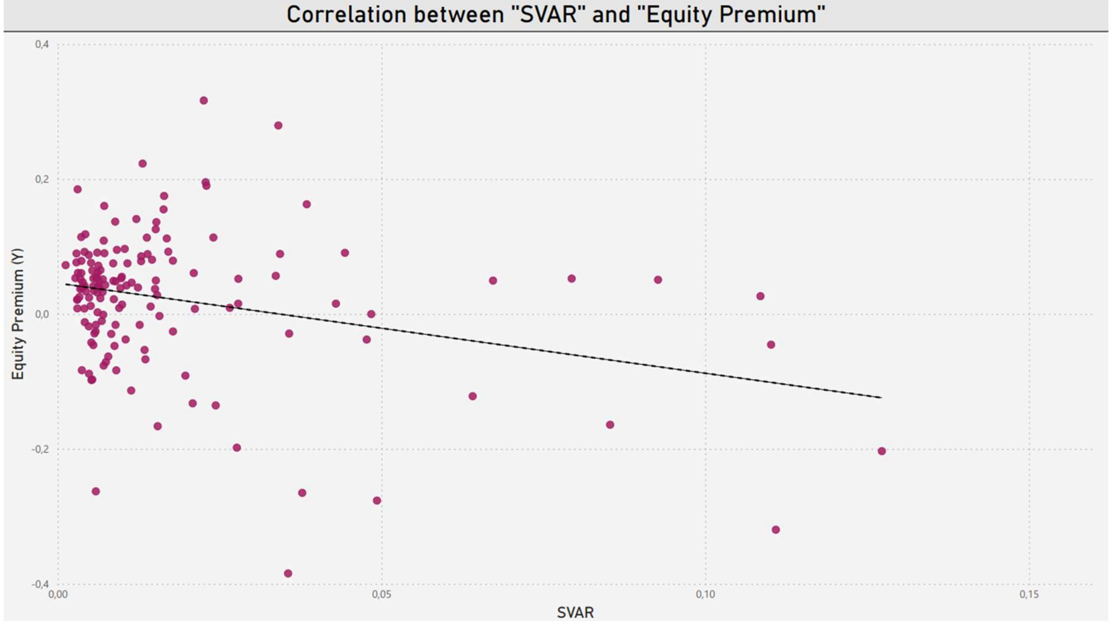


Figure 6 - Correlation between "SVAR" and "Equity Premium

The following are the full description of the behaviour of each variable not yet mentioned. As said above, the related graphs can be found in Appendix from 1 to 11.

- Relationship between BM and Equity Premium:

We note that the data are scattered. In addition, the trend line indicates a slightly positive relationship between the variables (Appendix 1).

- Relationship between DFY and Equity Premium:

The data are also more clustered. The trend line is slightly negative. The DFY and Equity Premium variables are therefore negatively correlated (Appendix 2).

- Relationship between INFL and Equity Premium:

It should be noted that most of the data points are grouped together, so there is less variability between them. The two variables are negatively correlated (Appendix 3).

- Relationship between log(DP) and Equity Premium:

On the one hand, the graph below shows that the points are widely scattered; there is therefore a great deal of variability between the data. On the other hand, the trend line is slightly negative, indicating a negative relationship between the variables (Appendix 4).

- Relationship between log(DY) and Equity Premium:

The points are scattered and the curve indicates a positive trend between the evolution of the variables log(DY) and Equity Premium (Appendix 5).

- Relationship between log(EP) and Equity Premium:

It can be seen that the data are more clustered, with fewer extreme values. There is therefore less variability between values. The curve shows a negative relationship between log(EP) and Equity Premium (Appendix 6).

- Relationship between LTR and Equity Premium:

Certain groups of data can be distinguished on the graph. The direct relationship between the two variables is slightly positive (Appendix 7).

- Relationship between LTY and Equity Premium:

The graph shows fairly scattered data and a slightly negative trend line (Appendix 8).

- Relationship between NTIS and Equity Premium:

The data points are scattered, and the curve shows a slightly negative trend (Appendix 9).

- Relationship between TBL and Equity Premium:

The graph clearly shows a clear dispersion of the data. The curve shows a modestly negative trend (Appendix 10).

- Relationship between TMS and Equity Premium:

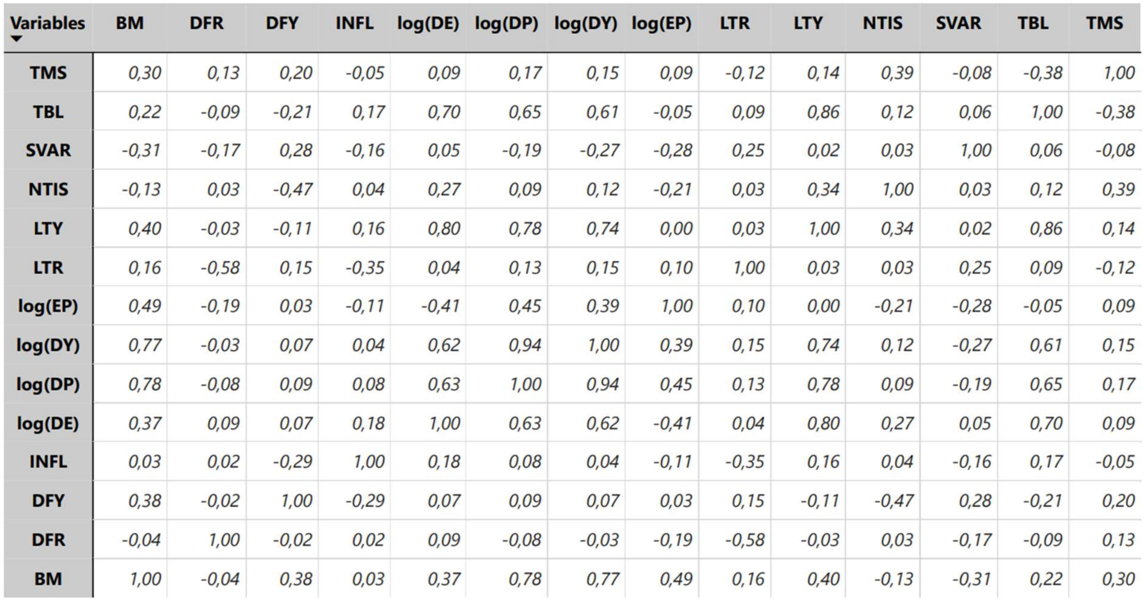
The graph also shows a clear dispersion of the data. The trend line is relatively flat, with a small negative trend (Appendix 11).

Overall, all the trendlines were identified as linear.

7.6.2. Correlation matrix

Similarly to what has been done, in RZS, table 2 presents the correlation matrix for the predictive regression model forecasts that are combined to create the overall forecasts. It is not surprising to find that the correlations among forecasts are generally high. However, many other correlations are notably low, and some are even negative. These observations suggest that combining forecasts can likely decrease the variance of the combined forecasts compared to those of the individual models. The reduction in variance, assuming it does not lead to a substantial increase in bias, also enhances the performance of the combination forecasts over the historical average forecast.

Table 2 illustrates the forecasts from individual predictive regression models for the period 1986:1–2022:4. The figure demonstrates that forecast combinations could reduce forecast variability. In general, the individual forecasts tend to be noisy and prone to misleading signals, which detracts from their forecasting accuracy. In contrast, the mean combination forecast might be more stable, showing potentially more realistic variations in magnitude.



Variables	BM	DFR	DFY	INFL	log(DE)	log(DP)	log(DY)	log(EP)	LTR	LTY	NTIS	SVAR	TBL	TMS
TMS	0,30	0,13	0,20	-0,05	0,09	0,17	0,15	0,09	-0,12	0,14	0,39	-0,08	-0,38	1,00
TBL	0,22	-0,09	-0,21	0,17	0,70	0,65	0,61	-0,05	0,09	0,86	0,12	0,06	1,00	-0,38
SVAR	-0,31	-0,17	0,28	-0,16	0,05	-0,19	-0,27	-0,28	0,25	0,02	0,03	1,00	0,06	-0,08
NTIS	-0,13	0,03	-0,47	0,04	0,27	0,09	0,12	-0,21	0,03	0,34	1,00	0,03	0,12	0,39
LTY	0,40	-0,03	-0,11	0,16	0,80	0,78	0,74	0,00	0,03	1,00	0,34	0,02	0,86	0,14
LTR	0,16	-0,58	0,15	-0,35	0,04	0,13	0,15	0,10	1,00	0,03	0,03	0,25	0,09	-0,12
log(EP)	0,49	-0,19	0,03	-0,11	-0,41	0,45	0,39	1,00	0,10	0,00	-0,21	-0,28	-0,05	0,09
log(DY)	0,77	-0,03	0,07	0,04	0,62	0,94	1,00	0,39	0,15	0,74	0,12	-0,27	0,61	0,15
log(DP)	0,78	-0,08	0,09	0,08	0,63	1,00	0,94	0,45	0,13	0,78	0,09	-0,19	0,65	0,17
log(DE)	0,37	0,09	0,07	0,18	1,00	0,63	0,62	-0,41	0,04	0,80	0,27	0,05	0,70	0,09
INFL	0,03	0,02	-0,29	1,00	0,18	0,08	0,04	-0,11	-0,35	0,16	0,04	-0,16	0,17	-0,05
DFY	0,38	-0,02	1,00	-0,29	0,07	0,09	0,07	0,03	0,15	-0,11	-0,47	0,28	-0,21	0,20
DFR	-0,04	1,00	-0,02	0,02	0,09	-0,08	-0,03	-0,19	-0,58	-0,03	0,03	-0,17	-0,09	0,13
BM	1,00	-0,04	0,38	0,03	0,37	0,78	0,77	0,49	0,16	0,40	-0,13	-0,31	0,22	0,30

Table 2 - Correlation matrix

7.7 Summary of the findings

This section provided a better understanding of the mechanics of the dataset.

On one hand, it is quite important to note that the scatter plots analysis shows that all the variables behave differently with the target variable, Y.

On the other hand, the correlation matrix allows us to understand how all the variables behave in relation to each other. They are all correlated in diverse ways, which means that there is potentially good synergy between the variables when they are combined. This combination reduces the volatility of predictions and provides a more accurate forecast over time.

This last point will be explored in the previous part of this research in order to test whether individual predictive regressors forecast better in an out-of-sample sample than the combination of these different individual variables.

8. Step 3: Methodology replication using RStudio

8.1 Introduction

This section, using RStudio, might be the one that was the most time-consuming to be completed. In this section, the core calculations of the research have been computed. Let's introduce this section.

A Matlab code was found on the Zhou's website[16] under the section "Forecasting Stock Returns". This code was built and never updated since the publication of their article, back in 2009. For this part of the study, the Matlab code was replicated using RStudio, this choice will be explained later.

The sole purpose of this code is to find if, as explained in 2009, a combination of multiple variables predicts the equity premium better than single predictors. Therefore, to do this, there was a need to calculate the coefficient of determination, the p-value and the utility of each single predictors and comparing them to the historical average. Afterwards, these single predictors are compared to some models using multiple predictors to see what the most relevant option is.

8.2 The choice of using RStudio instead of Matlab

Both RStudio and Matlab are programming and computing platforms for statistical computations and graphics. Let's dive into the programs in order to understand better the difference:

On one hand, RStudio is an integrated development environment for R and Python. RStudio is open-sourced and free to use. It can be used on normal desktops, such as Mac, Windows, and Linux. Its environment is quite complete and includes a console, syntax-highlighting tools, availability for plotting, history, debugging and workspace management (*Download RStudio | the Popular Open-Source IDE From Posit*, 2023).

On the other hand, Matlab offers an environment easy to use for iterative analysis and design processes. Its environment is composed of a programming language for matrix and array mathematics directly. Moreover, there is a live editor for creating scripts that combine code, output, and text in a ready-to-execute notebook (*MATLAB*, n.d.).

Both options could work for this thesis. However, I decided to work on RStudio because we used it during the bachelor classes and the master classes at HEC-Liège. Finally, RStudio is free while Matlab needs to be bought. This also motivated my choice, because RStudio's free nature and ease of access mean that the general public can revise and improve the codes in this paper as easily as they like.

---

[16] https://apps.olin.wustl.edu/faculty/zhou/zpublications.html

8.3 Explanation of the methodology

Before diving deep into the practical details on the methodology, let's explain what the main ideas behind this section are.

The goal is to find the nineteen distinct variables explained at the point 5.3.1 and to transform them in order to get the fourteen different variables explained at the point 7.4. and the equity premium variable. Overall, these are minor modifications and computations within the dataset.

Once the fourteen variables, which are the same as the ones in RSZ, are ready, the goal is to compare the R² of each variable against the predictability of the historical average of the equity premium.

Concretely, here is how to compute the out-of-sample R²:

$$\mathrm{R}_{\mathrm{OS}}^{2} = 1 - \frac{\sum_{\mathrm{k=q0+1}}^{\mathrm{q}} \left(\mathrm{r}_{\mathrm{m+k}} - \hat{\mathrm{r}}_{\mathrm{m+k}}\right)^{2}}{\sum_{\mathrm{k=q0+}}^{\mathrm{q}} \left(\mathrm{r}_{\mathrm{m+k}} - \bar{\mathrm{r}}_{\mathrm{m+k}}\right)^{2}}$$

Where,

- $\hat{r}$ is either an individual forecast based on the predictive regression model or a combination forecast.
- $\bar{r}$ is the historical average of the equity premium?
- m and k are periods.

This computes the mean squared prediction error for the different predictive models or combinations against the historical average forecast. When R² > 0, it means that R̂ is outperforming the historical average, which is what it would be nice to have.

As highlighted in RSZ: "a limitation to the R² measure is that it does not explicitly account for the risk borne by an investor over the out-of-sample period.". To solve this issue, it has been decided to compute the utility gains for a mean-variance investor.

$$\hat{\mathrm{v}}_{\mathrm{j}} = \hat{\mu}_{\mathrm{j}} - \left(\frac{1}{2}\right)\gamma\hat{\sigma}_{\mathrm{j}}^{2}$$

Where $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and variance over the out-of-sample period.

In this paper, the utility gain is measured as the difference between an investor that forecasts the equity premium using an individual model or combination and an investor that chooses the historical average to forecast the equity premium. This difference is multiplied by 400 to express it in average annualised percentage return.

In RSZ, it is said that: "the utility gain can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a predictive regression model or combination forecast relative to the information in the historical equity premium alone.".

The R² and the utility gain are computed for each individual predictor, for the combination and for the models in order to compare them later. The individual predictors are quite straightforward, and the models are explained later, but let's dive a bit into the combination.

Basically, the combination forecasts of $\hat{R}$ made at time t are weighted averages of the fourteen individual forecasts. There are three different combinations: mean, median and trimmed mean.

$$\hat{r}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \, \hat{r}_{i,t+1}$$

There are three different combinations: mean, median and trimmed mean.

- The mean sets ω = 1/14 for i = 1,...,14.
- The median value is the median of every $\hat{r}$.
- The trimmed mean combination forecast sets ω = 0 for the individual forecasts with the smallest and largest values and ω = 1/(N −2) for the remaining individual forecasts.

This methodology is similar to the one used in RSZ, as the goal is to update their paper and to compare the previous results they had. Therefore, it had been chosen to compute using the same methodology.

8.3.1 The "Forecasts_quarterly_log_version" files

Attached to this document, it is possible to find ten different files regarding the R code, the "forecasts_quarterly_log_version", the "forecasts_quarterly_version", the "nchoosek", the "perform_asset_allocation", the "nwest" and the "zscore" file.

Regarding the "forecasts_quarterly_log_version, there are three different files, namely:

- "Forecasts_quarterly_log_version_21_fin"
- "Forecasts_quarterly_log_version_22_fin"
- "Forecasts_quarterly_log_version_23_fin"

Each of these files are commented to have a better understanding of the computation made.

The first lines create an environment with two libraries and four sources, which will be explained later on. After loading the predictors on the environment, the first step is to create the variable 'equity_premium'.

Equity premium is defined by *Investopedia* as the "excess return that investing in the stock market provides over a risk-free rate" (Chen, 2023). This is characterised in this case by the above subtraction. The logarithmic expression is used for precision and measurement of the returns (*Arithmetic Returns Vs. Logarithmic Returns*, n.d.).

This line is important because the *equity_premium* is the 'Y' variable which is, in other word, the variable that the study is trying to predict in the best possible way.

After setting up the variables for the sum-of-the-parts model and the estimation of the full-sample parameters for the Campbell and Thompson restriction, which will both be explained below, the preliminaries start at line 104.

At this point, the *equity_premium* variable is renamed Y, while R and P_0 are respectively set up for being the in-sample and out-of-sample period. At line 115, six matrices are created to store different data later in the code.

The computations start at line 137. At this stage, the model is trained with the in-sample data. Firstly, to forecast each individual predictive regression. Then, to apply some Campbell and Thompson restrictions, which will be explained below. Once each individual predictive regression forecast has been made, the code focuses on the different models and combinations to be tested.

It does first the Kitchen sink forecast, then the Schwarz Information Criterion and two pooled forecasts: the simple average and the discount mean square prediction error (DMSPE[17]). To finish forecasting the models, there is the diffusion index forecast and then the sum-of-the-parts forecast. All the models will be explained later.

At line 313, a distinction is made between the economic period of recession and the economic period of expansion. This is decided by the National Bureau of Economic Research, henceforth NBER, variable. This variable is binary, it is either 0 or 1:

- 1 stands for recession time.
- 0 stands for expansion time.

Then, there are few lines on the computation of the forecast errors, the cumulative forecast errors, and the difference in cumulative squared forecast errors. After initialising matrices to store the coefficient of determination and the p-value, there are three different processes to calculate the coefficient of determination and the p-value. One for the overall period of time, one for the recession and one for the expansion.

Then, the similar process is done for the six different models. Once all the computations are made, the end of the code is used to collect the information calculated and arrange it correctly in an excel sheet so that it can be analysed efficiently.

There are three different files because there are three different in-sample periods and out-of-sample periods:

- Forecasts_quarterly_log_version_21_fin has an in-sample period ranging from 1986 to the end of 2001 and an out-of-sample period from 2002 to the end of 2022.

- Forecasts_quarterly_log_version_22_fin has an in-sample period ranging from 1986 to the end of 2008 and an out-of-sample period from 2010 to the end of 2022.

- Forecasts_quarterly_log_version_22_fin has an in-sample period ranging from 1986 to the end of 2019 and an out-of-sample period from 2021 to the end of 2022.

To summarise these three files, the goal is to find the out-of-sample coefficient of determination and the p-value of the single predictors and also the same information for the six different models, during the overall period of time and, afterwards, limited to the economic expansion period of time and the economic recession period of time.

---

[17] List of abbreviations

8.3.2 The “Forecasts_quarterly_version” files

There are three different files, namely:

- “Forecasts_quarterly_version_21_fin”,
- “Forecasts_quarterly_version_22_fin”,
- “Forecasts_quarterly_version_23_fin”.

As of the previous section, each of these files are commented to have a better understanding of the computation made.

Most of the code is the same as the code presented at the point 9.3.1, at least until the moment where a distinction is made between expansion and recession. Therefore, from line 309, the code differs.

The first step is to assign a relative risk-aversion and to implement a window size for estimation volatility. Then, a small function to do a mean-variance optimization. Afterwards, FC_VOL is a function to compute the estimated volatility in computing the variance of the actuals returns.

At line 334, U_HA[2] and U_HA[3] functions are being used: one to calculate the utility during expansions and one to calculate utility during recessions. Once these functions are set up, the code computes the utility for each single predictor for the three different periods: overall, recession and expansion. To finish this part, there is calculation of the utility difference between each single predictors and the historical average.

From line 406, the same calculations are made with the six different models that will be explained later, and then subtracted to the historical average of the utility.

