Norwegian School of Economics
Bergen, Fall 2023

Nowcasting GDP Growth on a Weekly Basis

Leveraging Comprehensive News Article Information and Macroeconomic Indicators

Jon Kvinnsland & Nicholas Foss
Supervisor: Håkon Otneim

Master thesis, Economics and Business Administration
Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

Timely economic indicators are crucial for effective macroeconomic decision-making. In particular, this applies during crises when economic shifts can be large and costly. In this thesis, we produce weekly Gross Domestic Product (GDP) growth estimates for 22 large Western economies using a neural network ensemble and a unique combination of macroeconomic data and text variables derived from 686 390 news articles. Our contribution to the literature on GDP growth nowcasting is investigating whether comprehensive information about the news coverage combined with macroeconomic indicators can be leveraged to improve nowcasting. Specifically, we combine sentiment analysis with macroeconomic variables and news topic analysis through zero-shot classification. We find that our model effectively captures the GDP growth trend in most countries during normal times and excels during crises, such as the Global Financial Crisis and the COVID-19 pandemic. Analyzing Shapley values revealed that the ensemble model identifies interaction effects and non-linearities, which enable it to capture sharp shifts during crises.

Contents

1 Introduction **7**

2 Rethinking Economic Indicators: the Emergence of High-Frequency Alternatives **10**
2.1 The shortcomings of traditional economic indicators during crises . . . 10
2.2 Current academic literature on text-based high-frequency indicators . . . 11
2.3 Addressing shortcomings in the existing textual analysis literature . . . 12
2.4 Our contributions to the existing literature . . . 12

3 Data Resources **14**
3.1 Selecting the macroeconomic variables . . . 14
3.2 Pre-processing of macroeconomic variables . . . 15
3.3 Selecting the newspapers for textual variables . . . 15
3.4 Including newspaper weighting variables . . . 16
3.5 Ensuring uniform textual analysis by translating non-English articles . . . 17
3.6 Deriving textual variables with sentiment analysis . . . 18
3.7 Deriving textual variables with zero-shot classification . . . 19
3.8 Accounting for time-based trends in sentiment and classification data . . . 20

4 Machine Learning Approach **21**
4.1 Using an ensemble of neural network models . . . 21
4.2 Describing the nowcasting algorithm for weekly GDP growth predictions . . . 21
4.3 Advantages and disadvantages of grouping countries during training . . . 23

5 Findings: Evaluating the Nowcasting Algorithm **25**
5.1 Assessing the efficacy of the nowcasting algorithm without grouping countries . . . 25
5.2 Comparing the nowcasting efficacy with and without grouping countries . . . 28
5.3 Comparing the nowcasting algorithm with an autoregressive baseline model . . . 33

6 Discussion: Understanding Variable Importance During Different Economic Conditions **38**
6.1 Explaining important features of Shapley values . . . 38
6.2 Uncovering the relevance of text data using Shapley values . . . 39
6.3 Key variables during the 2008 financial crisis and COVID-19 pandemic . . . 40
6.3.1 Capturing interaction effects can boost nowcasting performance during crises . . . 41

6.3.2 Capturing non-linear relationships also improves nowcasting performance during crises 42
6.4 Assessing which timespan provides the most insightful text variables 44

7 Conclusions and Future Research 46

A Appendix 51

List of Figures

3.1 Distribution of news articles per newspaper, 2007 - 2023. 17
3.2 Histogram of sentiment scores across all news articles. 18

4.1 Pseudocode of the nowcasting algorithm and a visual representation of the process. 22
4.2 GDP growth curves of larger Western economies. 23
4.3 GDP growth curves of smaller Western economies. 23

5.1 Nowcasts without grouping countries with monthly- and quarterly smoothed lines, United States. 25
5.2 GDP growth values and nowcasts in Q2 2020. 27
5.3 Nowcasts without grouping countries with monthly- and quarterly smoothed lines, Germany. 28
5.4 Nowcasts when grouping countries with monthly- and quarterly smoothed lines, Germany. . 29
5.5 Comparing GDP growth values and nowcasts with and without grouping countries in Q2 2020. 31
5.6 Autoregressive predictions and nowcasts when grouping countries with monthly- and quarterly smoothed lines, France. 34
5.7 Comparing GDP growth values with nowcasts and AR(4) predictions in Q2 2020. 36

6.1 Shapley summary plot for the period 2007 - 2023 39
6.2 Shapley value summary plots for the 2008 financial crisis and COVID-19 pandemic. 40
6.3 Contribution of most important variables for the nowcast of Q2 2020. 41
6.4 Partial dependency plots for Stock Market Text, Consumption Text, and Consumer Confidence Index. 42
6.5 Partial dependency plot for Pandemic Text and Recession Text. 42
6.6 Partial dependency plot for Pandemic Text and Hours Worked. 43
6.7 Partial dependency plot for Economic Growth and Recession. 44
6.8 Relative Shapley values for all text variables over the period 2007 - 2023. 44

A.1 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Australia. . . . 52
A.2 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Austria. 53
A.3 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Belgium. . . . 53
A.4 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Canada. 54

A.5 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Denmark. . . . 54
A.6 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Finland. 55
A.7 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, France. 55
A.8 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Germany. . . . 56
A.9 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Greece. 56
A.10 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Iceland. 57
A.11 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Ireland. 57
A.12 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Italy. 58
A.13 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Luxembourg. . 58
A.14 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Netherlands. . 59
A.15 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, New Zealand. . 59
A.16 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Norway. . . . 60
A.17 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Portugal. . . . 60
A.18 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Spain. 61
A.19 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Sweden. . . . 61
A.20 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Switzerland. . 62
A.21 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, United Kingdom. 62
A.22 AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, United States. . 63

List of Tables

2.1 Economic indicators and their characteristics. 10

3.1 Summary statistics of macroeconomic variables (unscaled). 15
3.2 News outlet information and summary statistics from news articles. 16
3.3 Descriptive statistics of sentiment scores across all news articles. 18
3.4 Zero-shot classification candidate labels and their descriptive statistics. 19

5.1 MSE scores for the grouping strategies, 2007-2023 and 2011-2019. 30
5.2 MSE scores for the grouping strategies, 2008 financial crisis and COVID-19 pandemic. . . . 32
5.3 MSE scores for nowcasting model and AR(4), 2007-2023 and 2011-2019. 35
5.4 MSE scores for nowcasting model and AR(4), Financial Crisis and COVID-19 pandemic. . . 37

A.1 Summary of Stock Market Indices and their Bloomberg Tickers. 51
A.2 Neural network (NN) candidate configurations, where N is the number of variables. 52

1. Introduction

Accurate and timely economic data is crucial for effective macroeconomic decision-making. Early access to precise information allows policymakers and market participants to implement effective policies, capitalize on market opportunities, and avoid losses. A key indicator applied for macroeconomic decision-making is the Gross Domestic Product (GDP). GDP describes the value of all produced products and services within a country's borders within a specific timeframe, making it an effective tool for evaluating the economic health of a country. Unfortunately, the most granular GDP estimates for most Western countries are quarterly values and are generally made available a couple of months after their corresponding timeframe. Consequently, policymakers and market participants are driven to alternative indicators, with more frequent releases and shorter delays.

Two prominent examples of more frequent indicators are the Purchasing Managers' Index (PMI) and the Consumer Confidence Index (CCI). These are survey-based indicators that reflect the purchasing managers' and customers' insights into the current and near-future state of the economy. These indicators are country-specific and released monthly, which is more frequent than most Western GDP values, but still too infrequent for real-time analysis of the economic state of a country. Additionally, they are not GDP substitutes, but instead provide information about GDP fluctuations. Furthermore, they possess limited capability to reflect the magnitude of economic changes, particularly during sudden shocks. Consequently, the search for timely and reliable indicators for assessing the current state of a country's economy continues.

News articles are potential sources of valuable information about a country's economy that have the benefit of being released on a near-daily basis. There are several reasons why we expect information regarding GDP to be present in news articles. Firstly, economic news articles cover topics that are directly or indirectly linked to GDP, such as unemployment rates, inflation, consumer spending, business investments, imports, and exports. Additionally, economic experts and analysts frequently contribute to news articles, offering insights into GDP trends and predictions. Lastly, political changes, international trade developments, and global crises are typically covered in the news. These events often include discussions about how they might influence economic growth, thus providing relevant information for GDP prediction.

By efficiently extracting information from news articles, it is possible to create economic indicators with weekly or daily updating frequencies. This is shown in the current literature, with important contributions including Ashwin, Kalamara, & Saiz (2021) and Woloszko (2020). We find that most of the current literature on applying textual analysis to nowcast GDP growth is dominated by sentiment analysis approaches. This involves extracting information about the polarity of texts to derive information. Other approaches include lexicon-based textual analysis, which consists of determining text topics based on the presence of certain words or phrases (Thorsrud, 2016). A frontier that has not yet been explored is whether combining information about news article contents with their polarity and concurrent macroeconomic data can be leveraged to create high-frequency economic indicators. Transformer methods, such as zero-shot classification and sentiment analysis, enable extracting a text's content information and its polarity, respectively. Zero-shot classification has the added benefit of considering the context when determining which topics are relevant to a text, which distinguishes it from the lexicon-based approach. Thus, this thesis seeks to add to the literature by answering the following research question:

"Can news topics, news sentiment, and macroeconomic data be combined to nowcast the GDP growth of Western economies on a weekly basis?"

Our research aims to investigate whether the combination of news topics, sentiment, and macroeconomic data can be applied to nowcast GDP growth weekly. The quarterly nowcasts are compared to the observed GDP growth values and the weekly are assessed visually. Our model's performance is compared against a baseline model. In addition, we want to explore how different news topics, sentiment scores, and macroeconomic variables correlate with GDP growth under different economic conditions, including crises and normal periods. To this end, we have conducted an analysis utilizing Shapley values. The geographical scope for the analysis enables comparison of nations with similar economic structures and operations, thereby minimizing the influence of unique country-specific variations for nowcasting GDP growth. Based on this, we concentrated on the largest Western economies, including 22 countries in total.

The method we used to analyze the research topic involved zero-shot classification and sentiment analysis on a corpus of 686 390 articles from The Guardian, New York Times, Handelsblatt, Der Spiegel, and Les Echos. Zero-shot classification produces scores of how relevant a candidate label (topic) is to an article, and sentiment analysis describes how positively or negatively loaded an article is. Thus, it was possible to create variables describing the contents and polarity of articles by aggregating the scores from the textual analysis. The macroeconomic and textual variables were used as inputs in an ensemble model, consisting of several neural networks (NNs), that produced the nowcasts. Nowcasts were generated for the 22 largest Western economies.

The ensemble model demonstrates strong performance in out-of-sample nowcasting scenarios. In quarterly measurements, the model achieves a Mean Square Error (MSE) of 1.661[1], outperforming an autoregressive baseline model with four lags, which records an MSE of 5.865[1]. The model adeptly captures most economic downturns and recoveries, including the periods around the 2008 financial crisis and the COVID-19 pandemic. It effectively identifies the timings of these crises, though it slightly underestimates the intense GDP downturn during the peak of the COVID-19 pandemic (Q2 2020). While there are occasional discrepancies in its estimates, the model could be a valuable asset for nowcasting GDP growth. Reflecting George Box's quote (2018), "Essentially, all models are wrong, but some are useful," this model may not provide an exact representation of GDP growth, but its estimates offer valuable insights, revealing the impact of various factors on GDP growth over different periods.

[1]Average MSE of quarterly predictions across all countries between 2007 and 2023. *See Table 5.3*

This paper comprises seven chapters, beginning with the introduction. In *Chapter 2* we discuss the significance of high-frequency indicators and present the significant contributions of this paper's model to existing literature. *Chapter 3* contains our description of the data sources used. In the fourth chapter, we delve into the machine learning methodology employed. *Chapter 5* contains our evaluation of the model's efficacy. In *Chapter 6*, we offer an in-depth analysis of the variables' importance to the model's output. The thesis concludes with our answer to the research question, on whether news topics, news sentiment, and macroeconomic data be combined to nowcast the GDP of Western economies weekly. The conclusion also includes our proposals for potential areas for further research.

2. Rethinking Economic Indicators: the Emergence of High-Frequency Alternatives

Traditional indicators are effective tools for decision-making under normal conditions, but many of them struggle during crises as they lack regular updates and fail to accurately measure the extent of crises (Claessens & Kose, 2023). The COVID-19 pandemic exposed these shortcomings with its unprecedented scale and rapid economic shifts. In general, policymakers and market participants have two types of indicators at their disposal when making decisions: (1) hard indicators and (2) soft indicators. Hard indicators, such as GDP, are quantitative and objective but suffer from low publication frequency and release delays of a couple of months. These limitations pose a major challenge for making decisions in times of rapid economic changes. On the other hand, soft indicators such as PMI and CCI are timelier but fail to accurately capture the magnitude of changes during crises. This suggests that there are other factors influencing GDP growth during crises that the traditional indicators overlook. *Table 2.1* displays the mentioned indicators and their characteristics.

Table 2.1: Economic indicators and their characteristics.

Indicator	**Type**	**Frequency**	**Release**	**Relationship to GDP**
GDP	Hard	Quarterly	1-2 months delay	-
Industrial production	Hard	Monthly	1-2 months delay	Linear
PMI	Soft	Monthly	1-3 days delay	Linear (non-linear during crises)
CCI	Soft	Monthly	1-3 days delayed or in advance	Linear (non-linear during crises)
News	High-frequency	Continuous	Immediate	Non-linear

2.1 The shortcomings of traditional economic indicators during crises

To illustrate the disadvantages of traditional hard and soft indicators during a crisis, consider the case of Norwegian policymakers who implemented a lockdown in mid-March. Following the lockdown's implementation on March 12th, the first available data came from soft indicators such as the PMIs on April 1st. These initial releases presented a mixed picture, highlighting the limitations of such measures during crises. The overall PMI showed a moderate decline (to 40.80), while the manufacturing PMI hit an all-time low (34.45), reflecting the uneven impact of the crisis (NIMA, 2022). It took more than seven weeks from the start of the lockdown for the first traditional, hard indicators to shed light on the activity in March. These indicators were released in May 2020, and showed a significant month-on-month decrease of 15.1% in household consumption and a 7.4% month-on-month decline in GDP (SSB, 2023).

2.2 Current academic literature on text-based high-frequency indicators

In recent years, the use of text data from news sources has emerged as an alternative candidate for GDP modeling over traditional indicators. Its capability for real-time updates and detailed economic perspectives makes it ideal for delivering high-frequent information that also captures other factors important for GDP growth during crises. For instance, Watson (2023) point out that hard indicators, which mainly concentrate on financial markets, are inadequate for predicting broader macroeconomic trends due to their inconsistency. They suggest that more varied and comprehensive information sources are necessary for forecasting these broader trends. Additionally, Ellingsen, Larsen, & Thorsrud (2016) argue that news articles more effectively reflect household expectations regarding future economic conditions compared to traditional soft indicators. This view is supported by Baker et al. (2016), who found that factors such as politics and uncertainty, typically underrepresented in standard economic data, can have a significant impact on GDP growth forecasts. Conclusively, text data offers potentially more comprehensive data that applies to both normal times and crises compared to current indicators.

Recent academic research has extensively explored the application of text data in economic analysis and GDP nowcasting. Baker, Bloom, & Davis (2016) demonstrated how uncertainty indices based on newspaper articles can forecast downturns in macroeconomic variables. Larsen & Thorsrud (2016) used topic-based sentiment indicators, also based on news articles, to predict economic variables like asset prices. Other studies, including those by Barbaglia, Consoli, & Manzan (2022), Kalamara et al. (2022), Shapiro, Sudhof, & Wilson (2022), Jardet & Meunier (2022), and Aguilar et al. (2021), have applied textual analysis on the sentiment of news articles to nowcast GDP, inflation, and unemployment. Barbaglia et al. (2023) expanded this research to also include unconventional variables such as Airbnb review numbers, air cargo, and air quality statistics, alongside measures of media attention.

Institutions like the European Central Bank (2021), International Monetary Fund (2022), Bank of England (2020), and OECD (2020) have explored using high-frequency data, including spending, travel, job postings, and textual sentiment analysis for real-time economic tracking, particularly during crises like the COVID-19 pandemic. Although these indicators are not the primary metrics for economic analysis used by the organizations mentioned above, the reality that they actively conduct research on them and release working papers highlights a strategic trend and shift in economic forecasting and analytical methods. Effectively, the use of diverse and high-frequent data, especially text-based materials such as news articles, combined with sophisticated machine learning methods, reveals a more comprehensive and accurate understanding of current economic trends.

2.3 Addressing shortcomings in the existing textual analysis literature

A major limitation in the existing research on textual analysis is its dependence on lexicon-based models, which struggle to capture text context. For example, in sentiment analysis, these models compare words to a preset list of terms with assigned sentiment scores (positive, negative, or neutral), and the text's overall sentiment is deduced by totaling these scores. A widely used dictionary for this purpose is the Valence Aware Dictionary and sEntiment Reasoner (VADER), which effectively interprets casual language, emojis, slang, and emphasis through its sentiment lexicon and rule-based approach (Hutto & Gilbert, 2014). For instance, Ashwin, Kalamara, & Saiz (2021) applied this dictionary to do sentiment analysis on text data and nowcast GDP growth. However, being a lexicon-based model, VADER faces challenges in grasping context and detecting intricate subtleties such as sarcasm or nuanced emotional expressions (Poria et al., 2020).

Furthermore, most studies nowcasting GDP growth with text data have concentrated on sentiment analysis, overlooking the actual content of the articles. This leads to limited insights into the specific topics driving GDP growth. Only a few researchers, including Thorsrud (2016) and Burri (2023), have investigated the relationship between news topics and GDP growth, providing clearer insights into how GDP-related variables change over time. However, these studies also use lexicon-Based Models to identify article topics by matching words to a defined list of terms. This is a significant limitation in the current research because it negates the context of each word in the article.

2.4 Our contributions to the existing literature

In this paper, we make several contributions to the existing body of literature. First, to address the context of the text in sentiment analysis, we have applied models based on the Transformer architecture. The Transformer architecture, introduced in the groundbreaking paper ”Attention Is All You Need” by researchers at Google Brain (2017), revolutionized the field of Natural Language Processing (NLP). It offers an innovative approach for sequence-to-sequence tasks like translation and text summarization. A prominent example of its application is the well-known ChatGPT models, launched by OpenAI in November 2022, which also utilize this architecture. The Transformer architecture is distinguished by its ”self-attention” mechanism, enabling it to assess the relevance of different words within a sentence. This capability allows it to effectively understand the context of single or multiple sentences. Moreover, its architecture is designed for efficient parallelization, making it highly scalable and capable of handling extensive datasets. This increased scalability boosts performance and enables a more nuanced understanding of word context in sentences. Consequently, it enhances the precision and clarity of detailed sentiment analysis tasks.

Following the sentiment analysis, we have employed a zero-shot classification analysis to classify articles on their content. Zero-shot classification analysis takes an article and a set of candidate labels as input and outputs a score between 0 and 1 for each candidate label, indicating how relevant that label is to the article. The technique employs the same Transformer architecture used in the sentiment analysis, thus benefiting from the same improvements in understanding context and processing large datasets efficiently. The model is initially trained to recognize and understand a variety of known categories. It then leverages this training to categorize new, unseen data, making it highly effective for text classification across a diverse set of labels. The zero-shot classification approach enables us to extract more information from news articles than simply the sentiment. Based on the classification scores, we can now evaluate how the relevance of various news topics in nowcasting GDP growth shifts during crises compared to normal periods.

We aim to improve conditions for nowcasting GDP growth by combining sentiment analysis with zero-shot classification. We believe this approach is beneficial because it provides the model with more comprehensive information about the news articles compared to the existing literature, by including both the sentiment and contents. An important reason why we believe this approach could enhance the model's preconditions for nowcasting is that it captures the relationship between sentiment and content. This is important because signals derived from an article's sentiment might be interpreted differently based on the contents of that article.

3. Data Resources

In this thesis, our selection of data resources was made with careful consideration, based on our belief that they are crucial for nowcasting the target variable. The target variable we want to nowcast is quarterly GDP growth expressed as year-on-year change. This variable is sourced from the Organization for Economic Co-operation and Development (OECD) for each country.

3.1 Selecting the macroeconomic variables

The macroeconomic variables we used as independent variables in this thesis include *inflation*, *unemployment*[1], the *monetary policy rate*[2], *current account balance*[3], *hours worked*[4], *consumer confidence index*[5], and the *stock market*[6]. All values are country-specific, meaning they describe the national conditions of the 22 observed countries in this thesis, as opposed to being international market indicators. The macroeconomic data is primarily sourced from the Organization for Economic Co-operation and Development (OECD), supplemented by Bloomberg for areas where the OECD data is incomplete.

We selected these variables due to their interconnectedness to GDP and its components: *consumption*, *investment*, *government spending*, and *net exports* (exports minus imports) (Stobierski, 2021). Inflation has a multifaceted impact on GDP, with an example being how it affects consumption by changing purchasing power. Unemployment is correlated with GDP growth because it reflects the size of the labor force, and influences purchasing power and the need for government spending on social welfare. Hours worked describe productivity and are normally positively correlated with GDP growth. The monetary policy rate affects GDP growth by regulating the cost of borrowing money. For instance, higher monetary policy rates deter investment and slow down consumption. The consumer confidence index (CCI), is a survey-based metric that describes the consumers' outlook, providing insights into near-future consumption. The stock market is viewed as a predictor of economic trends. A rising market may indicate a healthier economy, leading to increased consumer spending and investor optimism, which can enhance GDP growth. However, its reliability as an indicator is not absolute, as shown by historical events like the 2008 financial crisis. Lastly, the current account balance reflects a country's trade balance and is mainly related to net exports. A surplus or deficit in the current account can indicate trends affecting exports and imports, indirectly impacting other components of GDP through broader economic implications. The mentioned connections between variables and GDP are just a few examples and not an exhaustive explanation.

[1] Percentage of labor force without employment (seasonally adjusted) (OECD, 2023e).

[2] The cost of borrowing set by the national central bank (IMF, 2023a).

[3] A record of a country's international transactions (year-on-year percentage change) (OECD, 2023b).

[4] Total hours worked per year divided by the average number of employees (year-on-year percentage change) (OECD, 2023c).