As for the previous files, there are three different files because there are three different in-sample periods and out-of-sample periods:

- Forecasts_quarterly_version_21_fin has an in-sample period ranging from 1986 to the end of 2001 and an out-of-sample period from 2002 to the end of 2022.

- Forecasts_quarterly_version_22_fin has an in-sample period ranging from 1986 to the end of 2008 and an out-of-sample period from 2010 to the end of 2022.

- Forecasts_quarterly_version_22_fin has an in-sample period ranging from 1986 to the end of 2019 and an out-of-sample period from 2021 to the end of 2022.

To summarise, the goal is to find the utility gains of each single predictors and models and then to compare it with the historical average utility to compute the delta, during the overall period of time and, afterwards, limited to the economic expansion period of time and the economic recession period of time.

8.3.3 The “nchoosek” file

Nchoosek is a function that intends to mimic some functionality of the combination operation “n choose k”. This operation selects k unique items from a set of n items. This file is introduced in the six previously explained files and the function is used in a matrix, named “mask”.

8.3.4 The “nwest” file

The *nwest* function is an implementation of the Newey-West estimator, defined by the Havard university as an estimator which is used to compute robust standard errors in regression models to assess the presence of heteroscedasticity and autocorrelation (J. Stock, 2015).

8.3.5 The “perform_asset_allocation” file

The *perform_asset_allocation* function is designed to simulate an asset allocation strategy using principles from portfolio theory. It calculates the optimal weights for risky assets regarding the forecasts of the returns, the volatility and the risk-aversion. The function then computes the expected utility from implementing this allocation. According to *investopedia*, the modern portfolio theory is a strategic approach for selecting investments that aim to maximise returns while adhering to a defined level of risk tolerance. This mathematical framework assists investors in constructing a portfolio that optimises expected returns based on a specified risk level (Team, 2023).

8.3.6 The “zscore” file

The *zscore* function calculates the z-score for a given data vector.

8.4 Explanation of the three different out-of-sample periods

As said previously, the data used in this research is quarterly, starting from 1986 and ending in 2022. The model is trained on in-sample data and then, with the help of the out-of-sample data, it is compared to see the strength of the predictive regression.

It has been chosen that the three recessions that arose after the year 2000 will be the end of each in-sample period. In other words, the model is trained considering the crises to make it more robust and be trained in periods of high volatility to make the model as representative as possible. Here are the three different periods that can be found in this paper:

1) The first period is from 1986 to the end of 2001.
2) The second period is from 1986 to the end of 2009.
3) The third period is from 1986 to the end of 2020.

Those periods are based on the following data, from FRED. Indeed, in the graph below, when the value equals 1, it means that it is a period of recession indicator:

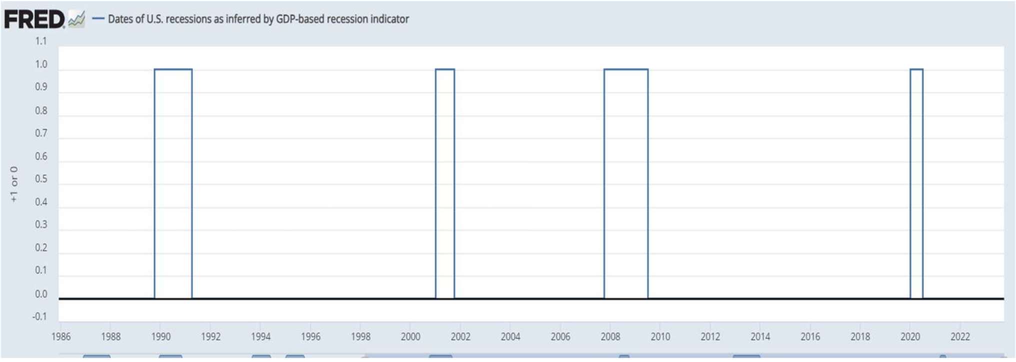


Figure 7 - Dates of U.S. recessions as inferred by GDP-based recession indicator (Dates of U.S. Recessions as Inferred by GDP-based Recession Indicator, 2024)

8.5 Explanation of the two scenarios

In the replicated code, the variables are doubled to fulfil two different scenarios, the one under Campbell and Thompson and the unrestricted one. These scenarios are both explained below.

One scenario comes from the fact that according to their paper, Campbell & Thompson (2007) shows the fact that: “Many predictive regressions beat the historical average return, once weak restrictions are imposed on the signs of coefficients and return forecasts.”. In the code replicated, the underlying of the Campbell and Thompson restrictions are the following: if a beta coefficient is expected to be positive but is estimated to be negative, the forecast is adjusted to use historical data instead. Additionally, any forecast that results in a negative value is set to zero to ensure all forecasts remain non-negative.

In the code, this ensures that if FC_ECON is negative, FC_OTHER_CT is set to zero. In the original paper (Rapach et al., 2009), the Campbell and Thompson (2008) restrictions are applied to individual predictive regression model forecasts to enhance out-of-sample performance for certain variables during specific periods. For instance, TBL and LTY showed improvement in the 1970s, and D/P and D/Y also benefited toward the end of the out-of-sample period.

The other scenario is the unrestricted one. This is basically without placing any restrictions on the calculation to let the variables be free to behave as they want.

8.6 Explanation of the different models

In the code explained above, there are different models. Each of them has different particularities. This section is dedicated to getting to know the different models, in order to better understand the results that will be given in the upcoming sections. Later in this paper, these models will be compared to the combination of the individual predictive regressors to see if they predict more accurately.

8.6.1 The individual predictive regressions forecast

This approach uses individual economic predictors in separate regressions to forecast the equity premium (Rapach et al., 2009).

It has some pros and some cons. In the pros, we can cite the straightforwardness of the use of the indicators. Indeed, it is easy to understand the impact of each predictor.

Concerning the cons of this model, the process of data mining for predictor variables interacts with the bias caused by spurious regression. The combination of these effects strengthens the overall bias because highly persistent data series are more likely to appear significant when searching for predictor variables. The simulations imply that many regressions in existing literature, which rely on individual predictor variables, might be spurious (Ferson et al., 2003).

Moreover, an individual economic variable could send a number of signals that are not necessarily true, which implies an inconceivable equity risk premium for certain periods of time as well as some uncertainty in the predictive method. The authors prefer combining multiple individual forecasts. Indeed, an average of the two forecasts should, in principle, have less volatility and be, therefore, more reliable (Rapach et al., 2009).

8.6.2 The Kitchen Sink forecast

Considering the vast and continuously growing set of predictors for the market equity premium, researchers might logically use a multivariate regression model that includes all available variables. This approach is frequently referred to as the "kitchen-sink" model in related literature (Yin, 2021).

It can therefore be argued that it must be comprehensive because it uses a broad array of information, and it must be potentially robust because it is created by the capture of complex relationships between the predictors.

Contrary to this expectation, Goyal, and Welch (2008) conducted a comprehensive analysis revealing that the kitchen-sink model consistently underperforms compared to the simple historical average benchmark in terms of predictive gains (Yin, 2021).

However, the same paper (Yin, 2021) showed that the very failure of the kitchen-sink might be caused by the presence of multicollinearity between some of the predictive variables. Multicollinearity is defined by *Investopedia* as the presence of strong intercorrelations between two or more independent variables in a multiple regression model. This phenomenon can distort results, complicating efforts to assess how effectively each independent variable predicts or explains the dependent variable (Hayes, 2024).

Multicollinearity typically results in wider confidence intervals, reducing the reliability of probabilities regarding the influence of independent variables in a model (Hayes, 2024).

In technical analysis, multicollinearity can cause erroneous assumptions about an investment, often arising from the use of multiple indicators of the same type to analyse a stock (Hayes, 2024).

8.6.3 The SIC forecast

The Schwarz Information Criterion (SIC[18]), also known as the Bayesian Information Criterion (BIC[19]), is a well-established method for model selection that prefers simpler models over more complex ones by imposing a penalty based on the number of parameters being estimated. The BIC can be calculated using:

$$\boxed{BIC = T_m - df_m ln\,(N)}$$

where Tm is the chi-square statistic for the hypothesised model. In this framework, a SIC greater than 0 favours the saturated model, which allows all observed variables to be intercorrelated without any assumed model structure. On the other hand, a SIC less than 0 supports the hypothesised model (Bauldry, 2015).

RSZ's explored a combination method where the weights are determined by the Schwarz Information Criterion calculated for each individual prediction regression model during the estimation period (Rapach et al., 2009). This approach is equivalent to setting the weights based on the approximate in-sample posterior model probabilities [Draper (1995)].

---

[18] List of abbreviations

[19] List of abbreviations

8.6.4 The pool forecast

Pooling is "the act of sharing or combining two or more things" (*Pooling*, 2024). Found in RSZ, this model is based on Stock and Watson (2004). In these methods, the weights used to combine forecasts at time t are determined by the past forecasting performance of each individual model during a designated out-of-sample period. The discount mean square prediction error (DMSPE) model uses the following weights:

$$\omega_{i,t} = \frac{\emptyset_{i,t}^{-1}}{\sum_{j=1}^{N} \emptyset_{j,t}^{-1}}$$

Where,

$$\emptyset_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} \left(r_{s+1} - \hat{r}_{i,s+1}\right)^2$$

It is good to note that θ is a discount factor. The pool forecast method allocates higher weights to individual predictive regression model forecasts that demonstrate lower mean square prediction error (MSPE) values, indicating superior forecasting performance, over the holdout out-of-sample period.

With θ set to 1, no discounting occurs, and this setting yields the optimal combination forecast originally proposed by Bates and Granger in 1969 (Bates & Granger, 1969). It is applicable when the individual forecasts are not correlated. Setting θ to less than 1 places more emphasis on the recent forecast accuracy of the model.

This paper uses two different cases of θ:

- When the value is 1, the variable is called POOL-AVG
- When the value is 0.9, the model is called pool-DMSPE.

8.6.5 The diffusion index forecast

According to Arthur Broida (1955): "diffusion indexes are simply proportions of "expanding series" out of a given set; they can be calculated for any set of time series whatsoever.". In other words, imagine you have a group of friends, and each of them has their own height growth chart. A diffusion index would tell us what fraction of your friends are still getting taller compared to the others. You can use this method with any group of friends to see how many are still growing.

In this paper, we use the diffusion index forecast with the principal component analysis (PCA[20]), which is a linear transformation that repositions the data into a new coordinate system. The resulting variables, called principal components, are linear combinations of the original variables. These components are uncorrelated, with the highest variance in the data captured along the first axis, the next highest variance along the second, and so forth (Sewell, 2007). The method aims to simplify multivariate data by reducing its dimensionality while retaining as much of the crucial information as possible. This is a type of unsupervised learning, which means it relies solely on the input data without considering any related target data (Sewell, 2007).

---

[20] List of abbreviations

The diffusion index uses the principal component because it can be generalised to oversee data irregularities as well as to handle very large numbers of data. It is also proved that principal components remain consistent with a very small number of data contamination occurrences when the data set is large (Stock & Watson, 2002).

8.6.6 The sum-of-the-parts forecast

The sum-of-the-parts (SOP[21]) is a method proposed by Ferreira and Santa-Clara (2011). It divides the stock market return into three components: the dividend-price ratio, the earnings growth rate, and the price-earnings ratio growth rate. Each component is forecasted separately, leveraging their distinct time series characteristics.

Since the dividend-price ratio is highly stable, it is forecasted using the current dividend-price ratio.
The earnings growth rate, which is generally unpredictable in the short term but shows predictable patterns over long periods (Binsbergen and Koijen, 2010), is forecasted using its long-term historical average, typically a 20-year moving average.
Finally, for simplicity, the SOP method assumes no growth in the price-earnings ratio, aligning closely with the random walk hypothesis for the dividend-price ratio. Therefore, the return forecast is the sum of the current dividend-price ratio and the long-term historical average of earnings growth.

A major concern regarding the findings is the possibility that the authors might have inadvertently chosen coefficients that closely resemble the in-sample estimates from the unrestricted predictive regression during the forecasting period. In that case, the out-of-sample R^2 would merely reflect the in-sample R^2.

To address this, the authors estimated the predictive regression and found that the in-sample estimated coefficients differ significantly from the SOP method's implied assumptions. This alleviates the concern that the SOP method relies on coefficient mining. By using restricted versions of the predictive regression, demonstrating that both the dividend-price ratio and earnings growth components contribute equally to the SOP method's performance.

An observation that the SOP method's performance remains robust, even with different estimates of the dividend-price ratio's persistence and average earnings growth (Ferreira & Santa-Clara, 2011).

8.7 Presentation of the results

Once the code has been finished, it is possible to run it. After running all the different files in RStudio, twelve different excel files will appear in the folder on the user's computer.

One file will be on the computation of the delta under the Campbell and Thompson restrictions. This file is named: "Delta_CT_quarterly".

Another one will be on the computation of the delta unrestricted, thus without any restrictions. This file is named: "Delta_unrestricted_quarterly".

The third file will be on the R^2 and the p-value under the Campbell and Thompson restrictions. This file is named: "ROS_CT_quarterly".

---

[21] List of abbreviations

The fourth file will be on the unrestricted R² and p-value, thus without any restrictions. This file is named: "ROS_unrestricted_quarterly".

Each of these files have been replicated three times regarding the three different out-of-sample periods, as explained above. All of this information has been stored in a single excel file, named: "Results.xlsx". This file contains three different sheets for each out-of-sample period with all the relevant data in it. It has been a bit improved in order to be easily readable. The risk aversion has been set to 5 in the utility function, which implies that the investor is very risk averse.

The given results are the same as the RStudio results. However, three lines have been added: mean, median and trimmed mean. These lines have been created in order to compare a combination of forecasts instead of a single predictor.

The functions are quite simple to implement. For the mean, the "average" function was used. Regarding the median, the "median" function was used. Finally, regarding the trimmed mean, the "trimmean" function was used. The threshold of the data to exclude is set to 5%.

Finally, to have a better understanding of the numbers below, it is good to note that it is against the historical average. Indeed, if the coefficient of determination is negative, it means that it predicts less well the equity premium than the historical average. The delta is the utility function against the historical average as well. A positive delta means that the investor has more utility using this predictor. Positive results are in blue while negative results are in orange, to make this file easier to read.

8.7.1 Results for the out-of-sample period: 2002-2022

8.7.1.1 Under Campbell and Thompson restrictions

1) **Overall**

Below is a table summarising the results under the Campbell and Thompson (C&T[22]) restrictions for the out-of-sample period between 2002 and 2022.

Regarding the coefficient of determination of each single predictors, it is possible to see that six out of fourteen indicators are negative, namely log(DP), log(EP), BM, NTIS, TMS and DFY. SVAR is null and DFR and INFL have a high number, respectively 8.63% and 9.32%.

Regarding the p-value of each single predictors, we can say that, except SVAR that cannot be computed and the p-value of INFL which is pretty low with 5%, the p-value is greatly diversified between 15% and 80% regarding the indicators.

Finally, regarding the utility value of each predictor, we can say that three indicators are negative, namely log(DE), LTY and TMS. Log(DY) has the greater delta, with 9.40% and INFL has the second best delta, with 8.36%.

Looking at the different combinations, the main characteristics that we can discuss are the fact that the R² is positive for each combination, with a utility gain for each combination as well and a p-value between 36% and 42%.

---

[22] List of abbreviations

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-0,95	0,74	0,89
log(DY)	1,15	0,36	9,40
log(EP)	-0,10	0,46	2,18
log(DE)	0,23	0,33	-0,22
SVAR	0,00	#NUM!	0,00
BM	-0,59	0,52	3,63
NTIS	-2,52	0,66	3,13
TBL	1,68	0,31	0,04
LTY	2,10	0,29	-1,58
LTR	2,34	0,12	3,06
TMS	-0,39	0,63	-0,71
DFY	-1,96	0,79	0,58
DFR	8,63	0,15	0,28
INFL	9,32	0,05	8,36
Mean	1,35	0,42	2,07
Median	0,12	0,36	0,73
Trimmed mean	1,35	0,42	2,07

Table 3 - Results under the C&T restrictions for the out-of-sample period 2002-2022

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	18,54	0,01	15,33
SIC	2,57	0,21	4,99
POOL-AVG	1,92	0,08	1,81
POOL-DMSPE	2,07	0,08	1,88
Diffusion index	-1,85	0,84	0,91
Sum-of-the-parts	1,94	0,22	-3,87

Table 4 - Summary of the six different models – C&T restrictions – Overall – 2002 to 2022

We can see that the Kitchen-sink model has the highest R^2 and the highest delta. To recall, this means that the kitchen-sink model predicts the equity premium better than the historical average and it gives the greatest utility for the investor.

However, it has a small p-value, below the threshold of 5%. On the other hand, the coefficient of determination of the Diffusion Index is negative and the delta of the sum-of-the-parts is negative.

2) During economic recession times

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2002 and 2022, during economic recession times.

Regarding the coefficient of determination of each single predictors, it is possible to see that seven indicators out of fourteen are negative. For log(DY) and NTIS, the numbers are quite high, being around -15%. On the other hand, LTY and DFR have quite high positive numbers, around 20%.

Regarding the p-value of each single predictor, we can say that INFL and log(DE) are really smaller than 5% while log(DP), log(DY), NTIS and DFY are higher than 95%.

Finally, regarding the utility value of each predictor, we can say that all the values are negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are the mean and the trimmed mean have positive R^2 while median has a negative one. The delta, however, is negative for the two means and equal to zero for the median.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-4,39	0,98	-0,67
log(DY)	-15,29	0,97	-8,42
log(EP)	-3,41	0,92	-7,83
log(DE)	1,66	0,01	0,00
SVAR	0,00	#NUM!	0,00
BM	-5,75	0,94	-3,55
NTIS	-15,51	0,96	-3,03
TBL	9,17	0,02	0,00
LTY	19,57	0,02	0,00
LTR	-0,93	0,70	0,00
TMS	0,61	0,06	0,00
DFY	-7,01	0,99	-0,27
DFR	21,08	0,15	0,00
INFL	2,45	0,00	0,00
Mean	0,16	0,52	-1,70
Median	-0,46	0,70	0,00
Trimmed mean	0,16	0,52	-1,70

Table 5 - Results under the C&T restrictions for the out-of-sample period 2002-2022, during economic recession times

Below is a table summing up the six different models in a table.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	71,23	0,10	0,00
SIC	-15,72	1,00	0,00
POOL-AVG	-2,81	0,86	0,00
POOL-DMSPE	-2,83	0,87	0,00
Diffusion index	-6,66	0,99	-0,73
Sum-of-the-parts	7,37	0,00	0,00

Table 6 - Summary of the six different models – C&T restriction – Recession – 2002 to 2022

The kitchen sink model has the greatest R², around 71% while sum-of-the-parts is also positive. All the other models have a negative coefficient of determination.

The p-value of SIC and Diffusion Index are above 95%, while the delta is equal to zero for each model, and negative for the diffusion index.

3) During economic expansion times

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2002 and 2022, during economic expansion times.

Regarding the coefficient of determination of each single predictors, it is possible to see that the values are quite extreme, with seven negative indicators, ranging from -41.46% to -3.47%. On the other hand, LTY, TBL, DFR and INFL have some quite high positive numbers, ranging from 24.46% to 60.67%.

Regarding the p-value of each single predictors, we can say that TBL, LTY, DFR and INFL have a p-value smaller than 5% and LTR, NTIS, BM, log(DY), log(DP) have a p-value higher than 95%.

Finally, regarding the utility value of each predictor, we can say that the utility value is negative for each predictor.

Looking at the different combinations, the main characteristics that we can discuss are the fact that the R² is positive for the mean and the trimmed mean while being negative for the median and the delta is negative for the two mean indicators and equal to zero for the median.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-11,77	1,00	-0,67
log(DY)	-41,46	1,00	-8,42
log(EP)	-29,04	0,92	-7,83
log(DE)	0,06	0,48	0,00
SVAR	0,00	#NUM!	0,00
BM	-28,79	1,00	-3,55
NTIS	-25,30	1,00	-3,03
TBL	27,53	0,02	0,00
LTY	60,67	0,01	0,00
LTR	-7,12	1,00	0,00
TMS	6,65	0,10	0,00
DFY	-3,47	0,75	-0,27
DFR	33,08	0,00	0,00
INFL	24,46	0,01	0,00
Mean	0,39	0,56	-1,70
Median	-1,73	0,75	0,00
Trimmed mean	0,39	0,56	-1,70

Table 7 - Results under the C&T restrictions for the out-of-sample period 2002-2022, during economic expansion times

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	-81,24	0,00	0,00
SIC	16,30	0,12	0,00
POOL-AVG	4,71	0,00	0,00
POOL-DMSPE	6,72	0,00	0,00
Diffusion index	-13,62	1,00	-0,73
Sum-of-the-parts	14,42	0,01	0,00

Table 8 - Summary of the six different models – C&T restrictions – Expansion – 2002 to 2022

Kitchen sink is very negative, around -81%, with the diffusion index. The other models are positive. Regarding the p-value, only SIC is between 5% and 95%. All the deltas are equal to zero, except the diffusion index, which is negative.

8.7.1.2 Under the unrestricted restrictions

1) **Overall**

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2002 and 2022.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-1,10	0,78	0,89
log(DY)	1,15	0,36	9,40
log(EP)	-0,10	0,46	2,18
log(DE)	0,23	0,33	-0,22
SVAR	-9,17	0,46	4,99
BM	-0,59	0,52	3,63
NTIS	-2,52	0,66	3,13
TBL	1,68	0,31	0,04
LTY	2,10	0,29	-1,58
LTR	2,34	0,12	3,06
TMS	-0,39	0,63	-0,71
DFY	-1,96	0,79	0,58
DFR	-4,06	0,28	0,28
INFL	10,59	0,05	8,36
Mean	-0,13	0,43	2,43
Median	-0,24	0,41	1,53
Trimmed mean	-0,13	0,43	2,43

Table 9 - Results under the unrestricted scenario for the out-of-sample period 2002-2022

Regarding the coefficient of determination of each single predictors, it is possible to see that eight indicators have a negative coefficient of determination, ranging from -10% to 0.59%. On the other hand, INFL has the greatest R^2, around 10%.

Regarding the p-value of each single predictors, we can say that this is pretty well distributed, except for INFL that has a p-value of 5%.

Finally, regarding the utility value of each predictor, we can say that only three indicators have a negative delta.

Looking at the different combinations, the main characteristics that we can discuss are the negative R^2 for each combination. However, the delta is positive for each combination.

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	20,98	0,004	15,33
SIC	-9,17	0,46	4,99
POOL-AVG	1,92	0,08	1,81
POOL-DMSPE	2,07	0,08	1,88
Diffusion index	-1,85	0,84	0,91
Sum-of-the-parts	1,94	0,22	-3,87

Table 10 - Summary of the six different models – Unrestricted – Overall – 2002 to 2022

It is possible to highlight that the kitchen sink model has the greatest R^2 and the greatest delta.

However, it has a very small p-value. Diffusion Index and SIC have negative R^2 but positive delta, while sum-of-the-parts is the only indicator with a negative delta.

2) During economic recession times

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2002 and 2022, during economic recession times.

Regarding the coefficient of determination of each single predictors, it is possible to see that nine indicators have a negative R^2, with the lowest being -54.74%. TBL and LTY have the higher R^2, respectively with 9.17% and 19.57%.

Regarding the p-value of each single predictors, we can say that INFL, LTY, TBL and log(DE) have a smaller p-value than 5%, while log(DP), log(DY), NTIS and DFY have a larger p-value than 95%.

Finally, regarding the utility value of each predictor, we can say that all the values are either negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are the negative coefficient of determination for each predictor as well as the negative delta for the two indicators representing the mean.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-4,39	0,98	0,00
log(DY)	-15,29	0,97	-4,74
log(EP)	-3,41	0,92	0,00
log(DE)	1,66	0,01	0,00
SVAR	-54,74	0,99999	-34,86
BM	-5,75	0,94	0,00
NTIS	-15,51	0,96	-2,29
TBL	9,17	0,02	0,00
LTY	19,57	0,02	0,00
LTR	-0,93	0,70	0,00
TMS	0,61	0,06	0,00
DFY	-7,01	0,99	0,00
DFR	-20,82	0,38	-34,86
INFL	2,45	0,00	0,00
Mean	-6,74	0,57	-5,48
Median	-3,90	0,81	0,00
Trimmed mean	-6,74	0,57	-5,48

Table 11 - Results under the unrestricted scenario for the out-of-sample period 2002-2022, during economic recession times

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	40,97	0,15	-34,86
SIC	-54,74	0,99999	-34,86
POOL-AVG	-2,81	0,86	-10,44
POOL-DMSPE	-2,83	0,87	-10,38
Diffusion index	-6,66	0,99	0,00
Sum-of-the-parts	7,37	0,002	0,00

Table 12 - Summary of the six different models – Unrestricted – Recession – 2002 to 2022

The R^2 of the kitchen sink model is the highest, around 41% while the other only positive R^2 is sum-of-the-parts. In any case, all the deltas are either equal to zero or negative.

The p-value of sum-of-the-parts is smaller than 5%, while the p-value of SIC is greater than 95%.

3) During economic expansion times

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2002 and 2022, during economic expansion times.

Regarding the coefficient of determination of each single predictors, it is possible to see that seven indicators are negative, with the lowest being -30%. On the other hand, LTY, DFR, TBL and INFL have a R² ranging between 24.46% and 60,67%.

Regarding the p-value of each single predictors, we can say that TBL, LTY, DFR and INFL have a p-value smaller than 5% and LTR, NTIS, BM, log(DY), log(DP) have a p-value higher than 95%.

Finally, regarding the utility value of each predictor, we can say that the utility value is negative for each predictor.

Looking at the different combinations, the main characteristics that we can discuss are the fact that the R² is positive for the mean and the trimmed mean while being negative for the median and the delta is negative for the two mean indicators and equal to zero for the median.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-11,77	1,00	-0,67
log(DY)	-41,46	1,00	-8,42
log(EP)	-29,04	0,92	-7,83
log(DE)	0,06	0,48	0,00
SVAR	16,30	0,12	0,00
BM	-28,79	1,00	-3,55
NTIS	-25,30	1,00	-3,03
TBL	27,53	0,02	0,00
LTY	60,67	0,01	0,00
LTR	-7,12	1,00	0,00
TMS	6,65	0,10	0,00
DFY	-3,47	0,75	-0,27
DFR	33,08	0,00	0,00
INFL	24,46	0,01	0,00
Mean	1,56	0,53	-1,70
Median	-1,70	0,62	0,00
Trimmed mean	1,56	0,53	-1,70

Table 13 - Results under the unrestricted scenario for the out-of-sample period 2002-2022, during economic expansion times

Below is a table summing up the six different models.

There are two negative coefficients of determination, which are kitchen sink and diffusion index, which also has a negative delta. All the other ones have a positive R^2 and a delta equal to zero.

The p-value is smaller than 5% for the kitchen sink, the POOL-AVG, the POOL-DMSPE and the sum-of-the-parts, while being greater than 95% for the diffusion index.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	-81,24	0,002	0,00
SIC	16,30	0,12	0,00
POOL-AVG	4,71	0,00	0,00
POOL-DMSPE	6,72	0,00	0,00
Diffusion index	-13,62	1,00	-0,73
Sum-of-the-parts	14,42	0,01	0,00

Table 14 - Summary of the six different models – Unrestricted – Expansion – 2002 to 2022

8.7.2 Results for the out-of-sample period: 2010-2022

8.7.2.1 Under Campbell and Thompson restrictions

1) Overall

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2010 and 2022.

Regarding the coefficient of determination of each single predictors, it is possible to see that six indicators are negative, with the lowest numbers being -2.52%. On the other hand, INFL and DFR have the highest results, being around 9%.

Regarding the p-value of each single predictors, we can say that INFL has a small value, equal to 5%. However, all the other results are between 5% and 95%.

Finally, regarding the utility value of each predictor, we can say that three indicators are negative, with log(DE), LTY and TMS. Log(DY) and INFL have the highest delta.

Looking at the different combinations, the main characteristics that we can discuss are the positivity of the delta and the R^2, which means that the combinations beat the historical average.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-0,95	0,74	0,89
log(DY)	1,15	0,36	9,40
log(EP)	-0,10	0,46	2,18
log(DE)	0,23	0,33	-0,22
SVAR	0,00	#NUM!	0,00
BM	-0,59	0,52	3,63
NTIS	-2,52	0,66	3,13
TBL	1,68	0,31	0,04
LTY	2,10	0,29	-1,58
LTR	2,34	0,12	3,06
TMS	-0,39	0,63	-0,71
DFY	-1,96	0,79	0,58
DFR	8,63	0,15	0,28
INFL	9,32	0,05	8,36
Mean	1,35	0,42	2,07
Median	0,12	0,36	0,73
Trimmed mean	1,35	0,42	2,07

Table 15 - Results under the C&T restrictions for the out-of-sample period 2010-2022

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	18,54	0,01	15,33
SIC	2,57	0,21	4,99
POOL-AVG	1,92	0,08	1,81
POOL-DMSPE	2,07	0,08	1,88
Diffusion index	-1,85	0,84	0,91
Sum-of-the-parts	1,94	0,22	-3,87

Table 16 - Summary of the six different models – C&T restrictions – Overall – 2010 to 2022

We can see that the Kitchen-sink model has the highest R^2 and the highest delta.

However, it has a small p-value, below the threshold of 5%. On the other hand, the coefficient of determination of the Diffusion Index is negative and the delta of the sum-of-the-parts is negative.

2) During economic recession times

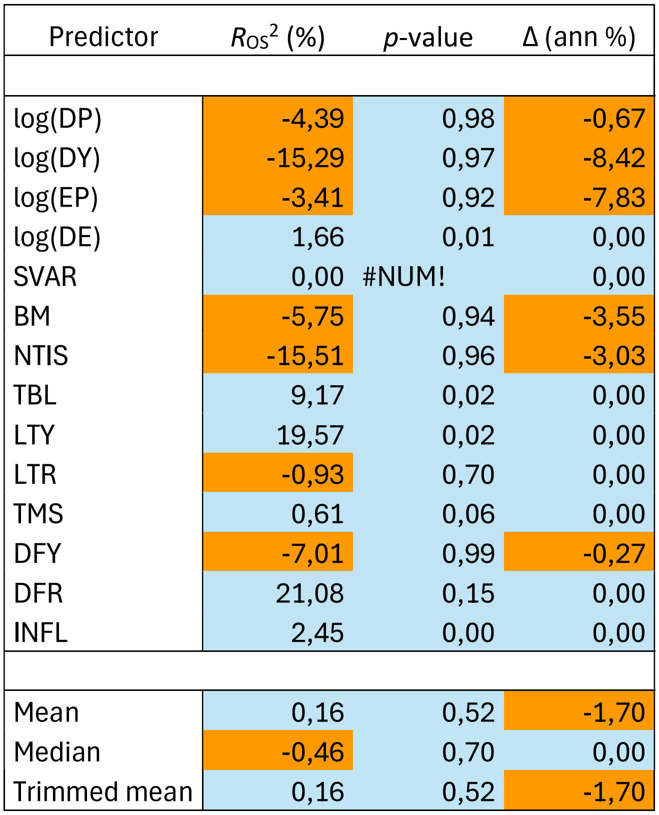
Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2010 and 2022, during economic recession times.

Regarding the coefficient of determination of each single predictors, it is possible to see that seven indicators out of fourteen are negative. For log(DY) and NTIS, the numbers are quite high, being around -15%. On the other hand, LTY and DFR have quite high positive number, around 20%.

Regarding the p-value of each single predictor, we can say that INFL and log(DE) are really smaller than 5% while log(DP), log(DY), NTIS and DFY are higher than 95%.

Finally, regarding the utility value of each predictor, we can say that all the values are negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are the mean and the trimmed mean have positive R² while median has a negative one. The delta, however, is negative for the two means and equal to zero for the median.



Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-4,39	0,98	-0,67
log(DY)	-15,29	0,97	-8,42
log(EP)	-3,41	0,92	-7,83
log(DE)	1,66	0,01	0,00
SVAR	0,00	#NUM!	0,00
BM	-5,75	0,94	-3,55
NTIS	-15,51	0,96	-3,03
TBL	9,17	0,02	0,00
LTY	19,57	0,02	0,00
LTR	-0,93	0,70	0,00
TMS	0,61	0,06	0,00
DFY	-7,01	0,99	-0,27
DFR	21,08	0,15	0,00
INFL	2,45	0,00	0,00
Mean	0,16	0,52	-1,70
Median	-0,46	0,70	0,00
Trimmed mean	0,16	0,52	-1,70

Table 17 - Results under the C&T restrictions for the out-of-sample period 2010-2022, during economic recession times

Below is a table summing up the six different models in a table.

The kitchen sink model has the greatest R², around 71% while sum-of-the-parts is also positive. All the other models have a negative coefficient of determination.

The p-value of SIC and Diffusion Index are above 95%, while the delta is equal to zero for each model, and negative for the diffusion index.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	71,23	0,10	0,00
SIC	-15,72	0,99999	0,00
POOL-AVG	-2,81	0,86	0,00
POOL-DMSPE	-2,83	0,87	0,00
Diffusion index	-6,66	0,99	-0,73
Sum-of-the-parts	7,37	0,002	0,00

Table 18 - Summary of the six different models – C&T restrictions – Recession – 2010 to 2022

3) During economic expansion times

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2010 and 2022, during economic expansion times.

Regarding the coefficient of determination of each single predictors, it is possible to see that the values are quite extreme, with seven negative indicators, ranging from -41.46% to -3.47%. On the other hand, LTY, TBL, DFR and INFL have some quite high positive numbers, ranging from 24.46% to 60.67%.

Regarding the p-value of each single predictors, we can say that TBL, LTY, DFR and INFL have a p-value smaller than 5% and LTR, NTIS, BM, log(DY), log(DP) have a p-value higher than 95%.

Finally, regarding the utility value of each predictor, we can say that the utility value is negative for each predictor.

Looking at the different combinations, the main characteristics that we can discuss are the fact that the R^2 is positive for the mean and the trimmed mean while being negative for the median and the delta is negative for the two mean indicators and equal to zero for the median.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-11,77	1,00	-0,67
log(DY)	-41,46	1,00	-8,42
log(EP)	-29,04	0,92	-7,83
log(DE)	0,06	0,48	0,00
SVAR	0,00	#NUM!	0,00
BM	-28,79	1,00	-3,55
NTIS	-25,30	1,00	-3,03
TBL	27,53	0,02	0,00
LTY	60,67	0,01	0,00
LTR	-7,12	1,00	0,00
TMS	6,65	0,10	0,00
DFY	-3,47	0,75	-0,27
DFR	33,08	0,00	0,00
INFL	24,46	0,01	0,00
Mean	0,39	0,56	-1,70
Median	-1,73	0,75	0,00
Trimmed mean	0,39	0,56	-1,70

Table 19 - Results under the C&T restrictions for the out-of-sample period 2010-2022, during economic expansion times

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	-81,24	0,002	0,00
SIC	16,30	0,12	0,00
POOL-AVG	4,71	0,00	0,00
POOL-DMSPE	6,72	0,00	0,00
Diffusion index	-13,62	1,00	-0,73
Sum-of-the-parts	14,42	0,01	0,00

Table 20 - Summary of the six different models – C&T restrictions – Expansion – 2010 to 2022

Kitchen sink is very negative, around -81%, with the diffusion index. The other models are positive.

Regarding the p-value, only SIC is between 5% and 95%. All the deltas are equal to zero, except the diffusion index, which is negative.

8.7.2.2 Under the unrestricted restrictions

1) Overall

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2010 and 2022.

Regarding the coefficient of determination of each single predictors, it is possible to see that eight indicators have a negative coefficient of determination, ranging from -10% to 0.59%. On the other hand, INFL has the greatest R^2, around 10%.

Regarding the p-value of each single predictors, we can say that this is pretty well distributed, except for INFL that has a p-value of 5%.

Finally, regarding the utility value of each predictor, we can say that only three indicators have a negative delta.

Looking at the different combinations, the main characteristics that we can discuss are the negative R^2 for each combination. However, the delta is positive for each combination.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-1,10	0,78	0,89
log(DY)	1,15	0,36	9,40
log(EP)	-0,10	0,46	2,18
log(DE)	0,23	0,33	-0,22
SVAR	-9,17	0,46	4,99
BM	-0,59	0,52	3,63
NTIS	-2,52	0,66	3,13
TBL	1,68	0,31	0,04
LTY	2,10	0,29	-1,58
LTR	2,34	0,12	3,06
TMS	-0,39	0,63	-0,71
DFY	-1,96	0,79	0,58
DFR	-4,06	0,28	0,28
INFL	10,59	0,05	8,36
Mean	-0,13	0,43	2,43
Median	-0,24	0,41	1,53
Trimmed mean	-0,13	0,43	2,43

Table 21 - Results under the unrestricted scenario for the out-of-sample period 2010-2022

Below is a table summing up the six different models.

It is possible to highlight that the kitchen sink model has the greatest R² and the greatest delta.

However, it has a very small p-value. Diffusion Index and SIC have negative R² but positive delta, while sum-of-the-parts is the only indicator with a negative delta.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	20,98	0,004	15,33
SIC	-9,17	0,46	4,99
POOL-AVG	1,92	0,08	1,81
POOL-DMSPE	2,07	0,08	1,88
Diffusion index	-1,85	0,84	0,91
Sum-of-the-parts	1,94	0,22	-3,87

Table 22 - Summary of the six different models – Unrestricted – Overall – 2010 to 2022

2) During economic recession times

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2010 and 2022, during economic recession times.

Regarding the coefficient of determination of each single predictors, it is possible to see that nine indicators have a negative R², with the lowest being -54.74%. TBL and LTY have the higher R², respectively with 9.17% and 19.57%.

Regarding the p-value of each single predictors, we can say that INFL, LTY, TBL and log(DE) have a smaller p-value than 5%, while log(DP), log(DY), NTIS and DFY have a larger p-value than 95%.

Finally, regarding the utility value of each predictor, we can say that all the values are either negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are the negative coefficient of determination for each predictor as well as the negative delta for the two indicators representing the mean.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-4,39	0,98	0,00
log(DY)	-15,29	0,97	-4,74
log(EP)	-3,41	0,92	0,00
log(DE)	1,66	0,01	0,00
SVAR	-54,74	0,99999	-34,86
BM	-5,75	0,94	0,00
NTIS	-15,51	0,96	-2,29
TBL	9,17	0,02	0,00
LTY	19,57	0,02	0,00
LTR	-0,93	0,70	0,00
TMS	0,61	0,06	0,00
DFY	-7,01	0,99	0,00
DFR	-20,82	0,38	-34,86
INFL	2,45	0,00	0,00
Mean	-6,74	0,57	-5,48
Median	-3,90	0,81	0,00
Trimmed mean	-6,74	0,57	-5,48

Table 23 - Results under the unrestricted scenario for the out-of-sample period 2010-2022, during economic recession times

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	40,97	0,15	-34,86
SIC	-54,74	0,99999	-34,86
POOL-AVG	-2,81	0,86	-10,44
POOL-DMSPE	-2,83	0,87	-10,38
Diffusion index	-6,66	0,99	0,00
Sum-of-the-parts	7,37	0,002	0,00

Table 24 - Summary of the six different models – Unrestricted – Recession – 2010 to 2022

The R^2 of the kitchen sink model is the highest, around 41% while the other only positive R^2 is sum-of-the-parts. In any case, all the deltas are either equal to zero or negative.