[5] Indication of future developments of households' consumption and saving, based on survey answers about their expected financial situation (OECD, 2023a).

[6] A country main stock market index, expressed as year-on-year change. See details *Table A.1* in *Appendix A*.

3.2 Pre-processing of macroeconomic variables

A few processing steps were necessary to prepare the variables for the nowcasting algorithm. Firstly, monthly macroeconomic data was converted into quarterly values to fit with the quarterly releases of GDP. Secondly, several variables were converted into percentage change values, to localize the change and remove trends. This pre-processing step was done for both the current account balance and hours worked variables. Thirdly, all variables have been scaled with standard scaling. This ensures that these variables have a mean of zero and a standard deviation of one. The scaling was incorporated sequentially as part of building the nowcasting algorithm, and the process is detailed in *Section 4.2*. Summary statistics of the macroeconomic variables, before the scaling, are provided in *Table 3.1*.

Table 3.1: Summary statistics of macroeconomic variables (unscaled).

Variable	**Mean**	**25th Perc.**	**75th Perc.**	**Std. Deviation**	**Max**	**Min**
CCI	94.86	97.56	100.82	20.14	144.87	-36.8
Inflation (%)	2.17	0.76	2.82	2.35	17.1	-2.76
Unemployment (%)	7.55	4.89	8.42	4.33	27.8	2.02
Monetary policy rate (%)	1.36	0	1.88	2.03	18	-0.75
Current account balance (Δ %)	0.15	0.14	0.14	0.02	1	0
Hours worked (Δ %)	-0.07	-0.43	0.3	2.15	21.96	-16.78
Stock market (Δ %)	0.11	0.02	0.15	0.12	0.9	0
GDP Growth (Δ %)	1.54	0.38	3.04	4.04	26.21	-21.94

The Canadian consumer confidence index and the hours worked values for Luxembourg could not be found in either OECD's or Bloomberg's database. These values were imputed with the mean value of all other countries at the corresponding date. Due to the small scale of the data, relative to its complexity, it was considered more appropriate to impute these values rather than removing the countries or variables in question.

3.3 Selecting the newspapers for textual variables

Our primary considerations when selecting newspapers to gather articles from included that the outlets must be relevant, have frequent publications, and be accessible. It was also considered beneficial to include news outlets from different countries, to get broader geographic coverage. With these criteria, the following five news outlets were selected: The Guardian, The New York Times, Les Echos, Handelsblatt, and Der Spiegel.

The Guardian and the New York Times provide articles daily, meaning that they meet the requirements for publication frequency. With regards to relevance, both The Guardian and The New York Times have a broad journalistic focus, as they cover everything from financial news to sporting events and political topics. However, both publishers host their own API solution that allows users to filter articles based on topics and dates. Thus, we could retrieve relevant articles with relative ease, which meant the relevancy and accessibility criteria were met also. For the United Kingdom, The Financial Times was considered as an alternative candidate due to their renowned coverage of economic topics, but accessing their articles proved expensive and difficult with the lack of an API. The same was the case for The Wall Street Journal, which was a potential substitute for The New York Times.

Les Echos did not offer an API, and we retrieved articles directly from their archives. The lack of an API meant that articles could not be filtered. However, Les Echos' distinct emphasis on financial news meant that no such filtering was necessary. Les Echos satisfied all requirements and was selected as an article source. Similarly to Les Echos, Handelsblatt did not have an API, and articles were gathered from their archives. As their archives did not include 2007, we used articles from Der Spiegel as substitutes.

The total article resources included 686 390 articles spanning from January 1st, 2007, to September 30th, 2023. We set a minimum threshold for articles per week during this entire interval at 50 publications per newspaper. This was a measure to ensure that the textual data was indeed evenly distributed. Additionally, we set a minimum and maximum threshold for article length at 200 and 5000 characters, respectively. This was to exclude articles that were too brief to yield meaningful information and to account for the increasing ambiguity in sentiment and content observed in lengthier articles. *Table 3.2* provides information about the selected news outlets and summary statistics from the articles.

Table 3.2: News outlet information and summary statistics from news articles.

Statistic	**Der Spiegel**	**Handelsblatt**	**Les Echos**	**The Guardian**	**New York Times**
Country of origin	Germany	Germany	France	United Kingdom	United States
Annual subscribers	700 000	170 000	140 000	1 000,000	10 000 000
Max. Article Length	5 000	5 000	5 000	5 000	5 000
Min. Article Length	200	205	200	202	200
Mean Article Length	2 244	2 444	2 534	2 887	3 196
25th Percentile	1 261	1 422	1 851	1 969	2 386
75th Percentile	3 115	3 366	3 275	3 843	4 135
Std. Deviation	1 213	1 212	1 089	1 169	1 128
Total Articles	22 819	145 054	327 781	91 681	99 055

3.4 Including newspaper weighting variables

Figure 3.1 illustrates the number of articles retrieved from each newspaper by quarter. A notable observation is how the composition of the article base changes across time. To account for possible implications of this inconsistency, we include a set of newspaper weighting variables as independent variables. The weighting variables are simply the proportion of the article base used to derive a set of textual variables that comes from a country. For instance, if the German weight for a period is 0.2, it means that 20% of the articles published in that period are from Handelsblatt or Der Spiegel. This could be a useful addition if the importance of certain variables is dependent on the composition of the article base. For instance, it could be the case that certain variables are particularly influential when the portion of articles supplied by the New York Times is high if the New York Times' coverage of those topics is a good proxy of the relevance of these variables.

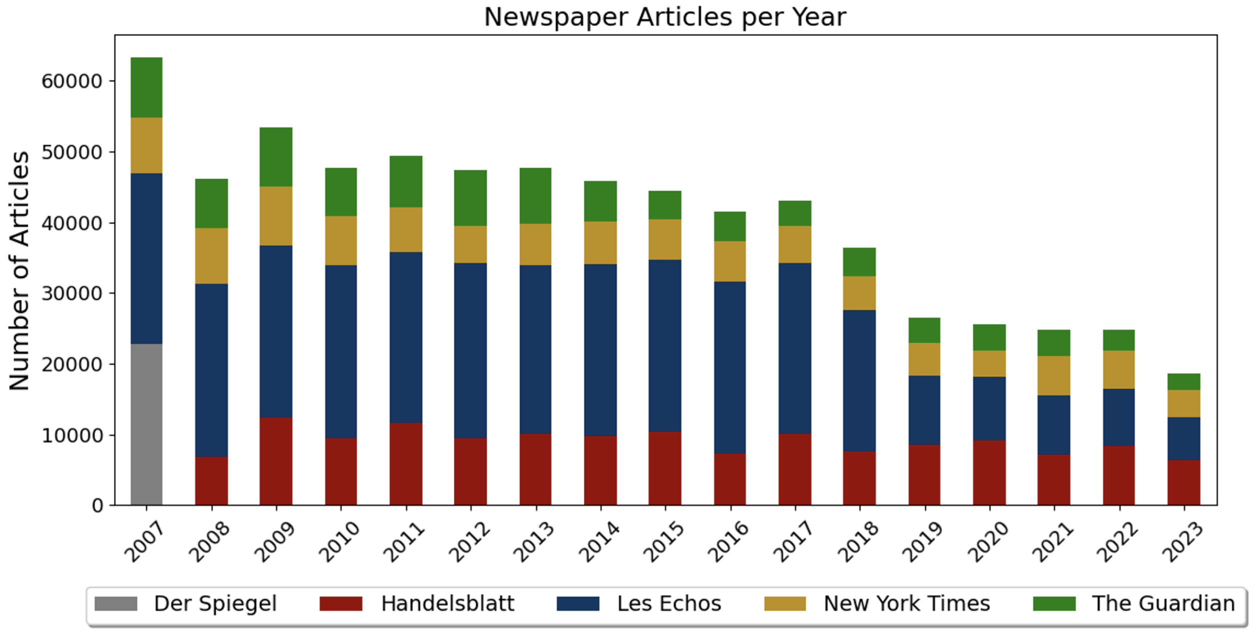

Figure 3.1: Distribution of news articles per newspaper, 2007 - 2023.

3.5 Ensuring uniform textual analysis by translating non-English articles

We translated all news articles into English as a measure to ensure uniformity in the textual analysis. Although we considered using specialized textual analysis models for German and French, that approach would involve the risk of incurring inconsistent results due to variances in model outputs.

Following our choice to translate articles, it is important to underscore the importance of having high-quality translation to preserve the content, sentiment, and context of an article. This is primarily because idiomatic expressions and linguistic nuances might not translate well to other languages. Preserving the context and sentiment of the translated articles is important to ensure that the interpretation of the news articles is as accurate as possible, thereby enabling the nowcasting algorithm for GDP growth to produce as precise estimates as possible. For this purpose, we considered the Google Trans API and a Transformer model for text translation. The Google Trans API offers relatively accurate translations with minimal processing time. On the other hand, the Transformer model for text translation is better at capturing the context and sentiment, but requires significantly more processing time. Due to the excessive processing time required by the Transformer model, it was infeasible to translate our extensive collection of articles with this method. As a result, we opted for the Google Trans API. To ensure that Google Trans API's translations were satisfactory, we manually assessed a series of random samples. Based on the analysis, we concluded that it produced adequate translations, and the benefits of the more elaborate models would be minimal even if the processing times were similar.

3.6 Deriving textual variables with sentiment analysis

The sentiment analysis used in this thesis outputs three scores per article: *negative*, *neutral*, and *positive*, each ranging from 0 to 1. Sentiment scores are positively correlated with the relevance of that particular sentiment type (negative, neutral, or positive). Additionally, the sentiment scores do not sum to 1, meaning an article can potentially emit weak signals of any sentiment type or strong signals of multiple types. Descriptive statistics of the sentiment scores (unscaled) across all news articles are provided in *Table 3.3* and *Figure 3.2*. Notable observations from the table include that most articles were considered neutral, and the general sentiment was more negative than positive. In terms of the informational content for the nowcasting model, it is promising that the standard deviation of the negative and positive scores are that large compared to the corresponding means.

Table 3.3: Descriptive statistics of sentiment scores across all news articles.

Statistic	Negative	Neutral	Positive
Mean	0.057	0.925	0.017
Median	0.054	0.932	0.014
25th Percentile	0.043	0.912	0.008
75th Percentile	0.068	0.946	0.022
Std. Deviation	0.022	0.03	0.014
Max	0.163	0.984	0.112
Min	0.014	0.781	0.001

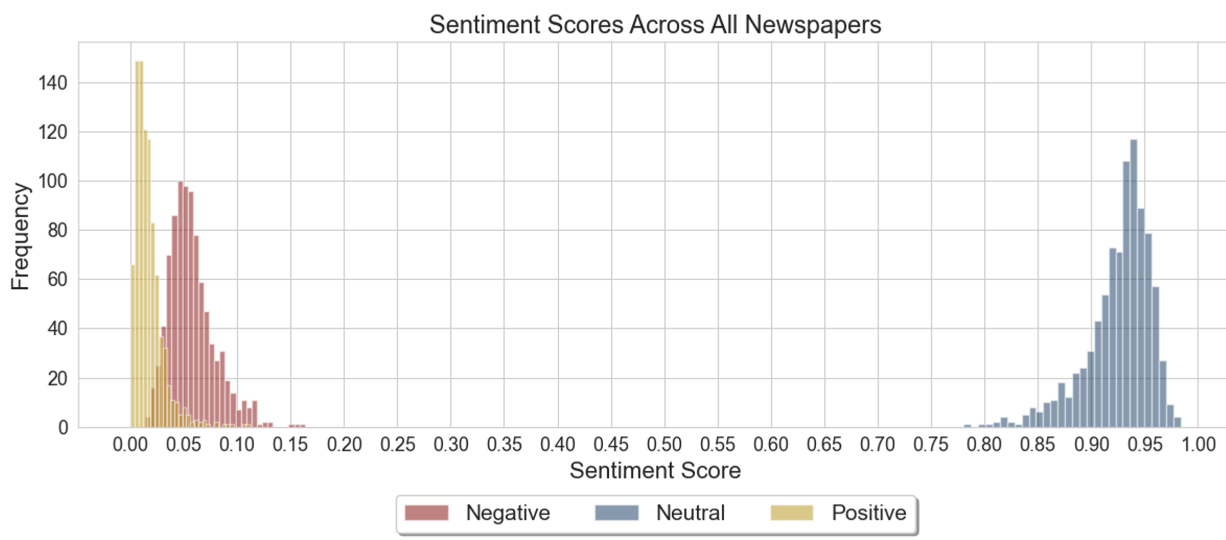

Figure 3.2: Histogram of sentiment scores across all news articles.

3.7 Deriving textual variables with zero-shot classification

Our selection of candidate labels for the zero-shot classification was based on the considerations typically made by leading economic policy organizations such as the OECD and ECB, when they create their economic outlook analyses (OECD, 2023d; IMF, 2023b). These candidate labels are provided in *Table 3.4*, and span categories including labor markets, financial markets, international trade, and consumption. Additionally, extraordinary events such as wars, pandemics, important elections, and recessions can have a significant effect on GDP growth and cause sudden shifts (Gourinchas, 2023). To account for this, a candidate label for each of the mentioned events is also included. All candidate labels, and corresponding, unscaled descriptive statistics across all news articles, are presented in the table below.

Table 3.4: Zero-shot classification candidate labels and their descriptive statistics.

Candidate Label	**Mean**	**Median**	**Std. Dev.**	**25th Perc.**	**75th Perc.**	**Max**	**Min**
Labor market	0.509	0.51	0.063	0.464	0.552	0.694	0.317
Unemployment	0.317	0.314	0.063	0.275	0.355	0.516	0.131
Wage growth	0.408	0.41	0.055	0.372	0.442	0.583	0.247
Inflation	0.377	0.375	0.044	0.346	0.406	0.525	0.263
Interest rate	0.388	0.393	0.05	0.357	0.424	0.544	0.232
Stock market	0.317	0.318	0.04	0.293	0.343	0.454	0.161
Credit market	0.417	0.418	0.052	0.382	0.45	0.607	0.269
Import	0.864	0.866	0.017	0.856	0.876	0.903	0.758
Export	0.632	0.628	0.048	0.599	0.665	0.801	0.502
Currency	0.468	0.467	0.048	0.435	0.498	0.663	0.335
Consumption	0.613	0.615	0.047	0.584	0.646	0.77	0.441
Energy prices	0.358	0.357	0.053	0.324	0.389	0.544	0.218
Housing market	0.389	0.387	0.051	0.354	0.42	0.573	0.26
Election	0.44	0.439	0.065	0.4	0.48	0.649	0.231
Pandemic	0.375	0.356	0.089	0.323	0.384	0.83	0.233
War	0.384	0.373	0.09	0.317	0.451	0.876	0.18
Recession	0.365	0.364	0.052	0.33	0.399	0.519	0.209
Economic growth	0.506	0.507	0.047	0.478	0.537	0.647	0.349
GDP	0.537	0.536	0.048	0.504	0.57	0.677	0.401

* Zero-shot classification scores range from 0 to 1 and correlate positively with the relevance of a label to a text.

Several candidate labels have corresponding macroeconomic variables, with examples including unemployment versus unemployment rate, interest rate versus monetary policy rate, and consumption versus CCI. We incorporated both textual and macroeconomic versions of these variables to explore whether the textual version could provide information as a high-frequency alternative or complement to its macroeconomic counterpart. The inclusion of both textual and macroeconomic versions of various variables also enables examining how the importance of these variables may shift in different economic contexts, potentially adding a deeper understanding of the relationship between these variables and GDP growth.

3.8 Accounting for time-based trends in sentiment and classification data

In order to capture nuances of how the sentiment scores and zero-shot classification labels have changed in the period preceding the nowcast we have calculated weekly, monthly and quarterly averages. For instance, in the case of the Negative Sentiment variable, we compute the average negative sentiment in the articles published over the past week, month, and quarter for each new nowcast. The weekly average provides the most recent information and is more apt to change between the weekly nowcasts because the entire article base is renewed. The quarterly averages describe the long-term trend and are only marginally changed from nowcast to nowcast. The monthly averages capture the sentiment with less impulsivity than the weekly and more flexibility than the quarterly. Including variables for different time windows also enable comparatively analyzing their time-dependent informational value. To align with the sentiment and classification labels, we have also applied the weekly, monthly, and quarterly averages to the newspaper weighting variables. Altogether, we have 9 sentiment variables (3 unique sentiment types), 57 zero-shot classification variables (19 unique candidate labels), and 12 newspaper weighting variables (4 unique newspapers).

4. Machine Learning Approach

4.1 Using an ensemble of neural network models

We considered neural networks (NNs) appropriate for nowcasting GDP growth due to their capacity to capture complex, non-linear correlations between economic indicators and GDP growth. Unlike traditional linear regression, NNs adeptly handle these intricate interdependencies. This particularly applies to multilayer perceptrons, which is the NN type used in this thesis, making them ideal for modeling and predicting economic trends vital for accurate GDP growth nowcasting. In the analysis covered in *Chapter 5*, a set of NN candidates is created with varying configurations. Information about the configurations is detailed in *Table A.2* in *Appendix A*.

In addition to having multiple benefits, NNs pose certain challenges. Their 'black box' nature complicates the interpretability of predictions, which is a significant issue in economic forecasting where understanding prediction rationale is vital. To address this, aggregated Shapley values from each NN in the ensemble are discussed in *Chapter 6*.

Another challenge is that NNs demand extensive, quality data for effective training. Our data set consists of only 66 labeled observations for each country, including 4 quarters in each year between 2007 and 2022 and 2 quarters in 2023. This data limitation might affect the models' effectiveness by increasing prediction variance. To mitigate this, an ensemble of NNs is used to produce predictions. Ensemble methods are generally used to reduce the variance of models with a tendency to overfit, by averaging the predictions of the included models into a single output (Makhijani, 2020). In the context of this thesis, using an ensemble mitigates the risk of a model being skewed by noisy variance. Another benefit of using an ensemble is to reduce the impact of randomness in the initialization of the NN models.

4.2 Describing the nowcasting algorithm for weekly GDP growth predictions

The primary objective of the NN ensemble is to offer weekly estimates of GDP growth. To achieve this, we have created a training and prediction scheme that simulates real-time nowcasts by only using independent variables available at the nowcasting time and training chronologically to avoid fitting the model to future events. Thus, it will be able to provide realistic insights into how effective the nowcasting algorithm would be in each nowcast. The scheme, referred to as the nowcasting algorithm, incorporates weekly predictions and quarterly training by chronologically processing the data in quarterly sections. The process is tailored to this thesis but draws inspiration from Woloszko's (2020) method for bridging the gap between quarterly and weekly prediction.

The process is illustrated and described in pseudocode in *Figure 4.1* and is repeated for each of the NNs that are candidates to be included in the ensemble. Each quarterly data section consists of the weekly, unlabeled observations first and the labeled observations for the corresponding quarter last. Thus, for each data section, the model makes predictions on all data in the current section, unlabeled and labeled. Then it is trained on the labeled data from that section. This process repeats for all data sections with only two exceptions. The model does not make predictions when processing the first data section, because it is not yet trained on any data. When processing the last data section, the model is not trained because there are no more predictions to make.

```
For each candidate model:
  For each data section:
    If first data section:
      1) Store unscaled data section
      2) Create Standard Scaler
      3) Scale labeled data from data section with Standard Scaler
      4) Train model on scaled, labeled data from data section
    If not first data section:
      1) Store unscaled data section
      2) Scale labeled and unlabeled data from data section with old Standard Scaler
      3) Create predictions on scaled, labeled and unlabeled data from data section
      4) Store predictions
      5) Create new Standard Scaler
      6) Use stored, unscaled data to fit and update the Standard Scaler for the next iteration
      If not last data section:
        1) Train model on labeled data from data section
```

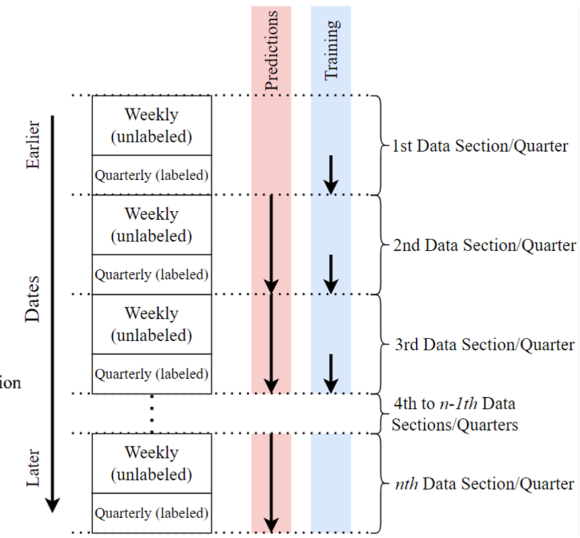

Figure 4.1: Pseudocode of the nowcasting algorithm and a visual representation of the process.

We use a standard scaler to scale the independent variables in each data section before the model makes predictions. Scaling was incorporated because NNs are sensitive to the scale of the input variables and our data set includes variables with highly different value ranges. The benefit of using standard scaling in particular is that it does not treat extreme values as outliers, thus allowing for better performance around downturns, which is beneficial in our use case (Woloszko, 2020).

We created 15 candidate NNs with different numbers of layers and neurons per layer, and when all candidates had made predictions, we computed their out-of-sample mean squared error (MSE) scores on the quarterly, labeled data. Then we selected the four NNs with the lowest overall MSE scores to constitute the ensemble model. The predictions of the ensemble were the mean predictions of the NNs included in the ensemble. It is important to recognize that since GDP values are released only quarterly, it is impossible to verify the accuracy of weekly estimates made between these quarterly releases. However, the primary objective of the ensemble model is not to ascertain the exactness of these interim estimates. Instead, the model aims to offer an informed estimate of the current GDP, drawing on the most recent quarterly macroeconomic data, as well as news data from the past week, month, and quarter.

4.3 Advantages and disadvantages of grouping countries during training

Examining the GDP growth curves of the countries in this study shows that some countries exhibit similar characteristics and trends, whereas others display distinct patterns. The curves are displayed in *Figures 4.2* and *4.3*. Addressing this diversity in GDP growth trajectories is crucial in optimizing the nowcasting algorithm.