The p-value of sum-of-the-parts is smaller than 5%, while the p-value of SIC is greater than 95%.

3) During economic expansion times

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2010 and 2022, during economic expansion times.

Regarding the coefficient of determination of each single predictors, it is possible to see that seven indicators are negative, with the lowest being -30%. On the other hand, LTY, DFR, TBL and INFL have a R² ranging between 24.46% and 60,67%.

Regarding the p-value of each single predictors, we can say that TBL, LTY, DFR and INFL have a p-value smaller than 5% and LTR, NTIS, BM, log(DY), log(DP) have a p-value higher than 95%.

Finally, regarding the utility value of each predictor, we can say that the utility value is negative for each predictor.

Looking at the different combinations, the main characteristics that we can discuss are the fact that the R² is positive for the mean and the trimmed mean while being negative for the median and the delta is negative for the two mean indicators and equal to zero for the median.

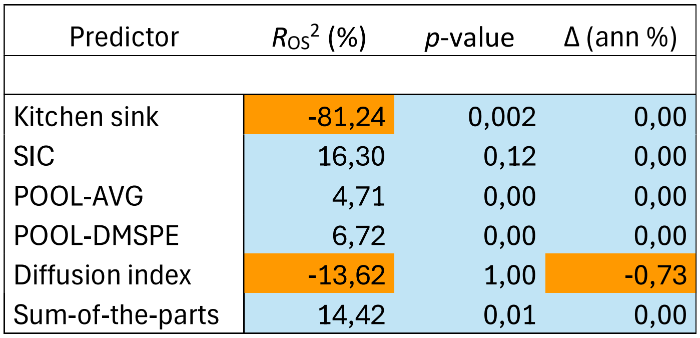
Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-11,77	1,00	-0,67
log(DY)	-41,46	1,00	-8,42
log(EP)	-29,04	0,92	-7,83
log(DE)	0,06	0,48	0,00
SVAR	16,30	0,12	0,00
BM	-28,79	1,00	-3,55
NTIS	-25,30	1,00	-3,03
TBL	27,53	0,02	0,00
LTY	60,67	0,01	0,00
LTR	-7,12	1,00	0,00
TMS	6,65	0,10	0,00
DFY	-3,47	0,75	-0,27
DFR	33,08	0,00	0,00
INFL	24,46	0,01	0,00
Mean	1,56	0,53	-1,70
Median	-1,70	0,62	0,00
Trimmed mean	1,56	0,53	-1,70

Table 25 - Results under the unrestricted scenario for the out-of-sample period 2010-2022, during economic expansion times

Below is a table summing up the six different models.

There are two negative coefficients of determination, which are kitchen sink and diffusion index, which also has a negative delta. All the other ones have a positive R² and a delta equal to zero.

The p-value is smaller than 5% for the kitchen sink, the POOL-AVG, the POOL-DMSPE and the sum-of-the-parts, while being greater than 95% for the diffusion index.



Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	-81,24	0,002	0,00
SIC	16,30	0,12	0,00
POOL-AVG	4,71	0,00	0,00
POOL-DMSPE	6,72	0,00	0,00
Diffusion index	-13,62	1,00	-0,73
Sum-of-the-parts	14,42	0,01	0,00

Table 26 - Summary of the six different models – Unrestricted – Expansion – 2010 to 2022

8.7.3 Results for the out-of-sample period: 2021-2022

8.7.3.1 Under Campbell and Thompson restrictions

1) Overall

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2021 and 2022.

Regarding the coefficient of determination of each single predictors, it is possible to see that four indicators are negative, with LTY being the lowest at around 10%. INFL and log(DY) are the highest numbers, around 11%.

Regarding the p-value of each single predictors, we can say that LTR has a small value of 5%, while all the other indicators are in between 5% and 95%.

Finally, regarding the utility value of each predictor, we can say that the delta is negative for all the indicators having a negative R², except for TBL which has a negative R² but a positive delta.

Looking at the different combinations, the main characteristics that we can discuss are the results are positive, with the two indicators related to the mean having better results than the median.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	1,31	0,09	1,27
log(DY)	11,39	0,08	14,90
log(EP)	3,23	0,18	4,67
log(DE)	-0,41	0,78	-0,28
SVAR	0,00	#NUM!	0,00
BM	3,55	0,19	5,77
NTIS	4,89	0,06	4,28
TBL	-3,38	0,65	0,06
LTY	-9,61	0,71	-2,01
LTR	4,43	0,05	4,01
TMS	-1,29	0,81	-0,91
DFY	0,46	0,40	0,78
DFR	1,36	0,36	4,10
INFL	11,54	0,07	11,77
Mean	1,96	0,34	3,46
Median	1,34	0,19	2,64
Trimmed mean	1,96	0,34	3,46

Table 27 - Results under the C&T restrictions for the out-of-sample period 2021-2022

Below is a table summing up the six different models.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	0,47	0,06	24,23
SIC	10,15	0,04	10,30
POOL-AVG	3,93	0,01	3,37
POOL-DMSPE	4,06	0,02	3,47
Diffusion index	1,11	0,14	1,31
Sum-of-the-parts	-1,34	0,60	-4,96

Table 28 - Summary of the six different models – C&T restrictions – Overall – 2021 to 2022

Only sum-of-the-parts has a negative R^2 and a negative delta. SIC has the greatest R^2, around 10%, while kitchen sink has the highest delta which is around 24.23%.

SIC, POOL-AVG and the POOL-DMSPE have a small p-value, below the threshold of 5%.

2) During economic recession times

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2021 and 2022, during economic recession times.

Regarding the coefficient of determination of each single predictors, it is possible to see that seven indicators are negative, while five indicators are higher than 20%.

Regarding the p-value of each single predictor, we can say that six are below the threshold of 5% and four are above 95%.

Finally, regarding the utility value of each predictor, we can say that all the deltas are negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are that R^2 is positive for the two indicators concerning the mean while the median is negative. The delta, however, is negative for each combination.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-0,94	0,92	-0,07
log(DY)	-44,09	1,00	-10,60
log(EP)	-22,99	1,00	-5,37
log(DE)	-0,44	0,60	0,00
SVAR	0,00	#NUM!	0,00
BM	-28,70	1,00	-4,34
NTIS	20,86	0,00	-0,44
TBL	32,52	0,00	0,00
LTY	52,94	0,00	0,00
LTR	-4,14	1,00	-0,25
TMS	4,49	0,01	0,00
DFY	15,44	0,00	0,00
DFR	28,15	0,01	0,00
INFL	-31,15	0,84	-15,96
Mean	1,57	0,49	-2,64
Median	-0,22	0,60	-0,04
Trimmed mean	1,57	0,49	-2,64

Table 29 - Results under the C&T restrictions for the out-of-sample period 2021-2022, during economic recession times

Below is a table summing up the six different models.

Only the diffusion index has a negative delta and R². Kitchen sink has the highest R², around 70.81%.

Regarding the p-value, kitchen sink, SIC, POOL-AVG, the POOL-DMSPE and sum-of-the-parts have a small p-value, below 5%. On the other hand, the diffusion index has a p-value above 95%.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	70,81	0,02	0,00
SIC	35,81	0,05	0,00
POOL-AVG	6,63	0,01	0,00
POOL-DMSPE	9,43	0,00	0,00
Diffusion index	-2,24	1,00	-0,06
Sum-of-the-parts	15,18	0,01	0,00

Table 30 - Summary of the six different models – C&T restrictions – Recession – 2021 to 2022

3) During economic expansion times

Below is a table summarising the results under the Campbell and Thompson restrictions for the out-of-sample period between 2021 and 2022, during economic expansion times.

Regarding the coefficient of determination of each single predictors, it is possible to see that eight indicators are negative, while three indicators are higher than 30%.

Regarding the p-value of each single predictors, we can say that four are below the threshold of 5% and five are above 95%.

Finally, regarding the utility value of each predictor, we can say that all the deltas are negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are that R² is positive for the two indicators concerning the mean while the median is negative. The delta, however, is negative for each combination.

Predictor	R_{OS}^2 (%)	p-value	Δ (ann %)
log(DP)	-4,67	0,92	-0,07
log(DY)	-46,40	0,99	-10,60
log(EP)	-28,18	0,98	-5,37
log(DE)	-0,81	0,81	0,00
SVAR	0,00	#NUM!	0,00
BM	-27,00	1,00	-4,34
NTIS	-0,99	0,51	-0,44
TBL	33,08	0,00	0,00
LTY	59,98	0,00	0,00
LTR	-6,24	0,98	-0,25
TMS	9,08	0,01	0,00
DFY	5,44	0,00	0,00
DFR	61,74	0,09	0,00
INFL	-43,74	1,00	-15,96
Mean	0,81	0,56	-2,64
Median	-0,90	0,81	-0,04
Trimmed mean	0,81	0,56	-2,64

Table 31 - Results under the C&T restrictions for the out-of-sample period 2021-2022, during economic expansion times

Below is a table summing up the six different models.

It is possible to see that the kitchen sink model and the diffusion index both have negative R^2. On the other hand, sum-of-the-parts scores the highest coefficient of determination, around 31.50%.

Regarding the p-value, all the models except diffusion index have a p-value below the threshold of 5%. All the deltas are equal to zero, except the diffusion index which has a slightly negative delta.

Predictor	R_{OS}^2 (%)	p-value	Δ (ann %)
Kitchen sink	-15,23	0,01	0,00
SIC	25,50	0,00	0,00
POOL-AVG	7,57	0,02	0,00
POOL-DMSPE	10,43	0,01	0,00
Diffusion index	-4,62	0,92	-0,06
Sum-of-the-parts	31,50	0,01	0,00

Table 32 - Summary of the six different models – C&T restrictions – Expansion – 2021 to 2022

8.7.3.2 Under the unrestricted restrictions

1) **Overall**

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2021 and 2022.

Regarding the coefficient of determination of each single predictors, it is possible to see that four indicators are negative, with LTY being the lowest at around 10%. INFL, SVAR and log(DY) are the highest numbers, around 12%.

Regarding the p-value of each single predictors, we can say that SVAR and LTR have a small value of 5%, while all the other indicators are in between 5% and 95%.

Finally, regarding the utility value of each predictor, we can say that the delta is negative for all the indicators having a negative R^2, except for TBL which has a negative R^2 but a positive delta.

Looking at the different combinations, the main characteristics that we can discuss are the results are positive, with the two indicators related to the mean having better results than the median.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	1,08	0,15	1,27
log(DY)	11,39	0,08	14,90
log(EP)	3,23	0,18	4,67
log(DE)	-0,41	0,78	-0,28
SVAR	10,25	0,04	10,30
BM	3,55	0,19	5,77
NTIS	4,89	0,06	4,28
TBL	-3,38	0,65	0,06
LTY	-9,61	0,71	-2,01
LTR	4,43	0,05	4,01
TMS	-1,29	0,81	-0,91
DFY	0,46	0,40	0,78
DFR	1,36	0,36	4,10
INFL	13,47	0,07	11,77
Mean	2,82	0,32	4,19
Median	2,30	0,18	4,06
Trimmed mean	2,82	0,32	4,19

Table 33 - Results under the unrestricted scenario for the out-of-sample period 2021- 2022

Below is a table summing up the six different models.

The coefficient of determination and the delta of sum-of-the-parts are negative. All the other models have positive numbers, especially the kitchen sink, which has a greatly positive R² and delta.

Focusing on the p-value, kitchen sink, SIC, POOL-AVG and the POOL-DMSPE have a p-value below 5%.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	18,13	0,02	24,23
SIC	10,25	0,04	10,30
POOL-AVG	3,93	0,01	3,37
POOL-DMSPE	4,06	0,02	3,47
Diffusion index	1,11	0,14	1,31
Sum-of-the-parts	-1,34	0,60	-4,96

Table 34 - Summary of the six models – Unrestricted – Overall – 2021 to 2022

2) During economic recession times

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2021 and 2022, during economic recession times.

Regarding the coefficient of determination of each single predictors, it is possible to see that seven indicators are negative, while five indicators are higher than 20%.

Regarding the p-value of each single predictor, we can say that seven are below the threshold of 5% and five are above 95%.

Finally, regarding the utility value of each predictor, we can say that all the deltas are negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are that R² is positive for all indicators. The delta, however, is negative or equal to zero for each combination.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-2,54	1,00	0,00
log(DY)	-44,09	0,9995	-8,39
log(EP)	-22,99	1,00	-1,74
log(DE)	-0,44	0,60	0,00
SVAR	35,81	0,05	0,00
BM	-28,70	0,999	-0,50
NTIS	20,86	0,002	0,00
TBL	32,52	0,001	0,00
LTY	52,94	0,001	0,00
LTR	-4,14	0,996	0,00
TMS	4,49	0,01	0,00
DFY	15,44	0,004	0,00
DFR	28,15	0,01	0,00
INFL	-31,15	0,84	-6,39
Mean	4,01	0,47	-1,21
Median	2,02	0,33	0,00
Trimmed mean	4,01	0,47	-1,21

Table 35 - Results under the unrestricted scenario for the out-of-sample 2021-2022, during economic recession times

Below is a table summing up the six different models.

Only the diffusion index has a negative R². Kitchen sink has the highest R², around 70.81%.

Regarding the p-value, kitchen sink, SIC, POOL-AVG, the POOL-DMSPE and sum-of-the-parts have a small p-value, below 5%. On the other hand, the diffusion index has a p-value above 95%.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	70,81	0,02	0,00
SIC	35,81	0,05	0,00
POOL-AVG	6,63	0,01	0,00
POOL-DMSPE	9,43	0,004	0,00
Diffusion index	-2,24	1,00	0,00
Sum-of-the-parts	15,18	0,008	0,00

Table 36 - Summary of the six different models – Unrestricted – Recession – 2021 to 2022

3) During economic expansion times

Below is a table summarising the results under the unrestricted scenario for the out-of-sample period between 2021 and 2022, during economic expansion times.

Regarding the coefficient of determination of each single predictors, it is possible to see that eight indicators are negative, while three indicators are higher than 30%.

Regarding the p-value of each single predictor, we can say that five are below the threshold of 5% and five are above 95%.

Finally, regarding the utility value of each predictor, we can say that all the deltas are negative or equal to zero.

Looking at the different combinations, the main characteristics that we can discuss are that R² is positive for the two indicators concerning the mean while the median is negative. The delta, however, is negative for each combination.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
log(DP)	-4,63	0,92	-0,07
log(DY)	-46,40	0,99	-10,60
log(EP)	-28,18	0,98	-5,37
log(DE)	-0,81	0,81	0,00
SVAR	25,50	0,00	0,00
BM	-27,00	1,00	-4,34
NTIS	-0,99	0,51	-0,44
TBL	33,08	0,00	0,00
LTY	59,98	0,00	0,00
LTR	-6,24	0,98	-0,25
TMS	9,08	0,01	0,00
DFY	5,44	0,00	0,00
DFR	61,74	0,09	0,00
INFL	-43,74	1,00	-15,96
Mean	2,63	0,52	-2,64
Median	-0,90	0,66	-0,04
Trimmed mean	2,63	0,52	-2,64

Table 37 - Results under the unrestricted scenario for the out-of-sample period 2021-2022, during economic expansion times

Below is a table summing up the six different models.

It is possible to see that the kitchen sink model and the diffusion index both have negative R^2. On the other hand, sum-of-the-parts scores the highest coefficient of determination, around 31.50%.

Regarding the p-value, all the models except diffusion index have a p-value below the threshold of 5%. All the deltas are equal to zero, except the diffusion index which has a slightly negative delta.

Predictor	R_{OS}^2 (%)	*p*-value	Δ (ann %)
Kitchen sink	-15,23	0,01	0,00
SIC	25,50	0,00	0,00
POOL-AVG	7,57	0,02	0,00
POOL-DMSPE	10,43	0,01	0,00
Diffusion index	-4,62	0,92	-0,06
Sum-of-the-parts	31,50	0,01	0,00

Table 38 - Summary of the six different models – Unrestricted – Expansion – 2021 to 2022

8.8 Discussion of findings

After having diligently presented the results in the previous section, let's dive deep in the findings that these numbers can give. This section will mainly focus on the R^2 results because they are the ones that are the most important regarding this study.

Before diving deep in the results, it is good to note that SVAR is always equal to zero under the Campbell and Thompson restrictions. This shows that the restrictions are working even though being consistently equal to zero, is not a normal occurrence. This could indicate that the code should be optimised somewhere. Moreover, the utility gain is not very reliable to some extent. Therefore, the findings are not really focused on the utility gain, but it is interesting to cross-check the results on the R^2 with the utility gain.

The results for two different out-of-sample periods, 2002 - 2022 and 2010 - 2022, are exactly the same. This is, most probably, not possible. Therefore, is it important to note that there must be a mistake within the code, resulting in the same results between two out-of-sample periods.

This being said, the results still give a good indication of the trend of the results. Indeed, the predictive ability of the variables are shown, even if not perfect, through the results obtained above. However, it has to be decided that these two periods are now considered as one single period for the continuation of the discussion of findings. To make it easier to follow, let's call this newly out-of-sample period: *"merged out of sample"*.

8.8.1 Regarding the individual predictors

As said above, let's just consider two out-of-sample periods. In each period, there are six different results regarding the economic times (overall, expansion or recession) and the restrictions (unrestricted and under the Campbell and Thompson restrictions). Let's have a look how the different individual indicators react according to the economic time.

For the merged out-of-sample, without making any distinction regarding the economic recession or expansion times, thus taking the full sample, there are eight negative R² under the unrestricted scenario, and only six negative values ranging from -9.17% to -0.1%. This means that, out of fourteen predictors and in the best case, 43% of the indicators are beaten by the historical average.

Regarding the economic expansion time, there are seven negative R² under both scenarios which means that 50% of the indicators are beaten by the historical average. However, the results are a bit more extreme here, with a R² ranging from -41.45% to 60.67%.

Regarding the economic recession time, nine indicators have a negative R² under the unrestricted scenario, while seven are negative under the Campbell and Thompson restrictions. The results are less extreme than for the expansion time, ranging from -20% to 20%, with a single indicator with -54%.

To summarise for the merged out-of-sample, if you want to use a single predictor, it is safer to use the general scenario under the Campbell and Thomspon restrictions.

Regarding the out-of-sample period 2021 - 2022 and taking the full sample, only four indicators are negative under both scenarios, with results ranging from -9.61% to 13.47%. This means that only four indicators are beaten by the historical average. This is the best results gotten so far.

Regarding the economic expansion time, there are seven negative indicators. Once again, the results are more volatile than for the full sample. Indeed, it ranges between -46.40% and 61.74%. Half of the indicators predict better than the historical average, while the other half have fewer great predictions.

Regarding the economic recession time, there are seven positive and seven negative indicators. They are a little bit less volatile than for the economic expansion time, ranging from -44.09% to 52.94%.

To summarise for the out-of-sample period from 2021 to 2022, if someone wants to use a single predictor, it is definitely safer to use the general scenario.

Concretely, it seems that the shorter the out-of-sample period, the better the individual predictors perform. Moreover, not a single predictor has a positive R² for all the different scenarios. This implies that the historical average, even if not perfect, is not consistently beaten by an individual variable. This is in line with RSZ and Goyal and Welch (2008).

8.8.2 About the combinations

As said above, let's just consider two out-of-sample periods. In each period, there are six different results regarding the economic times (overall, expansion or recession) and the restrictions (unrestricted and under the Campbell and Thompson restrictions). Let's have a look how the different combinations, mean, median and trimmed mean, react according to the economic time.

Looking at the merged out-of-sample and for the whole sample, the coefficient of determination is negative for the three combinations under the unrestricted scenario, while the R² is positive for the three combinations under the Campbell and Thompson scenario. For both cases, the predictions are near to the historical averages because they stand between -0.24% and 1.35%.