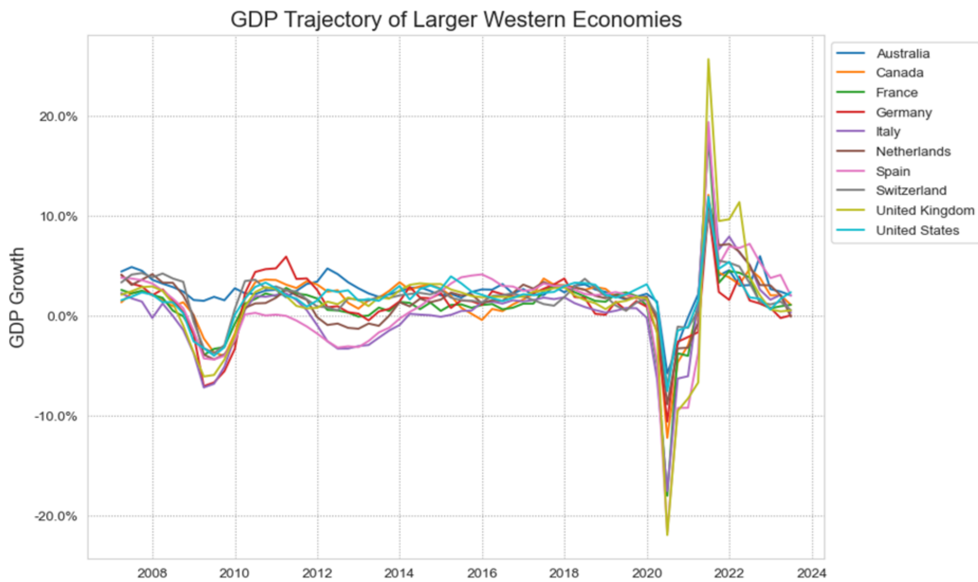

Figure 4.2: GDP growth curves of larger Western economies.

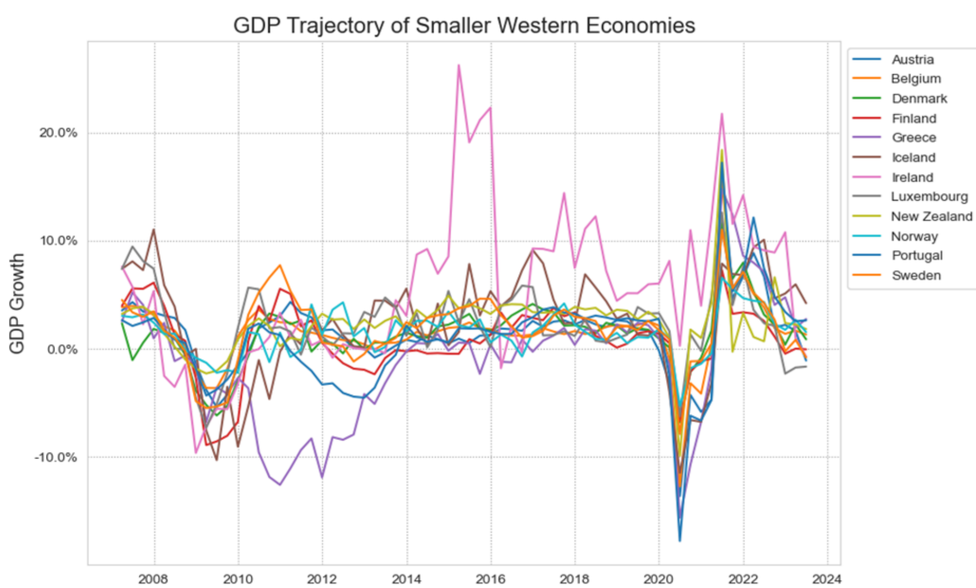

Figure 4.3: GDP growth curves of smaller Western economies.

One strategy is to include data from all 22 observed countries during the training of the nowcasting algorithm. This maximizes the amount of training data, but also the variance in the target variable. Thus, the nowcasting performance depends on the model's ability to distinguish between country-specific patterns and general trends. Woloszko (2020) points out that this approach might still lead to insufficient data for accurate differentiation in GDP growth trends. This is because higher variance further increases the need for training data, which implies that the benefit of maximizing training data could be outweighed by the disadvantage of maximizing variance.

Another strategy involves grouping countries with similar GDP growth trajectories during the training process. This mitigates the variance in the target variable but further reduces the training data. Grouping countries could be beneficial if the model is more affected by country-specific variance than the reduction in training data. An example could be classifying countries as either larger or smaller Western economies. The larger Western economies typically share major trends, except during events like the European debt crisis between 2012 and 2014, which mainly affected southern European nations (Kenton, 2021). Conversely, the smaller Western economies exhibit more diverse growth patterns. An example is Ireland's sharp GDP growth in 2015, driven by an influx of businesses attracted to favorable tax policies and minimal bureaucracy (Chapman, 2016).

In this thesis, we assess the impact of grouping countries on nowcasting performance, focusing on whether it enhances accuracy for larger Western economies with similar GDP patterns. Grouping might improve model performance for these countries, provided there is enough data. Conversely, the diverse economic profiles of smaller Western economies suggest that grouping may not significantly outperform a non-grouping method. *Chapter 5* offers a detailed comparison between the two strategies to ascertain the most effective nowcasting approach.

5. Findings: Evaluating the Nowcasting Algorithm

5.1 Assessing the efficacy of the nowcasting algorithm without grouping countries

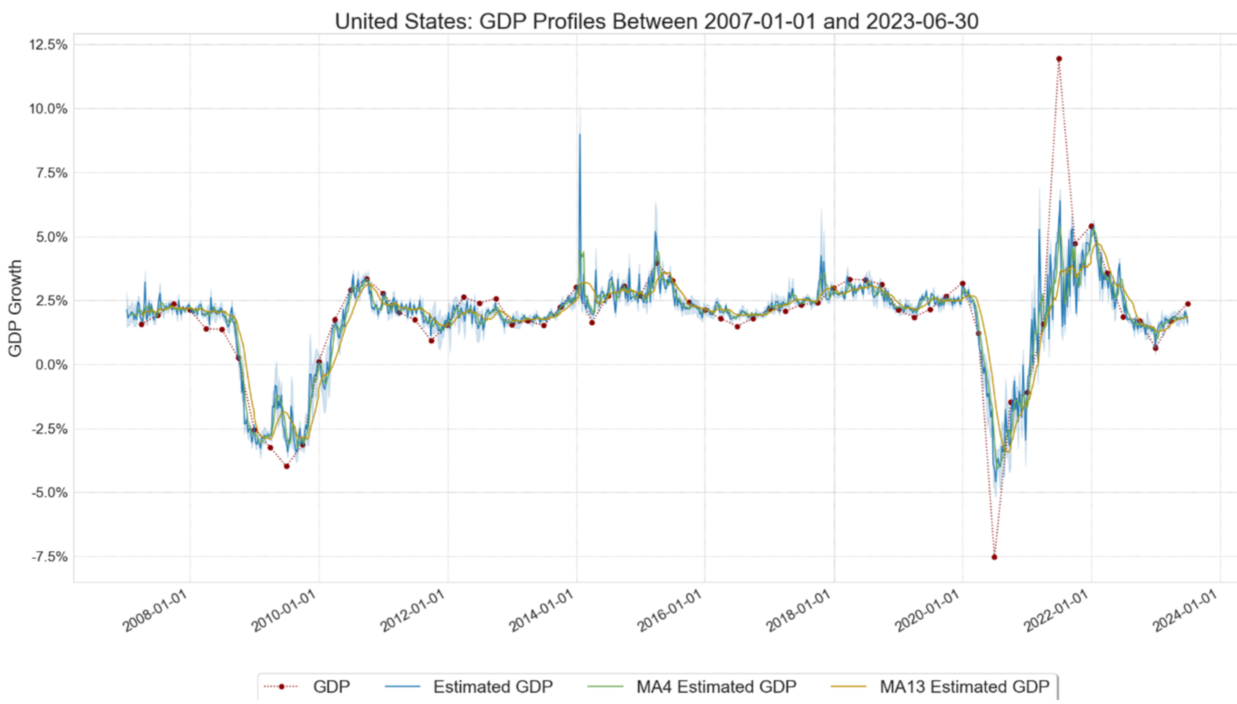

Figure 5.1: Nowcasts without grouping countries with monthly- and quarterly smoothed lines, United States.

Figure 5.1 depicts the GDP growth predictions of the nowcasting algorithm that was trained without grouping countries exemplified using the United States. This figure illustrates most of the tendencies we observed across all countries, and particularly the larger economies. Firstly, the nowcasting algorithm displays a promising ability to capture the general shape of the GDP growth curve. Although the prediction variance is high, causing the prediction line to fluctuate notably between nowcasts, the trend of the predictions rarely deviates persistently from the interpolated line between the recorded GDP growth values. The green and yellow lines show a 4-week (monthly) and 13-week (quarterly) moving average, respectively. These lines are included as the informational value of a single nowcast is diminished by the high variance. Thus, the nowcasting algorithm's predictions are more informative when considered in the context of a set of preceding predictions than when they are considered as stand-alone values.

During the interval between the 2008 financial crisis and the COVID-19 pandemic, the nowcasts are particularly well fitted to the GDP growth curve. This is unsurprising, considering how stable both the GDP growth values and the variables are in this period. The macroeconomic variables have particularly lower variance during "normal" times compared to during crises, and that is the case across all countries. A notable deviation is the spike that occurred around 2015. This spike, and a following period of increased variance, is observed for a majority of the countries. This interval perfectly coincides with an interval of particularly low values of the zero-shot classification variable, *import* (weekly). This variable has similar values in the entire dataset apart from this exact interval. This reflects how volatile the nowcasting can be to variation in the independent variables.

The model captures the trough of the 2008 financial crisis well, but there is a persistent deviance from the interpolation line of the GDP growth data. This is a feature that repeats in the nowcasts for most countries. One potential explanation could be that the model has relatively little training data at that point in the nowcast, making it less proficient at interpreting the relationship between the independent variables and the target variable. This deviance is considered a larger weakness than transient spikes, following the rationale that the predictions should be considered in the context of a set of previous predictions. Furthermore, it is an important weakness considering the increased importance of accurate GDP growth estimates during crises. In *Figure 5.1*, we can see that both the quarterly- and monthly- moving averages also deviate from the interpolation line due to the persistence of the errors. On the other hand, both the descent and rebound during this crisis are well-modeled, and the issue seems to be related to maintaining the low GDP growth predictions.

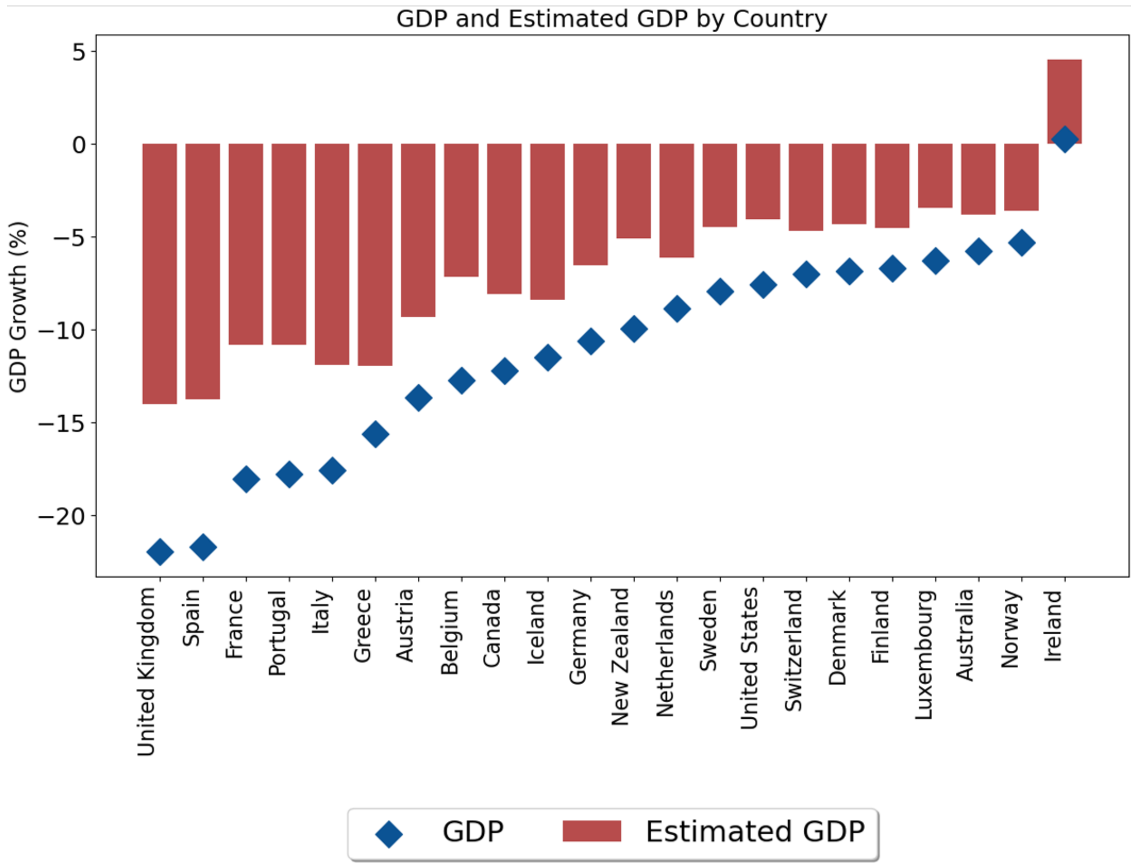

Figure 5.2: GDP growth values and nowcasts in Q2 2020.

During the COVID-19 pandemic, the nowcasts are not able to capture the depth of the trough as well as during the 2008 financial crisis. However, this is not surprising considering the unprecedented size of that shift combined with its sudden nature. Despite this, the nowcasts do capture a promising portion of the depth, and the beginning of the rebound is generally well-modeled also. *Figure 5.2* illustrates how well the model captured the deepest point during the COVID-19 pandemic, which was the second quarter of 2020 for most countries. Regarding the GDP growth peak following the COVID-19 pandemic, the nowcasting model performs generally poorly. Across most countries, the model reverts to nowcasting values resembling those in between crises of 0-5%. This could be explained by that shift being even more unprecedented than the negative shift of the COVID-19 pandemic. This peak is observable in most countries and the general trend is that it is significantly higher than the other GDP growth observation.

5.2 Comparing the nowcasting efficacy with and without grouping countries

Figures 5.3 and *5.4* depict the GDP growth predictions of the nowcasting algorithm for Germany without and with grouping countries during training, respectively. These figures effectively illustrate the differences between the grouping and no-grouping approach that we observed across most of the countries covered in this thesis. Our first observation is that the curves display a highly similar ability to capture the trajectory of the GDP growth curve. Additionally, the curve using the grouping strategy is slightly less volatile. The reduced variance is a trend we observed across most larger economies. This could be because of a reduction in country-specific trends being falsely attributed to other countries. This seems plausible, considering the similarity of the GDP curves of the larger economies and the distinguished nature of the smaller ones. Despite the reduced variance, we cannot conclude that one grouping strategy is notably better than the other, especially considering the potential of subjective bias when the results are this similar. Instead, we conclude that both strategies show promising capabilities of capturing the general GDP growth curve, with the strategy of grouping countries showing an indication that it could be beneficial when nowcasting for larger economies.

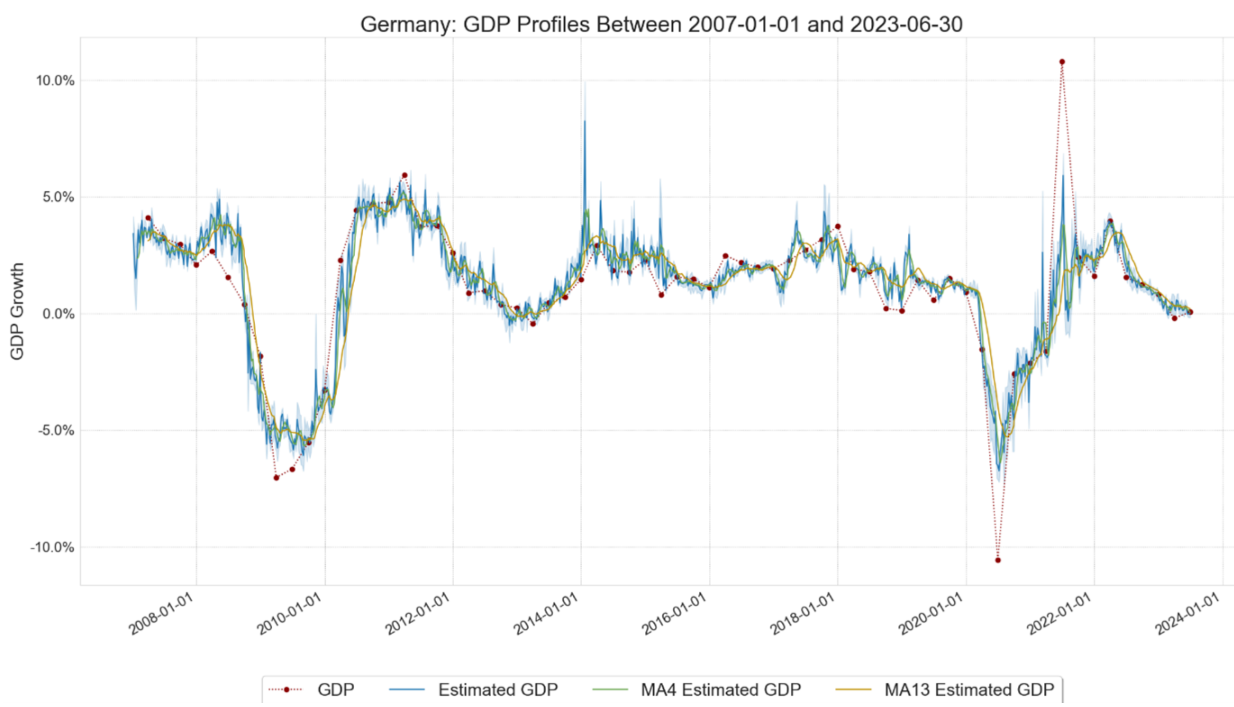

Figure 5.3: Nowcasts without grouping countries with monthly- and quarterly smoothed lines, Germany.

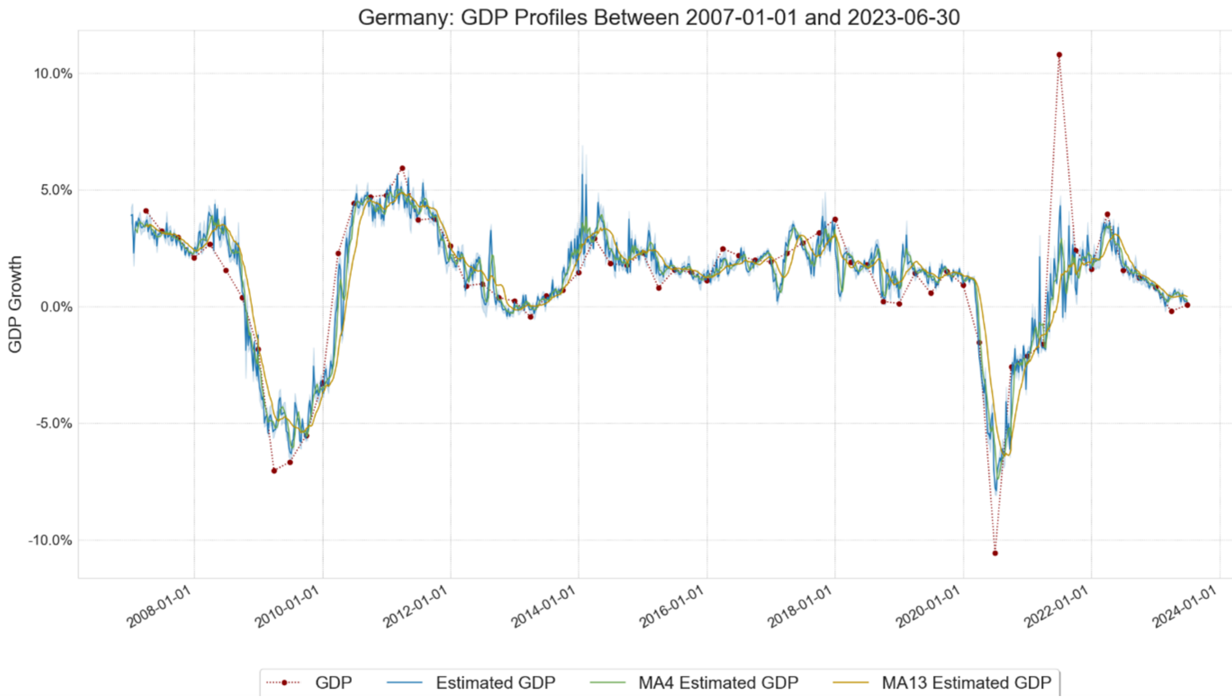

Figure 5.4: Nowcasts when grouping countries with monthly- and quarterly smoothed lines, Germany.

Table 5.1 contains the out-of-sample MSE scores across the entire timespan and during “normal” times (2011 – 2019). The larger economies are below the separation line, the smaller economies are above, and the countries we gathered news articles from are highlighted in grey. The aggregated mean MSE scores can be found at the bottom of the table. Studying the MSE scores of the nowcasting algorithm, the trend is clear that grouping countries during training could be beneficial in nowcasting for larger economies, while the opposite is the case for relatively smaller economies. However, despite the clear pattern, we must note that the individual differences are generally low.

Table 5.1: MSE scores for the grouping strategies, 2007-2023 and 2011-2019.

	All Dates		**Normal Times**	
Country	**MSE: No Group**	**MSE: Group**	**MSE: No Group**	**MSE: Group**
Austria	**1.824**	1.894	**0.258**	0.517
Belgium	**2.012**	2.226	**0.315**	0.343
Denmark	**0.704**	1.208	**0.284**	0.59
Finland	**1.002**	1.129	**0.415**	0.791
Greece	**1.591**	3.67	**0.563**	0.82
Iceland	**1.691**	1.814	**1.481**	1.865
Ireland	5.448	**3.996**	4.122	**3.461**
Luxembourg	**2.126**	2.171	**1.467**	1.797
New Zealand	2.383	**2.111**	**0.251**	0.34
Norway	**0.365**	0.811	**0.417**	0.732
Portugal	3.543	**2.579**	**0.434**	0.643
Sweden	**1.148**	1.383	**0.473**	0.636
Australia	**0.578**	0.678	0.163	**0.135**
Canada	1.106	**1.089**	0.205	**0.105**
France	3.138	**2.188**	0.222	**0.163**
Germany	1.449	**1.239**	0.332	**0.263**
Italy	2.598	**1.857**	0.531	**0.144**
Netherlands	**0.848**	1.297	0.157	**0.118**
Spain	4.082	**2.749**	0.316	**0.279**
Switzerland	**0.65**	0.84	0.165	**0.082**
United Kingdom	4.776	**3.512**	0.469	**0.32**
United States	0.951	**0.894**	**0.1**	0.127
Avg. MSE: Small	***1.602***	*1.886*	***0.564***	*0.81*
Avg. MSE: Large	*1.544*	***1.427***	*0.234*	***0.158***
Avg. MSE: All	***1.575***	*1.661*	***0.378***	*0.385*

During the COVID-19 pandemic, the grouping strategy generally displays an enhanced ability to capture more of the depth of the trough. This especially applies to the larger economies, but we observe this trend for some of the smaller economies also. This is illustrated in *Figure 5.5*, where the predictions with grouping are displayed with red bars and without grouping are shown in yellow (the bars overlap and are not stacked). Another trend to note from the figure is that grouping countries seems to enhance the model's ability to nowcast more extreme negative shifts. We could not find similar patterns of a grouping strategy performing notably better during the 2008 financial crisis.

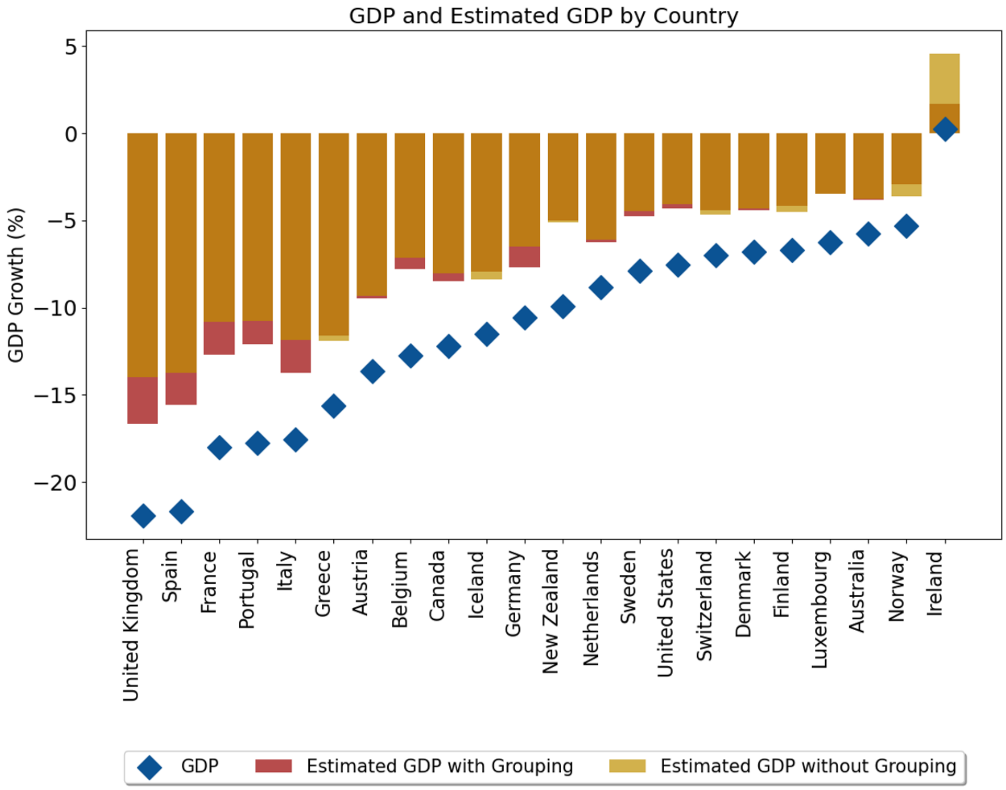

Figure 5.5: Comparing GDP growth values and nowcasts with and without grouping countries in Q2 2020.

Table 5.2 contains the out-of-sample MSE scores during the 2008 financial crisis and the COVID-19 pandemic. According to the MSE scores, grouping countries is beneficial for nowcasting GDP growth for the majority of the countries, regardless of their economic size. During the COVID-19 pandemic, we encounter the familiar pattern that grouping countries is advantageous for larger economies and oppositely for smaller economies. We note that the differences in MSE during the COVID-19 pandemic are generally more substantial than during the other periods analyzed, which gives a stronger indication of the proficiency of the grouping strategy.

Table 5.2: MSE scores for the grouping strategies, 2008 financial crisis and COVID-19 pandemic.

	Financial Crisis		**COVID-19**	
Country	**MSE: No Group**	**MSE: Group**	**MSE: No Group**	**MSE: Group**
Austria	0.201	**0.178**	7.707	**7.399**
Belgium	0.244	**0.203**	**8.397**	9.381
Denmark	0.344	**0.286**	**2.197**	3.848
Finland	1.838	**1.078**	**1.557**	2.058
Greece	**1.316**	3.471	**4.548**	11.226
Iceland	1.429	**0.633**	**2.53**	3.03
Ireland	**2.34**	2.649	12.409	**6.911**
Luxembourg	1.668	**1.421**	4.344	**3.99**
New Zealand	0.227	**0.11**	10.328	**8.951**
Norway	**0.07**	0.08	**0.569**	1.85
Portugal	**0.231**	0.249	15.322	**10.222**
Sweden	1.259	**0.93**	**2.756**	3.821
Australia	0.087	**0.049**	**2.205**	2.794
Canada	0.337	**0.256**	**4.302**	4.572
France	0.226	**0.17**	13.961	**9.699**
Germany	1.11	**0.593**	4.707	**4.488**
Italy	0.426	**0.355**	10.393	**7.977**
Netherlands	**0.182**	0.274	**3.386**	5.498
Spain	0.181	**0.169**	18.223	**12.05**
Switzerland	**0.251**	0.293	**2.356**	3.413
United Kingdom	0.248	**0.154**	21.025	**15.557**
United States	0.135	**0.085**	4.071	**3.79**
Avg. MSE: Small	*0.565*	***0.489***	***4.243***	*5.132*
Avg. MSE: Large	*0.248*	***0.195***	*6.175*	***5.987***
Avg. MSE: All	*0.388*	***0.322***	***5.032***	*5.504*

In evaluating the efficacy of the nowcasting algorithm's predictions, both quantitative and qualitative analyses fail to conclusively differentiate between the grouping strategies. However, there is evidence that grouping countries can be beneficial when nowcasting for larger economies both during crises and normal times. It is difficult to find a similar pattern for the smaller countries, as the optimal grouping strategy depends on the country and economic context. During crises most smaller countries benefit from being grouped together, while the opposite is the case during normal times. Consequently, the optimal grouping strategy depends on which countries is being nowcasted for and either strategy could be used in the following analysis. The following discussion will be based on the strategy of grouping similar economies. However, that does not conclude that it is a better option in general.

5.3 Comparing the nowcasting algorithm with an autoregressive baseline model

We used an autoregressive (AR) model to set a baseline performance to compare the nowcasting algorithm's efficacy against. AR models are statistical models that are based on the principle that historical values in a series have a significant influence on future outcomes. These models are usually denoted as AR(p), with *p* denoting the number of lagged observations used. Thus, an AR(4) model, which is the one we used in this thesis, makes predictions based on the 4 previous data points. In our case, those data points comprise the preceding year, because we have quarterly observations. In the context of economic forecasting, using an AR model as a baseline makes sense for several reasons. First, economic data often exhibit significant temporal dependencies, meaning past values can be strong predictors of future values. Furthermore, the AR framework offers a straightforward and interpretable approach, allowing for easy comparison with more complex models.

Figure 5.6 shows the GDP growth predictions for France of the AR(4) and nowcasting algorithm with monthly- and quarterly smoothed lines. The AR(4) predictions capture the general shape of the GDP growth curve, but it has a prominent lag. This is a feature of the model making predictions based on previous observations of the target variable. During "normal" times, and particularly in the interval between 2012 and 2017, the performance of the AR(4) model is indistinguishable from the nowcasting algorithm. Both models follow the GDP growth curve tightly. However, the AR(4) consistently underestimates GDP growth records during the last three years leading up to the COVID-19 pandemic. The persistence of this deviation makes it a distinguished weakness of the AR(4) model.

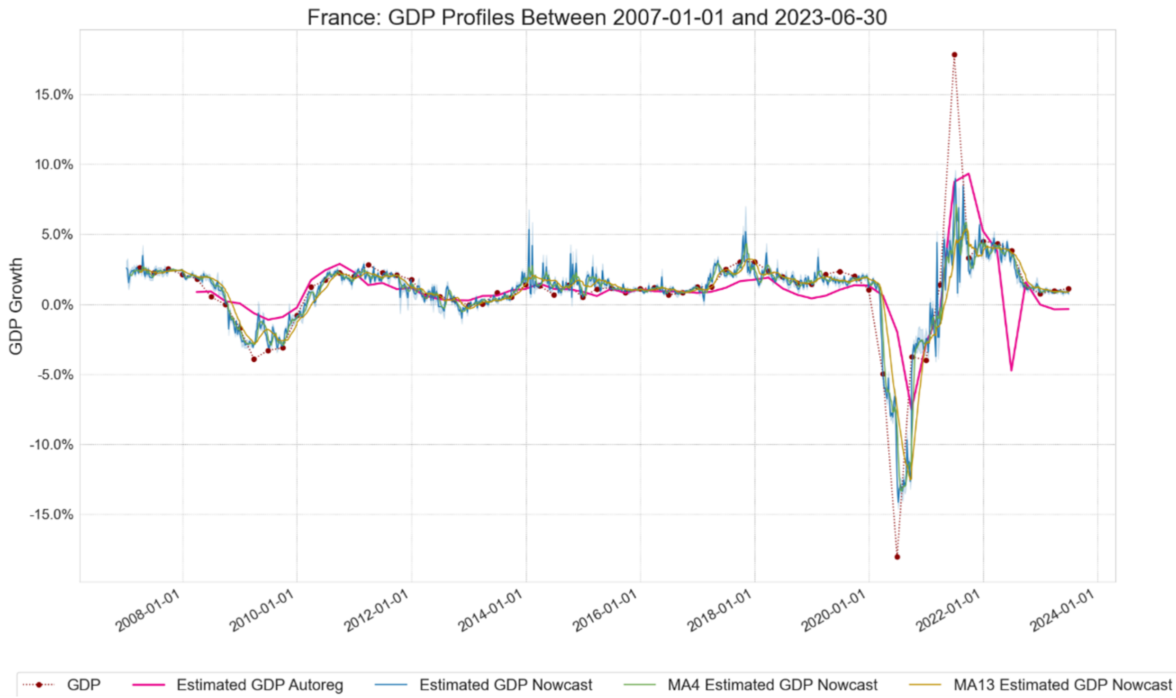

Figure 5.6: Autoregressive predictions and nowcasts when grouping countries with monthly- and quarterly smoothed lines, France.

The MSE scores of the nowcasting algorithm and the AR(4) during the entire timespan and during “normal” times (2011 – 2019) are displayed in *Table 5.3*. We can clearly see that the nowcasting algorithm outperforms the AR(4) model in terms of precision. The differences in MSE are substantial in most instances, particularly when considering the entire timespan. During “normal” times, the fluctuations in GDP growth are not large enough for the lagged nature of the AR(4) model to cause issues. Consequently, the MSE scores for both models are relatively low in this timespan. Thus, the findings from the MSE analysis reflect the visual examination of the predictions.

Table 5.3: MSE scores for nowcasting model and AR(4), 2007-2023 and 2011-2019.

	All Dates		Normal Times	
Country	**MSE: Autoreg**	**MSE: Group**	**MSE: Autoreg**	**MSE: Group**
Austria	5.346	**1.894**	**0.422**	0.517
Belgium	5.734	**2.226**	**0.224**	0.343
Denmark	3.152	**1.208**	0.622	**0.59**
Finland	3.044	**1.129**	**0.773**	0.791
Greece	10.906	**3.67**	2.627	**0.82**
Iceland	8.468	**1.814**	5.319	**1.865**
Ireland	25.165	**3.996**	27.878	**3.461**
Luxembourg	4.816	**2.171**	2.17	**1.797**
New Zealand	7.546	**2.111**	1.091	**0.34**
Norway	2.168	**0.811**	0.937	**0.732**
Portugal	10.628	**2.579**	1.851	**0.643**
Sweden	3.582	**1.383**	0.691	**0.636**
Australia	2.297	**0.678**	0.209	**0.135**
Canada	5.189	**1.089**	0.533	**0.105**
France	8.878	**2.188**	0.517	**0.163**
Germany	4.754	**1.239**	0.794	**0.263**
Italy	9.076	**1.857**	0.612	**0.144**
Netherlands	3.761	**1.297**	0.526	**0.118**
Spain	14.218	**2.749**	1.594	**0.279**
Switzerland	2.742	**0.84**	0.231	**0.082**
United Kingdom	16.942	**3.512**	0.601	**0.32**
United States	3.435	**0.894**	0.45	**0.127**
Avg. MSE	*5.865*	***1.661***	*0.888*	***0.385***

During both the 2008 financial crisis and the COVID-19 pandemic, the lagged property of the AR(4) predictions is particularly noticeable. The bottoms of its prediction trough in both crises are situated 1-2 quarters after the actual bottom, described by the interpolated GDP growth observations. Additionally, the AR(4) does not capture the depth of these troughs as well as the nowcasting algorithm. *Figure 5.7* shows a comparison of how the models performed during the second quarter of 2020.

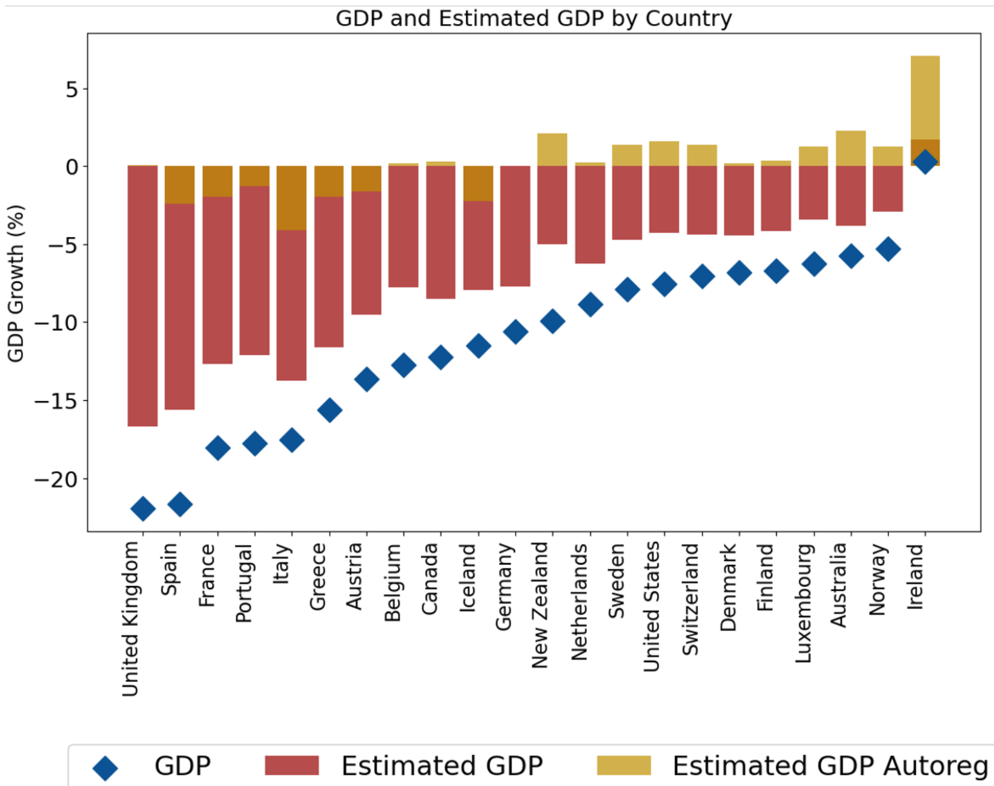

Figure 5.7: Comparing GDP growth values with nowcasts and AR(4) predictions in Q2 2020.

The out-of-sample MSE scores during the 2008 financial crisis and the COVID-19 pandemic are displayed in *Table 5.4*. Following the visual inspection of the GDP growth predictions, the AR(4) model has a significantly higher MSE during crises. This is caused by both a misalignment between its predictions and the GDP growth curve and a failure to capture the size of the shifts. This is a crucial weakness considering the importance of these models during crises. Thus, we conclude, from the visual assessment and quarterly MSE comparison, that our nowcasting algorithm is a superior option for producing weekly GDP growth estimates for the majority of the observed countries during normal times and for all countries during crises.

Table 5.4: MSE scores for nowcasting model and AR(4), Financial Crisis and COVID-19 pandemic.

	Financial Crisis		**COVID-19**	
Country	**MSE: Autoreg**	**MSE: Group**	**MSE: Autoreg**	**MSE: Group**
Austria	2.333	**0.178**	20.59	**7.399**
Belgium	2.003	**0.203**	23.101	**9.381**
Denmark	2.812	**0.286**	9.947	**3.848**
Finland	4.926	**1.078**	7.271	**2.058**
Greece	8.941	**3.471**	33.882	**11.226**
Iceland	11.254	**0.633**	14.176	3**.03**
Ireland	11.264	**2.649**	30.102	**6.911**
Luxembourg	5.826	**1.421**	10.753	**3.99**
New Zealand	5.376	**0.11**	26.003	**8.951**
Norway	2.299	**0.081**	5.22	**1.85**
Portugal	1.794	**0.249**	40.769	**10.222**
Sweden	3.241	**0.93**	11.309	**3.821**
Australia	0.24	**0.049**	9.429	**2.794**
Canada	2.523	**0.256**	19.447	**4.572**
France	2.181	**0.17**	36.118	**9.699**
Germany	4.837	**0.593**	14.864	**4.488**
Italy	3.25	**0.355**	35.834	**7.977**
Netherlands	2.115	**0.274**	13.49	**5.498**
Spain	2.031	**0.169**	57.125	**12.05**
Switzerland	2.747	**0.293**	9.194	**3.413**
United Kingdom	4.679	**0.154**	69.474	**15.557**
United States	3.044	**0.085**	11.448	**3.79**
Avg. MSE	3.172	**0.322**	18.394	**5.504**

6. Discussion: Understanding Variable Importance During Different Economic Conditions

6.1 Explaining important features of Shapley values

For an in-depth analysis of how the independent variables influence GDP growth predictions, during crises and normal periods, we have assessed their Shapley values. One of the advantages of Shapley values is that they facilitate both local and global interpretability. Local interpretability focuses on the explanation of individual predictions, whereas global interpretability is concerned with understanding the model's overall mechanics. Understanding what drives the predictions made by the neural network is crucial to ensure that the model is consistent with economic reasoning and does not depend on random patterns.

Shapley values, derived from cooperative game theory, provide a method for fairly distributing the "payout" or outcome among participants, which are analogous to variables in predictive modeling. These values quantify the contribution of each variable to the model's overall prediction. For a variable i and a subset of variables S excluding i, the marginal contribution $\Delta_i(S)$ is denoted as:

$$\Delta_i(S) = \text{Prediction with}(S \cup \{i\}) - \text{Prediction with } S$$

The Shapley Value $\phi(i)$ is then the average of all those marginal contributions:

$$\phi(i) = \text{Average of } \Delta_i(S) \text{ over all subsets } S$$

This comprehensive approach ensures that the effect of each variable is evaluated in the context of all possible variable interactions, leading to a fair and thorough assessment of its impact on the model's output. In other words, the sum of the Shapley values for all variables equals the total prediction made by the model. Mathematically, if the prediction of the model is P, and there are n variables with Shapley values, $\phi_1, \phi_2, \ldots, \phi_n$, then $P = \phi_1 + \phi_2 + \ldots + \phi_n$. Such a breakdown of predictions into Shapley values enables a detailed, local interpretation of the results. Furthermore, if two variables contribute equally to all possible combinations of variables, their Shapley values will also be equal.

6.2 Uncovering the relevance of text data using Shapley values

Figure 6.1 displays the ten most important variables ranked by Shapley values over the period from 2007 to 2023 in descending order. Each point on the graph represents the Shapley value of a variable to a particular prediction, as denoted by its position along the x-axis. The color range on the right side of the chart corresponds to the variable's value. To detect trends, we look for color patterns on either side of 0 on the x-axis. Color patterns indicate that the variable's effect on GDP growth predictions is dependent on its value.

Figure 6.1 reveals that textual data makes significant contributions to the nowcasting of GDP growth. This is demonstrated by the fact that only two of the ten most important variables are non-textual, namely the *Consumer Confidence Index* and *Hours Worked*. Furthermore, *Hours Worked* is the only traditional hard economic statistic ranking among the 10 most influential variables. This observation supports Stock & Watson's (2023) conclusion that hard indicators often fall short in predicting broader macroeconomic trends due to their inconsistency.