Regarding the economic expansion time, only the median is negative in both scenarios. However, under the unrestricted conditions, the negative result is less far from the historical average than under Campbell and Thompson, while the positive results are greater under the unrestricted scenario.

Looking now at the economic recession time, the three combinations have a negative R² under the unrestricted case, while only the median is negative under Campbell and Thompson. The mean and trimmed mean are only slightly positive, with a result of 0.16%.

Without regarding which combination is being used, the results under Campbell and Thompson for the whole out-of-sample period are convincing. However, they are not always convincing regarding all the other periods.

This part really highlights the fact that the unrestricted case is more volatile but predicts, statistically, more often better than under Campbell and Thompson.

Regarding the out-of-sample period 2021-2022 and focusing now on the whole data set, the coefficient of determination of the three combinations are positive, under both scenarios. They are better under the unrestricted scenario than under the Campbell and Thompson scenario.

Regarding the economic expansion, only the trimmed mean has a negative R² under both cases. However, the mean and trimmed mean are better under the unrestricted scenario than under the Campbell and Thompson restrictions.

Looking now at the recession time, all the indicators are positive under the unrestricted case, while only the median is negative under Campbell and Thompson. However, once again, the results are better under the unrestricted scenario.

There are few lessons to be learned from this section. Firstly, it seems that when the predictors do less great, they do better under Campbell and Thompson. But, on the other hand, when they do great, they do it better under the unrestricted scenario. This would mean that the Campbell and Thompson hypotheses would limit the damage but would have less upside potential when the indicators do well.

Secondly, as for the individual predictor section, the shorter the out-of-sample period, the greater the prediction. It is, most probably, because the model has been trained better with more information.

Thirdly, the combinations seem to predict, in general, better than the individual predictors, having a positive R² more often than the predictors. However, the results are less volatile. This implies that, for someone really risk averse, the potential upside is limited using the combinations.

8.8.3 About the models

As said above, let's just consider two out-of-sample periods. In each period, there are six different results regarding the economic times (overall, expansion or recession) and the restrictions (unrestricted and under the Campbell and Thompson restrictions). Let's have a look how the different models react according to the economic time.

Regarding the merged out-of-sample and focusing on the whole sample, two models have a negative coefficient of determination under the unrestricted case while only one has a negative R² under Campbell and Thompson. Except for the kitchen sink model, the positive results are quite not volatile, with the highest being 2.57%. The kitchen sink model has a R² of 20.98% and 18.54% in the scenarios. However, SIC has a negative coefficient of determination at around -9.17%, under the unrestricted case.

Focusing only on the economic expansion time, two indicators have negative results for the coefficient of determination. These are the kitchen sink model and the diffusion index models. The negative result of the kitchen sink is really extreme, with a value of -81.24% in both cases. The other models predict positive R^2.

Looking at the economic recession time, four models have negative results regarding the R^2, while only two have a positive coefficient of determination under both cases. The results are more volatile, with a R^2 ranging between -54.74% and 40.97%.

It is interesting to note that when the kitchen sink model predicts well, it predicts extremely well. However, when it predicts bad, it predicts extremely bad. Moreover, the sum-of-the-parts has a positive R^2 for each scenario, whatever the economic time.

Focusing on the out-of-sample period 2021 - 2022 and on the whole sample, all the indicators are positive, except for the sum-of-the-parts model which is slightly negative. The results are a bit more volatile under the unrestricted case, which gives a better upside, especially for the kitchen sink model.

Concerning only the economic expansion time, two indicators are negative, being kitchen sink and the diffusion index. The other one performs quite well, with the lowest positive R^2 being 7.57%.

Regarding the economic recession time, only one model performs poorly, with a R^2 being negative. Indeed, the diffusion index has a R^2 of -2.24%. On the other hand, the positive results are quite appealing. Indeed, the lowest one is 6.63% while the highest one is 70.81%.

We can summarise this section by saying that, firstly, overall, the models are not-too-bad predictors, especially when the out-of-sample period is short, except for the diffusion index which does overall poorly. Sum-of-the-parts does particularly well, with only being two times negative for with a value at around -1%. The kitchen sink model is really volatile. Indeed, it can predict very good or very bad and should be kept to the risk-averse investors.

Secondly, Campbell and Thompson hypotheses limit the damage when a model performs poorly but it also restricts the upside potential when the indicators do well.

8.8.4 Differences about the economic time

The economic time has been added to all the results to have a better overview of it. To limit the impact of the recessions over the predictors, the in-sample periods always end after a recession, in order to fully train the model.

On one hand, regarding the strategy of the investors, it can be interesting for someone to focus on indicators that perform well regarding his view on the market. For example, if someone is convinced that the market will be in recession next year, focusing on predictors that do great during recession times will be more useful. The kitchen sink is a good model for that. Indeed, it performs really well during recession time.

On the other hand, if an investor is convinced that the market will experience an expansion throughout the year, he should focus on a predictor that performs very well during expansion time. The SIC model is a notable example of it. A great counterexample is the kitchen sink model that performs really poorly during expansion time.

For someone more restrained, the strategy can be adapted with other predictors.

Finally, it has been designed to be useful to show the differences of the results for each economic period of time as well as for the whole sample in order for an investor to adapt as best as possible according to its view on the market and its risk-aversion.

8.8.5 Differences between the unrestricted and the under C&T restrictions scenario

Overall, the goal of testing different restrictions was to see if a scenario clearly outperforms. Even if we can say that the Campbell and Thompson restrictions are more conservative and, therefore, less risk averse, it is not possible to conclude that they are the perfect restrictions in order to consistently outperform the historical average.

However, the results shown clearly show that the R^2 tends to be less volatile under the Campbell and Thompson restrictions, which is important information for a risk-averse investor.

On the other hand, the unrestricted scenario has overall a better upside gain, for people ready to take more risks.

8.8.6 Summary of the main findings

Overall, this study has not found a predictor that consistently outperforms the historical average. However, multiple possibilities have been proposed in order to maximise the predictions according to the investor's view. Below can be found the main takeaways.

Firstly, the predictors do better if they are trained on a larger in-sample. This study was limited because the Nasdaq 100 was launched in 1985, but, in the future, time will pass, and predictions should become increasingly accurate.

Secondly, the Campbell and Thomspon restrictions can be valuable for non-risk-taking people. Indeed, most of the time, it limits the downside thanks to its more conservative approach. However, it also limits the upside, which is not suitable for risk-taking investors.

Having a look at the economic period of times can allow someone to choose a predictor that performs better during an expansion or recession time. However, it can be a difficult exercise because the future is, by definition, unpredictable. The kitchen sink model is a very good example of it: it performs greatly during recession but performs very poorly during expansion. Predicting the economic time could, then, be very useful but presents many risks and it should be carefully assessed if it is worth it.

Thirdly, the historical average forecast does not have to be thrown away. Indeed, it sometimes performs great and should be an indicator that any investor must keep in mind, mainly because it is easy to calculate.

Fourthly, the combinations and the models are, in general, safer than the individual predictors. Indeed, they have, statistically speaking, less often negative R^2. However, it really depends on multiple factors inherent to the market that could make this statement not always accurate.

Fifthly, the above results give a lot of different predictors possible, in different periods of time, with different restrictions possible. It is up to the investor to choose which one is the most suitable for him, according to his profile and risk-appetite.

Finally, no indicator is 100% convincing and consistently outperforms the historical average. Therefore, the claim made in the RSZ paper does not hold for the technology companies listed on the U.S. stock market. However, on the basis of the results obtained, it could be argued that the sum-of-the-parts models stand out in all the predictors tested above. Indeed, it has a very little downside when it predicts worse than the historical average but has a great upside when it does predict better.

8.9 Results comparison

Firstly, in RSZ[23], the authors highlight that: "numerous factors give rise to a highly uncertain, complex, and constantly evolving data-generating process for expected equity returns that is difficult to approximate with a single predictive regression model.". In other words, it is very difficult and very unlikely that a single predictor can generate reliable forecasts over time. Goyal and Welch (2008) states that the inconsistent performance of individual predictive regression models in out-of-sample tests is often attributed to structural instability.

Structural instability, in the context of economic and financial models, refers to a situation where the relationships between variables in a model change over time due to underlying changes in the economic or institutional environment. This can happen for several reasons such as:

- **Institutional and policy changes:** when there are changes in laws, regulations, or policies that impact economic activities, the previous relationships modelled between variables may no longer hold.

- **Technological advances:** innovations in technology can disrupt existing business models and economic relationships.

- **Economic shocks:** unexpected events like financial crises, the Covid-crisis, or geopolitical conflicts can cause sudden and profound changes in economic relationships and dynamics.

- **Changes in investor behaviour:** as investors learn from past experiences or adapt to new information and technologies, their behaviour might change, affecting the predictability of financial models based on historical data.

The implications for modelling and forecasting can be significant. That is why RSZ provides a solution which consists in combining multiple individual forecasts, which should enhance the certainty and the stability compared to the risk associated with a single model that only relies on one predictor.

Moreover, RSZ states that the equity premium forecast is more stable using their combinations rather than individual predictive regression models. This can be seen as a phenomenon of diversification which reduces the variance of the whole portfolio.

The above-mentioned statements clearly emphasise the results obtained in the previous sections. Indeed, it shows that the combinations tend to be less volatile and more reliable. It is also interesting to note that one of the main contributions of this research, which are the models, also prove that combining multiple variables allows the prediction to be, overall, more reliable for a well-defined model, such as the sum-of-the parts.

Secondly, RSZ suggests that combination forecasts are tied to the business-cycle fluctuations, which is also emphasised in this paper. Indeed, it is possible to see that, depending on the economic time, the predictive results are way different. Some models and individual predictors are more accurate during the expansion, while others are more accurate during the recession.

---

[23] List of abbreviations

To summarise, this comparison is very insightful as part of this thesis. Indeed, this allows us to conclude that our results are in the same vein as RSZ, albeit slightly different. The differences mainly come from the fact that this paper focuses on the technology stock of the companies listed in the United States and that the research has been done on a different time period. This is, therefore, normal to get different results. However, the main lesson to bear in mind after this section is that the results obtained above are plausible and are not fundamentally out of line with results already obtained in the scientific literature.

9. Step 4: Machine learning analysis using Google Colab and Python

9.1 Introduction to machine learning

The goal of the fourth step is to see if it is possible to optimise the combination of the variables given in the RSZ's paper. This optimisation could be done in multiple different ways, but it has been chosen to use machine learning to do it. Indeed, it was taught in the first year of our master, as part of the Business Analytics course, under Mr. Schyns, to use machine learning to optimise a model (Schyns, 2022).

According to Tiwari (2022), machine learning is a branch of artificial intelligence that equips systems with the ability to automatically make decisions based on previous experiences. This technology enables the recognition of images, analysis of speech or audio data, identification of hate speech in texts, and powers self-driving cars. Everyday applications of machine learning include the personalised recommendations you see on platforms like Netflix and Amazon, as well as virtual personal assistants such as Alexa and Siri.

More particularly, the supervised machine learning part is useful in this context. Timari (2022) defines it as "the most prevalent type of machine learning, which involves learning from examples". For instance, we can recognize a book because we were once taught that a collection of digitally typed papers, bound together with a hardcover at the top and bottom, constitutes a book. Similarly, supervised learning operates by examining the different characteristics of an object, alongside its corresponding label. These features are crucial for correctly identifying the object.

For example, if an item has all the features of a book but lacks digitally printed text or any text, is it still a book? Features are essential in accurately determining an object's classification. The more precise the features, the more accurate the predictions will be. Supervised learning employs specific algorithms to implement this concept in various ways. It includes two main techniques: classification, which assigns each observation to a predefined category, and regression, which is used for continuous outcomes like monthly salaries or daily sales totals.

To rephrase it, supervised machine learning can be relevant and useful for optimising a set of variables to predict the equity premium. That is why machine learning has been chosen as part of this paper. Indeed, in this context, supervised learning techniques can analyse historical financial data to identify patterns and relationships between variables and the equity premium.

9.2 Introduction to Python

For this part, the code responsible for the machine learning was made in Python. According to their own website (*What Is Python? Executive Summary*, n.d.), Python is: "an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python is simple, easy to learn syntax emphasises readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.".

In short, Python is a computer programming language used for many different tasks, including machine learning analysis. It is a general-purpose language, meaning that it is not specialised for any specific problems.

According to the website *machinelearningmodels (Nailman, 2023)*, Python has emerged as the preferred language for machine learning due to its multiple qualities in the programming sector. Here is a list of all the qualities presented by the website:

- Python has a simple and easy-to-understand syntax. This simplicity allows users to focus more on the algorithm itself, which is ideal for beginners who are starting. Moreover, the simplicity of the syntax makes it highly flexible, which is nice to maintain and modify the code.
- Python has plenty of machine learning libraries. Its wide range of libraries and frameworks for machine learning provide developers with powerful tools and possibilities to simplify the code using machine learning techniques.
- Python has a wide range of documentation and resources for machine learning such as tutorials or examples tailored to machine learning, which is very useful for beginners in the field.
- Python is platform-independent and can be run pretty easily on different systems such as Windows, macOS and Linux.

Lastly, we learned how to deal with Python in the Business Analytics lesson from Mr. Schyns in Master 1 (Schyns, 2022). Therefore, all in all, it has been decided to perform some machine learning using the programming language Python.

9.3 The choice of using Google Colab

Colaboratory: "allows you to write and execute Python in your browser, with zero configuration required, access to GPUs free of charge and easy sharing" (*Google Colab*, n.d.). Being a student without any profound knowledge of the technical configuration and working on multiple devices, the choice to use Google Colab was quite simple. Indeed, Colaboratory removes the necessity for intricate configuration setups and installations since it operates directly within the browser.

Moreover, it comes equipped with pre-installed Python libraries, eliminating the need for any setup procedures before utilisation.

Finally, we already worked with it last year with Mr. Schyns in the Business Analytics course (Schyns 2022). It has therefore been evaluated as the best tool to use.

9.4 Explanation of the methodology

As said previously, the goal of the Python analysis is to find a new set of predictors that could better predict the equity premium than the set proposed by RSZ. To do so, a machine learning code has been prepared in Google Colab. Here is a more detailed explanation:

The first step is to prepare the analysis. To do so, the dataset is stored on Google drive under the name of 'machinelearning.csv' and is imported within Colaboratory.

After a few lines that ensure the variables are all imported and that the equity premium, which is the target variable Y.

It is important to note that, from this line, the variables 'Risk-free rate' and 'Nasdaq 100 value-weighted return with dividends' have been deleted from the upcoming calculations. Indeed, as they are already represented in the Y, they will be kept no matter the results of the machine learning.

After a few non-necessary steps calculating the covariance and the multicollinearity, a step which computes the backward elimination is done. The significance level is put at 5% and it should give a number of variables that should optimise the prediction.

The second step is to create the out-of-sample period, and re-do a backward elimination which should, once again, give a set of optimised variables. At the end of the second step, the mean squared error and the R^2 error are computed in order to compare them with other models later.

The third step named "cross-validation" is done. This step assesses the model's performance across different subsets of the data. The goal of this part is to ensure the model performs consistently, whether it indicates good generalizability and robustness of the model's predictive ability. At this stage, all the relevant wanted information is known. However, the analysis does several other steps in order to ensure the accuracy of the results and the model. These steps are not necessary, but they ensure robustness and consistency in the results found.

The fourth step is about residual analysis. This is done to check if there are any patterns or systematic deviation from randomness, which could indicate potential issues with model assumptions such as homoscedasticity.

The fifth step is made to conduct more rigorous assessments of the model assumptions. Two tests have been chosen for it: the Breusch Pagan and the Shapiro-Wilk test. The Breusch Pagan's purpose is to determine if, in a regression model, we can find heteroscedasticity or not (Bobbitt, 2022). The Shapiro-Wilk test, for its part, is used to evaluate if the data are normally distributed. Normality does not imply that there is a standardised normal distribution, but it does mean that the sample has been generated from a Gaussian distribution (Malato, 2023).

The sixth step is testing the robustness of the model, using the Huber regressor. According to Huber (1964), this is a regression technique designed to be less sensitive to outliers because, in simple words, depending on the size of the errors, the distribution of these errors can look different. For smaller errors, it appears similar to the common normal distribution. But for larger errors, it changes to a Laplace distribution, indicating a higher chance of extreme error values. This dual behaviour is being captured or explained by looking at the loss as if it were representing these error distributions through the log-likelihood function.

The seventh step is about a random forest regressor which basically computes the predictive performance. *Towardsdatascience* explains that the Decision Tree algorithm is straightforward and easily interpretable, which sometimes limits its ability to fully capture complex patterns in data. In contrast, the Random Forest algorithm builds on this by aggregating multiple Decision Trees to enhance predictive accuracy and robustness. This ensemble approach, aptly named "Random Forest," leverages the collective strength of several trees, making it more powerful than a single Decision Tree. The "Random" in Random Forest refers to the randomness introduced in the creation of individual trees, ensuring that the ensemble does not just replicate the biases or errors of a single tree (K, 2021).

The eighth and ultimate step is about comparing the optimised model against the full model. Indeed, few variables have been selected as a result of this model in order to better predict the equity premium. Before testing it further, a first comparison is done in order to see if it reacts better than the full model.

This code has been replicated and adapted a few times in order to test it within multiple different out-of-sample periods. Precisely, it has been reproduced five times:

- One with the full sample cut in half. 50% of the sample was used to train the model, while 50% of the sample was actually kept for the out-of-sample prediction. This file is named: "50-50_fin.ipynb"

- One where 60% of the sample was used to train the model, while 40% of the sample was actually kept for the out-of-sample prediction. This file is named: "60-40_fin.ipynb"

- One where 70% of the sample was used to train the model, while 30% of the sample was actually kept for the out-of-sample prediction. This file is named: "70-30_fin.ipynb"

- One where 80% of the sample was used to train the model, while 20% of the sample was actually kept for the out-of-sample prediction. This file is named: "80-20_fin.ipynb"

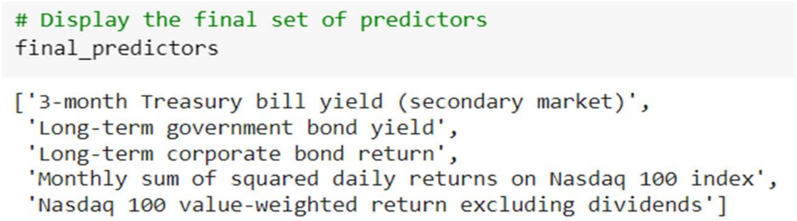
- One where 90% of the sample was used to train the model, while 10% of the sample was actually kept for the out-of-sample prediction. This file is named: "90-10_fin.ipynb"

The goal of testing the model with different out-of-sample, even if the cross-validation was already made, is to check whether new results and new combinations of predictors appear in order to not exclude any possible scenario.

9.5 Presentation of the results

Once the code has been created and running, some results appear. Here is a presentation of the main takeaways. Overall, the intermediate outcomes of the computations are not presented because they only assess the reliability of the results and do not represent the results themselves.

Firstly, let's have a look at the five chosen variables from the 70-30_fin file:



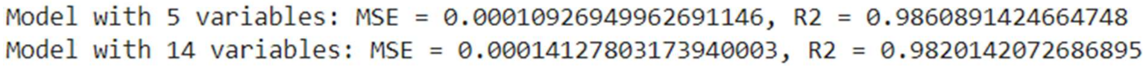
```
# Display the final set of predictors
final_predictors
```

```
['3-month Treasury bill yield (secondary market)',
 'Long-term government bond yield',
 'Long-term corporate bond return',
 'Monthly sum of squared daily returns on Nasdaq 100 index',
 'Nasdaq 100 value-weighted return excluding dividends']
```

- 3-month Treasury bill yield
- Long-term government bond yield
- Long-term corporate bond return
- Monthly sum of squared daily returns on Nasdaq 100 index (SVAR)
- Nasdaq 100 value-weighted return excluding dividends

Those variables are supposed to be the one that optimises the predictive model. To ensure the accuracy of the findings, the code computes the mean squared error (MSE) and the R² of the model for five different out-of-sample periods, as explained above.