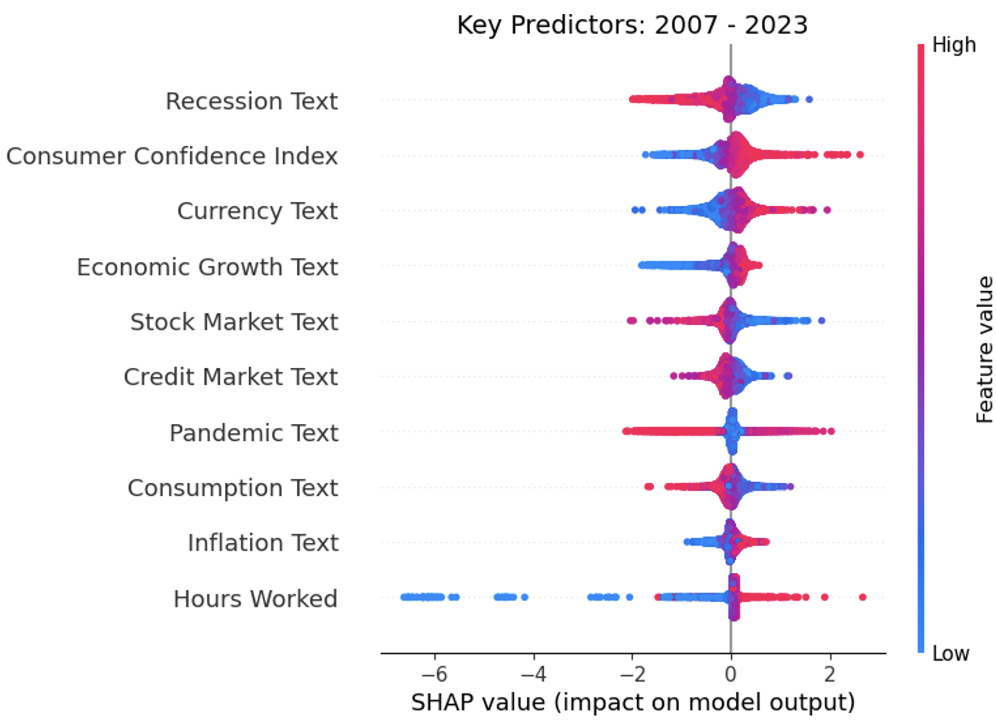

Figure 6.1: Shapley summary plot for the period 2007 - 2023

Our analysis of the Shapley values reveals several instances where the model's predictions are consistent with economic reasoning. For instance, the relationship between *Recession Text* and GDP growth reflects the typical economic impacts of a recession. The plot reveals that increased news coverage of recessions is correlated with a negative impact on GDP growth, while less coverage correlates with higher GDP growth. Similarly, greater values of *Consumer Confidence Index* are associated with increased GDP growth, whereas lower consumer confidence tends to be negatively correlated. This reflects the general trend of consumer sentiment, which is usually pessimistic during economic downturns and optimistic during growth phases. Regarding the variable *Economic Growth Text*, it appears that the lack of media attention towards economic growth tends to have a more pronounced negative correlation with GDP growth, compared to the positive contribution when economic growth is actively discussed in the news. This implies that the lack of media coverage on economic growth topics is a stronger indicator of GDP growth trends than the actual reporting on the subject.

6.3 Key variables during the 2008 financial crisis and COVID-19 pandemic

The model effectively distinguishes crisis periods from more stable times. During the 2008 financial crisis and the COVID-19 pandemic, the Shapley values reveal that the model recognizes specific variables driving GDP growth variations, as shown in *Figure 6.2*. *Recession Text* emerges as an important predictor in both crises, proving its effectiveness across different crisis types. Additionally, certain variables become more prominent during crises. *Pandemic Text*, for instance, gains significance during the COVID-19 pandemic, reflecting the crisis's nature. Similarly, *Hours Worked* takes on increased importance during the COVID-19 pandemic compared to normal periods and the 2008 financial crisis. On the other hand, *Negative Sentiment Text* shows greater relevance during the 2008 crisis than in other times, including the COVID-19 pandemic. The figure also shows a reduced relevance of the *Consumer Confidence Index* during crises, suggesting that frequent textual data takes on increased importance during times of crisis.

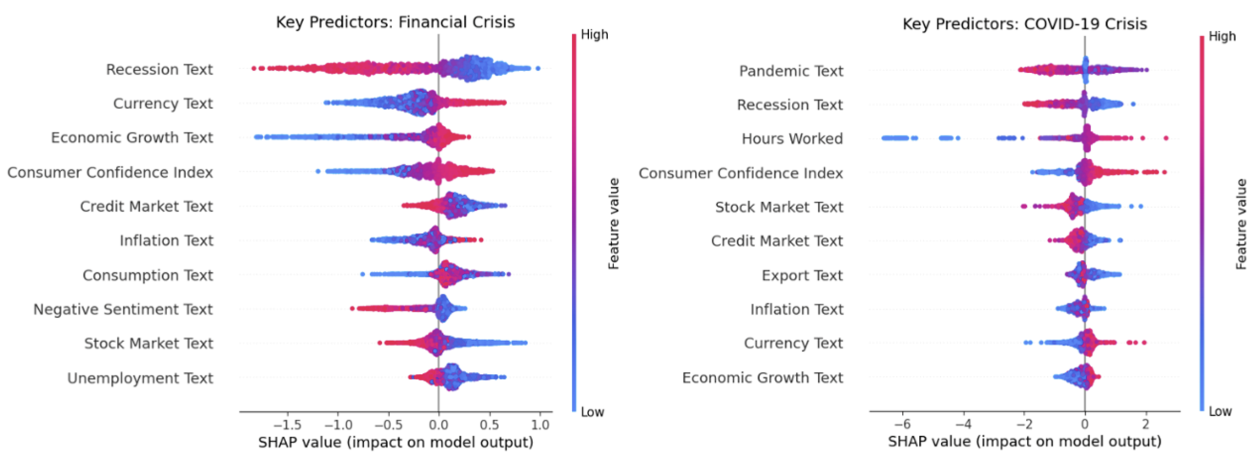

Figure 6.2: Shapley value summary plots for the 2008 financial crisis and COVID-19 pandemic.

The increased importance of the *Hours Worked* variable during the COVID-19 pandemic likely stems from the restrictions that led to widespread temporary business shutdowns, reducing overall working hours. In response, numerous countries implemented support and leave programs for those impacted by these restrictions. This allowed many people to maintain their employment status, even though they were not actively working. As a result, the number of hours worked was drastically reduced while the unemployment rates were not equally affected. This could also explain why neither unemployment variable features among the top 10 variables of the COVID-19 pandemic.

To delve deeper into the factors most influential in capturing the magnitude of our predictions, we examined the key drivers for nowcasting Q2 2020 for Germany, the United States, the United Kingdom, and France (see *Figure 6.3*). The plot reveals that *Hours Worked*, *Pandemic Text*, and *Recession Text* are the key variables driving changes in GDP growth, consistent with the findings from *Figure 6.2*.

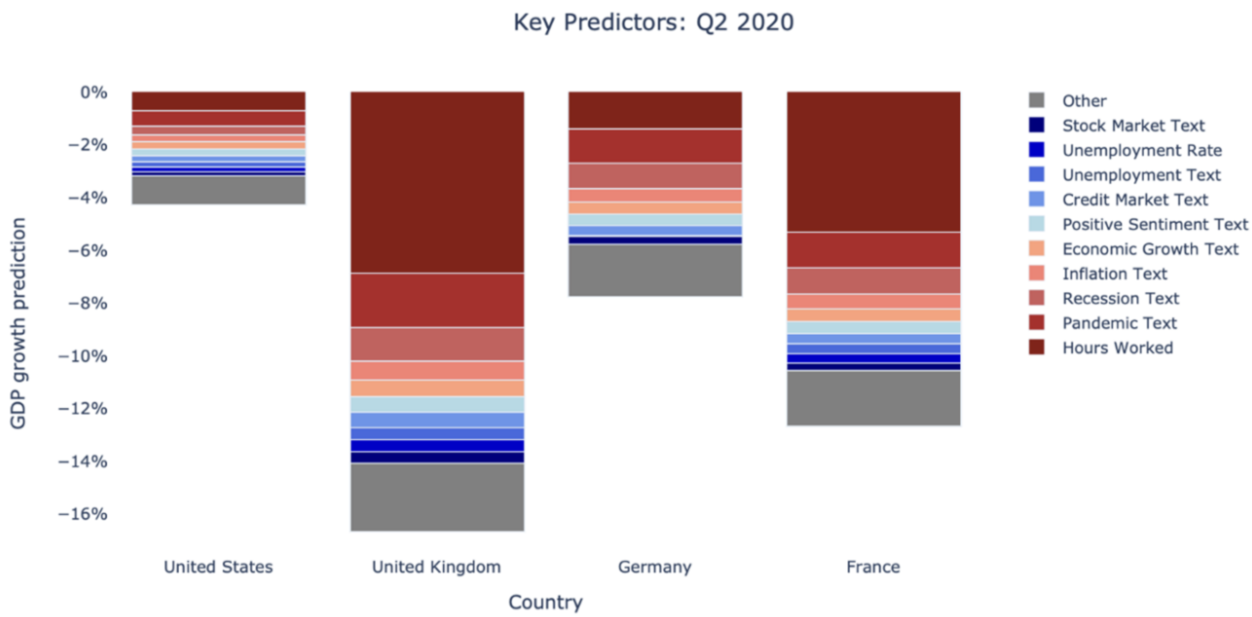

Figure 6.3: Contribution of most important variables for the nowcast of Q2 2020.

6.3.1 Capturing interaction effects can boost nowcasting performance during crises

The model's ability to capture interaction effects aids in its ability to nowcast during the sharp downturns of the COVID-19 pandemic. These effects are displayed in the partial dependency plots in *Figure 6.4*, and can be described as instances where the correlation of one independent variable and the target variable is affected by the value of another independent variable. The figure illustrates the interaction effects between *Pandemic Text* and *Stock Market Text*, *Consumption Text*, and *Consumer Confidence Index*. The x-axis and color scale describe the variable values and the y-axis describes the Shapley value. In all three instances displayed in the figure, we can see that the slope is steeper when the value of *Pandemic Text* is higher. This implies that when the value of *Pandemic Text* is high, changes in the other variables have a larger effect on the GDP growth predictions. For instance, when *Pandemic Text* is higher (red dots), an increase in *Stock Market Text* will lead to a relatively larger negative contribution to the GDP growth prediction. One way to interpret this is that when a pandemic is a more relevant topic in the news coverage, the model is more sensitive to changes in other variables. One possible explanation for this interaction effect could be the speed and magnitude of the pandemic that shifted the economic landscape into a more dynamic state. Thus, including *Pandemic Text*, might have boosted the model's performance during this period by capturing these interaction effects.

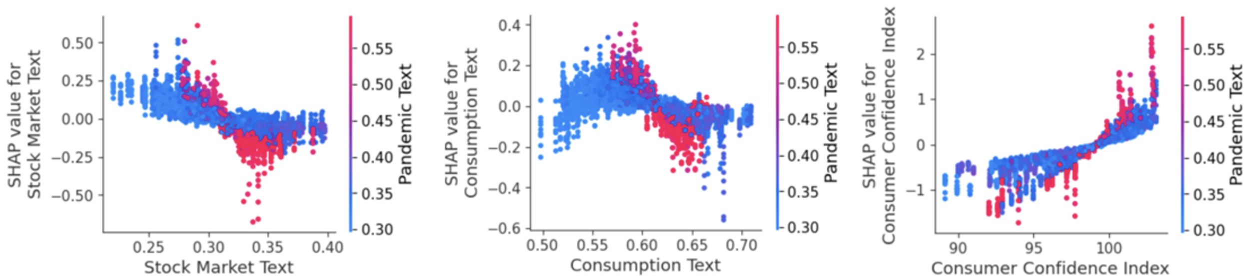

Figure 6.4: Partial dependency plots for Stock Market Text, Consumption Text, and Consumer Confidence Index.

6.3.2 Capturing non-linear relationships also improves nowcasting performance during crises

The model also captures non-linearities between independent variables and GDP growth, which allows the model to effectively differentiate between crisis and other periods. By including the variables *Recession Text* and *Pandemic Text*, the model uncovers non-linear relationships which enable it to adjust the way other variables affect its predictions. *Figure 6.5* displays the non-linear relationship between *Pandemic Text* and its impact on GDP growth, for varying values of *Recession Text*. In scenarios where the relationship is linear, this plot would typically present a straight line. However, due to the neural network's ability to capture non-linear relationships, the plot exhibits a curve that distinctly deviates from a straight-line pattern.

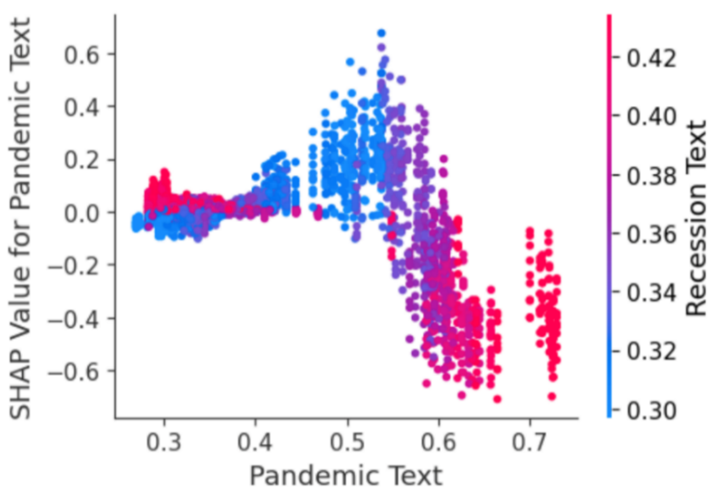

Figure 6.5: Partial dependency plot for Pandemic Text and Recession Text.

The plot suggests that when news focuses relatively more on the pandemic and less on a recession, the effect on GDP growth is positive. This finding likely reflects the pandemic's recovery phase where the economy starts to improve despite ongoing pandemic news. However, when news coverage also focuses on a recession, the pandemic's impact on GDP growth becomes negative. This observation likely reflects the early and most intense stages of the pandemic, characterized by significant economic impacts.

Figure 6.6 reveals that *Hours Worked* exhibits a non-linear relationship with GDP growth when pandemics are a relevant theme in the news coverage (red curvature), while the relationship is linear when it is not (blue line). Interestingly, a reduction in *Hours Worked*, is correlated with a greater impact on GDP growth than a comparable increase. This indicates that while the pandemic-related reduction in working hours is associated with a decline in GDP growth, a corresponding increase does not necessarily lead to a proportional increase in GDP growth. When the value of the *Pandemic Text* is high, only negative changes in *Hours Worked* affect the GDP growth, and this effect is negative. On the other hand, an increase in *Hours Worked* does not alter GDP growth predictions, as indicated by the red curve flattening out. This implies that overcoming the COVID-19 crisis requires more comprehensive measures than increasing working hours. Merely returning people to work is not enough to achieve complete economic recovery.

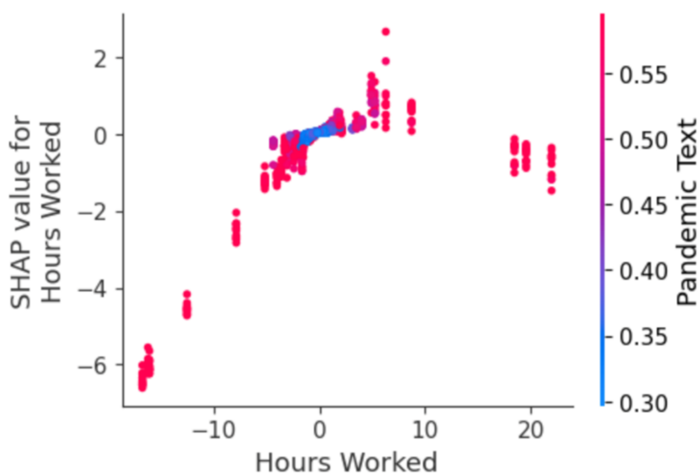

Figure 6.6: Partial dependency plot for Pandemic Text and Hours Worked.

A similar non-linear relationship with GDP growth is also evident for the Economic Growth Text variable. In *Figure 6.7*, the relationship between this variable and GDP growth is expressed for different values of *Recession Text*. For high values of *Recession Text*, the relationship between *Economic Growth Text* and GDP growth is non-linear (red curve), while the relationship is linear when *Recession Text* is low (blue line). This suggests that when the news coverage is relatively more dominated by recession and not economic growth, there is a strong negative effect on GDP growth predictions. On the other hand, when economic growth is a more prevalent topic, the effect on GDP growth is weakly positive. When recession is not a prominent topic, GDP growth is linearly, and positively correlated with the coverage of economic growth, but the slope is not particularly steep.

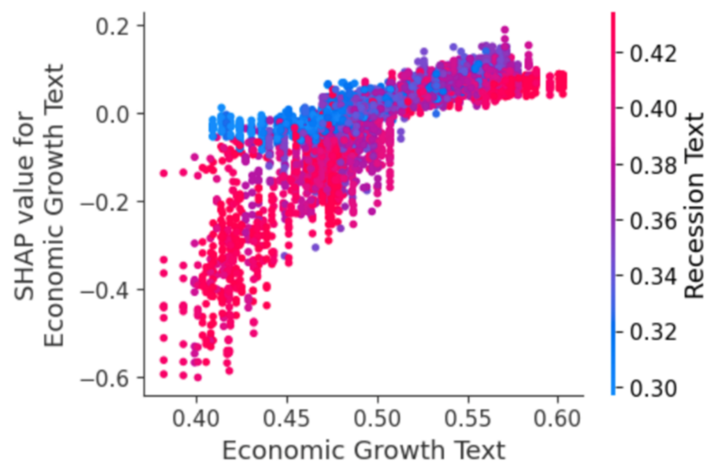

Figure 6.7: Partial dependency plot for Economic Growth and Recession.

6.4 Assessing which timespan provides the most insightful text variables

We assessed the Shapley values of the weekly, monthly, and quarterly variations of the zero-shot classification and sentiment variables, aiming to understand which timespan provided the most important signals when nowcasting GDP growth. *Figure 6.8* includes a bar for each of the variables, and each bar describes the proportion of the aggregated Shapley value belonging to each version of the respective variables.

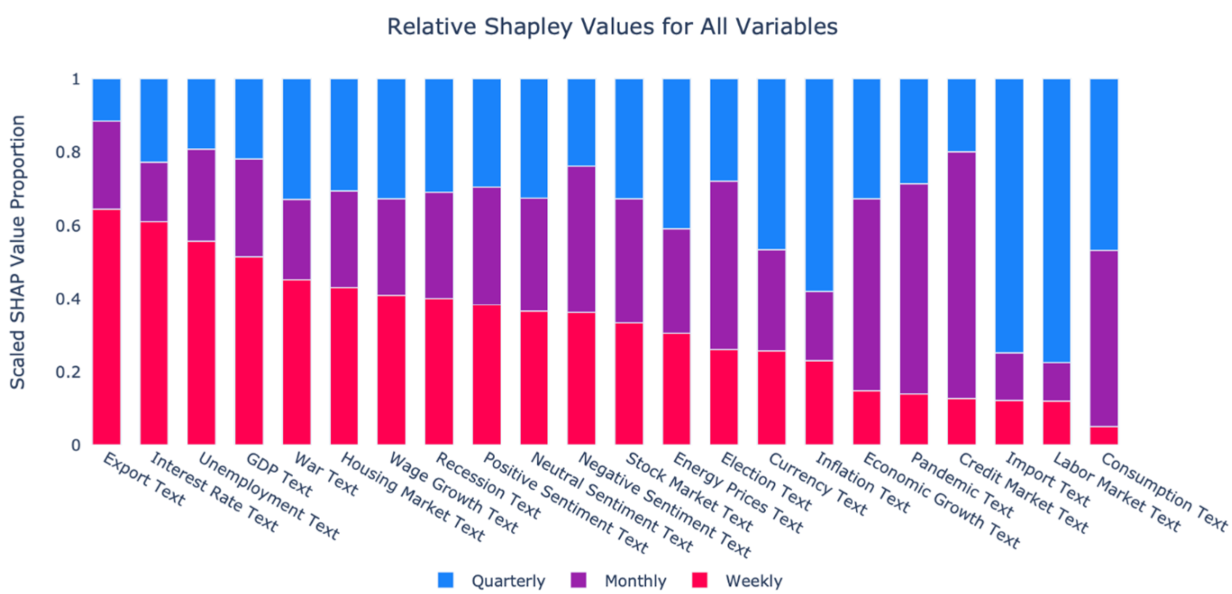

Figure 6.8: Relative Shapley values for all text variables over the period 2007 - 2023.

An important observation is that all timespans seem to contribute to the model's predictions, which implies that the model likely benefitted from including all versions of the text variables. However, we can see from the plot that there are notable individual differences concerning which timespan is most informative. The weekly versions are particularly important for the *Export Text*, *Interest Rate Text*, and *Unemployment Text* variables. This indicates that recency is an important factor in determining the importance of these variables and that longer timespans make the signals outdated. On the opposite side of the scale, variables such as *Consumption Text*, *Labor Market Text*, and *Import Text* seem to be most informative when aggregated over a longer period, and the signals from the shorter timespans might be too volatile to help the model. For the crisis variables, *Pandemic Text* and *Recession Text*, it appears that several timespans provide useful information. With regards to *Recession Text*, all three versions have similar aggregated Shapley values, which indicates that both long-term trends and recency are important factors. For *Pandemic Text*, the monthly version is the most important, which implies that there is a balance between capturing short-term fluctuations and long-term trends. Lastly, it is worth noting that several of these individual differences are difficult to rationalize with economic intuition. For instance, we would expect to see the same timespans to be best for *Import Text* and *Export Text*, considering these are tightly connected terms, but instead, they appear on opposite ends of *Figure 6.8*.

7. Conclusions and Future Research

In this study, we have advanced the field of GDP growth nowcasting using text data by using a unique combination of sentiment analysis, zero-shot classification, and macroeconomic data to generate weekly GDP growth nowcasts. Our results demonstrate that these variables can be leveraged by neural network ensemble to effectively nowcast GDP growth across the 22 largest Western economies in both stable periods and crises. The nowcasting model adeptly handles normal periods and distinguishes itself from the baseline model significantly in how well it captures the dynamics of the Financial Crisis and the initial downturn of the COVID-19 pandemic. On the other hand, the model struggles with capturing the rebound and GDP growth peak following the COVID-19 pandemic, and it exhibits notable variance between nowcasts. This necessitates evaluating nowcasts in the context of previous ones, rather than in isolation. Despite these challenges, our findings are promising and show that leveraging a mix of sentiment, zero-shot classification, and macroeconomic data can be used to estimate weekly GDP growth.

In our research we explored two methods of grouping countries during training, aiming to minimize the misattribution of country-specific variances. Our analysis showed that grouping countries by economic size yielded similar outcomes to ungrouped models. However, evidence suggests that grouping countries aids nowcasting for larger economies in both crisis and normal times. The best grouping strategy for smaller economies varies by country and economic context. Thus, the ideal grouping strategy depends on the specific countries being nowcasted and the economic situation. Regardless of the grouping strategy, the nowcasting model outperformed the AR(4) baseline model both in terms of quarterly MSE and visual analysis.

Our study of Shapley values reveals a key contribution to the existing literature by showcasing how the combination of conventional economic metrics and textual data can be applied for GDP growth nowcasting. A key finding is our discovery of how the interaction between these variables shifts in reaction to changing economic circumstances. By accounting for news content the model is enabled to identify interaction effects and non-linear relationships, which makes it apt to adjust how signals from other variables should be interpreted. For instance, we find that the slope of the relationship between *Consumer Confidence Index* and GDP growth is greater when the news media focuses on pandemic-related topics. On the other hand, we find that the *Consumer Confidence Index* had a relatively lower impact on GDP growth predictions compared to other text variables during crises. This underscores the importance of using high-frequency text data for nowcasting GDP growth during crises. Additionally, we find that when media coverage focuses relatively more on pandemics, the relationship of variables such as *Hours Worked* and *Economic Growth Text* with GDP growth changes from linear to non-linear.

Further research could explore expanding the geographical scope of the news article selection or including multiple newspapers per country, potentially making the information more relevant across all countries and enriching the content with diverse political opinions. Furthermore, including other macroeconomic variables or expanding the set of candidate labels for the zero-shot classification could uncover potentially important dynamics. Further research could also experiment with different variable subsets, which could mitigate noise and lead to the discovery of insightful combinations. Another frontier to explore could be optimizing grouping strategies to enhance model training, as we only cover two rudimentary options in this thesis and found indications that grouping could be beneficial for certain countries. Lastly, one could delve further into optimizing the hyperparameters of the nowcasting model, which could improve its ability to identify connections such as interaction effects and non-linear relationships, thus improving nowcasting.

Bibliography

Aguilar, P., Ghirelli, C., Pacce, M., & Urtasun, A. (2021). Can news help measure economic sentiment? an application in covid-19 times. *Economics Letters*.

Ashwin, J., Kalamara, E., & Saiz, L. (2021). Nowcasting euro area gdp with news sentiment: a tale of two crises. *ECB: Working Paper Series*.

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, (pp. 1593–1636).

Barbaglia, L., Consoli, S., & Manzan, S. (2022). Forecasting gdp in europe with textual data.

Barbaglia, L., Frattarolo, L., Onorante, L., Pericoli, F. M., Ratto, M., & Pezzoli, L. T. (2023). Testing big data in a big crisis: Nowcasting under covid-19. *International Journal of Forecasting*, (pp. 1548–1563).

Burri, M. (2023). Do daily lead texts help nowcasting gdp growth? *IRENE Working Paper*.

Chapman, B. (2016). Ireland's economy grows 26.3% in 2015 as corporations flock to low tax rate. Retrieved from: https://www.independent.co.uk/news/business/news/ireland-s-economy-grows-26-3-in-2015-as-corporations-flock-to-low-tax-rate-a7133321.html.

Claessens, S. & Kose, M. A. (2023). Recession: When bad times prevail. Retrieved from: https://www.imf.org/external/pubs/ft/fandd/basics/recess.htm.

Dauphin, M. J., Dybczak, M. K., Maneely, M., Sanjani, M. T., Suphaphiphat, M. N., Wang, Y., & Zhang, H. (2022). Nowcasting gdp-a scalable approach using dfm, machine learning and novel data, applied to european economies.

Gourinchas, P.-O. (2023). Resilient global economy still limping along, with growing divergences. Retrieved from: https://www.imf.org/en/Blogs/Articles/2023/10/10/resilient-global-economy-still-limping-along-with-growing-divergences.

Hutto, C. & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the international AAAI conference on web and social media*.

IMF (2023a). What is monetary policy and why is it important? Retrieved from: https://www.imf.org/en/About/Factsheets/Sheets/2023/monetary-policy-and-central-banking.

IMF (2023b). World economic outlook.

Jardet, C. & Meunier, B. (2022). Nowcasting world gdp growth with high-frequency data. *Journal of Forecasting*, (pp. 1181–1200).

Kalamara, E., Turrell, A., Redl, C., Kapetanios, G., & Kapadia, S. (2022). Making text count: economic forecasting using newspaper text. *Journal of Applied Econometrics*, (pp. 896–919).

Kenton, W. (2021). European sovereign debt crisis: Eurozone crisis causes, impacts. Retrieved from: https://www.investopedia.com/terms/e/european-sovereign-debt-crisis.asp.

Makhijani, C. (2020). Ensemble learning techniques. Retrieved from: https://towardsdatascience.com/ensemble-learning-techniques-6346db0c6ef8.

NIMA (2022). Dnb pmi gikk ned i januar. Retrieved from: https://www.nima.no/informasjon/nyheter/vis/?T=DNB

OECD (2023a). Consumer confidence index (cci). Retrieved from: https://data.oecd.org/leadind/consumer-confidence-index-cci.htm.

OECD (2023b). Current account balance. Retrieved from: https://data.oecd.org/trade/current-account-balance.htm.

OECD (2023c). Hours worked. Retrieved from: https://data.oecd.org/emp/hours-worked.htm.

OECD (2023d). Restoring growth. Retrieved from: https://www.oecd.org/economic-outlook/november-2023/.

OECD (2023e). Unemployment rate. Retrieved from: https://data.oecd.org/unemp/unemployment-rate.htm.

of England, B. (2020). How are we monitoring the economy during the covid-19 pandemic? Retrieved from: https://www.bankofengland.co.uk/bank-overground/2020/how-are-we-monitoring-the-economy-during-the-covid-19-pandemic.

Poria, S., Hazarika, D., Majumder, N., & Mihalcea, R. (2020). Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research. *IEEE Transactions on Affective Computing*.

School, I. C. B. (2018). “all models are wrong, but some are useful”. Retrieved from: https://www.imperial.ac.uk/business-school/news/all-models-are-wrong-some-are-useful/.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, (pp. 221–243).

SSB (2023). Nasjonalregnskap. Retrieved from: https://www.ssb.no/statbank/table/11721/.

Stobierski, T. (2021). What is gdp & why is it important? Retrieved from: https://online.hbs.edu/blog/post/why-is-gdp-important.

Thorsrud, L. A. (2016). Nowcasting using news topics. big data versus big bank. *Norges Bank*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

Watson, A. (2023). Leading daily and weekly newspapers in the united kingdom (uk) as of april 2023, by audience reach. Retrieved from: https://www.statista.com/statistics/278645/leading-newspapers-by-reach-in-the-united-kingdom-uk/.

Woloszko, N. (2020). Tracking activity in real time with google trends. *OECD Economics Department Working Papers*.

A. Appendix

Table A.1: Summary of Stock Market Indices and their Bloomberg Tickers.

Country	Stock Market Index	Bloomberg Ticker
Australia	S&P/ASX 200	AS51 Index
Austria	ATX	ATX Index
Belgium	BEL 20	BEL20 Index
Canada	S&P/TSX Composite Index	SPTSX Index
Denmark	OMX Copenhagen 20 Index (OMXC20)	KFX Index
Finland	OMX Helsinki 25 Index (OMXH25)	HEX25 Index
France	CAC 40	CAC Index
Germany	DAX 30	DAX Index
Greece	Athens Stock Exchange (ASE)	ASE Index
Iceland	OMX Iceland All-Share	ICESI Index
Ireland	ISEQ Overall Index	ISEQ Index
Italy	FTSE MIB	FTSEMIB Index
Luxembourg	N/A	LUXXX Index
Netherlands	AEX Index	AEX Index
New Zealand	NZX Market Trade	VONZC Index
Norway	OBX Index	OSEBX Index
Portugal	PSI 20	PSI20 Index
Spain	IBEX 35	IBEX Index
Sweden	OMX Stockholm 30 Index (OMXS30)	SBX Index
Switzerland	Swiss Market Index (SMI)	SMI Index
United Kingdom	FTSE 100	UKX Index
United States	S&P 500	SPX Index

Table A.2: Neural network (NN) candidate configurations, where N is the number of variables.

Model	**Layers**	**Neurons in Input Layer**	**Dropout**
1	2	$2 \times N$	0.1
2	3	$2 \times N + 10$	0.12
3	4	$2 \times N + 20$	0.14
4	5	$2 \times N + 30$	0.16
5	6	$2 \times N + 40$	0.18
6	3	$2 \times N$	0.2
7	4	$2 \times N + 8$	0.22
8	5	$2 \times N + 16$	0.24
9	6	$2 \times N + 24$	0.26
10	7	$2 \times N + 32$	0.28
11	4	$2 \times N$	-
12	5	$2 \times N + 6$	-
13	6	$2 \times N + 12$	-
14	7	$2 \times N + 18$	-
15	8	$2 \times N + 24$	-

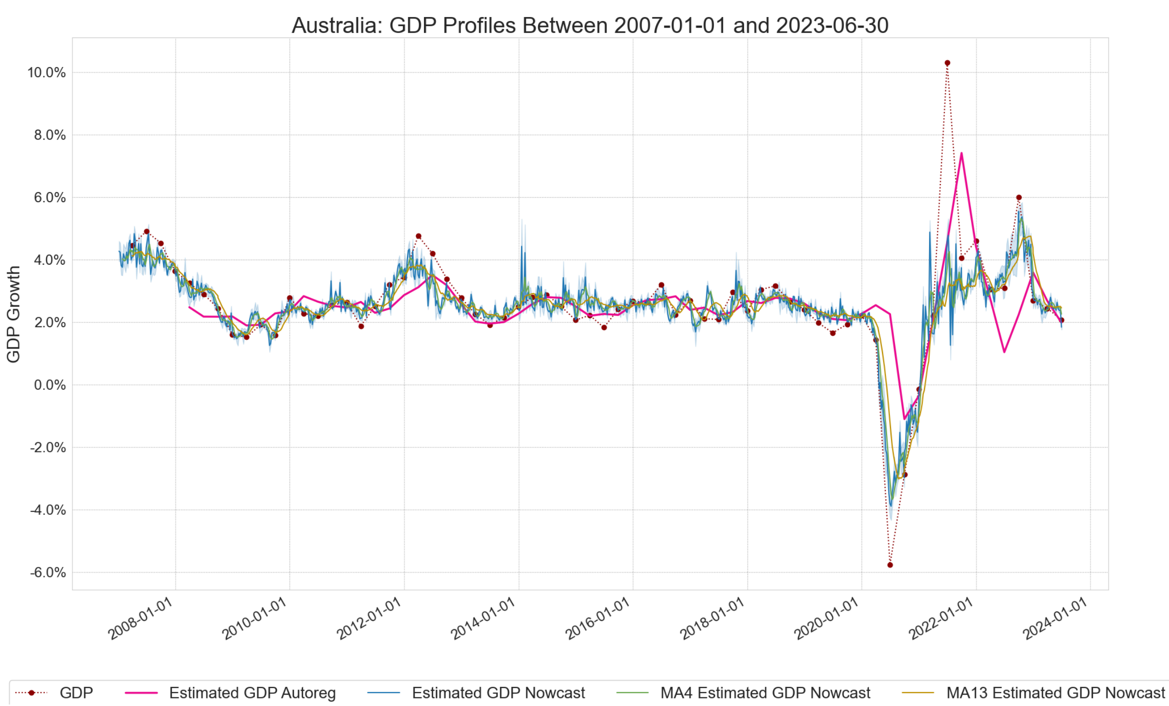

Figure A.1: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Australia.

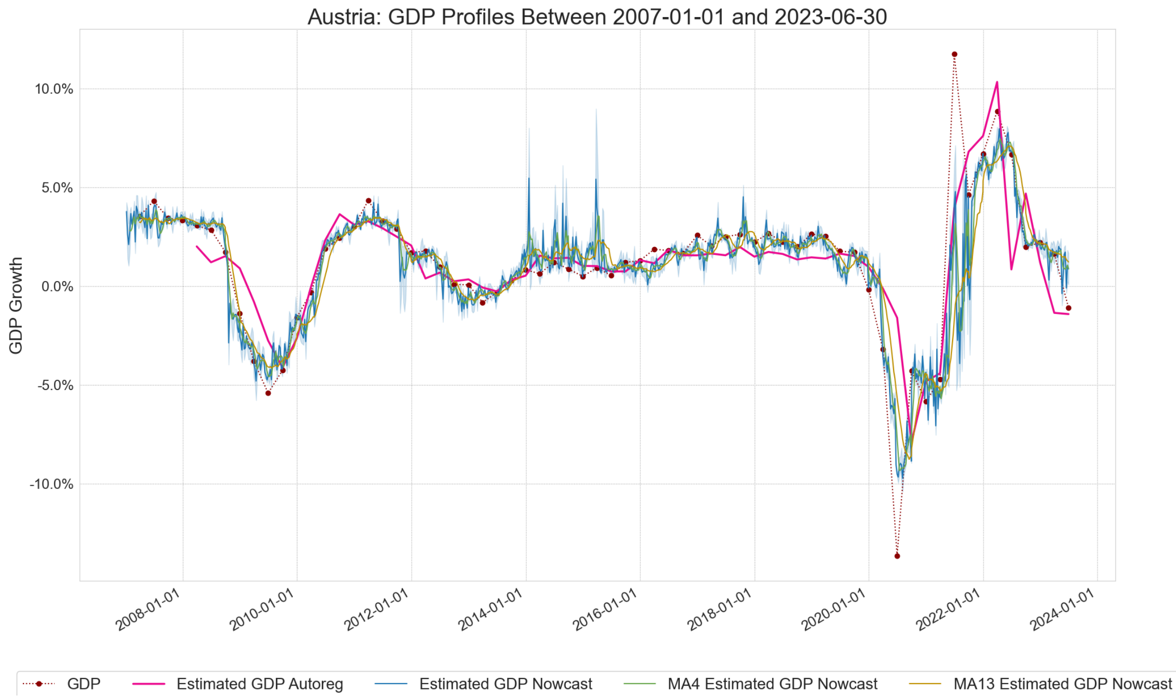

Figure A.2: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Austria.

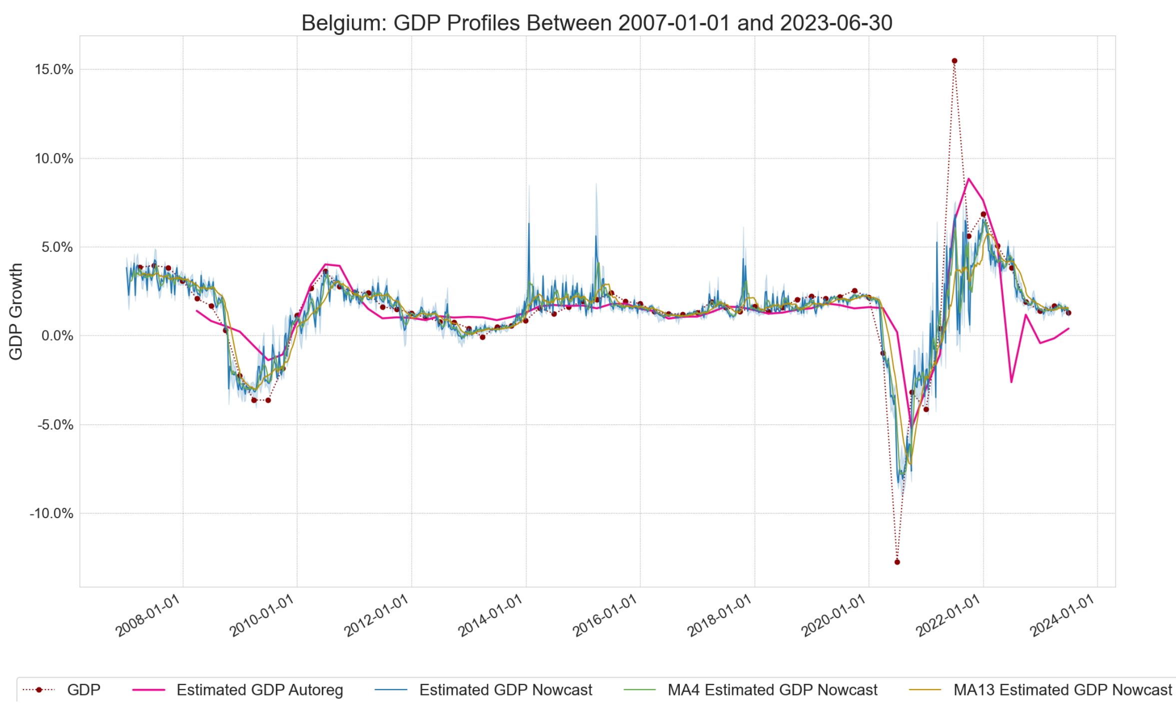

Figure A.3: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Belgium.

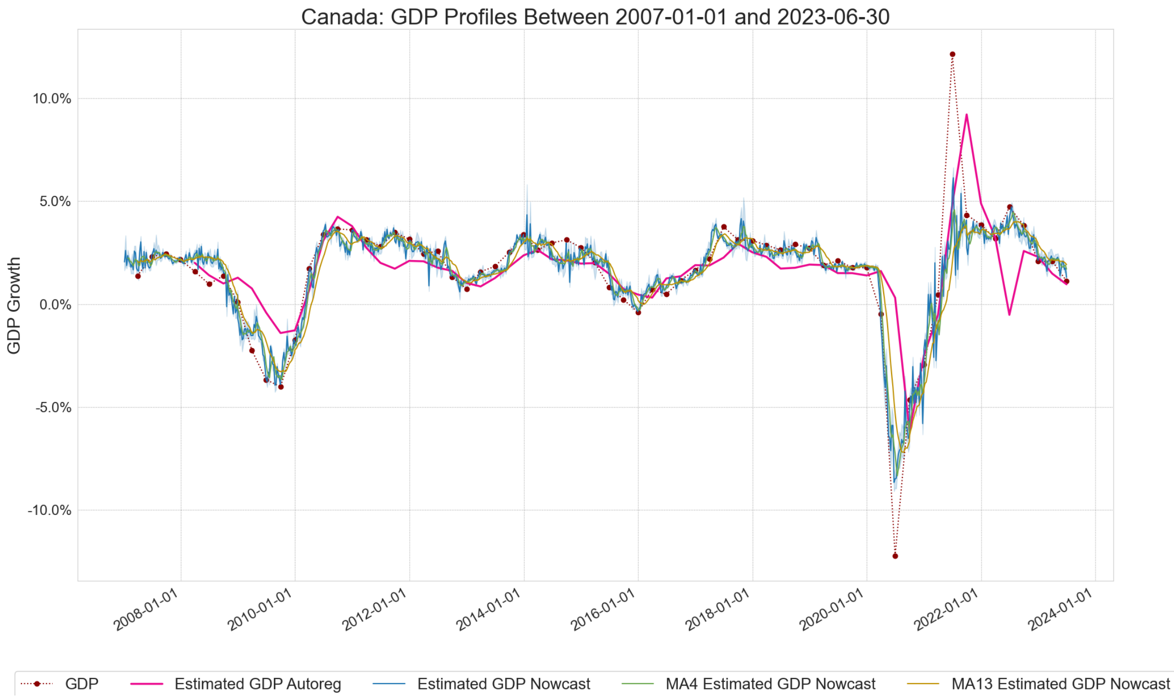

Figure A.4: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Canada.

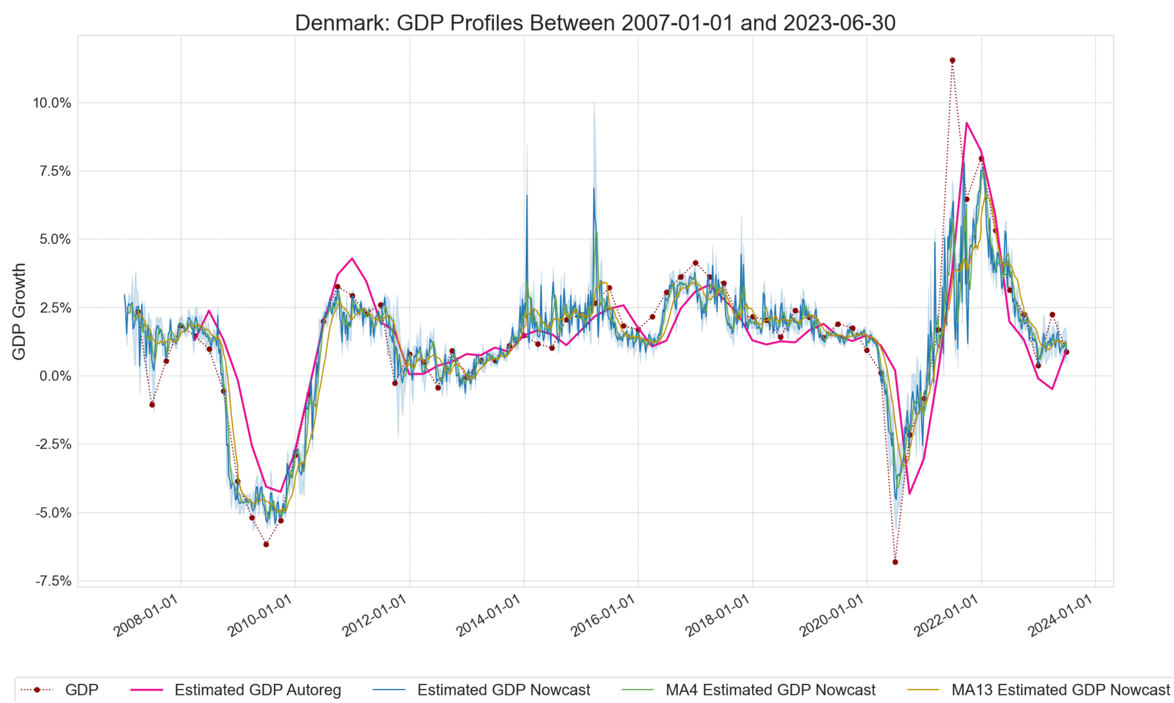

Figure A.5: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Denmark.