Here is an example from the 70-30_fin file:



```
Model with 5 variables: MSE = 0.00010926949962691146, R2 = 0.9860891424664748
Model with 14 variables: MSE = 0.00014127803173940003, R2 = 0.9820142072686895
```

On one hand, the smallest Mean Squared Error, as explained on the website *Stephen Allwright* (2022), is the most optimal situation. Indeed, a small MSE is considered optimal in many statistical and machine learning contexts because it indicates a close match between predicted values and actual values, reflecting a high level of model accuracy. As the goal is to minimise error and increase predictive accuracy, a small MSE serves as a strong indicator that the model is achieving these goals effectively.

On the other hand, as explained on the website *Stephen Allwrigh*t (2022b), when comparing two models, the higher R^2 value is generally considered better because it indicates that a greater proportion of variance in the dependent variable is explained by the independent variables in the model. A higher R^2 value means that more of the variance is captured by the model, suggesting it has a stronger explanatory power, which typically indicates that the model fits the data better.

Let’s now have a look at the results gotten from the five different out-of-sample periods. The results are not always the same due to the fact the out-of-sample periods are different but the R^2 and the MSE are always better with the limited set of variables model than the results with all the variables. To recall, a smaller MSE and a bigger R^2 than the one with the full set is more optimised.

As said previously, each file provides a different set of variables that should optimise the predictive model. Here is a list of the variables that have been selected by the model on Python:

- The variable “**Long-term corporate bond return**” appears in each of the five different out–of-sample periods.

- The variable “**Monthly sum of squared daily returns on Nasdaq 100 index**” appears in each of the five different out–of-sample periods.

- The variable “**Nasdaq 100 value-weighted return excluding dividends**” appears in each of the five different out–of-sample periods.

- The variable “**3-month Treasury bill yield (secondary market)**” appears in the four different out–of-sample periods. It only does not appear in the 50-50_fin machine learning computations.

- The variable “**Long-term government bond return**” appears in two different out–of-sample periods. It only appears in the 50-50_fin and 60-40_fin machine learning computations.

- The variable “**DJIA book-to-market value ratio**” appears in two different out–of-sample periods. It only appears in the 80-20_fin and 90-10_fin machine learning computations.

- The variable “**Long-term government bond yield**” appears in one different out–of-sample period. It only appears in the 70-30_fin machine learning computation.

9.6 Discussion of findings

Here are how the selected variables might impact the model's predictions of the equity premium and assess the model's overall performance based on the evaluation metrics, R^2 and MSE.

- **Long-term corporate bond return:** this variable helps the model understand the credit market conditions. Since corporate bonds and equities often are affected by similar economic factors but differ in their risk and return profiles, insights from the corporate bond market can enhance the model’s ability to forecast equity returns.

- **Monthly sum of squared daily returns on Nasdaq 100 index (SVAR):** the inclusion of this measure of volatility is particularly insightful because it quantifies market risk. Higher volatility often correlates with higher risk premiums. The model’s ability to integrate this variable can lead to more accurate predictions during turbulent market periods.

- **Nasdaq 100 value-weighted return excluding dividends:** this performance metric directly relates to the equity market's overall health and trends, excluding dividend distributions. By focusing on pure price returns, the model better isolates the factors that drive market price movements.

- **3-month Treasury bill yield:** changes in the Treasury bill yield are likely indicators of changes in the economic climate, such as central bank policy shifts, which directly affect equity returns. A model that can accurately capture this relationship will be better positioned to predict the equity premium.

- **DJIA book-to-market value ratio:** the book-to-market value ratio is a valuation metric that compares the book value of a company to its market value. A higher ratio often indicates that the stock may be priced lower than its actual value, which could signal potential for returns. For the Dow Jones Industrial Average (DJIA), this ratio provides a macroeconomic perspective on the valuation of large-cap U.S. equities. This could be useful in models whose aim is to capture cyclicality in equity returns.

- **Long-term government bond return:** the returns on long-term government bonds provide a benchmark for the risk-free rate over a longer horizon and reflect broader economic conditions, such as inflation expectations and the general investment climate. Since these bonds are considered low risk, their returns are often used as a safe comparison against more volatile equity investments. When bond returns are high, equities may seem less attractive unless they offer a significantly higher potential return to compensate for the increased risk.

- **Long-term government bond yield:** by including the yield on long-term government bonds, the model gains insights into the long-term interest rate expectations and overall economic sentiment. Since these yields often move with expectations of economic growth and inflation, they provide predictive power over stock returns and premiums.

These five variables should be the one that, put together, will optimise the combination of variables presented by RSZ.

Furthermore, it is good to note that there is a need to keep the variables ‘Risk-free rate’ and ‘Nasdaq 100 value-weighted return with dividends’ because they are the one that computes the equity premium, Y.

To summarise the results found in this section,

- If the goal is to forecast well the equity premium with a quite short training period, with the training period being respectively 60% or 50% of the whole examined period, one should focus on the variables: “**long-term corporate bond return**”, “**monthly sum of squared daily returns on Nasdaq 100 index**”, “**Nasdaq 100 value-weighted return excluding dividends**”, “**Long-term government bond return**” and eventually “**3-month Treasury bill yield (secondary market)**”. A model based only on these variables should bring a better forecast of the equity premium than a model trained on the fourteen variables.

- If the goal is to forecast well the equity premium with a training period being respectively 70% of the whole examined period, one should focus on the variables: “**long-term corporate bond return**”, “**monthly sum of squared daily returns on Nasdaq 100 index**”, “**Nasdaq 100 value-weighted return excluding dividends**”, “**Long-term government bond yield**” and “**3-month Treasury bill yield (secondary market)**”. A model based only on these variables should bring a better forecast of the equity premium than a model trained on the fourteen variables.

- If the goal is to forecast well the equity premium with a quite short out-of-sample period, with the training period being respectively 80% or 90% of the whole examined period, one should focus on the variables: “**long-term corporate bond return**”, “**monthly sum of squared daily returns on Nasdaq 100 index**”, “**Nasdaq 100 value-weighted return excluding dividends**”, “**DJIA book-to-market value ratio**” and “**3-month Treasury bill yield (secondary market)**”. A model based only on these variables should bring a better forecast of the equity premium than a model trained on the fourteen variables.

Therefore, an investor should decide between these variables according to his own goal and his short or long-term vision. Being able to adapt the calculation according to the goal of the investor is important in order to be as precise as possible.

10. Newly published article

10.1 Introduction

Amit Goyal, Ivo Welch and Athanasse Zafirov have published, in September 2023, a paper named "A comprehensive 2022 look at the empirical performance of equity premium prediction (Goyal et al., 2021). The goal of their paper is to criticise plenty of papers published in top finance journals. Indeed, since Goyal and Welch (2008), plenty of papers have proposed or even created their own variables with a confidence that academic finance has completed the problem of predicting equity premia.

As said previously in the introduction section, the last twenty-five years have provided complex yet fascinating opportunities for each newly proposed variable to test whether they have predictability characteristics or not. Overall, the paper examines twenty-nine variables proposed in twenty-six papers whereas the data ended between 2000 and 2017. Goyal, Welch and Zafirov had added ten years of extra data, ending in December 2021.

The mentioned paper is really interesting. Even if it focuses on the full market with the S&P 500 index and does not only focus on one geographical area, it, however, gives precious indications to compare the results with this research. Let's have a look at the main findings.

10.2 Main findings

Once again speaking about the paper published by Amit Goyal, Ivo Welch and Athanasse Zafirov, they found out that twenty variables out of twenty-nine variables deteriorated with data extended to 2021. Thirteen deteriorated so much that they do not have good in-sample performance.

They then had a look at what they call the "homologous specifications", where variables were required to predict the log-equity premium at their native frequencies without overlapping; three additional variables demonstrated a decline in their in-sample effectiveness, resulting in a total of ten remaining variables. Indeed, there were only thirteen variables left as the authors counted the 6-month delayed-release versions.

Of these ten variables, four also exhibited robust performance in out-of-sample tests, supporting the principle outlined and explained above by Campbell and Thompson in 2008 that predictions of equity premiums should never be negative. Those are, as they call it in their paper:

- **Tchi:** the 14 technical indicators in Neely, Rapach, Tu, and Zhou (2014)
- **Shtint:** the short-stock interest holdings in Rapach, Ringgenberg, and Zhou (2016)
- **Accrul:** aggregate accruals in Hirshleifer, Hou, and Teoh (2009)
- **Gpce:** fourth-quarter growth in personal consumption expenditures in Møller and Rangvid (2015).

In addition, three more variables from the pre-Goyal and Welch (2008) set also performed well:

- **Tby:** the Treasury bill rate (Campbell (1987))
- **I/k:** the investment-capital ratio (Cochrane (1991))
- **Eqis**: equity issuing activity (Baker and Wurgler (2000))

The authors go further saying that Campbell and Thompson (2008) provide reasons why a researcher might prefer to focus on in-sample prediction over the longest interval, rather than on in-sample subsets or both in-sample and out-of-sample prediction, as their study does. Their thinking typically applies when researchers are confident that their model is stable and well-understood over time. They argue that under the assumption that the model is accurate, tests can be designed to leverage this assumption, a concept that could be seen as a reflection of a strong theoretical prior. In such scenarios, tests for out-of-sample prediction have less sensitivity than a comprehensive in-sample test.

In contrast, their assumptions, and possibly those of many, though not all, are less certain. Martin and Nagel (2019) also advocate for out-of-samples tests, pointing out how non-linear dynamics can lead to temporarily misleading predictive in-sample relationships.

10.3 Comparison to the above-mentioned findings

These results, even if not directly linked to the results found in this research, are very insightful. Indeed, there are a number of interesting lessons to be learned from it.

Firstly, Amit Goyal, Ivo Welch and Athanasse Zafirov, prove that even results published in top finance journals and accepted by the scientific community can turn out to be irrelevant when compared and tested to another dataset than the one used in their scientific background paper (Goyal et al., 2021). This emphasises strongly with the results presented at the point 8.7 which differ from the results gotten by RSZ. However, it can also be seen the other way round in the sense that there is a possibility that the results found by RSZ are not relevant with the periods chosen in this research but might be relevant with other periods.

Secondly, one variable that stands out is the three-month Treasury bill yield. The *New-York Times* says that the Treasury market offers predictions for what will happen in the economy. For example, when the three-month bills have yielded higher than the ten-year note, it shows that investors are more worried in the short-term than the long-term regarding the economy, which is a quite good indicator of recession (Simonetti, 2022).

Moreover, *Investopedia* says wisely: “Treasury yields also show how investors assess the economy's prospects. The higher the yields on long-term U.S. Treasuries, the more confidence investors have in the economic outlook.” (Chen, 2024). These two statements are in agreement with what the computations made for this thesis highlight. Therefore, it can be said that the three-month Treasury bill yield should have good predictability abilities.

Finally, the results presented by Amit Goyal, Ivo Welch and Athanasse Zafirov (Goyal et al., 2021) show the importance of this paper. They clearly demonstrate the need for further research into equity premium prediction and the need to re-evaluate previous scientific research. Indeed, some results are only valid in their initial data set, and this is exactly why the study, which is now coming to an end, was carried out.

11. Conclusions

11.1 Summary of the findings sur R et sur le machine learning

This study explores the predictive models using RStudio to forecast equity premiums, focusing on a range of predictive models and their efficacy across different economic conditions. The main purpose of this study is to assess whether the statement and the findings of RSZ were still viable and to try to find a more efficient set of variables that could optimise the prediction of the equity premium.

The research is primarily done around key economic predictors such as the three-month Treasury bill yield, long-term government bond yields, long-term corporate bond returns, and the volatility captured through the monthly sum of squared daily returns on the Nasdaq 100 index (SVAR), ...

The research used rigorous metrics like R^2, utility gains and Mean Squared Error to evaluate model performance. A key finding was that models incorporating a combination of predictors such as mean, median, and trimmed mean generally outperformed those relying on single predictors, particularly in out-of-sample validation. These combination forecasts not only showed improved accuracy but also exhibited enhanced stability across varying economic conditions, highlighting their effectiveness in dealing with economic uncertainties. Another finding was the different models proposed at the point 8.6. which did great, especially for the sum-of-the-parts model.

Combining different economic predictors proved beneficial in reducing forecast variance and adapting to economic shifts, such as those between expansion and recession phases. This methodological approach highlighted the diversification benefit inherent in multivariable models, which tend to be more reliable than models based on single predictors. Indeed, the study emphasised that the effectiveness of predictive models varies significantly with economic conditions. Models that perform well during economic expansions may not necessarily do great during recessions. This variability shows the importance of developing adaptive models capable of adjusting to cyclical economic factors to maintain accuracy and reliability.

One of the major reasons identified was the structural instability in economic relationships, often altered by technological advancements and significant policy changes. These dynamics necessitate continuous updates and validation of model assumptions to ensure their applicability and accuracy over time.

Afterwards, a machine learning analysis was done in order to assess if it was possible to optimise the set of variables proposed by RSZ. The analysis concluded that it was indeed possible proposing different variables regarding the duration of the out-of-sample period.

However, this study is not without limitations. The generalisability of the findings is constrained by the data period and the specific economic conditions under which the models were tested. The models' performance during different economic conditions highlighted the need for adaptive strategies that consider cyclical economic factors.

Future research should therefore focus on expanding the datasets, exploring the integration of additional predictive indicators, and refining the combination methodologies to enhance the robustness and accuracy of the forecasts. It would also be beneficial to apply these models across various economic sectors to verify their effectiveness in diverse contexts.

Overall, the research confirms that while no single model uniformly outperforms others across all conditions, the strategic combination of various predictors can yield a more dependable forecasting tool. This conclusion supports the ongoing refinement and development of predictive models in financial economics, emphasising a tailored approach that considers specific market conditions and investor needs.

11.2 Implication and significance

The study conducted has several profound implications, impacting both theoretical and practical aspects of financial economics.

Firstly, the research demonstrates that combining multiple predictive models not only improves the accuracy of forecasts but also enhances their stability across various economic conditions. This is crucial for investment strategy development and portfolio management, where precision in forecasting equity returns is essential for effective risk management and asset allocation.

Secondly, one of the pivotal implications of this study is the emphasis on the necessity for adaptive models that can dynamically adjust to fluctuating economic cycles. Financial institutions, investors, and policymakers can benefit from robust forecasting tools that offer real-time responsiveness to economic shocks and market volatility, thereby facilitating more informed decision-making processes.

Thirdly, integrating predictors such as market volatility and long-term bond yields into forecasting models provides deeper insights into the risk factors influencing equity premiums. This enhanced understanding can change risk assessment practices, allowing for a more nuanced management of investment portfolio risks.

Fourthly, the implications of this study also extend to economic policymaking. Insights derived from predictive models can be used in developing policies aimed at stabilising financial markets. Analysing how macroeconomic indicators influence equity premiums offers valuable feedback on the effectiveness and impact of fiscal and monetary policies.

Fifthly, more reliable forecasting tools bolster investor confidence by providing clearer insights into market dynamics. This can encourage greater participation in equity markets, especially in environments characterised by volatility or economic uncertainty.

Finally, this study contributes to methodological innovation in financial modelling. It highlights the effectiveness of hybrid statistical models that combine various predictors to enhance predictive accuracy. This opens doors for further research into more sophisticated models that integrate diverse datasets and computational techniques.

In conclusion, this research not only advances the understanding of how to predict equity premiums but also catalyses further innovation in financial modelling. The implications are broad, affecting everything from strategic planning and policy formulation to investment decisions and risk management.

11.3 Limitations

This study presents several insightful findings, yet it is not without limitations that could impact the broader applicability and reliability of the results. Here is an articulated discussion on these limitations.

Firstly, a significant challenge in applying machine learning in financial forecasting is the risk of underfitting or overfitting, where models are finely tuned to perform well on training datasets but fail to generalise to new, unseen data. This phenomenon can lead to overly pessimistic or optimistic performance evaluations and might not accurately reflect the model's true predictive power in real-world scenarios.

Secondly, the accuracy and robustness of predictive models are heavily contingent upon the quality and completeness of the data used. Limitations in data availability, particularly for less transparent economic indicators or specific geographic markets, can impede the model's ability to accurately capture complex market dynamics. A good example is E12 that needed to be approximated because some values were missing.

Thirdly, the models rely on certain economic and market assumptions that may not remain valid over different periods or under different economic conditions. Changes in regulatory policies, shifts in economic climates, or structural changes in financial markets can alter fundamental relationships between variables, potentially diminishing the effectiveness of the models.

Fourthly, advanced machine learning models often suffer from high complexity, making them difficult to interpret. This lack of transparency limits the ability to fully understand the decision-making processes within the models, which is crucial for gaining trust and ensuring accountability in financial decision-making.

Fifthly, predictive models may not sufficiently account for sudden market shocks or so called “black swan”, as these events are often underrepresented in historical datasets. This can lead to underestimations of potential risks and associated losses, especially during highly volatile market conditions.

Sixthly, the R code was a replication of a code written by Zhuo in 2008. This lacks flexibility to be adapted to other datasets. This rigidity could introduce errors and mistakes within the results and also difficulties to expand the research to other variables.

Lastly, the findings and the performance of the models might not be generalizable across different sectors or geographical regions due to varying economic conditions, regulatory environments, and market behaviours. This limitation restricts the models' applicability beyond the specific contexts studied.

By recognizing and addressing these limitations, future research can aim to refine these models, making them more robust, transparent, and applicable across a broader range of scenarios in financial economics. This would not only enhance the predictive accuracy but also enhance the confidence of users in deploying these models for practical financial forecasting and decision-making.

11.4 Future research

Building on the insights garnered from the current study, there are different possible research for advancing this research area.

Firstly, expanding the predictor set to include non-traditional variables such as geopolitical events, sustainability metrics, and consumer sentiment indices could provide deeper insights into market dynamics. Research into these variables might reveal new causal relationships and correlations that could significantly enhance the predictive accuracy of equity premium models.

Secondly, applying the forecasting models to various industries and different geographical regions could validate the effectiveness of these models more broadly. This would help to understand how specific economic policies and sectoral shifts in different regions affect equity premiums, offering a more granulated perspective on global financial markets.

Thirdly, employing sophisticated algorithms such as deep learning, neural networks, and ensemble methods could potentially improve predictive performance. These advanced models are capable of detecting complex, non-linear patterns in large datasets, which could lead to more accurate forecasts.

Fourthly, developing models able to process real-time data to forecast equity premiums would be an innovative step forward. This approach would provide insights into the practical applicability and resilience of these models in a dynamic, real-time market environment.

Fifthly, integrating behavioural finance aspects to account for investor sentiment and psychological factors could augment the models' comprehensiveness. Exploring how emotional responses and irrational behaviours influence market movements could open up new predictive pathways.

Sixthly, double checking the new wart of variables proposed by Python in RStudio could bring a nice overview of the actual feasibility and improvement of the combinations and models.

Finally, with continual technological evolution impacting financial markets, studying the effects of innovations such as blockchain, high-frequency trading, and AI-driven investment strategies on market predictions could provide pivotal findings.

Each of these proposed research directions not only builds upon the current study's findings but also paves the way for significant advancements in the fields of financial modelling and economic forecasting. By exploring these options, future research could lead to more refined, reliable, and comprehensive tools for predicting market behaviours and investment decisions.

11.5 Overall conclusion

This study delves deeply into the capabilities of predictive models using RStudio to forecast equity premiums, with a focus on evaluating the viability of the RSZ model and exploring an optimised set of predictive variables. The study primarily investigates economic predictors such as Treasury bill yields, long-term bond yields, and volatility indices, using robust metrics like R^2 to assess model performance.