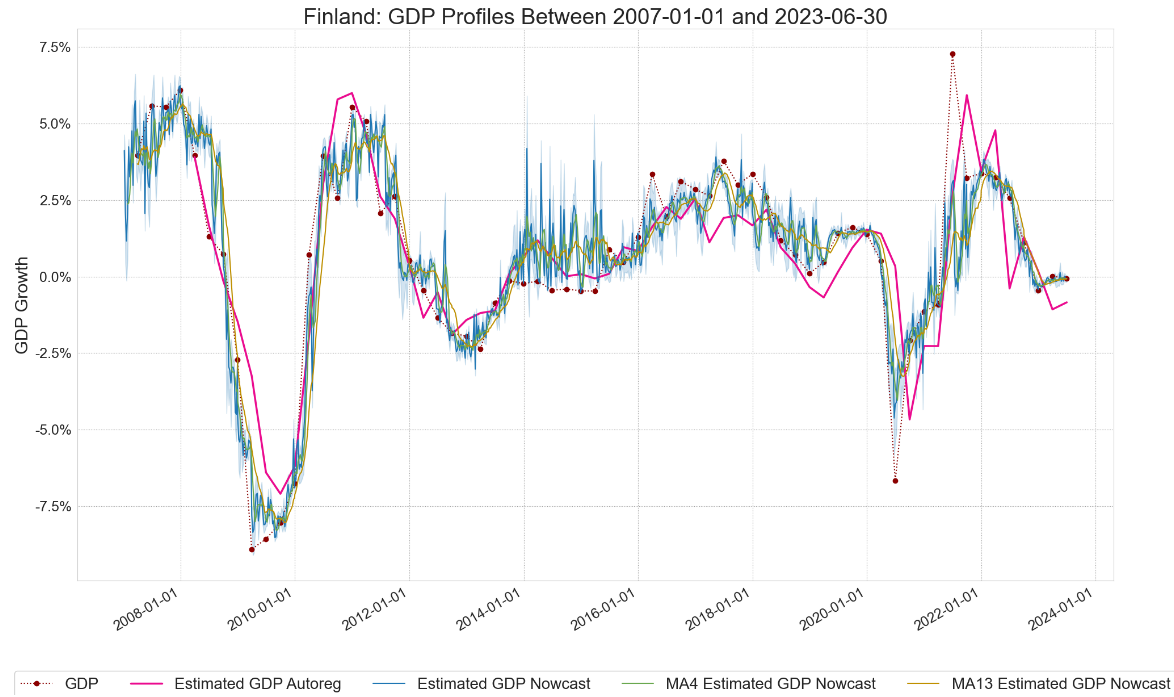

Figure A.6: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Finland.

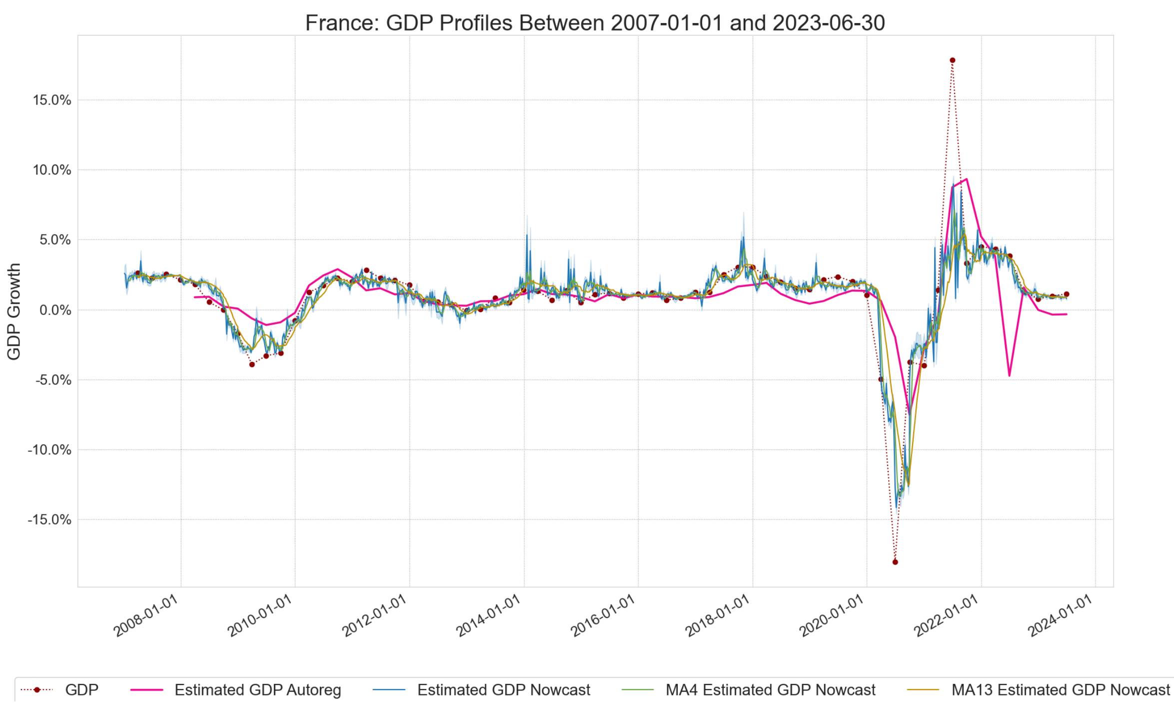

Figure A.7: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, France.

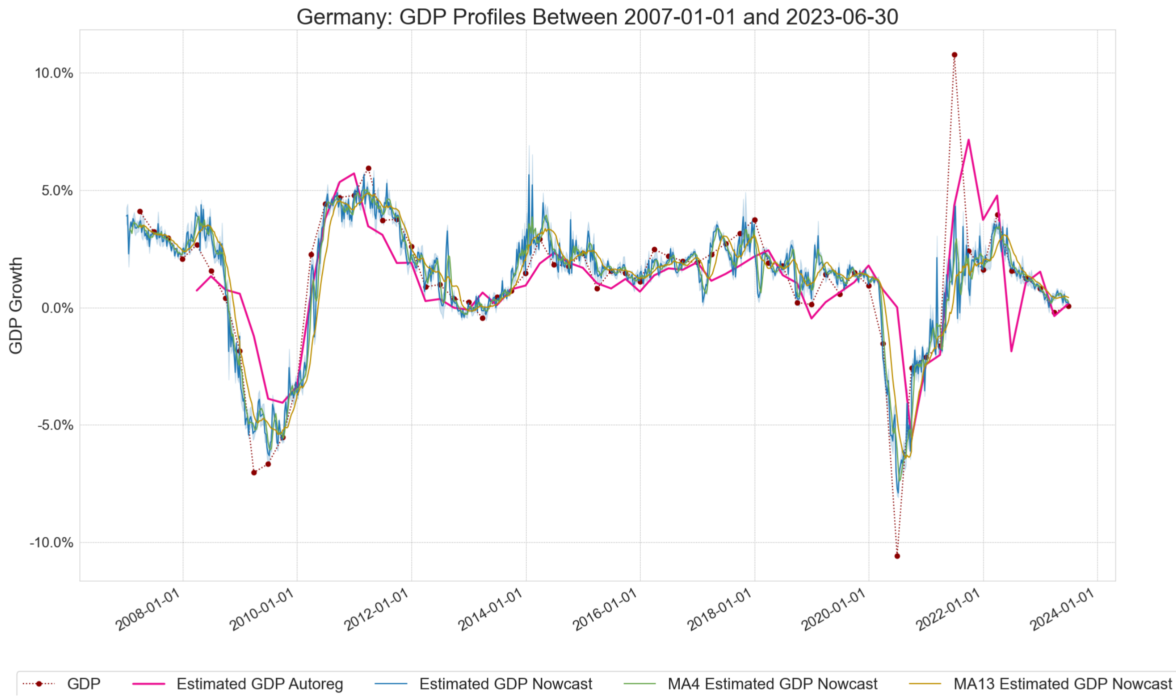

Figure A.8: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Germany.

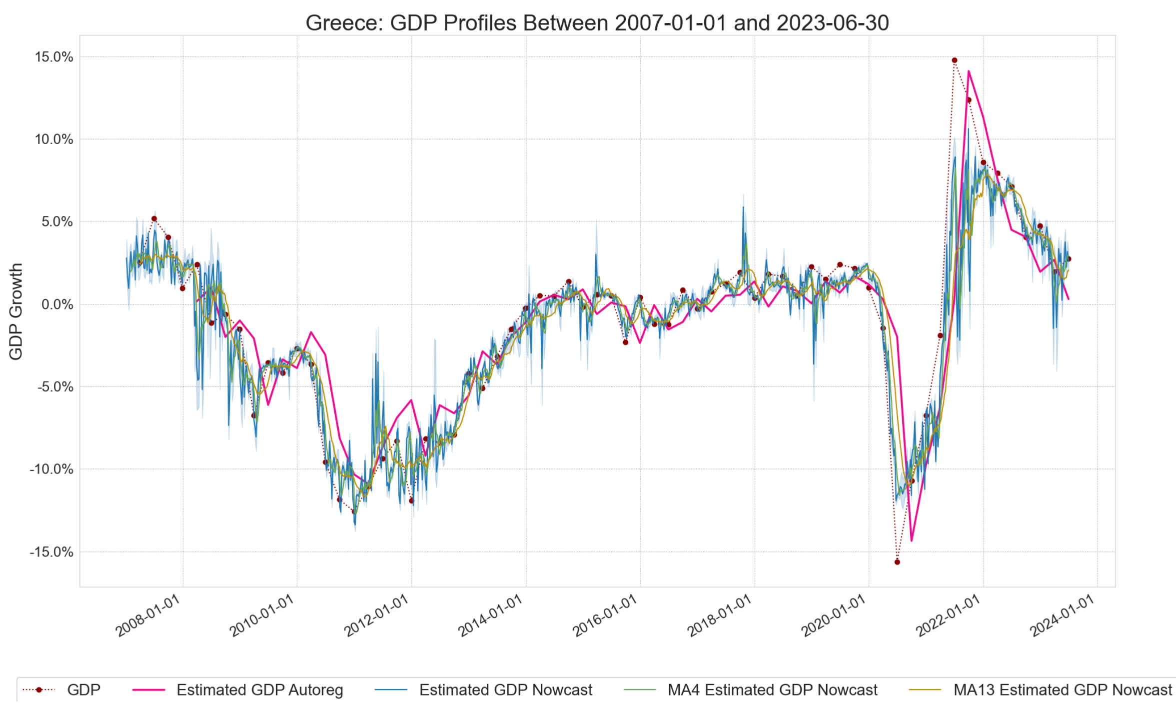

Figure A.9: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Greece.

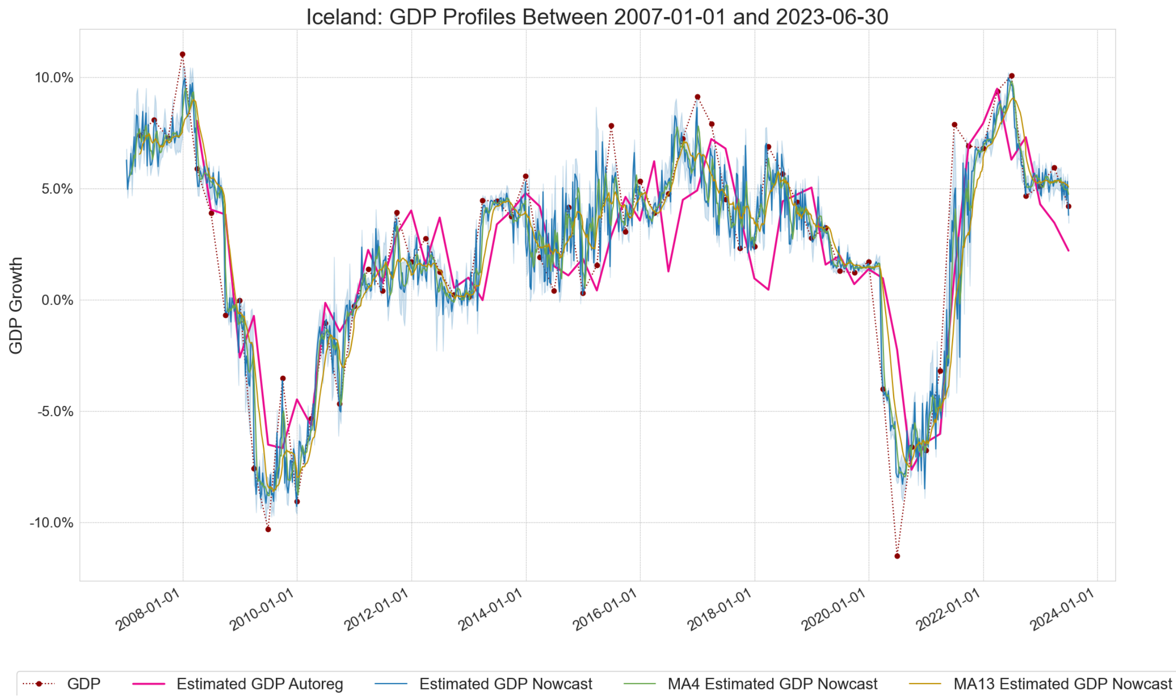

Figure A.10: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Iceland.

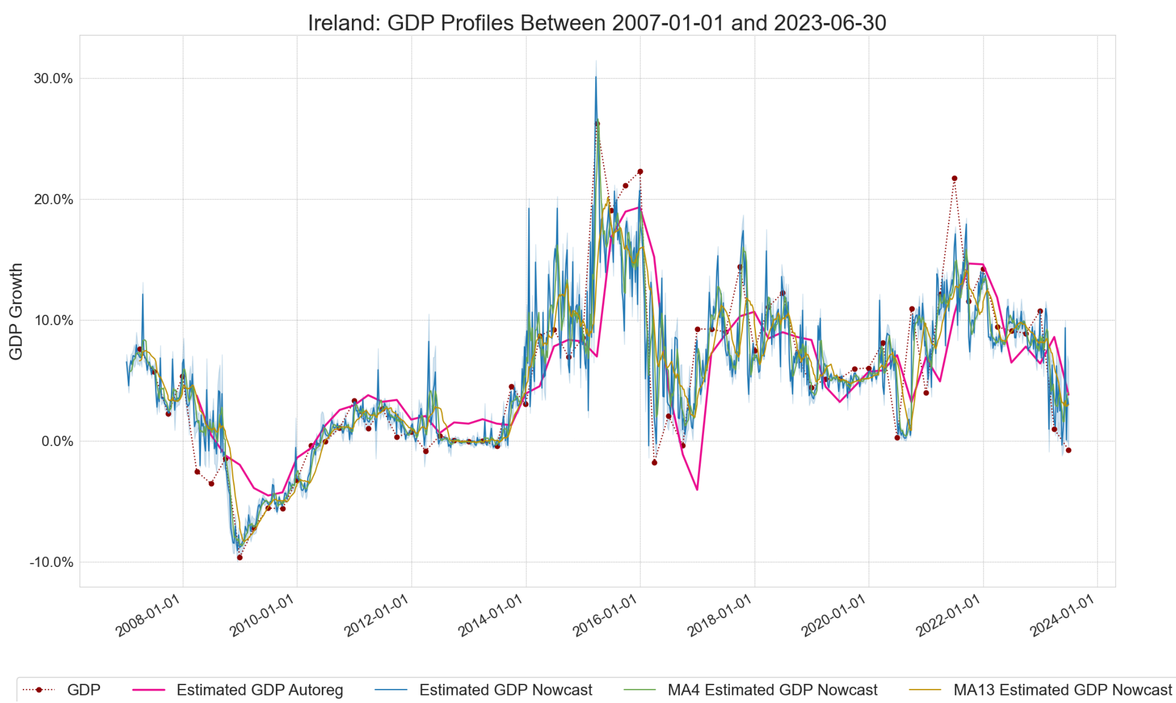

Figure A.11: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Ireland.

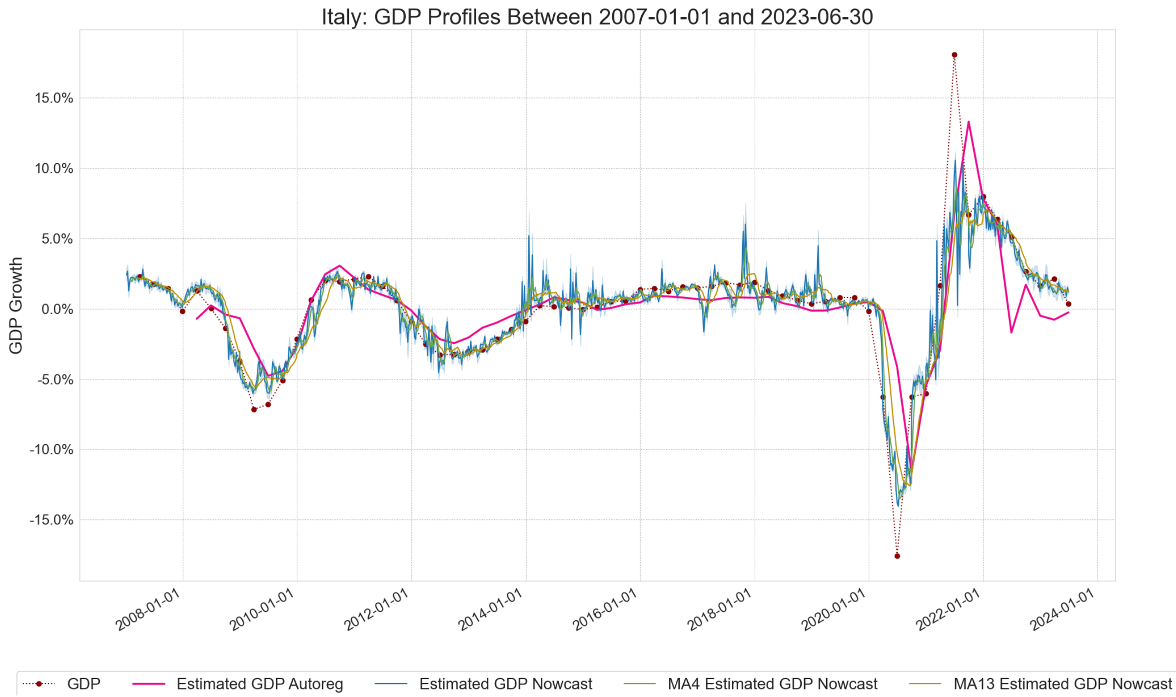

Figure A.12: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Italy.

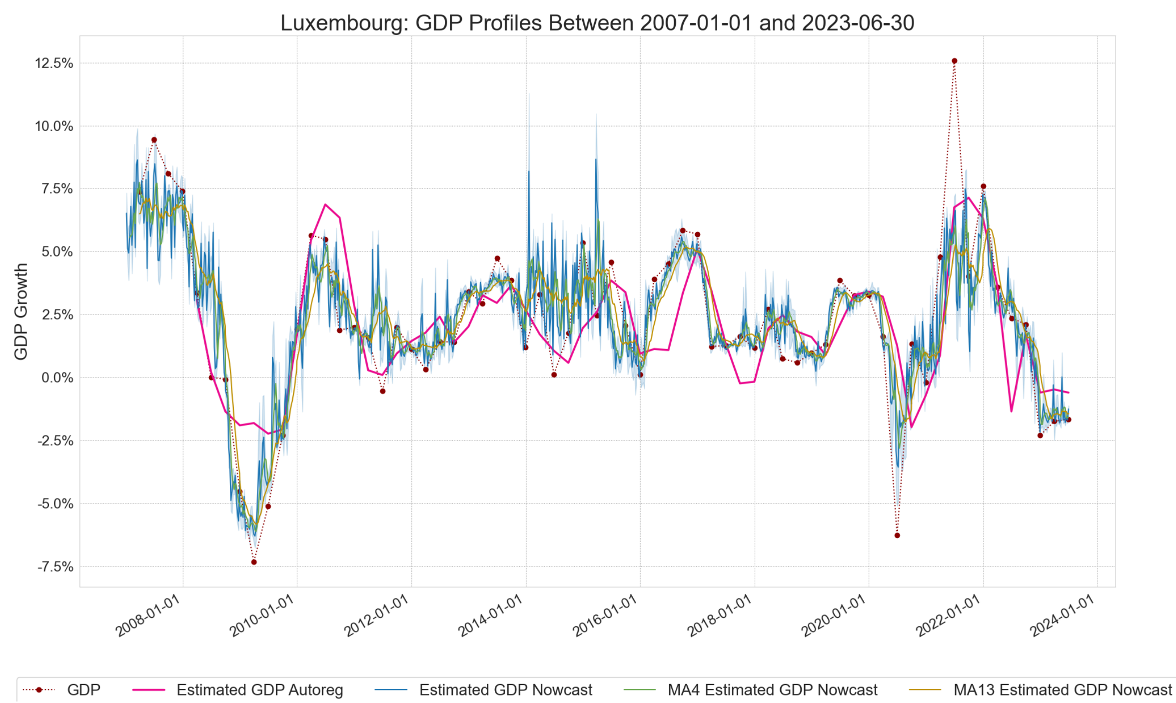

Figure A.13: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Luxembourg.

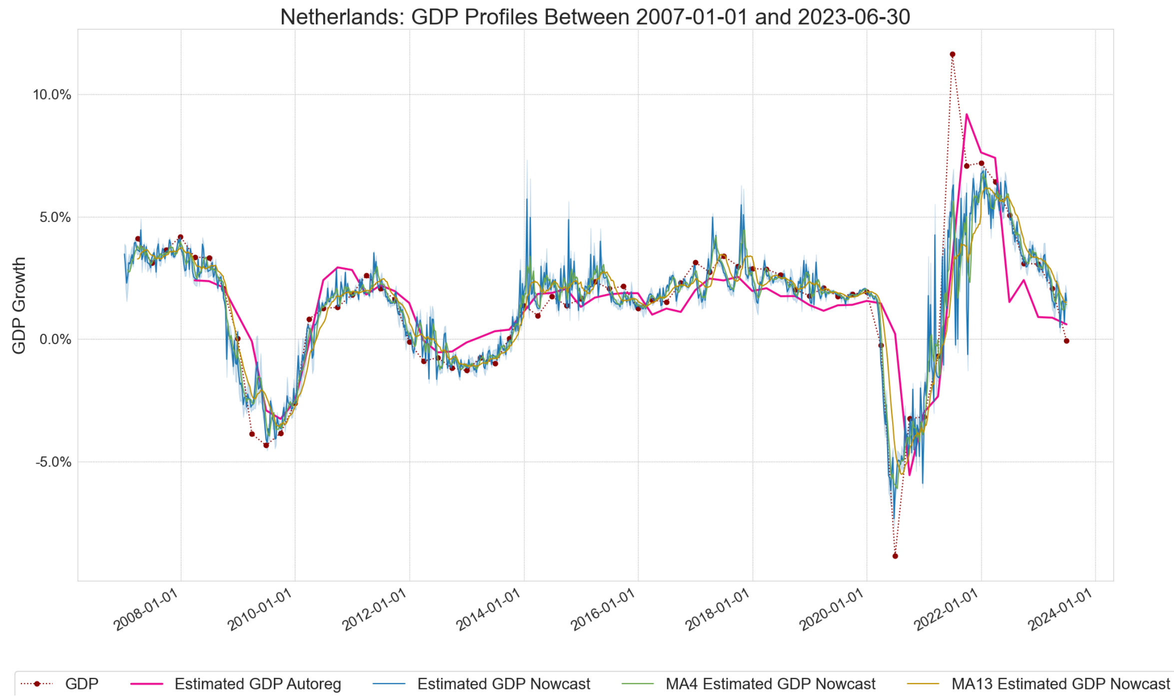

Figure A.14: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Netherlands.

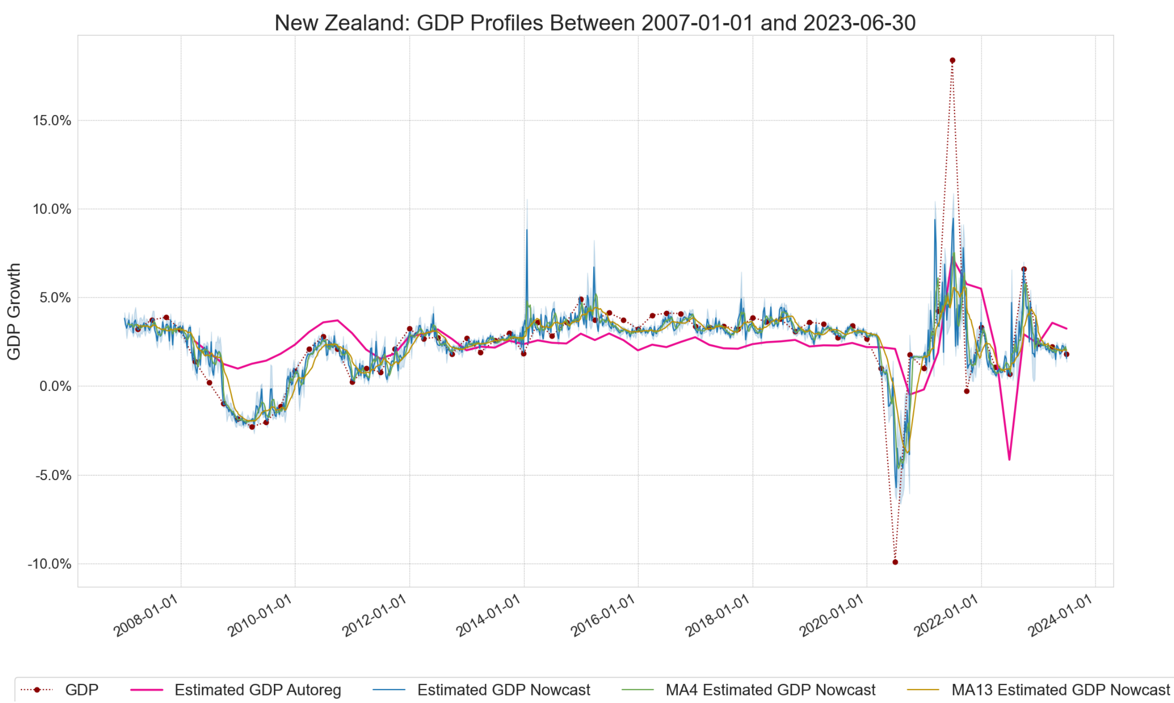

Figure A.15: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, New Zealand.

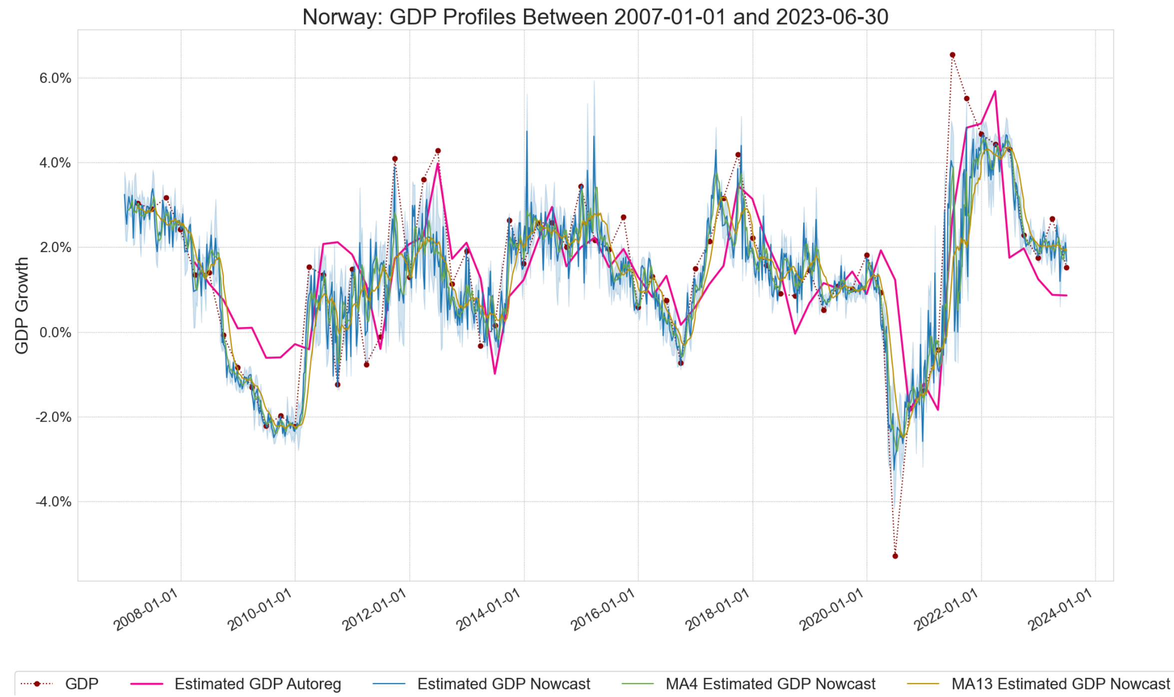

Figure A.16: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Norway.

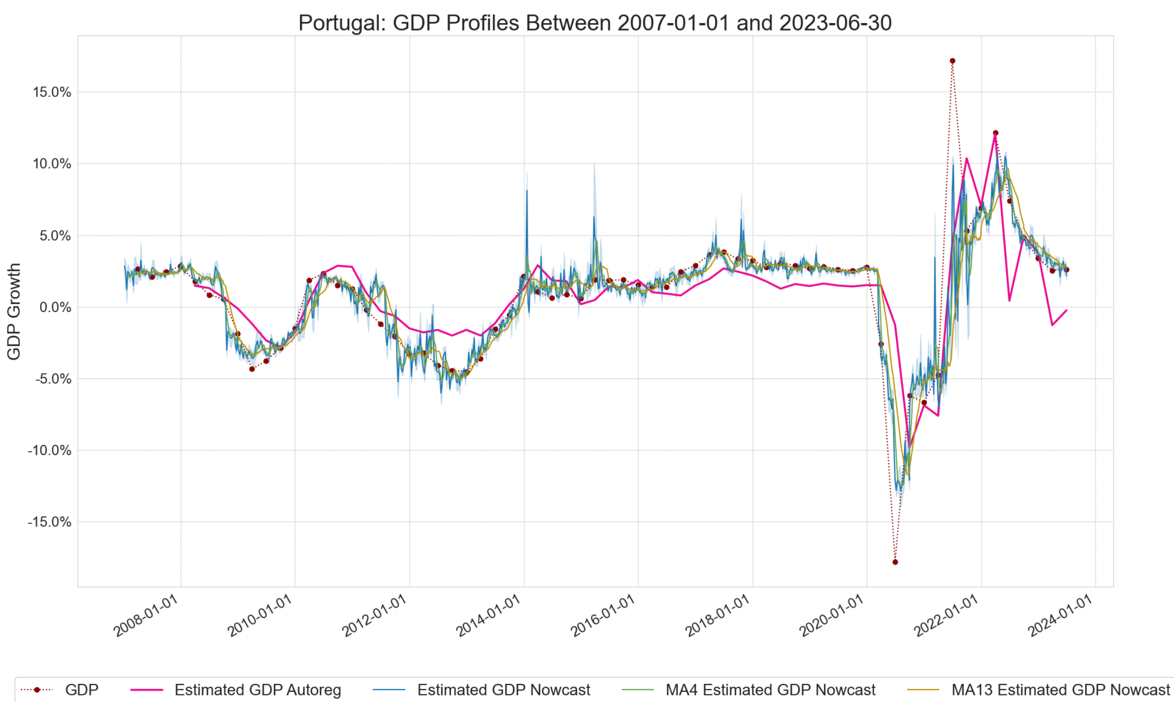

Figure A.17: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Portugal.

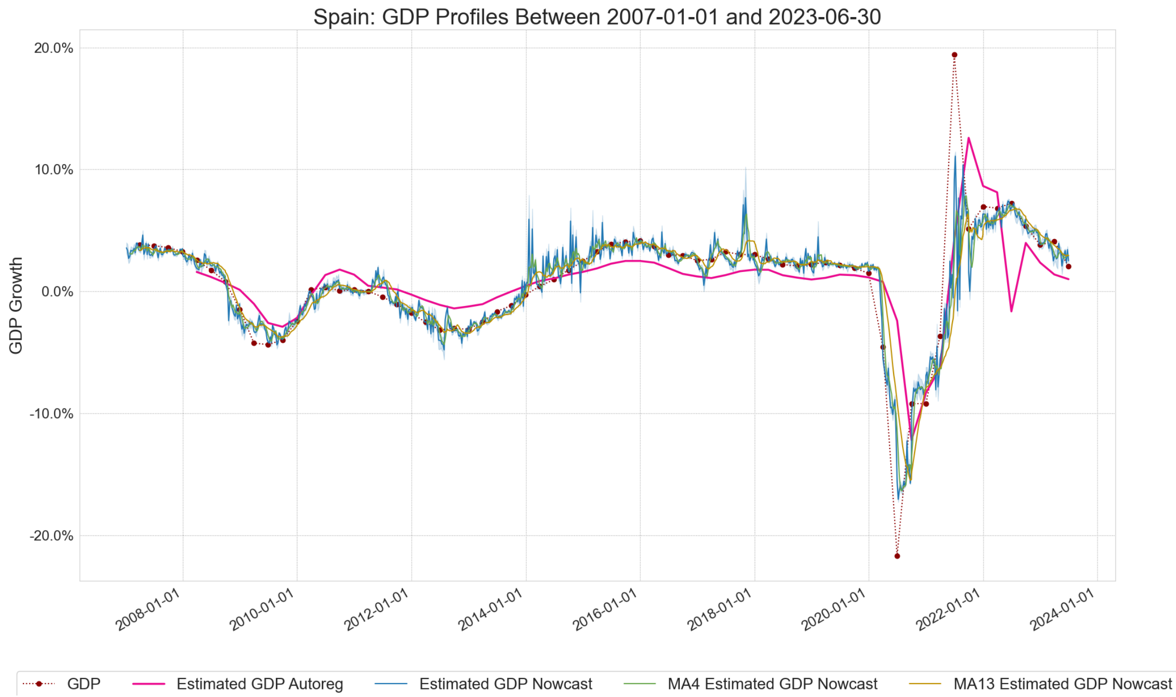

Figure A.18: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Spain.

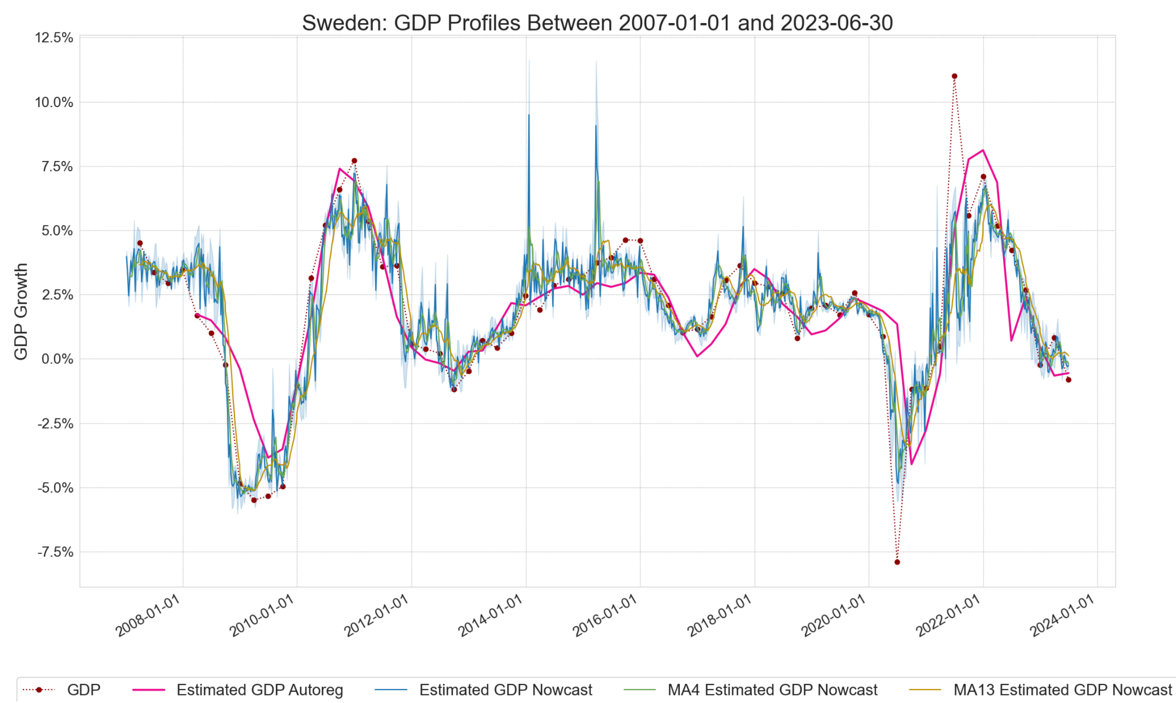

Figure A.19: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Sweden.

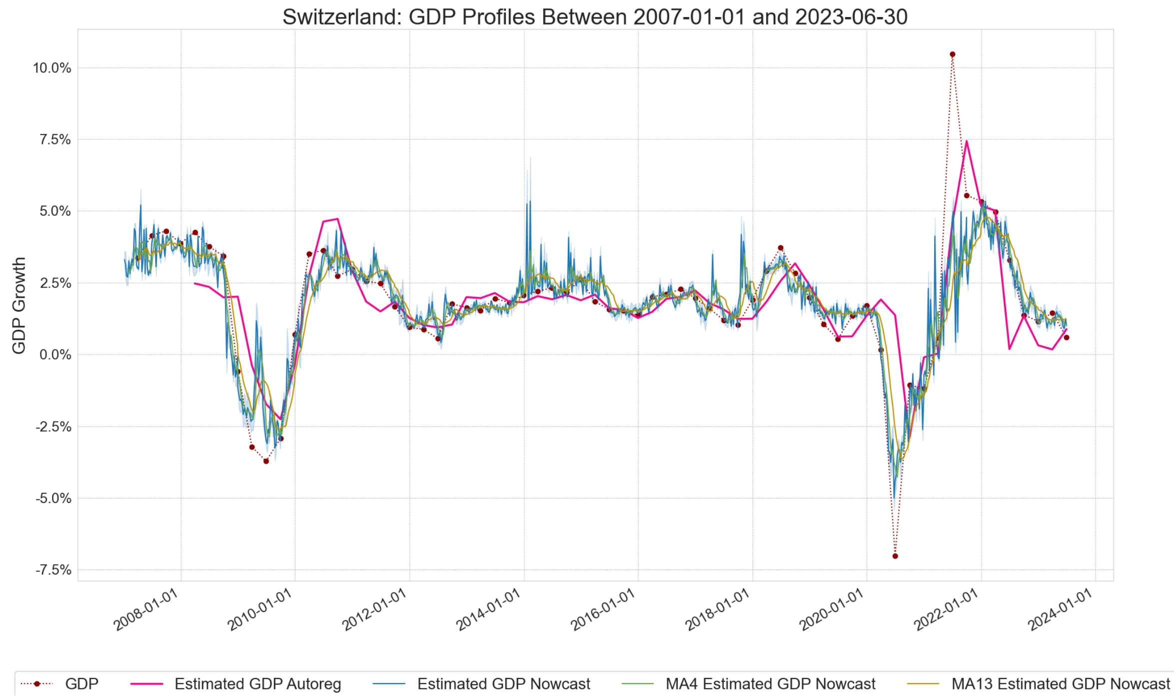

Figure A.20: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, Switzerland.

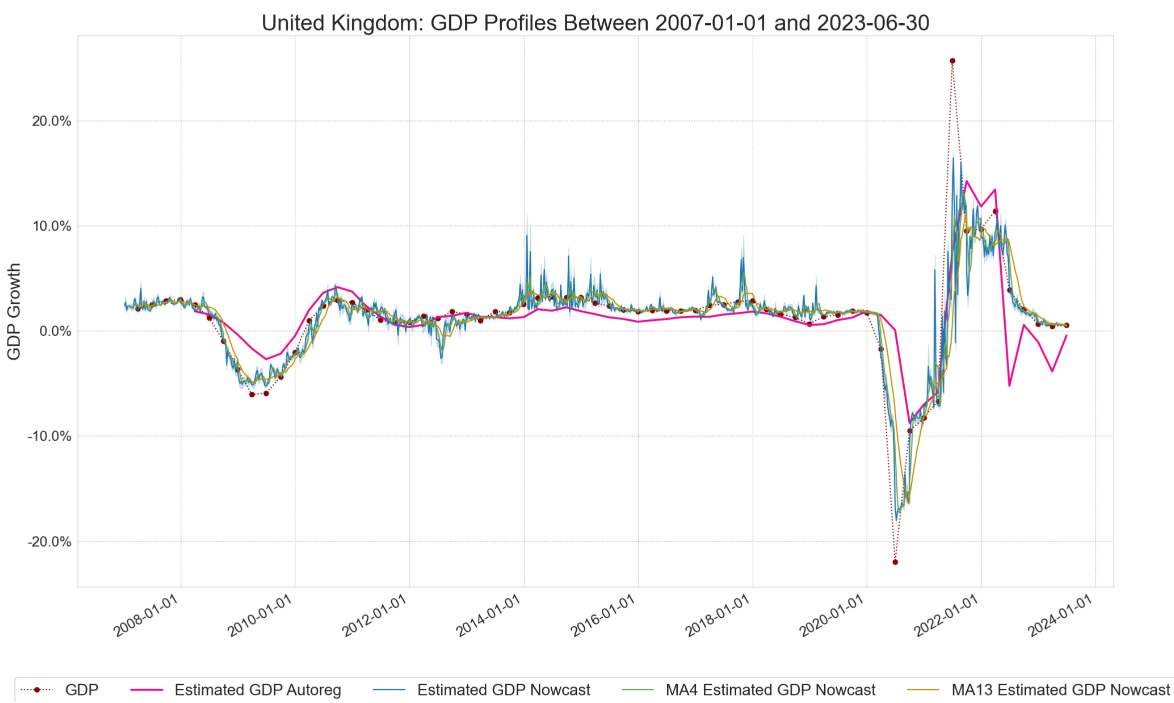

Figure A.21: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, United Kingdom.

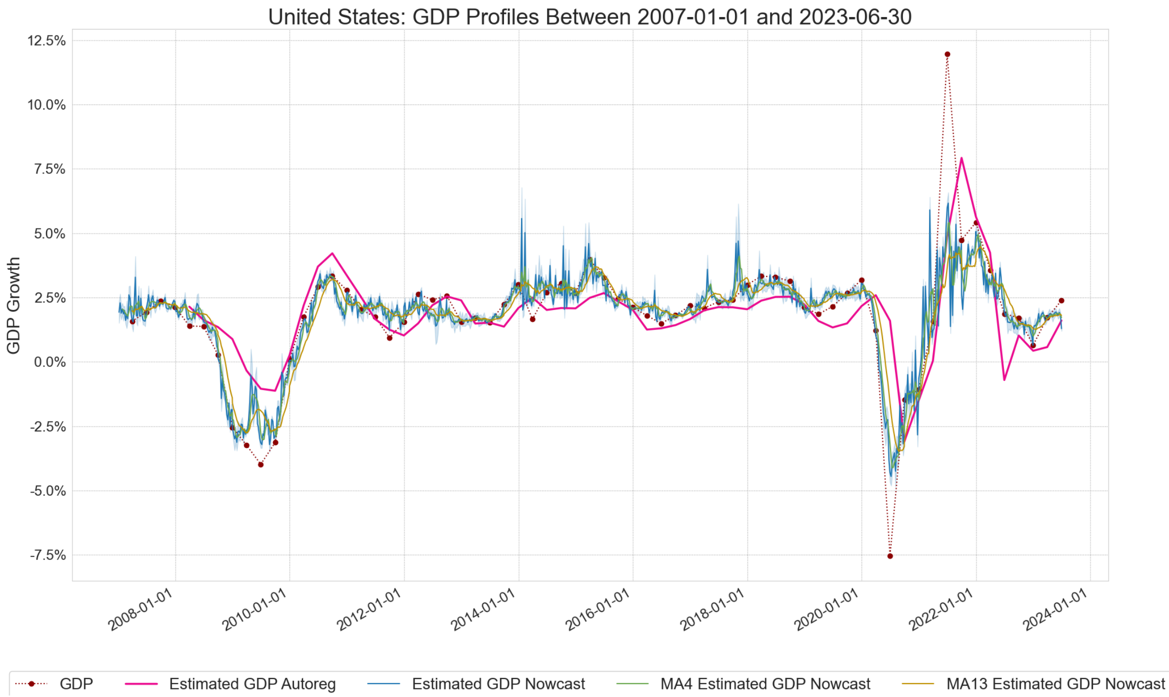

Figure A.22: AR(4) predictions and nowcasts with monthly- and quarterly smoothed lines, United States.