A significant finding of this research is the superior performance of combination forecasts and models, which integrate multiple predictors. These indicators not only offer enhanced accuracy but also demonstrate greater stability across various economic conditions, thus proving effective in navigating economic uncertainties. Such an approach underscores the value of diversified predictors in predictive models, which consistently outperform single-variable models, especially in dynamic economic scenarios. However, although the models and combination provide better prediction than the individual predictors for predicting equity premium, they do not consistently outperform the forecast using the historical average of equity premium.

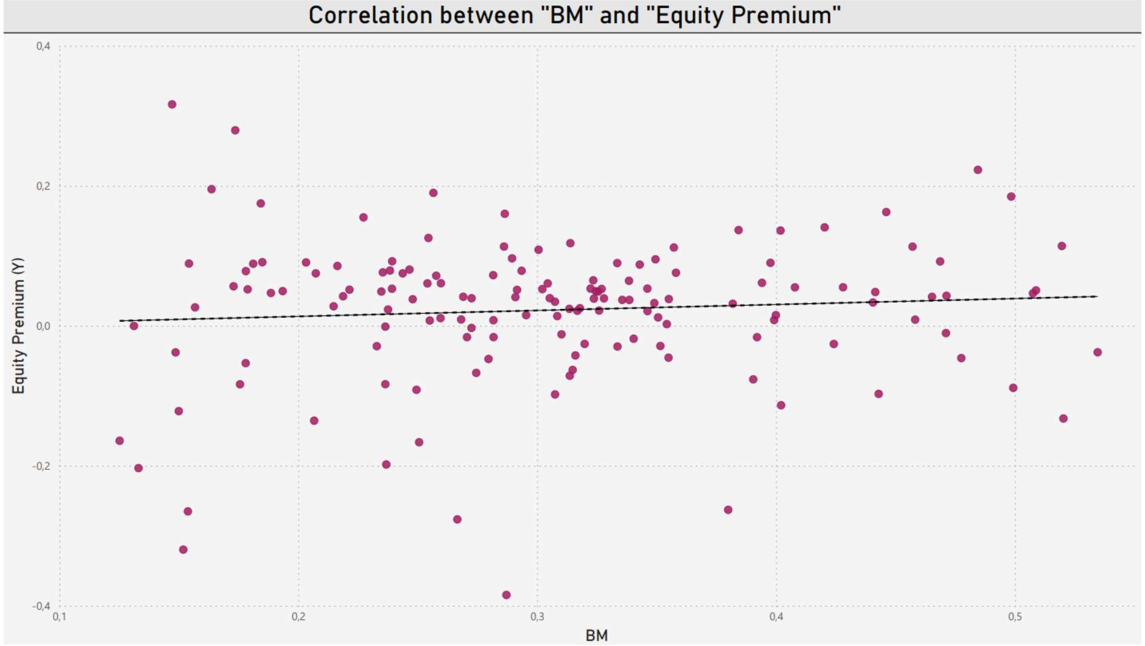
Moreover, the study explores machine learning techniques to further refine the predictive power of these models, leading to the identification of five critical variables that enhance forecasting accuracy. This methodological innovation introduces a new dimension to financial forecasting by integrating traditional economic indicators with advanced computational techniques.

However, the study is not deprived of limitations, primarily related to the specificity of the economic conditions and data periods tested. These constraints underscore the necessity for models that adapt to economic cycles and further suggest the expansion of datasets and predictive indicators in future research.

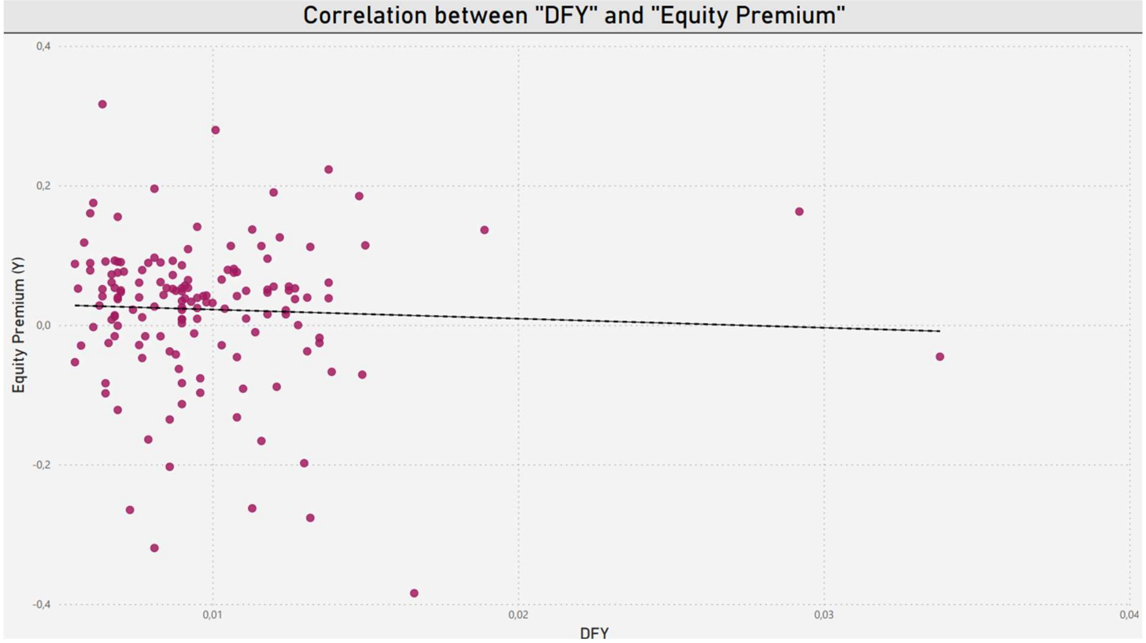
In conclusion, while no single model universally excels, the strategic integration of various predictors provides a more reliable forecasting framework. This study not only enriches the academic discourse on financial modelling but also offers practical insights for investors and policymakers, enhancing decision-making processes in financial markets. Future research should continue to refine these models, expand their applicability, and incorporate emerging technological trends to stay relevant in the ever-evolving financial landscape.

12. Appendices

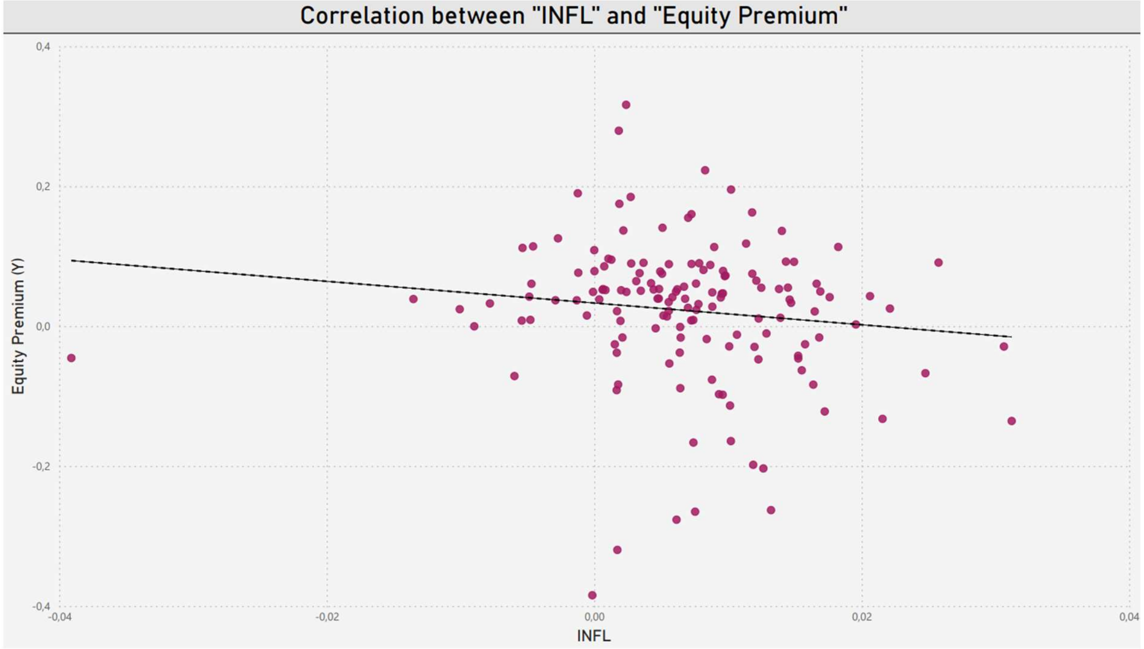
1) Appendix 1:



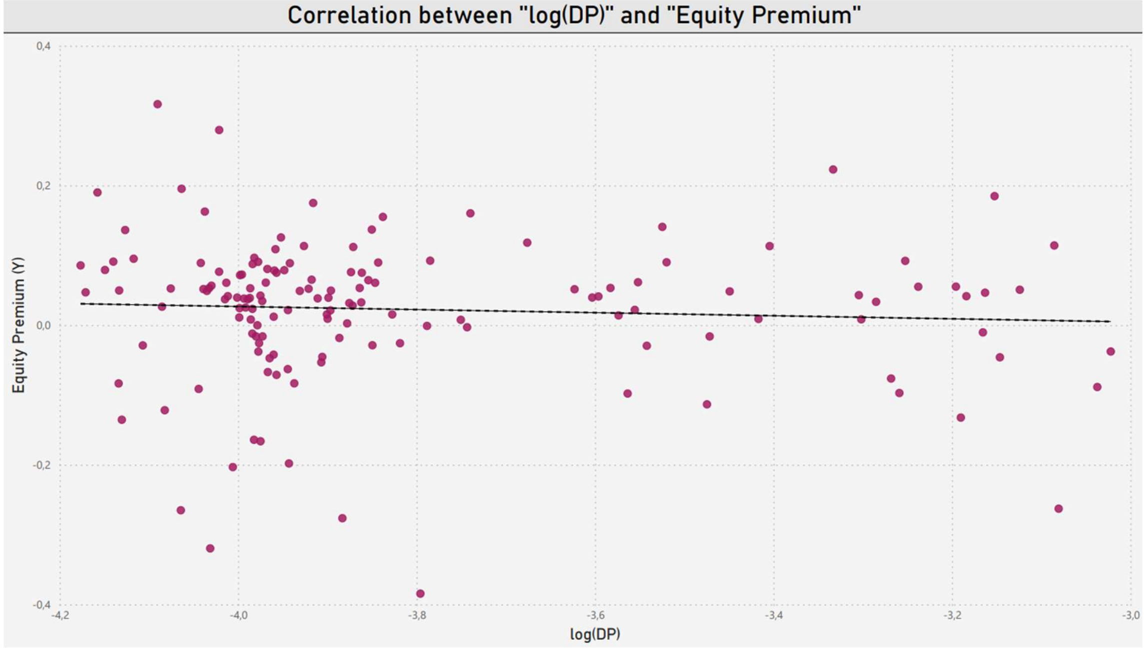
2) Appendix 2:



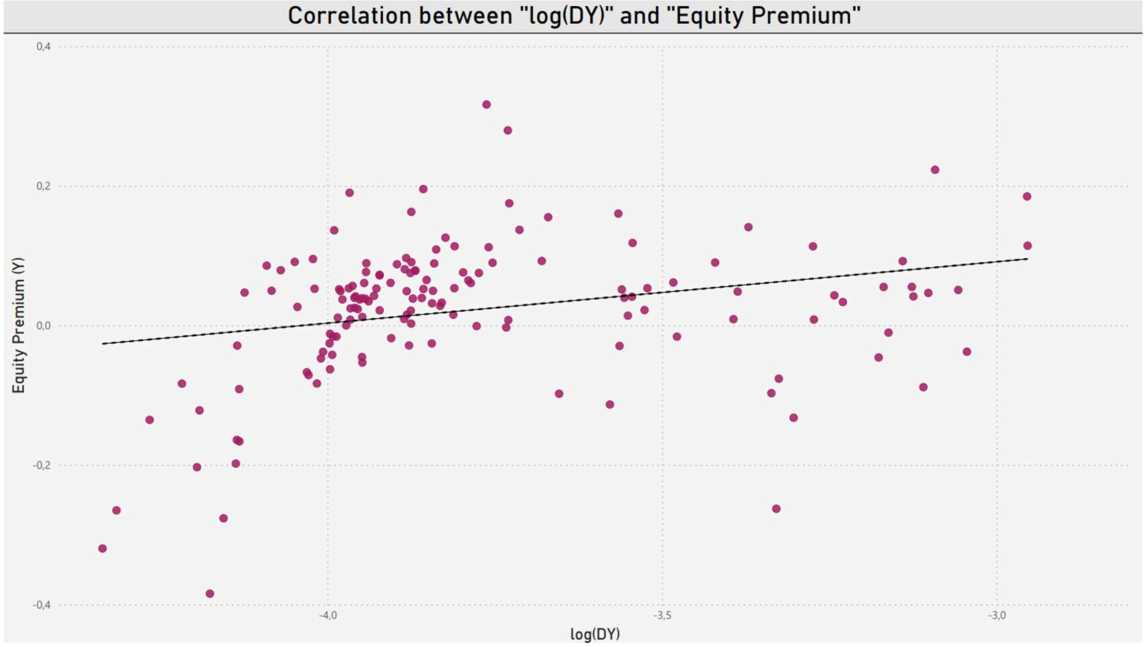
3) Appendix 3:



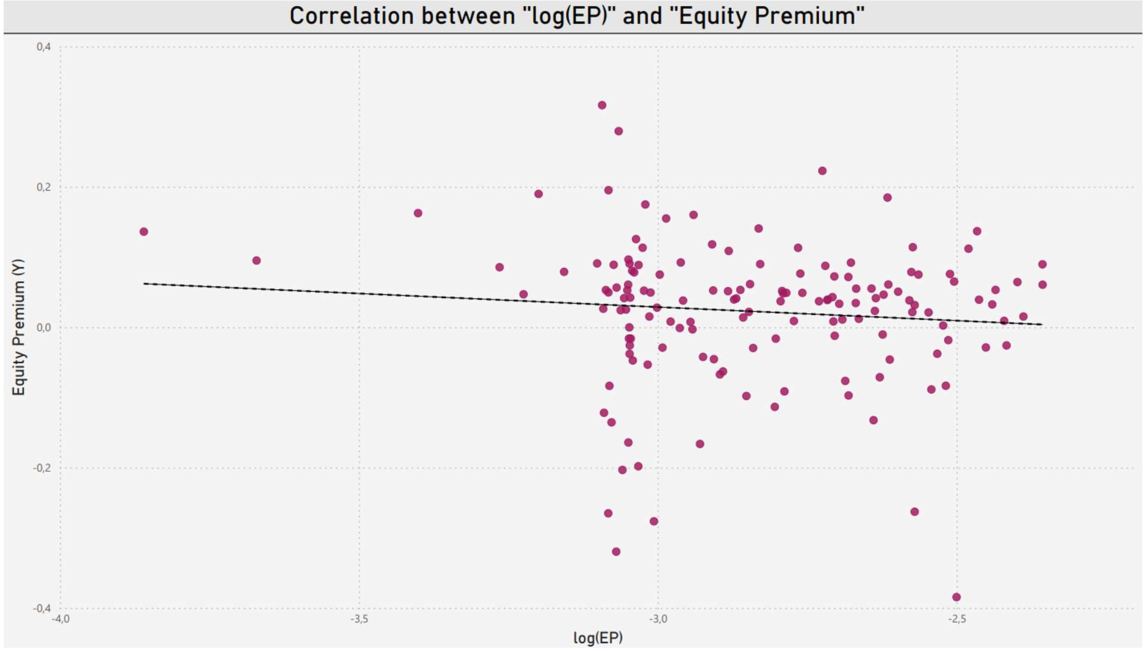
4) Appendix 4:



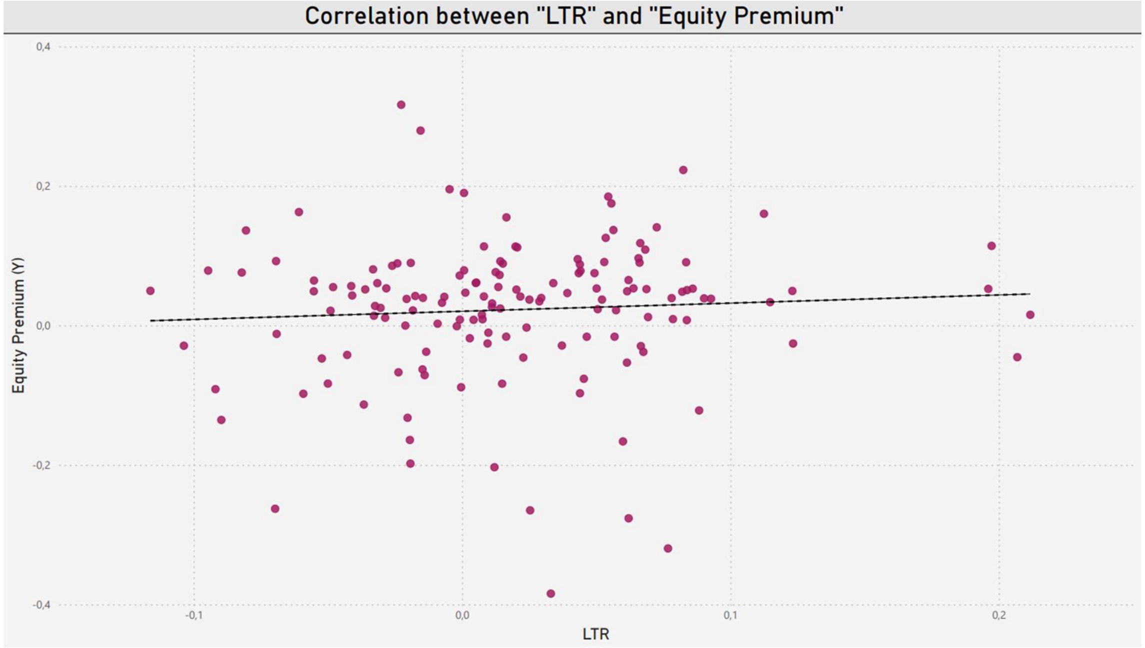
5) **Appendix 5:**



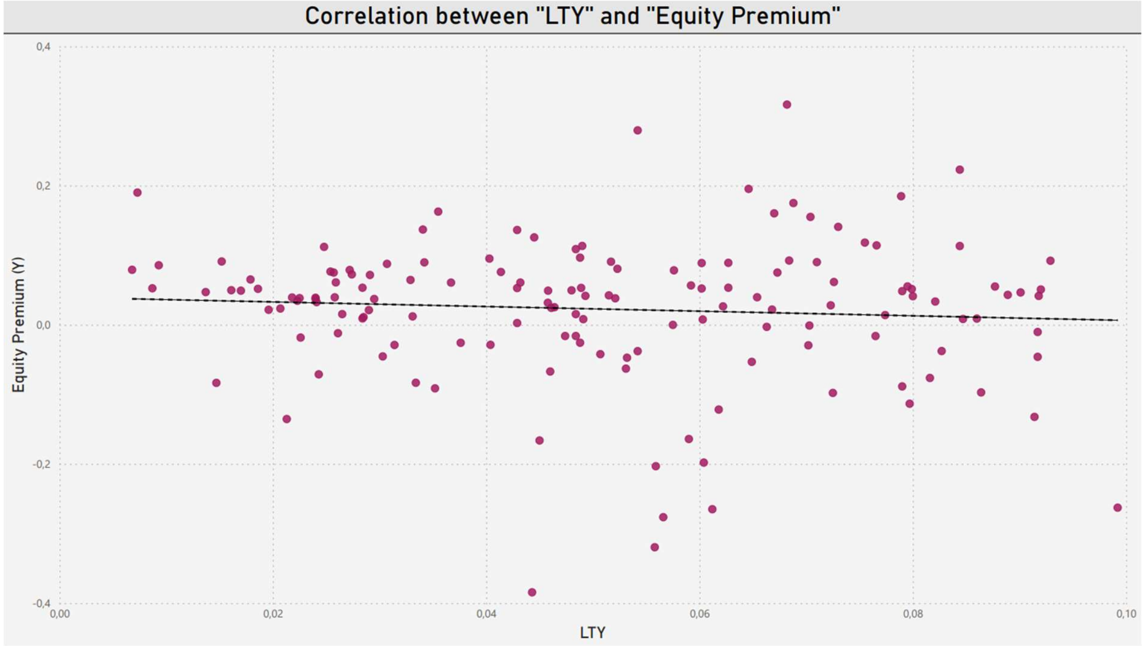
6) **Appendix 6:**



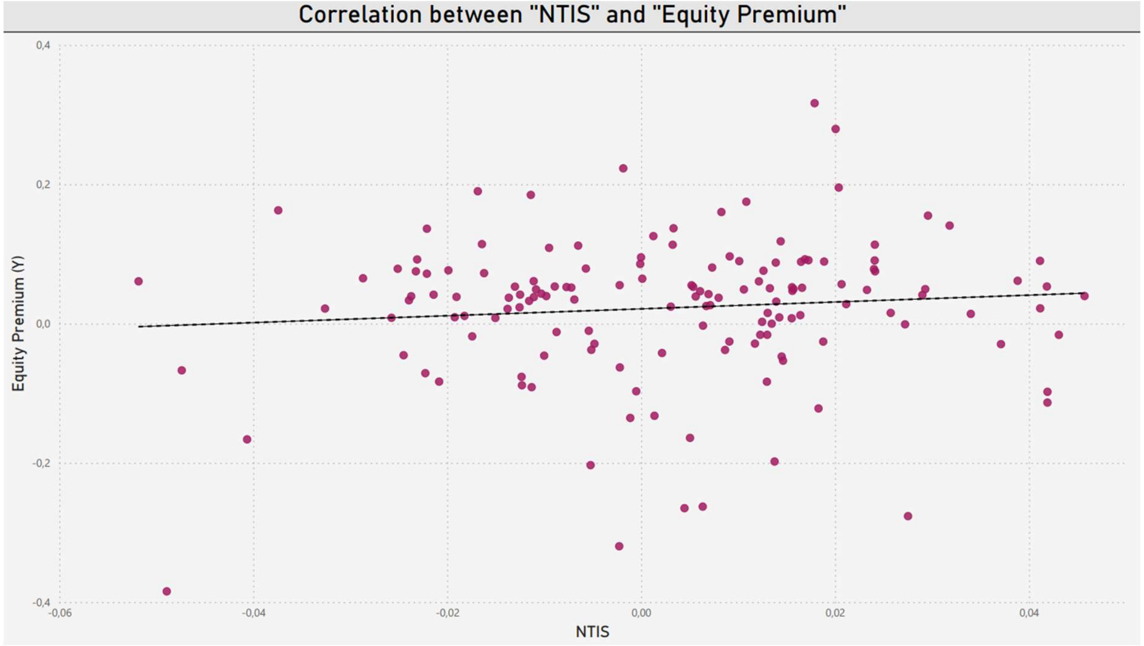
7) Appendix 7:



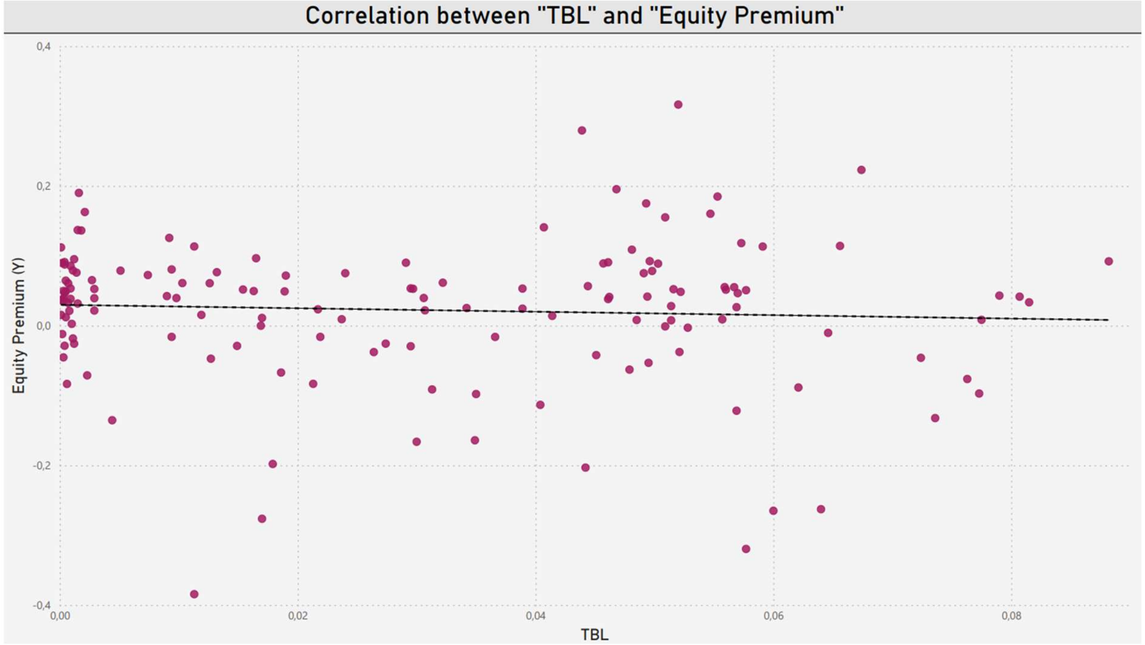
8) Appendix 8:



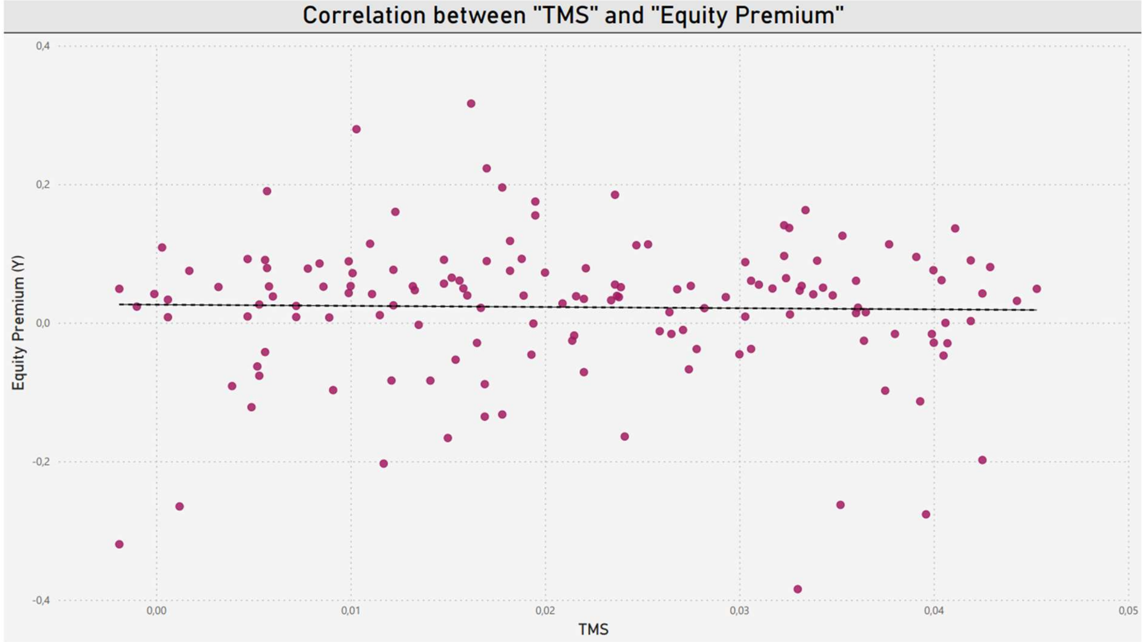
Appendix 9:



Appendix 10:



Appendix 11:


Correlation between "TMS" and "Equity Premium"

13. List of contacts

Hambuckers Julien	Professor at HEC	jhambuckers@uliege.be
Hubner Philippe	Professor at HEC	phubner@uliege.be
Zhou Gulfu	American economist	zhou@wustl.edu

14. Bibliography and references

Allwright, S. (2022a, December 6). *How to interpret MSE (simply explained)*. Stephen Allwright. https://stephenallwright.com/interpret-mse/

Allwright, S. (2022b, December 6). *How to interpret R Squared (simply explained)*. Stephen Allwright. https://stephenallwright.com/interpret-r-squared/

Amit Goyal. (n.d.). https://sites.google.com/view/agoyal145/?redirpath=

Amundi. (2023, June 28). *Retail Investors' Behaviour in the Digital Age: How Digitalisation is Impacting Investment Decisions*. Amundi Research Center. https://research-center.amundi.com/article/retail-investors-behaviour-digital-age-how-digitalisation-impacting-investment-decisions

Arithmetic Returns vs. Logarithmic Returns. (n.d.). Quantik. https://www.quantik.org/working-papers/arithmetic-returns-vs-logarithmic-returns/

Atlassian. (n.d.). *Mastering Scatter Plots: Visualize data correlations*. https://www.atlassian.com/data/charts/what-is-a-scatter-plot#:~:text=A%20scatter%20plot%20%28aka%20scatter%20chart%2C%20scatter%20graph%29,plots%20are%20used%20to%20observe%20relationships%20between%20variables

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468. https://doi.org/10.1057/jors.1969.103$

Bauldry, S. (2015). Structural equation modeling. In *Elsevier eBooks* (pp. 615–620). https://doi.org/10.1016/b978-0-08-097086-8.44055-9

Bobbitt, Z. (2022, January 25). *The Breusch-Pagan Test: Definition & example*. Statology. https://www.statology.org/breusch-pagan-test/

Broida, A. L. (1955). Diffusion indexes. *the American Statistician*, *9*(3), 7. https://doi.org/10.2307/2681458

Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies/ the Review of Financial Studies*, *21*(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

Chen, J. (2023, November 8). *What is equity Risk premium, and how do you calculate it?* Investopedia. https://www.investopedia.com/terms/e/equityriskpremium.asp#:~:text=Equity%20risk%20premium%20refers%20to%20an%20excess%20return,the%20level%20of%20risk%20in%20a%20particular%20portfolio

Chen, J. (2024, April 30). *Treasury yield: What it is and factors that affect it*. Investopedia. https://www.investopedia.com/terms/t/treasury-yield.asp

Contributeurs aux projets Wikimedia. (2024a, February 1). *S&P 500*. https://fr.wikipedia.org/wiki/S%26P_500

Contributeurs aux projets Wikimedia. (2024b, February 23). *NASDAQ-100*. https://fr.wikipedia.org/wiki/NASDAQ-100

Dates of U.S. recessions as inferred by GDP-based recession indicator. (2024, April 26). https://fred.stlouisfed.org/series/JHDUSRGDPBR

Dow, C. H., & Selden, G. C. (1920). Scientific stock speculation [New York, The Magazine of Wall street, c1920.]. In *New York, The Magazine of Wall street, c1920.* https://lccn.loc.gov/20012193

Download RStudio | The Popular Open-Source IDE from Posit. (2023, July 26). Posit. https://posit.co/products/open-source/rstudio/

Experts, T. (2023, July 3). In-Sample performance vs. Out-of-Sample performance (Explained). *How To Stop Writers Block*. https://predictivethought.com/in-sample-performance-vs-out-of-sample-performance-explained/

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, *100*(3), 514–537. https://doi.org/10.1016/j.jfineco.2011.02.003

Ferson, W., Sarkissian, S., & Simin, T. T. (2003). Spurious regressions in financial economics? *the Journal of Finance/ the Journal of Finance*, *58*(4), 1393–1413. https://doi.org/10.1111/1540-6261.00571

Google Colab. (n.d.). https://colab.research.google.com/#scrollTo=5fCEDCU_qrC0

Goval, A., & Welch, I. (2004). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *National Bureau of Economic Research*. https://doi.org/10.3386/w10483

Goyal, A., Welch, I., & Zafirov, A. (2021). A comprehensive look at the empirical performance ofEquity Premium Prediction II. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3929119

Hassan, M. (2023, August 15). *Descriptive Analytics - Methods, tools and examples*. Research Method. https://researchmethod.net/descriptive-analytics/

Hawkins, D. M. (2003). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12. https://doi.org/10.1021/ci0342472

Hayes, A. (2024, March 29). *Multicollinearity: Meaning, examples, and FAQs*. Investopedia. https://www.investopedia.com/terms/m/multicollinearity.asp

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Inoue, A., & Kilian, L. (2005). In-Sample or Out-of-Sample tests of predictability: Which one should we use? *Econometric Reviews*, *23*(4), 371–402. https://doi.org/10.1081/etc-200040785

Jousten, A. (2021). *Macroéconomie* [Slide show; Syllabus].

Université de Liège

K, G. M. (2021, December 15). Machine Learning Basics: Random forest regression - towards Data science. *Medium*. https://towards datascience.com/machine-learning-basics-random-forest-regression-be3e1e3bb91a

Kangralkar, S. (2023, July 5). Types of biases in data - towards data science. *Medium*. https://towardsdatascience.com/types-of-biases-in-data-cafc4f2634fb

Krishnan, V., Bharanidharan, & Krishnamoorthy. (2017). Research Data Analysis with Power BI. *INFLIBNET*.

Malato, G. (2023, May 3). *An introduction to the Shapiro-Wilk test for normality*. Built In. https://builtin.com/data-science/shapiro-wilk-test

Martin, I., & Nagel, S. (2019). *Market efficiency in the age of big data*. https://doi.org/10.3386/w26586

MATLAB. (n.d.). https://nl.mathworks.com/products/matlab.html

Mercurio, R., & Merrill, B. (2021). *Beginning Microsoft 365 Collaboration Apps: Working in the Microsoft Cloud*. Apress.

Nailman, A. (2023, May 12). *Why Python is the preferred language for machine learning*. Machine Learning Models. https://machinelearningmodels.org/why-python-is-the-preferred-language-for-machine-learning/

Overview for XNDX. (2024, May 10). https://indexes.nasdaqomx.com/Index/Overview/XNDX

pooling. (2024). https://dictionary.cambridge.org/dictionary/english/pooling

Qureshi, Z., & Woo, C. (2022). *Shifting paradigms: Growth, Finance, Jobs, and Inequality in the Digital Economy*. Brookings Institution Press.

Rapach, D. E., Strauss, J., & Zhou, G. (2009). Out-of-Sample Equity Premium Prediction: combination forecasts and links to the real economy. *Review of Financial Studies/ the Review of Financial Studies*, *23*(2), 821–862. https://doi.org/10.1093/rfs/hhp063

Rapach, D. E., & Zhou, G. (2013). Forecasting stock returns. In *Handbook of economic forecasting* (pp. 328–383). https://doi.org/10.1016/b978-0-444-53683-9.00006-2

Rapach, D., & Zhou, G. (2023). Forecasting stock returns. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4403635

Romer, P. M. (1989). *Endogenous technological change*.

Schyns, M. (2022). *Business Analytics* [Slide show; Syllabus]. Université de Liège.

Sewell, M. (2007). Principal Component Analysis. *Department of Computer Science University College London*.

Simonetti, I. (2022, November). The $24 trillion market that predicts and influences interest rates. *The New York Times*. https://www.nytimes.com/2022/11/02/business/treasury-yields-bond-market.html

Singh, S., & Jadhav, L. (2022, June 30). Data Analysis and Visualization of Sales Dataset using Power BI. *International Journal for Research in Applied Science and Engineering Technology*, *10*(6), 1749–1759. https://doi.org/10.22214/ijraset.2022.44132

Smith, T. (2022, April 2). *Disruptive Technology: definition, example, and how to invest*. Investopedia. https://www.investopedia.com/terms/d/disruptive-technology.asp#:~:text=Disruptive%20technology%20is%20an%20innovation%20that%20significantly%20alters,because%20it%20has%20attributes%20that%20are%20recognizably%20superior

Spiegel, M. (2008). Forecasting the equity premium: Where we stand today. *Review of Financial Studies/ the Review of Financial Studies*, *21*(4), 1453–1454. https://doi.org/10.1093/rfs/hhn069

Stock, J. (2015). *Heteroskedasticity- and Autocorrelation-Robust inference* [Slide show; AEA Continuing Education Lectures on Time Series]. Havard University. https://scholar.harvard.edu/files/stock/files/aea_2015_lecture4_har_rev.pdf

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. https://doi.org/10.1198/073500102317351921

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430. https://doi.org/10.1002/for.928

Team, I. (2023, August 29). *Modern Portfolio Theory: What MPT is and how Investors use it*. Investopedia. https://www.investopedia.com/terms/m/modernportfoliotheory.asp

Thomas, C. A. (2023, August 9). Technology competition between nations: Views from industry leaders. *Brookings*. https://www.brookings.edu/articles/technology-competition-between-nations-views-from-industry-leaders/

Tiwari, A. (2022). Supervised learning: From theory to applications. In *Elsevier eBooks* (pp. 23–32). https://doi.org/10.1016/b978-0-12-824054-0.00026-5

Tretina, K. (2023a, May 10). What is the Nasdaq 100 index? *Forbes Advisor*. https://www.forbes.com/advisor/investing/what-is-the-nasdaq-100-index/

Tretina, K. (2023b, August 9). What is the S&P 500? How does it work? *Forbes Advisor*. https://www.forbes.com/advisor/investing/what-is-sp-500/

Welch, I., & Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies/ the Review of Financial Studies*, *21*(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

What is Python? Executive Summary. (n.d.). Python.org. https://www.python.org/doc/essays/blurb/

Yin, A. (2021). Does the kitchen-sink model work forecasting the equity premium? *International Review of Finance*, *22*(1), 223–247. https://doi.org/10.1111/irfi.12352

Executive Summary

This thesis investigates whether the real economy continues to provide predictive signals for technology stock returns in U.S. companies during the digital era and economic crises. The study updates and extends previous research by focusing on modern economic indicators and their forecasting power amidst the evolving market landscape influenced by technological advancements and recent economic disruptions.

The research explores the viability of historical economic indicators in forecasting stock returns, particularly in the U.S. technology sector. It questions whether these indicators, which proved useful in past decades, still hold predictive power in a radically transformed economic and technological environment.

The thesis adopts a mixed-methods approach, combining quantitative analyses with econometric modelling. The data spans several decades, focusing on periods marked by significant economic events, such as the dot-com bubble, the 2008 financial crisis, and the COVID-19 pandemic, to assess the robustness of forecasting models under different economic conditions.

The study replicates and extends previous methodologies using RStudio, allowing for a comparison between past and current forecasting abilities of various economic indicators under different economic conditions. Leveraging Google Colab and Python, the thesis incorporates machine learning techniques to examine the predictive power of traditional and newly proposed economic indicators. This analysis aims to uncover complex nonlinear relationships that might be missed by conventional econometric models.

The findings suggest that the traditional indicators have diminished in predictive power. The discussion delves into the implications of the findings for investors and policymakers, emphasising the need for adaptive strategies that account for the rapid technological changes and their impact on the economic landscape. The thesis concludes that the real economy continues to provide valuable insights into technology stock returns, even if the predictive power has diminished. It calls for ongoing research to refine these indicators and adapt forecasting models to the changing economic and technological environment.

Suggestions for future research include exploring additional digital economy indicators and extending the analysis to global technology markets to validate the findings and enhance the generalizability of the forecasting models. This study contributes to the literature by updating forecasting models with contemporary economic indicators and by demonstrating the evolving relationship between the real economy and technology stock returns in the face of digital transformation and economic crises.

WORD COUNT = 26.066

HEC LIÈGE
Management School - Liège Université